Econometric Analysis of Panel Data

Badi H. Baltagi

Third Edition

Econometric Analysis of Panel Data

Badi H. Baltagi

Badi H. Baltagi earned his PhD in Economics at the University of Pennsylvania in 1979. He joined the faculty at Texas A&M University in 1988, having served previously on the faculty at the University of Houston. He is the author of *Econometric Analysis of Panel Data* and *Econometrics*, and editor of *A Companion to Theoretical Econometrics; Recent Developments in the Econometrics of Panel Data, Volumes I and II; Nonstationary Panels, Panel Cointegration, and Dynamic Panels*; and author or co-author of over 100 publications, all in leading economics and statistics journals. Professor Baltagi is the holder of the George Summey, Jr. Professor Chair in Liberal Arts and was awarded the Distinguished Achievement Award in Research. He is co-editor of *Empirical Economics*, and associate editor of *Journal of Econometrics* and *Econometric Reviews*. He is the replication editor of the *Journal of Applied Econometrics* and the series editor for *Contributions to Economic Analysis.* He is a fellow of the *Journal of Econometrics* and a recipient of the Plura Scripsit Award from *Econometric Theory*.

Econometric Analysis of Panel Data

Third edition

Badi H. Baltagi

John Wiley & Sons, Ltd

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wileyeurope.com or www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Cataloging-in-Publication Data

Baltagi, Badi H. (Badi Hani)
Econometric analysis of panel data / Badi H. Baltagi. — 3rd ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-470-01456-3 (pbk. : alk. paper)
1. Econometrics. 2. Panel analysis. I. Title.
HB139.B35 2005
330′.01′5195–dc22 2005006840

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN-13 978-0-470-01456-1
ISBN-10 0-470-01456-3

Typeset in 10/12pt Times by TechBooks, New Delhi, India
Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire

To My Wife, Phyllis

Contents

Preface

This book is intended for a graduate econometrics course on panel data. The prerequisites include a good background in mathematical statistics and econometrics at the level of Greene (2003). Matrix presentations are necessary for this topic.

Some of the major features of this book are that it provides an *up-to-date* coverage of panel data techniques, especially for serial correlation, spatial correlation, heteroskedasticity, seemingly unrelated regressions, simultaneous equations, dynamic models, incomplete panels, limited dependent variables and nonstationary panels. I have tried to keep things *simple*, illustrating the basic ideas using the *same notation* for a diverse literature with heterogeneous notation. Many of the estimation and testing techniques are illustrated with *data sets* which are available for classroom use on the Wiley web site (www.wiley.com/go/baltagi3e). The book also cites and summarizes several empirical studies using panel data techniques, so that the reader can relate the econometric methods with the economic applications. The book proceeds from single equation methods to simultaneous equation methods as in any standard econometrics text, so it should prove friendly to graduate students.

The book gives the basic coverage without being encyclopedic. There is an extensive amount of research in this area and not all topics are covered. The first conference on panel data was held in Paris more than 25 years ago, and this resulted in two volumes of the *Annales de l'INSEE* edited by Mazodier (1978). Since then, there have been eleven international conferences on panel data, the last one at Texas A&M University, College Station, Texas, June 2004.

In undertaking this revision, I benefited from teaching short panel data courses at the University of California-San Diego (2002); International Monetary Fund (IMF), Washington, DC (2004, 2005); University of Arizona (1996); University of Cincinnati (2004); Institute for Advanced Studies, Vienna (2001); University of Innsbruck (2002); Universidad del of Rosario, Bogotá (2003); Seoul National University (2002); Centro Interuniversitario de Econometria (CIDE)-Bertinoro (1998); Tor Vergata University-Rome (2002); Institute for Economic Research (IWH)-Halle (1997); European Central Bank, Frankfurt (2001); University of Mannheim (2002); Center for Economic Studies (CES-Ifo), Munich (2002); German Institute for Economic Research (DIW), Berlin (2004); University of Paris II, Pantheon (2000); International Modeling Conference on the Asia-Pacific Economy, Cairns, Australia (1996). The third edition, like the second, continues to use more empirical examples from the panel data literature to motivate the book. All proofs given in the appendices of the first edition have been deleted. There are worked out examples using Stata and EViews. The data sets as well as the output and programs to implement the estimation and testing procedures described in the book are provided on the Wiley web site (www.wiley.com/go/baltagi3e). Additional exercises have been added and solutions to selected exercises are provided on the Wiley web site. Problems and solutions published in *Econometric Theory* and used in this book are not given in the references, as in the previous editions, to save space. These can easily be traced to their source in the journal. For example, when the book refers to problem 99.4.3, this can be found in *Econometric Theory*, in the year 1999, issue 4, problem 3.

Several chapters have been revised and in some cases shortened or expanded upon. More specifically, Chapter 1 has been updated with web site addresses for panel data sources as well as more motivation for why one should use panel data. Chapters 2, 3 and 4 have empirical studies illustrated with Stata and EViews output. The material on heteroskedasticity in Chapter 5 is completely revised and updated with recent estimation and testing results. The material on serial correlation is illustrated with Stata and TSP. A simultaneous equation example using crime data is added to Chapter 7 and illustrated with Stata. The Hausman and Taylor method is also illustrated with Stata using PSID data to estimate an earnings equation. Chapter 8 updates the dynamic panel data literature using newly published papers and illustrates the estimation methods using a dynamic demand for cigarettes. Chapter 9 now includes Stata output on estimating a hedonic housing equation using unbalanced panel data. Chapter 10 has an update on spatial panels as well as heterogeneous panels. Chapter 11 updates the limited dependent variable panel data models with recent papers on the subject and adds an application on estimating nurses' labor supply in Norway. Chapter 12 on nonstationary panels is completely rewritten. The literature has continued to explode, with several theoretical results as well as influential empirical papers appearing in this period. An empirical illustration on purchasing power parity is added and illustrated with EViews. A new section surveys the literature on panel unit root tests allowing for cross-section correlation.

I would like to thank my co-authors for allowing me to draw freely on our joint work. In particular, I would like to thank Jan Askildsen, Georges Bresson, Young-Jae Chang, Peter Egger, Jim Griffin, Tor Helge Holmas, Chihwa Kao, Walter Krämer, Dan Levin, Dong Li, Qi Li, Michael Pfaffermayr, Nat Pinnoi, Alain Pirotte, Dan Rich, Seuck Heun Song and Ping Wu. Many colleagues who had direct and indirect influence on the contents of this book include Luc Anselin, George Battese, Anil Bera, Richard Blundell, Trevor Breusch, Chris Cornwell, Bill Griffiths, Cheng Hsiao, Max King, Kajal Lahiri, G.S. Maddala, Roberto Mariano, László Mátyás, Chiara Osbat, M. Hashem Pesaran, Peter C.B. Phillips, Peter Schmidt, Patrick Sevestre, Robin Sickles, Marno Verbeek, Tom Wansbeek and Arnold Zellner. Clint Cummins provided benchmark results for the examples in this book using TSP. David Drukker provided help with Stata on the Hausman and Taylor procedure as well as EC2SLS in Chapter 7. Also, the Baltagi and Wu LBI test in Chapter 9. Glenn Sueyoshi provided help with EViews on the panel unit root tests in Chapter 12. Thanks also go to Steve Hardman and Rachel Goodyear at Wiley for their efficient and professional editorial help, Teri Tenalio who typed numerous revisions of this book and my wife Phyllis whose encouragement and support gave me the required energy to complete this book. Responsibilities for errors and omissions are my own.

1 Introduction

1.1 PANEL DATA: SOME EXAMPLES

In this book, the term "panel data" refers to the pooling of observations on a cross-section of households, countries, firms, etc. over several time periods. This can be achieved by surveying a number of households or individuals and following them over time. Two well-known examples of US panel data are the Panel Study of Income Dynamics (PSID) collected by the Institute for Social Research at the University of Michigan (http://psidonline.isr.umich.edu) and the National Longitudinal Surveys (NLS) which is a set of surveys sponsored by the Bureau of Labor Statistics (http://www.bls.gov/nls/home.htm).

The PSID began in 1968 with 4800 families and has grown to more than 7000 families in 2001. By 2003, the PSID had collected information on more than 65 000 individuals spanning as much as 36 years of their lives. Annual interviews were conducted from 1968 to 1996. In 1997, this survey was redesigned for biennial data collection. In addition, the core sample was reduced and a refresher sample of post-1968 immigrant families and their adult children was introduced. The central focus of the data is economic and demographic. The list of variables include income, poverty status, public assistance in the form of food or housing, other financial matters (e.g. taxes, interhousehold transfers), family structure and demographic measures, labor market work, housework time, housing, geographic mobility, socioeconomic background and health. Other supplemental topics include housing and neighborhood characteristics, achievement motivation, child care, child support and child development, job training and job acquisition, retirement plans, health, kinship, wealth, education, military combat experience, risk tolerance, immigration history and time use.

The NLS, on the other hand, are a set of surveys designed to gather information at multiple points in time on labor market activities and other significant life events of several groups of men and women:

(1) The NLSY97 consists of a nationally representative sample of approximately 9000 youths who were 12–16 years old as of 1997. The NLSY97 is designed to document the transition from school to work and into adulthood. It collects extensive information about youths' labor market behavior and educational experiences over time.
(2) The NLSY79 consists of a nationally representative sample of 12 686 young men and women who were 14–24 years old in 1979. These individuals were interviewed annually through 1994 and are currently interviewed on a biennial basis.
(3) The NLSY79 children and young adults. This includes the biological children born to women in the NLSY79.
(4) The NLS of mature women and young women: these include a group of 5083 women who were between the ages of 30 and 44 in 1967. Also, 5159 women who were between the ages of 14 and 24 in 1968. Respondents in these cohorts continue to be interviewed on a biennial basis.

(5) The NLS of older men and young men: these include a group of 5020 men who were between the ages of 45 and 59 in 1966. Also, a group of 5225 men who were between the ages of 14 and 24 in 1966. Interviews for these two cohorts ceased in 1981.

The list of variables include information on schooling and career transitions, marriage and fertility, training investments, child care usage and drug and alcohol use. A large number of studies have used the NLS and PSID data sets. Labor journals in particular have numerous applications of these panels. Klevmarken (1989) cites a bibliography of 600 published articles and monographs that used the PSID data sets. These cover a wide range of topics including labor supply, earnings, family economic status and effects of transfer income programs, family composition changes, residential mobility, food consumption and housing.

Panels can also be constructed from the Current Population Survey (CPS), a monthly national household survey of about 50 000 households conducted by the Bureau of Census for the Bureau of Labor Statistics (http://www.bls.census.gov/cps/). This survey has been conducted for more than 50 years. Compared with the NLS and PSID data, the CPS contains fewer variables, spans a shorter period and does not follow movers. However, it covers a much larger sample and is representative of all demographic groups.

Although the US panels started in the 1960s, it was only in the 1980s that the European panels began setting up. In 1989, a special section of the *European Economic Review* published papers using the German Socio-Economic Panel (see Hujer and Schneider, 1989), the Swedish study of household market and nonmarket activities (see Björklund, 1989) and the Intomart Dutch panel of households (see Alessie, Kapteyn and Melenberg, 1989). The first wave of the German Socio-Economic Panel (GSOEP) was collected by the DIW (German Institute for Economic Research, Berlin) in 1984 and included 5921 West German households (www.diw.de/soep). This included 12 290 respondents. Standard demographic variables as well as wages, income, benefit payments, level of satisfaction with various aspects of life, hopes and fears, political involvement, etc. are collected. In 1990, 4453 adult respondents in 2179 households from East Germany were included in the GSOEP due to German unification. The attrition rate has been relatively low in GSOEP. Wagner, Burkhauser and Behringer (1993) report that through eight waves of the GSOEP, 54.9% of the original panel respondents have records without missing years. An inventory of national studies using panel data is given at (http://psidonline.isr.umich.edu/Guide/PanelStudies.aspx). These include the Belgian Socio-economic Panel (www.ufsia.ac.be/CSB/sep_nl.htm) which interviews a representative sample of 6471 Belgian households in 1985, 3800 in 1988 and 3800 in 1992 (including a new sample of 900 households). Also, 4632 households in 1997 (including a new sample of 2375 households). The British Household Panel Survey (BHPS) which is an annual survey of private households in Britain first collected in 1991 by the Institute for Social and Economic Research at the University of Essex (www.irc.essex.ac.uk/bhps). This is a national representative sample of some 5500 households and 10 300 individuals drawn from 250 areas of Great Britain. Data collected includes demographic and household characteristics, household organization, labor market, health, education, housing, consumption and income, social and political values. The Swiss Household Panel (SHP) whose first wave in 1999 interviewed 5074 households comprising 7799 individuals (www.unine.ch/psm). The Luxembourg Panel Socio-Economique "Liewen zu Letzebuerg" (PSELL I) (1985–94) is based on a representative sample of 2012 households and 6110 individuals. In 1994, the PSELL II expanded to 2978 households and 8232 individuals. The Swedish Panel Study Market and Non-market Activities (HUS) were collected in 1984, 1986, 1988, 1991, 1993, 1996 and 1998 (http://www.nek.uu.se/faculty/klevmark/hus.htm).

Data for 2619 individuals were collected on child care, housing, market work, income and wealth, tax reform (1993), willingness to pay for a good environment (1996), local taxes, public services and activities in the black economy (1998).

The European Community Household Panel (ECHP) is centrally designed and coordinated by the Statistical Office of the European Communities (EuroStat), see Peracchi (2002). The first wave was conducted in 1994 and included all current members of the EU except Austria, Finland and Sweden. Austria joined in 1995, Finland in 1996 and data for Sweden was obtained from the Swedish Living Conditions Survey. The project was launched to obtain comparable information across member countries on income, work and employment, poverty and social exclusion, housing, health, and many other diverse social indicators indicating living conditions of private households and persons. The EHCP was linked from the beginning to existing national panels (e.g. Belgium and Holland) or ran parallel to existing panels with similar content, namely GSOEP, PSELL and the BHPS. This survey ran from 1994 to 2001 (http://epunet.essex.ac.uk/echp.php).

Other panel studies include: the Canadian Survey of Labor Income Dynamics (SLID) collected by Statistics Canada (www.statcan.ca) which includes a sample of approximately 35 000 households located throughout all ten provinces. Years available are 1993–2000. The Japanese Panel Survey on Consumers (JPSC) collected in 1994 by the Institute for Research on Household Economics (www.kakeiken.or.jp). This is a national representative sample of 1500 women aged 24 and 34 years in 1993 (cohort A). In 1997, 500 women were added with ages between 24 and 27 (cohort B). Information gathered includes family composition, labor market behavior, income, consumption, savings, assets, liabilities, housing, consumer durables, household management, time use and satisfaction. The Russian Longitudinal Monitoring Survey (RLMS) collected in 1992 by the Carolina Population Center at the University of North Carolina (www.cpc.unc.edu/projects/rlms/home.html). The RLMS is a nationally representative household survey designed to measure the effects of Russian reforms on economic well-being. Data includes individual health and dietary intake, measurement of expenditures and service utilization and community level data including region-specific prices and community infrastructure. The Korea Labor and Income Panel Study (KLIPS) available for 1998–2001 surveys 5000 households and their members from seven metropolitan cities and urban areas in eight provinces (http://www.kli.re.kr/klips). The Household, Income and Labor Dynamics in Australia (HILDA) is a household panel survey whose first wave was conducted by the Melbourne Institute of Applied Economic and Social Research in 2001 (http://www.melbourneinstitute.com/hilda). This includes 7682 households with 13 969 members from 488 different neighboring regions across Australia. The Indonesia Family Life Survey (http://www.rand.org/FLS/IFLS) is available for 1993/94, 1997/98 and 2000. In 1993, this surveyed 7224 households living in 13 of the 26 provinces of Indonesia.

This list of panel data sets is by no means exhaustive but provides a good selection of panel data sets readily accessible for economic research. In contrast to these micro panel surveys, there are several studies on purchasing power parity (PPP) and growth convergence among countries utilizing macro panels. A well-utilized resource is the Penn World Tables available at www.nber.org. International trade studies utilizing panels using World Development Indicators are available from the World Bank at www.worldbank.org/data, Direction of Trade data and International Financial Statistics from the International Monetary Fund (www.imf.org). Several country-specific characteristics for these pooled country studies can be obtained from the CIA's "World Factbook" available on the web at http://www.odci.gov/cia/publications/factbook. For issues of nonstationarity in these long time-series macro panels, see Chapter 12.

Virtually every graduate text in econometrics contains a chapter or a major section on the econometrics of panel data. Recommended readings on this subject include Hsiao's (2003) Econometric Society monograph along with two chapters in the *Handbook of Econometrics*: chapter 22 by Chamberlain (1984) and chapter 53 by Arellano and Honoré (2001). Maddala (1993) edited two volumes collecting some of the classic articles on the subject. This collection of readings was updated with two more volumes covering the period 1992–2002 and edited by Baltagi (2002). Other books on the subject include Arellano (2003), Wooldridge (2002) and a handbook on the econometrics of panel data which in its second edition contained 33 chapters edited by Mátyás and Sevestre (1996). A book in honor of G.S. Maddala, edited by Hsiao et al. (1999); a book in honor of Pietro Balestra, edited by Krishnakumar and Ronchetti (2000); and a book with a nice historical perspective on panel data by Nerlove (2002). Recent survey papers include Baltagi and Kao (2000) and Hsiao (2001). Recent special issues of journals on panel data include two volumes of the *Annales D'Economie et de Statistique* edited by Sevestre (1999), a special issue of the *Oxford Bulletin of Economics and Statistics* edited by Banerjee (1999), two special issues (Volume 19, Numbers 3 and 4) of *Econometric Reviews* edited by Maasoumi and Heshmati, a special issue of *Advances in Econometrics* edited by Baltagi, Fomby and Hill (2000) and a special issue of *Empirical Economics* edited by Baltagi (2004).

The objective of this book is to provide a simple introduction to some of the basic issues of panel data analysis. It is intended for economists and social scientists with the usual background in statistics and econometrics. Panel data methods have been used in political science, see Beck and Katz (1995); in sociology, see England et al. (1988); in finance, see Brown, Kleidon and Marsh (1983) and Boehmer and Megginson (1990); and in marketing, see Erdem (1996) and Keane (1997). While restricting the focus of the book to basic topics may not do justice to this rapidly growing literature, it is nevertheless unavoidable in view of the space limitations of the book. Topics not covered in this book include duration models and hazard functions (see Heckman and Singer, 1985; Florens, Forgére and Monchart, 1996; Horowitz and Lee, 2004). Also, the frontier production function literature using panel data (see Schmidt and Sickles, 1984; Battese and Coelli, 1988; Cornwell, Schmidt and Sickles, 1990; Kumbhakar and Lovell, 2000; Koop and Steel, 2001) and the literature on time-varying parameters, random coefficients and Bayesian models, see Swamy and Tavlas (2001) and Hsiao (2003). The program evaluation literature, see Heckman, Ichimura and Todd (1998) and Abbring and Van den Berg (2004), to mention a few.

1.2 WHY SHOULD WE USE PANEL DATA? THEIR BENEFITS AND LIMITATIONS

Hsiao (2003) and Klevmarken (1989) list several benefits from using panel data. These include the following.

(1) Controlling for *individual heterogeneity*. Panel data suggests that individuals, firms, states or countries are heterogeneous. Time-series and cross-section studies not controlling this heterogeneity run the risk of obtaining biased results, e.g. see Moulton (1986, 1987). Let us demonstrate this with an empirical example. Baltagi and Levin (1992) consider cigarette demand across 46 American states for the years 1963–88. Consumption is modeled as a function of lagged consumption, price and income. These variables vary with states and time. However, there are a lot of other variables that may be state-invariant or time-invariant that may affect consumption. Let us call these Z_i and W_t, respectively. Examples of Z_i are religion and education. For the religion variable, one may not be able to get the percentage of the population

that is, say, Mormon in each state for every year, nor does one expect that to change much across time. The same holds true for the percentage of the population completing high school or a college degree. Examples of W_t include advertising on TV and radio. This advertising is nationwide and does not vary across states. In addition, some of these variables are difficult to measure or hard to obtain so that not all the Z_i or W_t variables are available for inclusion in the consumption equation. Omission of these variables leads to bias in the resulting estimates. Panel data are able to control for these state- and time-invariant variables whereas a time-series study or a cross-section study cannot. In fact, from the data one observes that Utah has less than half the average per capita consumption of cigarettes in the USA. This is because it is mostly a Mormon state, a religion that prohibits smoking. Controlling for Utah in a cross-section regression may be done with a dummy variable which has the effect of removing that state's observation from the regression. This would not be the case for panel data as we will shortly discover. In fact, with panel data, one might first difference the data to get rid of all Z_i-type variables and hence effectively control for all state-specific characteristics. This holds whether the Z_i are observable or not. Alternatively, the dummy variable for Utah controls for every state-specific effect that is distinctive of Utah without omitting the observations for Utah.

Another example is given by Hajivassiliou (1987) who studies the external debt repayments problem using a panel of 79 developing countries observed over the period 1970–82. These countries differ in terms of their colonial history, financial institutions, religious affiliations and political regimes. All of these country-specific variables affect the attitudes that these countries have with regards to borrowing and defaulting and the way they are treated by the lenders. Not accounting for this country heterogeneity causes serious misspecification.

Deaton (1995) gives another example from agricultural economics. This pertains to the question of whether small farms are more productive than large farms. OLS regressions of yield per hectare on inputs such as land, labor, fertilizer, farmer's education, etc. usually find that the sign of the estimate of the land coefficient is negative. These results imply that smaller farms are more productive. Some explanations from economic theory argue that higher output per head is an optimal response to uncertainty by small farmers, or that hired labor requires more monitoring than family labor. Deaton (1995) offers an alternative explanation. This regression suffers from the omission of unobserved heterogeneity, in this case "land quality", and this omitted variable is systematically correlated with the explanatory variable (farm size). In fact, farms in low-quality marginal areas (semi-desert) are typically large, while farms in high-quality land areas are often small. Deaton argues that while gardens add more value-added per hectare than a sheep station, this does not imply that sheep stations should be organized as gardens. In this case, differencing may not resolve the "small farms are productive" question since farm size will usually change little or not at all over short periods.

(2) Panel data give *more informative data, more variability, less collinearity among the variables, more degrees of freedom and more efficiency*. Time-series studies are plagued with multicollinearity; for example, in the case of demand for cigarettes above, there is high collinearity between price and income in the aggregate time series for the USA. This is less likely with a panel across American states since the cross-section dimension adds a lot of variability, adding more informative data on price and income. In fact, the variation in the data can be decomposed into variation between states of different sizes and characteristics, and variation within states. The former variation is usually bigger. With additional, more informative data one can produce more reliable parameter estimates. Of course, the same relationship has to hold for each state, i.e. the data have to be poolable. This is a testable assumption and one that we will tackle in due course.

(3) Panel data are better able to study the *dynamics of adjustment*. Cross-sectional distributions that look relatively stable hide a multitude of changes. Spells of unemployment, job turnover, residential and income mobility are better studied with panels. Panel data are also well suited to study the duration of economic states like unemployment and poverty, and if these panels are long enough, they can shed light on the speed of adjustments to economic policy changes. For example, in measuring unemployment, cross-sectional data can estimate what proportion of the population is unemployed at a point in time. Repeated cross-sections can show how this proportion changes over time. Only panel data can estimate what proportion of those who are unemployed in one period can remain unemployed in another period. Important policy questions like determining whether families' experiences of poverty, unemployment and welfare dependence are transitory or chronic necessitate the use of panels. Deaton (1995) argues that, unlike cross-sections, panel surveys yield data on *changes* for individuals or households. It allows us to observe *how* the individual living standards change during the development process. It enables us to determine *who* is benefiting from development. It also allows us to observe whether poverty and deprivation are transitory or long-lived, the income-dynamics question. Panels are also necessary for the estimation of intertemporal relations, lifecycle and intergenerational models. In fact, panels can relate the individual's experiences and behavior at one point in time to other experiences and behavior at another point in time. For example, in evaluating training programs, a group of participants and nonparticipants are observed before and after the implementation of the training program. This is a panel of at least two time periods and the basis for the "difference in differences" estimator usually applied in these studies; see Bertrand, Duflo and Mullainathan (2004).

(4) Panel data are better able to *identify and measure effects that are simply not detectable in pure cross-section or pure time-series data*. Suppose that we have a cross-section of women with a 50% average yearly labor force participation rate. This might be due to (a) each woman having a 50% chance of being in the labor force, in any given year, or (b) 50% of the women working all the time and 50% not at all. Case (a) has high turnover, while case (b) has no turnover. Only panel data could discriminate between these cases. Another example is the determination of whether union membership increases or decreases wages. This can be better answered as we observe a worker moving from union to nonunion jobs or vice versa. Holding the individual's characteristics constant, we will be better equipped to determine whether union membership affects wage and by how much. This analysis extends to the estimation of other types of wage differentials holding individuals' characteristics constant. For example, the estimation of wage premiums paid in dangerous or unpleasant jobs. Economists studying workers' levels of satisfaction run into the problem of anchoring in a cross-section study, see Winkelmann and Winkelmann (1998) in Chapter 11. The survey usually asks the question: "how satisfied are you with your life?" with zero meaning completely dissatisfied and 10 meaning completely satisfied. The problem is that each individual anchors their scale at different levels, rendering interpersonal comparisons of responses meaningless. However, in a panel study, where the metric used by individuals is time-invariant over the period of observation, one can avoid this problem since a difference (or fixed effects) estimator will make inference based only on intra- rather than interpersonal comparison of satisfaction.

(5) Panel data models allow us to *construct and test more complicated behavioral models than purely cross-section or time-series data*. For example, technical efficiency is better studied and modeled with panels (see Baltagi and Griffin, 1988b; Cornwell, Schmidt and Sickles, 1990; Kumbhakar and Lovell, 2000; Baltagi, Griffin and Rich, 1995; Koop and Steel, 2001). Also,

fewer restrictions can be imposed in panels on a distributed lag model than in a purely time-series study (see Hsiao, 2003).

(6) Micro panel data gathered on individuals, firms and households may be more accurately measured than similar variables measured at the macro level. *Biases resulting from aggregation over firms or individuals may be reduced or eliminated* (see Blundell, 1988; Klevmarken, 1989). For specific advantages and disadvantages of estimating life cycle models using micro panel data, see Blundell and Meghir (1990).

(7) Macro panel data on the other hand have a longer time series and unlike the problem of nonstandard distributions typical of unit roots tests in time-series analysis, Chapter 12 shows that panel unit root tests have standard asymptotic distributions.

Limitations of panel data include:

(1) *Design and data collection problems*. For an extensive discussion of problems that arise in designing panel surveys as well as data collection and data management issues see Kasprzyk et al. (1989). These include problems of coverage (incomplete account of the population of interest), nonresponse (due to lack of cooperation of the respondent or because of interviewer error), recall (respondent not remembering correctly), frequency of interviewing, interview spacing, reference period, the use of bounding and time-in-sample bias (see Bailar, 1989).[1]

(2) *Distortions of measurement errors*. Measurement errors may arise because of faulty responses due to unclear questions, memory errors, deliberate distortion of responses (e.g. prestige bias), inappropriate informants, misrecording of responses and interviewer effects (see Kalton, Kasprzyk and McMillen, 1989). Herriot and Spiers (1975), for example, match CPS and Internal Revenue Service data on earnings of the same individuals and show that there are discrepancies of at least 15% between the two sources of earnings for almost 30% of the matched sample. The validation study by Duncan and Hill (1985) on the PSID also illustrates the significance of the measurement error problem. They compare the responses of the employees of a large firm with the records of the employer. Duncan and Hill (1985) find small response biases except for work hours which are overestimated. The ratio of measurement error variance to the true variance is found to be 15% for annual earnings, 37% for annual work hours and 184% for average hourly earnings. These figures are for a one-year recall, i.e. 1983 for 1982, and are more than doubled with two years' recall. Brown and Light (1992) investigate the inconsistency in job tenure responses in the PSID and NLS. Cross-section data users have little choice but to believe the reported values of tenure (unless they have external information) while users of panel data can check for inconsistencies of tenure responses with elapsed time between interviews. For example, a respondent may claim to have three years of tenure in one interview and a year later claim six years. This should alert the user of this panel to the presence of measurement error. Brown and Light (1992) show that failure to use internally consistent tenure sequences can lead to misleading conclusions about the slope of wage-tenure profiles.

(3) *Selectivity problems*. These include:

(a) *Self-selectivity*. People choose not to work because the reservation wage is higher than the offered wage. In this case we observe the characteristics of these individuals but not their wage. Since only their wage is missing, the sample is censored. However, if we do not observe all data on these people this would be a truncated sample. An example of truncation is the New Jersey negative income tax experiment. We are only interested in poverty, and people with income larger than 1.5 times the poverty level are dropped from the sample. Inference from this truncated sample introduces bias that is not helped by more data, because of the truncation (see Hausman and Wise, 1979).

(b) *Nonresponse*. This can occur at the initial wave of the panel due to refusal to participate, nobody at home, untraced sample unit, and other reasons. Item (or partial) nonresponse occurs when one or more questions are left unanswered or are found not to provide a useful response. Complete nonresponse occurs when no information is available from the sampled household. Besides the efficiency loss due to missing data, this nonresponse can cause serious identification problems for the population parameters. Horowitz and Manski (1998) show that the seriousness of the problem is directly proportional to the amount of nonresponse. Nonresponse rates in the first wave of the European panels vary across countries from 10% in Greece and Italy where participation is compulsory, to 52% in Germany and 60% in Luxembourg. The overall nonresponse rate is 28%, see Peracchi (2002). The comparable nonresponse rate for the first wave of the PSID is 24%, for the BHPS (26%), for the GSOEP (38%) and for PSELL (35%).

(c) *Attrition*. While nonresponse occurs also in cross-section studies, it is a more serious problem in panels because subsequent waves of the panel are still subject to nonresponse. Respondents may die, or move, or find that the cost of responding is high. See Björklund (1989) and Ridder (1990, 1992) on the consequences of attrition. The degree of attrition varies depending on the panel studied; see Kalton, Kasprzyk and McMillen (1989) for several examples. In general, the overall rates of attrition increase from one wave to the next, but the rate of increase declines over time. Becketti et al. (1988) study the representativeness of the PSID after 14 years since it started. The authors find that only 40% of those originally in the sample in 1968 remained in the sample in 1981. However, they do find that as far as the dynamics of entry and exit are concerned, the PSID is still representative. Attrition rates between the first and second wave vary from 6% in Italy to 24% in the UK. The average attrition rate is about 10%. The comparable rates of attrition from the first to the second wave are 12% in the BHPS, 12.4% for the West German sample and 8.9% for the East German sample in the GSOEP and 15% for PSELL, see Peracchi (2002). In order to counter the effects of attrition, rotating panels are sometimes used, where a fixed percentage of the respondents are replaced in every wave to replenish the sample. More on rotating and pseudo-panels in Chapter 10. A special issue of the *Journal of Human Resources*, Spring 1998, is dedicated to attrition in longitudinal surveys.

(4) *Short time-series dimension*. Typical micro panels involve annual data covering a short time span for each individual. This means that asymptotic arguments rely crucially on the number of individuals tending to infinity. Increasing the time span of the panel is not without cost either. In fact, this increases the chances of attrition and increases the computational difficulty for limited dependent variable panel data models (see Chapter 11).

(5) *Cross-section dependence*. Macro panels on countries or regions with long time series that do not account for cross-country dependence may lead to misleading inference. Chapter 12 shows that several panel unit root tests suggested in the literature assumed cross-section independence. Accounting for cross-section dependence turns out to be important and affects inference. Alternative panel unit root tests are suggested that account for this dependence.

Panel data is not a panacea and will not solve all the problems that a time series or a cross-section study could not handle. Examples are given in Chapter 12, where we cite econometric studies arguing that panel data will yield more powerful unit root tests than individual time series. This in turn should help shed more light on the purchasing power parity and the growth convergence questions. In fact, this led to a flurry of empirical applications along with some sceptics who argued that panel data did not save the PPP or the growth convergence problem, see Maddala (1999), Maddala, Wu and Liu (2000) and Banerjee, Marcellino and Osbat (2004, 2005). Collecting panel data is quite costly, and there is always the question of how often one should interview respondents. Deaton (1995) argues that economic development is far from instantaneous, so that changes from one year to the next are probably too noisy and too short-term to be really useful. He concludes that the payoff for panel data is over long time periods, five years, ten years, or even longer. In contrast, for health and nutrition issues, especially those of children, one could argue the opposite case, i.e., those panels with a shorter time span are needed in order to monitor the health and development of these children.

This book will make the case that panel data provides several advantages worth its cost. However, as Griliches (1986) argued about economic data in general, the more we have of it, the more we demand of it. The economist using panel data or any data for that matter has to know its limitations.

NOTE

1. Bounding is used to prevent the shifting of events from outside the recall period into the recall period. Time-in-sample bias is observed when a significantly different level for a characteristic occurs in the first interview than in later interviews, when one would expect the same level.

2 The One-way Error Component Regression Model

2.1 INTRODUCTION

A panel data regression differs from a regular time-series or cross-section regression in that it has a double subscript on its variables, i.e.

$$y_{it} = \alpha + X'_{it}\beta + u_{it} \quad i = 1, \dots, N;\ t = 1, ..., T \tag{2.1}$$

with i denoting households, individuals, firms, countries, etc. and t denoting time. The i subscript, therefore, denotes the cross-section dimension whereas t denotes the time-series dimension. α is a scalar, β is $K \times 1$ and X_{it} is the itth observation on K explanatory variables. Most of the panel data applications utilize a one-way error component model for the disturbances, with

$$u_{it} = \mu_i + \nu_{it} \tag{2.2}$$

where μ_i denotes the *unobservable* individual-specific effect and ν_{it} denotes the remainder disturbance. For example, in an earnings equation in labor economics, y_{it} will measure earnings of the head of the household, whereas X_{it} may contain a set of variables like experience, education, union membership, sex, race, etc. Note that μ_i is time-invariant and it accounts for any individual-specific effect that is not included in the regression. In this case we could think of it as the individual's unobserved ability. The remainder disturbance ν_{it} varies with individuals and time and can be thought of as the usual disturbance in the regression. Alternatively, for a production function utilizing data on firms across time, y_{it} will measure output and X_{it} will measure inputs. The unobservable firm-specific effects will be captured by the μ_i and we can think of these as the unobservable entrepreneurial or managerial skills of the firm's executives. Early applications of error components in economics include Kuh (1959) on investment, Mundlak (1961) and Hoch (1962) on production functions and Balestra and Nerlove (1966) on demand for natural gas. In vector form (2.1) can be written as

$$y = \alpha\iota_{NT} + X\beta + u = Z\delta + u \tag{2.3}$$

where y is $NT \times 1$, X is $NT \times K$, $Z = [\iota_{NT}, X]$, $\delta' = (\alpha', \beta')$ and ι_{NT} is a vector of ones of dimension NT. Also, (2.2) can be written as

$$u = Z_\mu\mu + \nu \tag{2.4}$$

where $u' = (u_{11}, \dots, u_{1T}, u_{21}, \dots, u_{2T}, \dots, u_{N1}, \dots, u_{NT})$ with the observations stacked such that the slower index is over individuals and the faster index is over time. $Z_\mu = I_N \otimes \iota_T$ where I_N is an identity matrix of dimension N, ι_T is a vector of ones of dimension T and $\otimes$ denotes Kronecker product. Z_μ is a selector matrix of ones and zeros, or simply the matrix of individual dummies that one may include in the regression to estimate the μ_i if they are assumed to be fixed parameters. $\mu' = (\mu_1, \dots, \mu_N)$ and $\nu' = (\nu_{11}, \dots, \nu_{1T}, \dots, \nu_{N1}, \dots, \nu_{NT})$. Note that

$Z_\mu Z'_\mu = I_N \otimes J_T$ where J_T is a matrix of ones of dimension T and $P = Z_\mu(Z'_\mu Z_\mu)^{-1}Z'_\mu$, the projection matrix on Z_μ, reduces to $I_N \otimes \bar{J}_T$ where $\bar{J}_T = J_T/T$. P is a matrix which averages the observation across time for each individual, and $Q = I_{NT} - P$ is a matrix which obtains the deviations from individual means. For example, regressing y on the matrix of dummy variables Z_μ gets the predicted values Py which has a typical element $\overline{y}_{i.} = \sum_{t=1}^{T} y_{it}/T$ repeated T times for each individual. The residuals of this regression are given by Qy which has a typical element $\left(y_{it} - \overline{y}_{i.}\right)$. P and Q are (i) symmetric idempotent matrices, i.e. $P' = P$ and $P^2 = P$. This means that $\text{rank}(P) = \text{tr}(P) = N$ and $\text{rank}(Q) = \text{tr}(Q) = N(T-1)$. This uses the result that the rank of an idempotent matrix is equal to its trace (see Graybill, 1961, theorem 1.63). Also, (ii) P and Q are orthogonal, i.e. $PQ = 0$ and (iii) they sum to the identity matrix $P + Q = I_{NT}$. In fact, any two of these properties imply the third (see Graybill, 1961, theorem 1.68).

2.2 THE FIXED EFFECTS MODEL

In this case, the μ_i are assumed to be fixed parameters to be estimated and the remainder disturbances stochastic with ν_{it} independent and identically distributed $\text{IID}\left(0, \sigma_\nu^2\right)$. The X_{it} are assumed independent of the ν_{it} for all i and t. The fixed effects model is an appropriate specification if we are focusing on a specific set of N firms, say, IBM, GE, Westinghouse, etc. and our inference is restricted to the behavior of these sets of firms. Alternatively, it could be a set of N OECD countries, or N American states. Inference in this case is conditional on the particular N firms, countries or states that are observed. One can substitute the disturbances given by (2.4) into (2.3) to get

$$y = \alpha\iota_{NT} + X\beta + Z_\mu\mu + \nu = Z\delta + Z_\mu\mu + \nu \tag{2.5}$$

and then perform ordinary least squares (OLS) on (2.5) to get estimates of α, β and μ. Note that Z is $NT \times (K+1)$ and Z_μ, the matrix of individual dummies, is $NT \times N$. If N is large, (2.5) will include too many individual dummies, and the matrix to be inverted by OLS is large and of dimension $(N+K)$. In fact, since α and β are the parameters of interest, one can obtain the LSDV (least squares dummy variables) estimator from (2.5), by premultiplying the model by Q and performing OLS on the resulting transformed model:

$$Qy = QX\beta + Q\nu \tag{2.6}$$

This uses the fact that $QZ_\mu = Q\iota_{NT} = 0$, since $PZ_\mu = Z_\mu$. In other words, the Q matrix wipes out the individual effects. This is a regression of $\widetilde{y} = Qy$ with typical element $(y_{it} - \bar{y}_{i.})$ on $\widetilde{X} = QX$ with typical element $(X_{it,k} - \bar{X}_{i.,k})$ for the kth regressor, $k = 1, 2, \ldots, K$. This involves the inversion of a $(K \times K)$ matrix rather than $(N+K) \times (N+K)$ as in (2.5). The resulting OLS estimator is

$$\widetilde{\beta} = \left(X'QX\right)^{-1} X'Qy \tag{2.7}$$

with $\text{var}(\widetilde{\beta}) = \sigma_\nu^2(X'QX)^{-1} = \sigma_\nu^2(\widetilde{X}'\widetilde{X})^{-1}$. $\widetilde{\beta}$ could have been obtained from (2.5) using results on partitioned inverse or the Frisch–Waugh–Lovell theorem discussed in Davidson and MacKinnon (1993, p. 19). This uses the fact that P is the projection matrix on Z_μ and $Q = I_{NT} - P$ (see problem 2.1). In addition, generalized least squares (GLS) on (2.6), using the generalized inverse, will also yield $\widetilde{\beta}$ (see problem 2.2).

Note that for the simple regression

$$y_{it} = \alpha + \beta x_{it} + \mu_i + \nu_{it} \tag{2.8}$$

and averaging over time gives

$$\bar{y}_{i.} = \alpha + \beta \bar{x}_{i.} + \mu_i + \bar{\nu}_{i.} \tag{2.9}$$

Therefore, subtracting (2.9) from (2.8) gives

$$y_{it} - \bar{y}_{i.} = \beta(x_{it} - \bar{x}_{i.}) + (\nu_{it} - \bar{\nu}_{i.}) \tag{2.10}$$

Also, averaging across all observations in (2.8) gives

$$\bar{y}_{..} = \alpha + \beta \bar{x}_{..} + \bar{\nu}_{..} \tag{2.11}$$

where we utilized the restriction that $\sum_{i=1}^{N} \mu_i = 0$. This is an arbitrary restriction on the dummy variable coefficients to avoid the dummy variable trap, or perfect multicollinearity; see Suits (1984) for alternative formulations of this restriction. In fact only β and $(\alpha + \mu_i)$ are estimable from (2.8), and not α and μ_i separately, unless a restriction like $\sum_{i=1}^{N} \mu_i = 0$ is imposed. In this case, $\widetilde{\beta}$ is obtained from regression (2.10), $\widetilde{\alpha} = \bar{y}_{..} - \widetilde{\beta}\bar{x}_{..}$ can be recovered from (2.11) and $\widetilde{\mu}_i = \bar{y}_{i.} - \widetilde{\alpha} - \widetilde{\beta}\bar{x}_{i.}$ from (2.9). For large labor or consumer panels, where N is very large, regressions like (2.5) may not be feasible, since one is including $(N - 1)$ dummies in the regression. This fixed effects (FE) least squares, also known as least squares dummy variables (LSDV), suffers from a large loss of degrees of freedom. We are estimating $(N - 1)$ extra parameters, and too many dummies may aggravate the problem of multicollinearity among the regressors. In addition, this FE estimator cannot estimate the effect of any time-invariant variable like sex, race, religion, schooling or union participation. These time-invariant variables are wiped out by the Q transformation, the deviations from means transformation (see (2.10)). Alternatively, one can see that these time-invariant variables are spanned by the individual dummies in (2.5) and therefore any regression package attempting (2.5) will fail, signaling perfect multicollinearity. If (2.5) is the true model, LSDV is the best linear unbiased estimator (BLUE) as long as ν_{it} is the standard classical disturbance with mean 0 and variance–covariance matrix $\sigma_\nu^2 I_{NT}$. Note that as $T \to \infty$, the FE estimator is consistent. However, if T is fixed and $N \to \infty$ as is typical in short labor panels, then only the FE estimator of β is consistent; the FE estimators of the individual effects $(\alpha + \mu_i)$ are not consistent since the number of these parameters increases as N increases. This is the incidental parameter problem discussed by Neyman and Scott (1948) and reviewed more recently by Lancaster (2000). Note that when the true model is fixed effects as in (2.5), OLS on (2.1) yields biased and inconsistent estimates of the regression parameters. This is an omission variables bias due to the fact that OLS deletes the individual dummies when in fact they are relevant.

(1) *Testing for fixed effects*. One could test the joint significance of these dummies, i.e. $H_0; \mu_1 = \mu_2 = \cdots = \mu_{N-1} = 0$, by performing an F-test. (Testing for individual effects will be treated extensively in Chapter 4.) This is a simple Chow test with the restricted residual sums of squares (RRSS) being that of OLS on the pooled model and the unrestricted residual sums of squares (URSS) being that of the LSDV regression. If N is large, one can perform the Within transformation and use that residual sum of squares as the URSS. In this case

$$F_0 = \frac{(\text{RRSS} - \text{URSS})/(N-1)}{\text{URSS}/(NT - N - K)} \overset{H_0}{\sim} F_{N-1,N(T-1)-K} \tag{2.12}$$

(2) *Computational warning*. One computational caution for those using the Within regression given by (2.10). The s^2 of this regression as obtained from a typical regression package divides the residual sums of squares by $NT - K$ since the intercept and the dummies are not included. The proper s^2, say s^{*2} from the LSDV regression in (2.5), would divide the same residual sums of squares by $N(T-1) - K$. Therefore, one has to adjust the variances obtained from the Within regression (2.10) by multiplying the variance–covariance matrix by (s^{*2}/s^2) or simply by multiplying by $[NT - K]/[N(T-1) - K]$.

(3) *Robust estimates of the standard errors*. For the Within estimator, Arellano (1987) suggests a simple method for obtaining robust estimates of the standard errors that allow for a general variance–covariance matrix on the ν_{it} as in White (1980). One would stack the panel as an equation for each individual:

$$y_i = Z_i\delta + \mu_i\iota_T + \nu_i \tag{2.13}$$

where y_i is $T \times 1$, $Z_i = [\iota_T, X_i]$, X_i is $T \times K$, μ_i is a scalar, $\delta' = (\alpha, \beta')$, ι_T is a vector of ones of dimension T and ν_i is $T \times 1$. In general, $E(\nu_i\nu_i') = \Omega_i$ for $i = 1, 2, \ldots, N$, where Ω_i is a positive definite matrix of dimension T. We still assume $E(\nu_i\nu_j') = 0$, for $i \neq j$. T is assumed small and N large as in household or company panels, and the asymptotic results are performed for $N \to \infty$ and T fixed. Performing the Within transformation on this set of equations (2.13) one gets

$$\widetilde{y}_i = \widetilde{X}_i\beta + \widetilde{\nu}_i \tag{2.14}$$

where $\widetilde{y} = Qy$, $\widetilde{X} = QX$ and $\widetilde{\nu} = Q\nu$, with $\widetilde{y} = (\widetilde{y}_1', \ldots, \widetilde{y}_N')'$ and $\widetilde{y}_i = (I_T - \bar{J}_T)y_i$. Computing robust least squares on this system, as described by White (1980), under the restriction that each equation has the same β one gets the Within estimator of β which has the following asymptotic distribution:

$$N^{1/2}(\widetilde{\beta} - \beta) \sim N(0, M^{-1}VM^{-1}) \tag{2.15}$$

where $M = \text{plim}(\widetilde{X}'\widetilde{X})/N$ and $V = \text{plim}\sum_{i=1}^N(\widetilde{X}_i'\Omega_i\widetilde{X}_i)/N$. Note that $\widetilde{X}_i = (I_T - \bar{J}_T)X_i$ and $\widetilde{X}'Q\ \text{diag}[\Omega_i]Q\widetilde{X} = \widetilde{X}'\text{diag}[\Omega_i]\widetilde{X}$ (see problem 2.3). In this case, V is estimated by $\widetilde{V} = \sum_{i=1}^N \widetilde{X}_i'\widetilde{u}_i\widetilde{u}_i'\widetilde{X}_i/N$, where $\widetilde{u}_i = \widetilde{y}_i - \widetilde{X}_i\widetilde{\beta}$. Therefore, the robust asymptotic variance–covariance matrix of β is estimated by

$$\text{var}(\widetilde{\beta}) = (\widetilde{X}'\widetilde{X})^{-1}\left[\sum_{i=1}^N \widetilde{X}_i'\widetilde{u}_i\widetilde{u}_i'\widetilde{X}_i\right](\widetilde{X}'\widetilde{X})^{-1} \tag{2.16}$$

2.3 THE RANDOM EFFECTS MODEL

There are too many parameters in the fixed effects model and the loss of degrees of freedom can be avoided if the μ_i can be assumed random. In this case $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$, $\nu_{it} \sim \text{IID}(0, \sigma_\nu^2)$ and the μ_i are independent of the ν_{it}. In addition, the X_{it} are independent of the μ_i and ν_{it}, for all i and t. The random effects model is an appropriate specification if we are drawing N individuals randomly from a large population. This is usually the case for household panel studies. Care is taken in the design of the panel to make it "representative" of the population we are trying to make inferences about. In this case, N is usually large and a fixed effects model would lead to an enormous loss of degrees of freedom. The individual effect is characterized as random and inference pertains to the population from which this sample was randomly drawn.

But what is the population in this case? Nerlove and Balestra (1996) emphasize Haavelmo's (1944) view that the population "consists *not* of an infinity of individuals, in general, but of an infinity of *decisions*" that each individual might make. This view is consistent with a random effects specification. From (2.4), one can compute the variance–covariance matrix

$$\begin{aligned}\Omega = E(uu') &= Z_\mu E(\mu\mu')Z'_\mu + E(\nu\nu') \\ &= \sigma^2_\mu(I_N \otimes J_T) + \sigma^2_\nu(I_N \otimes I_T)\end{aligned} \tag{2.17}$$

This implies a homoskedastic variance $\text{var}(u_{it}) = \sigma^2_\mu + \sigma^2_\nu$ for all i and t, and an equicorrelated block-diagonal covariance matrix which exhibits serial correlation over time only between the disturbances of the same individual. In fact,

$$\begin{aligned}\text{cov}(u_{it}, u_{js}) &= \sigma^2_\mu + \sigma^2_\nu \quad \text{for} \quad i = j, t = s \\ &= \sigma^2_\mu \qquad\quad\ \ \text{for} \quad i = j, t \neq s\end{aligned}$$

and zero otherwise. This also means that the correlation coefficient between u_{it} and u_{js} is

$$\begin{aligned}\rho = \text{correl}(u_{it}, u_{js}) &= 1 \qquad\qquad\qquad \text{for}\, i = j, t = s \\ &= \sigma^2_\mu/(\sigma^2_\mu + \sigma^2_\nu) \quad \text{for}\, i = j, t \neq s\end{aligned}$$

and zero otherwise. In order to obtain the GLS estimator of the regression coefficients, we need Ω^{-1}. This is a huge matrix for typical panels and is of dimension $NT \times NT$. No brute force inversion should be attempted even if the researcher's application has a small N and T.[1] We will follow a simple trick devised by Wansbeek and Kapteyn (1982b, 1983) that allows the derivation of Ω^{-1} and $\Omega^{-1/2}$.[2] Essentially, one replaces J_T by $T\bar{J}_T$ and I_T by $(E_T + \bar{J}_T)$ where E_T is by definition $(I_T - \bar{J}_T)$. In this case

$$\Omega = T\sigma^2_\mu(I_N \otimes \bar{J}_T) + \sigma^2_\nu(I_N \otimes E_T) + \sigma^2_\nu(I_N \otimes \bar{J}_T)$$

Collecting terms with the same matrices, we get

$$\Omega = (T\sigma^2_\mu + \sigma^2_\nu)(I_N \otimes \bar{J}_T) + \sigma^2_\nu(I_N \otimes E_T) = \sigma^2_1 P + \sigma^2_\nu Q \tag{2.18}$$

where $\sigma^2_1 = T\sigma^2_\mu + \sigma^2_\nu$. (2.18) is the spectral decomposition representation of Ω, with σ^2_1 being the first unique characteristic root of Ω of multiplicity N and σ^2_ν the second unique characteristic root of Ω of multiplicity $N(T-1)$. It is easy to verify, using the properties of P and Q, that

$$\Omega^{-1} = \frac{1}{\sigma^2_1}P + \frac{1}{\sigma^2_\nu}Q \tag{2.19}$$

and

$$\Omega^{-1/2} = \frac{1}{\sigma_1}P + \frac{1}{\sigma_\nu}Q \tag{2.20}$$

In fact, $\Omega^r = (\sigma^2_1)^r P + (\sigma^2_\nu)^r Q$ where r is an arbitrary scalar. Now we can obtain GLS as a weighted least squares. Fuller and Battese (1973, 1974) suggested premultiplying the regression equation given in (2.3) by $\sigma_\nu\Omega^{-1/2} = Q + (\sigma_\nu/\sigma_1)P$ and performing OLS on the resulting transformed regression. In this case, $y^* = \sigma_\nu\Omega^{-1/2}y$ has a typical element $y_{it} - \theta\bar{y}_{i.}$ where $\theta = 1 - (\sigma_\nu/\sigma_1)$ (see problem 2.4). This transformed regression inverts a matrix of dimension $(K+1)$ and can easily be implemented using any regression package.

The best quadratic unbiased (BQU) estimators of the variance components arise naturally from the spectral decomposition of Ω. In fact, $Pu \sim (0, \sigma_1^2 P)$ and $Qu \sim (0, \sigma_\nu^2 Q)$ and

$$\widehat{\sigma}_1^2 = \frac{u'Pu}{\mathrm{tr}(P)} = T \sum_{i=1}^{N} \bar{u}_{i.}^2 / N \tag{2.21}$$

and

$$\widehat{\sigma}_\nu^2 = \frac{u'Qu}{\mathrm{tr}(Q)} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} (u_{it} - \bar{u}_{i.})^2}{N(T-1)} \tag{2.22}$$

provide the BQU estimators of σ_1^2 and σ_ν^2, respectively (see problem 2.5).

These are analyses of variance-type estimators of the variance components and are minimum variance-unbiased under normality of the disturbances (see Graybill, 1961). The true disturbances are not known and therefore (2.21) and (2.22) are not feasible. Wallace and Hussain (1969) suggest substituting OLS residual $\widehat{u}_{\mathrm{OLS}}$ instead of the true u. After all, under the random effects model, the OLS estimates are still unbiased and consistent, but no longer efficient. Amemiya (1971) shows that these estimators of the variance components have a different asymptotic distribution from that knowing the true disturbances. He suggests using the LSDV residuals instead of the OLS residuals. In this case $\widetilde{u} = y - \widetilde{\alpha}\iota_{NT} - X\widetilde{\beta}$ where $\widetilde{\alpha} = \bar{y}_{..} - \bar{X}'_{..}\widetilde{\beta}$ and $\bar{X}'_{..}$ is a $1 \times K$ vector of averages of all regressors. Substituting these $\widetilde{u}$ for u in (2.21) and (2.22) we get the Amemiya-type estimators of the variance components. The resulting estimates of the variance components have the same asymptotic distribution as that knowing the true disturbances:

$$\begin{pmatrix} \sqrt{NT}(\widehat{\sigma}_\nu^2 - \sigma_\nu^2) \\ \sqrt{N}(\widehat{\sigma}_\mu^2 - \sigma_\mu^2) \end{pmatrix} \sim N\left(0, \begin{pmatrix} 2\sigma_\nu^4 & 0 \\ 0 & 2\sigma_\mu^4 \end{pmatrix}\right) \tag{2.23}$$

where $\widehat{\sigma}_\mu^2 = (\widehat{\sigma}_1^2 - \widehat{\sigma}_\nu^2)/T$.[3]

Swamy and Arora (1972) suggest running two regressions to get estimates of the variance components from the corresponding mean square errors of these regressions. The first regression is the Within regression, given in (2.10), which yields the following s^2:

$$\widehat{\widehat{\sigma}}_\nu^2 = [y'Qy - y'QX(X'QX)^{-1}X'Qy]/[N(T-1) - K] \tag{2.24}$$

The second regression is the Between regression which runs the regression of averages across time, i.e.

$$\bar{y}_{i.} = \alpha + \bar{X}'_{i.}\beta + \bar{u}_{i.} \quad i = 1, \ldots, N \tag{2.25}$$

This is equivalent to premultiplying the model in (2.5) by P and running OLS. The only caution is that the latter regression has NT observations because it repeats the averages T times for each individual, while the cross-section regression in (2.25) is based on N observations. To remedy this, one can run the cross-section regression

$$\sqrt{T}\bar{y}_{i.} = \alpha\sqrt{T} + \sqrt{T}\bar{X}'_{i.}\beta + \sqrt{T}\bar{u}_{i.} \tag{2.26}$$

where one can easily verify that $\mathrm{var}(\sqrt{T}\bar{u}_{i.}) = \sigma_1^2$. This regression will yield an s^2 given by

$$\widehat{\widehat{\sigma}}_1^2 = (y'Py - y'PZ(Z'PZ)^{-1}Z'Py)/(N - K - 1) \tag{2.27}$$

Note that stacking the following two transformed regressions we just performed yields

$$\begin{pmatrix} Qy \\ Py \end{pmatrix} = \begin{pmatrix} QZ \\ PZ \end{pmatrix} \delta + \begin{pmatrix} Qu \\ Pu \end{pmatrix} \tag{2.28}$$

and the transformed error has mean 0 and variance–covariance matrix given by

$$\begin{pmatrix} \sigma_\nu^2 Q & 0 \\ 0 & \sigma_1^2 P \end{pmatrix}$$

Problem 2.7 asks the reader to verify that OLS on this system of $2NT$ observations yields OLS on the pooled model (2.3). Also, GLS on this system yields GLS on (2.3). Alternatively, one could get rid of the constant α by running the following stacked regressions:

$$\begin{pmatrix} Qy \\ (P - \bar{J}_{NT})y \end{pmatrix} = \begin{pmatrix} QX \\ (P - \bar{J}_{NT})X \end{pmatrix} \beta + \begin{pmatrix} Qu \\ (P - \bar{J}_{NT})u \end{pmatrix} \tag{2.29}$$

This follows from the fact that $Q\iota_{NT} = 0$ and $(P - \bar{J}_{NT})\iota_{NT} = 0$. The transformed error has zero mean and variance–covariance matrix

$$\begin{pmatrix} \sigma_\nu^2 Q & 0 \\ 0 & \sigma_1^2 (P - \bar{J}_{NT}) \end{pmatrix}$$

OLS on this system yields OLS on (2.3) and GLS on (2.29) yields GLS on (2.3). In fact,

$$\begin{aligned} \widehat{\beta}_{\text{GLS}} &= [(X'QX/\sigma_\nu^2) + X'(P - \bar{J}_{NT})X/\sigma_1^2]^{-1}[(X'Qy/\sigma_\nu^2) \\ &\quad + X'(P - \bar{J}_{NT})y/\sigma_1^2] \\ &= [W_{XX} + \phi^2 B_{XX}]^{-1}[W_{Xy} + \phi^2 B_{Xy}] \end{aligned} \tag{2.30}$$

with $\text{var}(\widehat{\beta}_{\text{GLS}}) = \sigma_\nu^2[W_{XX} + \phi^2 B_{XX}]^{-1}$. Note that $W_{XX} = X'QX$, $B_{XX} = X'(P - \bar{J}_{NT})X$ and $\phi^2 = \sigma_\nu^2/\sigma_1^2$. Also, the Within estimator of β is $\widetilde{\beta}_{\text{Within}} = W_{XX}^{-1} W_{Xy}$ and the Between estimator of β is $\widehat{\beta}_{\text{Between}} = B_{XX}^{-1} B_{Xy}$. This shows that $\widehat{\beta}_{\text{GLS}}$ is a matrix weighted average of $\widetilde{\beta}_{\text{Within}}$ and $\widehat{\beta}_{\text{Between}}$ weighing each estimate by the inverse of its corresponding variance. In fact

$$\widehat{\beta}_{\text{GLS}} = W_1 \widetilde{\beta}_{\text{Within}} + W_2 \widehat{\beta}_{\text{Between}} \tag{2.31}$$

where

$$W_1 = [W_{XX} + \phi^2 B_{XX}]^{-1} W_{XX}$$

and

$$W_2 = [W_{XX} + \phi^2 B_{XX}]^{-1}(\phi^2 B_{XX}) = I - W_1$$

This was demonstrated by Maddala (1971). Note that (i) if $\sigma_\mu^2 = 0$ then $\phi^2 = 1$ and $\widehat{\beta}_{\text{GLS}}$ reduces to $\widehat{\beta}_{\text{OLS}}$. (ii) If $T \to \infty$, then $\phi^2 \to 0$ and $\widehat{\beta}_{\text{GLS}}$ tends to $\widetilde{\beta}_{\text{Within}}$. Also, if W_{XX} is huge compared to B_{XX} then $\widehat{\beta}_{\text{GLS}}$ will be close to $\widetilde{\beta}_{\text{Within}}$. However, if B_{XX} dominates W_{XX} then $\widehat{\beta}_{\text{GLS}}$ tends to $\widehat{\beta}_{\text{Between}}$. In other words, the Within estimator ignores the Between variation, and the Between estimator ignores the Within variation. The OLS estimator gives equal weight to the Between and Within variations. From (2.30), it is clear that $\text{var}(\widetilde{\beta}_{\text{Within}}) - \text{var}(\widehat{\beta}_{\text{GLS}})$ is a positive semidefinite matrix, since ϕ^2 is positive. However, as $T \to \infty$ for any fixed N, $\phi^2 \to 0$ and both $\widehat{\beta}_{\text{GLS}}$ and $\widetilde{\beta}_{\text{Within}}$ have the same asymptotic variance.

Another estimator of the variance components was suggested by Nerlove (1971a). His suggestion is to estimate σ_μ^2 as $\sum_{i=1}^{N}(\widehat{\mu}_i - \widehat{\overline{\mu}})^2/(N-1)$ where $\widehat{\mu}_i$ are the dummy coefficients estimates from the LSDV regression. σ_ν^2 is estimated from the Within residual sums of squares divided by NT without correction for degrees of freedom.[4]

Note that, except for Nerlove's (1971a) method, one has to retrieve $\widehat{\sigma}_\mu^2$ as $(\widehat{\sigma}_1^2 - \widehat{\sigma}_\nu^2)/T$. In this case, there is no guarantee that the estimate of $\widehat{\sigma}_\mu^2$ would be nonnegative. Searle (1971) has an extensive discussion of the problem of negative estimates of the variance components in the biometrics literature. One solution is to replace these negative estimates by zero. This in fact is the suggestion of the Monte Carlo study by Maddala and Mount (1973). This study finds that negative estimates occurred only when the true σ_μ^2 was small and close to zero. In these cases OLS is still a viable estimator. Therefore, replacing negative $\widehat{\sigma}_\mu^2$ by zero is not a bad sin after all, and the problem is dismissed as not being serious.[5]

How about the properties of the various feasible GLS estimators of β? Under the random effects model, GLS based on the true variance components is BLUE, and all the feasible GLS estimators considered are asymptotically efficient as either N or $T \to \infty$. Maddala and Mount (1973) compared OLS, Within, Between, feasible GLS methods, MINQUE, Henderson's method III, true GLS and maximum likelihood estimation using their Monte Carlo study. They found little to choose among the various feasible GLS estimators in small samples and argued in favor of methods that were easier to compute. MINQUE was dismissed as more difficult to compute and the applied researcher given one shot at the data was warned to compute at least two methods of estimation, like an ANOVA feasible GLS and maximum likelihood to ensure that they do not yield drastically different results. If they do give different results, the authors diagnose misspecification.

Taylor (1980) derived exact finite sample results for the one-way error component model. He compared the Within estimator with the Swamy–Arora feasible GLS estimator. He found the following important results:

(1) Feasible GLS is more efficient than LSDV for all but the fewest degrees of freedom.
(2) The variance of feasible GLS is never more than 17% above the Cramer–Rao lower bound.
(3) More efficient estimators of the variance components do not necessarily yield more efficient feasible GLS estimators.

These finite sample results are confirmed by the Monte Carlo experiments carried out by Maddala and Mount (1973) and Baltagi (1981a).

Bellmann, Breitung and Wagner (1989) consider the bias in estimating the variance components using the Wallace and Hussain (1969) method due to the replacement of the true disturbances by OLS residuals, also the bias in the regression coefficients due to the use of estimated variance components rather than the true variance components. The magnitude of this bias is estimated using bootstrap methods for two economic applications. The first application relates product innovations, import pressure and factor inputs using a panel at the industry level. The second application estimates the earnings of 936 full-time working German males based on the first and second wave of the German Socio-Economic Panel. Only the first application revealed considerable bias in estimating σ_μ^2. However, this did not affect the bias much in the corresponding regression coefficients.

2.3.1 Fixed vs Random

Having discussed the fixed effects and the random effects models and the assumptions underlying them, the reader is left with the daunting question, which one to choose? This is not as easy a choice as it might seem. In fact, the fixed versus random effects issue has generated a hot debate in the biometrics and statistics literature which has spilled over into the panel data econometrics literature. Mundlak (1961) and Wallace and Hussain (1969) were early proponents of the fixed effects model and Balestra and Nerlove (1966) were advocates of the random error component model. In Chapter 4, we will study a specification test proposed by Hausman (1978) which is based on the difference between the fixed and random effects estimators. Unfortunately, applied researchers have interpreted a rejection as an adoption of the fixed effects model and nonrejection as an adoption of the random effects model.[6] Chamberlain (1984) showed that the fixed effects model imposes testable restrictions on the parameters of the reduced form model and one should check the validity of these restrictions before adopting the fixed effects model (see Chapter 4). Mundlak (1978) argued that the random effects model assumes exogeneity of *all* the regressors with the random individual effects. In contrast, the fixed effects model allows for endogeneity of *all* the regressors with these individual effects. So, it is an "all" or "nothing" choice of exogeneity of the regressors and the individual effects, see Chapter 7 for a more formal discussion of this subject. Hausman and Taylor (1981) allowed for *some* of the regressors to be correlated with the individual effects, as opposed to the all or nothing choice. These over-identification restrictions are testable using a Hausman-type test (see Chapter 7). For the applied researcher, performing fixed effects and random effects and the associated Hausman test reported in standard packages like Stata, LIMDEP, TSP, etc., the message is clear: Do not stop here. Test the restrictions implied by the fixed effects model derived by Chamberlain (1984) (see Chapter 4) and check whether a Hausman and Taylor (1981) specification might be a viable alternative (see Chapter 7).

2.4 MAXIMUM LIKELIHOOD ESTIMATION

Under normality of the disturbances, one can write the likelihood function as

$$L(\alpha,\beta,\phi^2,\sigma_\nu^2) = \text{ constant } - \frac{NT}{2}\log\sigma_\nu^2 + \frac{N}{2}\log\phi^2 - \frac{1}{2\sigma_\nu^2}u'\Sigma^{-1}u \tag{2.32}$$

where $\Omega = \sigma_\nu^2\Sigma$, $\phi^2 = \sigma_\nu^2/\sigma_1^2$ and $\Sigma = Q + \phi^{-2}P$ from (2.18). This uses the fact that $|\ \Omega\ |=$ product of its characteristic roots $= (\sigma_\nu^2)^{N(T-1)}\ (\sigma_1^2)^N = (\sigma_\nu^2)^{NT}\ (\phi^2)^{-N}$. Note that there is a one-to-one correspondence between ϕ^2 and σ_μ^2. In fact, $0 \le \sigma_\mu^2 < \infty$ translates into $0 < \phi^2 \le 1$. Brute force maximization of (2.32) leads to nonlinear first-order conditions (see Amemiya, 1971). Instead, Breusch (1987) concentrates the likelihood with respect to α and σ_ν^2. In this case, $\widehat{\alpha}_{mle} = \bar{y}_{..} - \bar{X}'_{..}\widehat{\beta}_{mle}$ and $\widehat{\sigma}^2_{\nu,mle} = (1/NT)\widehat{u}'\widehat{\Sigma}^{-1}\widehat{u}$ where $\widehat{u}$ and $\widehat{\Sigma}$ are based on maximum likelihood estimates of β, ϕ^2 and α. Let $d = y - X\widehat{\beta}_{mle}$ then $\widehat{\alpha}_{mle} = (1/NT)\iota'_{NT}d$ and $\widehat{u} = d - \iota_{NT}\widehat{\alpha}_{mle} = d - \bar{J}_{NT}d$. This implies that $\widehat{\sigma}^2_{\nu,mle}$ can be rewritten as

$$\widehat{\sigma}^2_{\nu,mle} = d'[Q + \phi^2(P - \bar{J}_{NT})]d/NT \tag{2.33}$$

and the concentrated likelihood becomes

$$L_C(\beta,\phi^2) = \text{ constant } - \frac{NT}{2}\log\{d'[Q + \phi^2(P - \bar{J}_{NT})]d\} + \frac{N}{2}\log\phi^2 \tag{2.34}$$

Maximizing (2.34) over ϕ^2, given β (see problem 2.9), yields

$$\widehat{\phi}^2 = \frac{d'Qd}{(T-1)d'(P-\bar{J}_{NT})d} = \frac{\sum\sum(d_{it}-\bar{d}_{i.})^2}{T(T-1)\sum(\bar{d}_{i.}-\bar{d}_{..})^2} \tag{2.35}$$

Maximizing (2.34) over β, given ϕ^2, yields

$$\widehat{\beta}_{mle} = [X'(Q+\phi^2(P-\bar{J}_{NT}))X]^{-1}X'[Q+\phi^2(P-\bar{J}_{NT})]y \tag{2.36}$$

One can iterate between β and ϕ^2 until convergence. Breusch (1987) shows that provided $T > 1$, any ith iteration β, call it β_i, gives $0 < \phi^2_{i+1} < \infty$ in the $(i+1)$th iteration. More importantly, Breusch (1987) shows that these ϕ^2_i have a "remarkable property" of forming a monotonic sequence. In fact, starting from the Within estimator of β, for $\phi^2 = 0$, the next ϕ^2 is finite and positive and starts a monotonically increasing sequence of ϕ^2. Similarly, starting from the Between estimator of β, for $(\phi^2 \to \infty)$ the next ϕ^2 is finite and positive and starts a monotonically decreasing sequence of ϕ^2. Hence, to guard against the possibility of a local maximum, Breusch (1987) suggests starting with $\widetilde{\beta}_{\text{Within}}$ and $\widehat{\beta}_{\text{Between}}$ and iterating. If these two sequences converge to the same maximum, then this is the global maximum. If one starts with $\widehat{\beta}_{\text{OLS}}$ for $\phi^2 = 1$, and the next iteration obtains a larger ϕ^2, then we have a local maximum at the boundary $\phi^2 = 1$. Maddala (1971) finds that there are at most two maxima for the likelihood $L(\phi^2)$ for $0 < \phi^2 \leq 1$. Hence, we have to guard against one local maximum.

2.5 PREDICTION

Suppose we want to predict S periods ahead for the ith individual. For the GLS model, knowing the variance–covariance structure of the disturbances, Goldberger (1962) showed that the best linear unbiased predictor (BLUP) of $y_{i,T+S}$ is

$$\widehat{y}_{i,T+S} = Z'_{i,T+S}\widehat{\delta}_{\text{GLS}} + w'\Omega^{-1}\widehat{u}_{\text{GLS}} \quad \text{for } s \geq 1 \tag{2.37}$$

where $\widehat{u}_{\text{GLS}} = y - Z\widehat{\delta}_{\text{GLS}}$ and $w = E(u_{i,T+S}u)$. Note that for period $T+S$

$$u_{i,T+S} = \mu_i + \nu_{i,T+S} \tag{2.38}$$

and $w = \sigma^2_\mu(l_i \otimes \iota_T)$ where l_i is the ith column of I_N, i.e., l_i is a vector that has 1 in the ith position and 0 elsewhere. In this case

$$w'\Omega^{-1} = \sigma^2_\mu(l'_i \otimes \iota'_T)\left[\frac{1}{\sigma^2_1}P + \frac{1}{\sigma^2_\nu}Q\right] = \frac{\sigma^2_\mu}{\sigma^2_1}(l'_i \otimes \iota'_T) \tag{2.39}$$

since $(l'_i \otimes \iota'_T)P = (l'_i \otimes \iota'_T)$ and $(l'_i \otimes \iota'_T)Q = 0$. Using (2.39), the typical element of $w'\Omega^{-1}\widehat{u}_{\text{GLS}}$ becomes $((T\sigma^2_\mu/\sigma^2_1)\overline{\widehat{u}}_{i.,\text{GLS}})$ where $\overline{\widehat{u}}_{i.,\text{GLS}} = \sum_{t=1}^{T}\widehat{u}_{it,\text{GLS}}/T$. Therefore, in (2.37), the BLUP for $y_{i,T+S}$ corrects the GLS prediction by a fraction of the mean of the GLS residuals corresponding to that ith individual. This predictor was considered by Taub (1979).

Baillie and Baltagi (1999) consider the practical situation of prediction from the error component regression model when the variance components are not known. They derive both theoretical and simulation evidence as to the relative efficiency of four alternative predictors: (i) an ordinary predictor, based on the optimal predictor given in (2.37), but with MLEs replacing population parameters; (ii) a truncated predictor that ignores the error component correction, given by the last term in (2.37), but uses MLEs for its regression parameters; (iii) a misspecified predictor which uses OLS estimates of the regression parameters; and (iv) a fixed effects predictor which assumes that the individual effects are fixed parameters that can be estimated. The asymptotic formula for MSE prediction are derived for all four predictors. Using numerical and simulation results, these are shown to perform adequately in realistic sample sizes ($N = 50$ and 500 and $T = 10$ and 20). Both the analytical and sampling results show that there are substantial gains in mean square error prediction by using the ordinary predictor instead of the misspecified or the truncated predictors, especially with increasing $\rho = \sigma_\mu^2/(\sigma_\mu^2 + \sigma_\nu^2)$ values. The reduction in MSE is about tenfold for $\rho = 0.9$ and a little more than twofold for $\rho = 0.6$ for various values of N and T. The fixed effects predictor performs remarkably well, being a close second to the ordinary predictor for all experiments. Simulation evidence confirms the importance of taking into account the individual effects when making predictions. The ordinary predictor and the fixed effects predictor outperform the truncated and misspecified predictors and are recommended in practice.

For an application in actuarial science to the problem of predicting future claims of a risk class, given past claims of that and related risk classes, see Frees, Young and Luo (1999). See also Chamberlain and Hirano (1999) who suggest optimal ways of combining an individual's personal earnings history with panel data on the earnings trajectories of other individuals to provide a conditional distribution for this individual's earnings.

2.6 EXAMPLES

2.6.1 Example 1: Grunfeld Investment Equation

Grunfeld (1958) considered the following investment equation:

$$I_{it} = \alpha + \beta_1 F_{it} + \beta_2 C_{it} + u_{it} \tag{2.40}$$

where I_{it} denotes real gross investment for firm i in year t, F_{it} is the real value of the firm (shares outstanding) and C_{it} is the real value of the capital stock. These panel data consist of 10 large US manufacturing firms over 20 years, 1935–54, and are available on the Wiley web site as Grunfeld.fil. This data set, even though dated, is of manageable size for classroom use and has been used by Zellner (1962) and Taylor (1980). Table 2.1 gives the OLS, Between

Table 2.1 Grunfeld's Data. One-way Error Component Results

	β_1	β_2	ρ	σ_μ	σ_ν
OLS	0.116	0.231			
	(0.006)*	(0.025)*			
Between	0.135	0.032			
	(0.029)	(0.191)			
Within	0.110	0.310			
	(0.012)	(0.017)			
WALHUS	0.110	0.308	0.73	87.36	53.75
	(0.011)	(0.017)			
AMEMIYA	0.110	0.308	0.71	83.52	52.77
	(0.010)	(0.017)			
SWAR	0.110	0.308	0.72	84.20	52.77
	(0.010)	(0.017)			
IMLE	0.110	0.308	0.70	80.30	52.49
	(0.010)	(0.017)			

* These are biased standard errors when the true model has error component disturbances (see Moulton, 1986).

and Within estimators for the slope coefficients along with their standard errors. The Between estimates are different from the Within estimates and a Hausman (1978) test based on their difference is given in Chapter 4. OLS and feasible GLS are matrix-weighted combinations of these two estimators. Table 2.1 reports three feasible GLS estimates of the regression coefficients along with the corresponding estimates of ρ, σ_μ and σ_ν. These are WALHUS, AMEMIYA and SWAR. EViews computes the Wallace and Hussain (1969) estimator as an option under the random effects panel data procedure. This EViews output is reproduced in Table 2.2. Similarly, Table 2.3 gives the EViews output for the Amemiya (1971) procedure which is named Wansbeek and Kapteyn (1989) in EViews, since the latter paper generalizes the Amemiya method to deal with unbalanced or incomplete panels, see Chapter 9. Table 2.4 gives the EViews output for the Swamy and Arora (1972) procedure. Note that in Table 2.4, $\widehat{\sigma}_\mu = 84.2$, $\widehat{\sigma}_\nu = 52.77$ and $\widehat{\rho} = \widehat{\sigma}_\mu^2/(\widehat{\sigma}_\mu^2 + \widehat{\sigma}_\nu^2) = 0.72$. This is not $\widehat{\theta}$, but the latter can be obtained as $\widehat{\theta} = 1 - (\widehat{\sigma}_\nu/\widehat{\sigma}_1) = 0.86$. Next, Breusch's (1987) iterative maximum likelihood estimation is performed (IMLE). This procedure converged to a global maximum in three to four iterations depending on whether one started from the Between or Within estimators. There is not much difference among the feasible GLS estimates or the iterative MLE and they are all close to the Within estimates. This is understandable given that $\widehat{\theta}$ for these estimators is close to 1.

Table 2.2 Grunfeld's Data: Wallace and Hussain RE Estimator

Dependent variable: I
Method: Panel EGLS (cross-section random effects)

Sample: 1935 1954
Cross-sections included: 10
Total panel (balanced) observations: 200
Wallace and Hussain estimator of component variances

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−57.86253	29.90492	−1.934883	0.0544
F	0.109789	0.010725	10.23698	0.0000
K	0.308183	0.017498	17.61207	0.0000

Effects Specification

Cross-section random S.D./rho	87.35803	0.7254
Idiosyncratic random S.D./rho	53.74518	0.2746

Weighted Statistics

R-squared	0.769410	Mean dependent variance	19.89203
Adjusted R-squared	0.767069	S.D. dependent variance	109.2808
S.E. of regression	52.74214	Sum squared residual	548001.4
F-statistic	328.6646	Durbin–Watson statistic	0.683829
Prob(F-statistic)	0.000000		

Unweighted Statistics

R-squared	0.803285	Mean dependent variance	145.9582
Sum squared residual	1841243	Durbin–Watson statistic	0.203525

Table 2.3 Grunfeld's Data: Amemiya/Wansbeek and Kapteyn RE Estimator

Dependent variable: I
Method: Panel EGLS (cross-section random effects)

Sample: 1935 1954
Cross-sections included: 10
Total panel (balanced) observations: 200
Wansbeek and Kapteyn estimator of component variances

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−57.82187	28.68562	−2.015710	0.0452
F	0.109778	0.010471	10.48387	0.0000
K	0.308081	0.017172	17.94062	0.0000
		Effects Specification		
Cross-section random S.D./rho			83.52354	0.7147
Idiosyncratic random S.D./rho			52.76797	0.2853

Weighted Statistics

R-squared	0.769544	Mean dependent variance	20.41664
Adjusted R-squared	0.767205	S.D. dependent variance	109.4431
S.E. of regression	52.80503	Sum squared residual	549309.2
F-statistic	328.9141	Durbin–Watson statistic	0.682171
Prob(F-statistic)	0.000000		

Unweighted Statistics

R-squared	0.803313	Mean dependent variance	145.9582
Sum squared residual	1840981	Durbin–Watson statistic	0.203545

2.6.2 Example 2: Gasoline Demand

Baltagi and Griffin (1983) considered the following gasoline demand equation:

$$\ln \frac{\text{Gas}}{\text{Car}} = \alpha + \beta_1 \ln \frac{Y}{N} + \beta_2 \ln \frac{P_{\text{MG}}}{P_{\text{GDP}}} + \beta_3 \ln \frac{\text{Car}}{N} + u \tag{2.41}$$

where Gas/Car is motor gasoline consumption per auto, Y/N is real per capita income, $P_{\text{MG}}/P_{\text{GDP}}$ is real motor gasoline price and Car$/N$ denotes the stock of cars per capita. This panel consists of annual observations across 18 OECD countries, covering the period 1960–78. The data for this example are given as Gasoline.dat on the Wiley web site. Table 2.5 gives the parameter estimates for OLS, Between, Within and three feasible GLS estimates of the slope coefficients along with their standard errors, and the corresponding estimates of ρ, σ_μ and σ_ν. Breusch's (1987) iterative maximum likelihood converged to a global maximum in four to six iterations depending on whether one starts from the Between or Within estimators. For the SWAR procedure, $\widehat{\sigma}_\mu = 0.196$, $\widehat{\sigma}_\nu = 0.092$, $\widehat{\rho} = 0.82$ and $\widehat{\theta} = 0.89$. Once again the estimates of θ are closer to 1 than 0, which explains why feasible GLS is closer to the Within estimator than the OLS estimator. The Between and OLS price elasticity estimates of gasoline demand are more than double those for the Within and feasible GLS estimators.

Table 2.4 Grunfeld's Data: Swamy and Arora RE Estimator

Dependent variable: I
Method: Panel EGLS (cross-section random effects)

Sample: 1935 1954
Cross-sections included: 10
Total panel (balanced) observations: 200
Swamy and Arora estimator of component variances

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−57.83441	28.88930	−2.001932	0.0467
F	0.109781	0.010489	10.46615	0.0000
K	0.308113	0.017175	17.93989	0.0000
		Effects Specification		
Cross-section random S.D./rho			84.20095	0.7180
Idiosyncratic random S.D./rho			52.76797	0.2820

Weighted Statistics

R-squared	0.769503	Mean dependent variance	20.25556
Adjusted R-squared	0.767163	S.D. dependent variance	109.3928
S.E. of regression	52.78556	Sum squared residual	548904.1
F-statistic	328.8369	Durbin–Watson statistic	0.682684
Prob(F-statistic)	0.000000		

Unweighted Statistics

R-squared	0.803304	Mean dependent variance	145.9582
Sum squared residual	1841062	Durbin–Watson statistic	0.203539

Table 2.5 Gasoline Demand Data. One-way Error Component Results

	β_1	β_2	β_3	ρ	σ_μ	σ_ν
OLS	0.890	−0.892	−0.763			
	(0.036)*	(0.030)*	(0.019)*			
Between	0.968	−0.964	−0.795			
	(0.156)	(0.133)	(0.082)			
Within	0.662	−0.322	−0.640			
	(0.073)	(0.044)	(0.030)			
WALHUS	0.545	−0.447	−0.605	0.75	0.197	0.113
	(0.066)	(0.046)	(0.029)			
AMEMIYA	0.602	−0.366	−0.621	0.93	0.344	0.092
	(0.066)	(0.042)	(0.027)			
SWAR	0.555	−0.402	−0.607	0.82	0.196	0.092
	(0.059)	(0.042)	(0.026)			
IMLE	0.588	−0.378	−0.616	0.91	0.292	0.092
	(0.066)	(0.046)	(0.029)			

* These are biased standard errors when the true model has error component disturbances (see Moulton, 1986).
Source: Baltagi and Griffin (1983). Reproduced by permission of Elsevier Science Publishers B.V. (North-Holland).

2.6.3 Example 3: Public Capital Productivity

Following Munnell (1990), Baltagi and Pinnoi (1995) considered the following Cobb–Douglas production function relationship investigating the productivity of public capital in private production:

$$\ln Y = \alpha + \beta_1 \ln K_1 + \beta_3 \ln K_2 + \beta_3 \ln L + \beta_4 \text{ Unemp } + u \tag{2.42}$$

where Y is gross state product, K_1 is public capital which includes highways and streets, water and sewer facilities and other public buildings and structures, K_2 is the private capital stock based on the Bureau of Economic Analysis national stock estimates, L is labor input measured as employment in nonagricultural payrolls, Unemp is the state unemployment rate included to capture business cycle effects. This panel consists of annual observations for 48 contiguous states over the period 1970–86. This data set was provided by Munnell (1990) and is given as Produc.prn on the Wiley web site. Table 2.6 gives the estimates for a one-way error component model. Note that both OLS and the Between estimators report that public capital is productive and significant in the states private production. In contrast, the Within and feasible GLS estimators find that public capital is insignificant. This result was also reported by Holtz-Eakin (1994) who found that after controlling for state-specific effects, the public-sector capital has no role in affecting private production.

Tables 2.7 and 2.8 give the Stata output reproducing the Between and Within estimates in Table 2.6. This is done using the xtreg command with options (,be) for between and (,fe) for fixed effects. Note that the fixed effects regression prints out the F-test for the significance of the state effects at the bottom of the output. This is the F-test described in (2.12). It tests whether all state dummy coefficients are equal and in this case it yields an $F(47,764) = 75.82$ which is statistically significant. This indicates that the state dummies are jointly significant. It also means that the OLS estimates which omit these state dummies suffer from an omission variables problem rendering them biased and inconsistent. Table 2.9 gives the Swamy and Arora (1972) estimate of the random effects model. This is the default option in Stata and is obtained from the xtreg command with option (,re). Finally, Table 2.10 gives the Stata output for the maximum likelihood estimator. These are obtained from the xtreg command with option (,mle).

Table 2.6 Public Capital Productivity Data. One-way Error Component Results

	β_1	β_2	β_3	β_4	ρ	σ_μ	σ_ν
OLS	0.155	0.309	0.594	−0.007			
	(0.017)*	(0.010)*	(0.014)*	(0.001)*			
Between	0.179	0.302	0.576	−0.004			
	(0.072)	(0.042)	(0.056)	(0.010)			
Within	−0.026	0.292	0.768	−0.005			
	(0.029)	(0.025)	(0.030)	(0.001)			
WALHUS	0.006	0.311	0.728	−0.006	0.82	0.082	0.039
	(0.024)	(0.020)	(0.025)	(0.001)			
AMEMIYA	0.002	0.309	0.733	−0.006	0.84	0.088	0.038
	(0.024)	(0.020)	(0.025)	(0.001)			
SWAR	0.004	0.311	0.730	−0.006	0.82	0.083	0.038
	(0.023)	(0.020)	(0.025)	(0.001)			
IMLE	0.003	0.310	0.731	−0.006	0.83	0.085	0.038
	(0.024)	(0.020)	(0.026)	(0.001)			

*These are biased standard errors when the true model has error component disturbances (see Moulton, 1986).

Table 2.7 Public Capital Productivity Data: The Between Estimator

```
. xtreg lny lnk1 lnk2 lnl u, be

Between regression (regression              Number of obs      =       816
  on group means)
Group variable (i) : stid                   Number of groups   =        48

R-sq:  within  = 0.9330                     Obs per group: min =        17
       between = 0.9939                                    avg =      17.0
       overall = 0.9925                                    max =        17

                                            F(4,43)            =   1754.11
sd(u_i + avg(e_i.))= 0.0832062              Prob > F           =    0.0000

---------------------------------------------------------------------
  lny |       Coef.   Std. Err.     t     P>|t|     [95% Conf. Interval]
------+--------------------------------------------------------------
 lnk1 |    .1793651   .0719719    2.49    0.017     .0342199   .3245104
 lnk2 |    .3019542   .0418215    7.22    0.000     .2176132   .3862953
  lnl |    .5761274   .0563746   10.22    0.000     .4624372   .6898176
    u |   -.0038903   .0099084   -0.39    0.697    -.0238724   .0160918
_cons |    1.589444   .2329796    6.82    0.000     1.119596   2.059292
---------------------------------------------------------------------
```

Table 2.8 Public Capital Productivity Data: Fixed Effects Estimator

```
. xtreg lny lnk1 lnk2 lnl u, fe

Fixed-effects (within) regression            Number of obs      =       816
Group variable (i) : stid                    Number of groups   =        48

R-sq:  within  = 0.9413                      Obs per group: min =        17
       between = 0.9921                                     avg =      17.0
       overall = 0.9910                                     max =        17

                                             F(4,764)           =   3064.81
corr(u_i, xb)  = 0.0608                      Prob > F           =    0.0000

---------------------------------------------------------------------
    lny |     Coef.    Std. Err.    t      P>|t|    [95% Conf. Interval]
--------+------------------------------------------------------------
   lnk1 |  -.0261493   .0290016   -0.90    0.368   -.0830815    .0307829
   lnk2 |   .2920067   .0251197   11.62    0.000    .2426949    .3413185
    lnl |   .7681595   .0300917   25.53    0.000    .7090872    .8272318
      u |  -.0052977   .0009887   -5.36    0.000   -.0072387   -.0033568
  _cons |   2.352898   .1748131   13.46    0.000    2.009727    2.696069
--------+------------------------------------------------------------
sigma_u |  .09057293
sigma_e |  .03813705
    rho |   .8494045   (fraction of variance due to u_i)
---------------------------------------------------------------------
F test that all u_i=0:   F(47, 764) =   75.82        Prob > F = 0.0000
```

Table 2.9 Public Capital Productivity Data: Swamy and Arora Estimator

```
. xtreg lny lnk1 lnk2 lnl u, re theta

Random-effects GLS regression               Number of obs      =       816
Group variable (i) : stid                   Number of groups   =        48

R-sq:  within  = 0.9412                     Obs per group: min =        17
       between = 0.9928                                    avg =      17.0
       overall = 0.9917                                    max =        17

Random effects u_i ~ Gaussian               Wald chi2(4)       = 19131.09
corr(u_i, X)       = 0 (assumed)            Prob > chi2        =   0.0000
theta              = 0.8888353

------------------------------------------------------------------------------
     lny |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
    lnk1 |   .0044388   .0234173    0.19   0.850     -.0414583    .0503359
    lnk2 |   .3105483   .0198047   15.68   0.000      .2717317    .3493649
     lnl |   .7296705   .0249202   29.28   0.000      .6808278    .7785132
       u |  -.0061725   .0009073   -6.80   0.000     -.0079507   -.0043942
   _cons |   2.135411   .1334615   16.00   0.000      1.873831     2.39699
---------+--------------------------------------------------------------------
 sigma_u |   .0826905
 sigma_e |  .03813705
     rho |  .82460109   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Table 2.10 Public Capital Productivity Data: The Maximum Likelihood Estimator

```
. xtreg lny lnk1 lnk2 lnl u, mle

Random-effects ML regression                Number of obs      =       816
Group variable (i) : stid                   Number of groups   =        48

Random effects u_i ~ Gaussian               Obs per group: min =        17
                                                           avg =      17.0
                                                           max =        17

                                            LR chi2(4)         =   2412.91
Log likelihood = 1401.9041                  Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      lny |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------+-------------------------------------------------------------------
     lnk1 |   .0031446   .0239185    0.13   0.895   -.0437348      .050024
     lnk2 |    .309811    .020081   15.43   0.000     .270453      .349169
      lnl |   .7313372   .0256936   28.46   0.000    .6809787     .7816957
        u |  -.0061382   .0009143   -6.71   0.000   -.0079302    -.0043462
    _cons |   2.143865   .1376582   15.57   0.000     1.87406     2.413671
----------+-------------------------------------------------------------------
 /sigma_u |    .085162   .0090452    9.42   0.000    .0674337     .1028903
 /sigma_e |   .0380836   .0009735   39.12   0.000    .0361756     .0399916
------------------------------------------------------------------------------
      rho |   .8333481   .0304597                     .7668537     .8861754
------------------------------------------------------------------------------
Likelihood ratio test of sigma_u=0: chibar2(01)= 1149.84 Prob>=
chibar2 = 0.000
```

2.7 SELECTED APPLICATIONS

There are far too many applications of the error component model in economics to be exhaustive and here we only want to refer the reader to a few applications. These include:

(1) Owusu-Gyapong (1986) who studied the strike activity of 60 Canadian manufacturing industries over the period 1967–79.
(2) Cardellichio (1990) who modeled the production behavior of 1147 sawmills in the state of Washington, observed biennially over the period 1972–84.
(3) Behrman and Deolalikar (1990) who estimated the effect of per capita income on the calorie intake using the panel data collected by the International Crops Research Institute for the Semi-Arid Tropics Village level studies in rural south India.
(4) Johnson and Lahiri (1992) who estimated a production function for ambulatory care using panel data on 30 health care centers in New York state over the years 1984–87.
(5) Conway and Kniesner (1992) who used the Panel Study of Income Dynamics to study the sensitivity of male labor supply function estimates to how the wage is measured and how the researcher models individual heterogeneity.
(6) Cornwell and Rupert (1997) who used panel data from the NLSY to show that much of the wage premium normally attributed to marriage is associated with unobservable individual effects that are correlated with marital status and wages.
(7) Lundberg and Rose (2002) who used panel data from the PSID to estimate the effects of children and the differential effects of sons and daughters on men's labor supply and hourly wage rate. Their fixed effects estimates indicate that, on average, a child increases a man's wage rate by 4.2% and his annual hours of work by 38 hours per year.
(8) Glick and Rose (2002) who studied the question of whether leaving a currency union reduces international trade. They used panel data on bilateral trade among 217 countries over the period 1948–97.

2.8 COMPUTATIONAL NOTE

There is no magical software written explicitly for all panel data estimation and testing procedures. For a software review of LIMDEP, RATS, SAS, TSP and GAUSS with special attention to the panel data procedures presented in this book, see Blanchard (1996). My students use SAS or Stata especially when large database management is needed. For hard to program estimation or testing methods, OX and GAUSS have a comparative advantage. Simple panel data estimators can be done with LIMDEP, TSP, EViews or Stata. In fact, the results reported in examples 2.1, 2.2 and 2.3 have been verified using TSP, EViews and Stata. Also, TSP, Stata and EViews use one or all three of these data sets as benchmarks to illustrate these panel methods.

NOTES

1. For example, if we observe $N = 20$ firms over $T = 5$ time periods, Ω will be 100 by 100.
2. See also Searle and Henderson (1979) for a systematic method for deriving the characteristic roots and vectors of Ω for any balanced error component model.
3. It is important to note that once one substitutes OLS or LSDV residuals in (2.21) and (2.22), the resulting estimators of the variance components are no longer unbiased. The degrees of freedom corrections required to make these estimators unbiased involve traces of matrices that depend on the data. These correction terms are given in Wallace and Hussain (1969) and Amemiya (1971), respectively. Alternatively, one can infer these correction terms from the more general unbalanced error component model considered in Chapter 9.

4. One can also apply Rao's (1970, 1972) MINQUE (minimum norm quadratic unbiased estimation) procedure or Henderson's method III as described by Fuller and Battese (1973). These methods are studied in detail in Baltagi (1995, Appendix 3) for the two-way error component model and in Chapter 9 for the unbalanced error component model. Unfortunately, these methods have not been widely used in the empirical economics literature.
5. Berzeg (1979) generalizes the one-way error component model to the case where the individual effects (μ_i) and the remainder disturbances (ν_{it}) are correlated for the same individual i. This specification ensures a nonnegative estimate of the error component variance. This is applied to the estimation of US demand for motor gasoline (see Berzeg, 1982).
6. Hsiao and Sun (2000) argue that fixed versus random effects specification is better treated as an issue of model selection rather than hypothesis testing. They suggest a recursive predictive density ratio as well as the Akaike and Schwartz information criteria for model selection. Monte Carlo results indicate that all three criteria perform well in finite samples. However, the Schwartz criterion was found to be the more reliable of the three.

PROBLEMS

2.1 Prove that $\widetilde{\beta}$ given in (2.7) can be obtained from OLS on (2.5) using results on partitioned inverse. This can easily be obtained using the Frisch–Waugh–Lovell theorem of Davidson and MacKinnon (1993, p. 19). *Hint*: This theorem states that the OLS estimate of β from (2.5) will be identical to the OLS estimate of β from (2.6). Also, the least squares residuals will be the same.

2.2 (a) Using generalized inverse, show that OLS or GLS on (2.6) yields $\widetilde{\beta}$, the Within estimator given in (2.7).
(b) Show that (2.6) satisfies the necessary and sufficient condition for OLS to be equivalent to GLS (see Baltagi, 1989). *Hint*: Show that $\text{var}(Q\nu) = \sigma_\nu^2 Q$ which is positive semidefinite and then use the fact that Q is idempotent and is its own generalized inverse.

2.3 Verify that by stacking the panel as an equation for each individual in (2.13) and performing the Within transformation as in (2.14) one gets the Within estimator as OLS on this system. Verify that the robust asymptotic $\text{var}(\widetilde{\beta})$ is the one given by (2.16).

2.4 (a) Verify (2.17) and check that $\Omega^{-1}\Omega = I$ using (2.18).
(b) Verify that $\Omega^{-1/2}\Omega^{-1/2} = \Omega^{-1}$ using (2.20) and (2.19).
(c) Premultiply y by $\sigma_\nu \Omega^{-1/2}$ from (2.20), and show that the typical element is $y_{it} - \theta \bar{y}_{i.}$ where $\theta = 1 - (\sigma_\nu/\sigma_1)$.

2.5 Using (2.21) and (2.22), show that $E(\widehat{\sigma}_1^2) = \sigma_1^2$, and $E(\widehat{\sigma}_\nu^2) = \sigma_\nu^2$. *Hint*: $E(u'Qu) = E\{\text{tr}(u'Qu)\} = E\{\text{tr}(uu'Q)\} = \text{tr}\{E(uu')Q\} = \text{tr}(\Omega Q)$.

2.6 (a) Show that $\widehat{\widehat{\sigma}}_\nu^2$, given in (2.24) is unbiased for σ_ν^2.
(b) Show that $\widehat{\widehat{\sigma}}_1^2$ given in (2.27) is unbiased for σ_1^2.

2.7 (a) Perform OLS on the system of equations given in (2.28) and show that the resulting estimator is $\widehat{\delta}_{\text{OLS}} = (Z'Z)^{-1}Z'y$.
(b) Perform GLS on the system of equations given in (2.28) and show that the resulting estimator is $\widehat{\delta}_{\text{GLS}} = (Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}y$ where Ω^{-1} is given in (2.19).

2.8 Using the $\text{var}(\widehat{\beta}_{\text{GLS}})$ expression below (2.30) and $\text{var}(\widetilde{\beta}_{\text{Within}}) = \sigma_\nu^2 W_{XX}^{-1}$, show that

$$(\text{var}(\widehat{\beta}_{\text{GLS}}))^{-1} - (\text{var}(\widetilde{\beta}_{\text{Within}}))^{-1} = \phi^2 B_{XX}/\sigma_\nu^2$$

which is positive semidefinite. Conclude that $\text{var}(\widetilde{\beta}_{\text{Within}}) - \text{var}(\widehat{\beta}_{\text{GLS}})$ is positive semidefinite.

2.9 (a) Using the concentrated likelihood function in (2.34), solve $\partial L_C/\partial\phi^2 = 0$ and verify (2.35).
(b) Solve $\partial L_C/\partial\beta = 0$ and verify (2.36).

2.10 (a) For the predictor $y_{i,T+S}$ given in (2.37), compute $E(u_{i,T+S}u_{it})$ for $t = 1, 2, \ldots, T$ and verify that $w = E(u_{i,T+S}u) = \sigma_\mu^2(l_i \otimes \iota_T)$ where l_i is the ith column of I_N.
(b) Verify (2.39) by showing that $(l_i' \otimes \iota_T')P = (l_i' \otimes \iota_T')$.

2.11 Using Grunfeld's data given as Grunfeld.fil on the Wiley web site, reproduce Table 2.1.

2.12 Using the gasoline demand data of Baltagi and Griffin (1983), given as Gasoline.dat on the Wiley web site, reproduce Table 2.5.

2.13 Using the Monte Carlo set-up for the one-way error component model, given in Maddala and Mount (1973), compare the various estimators of the variance components and regression coefficients studied in this chapter.

2.14 For the random one-way error component model given in (2.1) and (2.2), consider the OLS estimator of $\text{var}(u_{it}) = \sigma^2$, which is given by $s^2 = \widehat{u}'_{\text{OLS}}\widehat{u}_{\text{OLS}}/(n - K')$, where $n = NT$ and $K' = K + 1$.
(a) Show that $E(s^2) = \sigma^2 + \sigma_\mu^2[K' - \text{tr}(I_N \otimes J_T)P_x]/(n - K')$.
(b) Consider the inequalities given by Kiviet and Krämer (1992) which state that

$$\begin{aligned} 0 &\leq \text{mean of } (n - K') \text{ smallest roots of } \Omega \leq E(s^2) \\ &\leq \text{mean of } (n - K') \text{ largest roots of } \Omega \leq \text{tr}(\Omega)/(n - K') \end{aligned}$$

where $\Omega = E(uu')$. Show that for the one-way error component model, these bounds are

$$\begin{aligned} 0 &\leq \sigma_\nu^2 + (n - TK')\sigma_\mu^2/(n - K') \leq E(s^2) \leq \sigma_\nu^2 + n\sigma_\mu^2/(n - K') \\ &\leq n\sigma^2/(n - K') \end{aligned}$$

As $n \to \infty$, both bounds tend to σ^2, and s^2 is asymptotically unbiased, irrespective of the particular evolution of X. See Baltagi and Krämer (1994) for a proof of this result.

2.15 Using the public capital productivity data of Munnell (1990), given as Produc.prn on the Wiley web site, reproduce Table 2.6.

2.16 Using the Monte Carlo design of Baillie and Baltagi (1999), compare the four predictors described in Section 2.5.

2.17 *Heteroskedastic fixed effects models*. This is based on problem 96.5.1 in *Econometric Theory* by Baltagi (1996). Consider the fixed effects model

$$y_{it} = \alpha_i + u_{it} \quad i = 1, 2, \ldots, N; t = 1, 2, \ldots, T_i$$

where y_{it} denotes output in industry i at time t and α_i denotes the industry fixed effect. The disturbances u_{it} are assumed to be independent with heteroskedastic variances σ_i^2. Note that the data are unbalanced with different number of observations for each industry.
(a) Show that OLS and GLS estimates of α_i are identical.
(b) Let $\sigma^2 = \sum_{i=1}^N T_i\sigma_i^2/n$, where $n = \sum_{i=1}^N T_i$, be the average disturbance variance. Show that the GLS estimator of σ^2 is unbiased, whereas the OLS estimator of σ^2 is biased. Also show that this bias disappears if the data are balanced or the variances are homoskedastic.

(c) Define $\lambda_i^2 = \sigma_i^2/\sigma^2$ for $i = 1, 2, \ldots, N$. Show that for $\alpha' = (\alpha_1, \alpha_2, \ldots, \alpha_N)$

$$E[\text{estimated var}(\widehat{\alpha}_{\text{OLS}}) - \text{ true var}(\widehat{\alpha}_{\text{OLS}})]$$
$$= \sigma^2[(n - \sum_{i=1}^{N} \lambda_i^2)/(n - N)] \text{ diag } (1/T_i) - \sigma^2 \text{ diag } (\lambda_i^2/T_i)$$

This problem shows that in case there are no regressors in the unbalanced panel data model, fixed effects with heteroskedastic disturbances can be estimated by OLS, but one has to correct the standard errors. See solution 96.5.1 in *Econometric Theory* by Kleiber (1997).

3

The Two-way Error Component Regression Model

3.1 INTRODUCTION

Wallace and Hussain (1969), Nerlove (1971b) and Amemiya (1971), among others, considered the regression model given by (2.1), but with two-way error components disturbances:

$$u_{it} = \mu_i + \lambda_t + \nu_{it} \quad i = 1, \ldots, N;\ t = 1, \ldots, T \tag{3.1}$$

where μ_i denotes the unobservable individual effect discussed in Chapter 2, λ_t denotes the unobservable time effect and ν_{it} is the remainder stochastic disturbance term. Note that λ_t is individual-invariant and it accounts for any time-specific effect that is not included in the regression. For example, it could account for strike year effects that disrupt production; oil embargo effects that disrupt the supply of oil and affect its price; Surgeon General reports on the ill-effects of smoking, or government laws restricting smoking in public places, all of which could affect consumption behavior. In vector form, (3.1) can be written as

$$u = Z_\mu \mu + Z_\lambda \lambda + \nu \tag{3.2}$$

where Z_μ, μ and ν were defined earlier. $Z_\lambda = \iota_N \otimes I_T$ is the matrix of time dummies that one may include in the regression to estimate the λ_t if they are fixed parameters, and $\lambda' = (\lambda_1, \ldots, \lambda_T)$. Note that $Z_\lambda Z'_\lambda = J_N \otimes I_T$ and the projection on Z_λ is $Z_\lambda(Z'_\lambda Z_\lambda)^{-1}Z'_\lambda = \bar{J}_N \otimes I_T$. This last matrix averages the data over individuals, i.e., if we regress y on Z_λ, the predicted values are given by $(\bar{J}_N \otimes I_T)y$ which has typical element $\bar{y}_{.t} = \sum_{i=1}^{N} y_{it}/N$.

3.2 THE FIXED EFFECTS MODEL

If the μ_i *and* λ_t are assumed to be fixed parameters to be estimated and the remainder disturbances stochastic with $\nu_{it} \sim \text{IID}(0, \sigma_\nu^2)$, then (3.1) represents a two-way fixed effects error component model. The X_{it} are assumed independent of the ν_{it} for all i and t. Inference in this case is conditional on the particular N individuals and over the specific time periods observed. Recall that Z_λ, the matrix of time dummies, is $NT \times T$. If N or T is large, there will be too many dummy variables in the regression $\{(N-1) + (T-1)\}$ of them, and this causes an enormous loss in degrees of freedom. In addition, this attenuates the problem of multicollinearity among the regressors. Rather than invert a large $(N + T + K - 1)$ matrix, one can obtain the fixed effects estimates of β by performing the following Within transformation given by Wallace and Hussain (1969):

$$Q = E_N \otimes E_T = I_N \otimes I_T - I_N \otimes \bar{J}_T - \bar{J}_N \otimes I_T + \bar{J}_N \otimes \bar{J}_T \tag{3.3}$$

where $E_N = I_N - \bar{J}_N$ and $E_T = I_T - \bar{J}_T$. This transformation "sweeps" the μ_i and λ_t effects. In fact, $\widetilde{y} = Qy$ has a typical element $\widetilde{y}_{it} = (y_{it} - \bar{y}_{i.} - \bar{y}_{.t} + \bar{y}_{..})$ where $\bar{y}_{..} = \sum_i \sum_t y_{it}/NT$, and one would perform the regression of $\widetilde{y} = Qy$ on $\widetilde{X} = QX$ to get the Within estimator $\widetilde{\beta} = (X'QX)^{-1}X'Qy$.

Note that by averaging the simple regression given in (2.8) over individuals, we get

$$\bar{y}_{.t} = \alpha + \beta\bar{x}_{.t} + \lambda_t + \bar{\nu}_{.t} \tag{3.4}$$

where we have utilized the restriction that $\sum_i \mu_i = 0$ to avoid the dummy variable trap. Similarly the averages defined in (2.9) and (2.11) still hold using $\sum_t \lambda_t, = 0$, and one can deduce that

$$(y_{it} - \bar{y}_{i.} - \bar{y}_{.t} + \bar{y}_{..}) = (x_{it} - \bar{x}_{i.} - \bar{x}_{.t} + \bar{x}_{..})\beta + (\nu_{it} - \bar{\nu}_{i.} - \bar{\nu}_{.t} + \bar{\nu}_{..}) \tag{3.5}$$

OLS on this model gives $\widetilde{\beta}$, the Within estimator for the two-way model. Once again, the Within estimate of the intercept can be deduced from $\widetilde{\alpha} = \bar{y}_{..} - \widetilde{\beta}\bar{x}_{..}$ and those of μ_i and λ_t are given by

$$\widetilde{\mu}_i = (\bar{y}_{i.} - \bar{y}_{..}) - \widetilde{\beta}(\bar{x}_{i.} - \bar{x}_{..}) \tag{3.6}$$

$$\widetilde{\lambda}_t = (\bar{y}_{.t} - \bar{y}_{..}) - \widetilde{\beta}(\bar{x}_{.t} - \bar{x}_{..}) \tag{3.7}$$

Note that the Within estimator cannot estimate the effect of time-invariant and individual-invariant variables because the Q transformation wipes out these variables. If the true model is a two-way fixed effects model as in (3.2), then OLS on (2.1) yields biased and inconsistent estimates of the regression coefficients. OLS ignores both sets of dummy variables, whereas the one-way fixed effects estimator considered in Chapter 2 ignores only the time dummies. If these time dummies are statistically significant, the one-way fixed effects estimator will also suffer from omission bias.

3.2.1 Testing for Fixed Effects

As in the one-way error component model case, one can test for joint significance of the dummy variables:

$$H_0 : \mu_1 = \ldots = \mu_{N-1} = 0 \quad \text{and} \quad \lambda_1 = \ldots = \lambda_{T-1} = 0$$

The restricted residual sums of squares (RRSS) is that of pooled OLS and the unrestricted residual sums of squares (URSS) is that from the Within regression in (3.5). In this case,

$$F_1 = \frac{(\text{RRSS} - \text{URSS})/(N + T - 2)}{\text{URSS}/(N - 1)(T - 1) - K} \overset{H_0}{\sim} F_{(N+T-2),(N-1)(T-1)-K} \tag{3.8}$$

Next, one can test for the existence of individual effects allowing for time effects, i.e.

$$H_2 : \mu_1 = \ldots = \mu_{N-1} = 0 \quad \text{allowing} \quad \lambda_t \neq 0 \quad \text{for} \quad t = 1, \ldots, T - 1$$

The URSS is still the Within residual sum of squares. However, the RRSS is the regression with time-series dummies only, or the regression based upon

$$(y_{it} - \bar{y}_{.t}) = (x_{it} - \bar{x}_{.t})\beta + (u_{it} - \bar{u}_{.t}) \tag{3.9}$$

In this case the resulting F-statistic is $F_2 \overset{H_0}{\sim} F_{(N-1),(N-1)(T-1)-K}$. Note that F_2 differs from F_0 in (2.12) in testing for $\mu_i = 0$. The latter tests $H_0 : \mu_i = 0$ assuming that $\lambda_t = 0$, whereas the former tests $H_2 : \mu_i = 0$ allowing $\lambda_t \neq 0$ for $t = 1, \ldots, T - 1$. Similarly, one can test for the existence of time effects allowing for individual effects, i.e.

$$H_3 : \lambda_1 = \ldots = \lambda_{T-1} = 0 \quad \text{allowing} \quad \mu_i \neq 0;\ i = 1, \ldots, (N - 1)$$

The RRSS is given by the regression in (2.10), while the URSS is obtained from the regression (3.5). In this case, the resulting F-statistic is $F_3 \overset{H_0}{\sim} F_{(T-1),(N-1)(T-1)-K}$.

Computational Warning

As in the one-way model, s^2 from the regression in (3.5) as obtained from any standard regression package has to be adjusted for loss of degrees of freedom. In this case, one divides by $(N-1)(T-1)-K$ and multiplies by $(NT-K)$ to get the proper variance–covariance matrix of the Within estimator.

3.3 THE RANDOM EFFECTS MODEL

If $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$, $\lambda_t \sim \text{IID}(0, \sigma_\lambda^2)$ and $\nu_{it} \sim \text{IID}(0, \sigma_\nu^2)$ independent of each other, then this is the two-way *random* effects model. In addition, X_{it} is independent of μ_i, λ_t and ν_{it} for all i and t. Inference in this case pertains to the large population from which this sample was randomly drawn. From (3.2), one can compute the variance–covariance matrix

$$\begin{aligned}\Omega &= E(uu') = Z_\mu E(\mu\mu')Z'_\mu + Z_\lambda E(\lambda\lambda')Z'_\lambda + \sigma_\nu^2 I_{NT} \\ &= \sigma_\mu^2(I_N \otimes J_T) + \sigma_\lambda^2(J_N \otimes I_T) + \sigma_\nu^2(I_N \otimes I_T)\end{aligned} \tag{3.10}$$

The disturbances are homoskedastic with $\text{var}(u_{it}) = \sigma_\mu^2 + \sigma_\lambda^2 + \sigma_\nu^2$ for all i and t,

$$\begin{aligned}\text{cov}(u_{it}, u_{js}) &= \sigma_\mu^2 \quad i = j,\ t \neq s \\ &= \sigma_\lambda^2 \quad i \neq j,\ t = s\end{aligned} \tag{3.11}$$

and zero otherwise. This means that the correlation coefficient

$$\begin{aligned}\text{correl}(u_{it}, u_{js}) &= \sigma_\mu^2/(\sigma_\mu^2 + \sigma_\lambda^2 + \sigma_\nu^2) \quad & i = j,\ t \neq s \\ &= \sigma_\lambda^2/(\sigma_\mu^2 + \sigma_\lambda^2 + \sigma_\nu^2) \quad & i \neq j,\ t = s \\ &= 1 \quad & i = j,\ t = s \\ &= 0 \quad & i \neq j,\ t \neq s\end{aligned} \tag{3.12}$$

In order to get Ω^{-1}, we replace J_N by $N\bar{J}_N$, I_N by $E_N + \bar{J}_N$, J_T by $T\bar{J}_T$ and I_T by $E_T + \bar{J}_T$ and collect terms with the same matrices. This gives

$$\Omega = \sum_{i=1}^{4} \lambda_i Q_i \tag{3.13}$$

where $\lambda_1 = \sigma_\nu^2$, $\lambda_2 = T\sigma_\mu^2 + \sigma_\nu^2$, $\lambda_3 = N\sigma_\lambda^2 + \sigma_\nu^2$ and $\lambda_4 = T\sigma_\mu^2 + N\sigma_\lambda^2 + \sigma_\nu^2$. Correspondingly, $Q_1 = E_N \otimes E_T$, $Q_2 = E_N \otimes \bar{J}_T$, $Q_3 = \bar{J}_N \otimes E_T$ and $Q_4 = \bar{J}_N \otimes \bar{J}_T$, respectively. The λ_i are the distinct characteristic roots of Ω and the Q_i are the corresponding matrices of eigenprojectors. λ_1 is of multiplicity $(N-1)(T-1)$, λ_2 is of multiplicity $(N-1)$, λ_3 is of multiplicity $(T-1)$ and λ_4 is of multiplicity 1.[1] Each Q_i is symmetric and idempotent with its rank equal to its trace. Moreover, the Q_i are pairwise orthogonal and sum to the identity matrix. The advantages of this spectral decomposition are that

$$\Omega^r = \sum_{i=1}^{4} \lambda_i^r Q_i \tag{3.14}$$

where r is an arbitrary scalar so that

$$\sigma_\nu \Omega^{-1/2} = \sum_{i=1}^{4} (\sigma_\nu / \lambda_i^{1/2}) Q_i \tag{3.15}$$

and the typical element of $y^* = \sigma_\nu \Omega^{-1/2} y$ is given by

$$y_{it}^* = y_{it} - \theta_1 \bar{y}_{i.} - \theta_2 \bar{y}_{.t} + \theta_3 \bar{y}_{..} \tag{3.16}$$

where $\theta_1 = 1 - (\sigma_\nu / \lambda_2^{1/2})$, $\theta_2 = 1 - (\sigma_\nu / \lambda_3^{1/2})$ and $\theta_3 = \theta_1 + \theta_2 + (\sigma_\nu / \lambda_4^{1/2}) - 1$. As a result, GLS can be obtained as OLS of y^* on Z^*, where $Z^* = \sigma_\nu \Omega^{-1/2} Z$. This transformation was first derived by Fuller and Battese (1974), see also Baltagi (1993).

The best quadratic unbiased (BQU) estimators of the variance components arise naturally from the fact that $Q_i u \sim (0, \lambda_i Q_i)$. Hence,

$$\widehat{\lambda}_i = u' Q_i u / \mathrm{tr}(Q_i) \tag{3.17}$$

is the BQU estimator of λ_i for $i = 1, 2, 3$. These ANOVA estimators are minimum variance unbiased (MVU) under normality of the disturbances (see Graybill, 1961). As in the one-way error component model, one can obtain feasible estimates of the variance components by replacing the true disturbances by OLS residuals (see Wallace and Hussain, 1969). OLS is still an unbiased and consistent estimator under the random effects model, but it is inefficient and results in biased standard errors and t-statistics. Alternatively, one could substitute the Within residuals with $\widetilde{u} = y - \widetilde{\alpha}\iota_{NT} - X\widetilde{\beta}$, where $\widetilde{\alpha} = \bar{y}_{..} - \bar{X}'_{..}\widetilde{\beta}$ and $\widetilde{\beta}$ is obtained by the regression in (3.5). This is the method proposed by Amemiya (1971). In fact, Amemiya (1971) shows that the Wallace and Hussain (1969) estimates of the variance components have a different asymptotic distribution from that knowing the true disturbances, while the Amemiya (1971) estimates of the variance components have the same asymptotic distribution as that knowing the true disturbances:

$$\begin{pmatrix} \sqrt{NT}(\widehat{\sigma}_\nu^2 - \sigma_\nu^2) \\ \sqrt{N}(\widehat{\sigma}_\mu^2 - \sigma_\mu^2) \\ \sqrt{T}(\widehat{\sigma}_\lambda^2 - \sigma_\lambda^2) \end{pmatrix} \sim N \left(0, \begin{pmatrix} 2\sigma_\nu^4 & 0 & 0 \\ 0 & 2\sigma_\mu^4 & 0 \\ 0 & 0 & 2\sigma_\lambda^4 \end{pmatrix} \right) \tag{3.18}$$

Substituting OLS or Within residuals instead of the true disturbances in (3.17) introduces bias in the corresponding estimates of the variance components. The degrees of freedom corrections that make these estimates unbiased depend upon traces of matrices that involve the matrix of regressors X. These corrections are given in Wallace and Hussain (1969) and Amemiya (1971), respectively. Alternatively, one can infer these correction terms from the more general unbalanced error component model considered in Chapter 9.

Swamy and Arora (1972) suggest running three least squares regressions and estimating the variance components from the corresponding mean square errors of these regressions. The first regression corresponds to the Within regression which transforms the original model by $Q_1 = E_N \otimes E_T$. This is equivalent to the regression in (3.5), and yields the following estimate of σ_ν^2:

$$\widehat{\widehat{\lambda}}_1 = \widehat{\widehat{\sigma}}_\nu^2 = [y' Q_1 y - y' Q_1 X (X' Q_1 X)^{-1} X' Q_1 y] / [(N-1)(T-1) - K] \tag{3.19}$$

The second regression is the Between individuals regression which transforms the original model by $Q_2 = E_N \otimes \bar{J}_T$. This is equivalent to the regression of $(\bar{y}_{i.} - \bar{y}_{..})$ on $(\bar{X}_{i.} - \bar{X}_{..})$ and yields the following estimate of $\lambda_2 = T\sigma_\mu^2 + \sigma_\nu^2$:

$$\widehat{\widehat{\lambda}}_2 = [y'Q_2y - y'Q_2X(X'Q_2X)^{-1}X'Q_2y]/[(N-1)-K] \tag{3.20}$$

from which one obtains $\widehat{\widehat{\sigma}}^2_\mu = (\widehat{\widehat{\lambda}}_2 - \widehat{\widehat{\sigma}}^2_\nu)/T$. The third regression is the Between time-periods regression which transforms the original model by $Q_3 = \bar{J}_N \otimes E_T$. This is equivalent to the regression of $(\bar{y}_{.t} - \bar{y}_{..})$ on $(\bar{X}_{.t} - \bar{X}_{..})$ and yields the following estimate of $\lambda_3 = N\sigma_\lambda^2 + \sigma_\nu^2$:

$$\widehat{\widehat{\lambda}}_3 = [y'Q_3y - y'Q_3X(X'Q_3X)^{-1}X'Q_3y]/[(T-1)-K] \tag{3.21}$$

from which one obtains $\widehat{\widehat{\sigma}}^2_\lambda = (\widehat{\widehat{\lambda}}_3 - \widehat{\widehat{\sigma}}^2_\nu)/N$. Stacking the three transformed regressions just performed yields

$$\begin{pmatrix} Q_1y \\ Q_2y \\ Q_3y \end{pmatrix} = \begin{pmatrix} Q_1X \\ Q_2X \\ Q_3X \end{pmatrix} \beta + \begin{pmatrix} Q_1u \\ Q_2u \\ Q_3u \end{pmatrix} \tag{3.22}$$

since $Q_i \iota_{NT} = 0$ for $i = 1, 2, 3$, and the transformed error has mean 0 and variance–covariance matrix given by $\mathrm{diag}[\lambda_i Q_i]$ with $i = 1, 2, 3$. Problem 3.4 asks the reader to show that OLS on this system of $3NT$ observations yields the same estimator of β as OLS on the pooled model (2.3). Also, GLS on this system of equations (3.22) yields the same estimator of β as GLS on (2.3). In fact,

$$\begin{aligned} \widehat{\beta}_{\mathrm{GLS}} &= [(X'Q_1X)/\sigma_\nu^2 + (X'Q_2X)/\lambda_2 + (X'Q_3X)/\lambda_3]^{-1} \\ &\quad \times[(X'Q_1y)/\sigma_\nu^2 + (X'Q_2y)/\lambda_2 + (X'Q_3y)/\lambda_3] \\ &= [W_{XX} + \phi_2^2 B_{XX} + \phi_3^2 C_{XX}]^{-1}[W_{Xy} + \phi_2^2 B_{Xy} + \phi_3^2 C_{Xy}] \end{aligned} \tag{3.23}$$

with $\mathrm{var}(\widehat{\beta}_{\mathrm{GLS}}) = \sigma_\nu^2[W_{XX} + \phi_2^2 B_{XX} + \phi_3^2 C_{XX}]^{-1}$. Note that $W_{XX} = X'Q_1X$, $B_{XX} = X'Q_2X$ and $C_{XX} = X'Q_3X$ with $\phi_2^2 = \sigma_\nu^2/\lambda_2$, $\phi_3^2 = \sigma_\nu^2/\lambda_3$. Also, the Within estimator of β is $\widetilde{\beta}_W = W_{XX}^{-1}W_{Xy}$, the Between individuals estimator of β is $\widehat{\beta}_B = B_{XX}^{-1}B_{Xy}$ and the Between time-periods estimator of β is $\widehat{\beta}_C = C_{XX}^{-1}C_{Xy}$. This shows that $\widehat{\beta}_{\mathrm{GLS}}$ is a matrix-weighted average of $\widetilde{\beta}_W$, $\widehat{\beta}_B$ and $\widehat{\beta}_C$. In fact,

$$\widehat{\beta}_{\mathrm{GLS}} = W_1\widetilde{\beta}_W + W_2\widehat{\beta}_B + W_3\widehat{\beta}_C \tag{3.24}$$

where

$$\begin{aligned} W_1 &= [W_{XX} + \phi_2^2 B_{XX} + \phi_3^2 C_{XX}]^{-1} W_{XX} \\ W_2 &= [W_{XX} + \phi_2^2 B_{XX} + \phi_3^2 C_{XX}]^{-1} (\phi_2^2 B_{XX}) \\ W_3 &= [W_{XX} + \phi_2^2 B_{XX} + \phi_3^2 C_{XX}]^{-1} (\phi_3^2 C_{XX}) \end{aligned}$$

This was demonstrated by Maddala (1971). Note that (i) if $\sigma_\mu^2 = \sigma_\lambda^2 = 0$, $\phi_2^2 = \phi_3^2 = 1$ and $\widehat{\beta}_{\mathrm{GLS}}$ reduces to $\widehat{\beta}_{\mathrm{OLS}}$; (ii) as T and $N \to \infty$, ϕ_2^2 and $\phi_3^2 \to 0$ and $\widehat{\beta}_{\mathrm{GLS}}$ tends to $\widetilde{\beta}_W$; (iii) if $\phi_2^2 \to \infty$ with ϕ_3^2 finite, then $\widehat{\beta}_{\mathrm{GLS}}$ tends to $\widehat{\beta}_B$; (iv) if $\phi_3^2 \to \infty$ with ϕ_2^2 finite, then $\widehat{\beta}_{\mathrm{GLS}}$ tends to $\widehat{\beta}_C$.

Wallace and Hussain (1969) compare $\widehat{\beta}_{\mathrm{GLS}}$ and $\widetilde{\beta}_{\mathrm{Within}}$ in the case of nonstochastic (repetitive) X and find that both are (i) asymptotically normal, (ii) consistent and unbiased and that

(iii) $\widehat{\beta}_{\text{GLS}}$ has a smaller generalized variance (i.e. more efficient) in finite samples. In the case of nonstochastic (nonrepetitive) X they find that both $\widehat{\beta}_{\text{GLS}}$ and $\widetilde{\beta}_{\text{Within}}$ are consistent, asymptotically unbiased and have equivalent asymptotic variance–covariance matrices, as both N and $T \to \infty$. The last statement can be proved as follows: the limiting variance of the GLS estimator is

$$\frac{1}{NT} \lim_{\substack{N\to\infty \\ T\to\infty}} (X'\Omega^{-1}X/NT)^{-1} = \frac{1}{NT} \lim_{\substack{N\to\infty \\ T\to\infty}} \left[\sum_{i=1}^{3} \frac{1}{\lambda_i} (X'Q_i X/NT) \right]^{-1} \tag{3.25}$$

but the limit of the inverse is the inverse of the limit, and

$$\lim_{\substack{N\to\infty \\ T\to\infty}} \frac{X'Q_i X}{NT} \quad \text{for } i = 1, 2, 3 \tag{3.26}$$

all exist and are positive semidefinite, since $\lim_{\substack{N\to\infty \\ T\to\infty}} (X'X/NT)$ is assumed finite and positive definite. Hence

$$\lim_{\substack{N\to\infty \\ T\to\infty}} \frac{1}{(N\sigma_\lambda^2 + \sigma_\nu^2)} \left(\frac{X'Q_3 X}{NT} \right) = 0$$

and

$$\lim_{\substack{N\to\infty \\ T\to\infty}} \frac{1}{(T\sigma_\mu^2 + \sigma_\nu^2)} \left(\frac{X'Q_2 X}{NT} \right) = 0$$

Therefore the limiting variance of the GLS estimator becomes

$$\frac{1}{NT} \lim_{\substack{N\to\infty \\ T\to\infty}} \sigma_\nu^2 \left(\frac{X'Q_1 X}{NT} \right)^{-1}$$

which is the limiting variance of the Within estimator.

One can extend Nerlove's (1971a) method for the one-way model, by estimating σ_μ^2 as $\sum_{i=1}^{N}(\widehat{\mu}_i - \overline{\widehat{\mu}})^2/(N-1)$ and σ_λ^2 as $\sum_{t=1}^{T}(\widehat{\lambda}_t - \overline{\widehat{\lambda}})^2/(T-1)$ where the $\widehat{\mu}_i$ and $\widehat{\lambda}_t$ are obtained as coefficients from the least squares dummy variables regression (LSDV). σ_ν^2 is estimated from the Within residual sums of squares divided by NT. Baltagi (1995, appendix 3) develops two other methods of estimating the variance components. The first is Rao's (1970) minimum norm quadratic unbiased estimation (MINQUE) and the second is Henderson's method III as described by Fuller and Battese (1973). These methods require more notation and development and may be skipped in a brief course on this subject. Chapter 9 studies these estimation methods in the context of an unbalanced error component model.

Baltagi (1981a) performed a Monte Carlo study on a simple regression equation with two-way error component disturbances and studied the properties of the following estimators: OLS, the Within estimator and six feasible GLS estimators denoted by WALHUS, AMEMIYA, SWAR, MINQUE, FUBA and NERLOVE corresponding to the methods developed by Wallace and Hussain (1969), Amemiya (1971), Swamy and Arora (1972), Rao (1972), Fuller and Battese (1974) and Nerlove (1971a), respectively. The mean square error of these estimators was computed relative to that of true GLS, i.e. GLS knowing the true variance components.

To review some of the properties of these estimators: OLS is unbiased, but asymptotically inefficient, and its standard errors are biased; see Moulton (1986) for the extent of this bias in empirical applications. In contrast, the Within estimator is unbiased whether or not prior information about the variance components is available. It is also asymptotically equivalent to the GLS estimator in case of weakly nonstochastic exogenous variables. Early in the literature, Wallace and Hussain (1969) recommended the Within estimator for the practical researcher, based on theoretical considerations but more importantly for its ease of computation. In Wallace and Hussain's (1969, p. 66) words the "covariance estimators come off with a surprisingly clear bill of health". True GLS is BLUE, but the variance components are usually not known and have to be estimated. All of the feasible GLS estimators considered are asymptotically efficient. In fact, Prucha (1984) showed that as long as the estimate of σ_ν^2 is consistent, and the probability limits of the estimates σ_μ^2 and σ_λ^2 are finite, the corresponding feasible GLS estimator is asymptotically efficient. Also, Swamy and Arora (1972) proved the existence of a family of asymptotically efficient two-stage feasible GLS estimators of the regression coefficients. Therefore, based on asymptotics only, one cannot differentiate among these two-stage GLS estimators. This leaves undecided the question of which estimator is the best to use. Some analytical results were obtained by Swamy (1971) and Swamy and Arora (1972). These studies derived the relative efficiencies of (i) SWAR with respect to OLS, (ii) SWAR with respect to Within and (iii) Within with respect to OLS. Then, for various values of N, T, the variance components, the Between groups, Between time-periods and Within groups sums of squares of the independent variable, they tabulated these relative efficiency values (see Swamy, 1971, chapters II and III; Swamy and Arora, 1972, p. 272). Among their basic findings is the fact that, for small samples, SWAR is less efficient than OLS if σ_μ^2 and σ_λ^2 are small. Also, SWAR is less efficient than Within if σ_μ^2 and σ_λ^2 are large. The latter result is disconcerting, since Within, which uses only a part of the available data, is more efficient than SWAR, a feasible GLS estimator, which uses all of the available data.

3.3.1 Monte Carlo Experiment

Baltagi (1981a) considered the following simple regression equation:

$$y_{it} = \alpha + \beta x_{it} + u_{it} \tag{3.27}$$

with

$$u_{it} = \mu_i + \lambda_t + \nu_{it} \quad i = 1, \ldots, N;\ t = 1, \ldots, T \tag{3.28}$$

The exogenous variable x was generated by a similar method to that of Nerlove (1971a). Throughout the experiment $\alpha = 5$, $\beta = 0.5$, $N = 25$, $T = 10$ and $\sigma^2 = 20$. However, $\rho = \sigma_\mu^2/\sigma^2$ and $\omega = \sigma_\lambda^2/\sigma^2$ were varied over the set $(0, 0.01, 0.2, 0.4, 0.6, 0.8)$ such that $(1 - \rho - \omega)$ is always positive. In each experiment 100 replications were performed. For every replication $(NT + N + T)$ independent and identically distributed normal IIN(0, 1) random numbers were generated. The first N numbers were used to generate the μ_i as IIN$(0, \sigma_\mu^2)$. The second T numbers were used to generate the λ_t as IIN$(0, \sigma_\lambda^2)$ and the last NT numbers were used to generate the ν_{it} as IIN$(0, \sigma_\nu^2)$. For the estimation methods considered, the Monte Carlo results show the following:

(1) For the two-way model, the researcher should not label the problem of negative variance estimates "not serious" as in the one-way model. This is because we cannot distinguish between the case where the model is misspecified (i.e. with at least one of the variance components actually equal to zero) and the case where the model is properly specified (i.e. with at least one of the variance components relatively small but different from zero). Another important

reason is that we may not be able to distinguish between a case where OLS is equivalent to GLS according to the MSE criterion and a case where it is not. For these cases, the practical solution seems to be the replacement of a negative estimate by zero. Of course, this will affect the properties of the variance components estimates, especially if the actual variances are different from zero. The Monte Carlo results of Baltagi (1981a) report that the performance of the two-stage GLS methods is not seriously affected by this substitution.

(2) As long as the variance components are not relatively small and close to zero, there is always gain according to the MSE criterion in performing feasible GLS rather than least squares or least squares with dummy variables.

(3) All the two-stage GLS methods considered performed reasonably well according to the relative MSE criteria. However, none of these methods could claim to be the best for all the experiments performed. Most of these methods had relatively close MSEs which therefore made it difficult to choose among them. This same result was obtained in the one-way model by Maddala and Mount (1973).

(4) Better estimates of the variance components do not necessarily give better second-round estimates of the regression coefficients. This confirms the finite sample results obtained by Taylor (1980) and extends them from the one-way to the two-way model.

Finally, the recommendation given in Maddala and Mount (1973) is still valid, i.e. always perform more than one of the two-stage GLS procedures to see whether the estimates obtained differ widely.

3.4 MAXIMUM LIKELIHOOD ESTIMATION

In this case, the normality assumption is needed on our error structure. The loglikelihood function is given by

$$\log L = \text{ constant } - \frac{1}{2}\log \mid \Omega \mid - \frac{1}{2}(y - Z\gamma)'\Omega^{-1}(y - Z\gamma) \tag{3.29}$$

where Ω and Ω^{-1} were given in (3.13) and (3.14). The maximum likelihood estimators of $\gamma, \sigma_\nu^2, \sigma_\mu^2$ and σ_λ^2 are obtained by simultaneously solving the following normal equations:

$$\begin{aligned}
\frac{\partial \log L}{\partial \gamma} &= Z'\Omega^{-1}y - (Z'\Omega^{-1}Z)\gamma = 0\\
\frac{\partial \log L}{\partial \sigma_\nu^2} &= -\frac{1}{2}\operatorname{tr}\Omega^{-1} + \frac{1}{2}u'\Omega^{-2}u = 0\\
\frac{\partial \log L}{\partial \sigma_\mu^2} &= -\frac{1}{2}\operatorname{tr}\Omega^{-1}(I_N \otimes J_T) + \frac{1}{2}u'\Omega^{-2}(I_N \otimes J_T)u = 0\\
\frac{\partial \log L}{\partial \sigma_\lambda^2} &= -\frac{1}{2}\operatorname{tr}\Omega^{-1}(J_N \otimes I_T) + \frac{1}{2}u'\Omega^{-2}(J_N \otimes I_T)u = 0
\end{aligned} \tag{3.30}$$

Even if the u were observable, these would still be highly nonlinear and difficult to solve explicitly. However, Amemiya (1971) suggests an iterative scheme to solve (3.30). The resulting maximum likelihood estimates of the variance components are shown to be consistent and asymptotic normal with an asymptotic distribution given by (3.18).

Following Breusch (1987) one can write the likelihood for the two-way model as

$$L(\alpha, \beta, \sigma_\nu^2, \phi_2^2, \phi_3^2) = \text{constant} - (NT/2)\log \sigma_\nu^2 + (N/2)\log \phi_2^2 + (T/2)\log \phi_3^2 \\ -(1/2)\log[\phi_2^2 + \phi_3^2 - \phi_2^2\phi_3^2] - (1/2\sigma_\nu^2)u'\Sigma^{-1}u \tag{3.31}$$

where $\Omega = \sigma_\nu^2\Sigma = \sigma_\nu^2(\sum_{i=1}^4 Q_i/\phi_i^2)$ from (3.13) with $\phi_i^2 = \sigma_\nu^2/\lambda_i$ for $i = 1, \ldots, 4$. The likelihood (3.31) uses the fact that $| \Omega |^{-1} = (\sigma_\nu^2)^{-NT}(\phi_2^2)^{N-1}(\phi_3^2)^{T-1}\phi_4^2$. The feasibility conditions $\infty > \lambda_i \geq \sigma_\nu^2$ are equivalent to $0 < \phi_i^2 \leq 1$ for $i = 1, 2, 3, 4$. Following Breusch (1987), we define $d = y - X\beta$, therefore $u = d - \iota_{NT}\alpha$. Given arbitrary values of $\beta, \phi_2^2, \phi_3^2$, one can concentrate this likelihood function with respect to α and σ_ν^2. Estimates of α and σ_ν^2 are obtained later as $\widehat{\alpha} = \iota_{NT}'d/NT$ and $\widehat{\sigma}_\nu^2 = (u'\Sigma^{-1}u/NT)$. Substituting the maximum value of α in u one gets $u = d - \iota_{NT}\widehat{\alpha} = (I_{NT} - \bar{J}_{NT})d$. Also, using the fact that

$$(I_{NT} - \bar{J}_{NT})\Sigma^{-1}(I_{NT} - \bar{J}_{NT}) = Q_1 + \phi_2^2 Q_2 + \phi_3^2 Q_3$$

one gets $\widehat{\sigma}_\nu^2 = d'[Q_1 + \phi_2^2 Q_2 + \phi_3^2 Q_3]d/NT$, given β, ϕ_2^2 and ϕ_3^2. The concentrated likelihood function becomes

$$L_C(\beta, \phi_2^2, \phi_3^2) = \text{constant} - (NT/2)\log[d'(Q_1 + \phi_2^2 Q_2 + \phi_3^2 Q_3)d] \\ +(N/2)\log \phi_2^2 + (T/2)\log \phi_3^2 - (1/2)\log[\phi_2^2 + \phi_3^2 - \phi_2^2\phi_3^2] \tag{3.32}$$

Maximizing L_C over β, given ϕ_2^2 and ϕ_3^2, Baltagi and Li (1992a) get

$$\widehat{\beta} = [X'(Q_1 + \phi_2^2 Q_2 + \phi_3^2 Q_3)X]^{-1}X'(Q_1 + \phi_2^2 Q_2 + \phi_3^2 Q_3)y \tag{3.33}$$

which is the GLS estimator knowing ϕ_2^2 and ϕ_3^2. Similarly, maximizing L_C over ϕ_2^2, given β and ϕ_3^2, one gets[2]

$$\frac{\delta L_C}{\delta \phi_2^2} = -\frac{NT}{2}\frac{d'Q_2 d}{d'[Q_1 + \phi_2^2 Q_2 + \phi_3^2 Q_3]d} + \frac{N}{2}\frac{1}{\phi_2^2} - \frac{1}{2}\frac{(1 - \phi_3^2)}{[\phi_2^2 + \phi_3^2 - \phi_2^2\phi_3^2]} = 0 \tag{3.34}$$

which can be written as

$$a\phi_2^4 + b\phi_2^2 + c = 0 \tag{3.35}$$

where $a = -[N(T-1)+1](1-\phi_3^2)(d'Q_2 d)$, $b = (1-\phi_3^2)(N-1)d'[Q_1 + \phi_3^2 Q_3]d - \phi_3^2 (T-1)N(d'Q_2 d)$ and $c = N\phi_3^2 d'[Q_1 + \phi_3^2 Q_3]d$. We will *fix* ϕ_3^2, where $(0 < \phi_3^2 < 1)$ and focus on iterating between β and ϕ_2^2.[3] For a fixed ϕ_3^2, if $\phi_2^2 = 0$, then (3.33) becomes $\widehat{\beta}_{BW} = [X'(Q_1 + \phi_3^2 Q_3)X]^{-1}X'(Q_1 + \phi_3^2 Q_3)y$, which is a matrix-weighted average of the Within estimator $\widehat{\beta}_W = (X'Q_1X)^{-1}X'Q_1y$ and the Between time-periods estimator $\widehat{\beta}_C = (X'Q_3X)^{-1}X'Q_3y$. If $\phi_2^2 \to \infty$, with ϕ_3^2 fixed, then (3.33) reduces to the Between individuals estimator $\widehat{\beta}_B = (X'Q_2X)^{-1}X'Q_2y$. Using standard assumptions, Baltagi and Li (1992a) show that $a < 0$ and $c > 0$ in (3.35). Hence $b^2 - 4ac > b^2 > 0$, and the unique positive root of (3.35) is

$$\widehat{\phi_2^2} = \left[-b - \sqrt{b^2 - 4ac}\right]/2a = \left[b + \sqrt{b^2 + 4 \mid a \mid c}\right]/2 \mid a \mid \tag{3.36}$$

Since ϕ_3^2 is fixed, we let $\bar{Q}_1 = Q_1 + \phi_3^2 Q_3$, then (3.33) becomes

$$\widehat{\beta} = [X'(\bar{Q}_1 + \phi_2^2 Q_2)X]^{-1}X'(\bar{Q}_1 + \phi_2^2 Q_2)y \tag{3.37}$$

Iterated GLS can be obtained through the successive application of (3.36) and (3.37). Baltagi and Li (1992a) show that the update of $\phi_2^2(i+1)$ in the $(i+1)$th iteration will be positive and finite even if the initial $\beta(i)$ value is $\widehat{\beta}_{BW}$ (from $\phi_2^2(i)=0$) or $\widehat{\beta}_B$ (from the limit as $\phi_2^2(i)\to\infty$). More importantly, Breusch's (1987) "remarkable property" extends to the two-way error component model in the sense that the ϕ_2^2 form a monotonic sequence. Therefore, if one starts with $\widehat{\beta}_{BW}$, which corresponds to $\phi_2^2=0$, the sequence of ϕ_2^2 is strictly increasing. On the other hand, starting with $\widehat{\beta}_B$, which corresponds to $\phi_2^2\to\infty$, the sequence of ϕ_2^2 is strictly decreasing. This remarkable property allows the applied researcher to check for the possibility of multiple local maxima. For a fixed ϕ_3^2, starting with *both* $\widehat{\beta}_{BW}$ and $\widehat{\beta}_B$ as initial values, there is a single maximum if and only if both sequences of iterations converge to the same ϕ_2^2 estimate.[4] Since this result holds for any arbitrary ϕ_3^2 between zero and one, a search over ϕ_3^2 in this range will guard against multiple local maxima. Of course, there are other computationally more efficient maximum likelihood algorithms. In fact, two-way MLE can be implemented using TSP. The iterative algorithm described here is of value for pedagogical reasons as well as for guarding against a local maximum.

3.5 PREDICTION

How does the best linear unbiased predictor look for the ith individual, S periods ahead for the two-way model? From (3.1), for period $T+S$

$$u_{i,T+S}=\mu_i+\lambda_{T+S}+\nu_{i,T+S} \tag{3.38}$$

and

$$\begin{aligned} E(u_{i,T+S}u_{jt}) &= \sigma_\mu^2 \qquad \text{for } i=j \\ &= 0 \qquad \text{for } i\neq j \end{aligned} \tag{3.39}$$

and $t=1,2,\ldots,T$. Hence, for the BLUP given in (2.37), $w=E(u_{i,T+S}u)=\sigma_\mu^2(l_i\otimes\iota_T)$ remains the same where l_i is the ith column of I_N. However, Ω^{-1} is given by (3.14), and

$$w'\Omega^{-1}=\sigma_\mu^2(l_i'\otimes\iota_T')\left[\sum_{i=1}^{4}\frac{1}{\lambda_i}Q_i\right] \tag{3.40}$$

Using the fact that

$$\begin{aligned} (l_i'\otimes\iota_T')Q_1&=0 \qquad (l_i'\otimes\iota_T')Q_2=(l_i'\otimes\iota_T')-\iota_{NT}'/N \\ (l_i'\otimes\iota_T')Q_3&=0 \qquad (l_i'\otimes\iota_T')Q_4=\iota_{NT}'/N \end{aligned} \tag{3.41}$$

one gets

$$w'\Omega^{-1}=\frac{\sigma_\mu^2}{\lambda_2}[(l_i'\otimes\iota_T')-\iota_{NT}'/N]+\frac{\sigma_\mu^2}{\lambda_4}(\iota_{NT}'/N) \tag{3.42}$$

Therefore, the typical element of $w'\Omega^{-1}\widehat{u}_{\text{GLS}}$ where $\widehat{u}_{\text{GLS}}=y-Z\widehat{\delta}_{\text{GLS}}$ is

$$\frac{T\sigma_\mu^2}{(T\sigma_\mu^2+\sigma_\nu^2)}(\bar{\widehat{u}}_{i.,\text{GLS}}-\bar{\bar{\widehat{u}}}_{..,\text{GLS}})+\frac{T\sigma_\mu^2}{(T\sigma_\mu^2+N\sigma_\lambda^2+\sigma_\nu^2)}\bar{\bar{\widehat{u}}}_{..,\text{GLS}} \tag{3.43}$$

or

$$\frac{T\sigma_\mu^2}{(T\sigma_\mu^2+\sigma_\nu^2)}\overline{\overline{\widehat{u}}}_{i.,\mathrm{GLS}} + T\sigma_\mu^2\left[\frac{1}{\lambda_4}-\frac{1}{\lambda_2}\right]\overline{\overline{\widehat{u}}}_{..,\mathrm{GLS}}$$

where $\overline{\overline{\widehat{u}}}_{i.,\mathrm{GLS}} = \sum_{t=1}^{T}\widehat{u}_{it,\mathrm{GLS}}/T$ and $\overline{\overline{\widehat{u}}}_{..,\mathrm{GLS}} = \sum_i\sum_t \widehat{u}_{it,\mathrm{GLS}}/NT$. See problem 88.1.1 in *Econometric Theory* by Baltagi (1988) and its solution 88.1.1 by Koning (1989). In general, $\overline{\overline{\widehat{u}}}_{..,\mathrm{GLS}}$ is not necessarily zero. The GLS normal equations are $Z'\Omega^{-1}\widehat{u}_{\mathrm{GLS}} = 0$. However, if Z contains a constant, then $\iota_{NT}'\Omega^{-1}\widehat{u}_{\mathrm{GLS}} = 0$, and using the fact that $\iota_{NT}'\Omega^{-1} = \iota_{NT}'/\lambda_4$ from (3.14), one gets $\overline{\overline{\widehat{u}}}_{..,\mathrm{GLS}} = 0$. Hence, for the two-way model, if there is a constant in the model, the BLUP for $y_{i,T+S}$ corrects the GLS prediction by a fraction of the mean of the GLS residuals corresponding to that ith individual

$$\widehat{y}_{i,T+S} = Z_{i,T+S}'\widehat{\delta}_{\mathrm{GLS}} + \left(\frac{T\sigma_\mu^2}{T\sigma_\mu^2+\sigma_\nu^2}\right)\overline{\overline{\widehat{u}}}_{i.,\mathrm{GLS}} \tag{3.44}$$

This looks exactly like the BLUP for the one-way model but with a different Ω. If there is no constant in the model, the last term in (3.44) should be replaced by (3.43).

3.6 EXAMPLES

3.6.1 Example 1: Grunfeld Investment Equation

For Grunfeld's (1958) example considered in Chapter 2, the investment equation is estimated using a two-way error component model. Table 3.1 gives OLS, Within, three feasible GLS estimates and the iterative MLE for the slope coefficients. The Within estimator yields a $\widetilde{\beta}_1$ estimate at 0.118 (0.014) and a $\widetilde{\beta}_2$ estimate at 0.358 (0.023). In fact, Table 3.2 gives the EViews output for the two-way fixed effects estimator. This is performed under the panel option with fixed individual and fixed time effects. For the random effects estimators, both the SWAR and WALHUS report negative estimates of σ_λ^2 and this is replaced by zero. Table 3.3 gives the EViews output for the random effects estimator of the two-way error component model for the Wallace and Hussain (1969) option. Table 3.4 gives the EViews output for the Amemiya

Table 3.1 Grunfeld's Data. Two-way Error Component Results

	β_1	β_2	σ_μ	σ_λ	σ_ν
OLS	0.116	0.231			
	(0.006)*	(0.025)*			
Within	0.118	0.358			
	(0.014)	(0.023)			
WALHUS	0.110	0.308	87.31	0	55.33
	(0.010)	(0.017)			
AMEMIYA	0.111	0.324	89.26	15.78	51.72
	(0.011)	(0.019)			
SWAR	0.110	0.308	84.23	0	51.72
	(0.011)	(0.017)			
IMLE	0.110	0.309	80.41	3.87	52.35
	(0.010)	(0.020)			

*These are biased standard errors when the true model has error component disturbances (see Moulton, 1986).

Table 3.2 Grunfeld's Data. Two-way Within Estimator

Dependent variable: I
Method: Panel least squares

Sample: 1935 1954
Cross-sections included: 10
Total panel (balanced) observations: 200

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−80.16380	14.84402	−5.400409	0.0000
F	0.117716	0.013751	8.560354	0.0000
K	0.357916	0.022719	15.75404	0.0000

Effects Specification

Cross-section fixed (dummy variables)
Period fixed (dummy variables)

R-squared	0.951693	Mean dependent variance	145.9582
Adjusted R-squared	0.943118	S.D. dependent variance	216.8753
S.E. of regression	51.72452	Akaike information criterion	10.87132
Sum squared residual	452147.1	Schwarz criterion	11.38256
Loglikelihood	−1056.132	F-statistic	110.9829
Durbin–Watson statistic	0.719087	Prob(F-statistic)	0.000000

Table 3.3 Grunfeld's Data. Two-way Wallace and Hussain Estimator

Dependent variable: I
Method: Panel EGLS (two-way random effects)

Sample: 1935 1954
Cross-sections included: 10
Total panel (balanced) observations: 200
Wallace and Hussain estimator of component variances

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−57.81705	28.63258	−2.019275	0.0448
F	0.109776	0.010473	10.48183	0.0000
K	0.308069	0.017186	17.92575	0.0000

Effects Specification

Cross-section random S.D./rho	87.31428	0.7135
Period random S.D./rho	0.000000	0.0000
Idiosyncratic random S.D./rho	55.33298	0.2865

Weighted Statistics

R-squared	0.769560	Mean dependent variance	20.47837
Adjusted R-squared	0.767221	S.D. dependent variance	109.4624
S.E. of regression	52.81254	Sum squared residual	549465.3
F-statistic	328.9438	Durbin–Watson statistic	0.681973
Prob(F-statistic)	0.000000		

Unweighted Statistics

R-squared	0.803316	Mean dependent variance	145.9582
Sum squared residual	1840949	Durbin–Watson statistic	0.203548

Table 3.4 Grunfeld's Data. Two-way Amemiya/Wansbeek and Kapteyn Estimator

Dependent variable: I
Method: Panel EGLS (two-way random effects)

Sample: 1935 1954
Cross-sections included: 10
Total panel (balanced) observations: 200
Wansbeek and Kapteyn estimator of component variances

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−63.89217	30.53284	−2.092573	0.0377
F	0.111447	0.010963	10.16577	0.0000
K	0.323533	0.018767	17.23947	0.0000
		Effects Specification		
Cross-section random S.D./rho			89.26257	0.7315
Period random S.D./rho			15.77783	0.0229
Idiosyncratic random S.D./rho			51.72452	0.2456
		Weighted Statistics		
R-squared	0.748982	Mean dependent variance		18.61292
Adjusted R-squared	0.746433	S.D. dependent variance		101.7143
S.E. of regression	51.21864	Sum squared residual		516799.9
F-statistic	293.9017	Durbin–Watson statistic		0.675336
Prob(F-statistic)	0.000000			
		Unweighted Statistics		
R-squared	0.798309	Mean dependent variance		145.9582
Sum squared residual	1887813	Durbin–Watson statistic		0.199923

(1971) estimator. In this case the estimate of σ_λ is 15.8, the estimate of σ_μ is 89.3 and the estimate of σ_ν is 51.7. This means that the variance of the time effects is only 2.3% of the total variance, while the variance of the firm effects is 73.1% of the total variance, and the variance of the remainder effects is 24.6% of the total variance. Table 3.5 gives the EViews output for the Swamy and Arora (1972) estimator. The iterative maximum likelihood method yields $\widehat{\beta}_1$ at 0.110 (0.010) and $\widehat{\beta}_2$ at 0.309 (0.020). This was performed using TSP.

3.6.2 Example 2: Gasoline Demand

For the motor gasoline data in Baltagi and Griffin (1983) considered in Chapter 2, the gasoline demand equation is estimated using a two-way error component model. Table 3.6 gives OLS, Within, three feasible GLS estimates and iterative MLE for the slope coefficients. The Within estimator is drastically different from OLS. The WALHUS and SWAR methods yield negative estimates of σ_λ^2 and this is replaced by zero. IMLE is obtained using TSP.

3.6.3 Example 3: Public Capital Productivity

For the Munnell (1990) public capital data considered by Baltagi and Pinnoi (1995) in Chapter 2, the Cobb–Douglas production function is estimated using a two-way error component

Table 3.5 Grunfeld's Data. Two-way Swamy and Arora Estimator

Dependent variable: I
Method: Panel EGLS (two-way random effects)

Sample: 1935 1954
Cross-sections included: 10
Total panel (balanced) observations: 200
Swamy and Arora estimator of component variances

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−57.86538	29.39336	−1.968655	0.0504
F	0.109790	0.010528	10.42853	0.0000
K	0.308190	0.017171	17.94833	0.0000
		Effects Specification		
Cross-section random S.D./rho			84.23332	0.7262
Period random S.D./rho			0.000000	0.0000
Idiosyncratic random S.D./rho			51.72452	0.2738
		Weighted Statistics		
R-squared	0.769400	Mean dependent variance		19.85502
Adjusted R-squared	0.767059	S.D. dependent variance		109.2695
S.E. of regression	52.73776	Sum squared residual		547910.4
F-statistic	328.6473	Durbin–Watson statistic		0.683945
Prob(F-statistic)	0.000000			
		Unweighted Statistics		
R-squared	0.803283	Mean dependent variance		145.9582
Sum squared residual	1841262	Durbin–Watson statistic		0.203524

Table 3.6 Gasoline Demand Data. Two-way Error Component Results

	β_1	β_2	β_3	σ_μ	σ_λ	σ_ν
OLS	0.889	−0.892	−0.763			
	(0.036)*	(0.030)*	(0.019)			
Within	0.051	−0.193	−0.593			
	(0.091)	(0.043)	(0.028)			
WALHUS	0.545	−0.450	−0.605	0.197	0	0.115
	(0.056)	(0.039)	(0.025)			
AMEMIYA	0.170	−0.233	−0.602	0.423	0.131	0.081
	(0.080)	(0.041)	(0.026)			
SWAR	0.565	−0.405	−0.609	0.196	0	0.081
	(0.061)	(0.040)	(0.026)			
IMLE	0.231	−0.254	−0.606	0.361	0.095	0.082
	(0.091)	(0.045)	(0.026)			

*These are biased standard errors when the true model has error component disturbances (see Moulton, 1986).

Table 3.7 Public Capital Data. Two-way Error Component Results

	β_1	β_2	β_3	β_4	σ_μ	σ_λ	σ_ν
OLS	0.155	0.309	0.594	−0.007			
	(0.017)*	(0.010)*	(0.014)*	(0.001)*			
Within	−0.030	0.169	0.769	−0.004			
	(0.027)	(0.028)	(0.028)	(0.001)			
WALHUS	0.026	0.258	0.742	−0.005	0.082	0.016	0.036
	(0.023)	(0.021)	(0.024)	(0.001)			
AMEMIYA	0.002	0.217	0.770	−0.004	0.154	0.026	0.034
	(0.025)	(0.024)	(0.026)	(0.001)			
SWAR	0.018	0.266	0.745	−0.005	0.083	0.010	0.034
	(0.023)	(0.021)	(0.024)	(0.001)			
IMLE	0.020	0.250	0.750	−0.004	0.091	0.017	0.035
	(0.024)	(0.023)	(0.025)	(0.001)			

*These are biased standard errors when the true model has error component disturbances (see Moulton, 1986).

model. Table 3.7 gives OLS, Within, three feasible GLS estimates and iterative MLE for the slope coefficients. With the exception of OLS, estimates of the public capital coefficient are insignificant in this production function. Also, none of the feasible GLS estimators yield negative estimates of the variance components.

3.7 SELECTED APPLICATIONS

(1) For an application of the two-way fixed effects model to a study of the effects of foreign aid on public sector budgets of 46 developing countries observed over the period 1975–80, see Cashel-Cordo and Craig (1990).

(2) For an application of the two-way random effects model to study the determinants of secondary market prices for developing country syndicated loans, see Boehmer and Megginson (1990). Their panel consisted of 10 countries observed over 32 months beginning in July 1985 and ending in July 1988.

(3) Carpenter, Fazzari and Petersen (1998) estimate a two-way fixed effects model to provide evidence of the importance of the firm's financing constraints in explaining the dramatic cycles in inventory investments. Using quarterly firm panel data obtained from the Compustat tapes, they conclude that cash flow is much more successful than cash stocks or coverage in explaining inventory investment across firm size, different inventory cycles and different manufacturing sectors.

(4) Baltagi, Egger and Pfaffermayr (2003) consider an unbalanced panel of bilateral export flows from the EU15 countries, the USA and Japan to their 57 most important trading partners over the period 1986–98. They estimate a three-way gravity equation with importer, exporter and time fixed effects as well as pairwise interaction effects. These effects include time-invariant factors like distance, common borders, island nation, land-locked, common language, colonies, etc. These fixed effects as well as the interaction terms are found to be statistically significant. Omission of these effects can result in biased and misleading inference.

NOTES

1. These characteristic roots and eigenprojectors were first derived by Nerlove (1971b) for the two-way error component model. More details are given in appendix 1 of Baltagi (1995).

2. Alternatively, one can maximize L_C over ϕ_3^2, given β and ϕ_2^2. The results are symmetric and are left as an exercise. In fact, one can show (see problem 3.6) that ϕ_3^2 will satisfy a quadratic equation like (3.35) with N exchanging places with T, ϕ_2^2 replacing ϕ_3^2 and Q_2 exchanging places with Q_3 in a, b and c, respectively.
3. The case where $\phi_3^2 = 1$ corresponds to $\sigma_\lambda^2 = 0$, i.e. the one-way error component model where Breusch's (1987) results apply.
4. There will be no local maximum interior to $0 < \phi_2^2 \leq 1$ if starting from $\widehat{\beta}_{BW}$ we violate the nonnegative variance component requirement, $\phi_2^2 \leq 1$. In this case, one should set $\phi_2^2 = 1$.

PROBLEMS

3.1 (a) Prove that the Within estimator $\widetilde{\beta} = (X'QX)^{-1}X'Qy$ with Q defined in (3.3) can be obtained from OLS on the panel regression model (2.3) with disturbances defined in (3.2). *Hint*: Use the Frisch–Waugh–Lovell theorem of Davidson and MacKinnon (1993, p. 19). Also, the generalized inverse matrix result given in problem 9.6.

(b) *Within two-way is equivalent to two Withins one-way*. This is based on problem 98.5.2 in *Econometric Theory* by Baltagi (1998). Show that the Within two-way estimator of β can be obtained by applying two Within (one-way) transformations. The first is the Within transformation ignoring the time effects followed by the Within transformation ignoring the individual effects. Show that the order of these two Within (one-way) transformations is unimportant. Give an intuitive explanation for this result. See solution 98.5.2 in *Econometric Theory* by Li (1999).

3.2 (a) Using generalized inverse, show that OLS or GLS on (2.6) with Q defined in (3.3) yields $\widetilde{\beta}$, the Within estimator.

(b) Show that (2.6) with Q defined in (3.3) satisfies the necessary and sufficient condition for OLS to be equivalent to GLS (see Baltagi, 1989).

3.3 (a) Verify (3.10) and (3.13) and check that $\Omega^{-1}\Omega = I$ using (3.14).

(b) Verify that $\Omega^{-1/2}\Omega^{-1/2} = \Omega^{-1}$ using (3.14).

(c) Premultiply y by $\sigma_\nu \Omega^{-1/2}$ from (3.15) and show that the typical element is given by (3.16).

3.4 (a) Perform OLS on the system of equations given in (3.22) and show that the resulting estimate is $\widehat{\beta}_{\text{OLS}} = (X(I_{NT} - \bar{J}_{NT})X)^{-1}X'(I_{NT} - \bar{J}_{NT})y$.

(b) Perform GLS on this system of equations and show that $\widehat{\beta}_{\text{GLS}}$ reduces to the expression given by (3.23).

3.5 Show that the Swamy and Arora (1972) estimators of λ_1, λ_2 and λ_3 given by (3.19), (3.20) and (3.21) are unbiased for σ_ν^2, λ_2 and λ_3, respectively.

3.6 (a) Using the concentrated likelihood function in (3.32), solve $\partial L_C/\partial\beta = 0$, given ϕ_2^2 and ϕ_3^2, and verify (3.33).

(b) Solve $\partial L_C/\partial\phi_2^2 = 0$, given ϕ_3^2 and β, and verify (3.34).

(c) Solve $\partial L_C/\partial\phi_3^2 = 0$, given ϕ_2^2 and β, and show that the solution ϕ_3^2 satisfies

$$\overline{a}\phi_3^4 + \overline{b}\phi_3^2 + \overline{c} = 0$$

where

$$\overline{a} = -[T(N-1)+1](1-\phi_2^2)(d'Q_3d)$$
$$\overline{b} = (1-\phi_2^2)(T-1)d'[Q_1 + \phi_2^2 Q_2]d - \phi_2^2 T(N-1)d'Q_3d$$

and

$$\bar{c} = T\phi_2^2 d'(Q_1 + \phi_2^2 Q_2)d$$

Note that this is analogous to (3.35), with ϕ_2^2 replacing ϕ_3^2, N replacing T, and Q_2 replacing Q_3 and vice versa, wherever they occur.

3.7 *Predicting* $y_{i,T+S}$.
 (a) For the two-way error component model in (3.1), verify (3.39) and (3.42).
 (b) Also, show that if there is a constant in the regression, $\iota_{NT}'\Omega^{-1}\widehat{u}_{\text{GLS}} = 0$ and $\overline{\overline{\widehat{u}}}_{..,\text{GLS}} = 0$.

3.8 Using Grunfeld's data given on the Wiley web site as Grunfeld.fil, reproduce Table 3.1.

3.9 Using the gasoline demand data of Baltagi and Griffin (1983), given as Gasoline.dat on the Wiley web site, reproduce Table 3.6.

3.10 Using the public capital data of Munnell (1990), given as Produc.prn on the Wiley web site, reproduce Table 3.7.

3.11 Using the Monte Carlo set-up for the two-way error component model given in (3.27) and (3.28) (see Baltagi, 1981a), compare the various estimators of the variance components and regression coefficients studied in this chapter.

3.12 *Variance component estimation under misspecification*. This is based on problem 91.3.3 in *Econometric Theory* by Baltagi and Li (1991). This problem investigates the consequences of under- or overspecifying the error component model on the variance components estimates. Since the one-way and two-way error component models are popular in economics, we focus on the following two cases.
 (1) *Underspecification*. In this case, the true model is two-way, see (3.1):

$$u_{it} = \mu_i + \lambda_t + \nu_{it} \quad i = 1, \ldots, N;\ t = 1, \ldots, T$$

 while the estimated model is one-way, see (2.2):

$$u_{it} = \mu_i + \nu_{it}$$

 $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$, $\lambda_t \sim \text{IID}(0, \sigma_\lambda^2)$, $\nu_{it} \sim \text{IID}(0, \sigma_\nu^2)$ independent of each other and among themselves.
 (a) Knowing the true disturbances (u_{it}), show that the BQUE of σ_ν^2 for the misspecified one-way model is biased upwards, while the BQUE of σ_μ^2 remains unbiased.
 (b) Show that if the u_{it} are replaced by the one-way least squares dummy variables (LSDV) residuals, the variance component estimate of σ_ν^2 given in part (a) is inconsistent, while that of σ_μ^2 is consistent.
 (2) *Overspecification*. In this case, the true model is one-way, given by (2.2), while the estimated model is two-way, given by (3.1).
 (c) Knowing the true disturbances (u_{it}), show that the BQUE of σ_μ^2, σ_λ^2 and σ_ν^2 for the misspecified two-way model remain unbiased.
 (d) Show that if the u_{it} are replaced by the two-way (LSDV) residuals, the variance components estimates given in part (c) remain consistent. (*Hint*: See solution 91.3.3 in *Econometric Theory* by Baltagi and Li (1992). See also Deschamps (1991) who shows that an underspecified error component model yields inconsistent estimates of the coefficient variances.)

3.13 For the random two-way error component model described by (2.1) and (3.1), consider the OLS estimator of $\text{var}(u_{it}) = \sigma^2$, which is given by $s^2 = \widehat{u}_{\text{OLS}}'\widehat{u}_{\text{OLS}}/(n - K')$ where

$n = NT$ and $K' = K + 1$.

(a) Show that

$$E(s^2) = \sigma^2 - \sigma_\mu^2[\mathrm{tr}(I_N \otimes J_T)P_x - K']/(n - K')$$
$$-\sigma_\lambda^2[\mathrm{tr}(J_N \otimes I_T)P_x - K']/(n - K')$$

(b) Consider the inequalities given by Kiviet and Krämer (1992) which are reproduced in problem 2.14, part (b). Show that for the two-way error component model, these bounds are given by the following two cases.

(1) For $T\sigma_\mu^2 < N\sigma_\lambda^2$:

$$0 \leq \sigma_\nu^2 + \sigma_\mu^2(n - T)/(n - K') + \sigma_\lambda^2(n - NK')/(n - K') \leq E(s^2)$$
$$\leq \sigma_\nu^2 + \sigma_\mu^2[n/(n - K')] + \sigma_\lambda^2[n/(n - K')] \leq \sigma^2(n/n - K')$$

(2) For $T\sigma_\mu^2 > N\sigma_\lambda^2$:

$$0 \leq \sigma_\nu^2 + \sigma_\mu^2(n - TK')/(n - K') + \sigma_\lambda^2(n - N)/(n - K') \leq E(s^2)$$
$$\leq \sigma_\nu^2 + \sigma_\mu^2[n/(n - K')] + \sigma_\lambda^2[n/(n - K')] \leq \sigma^2(n/n - K')$$

In either case, as $n \to \infty$, both bounds tend to σ^2 and s^2 is asymptotically unbiased, irrespective of the particular evolution of X. See Baltagi and Krämer (1994) for a proof of this result.

3.14 *Nested effects*. This is based on problem 93.4.2 in *Econometric Theory* by Baltagi (1993). In many economic applications, the data may contain nested groupings. For example, data on firms may be grouped by industry, data on states by region and data on individuals by profession. In this case, one can control for unobserved industry and firm effects using a nested error component model. Consider the regression equation

$$y_{ijt} = x'_{ijt}\beta + u_{ijt} \quad \text{for} \quad i = 1, \ldots, M;\ j = 1, \ldots, N;\ t = 1, 2, \ldots, T$$

where y_{ijt} could denote the output of the jth firm in the ith industry for the tth time period. x_{ijt} denotes a vector of k inputs, and the disturbance is given by

$$u_{ijt} = \mu_i + \nu_{ij} + \epsilon_{ijt}$$

where $\mu_i \sim \mathrm{IID}(0, \sigma_\mu^2)$, $\nu_{ij} \sim \mathrm{IID}(0, \sigma_\nu^2)$ and $\epsilon_{ijt} \sim \mathrm{IID}(0, \sigma_\epsilon^2)$, independent of each other and among themselves. This assumes that there are M industries with N firms in each industry observed over T periods.

(1) Derive $\Omega = E(uu')$ and obtain $\Omega^{-1}$and $\Omega^{-1/2}$.

(2) Show that $y^* = \sigma_\epsilon\Omega^{-1/2}y$ has a typical element

$$y^*_{ijt} = (y_{ijt} - \theta_1\bar{y}_{ij.} + \theta_2\bar{y}_{i..})$$

where $\theta_1 = 1 - (\sigma_\epsilon/\sigma_1)$ with $\sigma_1^2 = (T\sigma_\nu^2 + \sigma_\epsilon^2)$; $\theta_2 = -(\sigma_\epsilon/\sigma_1) + (\sigma_\epsilon/\sigma_2)$ with $\sigma_2^2 = (NT\sigma_\mu^2 + T\sigma_\nu^2 + \sigma_\epsilon^2)$; $\bar{y}_{ij.} = \sum_{t=1}^T y_{ijt}/T$ and $\bar{y}_{i..} = \sum_{j=1}^N \sum_{t=1}^T y_{ijt}/NT$. See solution 93.4.2 in *Econometric Theory* by Xiong (1995).

3.15 Ghosh (1976) considered the following error component model:

$$u_{itq} = \mu_i + \lambda_t + \eta_q + \nu_{itq}$$

where $i = 1, \ldots, N$; $T = 1, \ldots, T$ and $q = 1, \ldots, M$. Ghosh (1976) argued that in international or interregional studies, there might be two rather than one cross-sectional components; for example, i might denote countries and q might be regions within that country. These four *independent* components are assumed to be random with $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$, $\lambda_t \sim \text{IID}(0, \sigma_\lambda^2)$, $\eta_q \sim \text{IID}(0, \sigma_\eta^2)$ and $\nu_{itq} \sim \text{IID}(0, \sigma_\nu^2)$. Order the observations such that the faster index is q, while the slower index is t, so that

$$u' = (u_{111}, \ldots, u_{11M}, u_{121}, \ldots, u_{12M}, \ldots, u_{1N1}, \ldots,$$
$$u_{1NM}, \ldots, u_{T11}, \ldots, u_{T1M}, \ldots, u_{TN1}, \ldots, u_{TNM})$$

(a) Show that the error has mean zero and variance–covariance matrix

$$\Omega = E(uu') = \sigma_\nu^2(I_T \otimes I_N \otimes I_M) + \sigma_\lambda^2(I_T \otimes J_N \otimes J_M)$$
$$+\sigma_\mu^2(J_T \otimes I_N \otimes J_M) + \sigma_\eta^2(J_T \otimes J_N \otimes I_M)$$

(b) Using the Wansbeek and Kapteyn (1982b) trick, show that $\Omega = \sum_{j=1}^5 \xi_j V_j$ where $\xi_1 = \sigma_\nu^2$, $\xi_2 = NM\sigma_\lambda^2 + \sigma_\nu^2$, $\xi_3 = TM\sigma_\mu^2 + \sigma_\nu^2$, $\xi_4 = NT\sigma_\eta^2 + \sigma_\nu^2$ and $\xi_5 = NM\sigma_\lambda^2 + TM\sigma_\mu^2 + NT\sigma_\eta^2 + \sigma_\nu^2$. Also

$$\begin{aligned}
V_1 &= I_T \otimes I_N \otimes I_M - I_T \otimes \bar{J}_N \otimes \bar{J}_M - \bar{J}_T \otimes I_N \otimes \bar{J}_M \\
&\quad -\bar{J}_T \otimes \bar{J}_N \otimes I_M + 2\bar{J}_T \otimes \bar{J}_N \otimes \bar{J}_M \\
V_2 &= E_T \otimes \bar{J}_N \otimes \bar{J}_M \quad \text{where} \quad E_T = I_T - \bar{J}_T \\
V_3 &= \bar{J}_T \otimes E_N \otimes \bar{J}_M \\
V_4 &= \bar{J}_T \otimes \bar{J}_N \otimes E_M \quad \text{and} \quad V_5 = \bar{J}_T \otimes \bar{J}_N \otimes \bar{J}_M
\end{aligned}$$

all symmetric and idempotent and sum to the identity matrix.

(c) Conclude that $\Omega^{-1} = \sum_{j=1}^5 (1/\xi_j)V_j$ and $\sigma_\nu\Omega^{-1/2} = \sum_{j=1}^5 (\sigma_\nu/\sqrt{\xi_j})V_j$ with the typical element of $\sigma_\nu\Omega^{-1/2}y$ being

$$y_{tiq} - \theta_1\bar{y}_{t..} - \theta_2\bar{y}_{.i.} - \theta_3\bar{y}_{..q} - \theta_4\bar{y}_{...}$$

where the dot indicates a sum over that index and a bar means an average. Here, $\theta_j = 1 - \sigma_\nu/\sqrt{\xi_{j+1}}$ for $j = 1, 2, 3$ while $\theta_4 = \theta_1 + \theta_2 + \theta_3 - 1 + (\sigma_\nu/\sqrt{\xi_5})$.

(d) Show that the BQU estimator of ξ_j is given by $u'V_ju/\text{tr}(V_j)$ for $j = 1, 2, 3, 4$. Show that BQU estimators of $\sigma_\nu^2, \sigma_\mu^2, \sigma_\eta^2$ and σ_λ^2 can be obtained using the one-to-one correspondence between the ξ_j and σ^2.

This problem is based on Baltagi (1987). For a generalization of this four-component model as well as an alternative class of decompositions of the variance–covariance matrix, see Wansbeek and Kapteyn (1982a). More recently, Davis (2001) gives an elegant generalization to the multi-way unbalanced error component model, see Chapter 9.

3.16 *A mixed-error component model.* This is based on problem 95.1.4 in *Econometric Theory* by Baltagi and Krämer (1995). Consider the panel data regression equation with a two-way mixed error component model described by (3.1) where the individual-specific effects are assumed to be random, with $\mu_i \sim (0, \sigma_\mu^2)$ and $\nu_{it} \sim (0, \sigma_\nu^2)$ independent of each other and among themselves. The time-specific effects, i.e. the λ_t's, are assumed to be fixed parameters to be estimated. In vector form, this can be written as

$$y = X\beta + Z_\lambda\lambda + w \tag{1}$$

where $Z_\lambda = \iota_N \otimes I_T$, and

$$w = Z_\mu \mu + \nu \tag{2}$$

with $Z_\mu = I_N \otimes \iota_T$. By applying the Frisch–Waugh–Lovell (FWL) theorem, one gets

$$Q_\lambda y = Q_\lambda X\beta + Q_\lambda w \tag{3}$$

where $Q_\lambda = E_N \otimes I_T$ with $E_N = I_N - \bar{J}_N$ and $\bar{J}_N = \iota_N \iota_N'/N$. This is the familiar Within time-effects transformation, with the typical element of $Q_\lambda y$ being $y_{it} - \bar{y}_{.t}$ and $\bar{y}_{.t} = \sum_{i=1}^{N} y_{it}/N$. Let $\Omega = E(ww')$, this is the familiar one-way error component variance–covariance matrix given in (2.17).

(a) Show that the GLS estimator of β obtained from (1) by premultiplying by $\Omega^{-1/2}$ first and then applying the FWL theorem yields the same estimator as GLS on (3) using the generalized inverse of $Q_\lambda \Omega Q_\lambda$. This is a special case of a more general result proved by Fiebig, Bartels and Krämer (1996).

(b) Show that pseudo-GLS on (3) using Ω rather than $Q_\lambda \Omega Q_\lambda$ for the variance of the disturbances yields the same estimator of β as found in part (a). In general, pseudo-GLS may not be the same as GLS, but Fiebig et al. (1996) provided a necessary and sufficient condition for this equivalence that is easy to check in this case. In fact, this amounts to checking whether $X' Q_\lambda \Omega^{-1} Z_\lambda = 0$. See solution 95.1.4 in *Econometric Theory* by Xiong (1996a).

For computational purposes, these results imply that one can perform the Within time-effects transformation to wipe out the matrix of time dummies and then do the usual Fuller–Battese (1974) transformation without worrying about the loss in efficiency of not using the proper variance–covariance matrix of the transformed disturbances.

4

Test of Hypotheses with Panel Data

4.1 TESTS FOR POOLABILITY OF THE DATA

The question of whether to pool the data or not naturally arises with panel data. The restricted model is the pooled model given by (2.3) representing a behavioral equation with the same parameters over time and across regions. The unrestricted model, however, is the same behavioral equation but with different parameters across time or across regions. For example, Balestra and Nerlove (1966) considered a dynamic demand equation for natural gas across 36 states over six years. In this case, the question of whether to pool or not to pool boils down to the question of whether the parameters of this demand equation vary from one year to the other over the six years of available data. One can have a behavioral equation whose parameters may vary across regions. For example, Baltagi and Griffin (1983) considered panel data on motor gasoline demand for 18 OECD countries. In this case, one is interested in testing whether the behavioral relationship predicting demand is the same across the 18 OECD countries, i.e. the parameters of the prediction equation do not vary from one country to the other.

These are but two examples of many economic applications where time-series and cross-section data may be pooled. Generally, most economic applications tend to be of the first type, i.e. with a large number of observations on individuals, firms, economic sectors, regions, industries and countries but only over a few time periods. In what follows, we study the tests for the poolability of the data for the case of pooling across regions keeping in mind that the other case of pooling over time can be obtained in a similar fashion.

For the unrestricted model, we have a regression equation for each region given by

$$y_i = Z_i\delta_i + u_i \qquad i = 1, 2, \dots, N \tag{4.1}$$

where $y_i' = (y_{i1}, \dots, y_{iT})$, $Z_i = [\iota_T, X_i]$ and X_i is $T \times K$. δ_i' is $1 \times (K+1)$ and u_i is $T \times 1$. The important thing to notice is that δ_i is different for every regional equation. We want to test the hypothesis $H_0 : \delta_i = \delta$ for all i, so that under H_0 we can write the restricted model given in (4.1) as

$$y = Z\delta + u \tag{4.2}$$

where $Z' = (Z_1', Z_2', \dots, Z_N')$ and $u' = (u_1', u_2', \dots, u_N')$. The unrestricted model can also be written as

$$y = \begin{pmatrix} Z_1 & 0 & \dots & 0 \\ 0 & Z_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & Z_N \end{pmatrix} \begin{pmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_N \end{pmatrix} + u = Z^*\delta^* + u \tag{4.3}$$

where $\delta^{*\prime} = (\delta_1', \delta_2', \dots, \delta_N')$ and $Z = Z^* I^*$ with $I^* = (\iota_N \otimes I_{K'})$, an $NK' \times K'$ matrix, with $K' = K + 1$. Hence the variables in Z are all linear combinations of the variables in Z^*.

4.1.1 Test for Poolability under $u \sim N(0, \sigma^2 I_{NT})$

Assumption 4.1 $u \sim N(0, \sigma^2 I_{NT})$

Under assumption 4.1, the minimum variance unbiased estimator for δ in equation (4.2) is

$$\widehat{\delta}_{\text{OLS}} = \widehat{\delta}_{mle} = (Z'Z)^{-1}Z'y \tag{4.4}$$

and therefore

$$y = Z\widehat{\delta}_{\text{OLS}} + e \tag{4.5}$$

implying that $e = (I_{NT} - Z(Z'Z)^{-1}Z')y = My = M(Z\delta + u) = Mu$ since $MZ = 0$. Similarly, under assumption 4.1, the MVU for δ_i is given by

$$\widehat{\delta}_{i,\text{OLS}} = \widehat{\delta}_{i,mle} = (Z_i'Z_i)^{-1}Z_i'y_i \tag{4.6}$$

and therefore

$$y_i = Z_i\widehat{\delta}_{i,\text{OLS}} + e_i \tag{4.7}$$

implying that $e_i = (I_T - Z_i(Z_i'Z_i)^{-1}Z_i')y_i = M_iy_i = M_i(Z_i\delta_i + u_i) = M_iu_i$ since $M_iZ_i = 0$, and this is true for $i = 1, 2, \ldots, N$. Also, let

$$M^* = I_{NT} - Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime} = \begin{pmatrix} M_1 & 0 & \ldots & 0 \\ 0 & M_2 & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & M_N \end{pmatrix}$$

One can easily deduce that $y = Z^*\widehat{\delta}^* + e^*$ with $e^* = M^*y = M^*u$ and $\widehat{\delta}^* = (Z^{*\prime}Z^*)^{-1}Z^{*\prime}y$. Note that both M and M^* are symmetric and idempotent with $MM^* = M^*$. This easily follows since

$$\begin{aligned} Z(Z'Z)^{-1}Z'Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime} &= Z(Z'Z)^{-1}I^{*\prime}Z^{*\prime}Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime} \\ &= Z(Z'Z)^{-1}Z' \end{aligned}$$

This uses the fact that $Z = Z^*I^*$. Under assumption 4.1, $e'e - e^{*\prime}e^* = u'(M - M^*)u$ and $e^{*\prime}e^* = u'M^*u$ are independent since $(M - M^*)M^* = 0$. Also, both quadratic forms when divided by σ^2 are distributed as χ^2 since $(M - M^*)$ and M^* are idempotent. Dividing these quadratic forms by their respective degrees of freedom and taking their ratio leads to the following test statistic:[1]

$$\begin{aligned} F_{obs} &= \frac{(e'e - e^{*\prime}e^*)/(\text{tr}(M) - \text{tr}(M^*))}{e^{*\prime}e^*/\text{tr}(M^*)} \\ F_{obs} &= \frac{(e'e - e_1'e_1 - e_2'e_2 - \ldots - e_N'e_N)/(N-1)K'}{(e_1'e_1 + e_2'e_2 + \ldots + e_N'e_N)/N(T - K')} \end{aligned} \tag{4.8}$$

Under H_0, F_{obs} is distributed as an $F((N-1)K', N(T-K'))$. Hence the critical region for this test is defined as

$$\{F_{obs} > F((N-1)K', NT - NK'; \alpha_0)\}$$

where α_0 denotes the level of significance of the test. This is exactly the Chow test presented by Chow (1960) extended to the case of N linear regressions. Therefore if an economist has reason to believe that assumption 4.1 is true, and wants to pool his data across regions, then it is recommended that he or she test for the poolability of the data using the Chow test given in (4.8). However, for the variance component model, $u \sim (0, \Omega)$ and not $(0, \sigma^2 I_{NT})$. Therefore, even if we assume normality on the disturbances two questions remain: (1) is the Chow test still the right test to perform when $u \sim N(0, \Omega)$? and (2) does the Chow statistic still have an F-distribution when $u \sim N(0, \Omega)$? The answer to the first question is no, the Chow test given in (4.8) is not the right test to perform. However, as will be shown later, a generalized Chow test will be the right test to perform. As for the second question, it is still relevant to ask because it highlights the problem of economists using the Chow test assuming erroneously that u is $N(0, \sigma^2 I_{NT})$ when in fact it is not. For example, Toyoda (1974), in treating the case where the u_i are heteroskedastic, found that the Chow statistic given by (4.8) has an approximate F-distribution where the degree of freedom of the denominator depends upon the true variances. Hence for specific values of these variances, Toyoda demonstrates how wrong it is to apply the Chow test in case of heteroskedastic variances.

Having posed the two questions above, we can proceed along two lines: the first is to find the approximate distribution of the Chow statistic (4.8) in case $u \sim N(0, \Omega)$ and therefore show how erroneous it is to use the Chow test in this case (this is not pursued in this book). The second route, and the more fruitful, is to derive the right test to perform for pooling the data in case $u \sim N(0, \Omega)$. This is done in the next subsection.

4.1.2 Test for Poolability under the General Assumption $u \sim N(0, \Omega)$

Assumption 4.2 $u \sim N(0, \Omega)$

In case Ω is known up to a scalar factor, the test statistic employed for the poolability of the data would be simple to derive. All we need to do is transform our model (under both the null and alternative hypotheses) such that the transformed disturbances have a variance of $\sigma^2 I_{NT}$, then apply the Chow test on the transformed model. The later step is legitimate because the transformed disturbances have homoskedastic variances and the analysis of the previous subsection applies in full. Given $\Omega = \sigma^2 \Sigma$, we premultiply the restricted model given in (4.2) by $\Sigma^{-1/2}$ and we call $\Sigma^{-1/2} y = \dot{y}$, $\Sigma^{-1/2} Z = \dot{Z}$ and $\Sigma^{-1/2} u = \dot{u}$. Hence

$$\dot{y} = \dot{Z}\delta + \dot{u} \tag{4.9}$$

with $E(\dot{u}\dot{u}') = \Sigma^{-1/2} E(uu') \Sigma^{-1/2'} = \sigma^2 I_{NT}$. Similarly, we premultiply the unrestricted model given in (4.3) by $\Sigma^{-1/2}$ and we call $\Sigma^{-1/2} Z^* = \dot{Z}^*$. Therefore

$$\dot{y} = \dot{Z}^* \delta^* + \dot{u} \tag{4.10}$$

with $E(\dot{u}\dot{u}') = \sigma^2 I_{NT}$.

At this stage, we can test $H_0 : \delta_i = \delta$ for every $i = 1, 2, \ldots, N$, simply by using the Chow statistic, only now on the transformed models (4.9) and (4.10) since they satisfy assumption 4.1 of homoskedasticity of the normal disturbances. Note that $\dot{Z} = \dot{Z}^* I^*$, which is simply obtained from $Z = Z^* I^*$ by premultiplying by $\Sigma^{-1/2}$. Defining $\dot{M} = I_{NT} - \dot{Z}(\dot{Z}'\dot{Z})^{-1}\dot{Z}'$ and $\dot{M}^* = I_{NT} - \dot{Z}^*(\dot{Z}^{*\prime}\dot{Z}^*)^{-1}\dot{Z}^{*\prime}$, it is easy to show that $\dot{M}$ and $\dot{M}^*$ are both symmetric, idempotent and such that $\dot{M}\dot{M}^* = \dot{M}^*$. Once again the conditions for lemma 2.2 of Fisher (1970) are satisfied, and the test statistic

$$\dot{F}_{obs} = \frac{(\dot{e}'\dot{e} - \dot{e}^{*\prime}\dot{e}^*)/(\text{tr}(\dot{M}) - \text{tr}(\dot{M}^*))}{\dot{e}^{*\prime}\dot{e}^*/\text{tr}(\dot{M}^*)} \sim F((N-1)K', N(T-K')) \tag{4.11}$$

where $\dot{e} = \dot{y} - \dot{Z}\widehat{\dot{\delta}}_{\text{OLS}}$ and $\widehat{\dot{\delta}}_{\text{OLS}} = (\dot{Z}'\dot{Z})^{-1}\dot{Z}'\dot{y}$ implying that $\dot{e} = \dot{M}\dot{y} = \dot{M}\dot{u}$. Similarly, $\dot{e}^* = \dot{y} - \dot{Z}^*\widehat{\dot{\delta}}^*_{\text{OLS}}$ and $\widehat{\dot{\delta}}^*_{\text{OLS}} = (\dot{Z}^{*\prime}\dot{Z}^*)^{-1}\dot{Z}^{*\prime}\dot{y}$ implying that $\dot{e}^* = \dot{M}^*\dot{y} = \dot{M}^*\dot{u}$. Using the fact that $\dot{M}$ and $\dot{M}^*$ are symmetric and idempotent, we can rewrite (4.11) as

$$\begin{aligned} \dot{F}_{obs} &= \frac{(\dot{y}'\dot{M}\dot{y} - \dot{y}'\dot{M}^*\dot{y})/(N-1)K'}{\dot{y}'\dot{M}^*\dot{y}/N(T-K')} \\ &= \frac{(y'\Sigma^{-1/2}\dot{M}\Sigma^{-1/2}y - y'\Sigma^{-1/2}\dot{M}^*\Sigma^{-1/2}y)/(N-1)K'}{y'\Sigma^{-1/2}\dot{M}^*\Sigma^{-1/2}y/N(T-K')} \end{aligned} \tag{4.12}$$

But

$$\dot{M} = I_{NT} - \Sigma^{-1/2}Z(Z'\Sigma^{-1}Z)^{-1}Z'\Sigma^{-1/2'}$$

and

$$\dot{M}^* = I_{NT} - \Sigma^{-1/2}Z^*(Z^{*\prime}\Sigma^{-1}Z^*)^{-1}Z^{*\prime}\Sigma^{-1/2'}$$

so that

$$\Sigma^{-1/2}\dot{M}\Sigma^{-1/2} = \Sigma^{-1} - \Sigma^{-1}Z(Z'\Sigma^{-1}Z)^{-1}Z'\Sigma^{-1}$$

and

$$\Sigma^{-1/2}\dot{M}^*\Sigma^{-1/2} = \Sigma^{-1} - \Sigma^{-1}Z^*(Z^{*\prime}\Sigma^{-1}Z^*)^{-1}Z^{*\prime}\Sigma^{-1}$$

Hence we can write (4.12) in the form

$$\dot{F}_{obs} = \frac{y'[\Sigma^{-1}(Z^*(Z^{*\prime}\Sigma^{-1}Z^*)^{-1}Z^{*\prime} - Z(Z'\Sigma^{-1}Z)^{-1}Z')\Sigma^{-1}]y/(N-1)K'}{(y'\Sigma^{-1}y - y'\Sigma^{-1}Z^*(Z^{*\prime}\Sigma^{-1}Z^*)^{-1}Z^{*\prime}\Sigma^{-1}y)/N(T-K')} \tag{4.13}$$

and $\dot{F}_{obs}$, has an F-distribution with $((N-1)K', N(T-K'))$ degrees of freedom. It is important to emphasize that (4.13) is operational only when Σ is known. This test is a special application of a general test for linear restrictions described in Roy (1957) and used by Zellner (1962) to test for aggregation bias in a set of seemingly unrelated regressions. In case Σ is unknown, we replace Σ in (4.13) by a consistent estimator (say $\widehat{\Sigma}$) and call the resulting test statistics $\widehat{\dot{F}}_{obs}$.

One of the main motivations behind pooling a time series of cross-sections is to widen our database in order to get better and more reliable estimates of the parameters of our model. Using the Chow test, the question of whether "to pool or not to pool" reduced to a test of the validity of the null hypothesis $H_0 : \delta_i = \delta$ for all i. Imposing these restrictions (true or false) will reduce the variance of the pooled estimator, but may introduce bias if these restrictions are false. This motivated Toro-Vizcarrondo and Wallace (1968, p. 560) to write, "if one is willing to accept some bias in trade for a reduction in variance, then even if the restriction is not true one might still prefer the restricted estimator". Baltagi (1995, pp. 54–58) discusses three mean square error criteria used in the literature to test whether the pooled estimator restricted by H_0 is better than the unrestricted estimator of δ^*. It is important to emphasize that these MSE criteria do not test whether H_0 is true or false, but help us to choose on "pragmatic grounds" between two sets of estimators of δ^* and hence achieve, in a sense, one of the main motivations behind pooling. A summary table of these MSE criteria is given by Wallace (1972, p. 697). McElroy (1977) extends these MSE criteria to the case where $u \sim N(0, \sigma^2\Sigma)$.

Monte Carlo Evidence

In the Monte Carlo study by Baltagi (1981a), the Chow test is performed given that the data are poolable and the model is generated as a two-way error component model. This test gave a high frequency of rejecting the null hypothesis when true. The reason for the poor performance of the Chow test is that it is applicable only under assumption 4.1 on the disturbances. This is violated under a random effects model with large variance components. For example, in testing the stability of cross-section regressions over time, the high frequency of type I error occurred whenever the variance components due to the time effects are not relatively small. Similarly, in testing the stability of time-series regressions across regions, the high frequency of type I error occurred whenever the variance components due to the cross-section effects are not relatively small.

Under this case of nonspherical disturbances, the proper test to perform is the Roy–Zellner test given by (4.13). Applying this test knowing the true variance components or using the Amemiya (1971) and the Wallace and Hussain (1969)-type estimates of the variance components leads to low frequencies of committing a type I error. Therefore, if pooling is contemplated using an error component model, then the Roy–Zellner test should be used rather than the Chow test.

The alternative MSE criteria, developed by Toro-Vizcarrondo and Wallace (1968) and Wallace (1972), were applied to the error component model in order to choose between the pooled and the unpooled estimators. These weaker criteria gave a lower frequency of committing a type I error than the Chow test, but their performance was still poor when compared to the Roy–Zellner test. McElroy's (1977) extension of these weaker MSE criteria to the case of nonspherical disturbances performed well when compared with the Roy–Zellner test, and is recommended.

Recently, Bun (2004) focused on testing the poolability hypothesis across cross-section units assuming constant coefficients over time. In particular, this testing applies to panel data with a limited number of cross-section units, like countries or states observed over a long time period, i.e., with T larger than N. Bun (2004) uses Monte Carlo experiments to examine the actual size of various asymptotic procedures for testing the poolability hypothesis. Dynamic regression models as well as nonspherical disturbances are considered. Results show that the classical asymptotic tests have poor finite sample performance, while their bootstrapped counterparts lead to more accurate inference. An empirical example is given using panel data on GDP growth and unemployment rates in 14 OECD countries over the period 1966–90. For this data set, it is shown that the classical asymptotic tests reject poolability while their bootstrap counterparts do not.

4.1.3 Examples

Example 1: Grunfeld Investment Equation

For the Grunfeld data, Chow's test for poolability across firms as in (4.1) gives an observed F-statistic of 27.75 and is distributed as $F(27, 170)$ under $H_0 : \delta_i = \delta$ for $i = 1, \ldots, N$. The RRSS = 1755850.48 is obtained from pooled OLS, and the URSS = 324728.47 is obtained from summing the RSS from 10 individual firm OLS regressions, each with 17 degrees of freedom. There are 27 restrictions and the test rejects poolability across firms for all coefficients. One can test for poolability of slopes only, allowing for varying intercepts. The restricted model is the Within regression with firm dummies. The RRSS = 523478, while the unrestricted regression is the same as above. The observed F-statistic is 5.78 which is distributed as $F(18, 170)$ under $H_0 : \beta_i = \beta$ for $i = 1, \ldots, N$. This again is significant at the 5% level and rejects poolability of the slopes across firms. Note that one could have tested poolability across time. The Chow test gives an observed value of 1.12 which is distributed as $F(57, 140)$. This does not reject poolability across time, but the unrestricted model is based on 20 regressions each with only 7 degrees of freedom. As is clear from the numerator degrees of freedom, this F-statistic tests 57 restrictions. The Roy–Zellner test for poolability across firms, allowing for one-way error component disturbances, yields an observed F-value of 4.35 which is distributed as $F(27, 170)$ under $H_0 : \delta_i = \delta$ for $i = 1, \ldots, N$. This still rejects poolability across firms even after allowing for one-way error component disturbances. The Roy–Zellner test for poolability over time, allowing for a one-way error component model, yields an F-value of 2.72 which is distributed as $F(57, 140)$ under $H_0 : \delta_t = \delta$ for $t = 1, \ldots, T$.

Example 2: Gasoline Demand

For the gasoline demand data in Baltagi and Griffin (1983), Chow's test for poolability across countries yields an observed F-statistic of 129.38 and is distributed as $F(68, 270)$ under $H_0 : \delta_i = \delta$ for $i = 1, \ldots, N$. This tests the stability of four time-series regression coefficients across 18 countries. The unrestricted SSE is based upon 18 OLS time-series regressions, one for each country. For the stability of the slope coefficients only, $H_0 : \beta_i = \beta$, an observed F-value of 27.33 is obtained which is distributed as $F(51, 270)$ under the null. Chow's test for poolability across time yields an F-value of 0.276 which is distributed as $F(72, 266)$ under $H_0 : \delta_t = \delta$ for $t = 1, \ldots, T$. This tests the stability of four cross-section regression coefficients across 19 time periods. The unrestricted SSE is based upon 19 OLS cross-section regressions, one for each year. This does not reject poolability across time periods. The Roy–Zellner test for poolability across countries, allowing for a one-way error component model, yields an F-value of 21.64 which is distributed as $F(68, 270)$ under $H_0 : \delta_i = \delta$ for $i = 1, \ldots, N$. The Roy–Zellner test for poolability across time yields an F-value of 1.66 which is distributed as $F(72, 266)$ under $H_0 : \delta_t = \delta$ for $t = 1, \ldots, T$. This rejects H_0 at the 5% level.

4.1.4 Other Tests for Poolability

Ziemer and Wetzstein (1983) suggest comparing pooled estimators (like $\widehat{\delta}_{\mathrm{OLS}}$) with nonpooled estimators (like $\widehat{\delta}_{i,\mathrm{OLS}}$) according to their forecast risk performance. Using a wilderness recreation demand model, they show that a Stein rule estimator gives a better forecast risk performance than the pooled or individual cross-section estimators. The Stein rule estimator for δ_i in (4.1) is given by

$$\widehat{\delta}_i^* = \widehat{\delta}_{\mathrm{OLS}} + \left(1 - \frac{c}{F_{obs}}\right)\left(\widehat{\delta}_{i,\mathrm{OLS}} - \widehat{\delta}_{\mathrm{OLS}}\right) \tag{4.14}$$

where $\widehat{\delta}_{i,\mathrm{OLS}}$ is given in (4.6) and $\widehat{\delta}_{\mathrm{OLS}}$ is given in (4.4). F_{obs} is the F-statistic to test $H_0 : \delta_i = \delta$, given in (4.8), and the constant c is given by $c = ((N-1)K' - 2)/(NT - NK' + 2)$. Note that $\widehat{\delta}_i^*$ shrinks $\widehat{\delta}_{i,\mathrm{OLS}}$ towards the pooled estimator $\widehat{\delta}_{\mathrm{OLS}}$. More recently, Maddala (1991) argued that shrinkage estimators appear to be better than either the pooled estimator or the individual cross-section estimators.

Brown, Durbin and Evans (1975) derived cumulative sum and cumulative sum of squares tests for structural change based on recursive residuals in a time-series regression. Han and Park (1989) extend these tests of structural change to the panel data case. They apply these tests to a study of US foreign trade of manufacturing goods. They find no evidence of structural change over the period 1958–76. Baltagi, Hidalgo and Li (1996) derive a nonparametric test for poolability which is robust to functional form misspecification. In particular, they consider the following nonparametric panel data model:

$$y_{it} = g_t(x_{it}) + \epsilon_{it} \quad i = 1, \ldots, N;\ t = 1, \ldots, T$$

where $g_t(.)$ is an unspecified functional form that may vary over time. x_{it} is a $k \times 1$ column vector of predetermined explanatory variables with ($p \geq 1$) variables being continuous and $k - p$ (≥ 0). Poolability of the data over time is equivalent to testing that $g_t(x) = g_s(x)$ almost everywhere for all t and $s = 1, 2, \ldots, T$ versus $g_t(x) \neq g_s(x)$ for some $t \neq s$ with probability greater than zero. The test statistic is shown to be consistent and asymptotically normal and is applied to an earnings equation using data from the PSID.

4.2 TESTS FOR INDIVIDUAL AND TIME EFFECTS

4.2.1 The Breusch–Pagan Test

For the random two-way error component model, Breusch and Pagan (1980) derived a Lagrange multiplier (LM) test to test $H_0 : \sigma_\mu^2 = \sigma_\lambda^2 = 0$. The loglikelihood function under normality of the disturbances is given by (3.29) as

$$L(\delta, \theta) = \text{constant} - \frac{1}{2}\log \mid \Omega \mid - \frac{1}{2}u'\Omega^{-1}u \tag{4.15}$$

where $\theta' = (\sigma_\mu^2, \sigma_\lambda^2, \sigma_\nu^2)$ and Ω is given by (3.10) as

$$\Omega = \sigma_\mu^2(I_N \otimes J_T) + \sigma_\lambda^2(J_N \otimes I_T) + \sigma_\nu^2 I_{NT} \tag{4.16}$$

The information matrix is block-diagonal between θ and δ. Since H_0 involves only θ, the part of the information matrix due to δ is ignored. In order to reconstruct the Breusch and Pagan (1980) LM statistic, we need the score $D(\widetilde{\theta}) = (\partial L/\partial\theta) \mid_{\widetilde{\theta}_{mle}}$, the first derivative of the likelihood with respect to θ, evaluated at the restricted MLE of θ under H_0, which is denoted by $\widetilde{\theta}_{mle}$. Hartley and Rao (1967) or Hemmerle and Hartley (1973) give a useful general formula to obtain $D(\theta)$:

$$\partial L/\partial\theta_r = \frac{1}{2}\text{tr}[\Omega^{-1}(\partial\Omega/\partial\theta_r)] + \frac{1}{2}[u'\Omega^{-1}(\partial\Omega/\partial\theta_r)\Omega^{-1}u] \tag{4.17}$$

for $r = 1, 2, 3$. Also, from (4.16), $(\partial\Omega/\partial\theta_r) = (I_N \otimes J_T)$ for $r = 1$; $(J_N \otimes I_T)$ for $r = 2$ and I_{NT} for $r = 3$. The restricted MLE of Ω under H_0 is $\widetilde{\Omega} = \widetilde{\sigma}_\nu^2 I_{NT}$ where $\widetilde{\sigma}_\nu^2 = \widetilde{u}'\widetilde{u}/NT$ and $\widetilde{u}$ are the OLS residuals. Using $\text{tr}(I_N \otimes J_T) = \text{tr}(J_N \otimes I_T) = \text{tr}(I_{NT}) = NT$, one gets

$$\begin{aligned} D(\widetilde{\theta}) &= \begin{bmatrix} -\frac{1}{2}\text{tr}[(I_N \otimes J_T)/\widetilde{\sigma}_\nu^2] + \frac{1}{2}[\widetilde{u}'(I_N \otimes J_T)\widetilde{u}/\widetilde{\sigma}_\nu^4] \\ -\frac{1}{2}\text{tr}[(J_N \otimes I_T)/\widetilde{\sigma}_\nu^2] + \frac{1}{2}[\widetilde{u}'(J_N \otimes I_T)\widetilde{u}/\widetilde{\sigma}_\nu^4] \\ -\frac{1}{2}\text{tr}[I_{NT}/\widetilde{\sigma}_\nu^2] + \frac{1}{2}[\widetilde{u}'\widetilde{u}/\widetilde{\sigma}_\nu^4] \end{bmatrix} \\ &= \frac{-NT}{2\widetilde{\sigma}_\nu^2}\begin{bmatrix} 1 - \dfrac{\widetilde{u}'(I_N \otimes J_T)\widetilde{u}}{\widetilde{u}'\widetilde{u}} \\ 1 - \dfrac{\widetilde{u}'(J_N \otimes I_T)\widetilde{u}}{\widetilde{u}'\widetilde{u}} \\ 0 \end{bmatrix} \end{aligned} \tag{4.18}$$

The information matrix for this model is

$$J(\theta) = E\left[\frac{\partial^2 L}{\partial\theta\partial\theta'}\right] = [J_{rs}] \quad \text{for } r, s = 1, 2, 3$$

where

$$J_{rs} = E[-\partial^2 L/\partial\theta_r\partial\theta_s] = \frac{1}{2}tr[\Omega^{-1}(\partial\Omega/\partial\theta_r)\Omega^{-1}(\partial\Omega/\partial\theta_s)] \tag{4.19}$$

(see Harville, 1977). Using $\widetilde{\Omega}^{-1} = (1/\widetilde{\sigma}_\nu^2)I_{NT}$ and $\text{tr}[(I_N \otimes J_T)(J_N \otimes I_T)] = \text{tr}(J_{NT}) = NT$, $\text{tr}(I_N \otimes J_T)^2 = NT^2$ and $\text{tr}(J_N \otimes I_T)^2 = N^2T$, one gets

$$\begin{aligned}\widetilde{J} &= \frac{1}{2\widetilde{\sigma}_\nu^4}\begin{bmatrix} \text{tr}(I_N \otimes J_T)^2 & \text{tr}(J_{NT}) & \text{tr}(I_N \otimes J_T) \\ \text{tr}(J_{NT}) & \text{tr}(J_N \otimes I_T)^2 & \text{tr}(J_N \otimes I_T) \\ \text{tr}(I_N \otimes J_T) & \text{tr}(J_N \otimes I_T) & \text{tr}(I_{NT}) \end{bmatrix} \\ &= \frac{NT}{2\widetilde{\sigma}_\nu^4}\begin{bmatrix} T & 1 & 1 \\ 1 & N & 1 \\ 1 & 1 & 1 \end{bmatrix}\end{aligned} \tag{4.20}$$

with

$$\widetilde{J}^{-1} = \frac{2\widetilde{\sigma}_\nu^4}{NT(N-1)(T-1)}\begin{bmatrix} (N-1) & 0 & (1-N) \\ 0 & (T-1) & (1-T) \\ (1-N) & (1-T) & (NT-1) \end{bmatrix} \tag{4.21}$$

Therefore

$$\begin{aligned}\text{LM} &= \widetilde{D}'\widetilde{J}^{-1}\widetilde{D} \\ &= \frac{NT}{2(N-1)(T-1)}\left[(N-1)\left[1 - \frac{\widetilde{u}'(I_N \otimes J_T)\widetilde{u}}{\widetilde{u}'\widetilde{u}}\right]^2\right. \\ &\quad \left. +(T-1)\left[1 - \frac{\widetilde{u}'(J_N \otimes I_T)\widetilde{u}}{\widetilde{u}'\widetilde{u}}\right]^2\right] \\ \text{LM} &= \text{LM}_1 + \text{LM}_2\end{aligned} \tag{4.22}$$

where

$$\text{LM}_1 = \frac{NT}{2(T-1)}\left[1 - \frac{\widetilde{u}'(I_N \otimes J_T)\widetilde{u}}{\widetilde{u}'\widetilde{u}}\right]^2 \tag{4.23}$$

and

$$\text{LM}_2 = \frac{NT}{2(N-1)}\left[1 - \frac{\widetilde{u}'(J_N \otimes I_T)\widetilde{u}}{\widetilde{u}'\widetilde{u}}\right]^2 \tag{4.24}$$

Under H_0, LM is asymptotically distributed as a χ_2^2. This LM test requires only OLS residuals and is easy to compute. This may explain its popularity. In addition, if one wants to test $H_0^a : \sigma_\mu^2 = 0$, following the derivation given above, one gets LM_1 which is asymptotically distributed under H_0^a as χ_1^2. Similarly, if one wants to test $H_0^b : \sigma_\lambda^2 = 0$, by symmetry, one gets LM_2 which is asymptotically distributed as χ_1^2 under H_0^b. This LM test performed well in Monte Carlo studies (see Baltagi, 1981a), except for small values of σ_μ^2 and σ_λ^2 close to zero.

These are precisely the cases where negative estimates of the variance components are most likely to occur.[2]

4.2.2 King and Wu, Honda and the Standardized Lagrange Multiplier Tests

One problem with the Breusch–Pagan test is that it assumes that the alternative hypothesis is two-sided when we know that the variance components are nonnegative. This means that the alternative hypotheses should be one-sided. Honda (1985) suggests a *uniformly most powerful* test for $H_0^a : \sigma_\mu^2 = 0$ which is based upon

$$\mathrm{HO} \equiv A = \sqrt{\frac{NT}{2(T-1)}}\left[\frac{\widetilde{u}'(I_N \otimes J_T)\widetilde{u}}{\widetilde{u}'\widetilde{u}} - 1\right] \overset{H_0^a}{\to} N(0,1) \tag{4.25}$$

Note that the square of this $N(0, 1)$ statistic is the Breusch and Pagan (1980) LM_1 test statistic given in (4.23). Honda (1985) finds that this test statistic is robust to nonnormality.[3] Moulton and Randolph (1989) showed that the asymptotic $N(0, 1)$ approximation for this one-sided LM statistic can be poor even in large samples. This occurs when the number of regressors is large or the intraclass correlation of some of the regressors is high. They suggest an alternative standardized Lagrange multiplier (SLM) test whose asymptotic critical values are generally closer to the exact critical values than those of the LM test. This SLM test statistic centers and scales the one-sided LM statistic so that its mean is zero and its variance is one:

$$\mathrm{SLM} = \frac{\mathrm{HO} - E(\mathrm{HO})}{\sqrt{\mathrm{var(\,HO)}}} = \frac{d - E(d)}{\sqrt{\mathrm{var}(d)}} \tag{4.26}$$

where $d = \widetilde{u}'D\widetilde{u}/\widetilde{u}'\widetilde{u}$ and $D = (I_N \otimes J_T)$. Using the results on moments of quadratic forms in regression residuals (see e.g. Evans and King, 1985), we get

$$E(d) = \mathrm{tr}(D\bar{P}_Z)/p \tag{4.27}$$

and

$$\mathrm{var}(d) = 2\{p\ \mathrm{tr}(D\bar{P}_Z)^2 - [\mathrm{tr}(D\bar{P}_Z)]^2\}/p^2(p+2) \tag{4.28}$$

where $p = n - (K + 1)$ and $\bar{P}_Z = I_n - Z(Z'Z)^{-1}Z'$. Under the null hypothesis H_0^a, SLM has an asymptotic $N(0, 1)$ distribution. King and Wu (1997) suggest a locally mean most powerful (LMMP) one-sided test, which for H_0^a coincides with Honda's (1985) uniformly most powerful test (see Baltagi, Chang and Li, 1992b).

Similarly, for $H_0^b : \sigma_\lambda^2 = 0$, the one-sided Honda-type LM test statistic is

$$B = \sqrt{\frac{NT}{2(N-1)}}\left[\frac{\widetilde{u}'(J_N \otimes I_T)\widetilde{u}}{\widetilde{u}'\widetilde{u}} - 1\right] \tag{4.29}$$

which is asymptotically distributed as $N(0, 1)$. Note that the square of this statistic is the corresponding two-sided LM test given by LM_2 in (4.24). This can be standardized as in (4.26) with $D = (J_N \otimes I_T)$. Also, the King and Wu (1997) LMMP test for H_0^b coincides with Honda's uniformly most powerful test given in (4.29).

For $H_0^c : \sigma_\mu^2 = \sigma_\lambda^2 = 0$, the two-sided LM test, given by Breusch and Pagan (1980), is $A^2 + B^2 \sim \chi^2(2)$. Honda (1985) does not derive a uniformly most powerful one-sided test for H_0^c, but he suggests a "handy" one-sided test given by $(A + B)/\sqrt{2}$ which is distributed as

$N(0, 1)$ under H_0^c. Following King and Wu (1997), Baltagi et al. (1992b) derived the LMMP one-sided test for H_0^c. This is given by

$$\text{KW} = \frac{\sqrt{T-1}}{\sqrt{N+T-2}} A + \frac{\sqrt{N-1}}{\sqrt{N+T-2}} B \tag{4.30}$$

which is distributed as $N(0, 1)$ under H_0^c. See problem 4.6.

Following the Moulton and Randolph (1989) standardization of the LM test for the one-way error component model, Honda (1991) suggested a standardization of his 'handy' one-sided test for the two-way error component model. In fact, for $\text{HO} = (A + B)/\sqrt{2}$, the SLM test is given by (4.26) with $d = \widetilde{u}' D \widetilde{u} / \widetilde{u}' \widetilde{u}$, and

$$D = \frac{1}{2} \sqrt{\frac{NT}{(T-1)}} (I_N \otimes J_T) + \frac{1}{2} \sqrt{\frac{NT}{(N-1)}} (J_N \otimes I_T) \tag{4.31}$$

Also, one can similarly standardize the KW test given in (4.30) by subtracting its mean and dividing by its standard deviation, as in (4.26), with $d = \widetilde{u}' D \widetilde{u} / \widetilde{u}' \widetilde{u}$ and

$$D = \frac{\sqrt{NT}}{\sqrt{2}\sqrt{N+T-2}} [(I_N \otimes J_T) + (J_N \otimes I_T)] \tag{4.32}$$

With this new D matrix, $E(d)$ and $\text{var}(d)$ can be computed using (4.27) and (4.28). Under $H_0^c : \sigma_\mu^2 = \sigma_\lambda^2 = 0$, these SLM statistics are asymptotically $N(0, 1)$ and their asymptotic critical values should be closer to the exact critical values than those of the corresponding unstandardized tests.

4.2.3 Gourieroux, Holly and Monfort Test

Note that A or B can be negative for a specific application, especially when one or both variance components are small and close to zero. Following Gourieroux, Holly and Monfort (1982), hereafter GHM, Baltagi et al. (1992b) proposed the following test for H_0^c:

$$\chi_m^2 = \begin{cases} A^2 + B^2 & \text{if } A > 0,\ B > 0 \\ A^2 & \text{if } A > 0,\ B \le 0 \\ B^2 & \text{if } A \le 0,\ B > 0 \\ 0 & \text{if } A \le 0,\ B \le 0 \end{cases} \tag{4.33}$$

χ_m^2 denotes the mixed χ^2 distribution. Under the null hypothesis,

$$\chi_m^2 \sim \left(\frac{1}{4}\right) \chi^2(0) + \left(\frac{1}{2}\right) \chi^2(1) + \left(\frac{1}{4}\right) \chi^2(2)$$

where $\chi^2(0)$ equals zero with probability one.[4] The weights $\left(\frac{1}{4}\right)$, $\left(\frac{1}{2}\right)$ and $\left(\frac{1}{4}\right)$ follow from the fact that A and B are asymptotically independent of each other and the results in Gourieroux et al. (1982). This proposed test has the advantage over the Honda and KW tests in that it is immune to the possible negative values of A and B.

4.2.4 Conditional LM Tests

When one uses HO given in (4.25) to test $H_0^a : \sigma_\mu^2 = 0$ one implicitly assumes that the time-specific effects do not exist. This may lead to incorrect decisions especially when the variance of the time effects (assumed to be zero) is large. To overcome this problem, Baltagi et al. (1992b) suggest testing the individual effects conditional on the time-specific effects (i.e. allowing $\sigma_\lambda^2 > 0$). The corresponding LM test for testing $H_0^d : \sigma_\mu^2 = 0$ (allowing $\sigma_\lambda^2 > 0$) is derived in appendix 2 of Baltagi et al. (1992b) and is given by

$$\text{LM}_\mu = \frac{\sqrt{2}\widetilde{\sigma}_2^2\widetilde{\sigma}_\nu^2}{\sqrt{T(T-1)[\widetilde{\sigma}_\nu^4 + (N-1)\widetilde{\sigma}_2^4]}}\widetilde{D}_\mu \tag{4.34}$$

where

$$\widetilde{D}_\mu = \frac{T}{2}\left\{\frac{1}{\widetilde{\sigma}_2^2}\left[\frac{\widetilde{u}'(\bar{J}_N \otimes \bar{J}_T)\widetilde{u}}{\widetilde{\sigma}_2^2} - 1\right] + \frac{(N-1)}{\widetilde{\sigma}_\nu^2}\left[\frac{\widetilde{u}'(E_N \otimes \bar{J}_T)\widetilde{u}}{(N-1)\widetilde{\sigma}_\nu^2} - 1\right]\right\} \tag{4.35}$$

with $\widetilde{\sigma}_2^2 = \widetilde{u}'(\bar{J}_N \otimes I_T)\widetilde{u}/T$ and $\widetilde{\sigma}_\nu^2 = \widetilde{u}'(E_N \otimes I_T)\widetilde{u}/T(N-1)$. LM_μ is asymptotically distributed as $N(0, 1)$ under H_0^d. The estimated disturbances $\widetilde{u}$ denote the one-way GLS residuals using the maximum likelihood estimates $\widetilde{\sigma}_\nu^2$ and $\widetilde{\sigma}_2^2$. One can easily check that if $\widetilde{\sigma}_\lambda^2 \to 0$, then $\widetilde{\sigma}_2^2 \to \widetilde{\sigma}_\nu^2$ and LM_μ given in (4.34), tends to the one-sided Honda test given in (4.25).

Similarly, the alternative LM test statistic for testing $H_0^e : \sigma_\lambda^2 = 0$ (allowing $\sigma_\mu^2 > 0$) can be obtained as follows:

$$\text{LM}_\lambda = \frac{\sqrt{2}\widetilde{\sigma}_1^2\widetilde{\sigma}_\nu^2}{\sqrt{N(N-1)[\widetilde{\sigma}_\nu^4 + (T-1)\widetilde{\sigma}_1^4]}}\widetilde{D}_\lambda \tag{4.36}$$

where

$$\widetilde{D}_\lambda = \frac{N}{2}\left\{\frac{1}{\widetilde{\sigma}_1^2}\left[\frac{\widetilde{u}'(\bar{J}_N \otimes \bar{J}_T)\widetilde{u}}{\widetilde{\sigma}_1^2} - 1\right] + \frac{(T-1)}{\widetilde{\sigma}_\nu^2}\left[\frac{\widetilde{u}'(\bar{J}_N \otimes E_T)\widetilde{u}}{(T-1)\widetilde{\sigma}_\nu^2} - 1\right]\right\} \tag{4.37}$$

with $\widetilde{\sigma}_1^2 = \widetilde{u}'(I_N \otimes \bar{J}_T)\widetilde{u}/N$, $\widetilde{\sigma}_\nu^2 = \widetilde{u}'(I_N \otimes E_T)\widetilde{u}/N(T-1)$. The test statistic LM_λ is asymptotically distributed as $N(0, 1)$ under H_0^e.

4.2.5 ANOVA *F* and the Likelihood Ratio Tests

Moulton and Randolph (1989) found that the ANOVA F-test, which tests the significance of the fixed effects, performs well for the one-way error component model. The ANOVA F-test statistics have the following familiar general form:

$$F = \frac{y'MD(D'MD)^- D'My/(p-r)}{y'Gy/[NT - (\widetilde{k} + p - r)]} \tag{4.38}$$

Under the null hypothesis, this statistic has a central F-distribution with $p - r$ and $NT - (\widetilde{k} + p - r)$ degrees of freedom. For $H_0^a : \sigma_\mu^2 = 0$, $D = I_N \otimes \iota_T$, $M = \bar{P}_Z$, $\widetilde{k} = K'$, $p = N$, $r = K' + N -$ rank(Z, D) and $G = \bar{P}_{(Z,D)}$ where $\bar{P}_Z = I - P_Z$ and $P_Z = Z(Z'Z)^{-1}Z'$. For details regarding other hypotheses, see Baltagi et al. (1992b).

The one-sided likelihood ratio (LR) tests all have the following form:

$$\text{LR} = -2\log\frac{l(\text{res})}{l(\text{unres})} \tag{4.39}$$

where $l(\text{res})$ denotes the restricted maximum likelihood value (under the null hypothesis), while $l(\text{unres})$ denotes the unrestricted maximum likelihood value. The LR tests require MLE estimators of the one-way and the two-way models and are comparatively more expensive than their LM counterparts. Under the null hypotheses considered, the LR test statistics have the same asymptotic distributions as their LM counterparts (see Gourieroux et al., 1982). More specifically, for H_0^a, H_0^b, H_0^d and H_0^e, LR $\sim (\frac{1}{2})\chi^2(0) + (\frac{1}{2})\chi^2(1)$ and for H_0^c, LR $\sim (\frac{1}{4})\chi^2(0) + (\frac{1}{2})\chi^2(1) + (\frac{1}{4})\chi^2(2)$.

4.2.6 Monte Carlo Results

Baltagi et al. (1992b) compared the performance of the above tests using Monte Carlo experiments on the two-way error component model described in Baltagi (1981a). Each experiment involved 1000 replications. For each replication, the following test statistics were computed: BP, Honda, KW, SLM, LR, GHM and the F-test statistics. The results can be summarized as follows: when $H_0^a : \sigma_\mu^2 = 0$ is true but σ_λ^2 is large, all the usual tests for H_0^a perform badly since they ignore the fact that $\sigma_\lambda^2 > 0$. In fact, the two-sided BP test performs the worst, over-rejecting the null, while HO, SLM, LR and F underestimate the nominal size. As σ_μ^2 gets large, all the tests perform well in rejecting the null hypothesis H_0^a. But, for small $\sigma_\mu^2 > 0$, the power of all the tests considered deteriorates as σ_λ^2 increases.

For testing $H_0^d : \sigma_\mu^2 = 0$ (allowing $\sigma_\lambda^2 > 0$), LM$_\mu$, LR and F perform well with their estimated size not significantly different from their nominal size. Also, for large σ_μ^2 all these tests have high power rejecting the null hypothesis in 98–100% of cases. The results also suggest that overspecifying the model, i.e. assuming the model is two-way ($\sigma_\lambda^2 > 0$) when in fact it is one-way ($\sigma_\lambda^2 = 0$), does not seem to hurt the power of these tests. Finally, the power of all tests improves as σ_λ^2 increases. This is in sharp contrast to the performance of the tests that ignore the fact that $\sigma_\lambda^2 > 0$. The Monte Carlo results strongly support the fact that one should not ignore the possibility that $\sigma_\lambda^2 > 0$ when testing $\sigma_\mu^2 = 0$. In fact, it may be better to overspecify the model rather than underspecify it in testing the variance components.

For the joint test $H_0^c : \sigma_\mu^2 = \sigma_\lambda^2 = 0$, the BP, HO, KW and LR significantly underestimate the nominal size, while GHM and the F-test have estimated sizes that are not significantly different from the nominal size. Negative values of A and B make the estimated size for HO and KW underestimate the nominal size. For these cases, the GHM test is immune to negative values of A and B, and performs well in the Monte Carlo experiments. Finally, the ANOVA F-tests perform reasonably well when compared to the LR and LM tests, for both the one-way and two-way models and are recommended. This confirms similar results on the F-statistic by Moulton and Randolph (1989) for the one-way error component model.

Baltagi, Bresson and Pirotte (2003b) compared the performance of the usual panel estimators and a pretest estimator for the two-way error component model using Monte Carlo experiments. The only type of misspecification considered is whether one or both variance components are actually zero. The pretest estimator is based on the application of the GHM test first, followed by the conditional LM tests of Baltagi et al. (1992b), i.e., LM$_\mu$ and LM$_\lambda$. If GHM does not reject the null, the pretest estimator reduces to OLS. If the null is rejected, LM$_\mu$ and LM$_\lambda$ are performed and depending on the outcome, the pretest estimator reduces to a one-way or two-way feasible GLS estimator. Some of the Monte Carlo results are the following: the correct FGLS or MLE are the best in terms of relative MSE performance with respect to true GLS. However, the researcher does not have perfect foresight regarding the true specification. The pretest estimator is a close second in MSE performance to the correct FGLS estimator for all type of misspecification considered. The wrong fixed effects or random effects FGLS estimators suffer from a huge loss of MSE. These results were checked for the nonnormality assumption as well as using double the significance levels (10% rather than 5%) for the pretest estimator.

4.2.7 An Illustrative Example

The Monte Carlo results show that the test statistics A and/or B take on large negative values quite often under some designs. A natural question is whether a large negative A and/or B is possible for real data. In this subsection, we apply the tests considered above to the Grunfeld (1958) investment equation. Table 4.1 gives the observed test statistics. The null hypotheses $H_0^c : \sigma_\mu^2 = \sigma_\lambda^2 = 0$, as well as $H_0^a : \sigma_\mu^2 = 0$ and $H_0^d : \sigma_\mu^2 = 0$ (allowing $\sigma_\lambda^2 > 0$) are rejected by all tests considered. Clearly, all the tests strongly suggest that there are individual-specific effects. However, for testing time-specific effects, except for the two-sided LM (BP) test which rejects $H_0^b : \sigma_\lambda^2 = 0$, all the tests suggest that there are no time-specific effects for this data. The conflict occurs because B takes on a large negative value (-2.540) for this data set. This means that the two-sided LM test is $B^2 = 6.454$, which is larger than the χ_1^2 critical value (3.841), whereas the one-sided LM, SLM, LR and F-tests for this hypothesis do not reject H_0^b. In fact, the LM_λ test proposed by Baltagi et al. (1992b) for testing $H_0^e : \sigma_\lambda^2 = 0$ (allowing $\sigma_\mu^2 > 0$) as well as the LR and F-tests for this hypothesis do not reject H_0^e. These data clearly support the use of the one-sided test in empirical applications. Stata reports the LM (BP) test for $H_0^a : \sigma_\mu^2 = 0$ using (xttest0). This is given in Table 4.2 for the Grunfeld data and computes the A^2 term in (4.23) of 798.16 which is the number reported in Table 4.1. Stata also reports the LR test for H_0^a at the bottom of the MLE results using (xtreg,mle). This replicates the observed LR test statistic of 193.04 in Table 4.1. The Stata output is not reproduced here but

Table 4.1 Test Results for the Grunfeld Example*

Null Hypothesis Tests	H_0^a $\sigma_\mu^2 = 0$	H_0^b $\sigma_\lambda^2 = 0$	H_0^c $\sigma_\mu^2 = \sigma_\lambda^2 = 0$	H_0^d $\sigma_\mu^2 = 0/\sigma_\lambda^2 > 0$	H_0^e $\sigma_\lambda^2 = 0/\sigma_\mu^2 > 0$
BP	798.162	6.454	804.615	—	—
	(3.841)	(3.841)	(5.991)		
HO	28.252	−2.540	18.181	—	—
	(1.645)	(1.645)	(1.645)		
KW	28.252	−2.540	21.832	—	—
	(1.645)	(1.645)	(1.645)		
SLM	32.661	−2.433	—	—	—
	(1.645)	(1.645)			
GHM	—	—	798.162	—	—
			(4.231)		
F	49.177	0.235	17.403	52.672	1.142
	(1.930)	(1.645)	(1.543)	(1.648)	(1.935)
LR	193.091	0	193.108	193.108	0.017
	(2.706)	(2.706)	(4.231)	(2.706)	(2.706)
LM_μ	—	—	—	28.252	—
				(2.706)	
LM_λ	—	—	—	—	0.110
					(2.706)

*Numbers in parentheses are asymptotic critical values at the 5% level.
Source: Baltagi et al. (1992b). Reproduced by permission of Elsevier Science Publishers B.V. (North Holland).

Table 4.2 Grunfeld's Data. Breusch and Pagan Lagrangian Multiplier Test.

```
. xttest0

Breusch and Pagan Lagrangian multiplier test for random effects:

        I[fn,t] = Xb + u[fn] + e[fn,t]

        Estimated results:
                         |       Var     sd = sqrt(Var)
                ---------+-----------------------------
                       I |    47034.89      216.8753
                       e |    2784.458      52.76796
                       u |      7089.8      84.20095

         Test:   Var(u) = 0
                              chi2(1) =   798.16
                          Prob > chi2 =     0.0000
```

one can refer to the Stata results in Table 2.10 where we reported the MLE for the public capital productivity data. The bottom of Table 2.10 reports the observed LR test statistic of 1149.84. This shows that the random state effects are significant and their variance is not 0. Also note that the fixed effects Stata output (xtreg,fe) reports the F-test for the significance of the fixed individual effects. For the Grunfeld data, this replicates the F(9,188) value of 49.18 which is the number reported in Table 4.1. The Stata output is not reproduced here, but one can refer to the Stata results in Table 2.8 where we reported the fixed effects estimates for the public capital productivity data. The bottom of Table 2.8 reports the observed F(47,764) value of 75.82. This shows that the fixed state effects are significant.

4.3 HAUSMAN'S SPECIFICATION TEST

A critical assumption in the error component regression model is that $E(u_{it}/X_{it}) = 0$. This is important given that the disturbances contain individual invariant effects (the μ_i) which are unobserved and may be correlated with the X_{it}. For example, in an earnings equation these μ_i may denote unobservable ability of the individual and this may be correlated with the schooling variable included on the right-hand side of this equation. In this case, $E(u_{it}/X_{it}) \neq 0$ and the GLS estimator $\widehat{\beta}_{\text{GLS}}$ becomes biased and inconsistent for β. However, the Within transformation wipes out these μ_i and leaves the Within estimator $\widetilde{\beta}_{\text{Within}}$ unbiased and consistent for β. Hausman (1978) suggests comparing $\widehat{\beta}_{\text{GLS}}$ and $\widetilde{\beta}_{\text{Within}}$, both of which are consistent under the null hypothesis $H_0 : E(u_{it}/X_{it}) = 0$, but which will have different probability limits if H_0 is not true. In fact, $\widetilde{\beta}_{\text{Within}}$ is consistent whether H_0 is true or not, while $\widehat{\beta}_{\text{GLS}}$ is BLUE, consistent and asymptotically efficient under H_0, but is inconsistent when H_0 is false. A natural test statistic would be based on $\widehat{q}_1 = \widehat{\beta}_{\text{GLS}} - \widetilde{\beta}_{\text{Within}}$. Under H_0, $\text{plim}\,\widehat{q}_1 = 0$ and $\text{cov}(\widehat{q}_1, \widehat{\beta}_{\text{GLS}}) = 0$.

Using the fact that $\widehat{\beta}_{\text{GLS}} - \beta = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}u$ and $\widetilde{\beta}_{\text{Within}} - \beta = (X'QX)^{-1}X'Qu$, one gets $E(\widehat{q}_1) = 0$ and

$$\begin{aligned}
\text{cov}(\widehat{\beta}_{\text{GLS}}, \widehat{q}_1) &= \text{var}(\widehat{\beta}_{\text{GLS}}) - \text{cov}(\widehat{\beta}_{\text{GLS}}, \widetilde{\beta}_{\text{Within}}) \\
&= (X'\Omega^{-1}X)^{-1} - (X'\Omega^{-1}X)^{-1}X\Omega^{-1}E(uu')QX(X'QX)^{-1} \\
&= (X'\Omega^{-1}X)^{-1} - (X'\Omega^{-1}X)^{-1} = 0
\end{aligned}$$

Using the fact that $\widetilde{\beta}_{\text{Within}} = \widehat{\beta}_{\text{GLS}} - \widehat{q}_1$, one gets

$$\text{var}(\widetilde{\beta}_{\text{Within}}) = \text{var}(\widehat{\beta}_{\text{GLS}}) + \text{var}(\widehat{q}_1)$$

since $\text{cov}(\widehat{\beta}_{\text{GLS}}, \widehat{q}_1) = 0$. Therefore

$$\text{var}(\widehat{q}_1) = \text{var}(\widetilde{\beta}_{\text{Within}}) - \text{var}(\widehat{\beta}_{\text{GLS}}) = \sigma_\nu^2(X'QX)^{-1} - (X'\Omega^{-1}X)^{-1} \tag{4.40}$$

Hence, the Hausman test statistic is given by

$$m_1 = \hat{q}_1'[\text{var}(\widehat{q}_1)]^{-1}\widehat{q}_1 \tag{4.41}$$

and under H_0 is asymptotically distributed as χ^2_K where K denotes the dimension of slope vector β. In order to make this test operational, Ω is replaced by a consistent estimator $\widehat{\Omega}$, and GLS by its corresponding feasible GLS.

An alternative asymptotically equivalent test can be obtained from the augmented regression

$$y^* = X^*\beta + \widetilde{X}\gamma + w \tag{4.42}$$

where $y^* = \sigma_\nu\Omega^{-1/2}y$, $X^* = \sigma_\nu\Omega^{-1/2}X$ and $\widetilde{X} = QX$. Hausman's test is now equivalent to testing whether $\gamma = 0$. This is a standard Wald test for the omission of the variables $\widetilde{X}$ from (4.42).[5] It is worthwhile to rederive this test. In fact, performing OLS on (4.42) yields

$$\begin{pmatrix} \widehat{\beta} \\ \widehat{\gamma} \end{pmatrix} = \begin{bmatrix} X'(Q+\phi^2 P)X & X'QX \\ X'QX & X'QX \end{bmatrix}^{-1} \begin{pmatrix} X'(Q+\phi^2 P)y \\ X'Qy \end{pmatrix} \tag{4.43}$$

where $\sigma_\nu\Omega^{-1/2} = Q + \phi P$ and $\phi = \sigma_\nu/\sigma_1$ (see (2.20)). Using partitioned inverse formulas, one can show that

$$\begin{pmatrix} \widehat{\beta} \\ \widehat{\gamma} \end{pmatrix} = \begin{bmatrix} E & -E \\ -E & (X'QX)^{-1} + E \end{bmatrix} \begin{pmatrix} X'(Q+\phi^2 P)y \\ X'Qy \end{pmatrix} \tag{4.44}$$

where $E = (X'PX)^{-1}/\phi^2$. This reduces to

$$\widehat{\beta} = \widehat{\beta}_{\text{Between}} = (X'PX)^{-1}X'Py \tag{4.45}$$

and

$$\widehat{\gamma} = \widetilde{\beta}_{\text{Within}} - \widehat{\beta}_{\text{Between}} \tag{4.46}$$

Substituting the Within and Between estimators of β into (4.46) one gets

$$\widehat{\gamma} = (X'QX)^{-1}X'Q\nu - (X'PX)^{-1}X'Pu \tag{4.47}$$

It is easy to show that $E(\widehat{\gamma}) = 0$ and

$$\begin{aligned} \text{var}(\widehat{\gamma}) &= E(\widehat{\gamma}\widehat{\gamma}') = \sigma_\nu^2(X'QX)^{-1} + \sigma_1^2(X'PX)^{-1} \\ &= \text{var}(\widetilde{\beta}_{\text{Within}}) + \text{var}(\widehat{\beta}_{\text{Between}}) \end{aligned} \tag{4.48}$$

since the cross-product terms are zero. The test for $\gamma = 0$ is based on $\widehat{\gamma} = \widetilde{\beta}_{\text{Within}} - \widehat{\beta}_{\text{Between}} = 0$ and the corresponding test statistic would therefore be $\widehat{\gamma}'(\text{var}(\widehat{\gamma}))^{-1}\widehat{\gamma}$, which looks different from the Hausman m_1 statistic given in (4.41). These tests are numerically exactly identical (see Hausman and Taylor, 1981). In fact, Hausman and Taylor (1981) showed that H_0 can be tested using any of the following three paired differences: $\widehat{q}_1 = \widehat{\beta}_{\text{GLS}} - \widetilde{\beta}_{\text{Within}}$; $\widehat{q}_2 = \widehat{\beta}_{\text{GLS}} - \widehat{\beta}_{\text{Between}}$; or $\widehat{q}_3 = \widetilde{\beta}_{\text{Within}} - \widehat{\beta}_{\text{Between}}$. The corresponding test statistics can be computed as $m_i = \widehat{q}_i'V_i^{-1}\widehat{q}_i$, where $V_i = \text{var}(\widehat{q}_i)$. These are asymptotically distributed as χ^2_K for $i = 1, 2, 3$ under H_0.[6] Hausman and Taylor (1981) proved that these three tests differ from each other by nonsingular matrices. This easily follows from the fact that

$$\widehat{\beta}_{\text{GLS}} = W_1\widetilde{\beta}_{\text{Within}} + (I - W_1)\widehat{\beta}_{\text{Between}}$$

derived in (2.31). So $\widehat{q}_1 = \widehat{\beta}_{\text{GLS}} - \widetilde{\beta}_{\text{Within}} = (I - W_1)(\widehat{\beta}_{\text{Between}} - \widetilde{\beta}_{\text{Within}}) = \Gamma\widehat{q}_3$, where $\Gamma = W_1 - I$. Also, $\text{var}(\widehat{q}_1) = \Gamma\text{var}(\widehat{q}_3)\Gamma'$ and

$$\begin{aligned} m_1 &= \widehat{q}_1'[\text{var}(\widehat{q}_1)]^{-1}\widehat{q}_1 = \widehat{q}_3'\Gamma'[\Gamma\text{var}(\widehat{q}_3)\Gamma']^{-1}\Gamma\widehat{q}_3 \\ &= \widehat{q}_3'[\text{var}(\widehat{q}_3)]^{-1}\widehat{q}_3 = m_3 \end{aligned}$$

This proves that m_1 and m_3 are numerically exactly identical. Similarly one can show that m_2 is numerically exactly identical to m_1 and m_3. In fact, problem 4.13 shows that these m_i are also exactly numerically identical to $m_4 = \widehat{q}_4'V_4^{-1}\widehat{q}_4$ where $\widehat{q}_4 = \widehat{\beta}_{\text{GLS}} - \widehat{\beta}_{\text{OLS}}$ and $V_4 = \text{var}(\widehat{q}_4)$. In the Monte Carlo study by Baltagi (1981a), the Hausman test is performed given that the exogeneity assumption is true. This test performed well with a low frequency of type I errors.

More recently, Arellano (1993) provided an alternative variable addition test to the Hausman test which is robust to autocorrelation and heteroskedasticity of arbitrary form. In particular, Arellano (1993) suggests constructing the following regression:

$$\begin{pmatrix} y_i^+ \\ \bar{y}_i \end{pmatrix} = \begin{bmatrix} X_i^+ & 0 \\ \bar{X}_i' & \bar{X}_i' \end{bmatrix}\begin{pmatrix} \beta \\ \gamma \end{pmatrix} + \begin{pmatrix} u_i^+ \\ \bar{u}_i \end{pmatrix} \tag{4.49}$$

where $y_i^+ = (y_{i1}^+, \ldots, y_{iT}^+)'$ and $X_i^+ = (X_{i1}^+, \ldots, X_{iT}^+)'$ is a $T \times K$ matrix and $u_i^+ = (u_{i1}^+, \ldots, u_{iT}^+)'$. Also

$$y_{it}^+ = \left[\frac{T-t}{T-t+1}\right]^{1/2}\left[y_{it} - \frac{1}{(T-t)}(y_{i,t+1} + \ldots + y_{iT})\right] \qquad t = 1, 2, \ldots, T-1$$

defines the forward orthogonal deviations operator, $\bar{y}_i = \Sigma_{t=1}^T y_{it}/T$, X_{it}^+, $\bar{X}_i$, u_{it}^+ and $\bar{u}_i$ are similarly defined. OLS on this model yields $\widehat{\beta} = \widetilde{\beta}_{\text{Within}}$ and $\widehat{\gamma} = \widehat{\beta}_{\text{Between}} - \widetilde{\beta}_{\text{Within}}$. Therefore, Hausman's test can be obtained from the artificial regression (4.49) by testing for $\gamma = 0$. If the disturbances are heteroskedastic and/or serially correlated, then neither $\widetilde{\beta}_{\text{Within}}$ nor $\widehat{\beta}_{\text{GLS}}$ are optimal under the null or alternative. Also, the standard formulae for the asymptotic variances of these estimators are no longer valid. Moreover, these estimators cannot be ranked in terms of efficiency so that the var(q) is not the difference of the two variances $\text{var}(\widetilde{\beta}_W) - \text{var}(\widehat{\beta}_{\text{GLS}})$. Arellano (1993) suggests using White's (1984) robust variance–covariance matrix from OLS on (4.49) and applying a standard Wald test for $\gamma = 0$ using these robust standard errors. This can easily be calculated using any standard regression package that computes White robust standard errors. This test is asymptotically distributed as χ^2_K under the null.

Chamberlain (1982) showed that the fixed effects specification imposes testable restrictions on the coefficients from regressions of all leads and lags of dependent variables on all leads and lags of independent variables. Chamberlain specified the relationship between the unobserved individual effects and X_{it} as follows:

$$\mu_i = X_{i1}'\lambda_1 + \ldots + X_{iT}'\lambda_T + \varepsilon_i \tag{4.50}$$

where each λ_t is of dimension $K \times 1$ for $t = 1, 2, \ldots, T$. Let $y_i' = (y_{i1}, \ldots, y_{iT})$ and $X_i' = (X_{i1}', \ldots, X_{iT}')$ and denote the "reduced form" regression of y_i' on X_i' by

$$y_i' = X_i'\pi + \eta_i \tag{4.51}$$

The restrictions between the reduced form and structural parameters are given by

$$\pi = (I_T \otimes \beta) + \lambda \iota_T' \tag{4.52}$$

with $\lambda' = (\lambda_1', \ldots, \lambda_T')$.[7] Chamberlain (1982) suggested estimation and testing be carried out using the minimum chi-square method where the minimand is a χ^2 goodness-of-fit statistic for the restrictions on the reduced form. However, Angrist and Newey (1991) showed that this minimand can be obtained as the sum of T terms. Each term of this sum is simply the degrees of freedom times the R^2 from a regression of the Within residuals for a particular period on all leads and lags of the independent variables. Angrist and Newey (1991) illustrate this test using two examples. The first example estimates and tests a number of models for the union–wage effect using five years of data from the National Longitudinal Survey of Youth (NLSY). They find that the assumption of fixed effects in an equation for union–wage effects is not at odds with the data. The second example considers a conventional human capital earnings function. They find that the fixed effects estimates of the return to schooling in the NLSY are roughly twice those of ordinary least squares. However, the over-identification test suggest that the fixed effects assumption may be inappropriate for this model.

Modifying the set of additional variables in (4.49) so that the set of K additional regressors are replaced by KT additional regressors Arellano (1993) obtains

$$\begin{pmatrix} y_i^+ \\ \bar{y}_i \end{pmatrix} = \begin{bmatrix} X_i^+ & 0 \\ \bar{X}_i' & X_i' \end{bmatrix} \begin{pmatrix} \beta \\ \lambda \end{pmatrix} + \begin{pmatrix} u_i^+ \\ \bar{u}_i \end{pmatrix} \tag{4.53}$$

where $X_i = (X_{i1}', \ldots, X_{iT}')'$ and λ is $KT \times 1$. Chamberlain's (1982) test of correlated effects based on the reduced form approach turns out to be equivalent to testing for $\lambda = 0$ in (4.53). Once again this can be made robust to an arbitrary form of serial correlation and heteroskedasticity by using a Wald test for $\lambda = 0$ using White's (1984) robust standard errors. This test is asymptotically distributed as χ^2_{TK}. Note that this clarifies the relationship between the Hausman specification test and Chamberlain omnibus goodness-of-fit test. In fact, both tests can be computed as Wald tests from the artificial regressions in (4.49) and (4.53). Hausman's test can be considered as a special case of the Chamberlain test for $\lambda_1 = \lambda_2 = \ldots = \lambda_T = \gamma/T$. Arellano (1993) extends this analysis to dynamic models and to the case where some of the explanatory variables are known to be uncorrelated or weakly correlated with the individual effects.

Recently, Ahn and Low (1996) showed that Hausman's test statistic can be obtained from the artificial regression of the GLS residuals $(y_{it}^* - X_{it}^{*\prime}\widehat{\beta}_{\text{GLS}})$ on $\widetilde{X}$ and $\bar{X}$, where $\widetilde{X}$ has typical element $\widetilde{X}_{it,k}$ and $\bar{X}$ is the matrix of regressors averaged over time. The test statistic is NT times the R^2 of this regression. Using (4.42), one can test $H_0 : \gamma = 0$ by running the Gauss–Newton regression (GNR) evaluated at the restricted estimators under the null. Knowing θ, the restricted estimates yield $\widehat{\beta} = \widehat{\beta}_{\text{GLS}}$ and $\widehat{\gamma} = 0$. Therefore, the GNR on (4.42) regresses the GLS residuals $(y_{it}^* - X_{it}^{*\prime}\widehat{\beta}_{\text{GLS}})$ on the derivatives of the regression function with respect to β and γ evaluated at $\widehat{\beta}_{\text{GLS}}$ and $\widehat{\gamma} = 0$. These regressors are X_{it}^* and $\widetilde{X}_{it}$, respectively. But X_{it}^* and $\widetilde{X}_{it}$ span the same space as $\widetilde{X}_{it}$ and $\bar{X}_{i.}$. This follows immediately from the definition of X_{it}^* and

$\widetilde{X}_{it}$ given above. Hence, this GNR yields the same regression sums of squares and therefore, the same Hausman test statistic as that proposed by Ahn and Low (1996), see problem 97.4.1 in *Econometric Theory* by Baltagi (1997).

Ahn and Low (1996) argue that Hausman's test can be generalized to test that each X_{it} is uncorrelated with μ_i and not simply that $\bar{X}_i$ is uncorrelated with μ_i. In this case, one computes NT times R^2 of the regression of GLS residuals $(y_{it}^* - X_{it}^{*\prime}\widehat{\beta}_{\text{GLS}})$ on $\widetilde{X}_{it}$ and $[X'_{i1}, \ldots, X'_{iT}]$. This LM statistic is identical to Arellano's (1993) Wald statistic described earlier if the same estimates of the variance components are used. Ahn and Low (1996) argue that this test is recommended for testing the joint hypothesis of exogeneity of the regressors and the stability of the regression parameters over time. If the regression parameters are nonstationary over time, both $\widehat{\beta}_{\text{GLS}}$ and $\widetilde{\beta}_{\text{Within}}$ are inconsistent even though the regressors are exogenous. Monte Carlo experiments were performed that showed that both the Hausman test and the Ahn and Low (1996) test have good power in detecting endogeneity of the regressors. However, the latter test dominates if the coefficients of the regressors are nonstationary. For Ahn and Low (1996), rejection of the null does not necessarily favor the Within estimator since the latter estimator may be inconsistent. In this case, the authors recommend performing Chamberlain's (1982) test or the equivalent test proposed by Angrist and Newey (1991).

4.3.1 Example 1: Grunfeld Investment Equation

For the Grunfeld data, the Within estimates are given by $(\widetilde{\beta}_1, \widetilde{\beta}_2) = (0.1101238, 0.310065)$ with a variance–covariance matrix:

$$\text{var}(\widetilde{\beta}_{\text{Within}}) = \begin{bmatrix} 0.14058 & -0.077468 \\ & 0.3011788 \end{bmatrix} \times 10^{-3}$$

The Between estimates are given by $(0.1346461, 0.03203147)$ with variance–covariance matrix:

$$\text{var}(\widehat{\beta}_{\text{Between}}) = \begin{bmatrix} 0.82630142 & -3.7002477 \\ & 36.4572431 \end{bmatrix} \times 10^{-3}$$

The resulting Hausman test statistic based on (4.46) and (4.48) and labeled as m_3 yields an observed χ_2^2 statistic of 2.131. This is not significant at the 5% level and we do not reject the null hypothesis of no correlation between the individual effects and the X_{it}. As a cautionary note, one should not use the Hausman command in Stata to perform the Hausman test based on a contrast between the fixed effects (FE) and Between (BE) estimators. This will automatically subtract the variance–covariance matrices of the two estimators, rather than add them as required in (4.48). However, the Hausman test statistic can be properly computed in Stata based upon the contrast between the RE (feasible GLS) estimator and fixed effects (FE). This is the Hausman statistic labeled as m_1 in (4.41) based on the contrast $\widehat{q}_1$ and $\text{var}(\widehat{q}_1)$ given in (4.40). Table 4.3 gives the Stata output using the Hausman command which computes (4.41). This yields an m_1 statistic of 2.33 which is distributed as χ_2^2. This does not reject the null hypothesis as obtained using m_3. Note that the feasible GLS estimator in Stata is SWAR and is computed whenever the RE option is invoked. One can also compute m_2 based on $\widehat{q}_2$ which is the contrast between the SWAR feasible GLS estimator and the Between estimator. Table 4.4 gives the Stata output that replicates this Hausman test yielding an m_2 statistic of 2.13. As expected, this statistic is not significant and does not reject the null hypothesis as obtained using m_1 and m_3. Hence, one does not reject the null hypothesis that the RE estimator is consistent. Finally, the augmented regression, given in (4.42) based on the SWAR feasible GLS estimates of θ,

Table 4.3 Grunfeld's Data. Hausman Test FE vs RE

```
. hausman fe re

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       fe           re         Difference          S.E.
-------------+----------------------------------------------------------------
           F |    .1101238     .1097811       .0003427        .0055213
           C |    .3100653     .308113        .0019524        .0024516
------------------------------------------------------------------------------
                    b = consistent under Ho and Ha; obtained from xtreg
     B = inconsistent under Ha, efficient under Ho; obtained from xtreg

     Test:  Ho:  difference in coefficients not systematic

                    chi2(2) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                            =        2.33
                  Prob>chi2 =      0.3119
```

yields the following estimates: $\widehat{\beta} = (0.135, 0.032)$ and $\widehat{\gamma} = (-0.025, 0.278)$ with an observed F-value for $H_0 : \gamma = 0$ equal to 1.066. This is distributed under H_0 as an $F(2, 195)$. This is again not significant at the 5% level and leads to nonrejection of H_0.

4.3.2 Example 2: Gasoline Demand

For the Baltagi and Griffin (1983) gasoline data, the Within estimates are given by $(\widetilde{\beta}_1, \widetilde{\beta}_2, \widetilde{\beta}_3) = (0.66128, -0.32130, -0.64015)$ with variance–covariance matrix:

$$\operatorname{var}(\widetilde{\beta}_{\text{Within}}) = \begin{bmatrix} 0.539 & 0.029 & -0.205 \\ & 0.194 & 0.009 \\ & & 0.088 \end{bmatrix} \times 10^{-2}$$

Table 4.4 Grunfeld's Data. Hausman Test BE vs RE

```
. hausman be re

                 ---- Coefficients ----
           |      (b)         (B)            (b-B)     sqrt(diag(V_b-V_B))
           |       be          re         Difference          S.E.
-----------+----------------------------------------------------------------
         F |    .1346461   .1097811       .0248649         .026762
         C |    .0320315    .308113      -.2760815         .1901633
----------------------------------------------------------------------------
                    b = consistent under Ho and Ha; obtained from xtreg
     B = inconsistent under Ha, efficient under Ho; obtained from xtreg

         Test:  Ho:  difference in coefficients not systematic

                        chi2(2) = (b-B)'[(V_b-V_B) ^ (-1)](b-B)
                                =        2.13
                      Prob>chi2 =      0.3445
```

The Between estimates are given by $(0.96737, -0.96329, -0.79513)$ with variance–covariance matrix:

$$\operatorname{var}(\widehat{\beta}_{\text{Between}}) = \begin{bmatrix} 2.422 & -1.694 & -1.056 \\ & 1.766 & 0.883 \\ & & 0.680 \end{bmatrix} \times 10^{-2}$$

The resulting Hausman χ_3^2 test statistic is $m_3 = 26.507$ which is significant. Hence we reject the null hypothesis of no correlation between the individual effects and the X_{it}. One can similarly compute $m_2 = 27.45$, based on the contrast between the SWAR feasible GLS estimator and the Between estimator, and $m_1 = 302.8$ based on the contrast between the SWAR feasible GLS estimator and the fixed effects estimator. These were obtained using Stata. Although m_1 gives a drastically different value of the Hausman statistic than m_2 or m_3, all three test statistics lead to the same decision. The null is rejected and the RE estimator is not consistent. The augmented regression, given in (4.42) based on the iterative MLE estimate of θ, yields the following estimates: $\widehat{\beta}_{\text{Between}} = (0.967, -0.963, -0.795)$ and $\widehat{\gamma} = \widetilde{\beta}_{\text{Within}} - \widehat{\beta}_{\text{Between}} = (-0.306, 0.642, 0.155)$ with an observed F-value for $H_0 : \gamma = 0$ equal to 4.821. This is distributed under H_0 as an $F(3, 335)$, and leads to the rejection of H_0.

4.3.3 Example 3: Strike Activity

Owusu-Gyapong (1986) considered panel data on strike activity in 60 Canadian manufacturing industries for the period 1967–79. A one-way error component model is used and OLS, Within and GLS estimates are obtained. With $K' = 12$ regressors, $N = 60$ and $T = 13$, an F-test for the significance of industry-specific effects described in (2.12) yields an F-value of 5.56. This is distributed as $F_{59,709}$ under the null hypothesis of zero industry-specific effects. The null is soundly rejected and the Within estimator is preferred to the OLS estimator. Next, $H_0 : \sigma_\mu^2 = 0$ is tested using the Breusch and Pagan (1980) two-sided LM test given as LM_1 in (4.23). This yields a χ^2-value of 21.4, which is distributed as χ_1^2 under the null hypothesis of zero random effects. The null is soundly rejected and the GLS estimator is preferred to the OLS estimator. Finally, for a choice between the Within and GLS estimators, the author performs a Hausman (1978)-type test to test $H_0 : E(\mu_i / X_{it}) = 0$. This is based on the difference between the Within and GLS estimators as described in (4.41) and yields a χ^2 value equal to 3.84. This is distributed as χ_{11}^2 under the null and is not significant. The Hausman test was also run as an augmented regression-type test described in (4.42). This also did not reject the null of no correlation between the μ_i and the regressors. Based on these results, Owusu-Gyapong (1986) chose GLS as the preferred estimator.

4.3.4 Example 4: Production Behavior of Sawmills

Cardellichio (1990) estimated the production behavior of 1147 sawmills in the state of Washington observed biennially over the period 1972–84. A one-way error component model is used and OLS, Within and GLS estimates are presented. With $K' = 21$ regressors, $N = 1147$ and $T = 7$, an F-test for the stability of the slope parameters over time was performed which was not significant at the 5% level. In addition, an F-test for the significance of sawmill effects described in (2.12) was performed which rejected the null at the 1% significance level. Finally, a Hausman test was performed and it rejected the null at the 1% significance level. Cardellichio (1990) concluded that the regression slopes are stable over time, sawmill dummies should be included and the Within estimator is preferable to OLS and GLS since the orthogonality assumption between the regressors and the sawmill effects is rejected.

4.3.5 Example 5: The Marriage Wage Premium

Cornwell and Rupert (1997) estimated the wage premium attributed to marriage using the 1971, 1976, 1978 and 1980 waves of the NLSY. They find that the Within estimates of the marriage premium are smaller than those obtained from feasible GLS. A Hausman test based on the difference between these two estimators rejects the null hypothesis. This indicates the possibility of important omitted individual-specific characteristics which are correlated with both marriage and the wage rate. They conclude that the marriage premium is purely an intercept shift and no more than 5% to 7% . They also cast doubt on the interpretation that marriage enhances productivity through specialization.

4.3.6 Example 6: Currency Union and Trade

Glick and Rose (2002) consider the question of whether leaving a currency union reduces international trade. Using annual data on bilateral trade among 217 countries from 1948 through 1997, they estimate an augmented gravity model controlling for several factors. These include real GDP, distance, land mass, common language, sharing a land border, whether they belong to the same regional trade agreement, land-locked, island nations, common colonizer, current colony, ever a colony and whether they remained part of the same nation. The focus variable is a binary variable which is unity if country i and country j use the same currency at time t. They apply OLS, FE, RE, and their preferred estimator is FE based on the Hausman test. They find that a pair of countries which joined/left a currency union experienced a near-doubling/halving of bilateral trade. The data set along with the Stata logs are available on Rose's web site, see problem 4.19.

4.3.7 Hausman's Test for the Two-way Model

For the two-way error component model, Hausman's (1978) test can still be based on the difference between the fixed effects estimator (with both time and individual dummies) and the two-way random effects GLS estimator. Also, the augmented regression, given in (4.42), can still be used as long as the Within and GLS transformations used are those for the two-way error component model. But, what about the equivalent tests described for the one-way model? Do they extend to the two-way model? Not quite. Kang (1985) showed that a similar equivalence for the Hausman test does not hold for the two-way error component model, since there would be two Between estimators, one between time periods $\widehat{\beta}_T$ and one between cross-sections $\widehat{\beta}_C$. Also, $\widehat{\beta}_{\text{GLS}}$ is a weighted combination of $\widehat{\beta}_T, \widehat{\beta}_C$ and the Within estimator $\widetilde{\beta}_W$. Kang (1985) shows that the Hausman test based on $(\widetilde{\beta}_W - \widehat{\beta}_{\text{GLS}})$ is not equivalent to that based on $(\widehat{\beta}_C - \widehat{\beta}_{\text{GLS}})$ nor that based on $(\widehat{\beta}_T - \widehat{\beta}_{\text{GLS}})$. But there are other types of equivalencies (see Kang's table 2). More importantly, Kang classifies five testable hypotheses:

(1) Assume that μ_i are fixed and test $E(\lambda_t/X_{it}) = 0$ based upon $\widetilde{\beta}_W - \widehat{\beta}_T$.
(2) Assume the μ_i are random and test $E(\lambda_t/X_{it}) = 0$ based upon $\widehat{\beta}_T - \widehat{\beta}_{\text{GLS}}$.
(3) Assume the λ_t are fixed and test $E(\mu_i/X_{it}) = 0$ based upon $\widetilde{\beta}_W - \widehat{\beta}_C$.

(4) Assume the λ_t, are random and test $E(\mu_i/X_{it}) = 0$ based upon $\widehat{\beta}_C - \widehat{\beta}_{\text{GLS}}$.
(5) Compare two estimators, one which assumes both the μ_i and λ_T are fixed, and another that assumes both are random such that $E(\lambda_t/X_{it}) = E(\mu_i/X_{it}) = 0$. This test is based upon $\widehat{\beta}_{\text{GLS}} - \widetilde{\beta}_W$.

4.4 FURTHER READING

Li and Stengos (1992) proposed a Hausman specification test based on root-N consistent semiparametric estimators. Also, Baltagi and Chang (1996) proposed a simple ANOVA F-statistic based on recursive residuals to test for random individual effects and studied its size and power using Monte Carlo experiments. Chesher (1984) derived a score test for neglected heterogeneity, which is viewed as causing parameter variation. Also, Hamerle (1990) and Orme (1993) suggest a score test for neglected heterogeneity for qualitative limited dependent variable panel data models.

The normality assumption on the error components disturbances may be untenable. Horowitz and Markatou (1996) show how to carry out nonparametric estimation of the densities of the error components. Using data from the Current Population Survey, they estimate an earnings model and show that the probability that individuals with low earnings will become high earners in the future are much lower than that obtained under the assumption of normality. One drawback of this nonparametric estimator is its slow convergence at a rate of $1/(\log N)$ where N is the number of individuals. Monte Carlo results suggest that this estimator should be used for N larger than 1000. Blanchard and Mátyás (1996) perform Monte Carlo simulations to study the robustness of several tests for individual effects with respect to nonnormality of the disturbances. The alternative distributions considered are the exponential, lognormal, $t(5)$ and Cauchy distributions. The main findings are that the F-test is robust against nonnormality while the one-sided and two-sided LM and LR tests are sensitive to nonnormality.

Davidson and MacKinnon (1993) showed that the double-length artificial regression (DLR) can be very useful in choosing between, and testing the specification of, models that are linear or loglinear in the dependent variable. Baltagi (1997) extends this DLR to panel data regressions, where the choice between linear and loglinear models is complicated by the presence of error components. This DLR can easily be extended to test jointly for functional form and random individual effects (see problem 97.1.3 in *Econometric Theory* by Baltagi (1997) and its solution by Li (1998)).

NOTES

1. An elegant presentation of this F-statistic is given in Fisher (1970).
2. Baltagi (1996) shows that testing for random individual and time effects can be obtained from a variable addition test using two extra variables, one that involves the average of least squares residuals over time and another that involves the average of these residuals across individuals. In fact, this test applies to *nonlinear* regression models with error components disturbances. This variable addition test is an application of the Gauss–Newton regression (GNR) described in detail in Davidson and MacKinnon (1993). For other applications of the GNR in panel data, see Baltagi (1999).
3. Häggström (2002) studies the properties of Honda's tests for random individual effects in nonlinear regression models. Two corrections for Honda's test statistic are suggested when random time effects are present.
4. Critical values for the mixed χ^2_m are 7.289, 4.321 and 2.952 for $\alpha = 0.01, 0.05$ and 0.1, respectively.

5. Hausman (1978) tests $\gamma = 0$ from (4.42) using an F-statistic. The restricted regression yields OLS of y^* on X^*. This is the Fuller and Battese (1973) regression yielding GLS as described below (2.20). The unrestricted regression adds the matrix of Within regressors $\widetilde{X}$ as in (4.42).
6. For an important discussion of what null hypothesis is actually being tested using the Hausman test, see Holly (1982).
7. For more on the Chamberlain approach, read Crépon and Mairesse (1996).

PROBLEMS

4.1 Verify the relationship between M and M^*, i.e. $MM^* = M^*$, given below (4.7). *Hint*: Use the fact that $Z = Z^* I^*$ with $I^* = (\iota_N \otimes I_{K'})$.

4.2 Verify that $\dot{M}$ and $\dot{M}^*$ defined below (4.10) are both symmetric, idempotent and satisfy $\dot{M}\dot{M}^* = \dot{M}^*$.

4.3 For Grunfeld's data given as Grunfeld.fil on the Wiley web site, verify the testing for poolability results given in example 1, section 4.1.3.

4.4 For the gasoline data given as Gasoline.dat on the Wiley web site, verify the testing for poolability results given in example 2, section 4.1.3.

4.5 Under normality of the disturbances, show that for the likelihood function given in (4.15):
(a) The information matrix is block-diagonal between $\theta' = (\sigma_\mu^2, \sigma_\lambda^2, \sigma_\nu^2)$ and δ.
(b) For $H_0^c : \sigma_\mu^2 = \sigma_\lambda^2 = 0$, verify (4.18), (4.20) and (4.22).

4.6 Using the results of Baltagi et al. (1992b), verify that the King–Wu (1997) test for $H_0^c : \sigma_\mu^2 = \sigma_\lambda^2 = 0$ is given by (4.30).

4.7 For $H_0^c : \sigma_\mu^2 = \sigma_\lambda^2 = 0$: (a) Verify that the standardized Lagrange multiplier (SLM) test statistics for Honda's (1991) $(A + B)/\sqrt{2}$ statistic is as described by (4.26) and (4.31). (b) Also, verify that the King and Wu (1997) standardized test statistic is as described by (4.26) and (4.32).

4.8 Using the Monte Carlo set-up for the two-way error component model described in Baltagi (1981a):
(a) Compare the performance of the Chow F-test and the Roy–Zellner test for various values of the variance components.
(b) Compare the performance of the BP, KW, SLM, LR, GHM and F-test statistics as done in Baltagi et al. (1992b).
(c) Perform Hausman's specification test and discuss its size for the various experiments conducted.

4.9 For the Grunfeld data, replicate Table 4.1.

4.10 For the gasoline data, derive a similar table to test the hypotheses given in Table 4.1.

4.11 For the public capital data, derive a similar table to test the hypotheses given in Table 4.1.

4.12 Using partitioned inverse on (4.43), verify (4.44) and deduce (4.45) and (4.46).

4.13 (a) Verify that m_2 is numerically exactly identical to m_1 and m_3, where $m_i = \widehat{q}_i' V_i^{-1} \widehat{q}_i$ defined below (4.48).
(b) Verify that these are also exactly numerically identical to $m_4 = \widehat{q}_4' V_4^{-1} \widehat{q}_4$ where $\widehat{q}_4 = \widehat{\beta}_{\text{GLS}} - \widehat{\beta}_{\text{OLS}}$ and $V_4 = \text{var}(\widehat{q}_4)$. *Hint*: See problem 89.3.3 in *Econometric Theory* by Baltagi (1989) and its solution by Koning (1990).

4.14 *Testing for correlated effects in panels.* This is based on problem 95.2.5 in *Econometric Theory* by Baltagi (1995). This problem asks the reader to show that Hausman's test, studied in section 4.3, can be derived from Arellano's (1993) extended regression by using an alternative transformation of the data. In particular, consider the transformation given by $H = (C', \iota_T/T)'$ where C is the first $(T-1)$ rows of the Within transformation $E_T = I_T - \bar{J}_T$, I_T is an identity matrix of dimension T and $\bar{J}_T = \iota_T \iota_T'/T$ with ι_T a vector of 1's of dimension T.

(a) Show that the matrix C satisfies the following properties: $C\iota_T = 0$, $C'(CC')^{-1}C = I_T - \bar{J}_T$; see Arellano and Bover (1995).

(b) For the transformed model $y_i^+ = Hy_i = (y_i^{*\prime}, \bar{y}_i)'$, where $y_i^* = Cy_i$ and $\bar{y}_i = \Sigma_{t=1}^T y_{it}/T$. The typical element of y_i^* is given by $y_{it}^* = [y_{it} - \bar{y}_i]$ for $t = 1, 2, \ldots, T-1$. Consider the extended regression similar to (4.49)

$$\begin{bmatrix} y_i^* \\ \bar{y}_i \end{bmatrix} = \begin{bmatrix} X_i^{*\prime} & 0 \\ \bar{X}_i' & \bar{X}_i' \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} + \begin{bmatrix} u_i^* \\ \bar{u}_i \end{bmatrix}$$

and show that GLS on this extended regression yields $\widehat{\beta} = \widehat{\beta}_{\text{Within}}$ and $\widehat{\gamma} = \widehat{\beta}_{\text{Between}} - \widehat{\beta}_{\text{Within}}$, where $\widehat{\beta}_{\text{Within}}$ and $\widehat{\beta}_{\text{Between}}$ are the familiar panel data estimators. Conclude that Hausman's test for $H_0: E(\mu_i/X_i) = 0$ can be based on a test for $\gamma = 0$, as shown by Arellano (1993). See solution 95.2.5 in *Econometric Theory* by Xiong (1996).

4.15 For the Grunfeld data, replicate the Hausman test results given in example 1 of section 4.3.

4.16 For the gasoline demand data, replicate the Hausman test results given in example 2 of section 4.3.

4.17 Perform Hausman's test for the public capital data.

4.18 *The relative efficiency of the Between estimator with respect to the Within estimator.* This is based on problem 99.4.3 in *Econometric Theory* by Baltagi (1999). Consider the simple panel data regression model

$$y_{it} = \alpha + \beta x_{it} + u_{it} \quad i = 1, 2, \ldots, N;\ t = 1, 2, \ldots, T \tag{1}$$

where α and β are scalars. Subtract the mean equation to get rid of the constant

$$y_{it} - \bar{y}_{..} = \beta(x_{it} - \bar{x}_{..}) + u_{it} - \bar{u}_{..} \tag{2}$$

where $\bar{x}_{..} = \Sigma_{i=1}^N \Sigma_{t=1}^T x_{it}/NT$ and $\bar{y}_{..}$ and $\bar{u}_{..}$ are similarly defined. Add and subtract $\bar{x}_{i.}$ from the regressor in parentheses and rearrange

$$y_{it} - \bar{y}_{..} = \beta(x_{it} - \bar{x}_{i.}) + \beta(\bar{x}_{i.} - \bar{x}_{..}) + u_{it} - \bar{u}_{..} \tag{3}$$

where $\bar{x}_{i.} = \Sigma_{t=1}^T x_{it}/T$. Now run the unrestricted least squares regression

$$y_{it} - \bar{y}_{..} = \beta_w(x_{it} - \bar{x}_{i.}) + \beta_b(\bar{x}_{i.} - \bar{x}_{..}) + u_{it} - \bar{u}_{..} \tag{4}$$

where β_w is not necessarily equal to β_b.

(a) Show that the least squares estimator of β_w from (4) is the Within estimator and that of β_b is the Between estimator.

(b) Show that if $u_{it} = \mu_i + \nu_{it}$ where $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$ and $\nu_{it} \sim \text{IID}(0, \sigma_\nu^2)$ independent of each other and among themselves, then ordinary least squares (OLS) is equivalent to generalized least squares (GLS) on (4).

(c) Show that for model (1), the relative efficiency of the Between estimator with respect to the Within estimator is equal to $(B_{XX}/W_{XX})[(1-\rho)/(T\rho + (1-\rho))]$, where $W_{XX} = \Sigma_{i=1}^N \Sigma_{t=1}^T (x_{it} - \bar{x}_{i.})^2$ denotes the Within variation and $B_{XX} = T\Sigma_{i=1}^N (\bar{x}_{i.} -$

$\bar{x}_{..})^2$ denotes the Between variation. Also, $\rho = \sigma_\mu^2/(\sigma_\mu^2 + \sigma_\nu^2)$ denotes the equicorrelation coefficient.

(d) Show that the square of the t-statistic used to test $H_0 : \beta_w = \beta_b$ in (4) yields exactly Hausman's (1978) specification test. See solution 99.4.3 in *Econometric Theory* by Gurmu (2000).

4.19 Using the Glick and Rose (2002) data set, downloadable from Rose's web site at http://haas.berkeley.edu)

(a) Replicate their results for the FE, RE, Between and MLE estimators reported in table 4 of their paper.

(b) Perform the Hausman test based on FE vs RE as well as Between vs RE using Stata.

5
Heteroskedasticity and Serial Correlation in the Error Component Model

5.1 HETEROSKEDASTICITY

The standard error component model given by equations (2.1) and (2.2) assumes that the regression disturbances are homoskedastic with the same variance across time and individuals. This may be a restrictive assumption for panels, where the cross-sectional units may be of varying size and as a result may exhibit different variation. For example, when dealing with gasoline demand across OECD countries, steam electric generation across various size utilities or estimating cost functions for various US airline firms, one should expect to find heteroskedasticity in the disturbance term. Assuming homoskedastic disturbances when heteroskedasticity is present will still result in consistent estimates of the regression coefficients, but these estimates will not be efficient. Also, the standard errors of these estimates will be biased and one should compute robust standard errors correcting for the possible presence of heteroskedasticity. In this section, we relax the assumption of homoskedasticity of the disturbances and introduce heteroskedasticity through the μ_i as first suggested by Mazodier and Trognon (1978). Next, we suggest an alternative heteroskedastic error component specification, where only the ν_{it} are heteroskedastic. We derive the true GLS transformation for these two models. We also consider two adaptive heteroskedastic estimators based on these models where the heteroskedasticity is of unknown form. These adaptive heteroskedastic estimators were suggested by Li and Stengos (1994) and Roy (2002).

Mazodier and Trognon (1978) generalized the homoskedastic error component model to the case where the μ_i are heteroskedastic, i.e. $\mu_i \sim (0, w_i^2)$ for $i = 1, \ldots, N$, but $\nu_{it} \sim \text{IID}(0, \sigma_\nu^2)$. In vector form, $\mu \sim (0, \Sigma_\mu)$ where $\Sigma_\mu = \text{diag}[w_i^2]$ is a diagonal matrix of dimension $N \times N$, and $\nu \sim (0, \sigma_\nu^2 I_{NT})$. Therefore, using (2.4), one gets

$$\Omega = E(uu') = Z_\mu \Sigma_\mu Z'_\mu + \sigma_\nu^2 I_{NT} \tag{5.1}$$

This can be written as

$$\Omega = \text{diag}[w_i^2] \otimes J_T + \text{diag}[\sigma_\nu^2] \otimes I_T \tag{5.2}$$

where $\text{diag}[\sigma_\nu^2]$ is also of dimension $N \times N$. Using the Wansbeek and Kapteyn (1982b, 1983) trick, Baltagi and Griffin (1988a) derived the corresponding Fuller and Battese (1974) transformation as follows:

$$\Omega = \text{diag}[Tw_i^2 + \sigma_\nu^2] \otimes \bar{J}_T + \text{diag}[\sigma_\nu^2] \otimes E_T$$

Therefore

$$\Omega^r = \text{diag}[(\tau_i^2)^r] \otimes \bar{J}_T + \text{diag}[(\sigma_\nu^2)^r] \otimes E_T \tag{5.3}$$

with $\tau_i^2 = Tw_i^2 + \sigma_\nu^2$, and r is any arbitrary scalar. The Fuller–Battese transformation for the heteroskedastic case premultiplies the model by

$$\sigma_\nu \Omega^{-1/2} = \text{diag}[\sigma_\nu/\tau_i] \otimes \bar{J}_T + (I_N \otimes E_T) \tag{5.4}$$

Hence, $y^* = \sigma_\nu \Omega^{-1/2} y$ has a typical element $y^*_{it} = y_{it} - \theta_i \overline{y}_{i.}$ where $\theta_i = 1 - (\sigma_\nu/\tau_i)$ for $i = 1, \ldots, N$.

Baltagi and Griffin (1988a) provided feasible GLS estimators including Rao's (1970, 1972) MINQUE estimators for this model. Phillips (2003) argues that this model suffers from the incidental parameters problem and the variance estimates of μ_i (the ω_i^2) cannot be estimated consistently, so there is no guarantee that feasible GLS and true GLS will have the same asymptotic distributions. Instead, he suggests a stratified error component model where the variances change across strata and provides an EM algorithm to estimate it. It is important to note that Mazodier and Trognon (1978) had already suggested stratification in a two-way heteroskedastic error component model. Also, that one can specify parametric variance functions which avoid the incidental parameter problem and then apply the GLS transformation described above. As in the cross-section heteroskedastic case, one has to know the variables that determine heteroskedasticity, but not necessarily the form. Adaptive estimation of heteroskedasticity of unknown form has been suggested for this model by Roy (2002). This follows similar literature on adaptive estimation for cross-section data.

Alternatively, one could keep the μ_i homoskedastic with $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$ and impose the heteroskedasticity on the ν_{it}, i.e. $\nu_{it} \sim (0, w_i^2)$ (see problem 88.2.2 by Baltagi (1988) and its solution by Wansbeek (1989) in *Econometric Theory*). In this case, using (2.4) one obtains

$$\Omega = E(uu') = \text{diag}[\sigma_\mu^2] \otimes J_T + \text{diag}[w_i^2] \otimes I_T \tag{5.5}$$

Replacing J_T by $T\bar{J}_T$ and I_T by $E_T + \bar{J}_T$, we get

$$\Omega = \text{diag}[T\sigma_\mu^2 + w_i^2] \otimes \bar{J}_T + \text{diag}[w_i^2] \otimes E_T$$

and

$$\Omega^r = \text{diag}[(\tau_i^2)^r] \otimes \bar{J}_T + \text{diag}[(w_i^2)^r] \otimes E_T \tag{5.6}$$

where $\tau_i^2 = T\sigma_\mu^2 + w_i^2$, and r is an arbitrary scalar. Therefore

$$\Omega^{-1/2} = \text{diag}[1/\tau_i] \otimes \bar{J}_T + \text{diag}[1/w_i] \otimes E_T \tag{5.7}$$

and $y^* = \Omega^{-1/2} y$ has a typical element

$$y^*_{it} = (\bar{y}_{i.}/\tau_i) + (y_{it} - \bar{y}_{i.})/w_i$$

Upon rearranging terms, we get

$$y^*_{it} = \frac{1}{w_i}(y_{it} - \theta_i \bar{y}_{i.}) \quad \text{where} \quad \theta_i = 1 - (w_i/\tau_i)$$

One can argue that heteroskedasticity will contaminate both μ_i and ν_{it} and it is hard to claim that it is in one component and not the other. Randolph (1988) gives the GLS transformation for a more general heteroskedastic model where *both* the μ_i and the ν_{it} are assumed heteroskedastic in the context of an unbalanced panel. In this case, the var$(\mu_i) = \sigma_i^2$ and $E(\nu\nu') = \text{diag}[\sigma_{it}^2]$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T_i$. More recently, Li and Stengos (1994) considered the regression model given by (2.1) and (2.2) with $\mu_i \sim \text{IID}\left(0, \sigma_\nu^2\right)$ and $E\left[\nu_{it}|X'_{it}\right] = 0$ with

$\text{var}\left[\nu_{it}|X'_{it}\right] = \gamma\left(X'_{it}\right) \equiv \gamma_{it}$. So that the heteroskedasticity is on the remainder error term and it is of an unknown form.

Therefore $\sigma^2_{it} = E\left[u^2_{it}|X_{it}\right] = \sigma^2_\mu + \gamma_{it}$ and the proposed estimator of σ^2_μ is given by:

$$\widehat{\sigma}^2_\mu = \frac{\sum\limits_{i=1}^{N}\sum\limits_{t\neq s}^{T}\widehat{u}_{it}\widehat{u}_{is}}{NT(T-1)}$$

where $\widehat{u}_{it}$ denotes the OLS residual. Also

$$\widehat{\gamma}_{it} = \frac{\sum\limits_{j=1}^{N}\sum\limits_{s=1}^{T}\widehat{u}^2_{js}K_{it,js}}{\sum\limits_{j=1}^{N}\sum\limits_{s=1}^{T}K_{it,js}} - \widehat{\sigma}^2_\mu$$

where the kernel function is given by $K_{it,js} = K\left(\frac{X'_{it}-X'_{js}}{h}\right)$ and h is the smoothing parameter. These estimators of the variance components are used to construct a feasible adaptive GLS estimator of β which they denote by GLSAD. The computation of their feasible GLS estimator is simplified into an OLS regression using a recursive transformation that reduces the general heteroskedastic error components structure into classical errors, see Li and Stengos (1994) for details.

Roy (2002) considered the alternative heteroskedastic model $E\left[\mu_i|\overline{X}'_{i.}\right] = 0$ with

$$\text{var}\left[\mu_i|\overline{X}'_{i.}\right] = \omega\left(\overline{X}'_{i.}\right) \equiv \omega_i$$

with $\overline{X}'_{i.} = \sum\limits_{t=1}^{T}X'_{it}/T$ and $\upsilon_{it} \sim \text{IID}\left(0, \sigma^2_\nu\right)$. So that the heteroskedasticity is on the individual specific error component and it is of an unknown form. Roy (2002) used the usual estimator of σ^2_ν which is the MSE of the Within regression, see (2.24), and this can be written as

$$\widehat{\sigma}^2_\nu = \frac{\sum\limits_{i=1}^{N}\sum\limits_{t=1}^{T}\left[\left(y_{it}-\overline{y}_{i.}\right)-\left(X_{it}-\overline{X}_{i.}\right)'\widetilde{\beta}\right]^2}{N(T-1)-k}$$

where $\widetilde{\beta}$ is the fixed effects or Within estimator of β given in (2.7). Also

$$\widehat{\omega}_i = \frac{\sum\limits_{j=1}^{N}\sum\limits_{t=1}^{T}\widehat{u}^2_{jt}K_{i.,j.}}{\sum\limits_{j=1}^{N}\sum\limits_{t=1}^{T}K_{i.,j.}} - \widehat{\sigma}^2_\nu$$

where the kernel function is given by:

$$K_{i.,j.} = K\left(\frac{\overline{X}'_{i.}-\overline{X}'_{j.}}{h}\right)$$

Using these estimators of the variance components, Roy (2002) computed a feasible GLS estimator using the transformation derived by Baltagi and Griffin (1988a) and given in (5.4). This was denoted by EGLS.

Both Li and Stengos (1994) and Roy (2002) performed Monte Carlo experiments based on the simple regression model given in (2.8). They compared the following estimators: (1) OLS; (2) fixed effects or Within estimator (Within); (3) the conventional GLS estimator for the one-way error component model that assumes the error term u_{it} is homoskedastic (GLSH); and (4) their own adaptive heteroskedastic estimator denoted by (EGLS) for Roy (2002) and (GLSAD) for Li and Stengos (1994). Li and Stengos (1994) found that their adaptive estimator outperforms all the other estimators in terms of relative MSE with respect to true GLS for $N = 50$, 100 and $T = 3$ and for moderate to severe degrees of heteroskedasticity. Roy (2002) also found that her adaptive estimator performs well, although it was outperformed by fixed effects in some cases where there were moderate and severe degrees of heteroskedasticity. Recently, Baltagi, Bresson and Pirotte (2005a) checked the sensitivity of the two proposed adaptive heteroskedastic estimators under misspecification of the form of heteroskedasticity. In particular, they ran Monte Carlo experiments using the heteroskedasticity set-up of Li and Stengos (1994) to see how the misspecified Roy (2002) estimator performs. Next, they used the heteroskedasticity set-up of Roy (2002) to see how the misspecified Li and Stengos (1994) estimator performs. They also checked the sensitivity of these results to the choice of the smoothing parameters, the sample size and the degree of heteroskedasticity. Baltagi et al. (2005a) found that in terms of loss in efficiency, misspecifying the adaptive form of heteroskedasticity can be costly when the Li and Stengos (1994) model is correct and the researcher performs the Roy (2002) estimator. This loss in efficiency is smaller when the true model is that of Roy (2002) and one performs the Li and Stengos (1994) estimator. The latter statement is true as long as the choice of bandwidth is not too small. Both papers also reported the 5% size performance of the estimated t-ratios of the slope coefficient. Li and Stengos (1994) found that only GLSAD had the correct size while OLS, GLSH and Within over-rejected the null hypothesis. Roy (2002) found that GLSH and EGLS had the correct size no matter what choice of h was used. Baltagi et al. (2005a) found that OLS and GLSAD (small h) tend to over-reject the null when true no matter what form of adaptive heteroskedasticity. In contrast, GLSH, EGLS and Within have size not significantly different from 5% when the true model is that of Roy (2002) and slightly over-reject (7–8%) when the true model is that of Li and Stengos (1994).

In Chapter 2, we pointed out that Arellano (1987) gave a neat way of obtaining standard errors for the fixed effects estimator that are robust to heteroskedasticity and serial correlation of arbitrary form, see equation (2.16). In Chapter 4, we discussed how Arellano (1993) suggested a Hausman (1978) test as well as a Chamberlain (1982) omnibus goodness-of-fit test that are robust to heteroskedasticity and serial correlation of arbitrary form, see equations (4.49) and (4.53). Li and Stengos (1994) suggested a modified Breusch and Pagan test for significance of the random individual effects, i.e., $H_0 : \sigma_\mu^2 = 0$, which is robust to heteroskedasticity of unknown form in the remainder error term.

5.1.1 Testing for Homoskedasticity in an Error Component Model

Verbon (1980) derived a Lagrange multiplier test for the null hypothesis of homoskedasticity against the heteroskedastic alternative $\mu_i \sim \left(0, \sigma_{\mu_i}^2\right)$ and $\nu_{it} \sim \left(0, \sigma_{\nu_i}^2\right)$. In Verbon's model, however, $\sigma_{\mu_i}^2$ and $\sigma_{\nu_i}^2$ are, up to a multiplicative constant, *identical* parametric functions of time-invariant exogenous variables Z_i, i.e., $\sigma_{\mu_i}^2 = \sigma_\mu^2 f(Z_i\theta_2)$ and $\sigma_{\nu_i}^2 = \sigma_\nu^2 f(Z_i\theta_1)$. Lejeune (1996), on the other hand, dealt with maximum likelihood estimation and Lagrange multiplier testing of a general heteroskedastic one-way error components regression model assuming that $\mu_i \sim \left(0, \sigma^2_{\mu_i}\right)$ and $\nu_{it} \sim \left(0, \sigma^2_{\nu_{it}}\right)$ where $\sigma^2_{\mu_i}$ and $\sigma^2_{\nu_{it}}$ are *distinct* parametric functions of exogenous variables Z_{it} and F_i, i.e., $\sigma^2_{\nu_{it}} = \sigma^2_\nu h_\nu\left(Z_{it}\theta_1\right)$ and $\sigma^2_{\mu_i} = \sigma^2_\mu h_\mu\left(F_i\theta_2\right)$. In the context of incomplete panels, Lejeune (1996) derived two joint LM tests for no individual effects and homoskedasticity in the remainder error term. The first LM test considers a random effects one-way error component model with $\mu_i \sim \text{IIN}\left(0, \sigma^2_\mu\right)$ and a remainder error term that is heteroskedastic $\nu_{it} \sim N\left(0, \sigma^2_{\nu_{it}}\right)$ with $\sigma^2_{\nu_{it}} = \sigma^2_\nu h_\nu\left(Z_{it}\theta_1\right)$. The joint hypothesis $H_0 : \theta_1 = \sigma^2_\mu = 0$ renders OLS the restricted MLE. Lejeune argued that there is no need to consider a variance function for μ_i since one is testing σ^2_μ equal to zero. While the computation of the LM test statistic is simplified under this assumption, i.e., $\mu_i \sim \text{IIN}\left(0, \sigma^2_\mu\right)$, this is not in the original spirit of Lejeune's ML estimation where both μ_i and ν_{it} have general variance functions. Lejeune's second LM test considers a fixed effects one-way error component model where μ_i is a fixed parameter to be estimated and the remainder error term is heteroskedastic with $\nu_{it} \sim N\left(0, \sigma^2_{\nu_{it}}\right)$ and $\sigma^2_{\nu_{it}} = \sigma^2_\nu h_\nu\left(Z_{it}\theta_1\right)$. The joint hypothesis is $H_0 : \mu_i = \theta_1 = 0$ for all $i = 1, 2, \ldots, N$. This renders OLS the restricted MLE.

Holly and Gardiol (2000) derived a score test for homoskedasticity in a one-way error component model where the alternative model is that the μ_i's are independent and distributed as $N(0, \sigma^2_{\mu_i})$ where $\sigma^2_{\mu_i} = \sigma^2_\mu h_\mu\left(F_i\theta_2\right)$. Here, F_i is a vector of p explanatory variables such that $F_i\theta_2$ does not contain a constant term and h_μ is a strictly positive twice differentiable function satisfying $h_\mu(0) = 1$ with $h'_\mu(0) \neq 0$ and $h''_\mu(0) \neq 0$. The score test statistic for $H_0 : \theta_2 = 0$ turns out to be one half the explained sum of squares of the OLS regression of $(\hat{s}/\bar{s}) - \iota_N$ against the p regressors in F as in the Breusch and Pagan test for homoskedasticity. Here $\hat{s}_i = \hat{u}'_i \bar{J}_T \hat{u}_i$ and $\bar{s} = \sum_{i=1}^N \hat{s}_i/N$ where $\widehat{u}$ denotes the maximum likelihood residuals from the restricted model under $H_0 : \theta_2 = 0$. This is a one-way homoskedastic error component model with $\mu_i \sim N(0, \sigma^2_\mu)$. The reader is asked to verify this result in problem 5.3.

In the spirit of the general heteroskedastic model of Randolph (1988) and Lejeune (1996), Baltagi, Bresson and Pirotte (2005b) derived a *joint* Lagrange multiplier test for homoskedasticity, i.e., $H_0 : \theta_1 = \theta_2 = 0$. Under the null hypothesis, the model is a homoskedastic one-way error component regression model. Note that this is different from Lejeune (1996), where under his null, $\sigma^2_\mu = 0$, so that the restricted MLE is OLS and not MLE on a one-way homoskedastic error component model. Allowing for $\sigma^2_\mu > 0$ is more likely to be the case in panel data where heterogeneity across the individuals is likely to be present even if heteroskedasticity is not. The model under the null is exactly that of Holly and Gardiol (2000), but it is more general under the alternative since it does not assume a homoskedastic remainder error term. Next, Baltagi et al. (2005b) derived an LM test for the null hypothesis of homoskedasticity of the individual random effects assuming homoskedasticity of the remainder error term, i.e., $\theta_2 = 0 \mid \theta_1 = 0$. Not surprisingly, they get the Holly and Gardiol (2000) LM test. Last but not least, Baltagi et al. (2005b) derived an LM test for the null hypothesis of homoskedasticity of the remainder error term assuming homoskedasticity of the individual effects, *i.e.*, $\theta_1 = 0 \mid \theta_2 = 0$. The details for the derivations and the resulting statisitics are not provided here and the reader is referred to their paper. Monte Carlo experiments showed that the joint LM test performed well when both error components were heteroskedastic, and performed second best when one of the components was homoskedastic while the other was not. In contrast, the marginal LM tests performed best when heteroskedasticity was present in the right error component. They yielded misleading results if heteroskedasticity was present in the wrong error component.

5.2 SERIAL CORRELATION

The classical error component disturbances given by (2.2) assume that the only correlation over time is due to the presence of the same individual across the panel. In Chapter 2, this equicorrelation coefficient was shown to be correl$(u_{it}, u_{is}) = \sigma_\mu^2/(\sigma_\mu^2 + \sigma_\nu^2)$ for $t \neq s$. Note that it is the same no matter how far t is from s. This may be a restrictive assumption for economic relationships, like investment or consumption, where an unobserved shock this period will affect the behavioral relationship for at least the next few periods. This type of serial correlation is not allowed for in the simple error component model. Ignoring serial correlation when it is present results in consistent but inefficient estimates of the regression coefficients and biased standard errors. This section introduces serial correlation in the ν_{it}, first as an autoregressive process of order one AR(1), as in the Lillard and Willis (1978) study on earnings. Next, as a second-order autoregressive process AR(2), also as a special fourth-order autoregressive process AR(4) for quarterly data and finally as a first-order moving average MA(1) process. For all these serial correlation specifications, a simple generalization of the Fuller and Battese (1973) transformation is derived and the implications for predictions are given. Testing for individual effects and serial correlation is taken up in the last subsection.

5.2.1 The AR(1) Process

Lillard and Willis (1978) generalized the error component model to the serially correlated case, by assuming that the remainder disturbances (the ν_{it}) follow an AR(1) process. In this case $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$, whereas

$$\nu_{it} = \rho \nu_{i,t-1} + \epsilon_{it} \tag{5.8}$$

$|\rho| < 1$ and $\epsilon_{it} \sim \text{IID}(0, \sigma_\epsilon^2)$. The μ_i are independent of the ν_{it} and $\nu_{i0} \sim (0, \sigma_\epsilon^2/(1-\rho^2))$. Baltagi and Li (1991a) derived the corresponding Fuller and Battese (1974) transformation for this model. First, one applies the Prais–Winsten (PW) transformation matrix

$$C = \begin{bmatrix} (1-\rho^2)^{1/2} & 0 & 0 & \cdots & 0 & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\rho & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -\rho & 1 \end{bmatrix}$$

to transform the remainder AR(1) disturbances into serially uncorrelated classical errors. For panel data, this has to be applied for N individuals. The transformed regression disturbances are in vector form

$$u^* = (I_N \otimes C)u = (I_N \otimes C\iota_T)\mu + (I_N \otimes C)\nu \tag{5.9}$$

Using the fact that $C\iota_T = (1-\rho)\iota_T^\alpha$, where $\iota_T^{\alpha\prime} = (\alpha, \iota_{T-1}^\prime)$ and $\alpha = \sqrt{(1+\rho)/(1-\rho)}$, one can rewrite (5.9) as

$$u^* = (1-\rho)(I_N \otimes \iota_T^\alpha)\mu + (I_N \otimes C)\nu \tag{5.10}$$

Therefore, the variance–covariance matrix of the transformed disturbances is

$$\Omega^* = E(u^* u^{*\prime}) = \sigma_\mu^2 (1-\rho)^2 [I_N \otimes \iota_T^\alpha \iota_T^{\alpha\prime}] + \sigma_\epsilon^2 (I_N \otimes I_T)$$

since $(I_N \otimes C)E(\nu\nu')(I_N \otimes C') = \sigma_\epsilon^2(I_N \otimes I_T)$. Alternatively, this can be rewritten as

$$\Omega^* = d^2\sigma_\mu^2(1-\rho)^2[I_N \otimes \iota_T^\alpha \iota_T^{\alpha\prime}/d^2] + \sigma_\epsilon^2(I_N \otimes I_T) \tag{5.11}$$

where $d^2 = \iota_T^{\alpha\prime}\iota_T^\alpha = \alpha^2 + (T-1)$. This replaces $J_T^\alpha = \iota_T^\alpha\iota_T^{\alpha\prime}$ by $d^2\bar{J}_T^\alpha$, its idempotent counterpart, where $\bar{J}_T^\alpha = \iota_T^\alpha\iota_T^{\alpha\prime}/d^2$. Extending the Wansbeek and Kapteyn trick, we replace I_T by $E_T^\alpha + \bar{J}_T^\alpha$, where $E_T^\alpha = I_T - \bar{J}_T^\alpha$. Collecting terms with the same matrices, one obtains the spectral decomposition of Ω^*,

$$\Omega^* = \sigma_\alpha^2(I_N \otimes \bar{J}_T^\alpha) + \sigma_\epsilon^2(I_N \otimes E_T^\alpha) \tag{5.12}$$

where $\sigma_\alpha^2 = d^2\sigma_\mu^2(1-\rho)^2 + \sigma_\epsilon^2$. Therefore

$$\sigma_\epsilon\Omega^{*-1/2} = (\sigma_\epsilon/\sigma_\alpha)(I_N \otimes \bar{J}_T^\alpha) + (I_N \otimes E_T^\alpha) = I_N \otimes I_T - \theta_\alpha(I_N \otimes \bar{J}_T^\alpha) \tag{5.13}$$

where $\theta_\alpha = 1 - (\sigma_\epsilon/\sigma_\alpha)$.

Premultiplying the PW transformed observations $y^* = (I_N \otimes C)y$ by $\sigma_\epsilon\Omega^{*-1/2}$ one gets $y^{**} = \sigma_\epsilon\Omega^{*-1/2}y^*$. The typical elements of $y^{**} = \sigma_\epsilon\Omega^{*-1/2}y^*$ are given by

$$(y_{i1}^* - \theta_\alpha\alpha b_i, y_{i2}^* - \theta_\alpha b_i, \ldots, y_{iT}^* - \theta_\alpha b_i)' \tag{5.14}$$

where $b_i = [(\alpha y_{i1}^* + \sum_2^T y_{it}^*)/d^2]$ for $i = 1, \ldots, N$.[1] The first observation gets special attention in the AR(1) error component model. First, the PW transformation gives it a special weight $\sqrt{1-\rho^2}$ in y^*. Second, the Fuller and Battese transformation gives it a special weight $\alpha = \sqrt{(1+\rho)/(1-\rho)}$ in computing the weighted average b_i and the pseudo-difference in (5.14). Note that (i) if $\rho = 0$, then $\alpha = 1$, $d^2 = T$, $\sigma_\alpha^2 = \sigma_1^2$ and $\theta_\alpha = \theta$. Therefore, the typical element of y_{it}^{**} reverts to the familiar $(y_{it} - \theta\bar{y}_{i.})$ transformation for the one-way error component model with no serial correlation. (ii) If $\sigma_\mu^2 = 0$, then $\sigma_\alpha^2 = \sigma_\epsilon^2$ and $\theta_\alpha = 0$. Therefore, the typical element of y_{it}^{**} reverts to the PW transformation y_{it}^*.

The BQU estimators of the variance components arise naturally from the spectral decomposition of Ω^*. In fact, $(I_N \otimes E_T^\alpha)u^* \sim (0, \sigma_\epsilon^2[I_N \otimes E_T^\alpha])$ and $(I_N \otimes \bar{J}_T^\alpha)u^* \sim (0, \sigma_\alpha^2[I_N \otimes \bar{J}_T^\alpha])$ and

$$\widehat{\sigma}_\epsilon^2 = u^{*\prime}(I_N \otimes E_T^\alpha)u^*/N(T-1) \quad \text{and} \quad \widehat{\sigma}_\alpha^2 = u^{*\prime}(I_N \otimes \bar{J}_T^\alpha)u^*/N \tag{5.15}$$

provide the BQU estimators of σ_ϵ^2 and σ_α^2, respectively. Baltagi and Li (1991a) suggest estimating ρ from Within residuals $\widetilde{\nu}_{it}$ as $\widetilde{\rho} = \sum_{i=1}^N \sum_{t=1}^T \widetilde{\nu}_{it}\widetilde{\nu}_{i,t-1}/\sum_{i=1}^N \sum_{t=2}^T \widetilde{\nu}_{i,t-1}^2$. Then, $\widehat{\sigma}_\epsilon^2$ and $\widehat{\sigma}_\alpha^2$ are estimated from (5.15) by substituting OLS residuals $\hat{u}^*$ from the PW transformed equation using $\widetilde{\rho}$. Using Monte Carlo experiments, Baltagi and Li (1997) found that $\widetilde{\rho}$ performs poorly for small T and recommended an alternative estimator of ρ which is based on the autocovariance function $Q_s = E(u_{it}u_{i,t-s})$. For the AR(1) model given in (5.8), it is easy to show that $Q_s = \sigma_\mu^2 + \sigma_\nu^2\rho^s$. From Q_0, Q_1 and Q_2, one can easily show that $\rho + 1 = (Q_0 - Q_2)/(Q_0 - Q_1)$. Hence, a consistent estimator of ρ (for large N) is given by

$$\widehat{\rho} = \frac{\widetilde{Q}_0 - \widetilde{Q}_2}{\widetilde{Q}_0 - \widetilde{Q}_1} - 1 = \frac{\widetilde{Q}_1 - \widetilde{Q}_2}{\widetilde{Q}_0 - \widetilde{Q}_1}$$

where $\widetilde{Q}_s = \sum_{i=1}^N \sum_{t=s+1}^T \widehat{u}_{it}\widehat{u}_{i,t-s}/N(T-s)$ and $\widehat{u}_{it}$ denotes the OLS residuals on (2.1). $\widehat{\sigma}_\epsilon^2$ and $\widehat{\sigma}_\alpha^2$ are estimated from (5.15) by substituting OLS residuals $\hat{u}^*$ from the PW transformed equation using $\widehat{\rho}$ rather than $\widetilde{\rho}$.

Therefore, the estimation of an AR(1) serially correlated error component model is considerably simplified by (i) applying the PW transformation in the first step, as is usually done in the time-series literature, and (ii) subtracting a pseudo-average from these transformed data as in (5.14) in the second step.

Empirical Applications

Lillard and Weiss (1979) apply the first-order autoregressive error component model to study the sources of variation in the earnings of American scientists over the decade 1960–70. The disturbances are assumed to be of the form

$$u_{it} = \mu_i + \xi_i(t - \bar{t}) + \nu_{it}$$

with $\nu_{it} = \rho\nu_{i,t-1} + \epsilon_{it}$ as in (5.8), $\epsilon_{it} \sim \text{IID}(0, \sigma_\epsilon^2)$ and

$$\begin{pmatrix} \mu_i \\ \xi_i \end{pmatrix} \sim (0, \Sigma_{\mu\xi})$$

Unlike the individual effect μ_i which represents unmeasured characteristics like ability that affect the levels of earnings and persist throughout the period of observation, ξ_i represents the effect of omitted variables which affect the growth in earnings. ξ_i could be the individual's learning ability, so it is highly likely that μ_i and ξ_i are correlated. Lillard and Weiss (1979) derive the MLE and GLS for this model and offer two generalizations for the error structure.

Berry, Gottschalk and Wissoker (1988) apply the one-way error component model with first-order autoregressive remainder disturbances to study the impact of plant closing on the mean and variance of log earnings. The data are drawn from the Panel Study of Income Dynamics (PSID) and includes male heads of households who were less than 65 years old and not retired. The sample period considered spans seven years (1975–81) and allows observation over the pre- and post-displacement earnings histories. The sample is not limited only to displaced workers and therefore naturally provides a control group. Their findings show that during the period of displacement, mean earnings decline while the variance of earnings increases sharply. This causes a dramatic increase in the proportion of persons earning less than \$10 000. However, this is temporary, as the mean earnings increase in the post-displacement period and the variance of earnings declines back to its pre-displacement level.

5.2.2 The AR(2) Process

This simple transformation can be extended to allow for an AR(2) process on the ν_{it}, i.e.

$$\nu_{it} = \rho_1\nu_{i,t-1} + \rho_2\nu_{i,t-2} + \epsilon_{it} \tag{5.16}$$

where $\epsilon_{it} \sim \text{IIN}(0, \sigma_\epsilon^2)$, $|\,\rho_2\,| < 1$ and $|\,\rho_1\,| < (1 - \rho_2)$. Let $E(\nu_i\nu_i') = \sigma_\epsilon^2 V$, where $\nu_i' = (\nu_{i1}, \ldots, \nu_{iT})$ and note that V is invariant to $i = 1, \ldots, N$. The unique $T \times T$ lower triangular matrix C with positive diagonal elements which satisfies $CVC' = I_T$ is given by

$$C = \begin{bmatrix} \gamma_0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ -\gamma_2 & \gamma_1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ -\rho_2 & -\rho_1 & 1 & 0 & \dots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -\rho_2 & -\rho_1 & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & -\rho_2 & -\rho_1 & 1 \end{bmatrix}$$

where $\gamma_0 = \sigma_\epsilon/\sigma_\nu$, $\gamma_1 = \sqrt{1-\rho_2^2}$, $\gamma_2 = \gamma_1[\rho_1/(1-\rho_2)]$ and $\sigma_\nu^2 = \sigma_\epsilon^2(1-\rho_2)/(1+\rho_2)[(1-\rho_2)^2 - \rho_1^2]$. The transformed disturbances are given by

$$u^* = (I_N \otimes C)u = (1-\rho_1-\rho_2)(I_N \otimes \iota_T^\alpha)\mu + (I_N \otimes C)\nu \tag{5.17}$$

Using the fact that $C\iota_T = (1-\rho_1-\rho_2)\times$(the new ι_T^α) where $\iota_T^{\alpha\prime} = (\alpha_1, \alpha_2, \iota_{T-2}')$, $\alpha_1 = \sigma_\epsilon/\sigma_\nu(1-\rho_1-\rho_2)$, and $\alpha_2 = \sqrt{(1+\rho_2)/(1-\rho_2)}$.

Similarly, one can define

$$d^2 = \iota_T^{\alpha\prime}\iota_T^\alpha = \alpha_1^2 + \alpha_2^2 + (T-2),\ J_T^\alpha,\ E_T^\alpha,\ \text{etc.}$$

as in section 5.2.1, to obtain

$$\Omega^* = d^2\sigma_\mu^2(1-\rho_1-\rho_2)^2[I_N \otimes \bar{J}_T^\alpha] + \sigma_\epsilon^2[I_N \otimes I_T] \tag{5.18}$$

as in (5.11). The only difference is that $(1-\rho_1-\rho_2)$ replaces $(1-\rho)$ and ι_T^α is defined in terms of α_1 and α_2 rather than α. Similarly, one can obtain $\sigma_\epsilon\Omega^{*-1/2}$ as in (5.13) with $\sigma_\alpha^2 = d^2\sigma_\mu^2(1-\rho_1-\rho_2)^2 + \sigma_\epsilon^2$. The typical elements of $y^{**} = \sigma_\epsilon\Omega^{*-1/2}y^*$ are given by

$$(y_{i1}^* - \theta_\alpha\alpha_1 b_i,\ y_{i2}^* - \theta_\alpha\alpha_2 b_i,\ y_{i3}^* - \theta_\alpha b_i, \dots, y_{iT}^* - \theta_\alpha b_i) \tag{5.19}$$

where $b_i = [(\alpha_1 y_{i1}^* + \alpha_2 y_{i2}^* + \sum_3^T y_{it}^*)/d^2]$. The first two observations get special attention in the AR(2) error component model. First in the matrix C defined above (5.17) and second in computing the average b_i and the Fuller and Battese transformation in (5.19). Therefore, one can obtain GLS on this model by (i) transforming the data as in the time-series literature by the C matrix defined above (5.17) and (ii) subtracting a pseudo-average in the second step as in (5.19).

5.2.3 The AR(4) Process for Quarterly Data

Consider the specialized AR(4) process for quarterly data, i.e. $\nu_{it} = \rho\nu_{i,t-4} + \epsilon_{it}$, where $|\rho| < 1$ and $\epsilon_{it} \sim \text{IIN}(0, \sigma_\epsilon^2)$. The C matrix for this process can be defined as follows: $u_i^* = Cu_i$ where

$$\begin{aligned} u_{it}^* &= \sqrt{1-\rho^2}\, u_{it} && \text{for } t = 1,2,3,4 \\ &= u_{it} - \rho u_{i,t-4} && \text{for } t = 5,6,\dots,T \end{aligned} \tag{5.20}$$

This means that the μ_i component of u_{it} gets transformed as $\sqrt{1-\rho^2}\,\mu_i$ for $t = 1,2,3,4$ and as $(1-\rho)\mu_i$ for $t = 5,6,\dots,T$. This can be rewritten as $\alpha(1-\rho)\mu_i$ for $t = 1,2,3,4$ where $\alpha = \sqrt{(1+\rho)/(1-\rho)}$, and $(1-\rho)\mu_i$ for $t = 5,\dots,T$. So that $u^* = (I_N \otimes C)u$ is given by (5.9) with a new C, the same α, but $\iota_T^{\alpha\prime} = (\alpha, \alpha, \alpha, \alpha, \iota_{T-4}')$, $d^2 = \iota_T^{\alpha\prime}\iota_T^\alpha = 4\alpha^2 + (T-4)$, and the derivations Ω^* and $\sigma_\epsilon \Omega^{*-1/2}$ in (5.12) and (5.13) are the same. The typical elements of $y^{**} = \sigma_\epsilon \Omega^{*-1/2} y^*$ are given by

$$(\overset{*}{y}_{i1} - \theta_\alpha \alpha b_i, \ldots, \overset{*}{y}_{i4} - \theta_\alpha b_i, \overset{*}{y}_{i5} - \theta_\alpha b_i, \ldots, \overset{*}{y}_{iT} - \theta_\alpha b_i) \tag{5.21}$$

where $b_i = [(\alpha(\sum_{t=1}^{4} \overset{*}{y}_{it}) + \sum_{t=5}^{T} \overset{*}{y}_{it})/d^2]$. Once again, GLS can easily be computed by applying (5.20) to the data in the first step and (5.21) in the second step.

5.2.4 The MA(1) Process

For the MA(1) model, defined by

$$\nu_{it} = \epsilon_{it} + \lambda \epsilon_{i,t-1} \tag{5.22}$$

where $\epsilon_{it} \sim \text{IIN}(0, \sigma_\epsilon^2)$ and $| \lambda |< 1$, Balestra (1980) gives the following C matrix, $C = D^{-1/2} P$ where $D = \text{diag}\{a_t, a_{t-1}\}$ for $t = 1, \ldots, T$,

$$P = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ \lambda & a_1 & 0 & \ldots & 0 \\ \lambda^2 & a_1\lambda & a_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \lambda^{T-1} & a_1\lambda^{T-2} & a_2\lambda^{T-3} & \ldots & a_{T-1} \end{bmatrix}$$

and $a_t = 1 + \lambda^2 + \ldots + \lambda^{2t}$ with $a_0 = 1$. For this C matrix, one can show that the new $\iota_T^\alpha = C\iota_T = (\alpha_1, \alpha_2, \ldots, \alpha_T)'$ where these α_t can be solved for recursively as follows:

$$\begin{aligned} \alpha_1 &= (a_0/a_1)^{1/2} \\ \alpha_t &= \lambda(a_{t-2}/a_{t-1})^{1/2}\alpha_{t-1} + (a_{t-1}/a_t)^{1/2} \qquad t = 2, \ldots, T \end{aligned} \tag{5.23}$$

Therefore, $d^2 = \iota_T^{\alpha\prime} \iota_T^\alpha = \sum_{t=1}^{T} \alpha_t^2$, $\sigma_\alpha^2 = d^2\sigma_\mu^2 + \sigma_\epsilon^2$ and the spectral decomposition of Ω^* is the same as that given in (5.12), with the newly defined ι_T^α and σ_α^2. The typical elements of $y^{**} = \sigma_\epsilon \Omega^{*-1/2} y^*$ are given by

$$(\overset{*}{y}_{i1} - \theta_\alpha \alpha_1 b_i, \ldots, \overset{*}{y}_{iT} - \theta_\alpha \alpha_T b_i) \tag{5.24}$$

where $b_i = [\sum_{t=1}^{T} \alpha_t \overset{*}{y}_{it}/d^2]$. Therefore, for an MA(1) error component model, one applies the recursive transformation given in (5.23) in the first step and subtracts a pseudo-average described in (5.24) in the second step; see Baltagi and Li (1992b) for more details. In order to implement the estimation of an error component model with MA(1) remainder errors, Baltagi and Li (1997) proposed an alternative transformation that is simple to compute and requires only least squares. This can be summarized as follows.

Let $\gamma_s = E(\nu_{it}\nu_{i,t-s})$ denote the autocovariance function of ν_{it} and $r = \gamma_1/\gamma_0$. Note that when ν_{it} follows an MA(1) process, we have $Q_s = \sigma_\mu^2 + \gamma_s$ for $s = 0, 1$ and $Q_s = \sigma_\mu^2$ for $s > 1$. Hence we have $\gamma_\tau = Q_\tau - Q_s (\tau = 0, 1)$ for some $s > 1$.

Step 1. Compute $y_{i1}^* = y_{i1}/\sqrt{g_1}$ and $y_{it}^* = [y_{it} - (ry_{i,t-1}^*)/\sqrt{g_{t-1}}]/\sqrt{g_t}$ for $t = 2, \ldots, T$, where $g_1 = 1$ and $g_t = 1 - r^2/g_{t-1}$ for $t = 2, \ldots, T$. Note that this transformation depends only on r, which can be estimated by $\widehat{r} = \widehat{\gamma}_1/\widehat{\gamma}_0 = (\widetilde{Q}_1 - \widetilde{Q}_s)/(\widetilde{Q}_0 - \widetilde{Q}_s)$ for some $s > 1$.

Step 2. Compute y^{**} using the result that $\iota_T^{\alpha} = C\iota_T = (\alpha_1, \ldots, \alpha_T)'$ with $\alpha_1 = 1$ and $\alpha_t = [1 - r/\sqrt{g_{t-1}}]/\sqrt{g_t}$ for $t = 2, \ldots, T$. Note that in this case $\sigma^2 = \gamma_0$. The estimators of σ_α^2 and σ^2 are simply given by $\widehat{\sigma}_\alpha^2 = (\sum_{t=1}^{T} \widehat{\alpha}_t^2)\widehat{\sigma}_\mu^2 + \widehat{\sigma}^2$ and $\widehat{\sigma}^2 = \widehat{\gamma}_0 = \widetilde{Q}_0 - \widetilde{Q}_s$ for some $s > 1$ with $\widehat{\sigma}_\mu^2 = \widetilde{Q}_s$ for some $s > 1$. Finally $\widehat{\delta} = 1 - \sqrt{\widehat{\gamma}_0/\widehat{\sigma}_\alpha^2}$. Again, the OLS estimator on the (**) transformed equation is equivalent to GLS on (2.1).

The advantages of this approach are by now evident: $\widehat{\sigma}^2 = \widehat{\gamma}_0$ is trivially obtained from OLS residuals. This is because we did not choose $\sigma_\epsilon^2 = \sigma^2$ as in Baltagi and Li (1991a). Next we estimated γ's rather than the moving average parameter λ. The $\widehat{\gamma}$'s require only linear least squares, whereas $\widehat{\lambda}$ requires nonlinear least squares. Finally, our proposed estimation procedure requires simple recursive transformations that are very easy to program. This should prove useful for panel data users.

In summary, a simple transformation for the one-way error component model with serial correlation can easily be generalized to any error process generating the remainder disturbances ν_{it} as long as there exists a *simple* $T \times T$ matrix C such that the transformation $(I_N \otimes C)\nu$ has zero mean and variance $\sigma^2 I_{NT}$.

Step 1. Perform the C transformation on the observations of each individual $y_i' = (y_{i1}, \ldots, y_{iT})$ to obtain $y_i^* = Cy_i$ free of serial correlation.

Step 2. Perform another transformation on the y_{it}^*'s, obtained in step 1, which subtracts from y_{it}^* a fraction of a weighted average of observations on y_{it}^*, i.e.,

$$y_{it}^{**} = y_{it}^* - \theta_\alpha \alpha_t (\Sigma_{s=1}^{T} \alpha_s y_{is}^*)/(\Sigma_{s=1}^{T} \alpha_s^2)$$

where the α_t's are the elements of $\iota_T^{\alpha} = C\iota_T \equiv (\alpha_1, \alpha_2, \ldots, \alpha_T)'$ and $\theta_\alpha = 1 - (\sigma/\sigma_\alpha)$ with $\sigma_\alpha^2 = \sigma_\mu^2(\Sigma_{t=1}^{T}\alpha_t^2) + \sigma^2$. See Baltagi and Li (1994) for an extension to the MA(q) case and Galbraith and Zinde-Walsh (1995) for an extension to the ARMA(p,q) case.

5.2.5 Unequally Spaced Panels with AR(1) Disturbances

Some panel data sets cannot be collected every period due to lack of resources or cuts in funding. Instead, these panels are collected over unequally spaced time intervals. For example, a panel of households could be collected over unequally spaced years rather than annually. This is also likely when collecting data on countries, states or firms where, in certain years, the data are not recorded, are hard to obtain, or are simply missing. Other common examples are panel data sets using daily data from the stock market, including stock prices, commodity prices, futures, etc. These panel data sets are unequally spaced when the market closes on weekends and holidays. This is also common for housing resale data where the pattern of resales for each house occurs at different time periods and the panel is unbalanced because we observe different numbers of resales for each house. Baltagi and Wu (1999) extend the Baltagi and Li (1991a) results to the estimation of an unequally spaced panel data regression model with AR(1) remainder disturbances. A feasible generalized least squares procedure is proposed as a weighted least squares that can handle a wide range of unequally spaced panel data patterns. This procedure is simple to compute and provides natural estimates of the serial correlation and variance components parameters. Baltagi and Wu (1999) also provide a locally best invariant (LBI) test for zero first-order serial correlation against positive or negative serial

Table 5.1 Grunfeld's Data. Random Effects and AR(1) Remainder Disturbances

```
. xtregar I F C , re lbi

RE GLS regression with AR(1)           Number of obs      =        200
    disturbances
Group variable (i): fn                 Number of groups   =         10

R-sq:  within  = 0.7649                 Obs per group: min =         20
       between = 0.8068                                avg =       20.0
       overall = 0.7967                                max =         20

                                       Wald chi2(3)       =     360.31
corr(u_i, Xb)      = 0 (assumed)        Prob > chi2        =     0.0000

------------------------------------------------------------------------
         I |     Coef.  Std. Err.     z    P>|z|  [95% Conf. Interval]
----------+-------------------------------------------------------------
         F |  .0949215  .0082168  11.55   0.000    .0788168   .1110262
         C |  .3196589  .0258618  12.36   0.000    .2689707   .3703471
     _cons | -44.38123  26.97525  -1.65   0.100   -97.25175   8.489292
----------+-------------------------------------------------------------
    rho_ar | .67210608  (estimated autocorrelation coefficient)
   sigma_u | 74.517098
   sigma_e | 41.482494
   rho_fov |  .7634186  (fraction of variance due to u_i)
     theta | .67315699
------------------------------------------------------------------------
modified Bhargava et al. Durbin-Watson = .6844797
Baltagi-Wu LBI = .95635623
```

correlation in case of unequally spaced panel data. Details are given in that paper. This is programed in Stata under the (xtregar,re lbi) command. Table 5.1 gives the Stata output for Grunfeld's investment equation, given in (2.40), with random effects and an AR(1) remainder disturbance term. The bottom of Table 5.1 produces the Baltagi–Wu LBI statistic of 0.956 and the Bhargava, Franzini and Narendranathan (1982) Durbin–Watson statistic for zero first-order serial correlation described in (5.44) below. Both tests reject the null hypothesis of no first-order serial correlation. The estimate of ρ for the AR(1) remainder disturbances is 0.67 while $\widehat{\sigma}_\mu = 74.52$ and $\widehat{\sigma}_\nu = 41.48$. Note that $\widehat{\beta}_1$ in (2.41) drops from 0.110 for a typical random effects GLS estimator reported in Table 2.1 to 0.095 for the random effects GLS estimator with AR(1) remainder disturbances in Table 5.1. This is contrasted to an increase in $\widehat{\beta}_2$ from 0.308 in Table 2.1 to 0.320 in Table 2.5. Table 5.2 gives the TSP output for the maximum likelihood estimates of the random effects model with AR(1) remainder disturbances under the normality assumption. The results are similar to the feasible GLS estimates reported in Table 5.1. Note that if we have missing data on say 1951 and 1952, Stata computes this unequally spaced panel estimation for the random effects with AR(1) disturbances. Table 5.3 reproduces this output. Note that it is based on 180 observations, due to the loss of two years of data for all 10 firms. The Baltagi–Wu LBI statistic is 1.139 and the Bhargava et al. (1982) Durbin–Watson statistic is 0.807, exactly as reported in table 1 of Baltagi and Wu (1999, p. 822). Both test statistics

Table 5.2 Grunfeld's Data. MLE Random Effects with AR(1) Disturbances

```
Balanced data:  N=    10, T_I=   20, NOB=    200
Working space used: 3981
CONVERGENCE ACHIEVED AFTER  13 ITERATIONS
      32 FUNCTION EVALUATIONS.

Schwarz B.I.C. = 1052.412710  Log likelihood = -1039.166917
```

Parameter	Estimate	Standard Error	t-statistic	P-value
C	-40.79118966	29.05697837	-1.403834533	[.160]
F	.0937033982	.7963697796E-02	11.76631769	[.000]
K	.3135856916	.0319818319	9.805119753	[.000]
RHO	.8155980082	.0711931733	11.45612662	[.000]
RHO_I	.7580118599	.1187601536	6.382712018	[.000]
SIGMA2	6958.604792	3306.005910	2.104837372	[.035]

Standard Errors computed from analytic second derivatives (Newton)

Parameter	Estimate	Standard Error	t-statistic	P-value
S2_I	5274.704961	3318.394021	1.589535458	[.112]
S2_IT	1683.899831	174.4331156	9.653555893	[.000]

This TSP output is available at (http://www.stanford.edu/~clint /bench/grar1rei.out).

reject the null hypothesis of no first-order serial correlation. Problem 5.19 asks the reader to replicate these results for other patterns of missing observations.

5.2.6 Prediction

In section 2.5 we derived Goldberger's (1962) BLUP of $y_{i,T+S}$ for the one-way error component model without serial correlation. For ease of reference, we reproduce equation (2.37) for predicting one period ahead for the ith individual

$$\widehat{y}_{i,T+1} = Z'_{i,T+1}\widehat{\delta}_{\text{GLS}} + w'\Omega^{-1}\widehat{u}_{\text{GLS}} \tag{5.25}$$

where $\widehat{u}_{\text{GLS}} = y - Z\widehat{\delta}_{\text{GLS}}$ and $w = E(u_{i,T+1}u)$. For the AR(1) model with no error components, a standard result is that the last term in (5.25) reduces to $\rho\widehat{u}_{i,T}$, where $\widehat{u}_{i,T}$ is the Tth GLS residual for the ith individual. For the one-way error component model without serial correlation (see Taub, 1979 or section 2.5), the last term of (5.25) reduces to $[T\sigma_\mu^2/(T\sigma_\mu^2 + \sigma_\nu^2)]\overline{\widehat{u}}_{i.}$, where $\overline{\widehat{u}}_{i.} = \sum_{t=1}^{T}\widehat{u}_{it}/T$ is the average of the ith individual's GLS residuals. This section summarizes the Baltagi and Li (1992b) derivation of the last term of (5.25) when *both* error components and serial correlation are present. This provides the applied researcher with a simple way of augmenting the GLS predictions obtained from the Fuller and Battese (1973) transformation described above.

Table 5.3 Grunfeld's Data. Unequally Spaced Panel

```
.  xtregar I F C if yr!=1951 & yr!= 1952 , re lbi

RE GLS regression with AR(1)          Number of obs      =        180
  disturbances
Group variable (i): fn                Number of groups   =         10

R-sq:  within  = 0.7766                Obs per group: min =         18
       between = 0.8112                               avg =       18.0
       overall = 0.8024                               max =         18

                                      Wald chi2(3)       =     341.38
corr(u_i, Xb)      = 0 (assumed)      Prob > chi2        =     0.0000
-------------------------------------------------------------------
        I |    Coef.  Std. Err.      z   P>|z|  [95% Conf. Interval]
----------+--------------------------------------------------------
        F |  .0919986  .0083459  11.02   0.000   .0756409   .1083563
        C |  .3243706  .0266376  12.18   0.000   .2721618   .3765793
    _cons | -43.01923  27.05662  -1.59   0.112  -96.04924   10.01077
----------+--------------------------------------------------------
   rho_ar | .68934342  (estimated autocorrelation coefficient)
  sigma_u | 74.002133
  sigma_e | 41.535675
  rho_fov | .76043802  (fraction of variance due to u_i)
    theta |  .6551959
-------------------------------------------------------------------
modified Bhargava et al. Durbin-Watson = .80652308
Baltagi-Wu LBI = 1.1394026
```

For the one-way error component model with AR(1) remainder disturbances, considered in section 5.2.1, Baltagi and Li (1992b) find that

$$w'\Omega^{-1}\widehat{u}_{\text{GLS}} = \rho\widehat{u}_{i,T} + \left(\frac{(1-\rho)^2\sigma_\mu^2}{\sigma_\alpha^2}\right)\left[\alpha\hat{u}_{i1}^* + \sum_{t=2}^{T}\hat{u}_{it}^*\right] \tag{5.26}$$

Note that the first PW-transformed GLS residual receives an α weight in averaging across the ith individual's residuals in (5.26). (i) If $\sigma_\mu^2 = 0$, so that only serial correlation is present, (5.26) reduces to $\rho\widehat{u}_{i,T}$. Similarly, (ii) if $\rho = 0$, so that only error components are present, (5.26) reduces to $[T\sigma_\mu^2/(T\sigma_\mu^2+\sigma_\nu^2)]\overline{\widehat{u}}_{i.}$.

For the one-way error component model with remainder disturbances following an AR(2) process, considered in section 5.2.2, Baltagi and Li (1992b) find that

$$\begin{aligned} w'\Omega^{-1}\widehat{u}_{\text{GLS}} = {} & \rho_1\widehat{u}_{i,T-1} + \rho_2\widehat{u}_{i,T-2} \\ & + \left[\frac{(1-\rho_1-\rho_2)^2\sigma_\mu^2}{\sigma_\alpha^2}\right]\left[\alpha_1\hat{u}_{i1}^* + \alpha_2\hat{u}_{i2}^* + \sum_{t=3}^{T}\hat{u}_{it}^*\right] \end{aligned} \tag{5.27}$$

where

$$\alpha_1 = \sigma_\epsilon/\sigma_\nu(1-\rho_1-\rho_2) \quad \alpha_2 = \sqrt{(1+\rho_2)/(1-\rho_2)}$$
$$\sigma_\alpha^2 = d^2\sigma_\mu^2(1-\rho_1-\rho_2)^2 + \sigma_\epsilon^2$$
$$d^2 = \alpha_1^2 + \alpha_2^2 + (T-2)$$

and

$$\begin{aligned}
\hat{u}_{i1}^{*} &= (\sigma_\epsilon/\sigma_\nu)\widehat{u}_{i1} \\
\hat{u}_{i2}^{*} &= \sqrt{1-\rho_2^2}\,[\widehat{u}_{i2} - (\rho_1/(1-\rho_2))\widehat{u}_{i1}] \\
\hat{u}_{it}^{*} &= \widehat{u}_{it} - \rho_1\widehat{u}_{i,t-1} - \rho_2\widehat{u}_{i,t-2} \quad \text{for } t = 3, \ldots, T
\end{aligned}$$

Note that if $\rho_2 = 0$, this predictor reduces to (5.26). Also, note that for this predictor, the first two residuals are weighted differently when averaging across the ith individual's residuals in (5.27).

For the one-way error component model with remainder disturbances following the specialized AR(4) process for quarterly data, considered in section 5.2.3, Baltagi and Li (1992b) find that

$$w'\Omega^{-1}\widehat{u}_{\text{GLS}} = \rho\widehat{u}_{i,T-3} + \left[\frac{(1-\rho)^2\sigma_\mu^2}{\sigma_\alpha^2}\right]\left[\alpha\sum_{t=1}^{4}\hat{u}_{it}^{*} + \sum_{t=5}^{T}\hat{u}_{it}^{*}\right] \tag{5.28}$$

where $\alpha = \sqrt{(1+\rho)/(1-\rho)}$, $\sigma_\alpha^2 = d^2(1-\rho)^2\sigma_\mu^2 + \sigma_\epsilon^2$, $d^2 = 4\alpha^2 + (T-4)$ and

$$\begin{aligned}
u_{it}^{*} &= \sqrt{1-\rho^2}\,u_{it} \qquad \text{for } t = 1, 2, 3, 4 \\
&= u_{it} - \rho u_{i,t-4} \quad \text{for } t = 5, 6, \ldots, T
\end{aligned}$$

Note, for this predictor, that the first four quarterly residuals are weighted by α when averaging across the ith individual's residuals in (5.28).

Finally, for the one-way error component model with remainder disturbances following an MA(1) process, considered in section 5.2.4, Baltagi and Li (1992c) find that

$$\begin{aligned}
w'\Omega^{-1}\widehat{u}_{\text{GLS}} = {}& -\lambda\left(\frac{a_{T-1}}{a_T}\right)^{1/2}\hat{u}_{iT}^{*} \\
&+ \left[1 + \lambda\left(\frac{a_{T-1}}{a_T}\right)^{1/2}\alpha_T\right]\left(\frac{\sigma_\mu^2}{\sigma_\alpha^2}\right)\left[\sum_{t=1}^{T}\alpha_t\hat{u}_{it}^{*}\right]
\end{aligned} \tag{5.29}$$

where the $\hat{u}_{it}^{*}$ can be solved for recursively as follows:

$$\begin{aligned}
\hat{u}_{i1}^{*} &= (a_0/a_1)^{1/2}\hat{u}_{i1} \\
\hat{u}_{it}^{*} &= \lambda(a_{t-2}/a_{t-1})^{1/2}\hat{u}_{i,t-1}^{*} + (a_{t-1}/a_t)^{1/2}\hat{u}_{i,t} \quad t = 2, \ldots, T
\end{aligned}$$

If $\lambda = 0$, then from (5.23) $a_t = \alpha_t = 1$ for all t and (5.29) reduces to the predictor for the error component model with no serial correlation. If $\sigma_\mu^2 = 0$, the second term in (5.29) drops out and the predictor reduces to that of the MA(1) process.

5.2.7 Testing for Serial Correlation and Individual Effects

In this section, we address the problem of *jointly* testing for serial correlation and individual effects. Baltagi and Li (1995) derived three LM statistics for an error component model with first-order serially correlated errors. The first LM statistic jointly tests for zero first-order serial correlation and random individual effects. The second LM statistic tests for zero first-order serial correlation assuming fixed individual effects, and the third LM statistic tests for zero first-order serial correlation assuming random individual effects. In all three cases, Baltagi and

Li (1995) showed that the corresponding LM statistic is the *same* whether the alternative is AR(1) or MA(1). Also, Baltagi and Li (1995) derived two extensions of the Burke, Godfrey and Termayne (1990) AR(1) vs MA(1) test from the time series to the panel data literature. The first extension tests the null of AR(1) disturbances against MA(1) disturbances, and the second the null of MA(1) disturbances against AR(1) disturbances in an error component model. These tests are computationally simple, requiring only OLS or Within residuals. In what follows, we briefly review the basic ideas behind these tests.

Consider the panel data regression given in (2.3)

$$y_{it} = Z_{it}'\delta + u_{it} \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T \tag{5.30}$$

where δ is a $(K + 1) \times 1$ vector of regression coefficients including the intercept. The disturbance follows a one-way error component model

$$u_{it} = \mu_i + \nu_{it} \tag{5.31}$$

where $\mu_i \sim \text{IIN}(0, \sigma_\mu^2)$ and the remainder disturbance follows a stationary AR(1) process: $\nu_{it} = \rho\nu_{i,t-1} + \epsilon_{it}$ with $|\rho| < 1$, or an MA(1) process: $\nu_{it} = \epsilon_{it} + \lambda\epsilon_{i,t-1}$ with $|\lambda| < 1$, and $\epsilon_{it} \sim \text{IIN}(0, \sigma_\epsilon^2)$. In what follows, we will show that the joint LM test statistic for $H_1^a : \sigma_\mu^2 = 0$; $\lambda = 0$ is the same as that for $H_1^b : \sigma_\mu^2 = 0$; $\rho = 0$.

A Joint LM Test for Serial Correlation and Random Individual Effects

Let us consider the joint LM test for the error component model where the remainder disturbances follow an MA(1) process. In this case, the variance–covariance matrix of the disturbances is given by

$$\Omega = E(uu') = \sigma_\mu^2 I_N \otimes J_T + \sigma_\epsilon^2 I_N \otimes V_\lambda \tag{5.32}$$

where

$$V_\lambda = \begin{pmatrix} 1+\lambda^2 & \lambda & 0 & \dots & 0 \\ \lambda & 1+\lambda^2 & \lambda & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1+\lambda^2 \end{pmatrix} \tag{5.33}$$

and the loglikelihood function is given by $L(\delta, \theta)$ in (4.15) with $\theta = (\lambda, \sigma_\mu^2, \sigma_\epsilon^2)'$. In order to construct the LM test statistic for $H_1^a : \sigma_\mu^2 = 0; \lambda = 0$, one needs $D(\theta) = \partial L(\theta)/\partial\theta$ and the information matrix $J(\theta) = E[\partial^2 L(\theta)/\partial\theta\partial\theta']$ evaluated at the restricted maximum likelihood estimator $\widehat{\theta}$. Note that under the null hypothesis $\Omega^{-1} = (1/\sigma_\epsilon^2)I_{NT}$. Using the general Hemmerle and Hartley (1973) formula given in (4.17), one gets the scores

$$\begin{aligned} \partial L(\theta)/\partial\lambda &= NT \sum_{i=1}^{N}\sum_{t=2}^{T} \widehat{u}_{it}\widehat{u}_{i,t-1} \Big/ \sum_{i=1}^{N}\sum_{t=2}^{T} \widehat{u}_{it}^2 \equiv NT(\widehat{u}'\widehat{u}_{-1}/\widehat{u}'\widehat{u}) \qquad (5.34) \\ \partial L(\theta)/\partial\sigma_\mu^2 &= -(NT/2\widehat{\sigma}_\epsilon^2)[1 - \widehat{u}'(I_N \otimes J_T)\widehat{u}/(\widehat{u}'\widehat{u})] \end{aligned}$$

where $\widehat{u}$ denotes the OLS residuals and $\widehat{\sigma}_\epsilon^2 = \widehat{u}'\widehat{u}/NT$. Using (4.19), see Harville (1977), one gets the information matrix

$$\widehat{J} = (NT/2\widehat{\sigma}_\epsilon^4)\begin{pmatrix} T & 2(T-1)\widehat{\sigma}_\epsilon^2/T & 1 \\ 2(T-1)\widehat{\sigma}_\epsilon^2/T & 2\widehat{\sigma}_\epsilon^4(T-1)/T & 0 \\ 1 & 0 & 1 \end{pmatrix} \tag{5.35}$$

Hence the LM statistic for the null hypothesis $H_1^a : \sigma_\mu^2 = 0; \lambda = 0$ is given by

$$\mathrm{LM}_1 = \widehat{D}'\widehat{J}^{-1}\widehat{D} = \frac{NT^2}{2(T-1)(T-2)}[A^2 - 4AB + 2TB^2] \tag{5.36}$$

where $A = [\widehat{u}'(I_N \otimes J_T)\widehat{u}/(\widehat{u}'\widehat{u})] - 1$ and $B = (\widehat{u}'\widehat{u}_{-1}/\widehat{u}'\widehat{u})$. This is asymptotically distributed (for large N) as χ_2^2 under H_1^a.

It remains to show that LM_1 is exactly the same as the joint test statistic for $H_1^b : \sigma_\mu^2 = 0; \rho = 0$, where the remainder disturbances follow an AR(1) process (see Baltagi and Li, 1991b). In fact, if we repeat the derivation given in (5.32)–(5.36), the only difference is to replace the V_λ matrix by its AR(1) counterpart

$$V_\rho = \begin{pmatrix} 1 & \rho & \dots & \rho^{T-1} \\ \rho & 1 & \dots & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \dots & 1 \end{pmatrix}$$

Note that under the null hypothesis, we have $(V_\rho)_{\rho=0} = I_T = (V_\lambda)_{\lambda=0}$ and

$$(\partial V_\rho/\partial\rho)_{\rho=0} = G = (\partial V_\lambda/\partial\lambda)_{\lambda=0}$$

where G is the bidiagonal matrix with bidiagonal elements all equal to one. Using these results, problem 5.14 asks the reader to verify that the resulting joint LM test statistic is the same whether the residual disturbances follow an AR(1) or an MA(1) process. Hence, the joint LM test statistic for random individual effects and first-order serial correlation is independent of the form of serial correlation, whether it is AR(1) or MA(1). This extends the Breusch and Godfrey (1981) result from a time series regression to a panel data regression using an error component model.

Note that the A^2 term is the basis for the LM test statistic for $H_2 : \sigma_\mu^2 = 0$ assuming there is no serial correlation (see Breusch and Pagan, 1980). In fact, $\mathrm{LM}_2 = \sqrt{NT/2(T-1)}A$ is asymptotically distributed (for large N) as $N(0, 1)$ under H_2 against the one-sided alternative $H_2' : \sigma_\mu^2 > 0$, see (4.25). Also, the B^2 term is the basis for the LM test statistic for $H_3 : \rho = 0$ (or $\lambda = 0$) assuming there are no individual effects (see Breusch and Godfrey, 1981). In fact, $\mathrm{LM}_3 = \sqrt{NT^2/(T-1)}B$ is asymptotically distributed (for large N) as $N(0, 1)$ under H_3 against the one-sided alternative $H_3' : \rho$ (or λ) > 0. The presence of an interaction term in the joint LM test statistic, given in (5.36), emphasizes the importance of the joint test when both serial correlation and random individual effects are suspected. However, when T is large the interaction term becomes negligible.

Note that all the LM tests considered assume that the underlying null hypothesis is that of white noise disturbances. However, in panel data applications, especially with large labor panels, one is concerned with individual effects and is guaranteed their existence. In this case, it is inappropriate to test for serial correlation assuming no individual effects as is done in H_3. In fact, if one uses LM_3 to test for serial correlation, one is very likely to reject the null hypothesis of H_3 even if the null is true. This is because the μ_i are correlated for the same individual across time and this will contribute to rejecting the null of no serial correlation.

An LM Test for First-order Serial Correlation in a Random Effects Model

Baltagi and Li (1995) also derived a conditional LM test for first-order serial correlation given the existence of random individual effects. In case of an AR(1) model, the null hypothesis is $H_4^b\ :\ \rho = 0$ (given $\sigma_\mu^2 > 0$) vs $H_4^{b\prime}\ :\ \rho \neq 0$ (given $\sigma_\mu^2 > 0$). The variance–covariance matrix (under the alternative) is

$$\Omega_1 = \sigma_\mu^2(I_N \otimes J_T) + \sigma_\nu^2(I_N \otimes V_\rho) \tag{5.37}$$

Under the null hypothesis H_4^b, we have

$$\begin{aligned}
(\Omega_1^{-1})_{\rho=0} &= (1/\sigma_\epsilon^2)I_N \otimes E_T + (1/\sigma_1^2)I_N \otimes \bar{J}_T \\
(\partial\Omega_1/\partial\rho)\mid_{\rho=0} &= \sigma_\epsilon^2(I_N \otimes G) \\
(\partial\Omega_1/\partial\sigma_\mu^2)\mid_{\rho=0} &= (I_N \otimes J_T) \\
(\partial\Omega_1/\partial\sigma_\epsilon^2)\mid_{\rho=0} &= (I_N \otimes I_T)
\end{aligned}$$

where $\bar{J}_T = \iota_T\iota_T'/T$, $E_T = I_T - \bar{J}_T$, G is a bidiagonal matrix with bidiagonal elements all equal to one, and $\sigma_1^2 = T\sigma_\mu^2 + \sigma_\epsilon^2$.

When the first-order serial correlation is of the MA(1) type, the null hypothesis becomes $H_4^a\ :\ \lambda = 0$ (given that $\sigma_\mu^2 > 0$) vs $H_4^{a\prime}\ :\ \lambda \neq 0$ (given that $\sigma_\mu^2 > 0$). In this case, the variance–covariance matrix is

$$\Omega_2 = \sigma_\mu^2(I_N \otimes J_T) + \sigma_\epsilon^2(I_N \otimes V_\lambda) \tag{5.38}$$

and under the null hypothesis H_4^a,

$$\begin{aligned}
(\Omega_2^{-1})_{\lambda=0} &= (1/\sigma_\epsilon^2)(I_N \otimes E_T) + (1/\sigma_1^2)(I_N \otimes \bar{J}_T) = (\Omega_1^{-1})_{\rho=0} \\
(\partial\Omega_2/\partial\lambda)_{\lambda=0} &= \sigma_\epsilon^2(I_N \otimes G) = (\partial\Omega_1/\partial\rho)\mid_{\rho=0} \\
(\partial\Omega_2/\partial\sigma_\mu^2)\mid_{\lambda=0} &= (I_N \otimes J_T) = (\partial\Omega_1/\partial\sigma_\mu^2)\mid_{\rho=0} \\
(\partial\Omega_2/\partial\sigma_\epsilon^2)\mid_{\lambda=0} &= (I_N \otimes I_T) = (\partial\Omega_1/\partial\sigma_\epsilon^2)\mid_{\rho=0}
\end{aligned}$$

Using these results, problem 5.15 asks the reader to verify that the test statistic for H_4^a is the same as that for H_4^b. This conditional LM statistic, call it LM_4, is not given here but is derived in Baltagi and Li (1995).

To summarize, the conditional LM test statistics for testing first-order serial correlation, assuming random individual effects, are invariant to the form of serial correlation (i.e. whether it is AR(1) or MA(1)). Also, these conditional LM tests require restricted mle of a one-way error component model with random individual effects rather than OLS estimates as is usual with LM tests.

Bera, Sosa-Escudero and Yoon (2001) criticize this loss of simplicity in computation of LM tests that use OLS residuals and suggest an adjustment of these LM tests that are robust to local misspecification. Instead of $\text{LM}_\mu = NTA^2/2(T-1) = \text{LM}_2^2$ for testing $H_2 : \sigma_\mu^2 = 0$ which ignores the possible presence of serial correlation, they suggest computing

$$\text{LM}_\mu^* = \frac{NT(2B - A)^2}{2(T-1)(1-(2/T))}$$

This test essentially modifies LM_μ by correcting the mean and variance of the score $\partial L/\partial\sigma_\mu^2$ for its asymptotic correlation with $\partial L/\partial\rho$. Under the null hypothesis, LM_μ^* is asymptotically distributed as χ_1^2. Under local misspecification, this adjusted test statistic is equivalent to Neyman's $C(\alpha)$ test and shares its optimality properties. Similarly, they suggest computing

$$\mathrm{LM}_\rho^* = \frac{NT^2[B-(A/T)]^2}{(T-1)(1-(2/T))}$$

instead of $\mathrm{LM}_\rho = \mathrm{NT}^2 B^2/(T-1) = \mathrm{LM}_3^2$ to test $H_3 : \rho = 0$, against the alternative that $\rho \neq 0$, ignoring the presence of random individual effects. They also show that

$$\mathrm{LM}_\mu^* + \mathrm{LM}_\rho = \mathrm{LM}_\rho^* + \mathrm{LM}_\mu = \mathrm{LM}_1$$

where LM_1 is the joint LM test given in (5.36). In other words, the two-directional LM test for σ_μ^2 and ρ can be decomposed into the sum of the *adjusted* one-directional test of one type of alternative and the *unadjusted* form of the other hypothesis. Bera et al. (2001) argue that these tests use only OLS residuals and are easier to compute than the conditional LM tests derived by Baltagi and Li (1995). Bera et al. (2001) perform Monte Carlo experiments that show the usefulness of these modified Rao–Score tests in guarding against *local* misspecification.

For the Grunfeld data, we computed $\mathrm{LM}_\mu = 798.162$ in Table 4.2 using the xttest0 command in Stata. Using TSP, $\mathrm{LM}_\rho = 143.523$, $\mathrm{LM}_\mu^* = 664.948$, $\mathrm{LM}_\rho^* = 10.310$ and the joint LM_1 statistic in (5.36) is 808.471. The joint test rejects the null of no first-order serial correlation and no random firm effects. The one-directional tests LM_ρ and LM_ρ^* reject the null of no first-order serial correlation, while the one-directional tests LM_μ and LM_μ^* reject the null of no random firm effects.

An LM Test for First-order Serial Correlation in a Fixed Effects Model

The model is the same as (5.30), and the null hypothesis is $H_5^b : \rho = 0$ given that the μ_i are fixed parameters. Writing each individual's variables in a $T \times 1$ vector form, we have

$$y_i = Z_i\delta + \mu_i\iota_T + \nu_i \tag{5.39}$$

where $y_i = (y_{i1}, y_{i2}, \ldots, y_{iT})'$, Z_i is $T \times (K+1)$ and ν_i is $T \times 1$. $\nu_i \sim N(0, \Omega_\rho)$ where $\Omega_\rho = \sigma_\epsilon^2 V_\rho$ for the AR(1) disturbances. The loglikelihood function is

$$\begin{aligned} L(\delta, \rho, \mu, \sigma_\epsilon^2) = {} & \text{constant} - \frac{1}{2}\log|\,\Omega\,| \\ & -\frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{N}[(y_i - Z_i\delta - \mu_i\iota_T)'V_\rho^{-1}(y_i - Z_i\delta - \mu_i\iota_T)] \end{aligned} \tag{5.40}$$

where $\Omega = I_N \otimes \Omega_\rho$ is the variance–covariance matrix of $\nu' = (\nu_1', \ldots, \nu_N')$. One can easily check that the maximum likelihood estimator of μ_i is given by $\widehat{\mu}_i = \{(\iota_T' V_\rho^{-1}\iota_T)^{-1}[\iota_T' V_\rho^{-1}(y_i - Z_i\widehat{\delta})]\}_{\rho=0} = \bar{y}_{i.} - \bar{Z}_{i.}'\widehat{\delta}$, where $\widehat{\delta}$ is the maximum likelihood estimator of δ, $\bar{y}_{i.} = \sum_{t=1}^{T} y_{it}/T$ and $\bar{Z}_{i.}$ is a $(K+1)\times 1$ vector of averages of Z_{it} across time.

Write the loglikelihood function in vector form of ν as

$$L(\delta, \mu, \theta) = \text{constant} - \frac{1}{2}\log|\,\Omega\,| - \frac{1}{2}\nu'\Omega^{-1}\nu \tag{5.41}$$

where $\theta' = (\rho, \sigma_\epsilon^2)$. Now (5.41) has a similar form to (4.15). By following a similar derivation as that given earlier, one can easily verify that the LM test statistic for testing H_5^b is

$$\text{LM} = [NT^2/(T-1)](\widehat{\nu}'\widehat{\nu}_{-1}/\widehat{\nu}'\widehat{\nu})^2 \tag{5.42}$$

which is asymptotically distributed (for large T) as χ_1^2 under the null hypothesis H_5^b. Note that $\widehat{\nu}_{it} = y_{it} - Z_{it}'\widehat{\delta} - \widehat{\mu}_i = (\widetilde{y}_{it} - \widetilde{Z}_{it}'\widehat{\delta}) + (\bar{y}_{i.} - \bar{Z}_{i.}'\widehat{\delta} - \widehat{\mu}_i)$ where $\widetilde{y}_{it} = y_{it} - \bar{y}_{i.}$ is the usual Within transformation. Under the null of $\rho = 0$, the last term in parentheses is zero since $\{\widehat{\mu}_i\}_{\rho=0} = \bar{y}_{i.} - \bar{Z}_{i.}'\widehat{\delta}$ and $\{\widehat{\nu}_{it}\}_{\rho=0} = \widetilde{y}_{it} - \widetilde{Z}_{it}\widehat{\delta} = \widetilde{\nu}_{it}$. Therefore, the LM statistic given in (5.42) can be expressed in terms of the usual Within residuals (the $\widetilde{\nu}$) and the one-sided test for H_5^b (corresponding to the alternative $\rho > 0$) is

$$\text{LM}_5 = \sqrt{NT^2/(T-1)}(\widetilde{\nu}'\widetilde{\nu}_{-1}/\widetilde{\nu}'\widetilde{\nu}) \tag{5.43}$$

This is asymptotically distributed (for large T) as $N(0, 1)$.

By a similar argument, one can show that the LM test statistic for $H_5^a : \lambda = 0$, in a fixed effects model with MA(1) residual disturbances, is identical to LM_5.

Note also that LM_5 differs from LM_3 only by the fact that the Within residuals $\widetilde{\nu}$ (in LM_5) replace the OLS residuals $\widehat{u}$ (in LM_3). Since the Within transformation wipes out the individual effects whether fixed or random, one can also use (5.43) to test for serial correlation in the random effects models.

The Durbin–Watson Statistic for Panel Data

For the fixed effects model described in (5.39) with ν_{it} following an AR(1) process, Bhargava, Franzini and Narendranathan (1982), hereafter BFN, suggested testing for $H_0 : \rho = 0$ against the alternative that $|\rho| < 1$, using the Durbin–Watson statistic only based on the Within residuals (the $\widetilde{\nu}_{it}$) rather than OLS residuals:

$$d_p = \sum_{i=1}^{N}\sum_{t=2}^{T}(\widetilde{\nu}_{it} - \widetilde{\nu}_{i,t-1})^2 / \sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{\nu}_{it}^2 \tag{5.44}$$

BFN showed that for arbitrary regressors, d_p is a locally most powerful invariant test in the neighborhood of $\rho = 0$. They argued that exact critical values using the Imhof routine are both impractical and unnecessary for panel data since they involve the computation of the nonzero eigenvalues of a large $NT \times NT$ matrix. Instead, BFN show how one can easily compute upper and lower bounds of d_p, and they tabulate the 5% levels for $N = 50, 100, 150, 250, 500, 1000$, $T = 6, 10$ and $k = 1, 3, 5, 7, 9, 11, 13, 15$. BFN remark that d_p would rarely be inconclusive since the bounds will be very tight even for moderate values of N. Also, for very large N, BFN argue that it is not necessary to compute these bounds, but simply test whether d_p is less than two when testing against positive serial correlation.

BFN also suggested the Berenblut–Webb statistic to test $H_0 : \rho = 0$ because it is a locally most powerful invariant test in the neighborhood of $\rho = 1$. This is given by

$$g_p = \sum_{i=1}^{N}\sum_{t=2}^{T}\Delta\widehat{u}_{it}^2 / \sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{\nu}_{it}^2 \tag{5.45}$$

where $\Delta\widehat{u}_{it}$ denotes the OLS residuals obtained from the first-differenced version of the regression equation given in (5.30), and $\widetilde{\nu}_{it}$ denotes the Within residuals. BFN show that g_p and d_p

have similar exact powers when $N = 30$, $T = 10$, $k = 2$, $\alpha = 0.05$ and $\rho = 0.25, 0.40, 0.50$. Also, the two tests are equivalent if N is large.

BFN also suggest a test for random walk residuals, i.e. $H_0 : \rho = 1$ vs the alternative that $|\rho| < 1$. This is based on the statistic

$$R_p = \frac{\Delta\widehat{u}'\Delta\widehat{u}}{(\Delta\widehat{u})'F^*(\Delta\widehat{u})} \tag{5.46}$$

where $\Delta\widehat{u}$ are the differenced OLS residuals used in g_p. $F^* = I_N \otimes F$ with F being a $(T-1) \times (T-1)$ symmetric matrix with elements given by

$$F_{js} = (T - j)s/T \quad \text{if } j \geq s \quad (j, s = 1, \ldots, T - 1)$$

For general regressors, BFN show that $R_p \leq g_p \leq d_p$ where g_p and d_p are now being considered under the random walk null hypothesis. BFN also tabulate 5% lower and upper bounds for R_p and suggest that the bounds for R_p may be used in practice for g_p and d_p. However, when $N \to \infty$, as in typical panels, all three tests are equivalent, $R_p = g_p = d_p$, and BFN recommend only the Durbin–Watson d_p be calculated for testing the random walk hypothesis.

Testing AR(1) Against MA(1) in an Error Component Model

Testing AR(1) against MA(1) has been studied extensively in the time series literature; see King and McAleer (1987) for a Monte Carlo comparison of nonnested, approximate point optimal, as well as LM tests.[3] In fact, King and McAleer (1987) found that the nonnested tests perform poorly in small samples, while King's (1983) point optimal test performs the best. Recently Burke, Godfrey and Termayne (1990) (hereafter BGT) derived a simple test to distinguish between AR(1) and MA(1) processes. Baltagi and Li (1995) proposed two extensions of the BGT test to the error component model. These tests are simple to implement, requiring Within or OLS residuals.

The basic idea of the BGT test is as follows: under the null hypothesis of an AR(1) process, the remainder error term ν_{it} satisfies

$$\text{correl}(\nu_{it}, \nu_{i,t-\tau}) = \rho_\tau = (\rho_1)^\tau \qquad \tau = 1, 2, \ldots \tag{5.47}$$

Therefore, under the null hypothesis

$$\rho_2 - (\rho_1)^2 = 0 \tag{5.48}$$

Under the alternative hypothesis of an MA(1) process on ν_{it}, $\rho_2 = 0$ and hence $\rho_2 - (\rho_1)^2 < 0$. Therefore, BGT recommend a test statistic based on (5.48) using estimates of ρ obtained from OLS residuals. One problem remains. King (1983) suggests that any "good" test should have a size which tends to zero, asymptotically, for $\rho > 0.5$. The test based on (5.48) does not guarantee this property. To remedy this, BGT proposed supplementing (5.48) with the decision to accept the null hypothesis of AR(1) if $\widehat{\rho}_1 > \frac{1}{2} + 1/\sqrt{T}$.

In an error component model, the Within transformation wipes out the individual effects, and one can use the Within residuals of $\widetilde{u}_{it}\ (= \widetilde{\nu}_{it})$ instead of OLS residuals $\widehat{u}_{it}$ to construct the BGT test. Let

$$(\widetilde{\rho}_1)_i = \sum_{t=2}^{T} \widetilde{u}_{it}\widetilde{u}_{i,t-1} / \sum_{t=1}^{T} \widetilde{u}_{it}^2$$

and

$$(\widetilde{\rho}_2)_i = \sum_{t=3}^{T} \widetilde{u}_{it}\widetilde{u}_{i,t-2} / \sum_{t=1}^{T} \widetilde{u}_{it}^2 \quad \text{for } i = 1, \ldots, N$$

The following test statistic, based on (5.48),

$$\widetilde{\gamma}_i = \sqrt{T}[(\widetilde{\rho}_2)_i - (\widetilde{\rho}_1^2)_i]/[1 - (\widetilde{\rho}_2)_i] \tag{5.49}$$

is asymptotically distributed (for large T) as $N(0, 1)$ under the null hypothesis of an AR(1). Using the data on all N individuals, we can construct a generalized BGT test statistic for the error component model

$$\widetilde{\gamma} = \sqrt{N}\left(\sum_{i=1}^{N} \widetilde{\gamma}_i / N\right) = \sqrt{NT} \sum_{i=1}^{N} \left[\frac{(\widetilde{\rho}_2)_i - (\widetilde{\rho}_1^2)_i}{1 - (\widetilde{\rho}_2)_i}\right] / N \tag{5.50}$$

$\widetilde{\gamma}_i$ are independent for different i since the $\widetilde{u}_i$ are independent. Hence $\widetilde{\gamma}$ is also asymptotically distributed (for large T) as $N(0, 1)$ under the null hypothesis of an AR(1) process. The test statistic (5.50) is supplemented by

$$\widetilde{r}_1 = \sum_{i=1}^{N} (\widetilde{r}_1)_i / N \equiv \frac{1}{N} \sum_{i=1}^{N} \left[\sum_{t=2}^{T} \widetilde{u}_{it}\widetilde{u}_{i,t-1} / \sum_{t=1}^{T} \widetilde{u}_{it}^2\right] \tag{5.51}$$

and the Baltagi and Li (1995) proposed BGT_1 test can be summarized as follows:

(1) Use the Within residuals $\widetilde{u}_{it}$ to calculate $\widetilde{\gamma}$ and $\widetilde{r}_1$ from (5.50) and (5.51).
(2) Accept the AR(1) model if $\widetilde{\gamma} > c_\alpha$, or $\widetilde{r}_1 > \frac{1}{2} + 1/\sqrt{T}$, where $\Pr[N(0, 1) \leq c_\alpha] = \alpha$.

The bias in estimating ρ_s ($s = 1, 2$) by using Within residuals is of $O(1/T)$ as $N \to \infty$ (see Nickell, 1981). Therefore, BGT_1 may not perform well for small T. Since for typical labor panels, N is large and T is small, it would be desirable if an alternative simple test can be derived which performs well for large N rather than large T. In the next section we will give such a test.

An Alternative BGT-type Test for Testing AR(1) vs MA(1)

Let the null hypothesis be $H_7 : \nu_{it} = \epsilon_{it} + \lambda\epsilon_{i,t-1}$ and the alternative be $H_7' : \nu_{it} = \rho\nu_{i,t-1} + \epsilon_{it}$, where $\epsilon_{it} \sim N(0, \sigma_\epsilon^2)$. Note that this test differs from the BGT_1 test in that the null hypothesis is MA(1) rather than AR(1). The alternative BGT-type test uses autocorrelation estimates derived from OLS residuals and can be motivated as follows. Let

$$Q_0 = \frac{\sum\sum u_{it}^2}{NT} = u'u/NT$$

and

$$Q_s = \frac{\sum\sum u_{it}u_{i,t-s}}{N(T-s)} = u'(I_N \otimes G_s)u/N(T-s) \quad \text{for } s = 1, \ldots, S$$

where $G_s = \frac{1}{2}$ Toeplitz(ι_s), ι_s is a vector of zeros with the $(s+1)$th element being one. $s = 1, \ldots, S$ with $S \leq (T-1)$ and S is finite.[4] Given the true residuals (the u), and assuming

$$\left[\frac{u'Au}{n} - E\left(\frac{u'Au}{n}\right)\right] \xrightarrow{P} 0$$

where $n = NT$ and A is an arbitrary symmetric matrix, Baltagi and Li (1995) proved the following results, as $N \to \infty$:

(1) For the MA(1) model

$$\begin{aligned} \text{plim } Q_0 &= \sigma_\mu^2 + \sigma_\nu^2 = \sigma_\mu^2 + \sigma_\epsilon^2(1+\lambda^2) \\ \text{plim } Q_1 &= \sigma_\mu^2 + \lambda\sigma_\epsilon^2 \\ \text{plim } Q_s &= \sigma_\mu^2 \quad \text{for } s = 2, \ldots, S \end{aligned} \tag{5.52}$$

(2) For the AR(1) model

$$\begin{aligned} \text{plim } Q_0 &= \sigma_\mu^2 + \sigma_\nu^2 \\ \text{plim } Q_s &= \sigma_\mu^2 + \rho^s\sigma_\nu^2 \quad \text{for } s = 1, \ldots, S \end{aligned} \tag{5.53}$$

see problem 5.17. Baltagi and Li (1995) showed that for large N one can distinguish the AR(1) process from the MA(1) process based on the information obtained from $Q_s - Q_{s+l}$, for $s \geq 2$ and $l \geq 1$. To see this, note that $\text{plim}(Q_s - Q_{s+l}) = 0$ for the MA(1) process and $\text{plim}(Q_s - Q_{s+l}) = \sigma_\nu^2\rho^s(1-\rho^l) > 0$ for the AR(1) process.

Hence, Baltagi and Li (1995) suggest an asymptotic test of H_7 against H_7' based upon

$$\gamma = \sqrt{N/V}(Q_2 - Q_3) \tag{5.54}$$

where $V = 2\,\text{tr}\{[(\sigma_\mu^2 J_T + \sigma_\epsilon^2 V_\lambda)(G_2/(T-2) - G_3/(T-3))]^2\}$. Under some regularity conditions, γ is asymptotically distributed (for large N) as $N(0,1)$ under the null hypothesis of an MA(1) process.[5] In order to calculate V, we note that for the MA(1) process, $\sigma_\nu^2 = \sigma_\epsilon^2(1+\lambda^2)$ and $\sigma_\epsilon^2 V_\lambda = \sigma_\nu^2 I_T + \sigma_\epsilon^2\lambda G$. Therefore we do not need to estimate λ in order to compute the test statistic γ, all we need to get are some consistent estimators for σ_ν^2, $\lambda\sigma_\epsilon^2$ and σ_μ^2. These are obtained as follows:

$$\begin{aligned} \widehat{\sigma}_\nu^2 &= \widehat{Q}_0 - \widehat{Q}_2 \\ \lambda\widehat{\sigma}_\epsilon^2 &= \widehat{Q}_0 - \widehat{Q}_1 \\ \widehat{\sigma}_\mu^2 &= \widehat{Q}_2 \end{aligned}$$

where $\widehat{Q}_s$ are obtained from Q_s by replacing u_{it} by the OLS residuals $\widehat{u}_{it}$. Substituting these consistent estimators into V we get $\widehat{V}$, and the test statistic γ becomes

$$\widehat{\gamma} = \sqrt{N/\widehat{V}}(\widehat{Q}_2 - \widehat{Q}_3) \tag{5.55}$$

where

$$(\widehat{Q}_2 - \widehat{Q}_3) = \sum_{i=1}^{N}\sum_{t=3}^{N} \widehat{u}_{it}\widehat{u}_{i,t-2}/N(T-2) - \sum_{i=1}^{N}\sum_{t=4}^{T} \widehat{u}_{it}\widehat{u}_{i,t-3}/N(T-3)$$

and

$$\widehat{V} = 2\,\text{tr}\{[(\widehat{\sigma}_\mu^2 J_T + \widehat{\sigma}_\nu^2 I_T + \sigma_\epsilon^2\widehat{\lambda}G)/(G_2/(T-2) + G_3/(T-3))]^2\}$$

$\widehat{\gamma}$ is asymptotically distributed (for large N) as $N(0,1)$ under the null hypothesis H_7 and is referred to as the BGT_2 test.

Baltagi and Li (1995) perform extensive Monte Carlo experiments using the regression model set-up considered in Chapter 4. However, the remainder disturbances are now allowed

Table 5.4 Testing for Serial Correlation and Individual Effects

	Null Hypothesis H_0	Alternative Hypothesis H_A	Test Statistics	Asymptotic Distribution under H_0
1a	$H_1^a : \sigma_\mu^2 = 0; \lambda = 0$	σ_μ^2 or $\lambda \neq 0$	LM_1	χ_2^2
1b	$H_1^b : \sigma_\mu^2 = 0; \rho = 0$	σ_μ^2 or $\rho \neq 0$	LM_1	χ_2^2
2	$H_2 : \sigma_\mu^2 = 0$	$\sigma_\mu^2 > 0$	LM_2	$N(0, 1)$
3a	$H_3^a : \lambda = 0$	$\lambda > 0$	LM_3	$N(0, 1)$
3b	$H_3^b : \rho = 0$	$\rho > 0$	LM_3	$N(0, 1)$
4a	$H_4^a : \lambda = 0 (\sigma_\mu^2 > 0)$	$\lambda > 0 (\sigma_\mu^2 > 0)$	LM_4	$N(0, 1)$
4b	$H_4^b : \rho = 0 (\sigma_\mu^2 > 0)$	$\rho > 0 (\sigma_\mu^2 > 0)$	LM_4	$N(0, 1)$
5a	$H_5^a : \lambda = 0 (\mu_i \text{ fixed})$	$\lambda > 0 (\mu_i \text{ fixed})$	LM_5	$N(0, 1)$
5b	$H_5^b : \rho = 0 (\mu_i \text{ fixed})$	$\rho > 0 (\mu_i \text{ fixed})$	LM_5	$N(0, 1)$
6	H_6 : AR(1)	MA(1)	BGT_1	$N(0, 1)$
7	H_7 : MA(1)	AR(1)	BGT_2	$N(0, 1)$

Source: Baltagi and Li (1995). Reproduced by permission of Elsevier Science Publishers B.V. (North Holland).

to follow the AR(1) or MA(1) process. Table 5.4 gives a summary of all tests considered. Their main results can be summarized as follows.

(1) The joint LM_1 test performs well in testing the null of $H_1 : \rho = \sigma_\mu^2 = 0$. Its estimated size is not statistically different from its nominal size. Let $\omega = \sigma_\mu^2/\sigma^2$ denote the proportion of the total variance that is due to individual effects. Baltagi and Li (1995) find that in the presence of large individual effects ($\omega > 0.2$), or high serial correlation ρ (or λ) > 0.2, LM_1 has high power rejecting the null in 99−100% of cases. It only has low power when $\omega = 0$ and ρ (or λ) $= 0.2$, or when $\omega = 0.2$ and ρ (or λ) $= 0$.

(2) The test statistic LM_2 for testing $H_2 : \sigma_\mu^2 = 0$ implicitly assumes that ρ (or λ) $= 0$. When ρ is indeed equal to zero, this test performs well. However, as ρ moves away from zero and increases, this test tends to be biased in favor of rejecting the null. This is because a large serial correlation coefficient (i.e. large ρ) contributes to a large correlation among the individuals in the sample, even though $\sigma_\mu^2 = 0$. For example, when the null is true ($\sigma_\mu^2 = 0$) but $\rho = 0.9$, LM_2 rejects in 100% of cases. Similar results are obtained in case ν_{it} follows an MA(1) process. In general, the presence of positive serial correlation tends to bias the case in favor of finding nonzero individual effects.

(3) Similarly, the LM_3 test for testing $H_3 : \rho = 0$ implicitly assumes $\sigma_\mu^2 = 0$. This test performs well when $\sigma_\mu^2 = 0$. However, as σ_μ^2 increases, the performance of this test deteriorates. For example, when the null is true ($\rho = 0$) but $\omega = 0.9$, LM_3 rejects the null hypothesis in 100% of cases. The large correlation among the μ_i contributes to the rejection of the null hypothesis of no serial correlation. These results strongly indicate that one should not ignore the individual effects when testing for serial correlation.

(4) In contrast to LM_3, both LM_4 and LM_5 take into account the presence of individual effects. For large values of ρ or λ (greater than 0.4), both LM_4 and LM_5 have high power, rejecting the null more than 99% of the time. However, the estimated size of LM_4 is closer to the 5% nominal value than that of LM_5. In addition, Baltagi and Li (1995) show that Bhargava et al.'s (1982) modified Durbin–Watson performs better than LM_5 and is recommended.

(5) The BGT_1 test, which uses Within residuals and tests the null of an AR(1) against the alternative of an MA(1), performs well if $T \geq 60$ and $T > N$. However, when T is small, or T is of moderate size but N is large, BGT_1 will tend to over-reject the null hypothesis. Therefore

BGT_1 is not recommended for these cases. For typical labor panels, N is large and T is small. For these cases, Baltagi and Li (1995) recommend the BGT_2 test, which uses OLS residuals and tests the null of an MA(1) against the alternative of an AR(1). This test performs well when N is large and does not rely on T to achieve its asymptotic distribution. The Monte Carlo results show that BGT_2's performance improves as either N or T increases.

Baltagi and Li (1997) perform Monte Carlo experiments to compare the finite sample relative efficiency of a number of pure and *pre-test* estimators for an error component model with remainder disturbances that are generated by an AR(1) or an MA(1) process. These estimators are: (1) OLS; (2) the Within estimator; (3) conventional GLS which ignores the serial correlation in the remainder disturbances but accounts for the random error components structure – this is denoted by CGLS; (4) GLS assuming random error components with the remainder disturbances following an MA(1) process – this is denoted by GLSM; (5) GLS assuming random error components with the remainder disturbances following an AR(1) process – this is denoted by GLSA; (6) a pre-test estimator which is based on the results of two tests – this is denoted by PRE. The first test is LM_4 which tests for the presence of serial correlation given the existence of random individual effects. If the null is not rejected, this estimator reduces to conventional GLS. In case serial correlation is found, the BGT_2 test is performed to distinguish between the AR(1) and MA(1) process and GLSA or GLSM is performed. (7) A generalized method of moments (GMM) estimator, where the error component structure of the disturbances is ignored and a general variance–covariance matrix is estimated across the time dimension. Finally (8) true GLS, which is denoted by TGLS, is obtained for comparison purposes. In fact, the relative efficiency of each estimator is obtained by dividing its MSE by that of TGLS. It is important to emphasize that all the estimators considered are consistent as long as the explanatory variables and the disturbances are uncorrelated, as $N \to \infty$, with T fixed. The primary concern here is with their small sample properties. The results show that the correct GLS procedure is always the best, but the researcher does not have perfect foresight on which one it is: GLSA for an AR(1) process, or GLSM for an MA(1) process. In this case, the pre-test estimator is a viable alternative given that its performance is a close second to correct GLS whether the true serial correlation process is AR(1) or MA(1).

5.2.8 Extensions

Other extensions include the fixed effects model with AR(1) remainder disturbances considered by Bhargava et al. (1982), and also Kiefer (1980) and Schmidt (1983) who extend the fixed effects model to cover cases with an arbitrary intertemporal covariance matrix. There is an extension to the MA(q) case, by Baltagi and Li (1994) and a treatment of the autoregressive moving average ARMA(p, q) case on the ν_{it}, by MaCurdy (1982) and more recently Galbraith and Zinde-Walsh (1995). For an extension to the two-way model with serially correlated disturbances, see Revankar (1979) who considers the case where the λ_t follow an AR(1) process. Also, Karlsson and Skoglund (2004) for the two-way error component model with an ARMA process on the time-specific effects. They derive the maximum likelihood estimator under normality of the disturbances and propose LM tests for serial correlation and for the choice between the AR(1) and MA(1) specification for the time-specific effects following Baltagi and Li (1995). Magnus and Woodland (1988) generalize this Revankar (1979) model to the multivariate error component model case and derive the corresponding maximum likelihood estimator. Chamberlain (1982, 1984) allows for arbitrary serial correlation and heteroskedastic patterns by viewing each time period as an equation and treating the panel as a multivariate set-up. Testing for serial correlation in a dynamic error component model will be studied in Chapter 8. Li and Hsiao (1998) propose three test statistics in the context of a semiparametric partially linear panel data model

$$y_{it} = x'_{it}\beta + \theta(w_{it}) + u_{it}$$

with u_{it} satisfying $E(u_{it}/w_{it}, x_{it}) = 0$. The functional form of $\theta(.)$ is unknown and $E(u^2_{it}/x_{it}, w_{it})$ is not specified. The first test statistic tests the null of zero first-order serial correlation. The second tests for the presence of higher-order serial correlation and the third tests for the presence of individual effects. The asymptotics are carried out for $N \to \infty$ and fixed T. Monte Carlo experiments are performed to study the finite sample performance of these tests. More recently, Hong and Kao (2004) suggest wavelet-based testing for serial correlation of unknown form in panel data.

NOTES

1. An alternative derivation of this transformation is given by Wansbeek (1992). Bhargava, et al. (1982) give the corresponding transformation for the one-way error component model with fixed effects and first-order autoregressive disturbances.
2. Let $a = (a_1, a_2, \ldots, a_n)'$ denote an arbitrary $n \times 1$ vector, then Toeplitz(a) is an $n \times n$ symmetric matrix generated from the $n \times 1$ vector a with the diagonal elements all equal to a_1, second diagonal elements equal to a_2, etc.
3. Obviously, there are many different ways to construct such a test. For example, we can use $Q_2 + Q_3 - 2Q_4$ instead of $Q_2 - Q_3$ to define the γ test. In this case

$$V = 2\text{tr}\{[(\sigma^2_\mu J_T + \sigma^2_\epsilon V_\lambda)(G_2/(T-2) + G_3/(T-3) - 2G_4/(T-4))]^2\}$$

PROBLEMS

5.1 (a) For the one-way error component model with heteroskedastic μ_i, i.e. $\mu_i \sim (0, w^2_i)$, verify that $\Omega = E(uu')$ is given by (5.1) and (5.2).
(b) Using the Wansbeek and Kapteyn (1982b) trick show that Ω can also be written as in (5.3).

5.2 (a) Using (5.3) and (5.4), verify that $\Omega\Omega^{-1} = I$ and that $\Omega^{-1/2}\Omega^{-1/2} = \Omega^{-1}$.
(b) Show that $y^* = \sigma_\nu\Omega^{-1/2}y$ has a typical element $y^*_{it} = y_{it} - \theta_i\bar{y}_{i.}$ where $\theta_i = 1 - (\sigma_\nu/\tau_i)$ and $\tau^2_i = Tw^2_i + \sigma^2_\nu$ for $i = 1, \ldots, N$.

5.3 Holly and Gardiol (2000) derived a score test for homoskedasticity in a one-way error component model where the alternative model is that the μ_i's are independent and distributed as $N(0, \sigma^2_{\mu_i})$ where $\sigma^2_{\mu_i} = \sigma^2_\mu h_\mu(F_i\theta_2)$. Here, F_i is a vector of p explanatory variables such that $F_i\theta_2$ does not contain a constant term and h_μ is a strictly positive twice-differentiable function satisfying $h_\mu(0) = 1$ with $h'_\mu(0) \neq 0$ and $h''_\mu(0) \neq 0$. Show that the score test statistic for $H_0: \theta_2 = 0$ is equal to one half the explained sum of squares of the OLS regression of $(\hat{s}/\bar{s}) - \iota_N$ against the p regressors in F as in the Breusch and Pagan test for homoskedasticity. Here $\hat{s}_i = \hat{u}'_i\bar{J}_T\hat{u}_i$ and $\bar{s} = \sum^N_{i=1}\hat{s}_i/N$ where $\widehat{u}$ denote the maximum likelihood residuals from the restricted model under $H_0: \theta_2 = 0$.

5.4 (a) For the one-way error component model with heteroskedastic remainder disturbances, i.e. $\nu_{it} \sim (0, w^2_i)$, verify that $\Omega = E(uu')$ is given by (5.5).

(b) Using the Wansbeek and Kapteyn (1982b) trick show that Ω can also be written as in (5.6).

5.5 (a) Using (5.6) and (5.7), verify that $\Omega\Omega^{-1} = I$ and $\Omega^{-1/2}\Omega^{-1/2} = \Omega^{-1}$.

(b) Show that $y^* = \Omega^{-1/2}y$ has a typical element $y_{it}^* = (y_{it} - \theta_i \bar{y}_{i.})/w_i$ where $\theta_i = 1 - (w_i/\tau_i)$ and $\tau_i^2 = T\sigma_\mu^2 + w_i^2$ for $i = 1, \ldots, N$.

5.6 (a) For the one-way error component model with remainder disturbances ν_{it} following a stationary AR(1) process as in (5.8), verify that $\Omega^* = E(u^* u^{*\prime})$ is that given by (5.11).

(b) Using the Wansbeek and Kapteyn (1982b) trick, show that Ω^* can be written as in (5.12).

5.7 (a) Using (5.12) and (5.13), verify that $\Omega^*\Omega^{*-1} = I$ and $\Omega^{*-1/2}\Omega^{*-1/2} = \Omega^{*-1}$.

(b) Show that $y^{**} = \sigma_\epsilon \Omega^{*-1/2} y^*$ has a typical element given by (5.14).

(c) Show that for $\rho = 0$, (5.14) reduces to $(y_{it} - \theta \bar{y}_{i.})$.

(d) Show that for $\sigma_\mu^2 = 0$, (5.14) reduces to y_{it}^*.

5.8 Prove that $\widehat{\sigma}_\epsilon^2$ and $\widehat{\sigma}_\alpha^2$ given by (5.15) are unbiased for σ_ϵ^2 and σ_α^2, respectively.

5.9 (a) For the one-way error component model with remainder disturbances ν_{it} following a stationary AR(2) process as in (5.16), verify that $\Omega^* = E(u^* u^{*\prime})$ is that given by (5.18).

(b) Show that $y^{**} = \sigma_\epsilon \Omega^{*-1/2} y^*$ has a typical element given by (5.19).

5.10 For the one-way error component model with remainder disturbances ν_{it} following a specialized AR(4) process $\nu_{it} = \rho\nu_{i,t-4} + \epsilon_{it}$ with $| \rho |< 1$ and $\epsilon_{it} \sim \text{IIN}(0, \sigma_\epsilon^2)$, verify that $y^{**} = \sigma_\epsilon \Omega^{-1/2} y^*$ is given by (5.21).

5.11 For the one-way error component model with remainder disturbances ν_{it} following an MA(1) process given by (5.22), verify that $y^{**} = \sigma_\epsilon \Omega^{-1/2} y^*$ is given by (5.24).

5.12 For the BLU predictor of $y_{i,T+1}$ given in (5.25), show that when ν_{it} follows:

(a) the AR(1) process, the GLS predictor is corrected by the term in (5.26);

(b) the AR(2) process, the GLS predictor is corrected by the term given in (5.27);

(c) the specialized AR(4) process, the GLS predictor is corrected by the term given in (5.28);

(d) the MA(1) process, the GLS predictor is corrected by the term given in (5.29).

5.13 Using (4.17) and (4.19), verify (5.34) and (5.35) and derive the LM_1 statistic given in (5.36).

5.14 (a) Verify that $(\partial V_\rho/\partial\rho)_{\rho=0} = G = (\partial V_\lambda/\partial\lambda)_{\lambda=0}$ where G is the bidiagonal matrix with bidiagonal elements all equal to one.

(b) Using this result verify that the joint LM statistic given in (5.36) is the same whether the residual disturbances follow an AR(1) or an MA(1) process, i.e., the joint LM test statistic for $H_1^a : \sigma_\mu^2 = 0; \lambda = 0$ is the same as that for $H_1^b : \sigma_\mu^2 = 0; \rho = 0$.

5.15 For $H_4^b : \rho = 0$ (given $\sigma_\mu^2 > 0$):

(a) Derive the score, the information matrix and the LM statistic for H_4^b.

(b) Verify that for $H_4^a : \lambda = 0$ (given $\sigma_\mu^2 > 0$) one obtains the same LM statistic as in part (a).

5.16 For $H_5^b : \rho = 0$ (given the μ_i are fixed):

(a) Verify that the likelihood is given by (5.40) and derive the MLE of the μ_i.

(b) Using (5.34) and (5.35), verify that the LM statistic for H_5^b is given by (5.42).

(c) Verify that for $H_5^a : \lambda = 0$ (given the μ_i are fixed) one obtains the same LM statistic as in (5.42).

5.17 (a) Verify (5.52) for the MA(1) model. *Hint*: Use the fact that lim $E(u'u)/(NT) =$ lim $\text{tr}(\Omega)/(NT)$ for deriving plim Q_0. Similarly, use the fact that

$$\lim E(u'(I_N \otimes G_1)u)/N(T-1) = \lim \text{tr}[\Omega(I_N \otimes G_1)]/N(T-1)$$

for deriving plim Q_1. Also,

$$\lim E(u'(I_N \otimes G_s)u)/N(T-s) = \lim \text{tr}[\Omega(I_N \otimes G_s)]/N(T-s)$$

for deriving plim Q_s for $s = 2, \ldots, S$.

(b) Verify (5.53) for the AR(1) model.

5.18 Using the Monte Carlo set-up in Baltagi and Li (1995), study the performance of the tests proposed in Table 5.4.

5.19 For the Grunfeld data:

(a) Perform the tests described in Table 5.4.

(b) Using the unbalanced patterns described in table 1 of Baltagi and Wu (1999), replicate the Baltagi–Wu LBI and Bhargava et al. (1982) Durbin–Watson test statistics reported in that table. This can easily be done using the (xtregar,re lbi) command in Stata.

5.20 For the gasoline data given on the Wiley web site, perform the tests described in Table 5.4.

5.21 For the public capital data, given on the Wiley web site, perform the tests described in Table 5.4.

6 Seemingly Unrelated Regressions with Error Components

6.1 THE ONE-WAY MODEL

In several instances in economics, one needs to estimate a set of equations. This could be a set of demand equations, across different sectors, industries or regions. Other examples include the estimation of a translog cost function along with the corresponding cost share equations. In these cases, Zellner's (1962) seemingly unrelated regressions (SUR) approach is popular since it captures the efficiency due to the correlation of the disturbances across equations. Applications of the SUR procedure with time-series or cross-section data are too numerous to cite. In this chapter, we focus on the estimation of a set of SUR equations with panel data.

Avery (1977) seems to be the first to consider the SUR model with error component disturbances. In this case, we have a set of M equations

$$y_j = Z_j\delta_j + u_j \quad j = 1, \ldots, M \tag{6.1}$$

where y_j is $NT \times 1$, Z_j is $NT \times k'_j$, $\delta'_j = (\alpha_j, \beta'_j)$, β_j is $k_j \times 1$ and $k'_j = k_j + 1$ with

$$u_j = Z_\mu \mu_j + \nu_j \quad j = 1, \ldots, M \tag{6.2}$$

where $Z_\mu = (I_N \otimes \iota_T)$ and $\mu'_j = (\mu_{1j}, \mu_{2j}, \ldots, \mu_{Nj})$ and $\nu'_j = (\nu_{11j}, \ldots, \nu_{1Tj}, \ldots, \nu_{N1j}, \ldots, \nu_{NTj})$ are random vectors with zero means and covariance matrix

$$E\begin{pmatrix}\mu_j \\ \nu_j\end{pmatrix}(\mu'_l, \nu'_l) = \begin{bmatrix}\sigma^2_{\mu_{jl}} I_N & 0 \\ 0 & \sigma^2_{\nu_{jl}} I_{NT}\end{bmatrix} \tag{6.3}$$

for $j, l = 1, 2, \ldots, M$. This can be justified as follows: $\mu \sim (0, \Sigma_\mu \otimes I_N)$ and $\nu \sim (0, \Sigma_\nu \otimes I_{NT})$ where $\mu' = (\mu'_1, \mu'_2, \ldots, \mu'_M)$, $\nu' = (\nu'_1, \nu'_2, \ldots, \nu'_M)$, $\Sigma_\mu = [\sigma^2_{\mu_{jl}}]$ and $\Sigma_\nu = [\sigma^2_{\nu_{jl}}]$ for $j, l = 1, 2, \ldots, M$. In other words, *each* error component follows the same standard Zellner (1962) SUR assumptions imposed on classical disturbances. Using (6.2), it follows that

$$\Omega_{jl} = E(u_j u'_l) = \sigma^2_{\mu_{jl}}(I_N \otimes J_T) + \sigma^2_{\nu_{jl}}(I_N \otimes I_T) \tag{6.4}$$

In this case, the covariance matrix between the disturbances of different equations has the same one-way error component form. Except now, there are additional *cross-equations* variance components to be estimated. The variance–covariance matrix for the set of M equations is given by

$$\Omega = E(uu') = \Sigma_\mu \otimes (I_N \otimes J_T) + \Sigma_\nu \otimes (I_N \otimes I_T) \tag{6.5}$$

where $u' = (u'_1, u'_2, \ldots, u'_M)$ is a $1 \times MNT$ vector of disturbances with u_j defined in (6.2) for $j = 1, 2, \ldots, M$. $\Sigma_\mu = [\sigma^2_{\mu_{jl}}]$ and $\Sigma_\nu = [\sigma^2_{\nu_{jl}}]$ are both $M \times M$ matrices. Replacing J_T by

$T\bar{J}_T$ and I_T by $E_T + \bar{J}_T$, and collecting terms one gets

$$\begin{aligned}\Omega &= (T\Sigma_\mu + \Sigma_\nu) \otimes (I_N \otimes \bar{J}_T) + \Sigma_\nu \otimes (I_N \otimes E_T) \\ &= \Sigma_1 \otimes P + \Sigma_\nu \otimes Q\end{aligned} \tag{6.6}$$

where $\Sigma_1 = T\Sigma_\mu + \Sigma_\nu$. Also, $P = I_N \otimes \bar{J}_T$ and $Q = I_{NT} - P$ were defined below (2.4). (6.6) is the spectral decomposition of Ω derived by Baltagi (1980), which means that

$$\Omega^r = \Sigma_1^r \otimes P + \Sigma_\nu^r \otimes Q \tag{6.7}$$

where r is an arbitrary scalar (see also Magnus, 1982). For $r = -1$, one gets the inverse Ω^{-1} and for $r = -\frac{1}{2}$ one gets

$$\Omega^{-1/2} = \Sigma_1^{-1/2} \otimes P + \Sigma_\nu^{-1/2} \otimes Q \tag{6.8}$$

Kinal and Lahiri (1990) suggest obtaining the Cholesky decomposition of Σ_ν and Σ_1 in (6.8) to reduce the computation and simplify the transformation of the system.

One can estimate Σ_ν by $\widehat{\Sigma}_\nu = U'QU/N(T-1)$ and Σ_1 by $\widehat{\Sigma}_1 = U'PU/N$ where $U = [u_1, \ldots, u_M]$ is the $NT \times M$ matrix of disturbances for all M equations. Problem 6.7 asks the reader to verify that knowing U, $\widehat{\Sigma}_\nu$ and $\widehat{\Sigma}_1$ are unbiased estimates of Σ_ν and Σ_1, respectively. For feasible GLS estimates of the variance components, Avery (1977), following Wallace and Hussain (1969) in the single equation case, recommends replacing U by OLS residuals, while Baltagi (1980), following Amemiya's (1971) suggestion for the single equation case, recommends replacing U by Within-type residuals.

For this model, a block-diagonal Ω makes GLS on the whole system equivalent to GLS on each equation separately, see problem 6.3. However, when the same X appear in each equation, GLS on the whole system is not equivalent to GLS on each equation separately (see Avery, 1977). As in the single equation case, if N and $T \to \infty$, then the Within estimator of this system is asymptotically efficient and has the same asymptotic variance–covariance matrix as the GLS estimator. In fact, Prucha (1984) shows that as long as Σ_ν is estimated consistently and the estimate of Σ_μ has a finite positive definite limit then the corresponding feasible SUR–GLS estimator is asymptotically efficient. This implies the existence of a large family of asymptotically efficient estimators of the regression coefficients.

6.2 THE TWO-WAY MODEL

It is easy to extend the analysis to a two-way error component structure across the system of equations. In this case (6.2) becomes

$$u_j = Z_\mu \mu_j + Z_\lambda \lambda_j + \nu_j \quad j = 1, \ldots, M \tag{6.9}$$

where $\lambda_j' = (\lambda_{1j}, \ldots, \lambda_{Tj})$ is a random vector with zero mean and covariance matrix given by the following:

$$E\begin{pmatrix}\mu_j \\ \lambda_j \\ \nu_j\end{pmatrix}(\mu_l', \lambda_l', \nu_l') = \begin{bmatrix}\sigma^2_{\mu_{jl}} I_N & 0 & 0 \\ 0 & \sigma^2_{\lambda_{jl}} I_T & 0 \\ 0 & 0 & \sigma^2_{\nu_{jl}} I_{NT}\end{bmatrix} \tag{6.10}$$

for $j, l = 1, 2, \ldots, M$. In this case, $\lambda \sim (0, \Sigma_\lambda \otimes I_T)$ where $\lambda' = (\lambda_1, \lambda_2, \ldots, \lambda_T)$ and $\Sigma_\lambda = [\sigma^2_{\lambda_{jl}}]$ is $M \times M$. Like μ and ν, the λ follow a standard Zellner SUR-type assumption.

Therefore

$$\Omega_{jl} = E(u_j u_l') = \sigma^2_{\mu_{jl}}(I_N \otimes J_T) + \sigma^2_{\lambda_{jl}}(J_N \otimes I_T) + \sigma^2_{\nu_{jl}}(I_N \otimes I_T) \tag{6.11}$$

As in the one-way SUR model, the covariance between the disturbances of different equations has the same two-way error component form. Except now, there are additional cross-equations variance components to be estimated. The variance–covariance matrix of the system of M equations is given by

$$\Omega = E(uu') = \Sigma_\mu \otimes (I_N \otimes J_T) + \Sigma_\lambda \otimes (J_N \otimes I_T) + \Sigma_\nu \otimes (I_N \otimes I_T) \tag{6.12}$$

where $u' = (u_1', u_2', \ldots, u_M')$ with u_j defined in (6.9). Using the Wansbeek and Kapteyn (1982b) trick one gets (see problem 6.5):

$$\Omega = \sum_{i=1}^{4} \Lambda_i \otimes Q_i \tag{6.13}$$

where $\Lambda_1 = \Sigma_\nu$, $\Lambda_2 = T\Sigma_\mu + \Sigma_\nu$, $\Lambda_3 = N\Sigma_\lambda + \Sigma_\nu$ and $\Lambda_4 = T\Sigma_\mu + N\Sigma_\lambda + \Sigma_\nu$, with Q_i defined below (3.13). This is the spectral decomposition of Ω (see Baltagi, 1980), with

$$\Omega^r = \sum_{i=1}^{4} \Lambda_i^r \otimes Q_i \tag{6.14}$$

for r an arbitrary scalar. When $r = -1$ one gets the inverse Ω^{-1} and when $r = -\frac{1}{2}$ one gets

$$\Omega^{-1/2} = \sum_{i=1}^{4} \Lambda_i^{-1/2} \otimes Q_i \tag{6.15}$$

Once again, the Cholesky decompositions of the Λ_i can be obtained in (6.15) to reduce the computation and simplify the transformation of the system (see Kinal and Lahiri, 1990). Knowing the true disturbances U, quadratic unbiased estimates of the variance components are obtained from

$$\widehat{\Sigma}_\nu = \frac{U'Q_1U}{(N-1)(T-1)}, \quad \widehat{\Lambda}_2 = \frac{U'Q_2U}{(N-1)} \quad \text{and} \quad \widehat{\Lambda}_3 = \frac{U'Q_3U}{(T-1)} \tag{6.16}$$

see problem 6.7. Feasible estimates of (6.16) are obtained by replacing U by OLS residuals or Within-type residuals. One should check for positive definite estimates of Σ_μ and Σ_λ before proceeding. The Within estimator has the same asymptotic variance–covariance matrix as GLS when N and $T \to \infty$. Also, as long as the estimate of Σ_ν is consistent and the estimates of Σ_μ and Σ_λ have a finite positive definite probability limit, the corresponding feasible SUR–GLS estimate of the regression coefficients is asymptotically efficient.

6.3 APPLICATIONS AND EXTENSIONS

Verbon (1980) applies the SUR procedure with one-way error components to a set of four labor demand equations, using data from the Netherlands on 18 industries over 10 semiannual periods covering the period 1972–79. Verbon (1980) extends the above error component specification to allow for heteroskedasticity in the individual effects modeled as a simple function of p time-invariant variables. He applies a Breusch and Pagan (1979) LM test to check for the existence of heteroskedasticity.

Beierlein, Dunn and McConnon (1981) estimated the demand for electricity and natural gas in the northeastern United States using a SUR model with two-way error component disturbances. The data were collected for nine states comprising the Census Bureau's northeastern region of the USA for the period 1967–77. Six equations were considered corresponding to the various sectors considered. These were residential gas, residential electric, commercial gas, commercial electric, industrial gas and industrial electric. Comparison of the error components SUR estimates with those obtained from OLS and single equation error component procedures showed substantial improvement in the estimates and a sizable reduction in the empirical standard errors.

Brown et al. (1983) apply the SUR model with error components to study the size-related anomalies in stock returns. Previous empirical evidence has shown that small firms tend to yield returns greater than those predicted by the capital asset pricing model. Brown et al. (1983) used a panel of 566 firms observed quarterly over the period June 1967 to December 1975. They find that size effects are sensitive to the time period studied.

Howrey and Varian (1984) apply the SUR with one-way error component disturbances to the estimation of a system of demand equations for electricity by time of day. Their data are based on the records of 60 households whose electricity usage was recorded over a five-month period in 1976 by the Arizona Public Service Company. Using these panel data, the authors calculate the fraction of the population which would prefer such pricing policies to flat rate pricing.

Magnus (1982) derives the maximum likelihood estimator for the linear and nonlinear multivariate error component model under various assumptions on the errors. Sickles (1985) applies Magnus's multivariate nonlinear error components analysis to model the technology and specific factor productivity growth in the US airline industry.

Wan, Griffiths and Anderson (1992) apply a SUR model with two-way error component disturbances that are heteroskedastic to estimate the rice, maize and wheat production in China. These production functions allow for positive or negative marginal risks of output. The panel data cover 28 regions of China over the 1980–83 period. Their findings indicate that increases in chemical fertilizer and sown area generally increase the output variance. However, organic fertilizer and irrigation help stabilize Chinese cereal production.

Baltagi et al. (1995) estimate a SUR model consisting of a translog variable cost function and its corresponding input share equations for labor, fuel and material. The panel data consists of 24 US airlines over the period 1971–86. Firm and time dummies are included in the variable cost equation, and symmetry as well as adding-up restrictions on the share equations are imposed. A general Solow-type index of technical change is estimated and its determinants are in turn analyzed. One of the main findings of this study is that despite the slowing of productivity growth in the 1980s, deregulation does appear to have stimulated technical change due to more efficient route structure.

Biorn (2004) considers the problem of estimating a system of regression equations with random individual effects from unbalanced panel data. The unbalancedness is due to random attrition. Biorn (2004) shows that GLS on this system can be interpreted as a matrix-weighted average of group-specific GLS estimators with weights equal to the inverse of their respective variance–covariance matrices. The grouping of individuals in the panel is according to the number of times they are observed (not necessarily the same period and not necessarily consecutive periods). Biorn also derives a stepwise algorithm for obtaining the MLE under normality of the disturbances.

PROBLEMS

6.1 Using the one-way error component structure on the disturbances of the jth equation given in (6.2) and (6.3), verify that Ω_{jl}, the variance–covariance matrix between the jth and lth equation disturbances, is given by (6.4).

6.2 Using (6.6) and (6.7), verify that $\Omega\Omega^{-1} = I$ and $\Omega^{-1/2}\Omega^{-1/2} = \Omega^{-1}$.

6.3 Consider a set of two equations with one-way error components disturbances.
  (a) Show that if the variance–covariance matrix between the equations is block-diagonal, then GLS on the system is equivalent to GLS on each equation separately (see Avery, 1977; Baltagi, 1980).
  (b) Show that if the explanatory variables are the same across the two equations, GLS on the system does not necessarily revert to GLS on each equation separately (see Avery, 1977; Baltagi, 1980).
  (c) Does your answer to parts (a) and (b) change if the disturbances followed a two-way error component model?

6.4 Using the two-way error component structure on the disturbances of the jth equation given in (6.9) and (6.10), verify that Ω_{jl}, the variance–covariance matrix between the jth and lth equation disturbances, is given by (6.11).

6.5 Using the form of Ω given in (6.12) and the Wansbeek and Kapteyn (1982b) trick, verify (6.13).

6.6 Using (6.13) and (6.14), verify that $\Omega\Omega^{-1} = I$ and $\Omega^{-1/2}\Omega^{-1/2} = \Omega^{-1}$.

6.7 (a) Using (6.6), verify that $\widehat{\Sigma}_\nu = U'QU/N(T-1)$ and $\widehat{\Sigma}_1 = U'PU/N$ yield unbiased estimates of Σ_ν and Σ_1, respectively.
  (b) Using (6.13), verify that (6.16) results in unbiased estimates of Σ_ν, Λ_2 and Λ_3, respectively.

7 Simultaneous Equations with Error Components

7.1 SINGLE EQUATION ESTIMATION

Endogeneity of the right-hand regressors is a serious problem in econometrics. By endogeneity we mean the correlation of the right-hand side regressors and the disturbances. This may be due to the omission of relevant variables, measurement error, sample selectivity, self-selection or other reasons. Endogeneity causes inconsistency of the usual OLS estimates and requires instrumental variable (IV) methods like two-stage least squares (2SLS) to obtain consistent parameter estimates. The applied literature is full of examples of endogeneity: demand and supply equations for labor, money, goods and commodities to mention a few. Also, behavioral relationships like consumption, production, investment, import and export are just a few more examples in economics where endogeneity is suspected. We assume that the reader is familiar with the identification and estimation of a single equation and a system of simultaneous equations. In this chapter we focus on the estimation of simultaneous equations using panel data.

Consider the following first structural equation of a simultaneous equation model:

$$y_1 = Z_1\delta_1 + u_1 \tag{7.1}$$

where $Z_1 = [Y_1, X_1]$ and $\delta_1' = (\gamma_1', \beta_1')$. As in the standard simultaneous equation literature, Y_1 is the set of g_1 right-hand side endogenous variables, and X_1 is the set of k_1 included exogenous variables. Let $X = [X_1, X_2]$ be the set of all exogenous variables in the system. This equation is identified with k_2, the number of excluded exogenous variables from the first equation (X_2) being larger than or equal to g_1.

Throughout this chapter we will focus on the one-way error component model

$$u_1 = Z_\mu\mu_1 + \nu_1 \tag{7.2}$$

where $Z_\mu = (I_N \otimes \iota_T)$ and $\mu_1' = (\mu_{11}, \ldots, \mu_{N1})$ and $\nu_1' = (\nu_{111}, \ldots, \nu_{NT1})$ are random vectors with zero means and covariance matrix

$$E\begin{pmatrix}\mu_1\\ \nu_1\end{pmatrix}(\mu_1', \nu_1') = \begin{bmatrix}\sigma^2_{\mu_{11}} I_N & 0\\ 0 & \sigma^2_{\nu_{11}} I_{NT}\end{bmatrix} \tag{7.3}$$

This differs from the SUR set-up in Chapter 6 only in the fact that there are right-hand side endogenous variables in Z_1.[1] In this case,

$$E(u_1u_1') = \Omega_{11} = \sigma^2_{\nu_{11}} I_{NT} + \sigma^2_{\mu_{11}}(I_N \otimes J_T) \tag{7.4}$$

In other words, the first structural equation has the typical variance–covariance matrix of a one-way error component model described in Chapter 2. The only difference is that now a double subscript (1, 1) is attached to the variance components to specify that this is the first equation. One can transform (7.1) by $Q = I_{NT} - P$ with $P = I_N \otimes \bar{J}_T$, to get

$$Qy_1 = QZ_1\delta_1 + Qu_1 \tag{7.5}$$

Let $\widetilde{y}_1 = Qy_1$ and $\widetilde{Z}_1 = QZ_1$. Performing 2SLS on (7.5) with $\widetilde{X} = QX$ as the set of instruments, one gets Within 2SLS (or fixed effects 2SLS)

$$\widetilde{\delta}_{1,\text{W2SLS}} = (\widetilde{Z}_1' P_{\widetilde{X}} \widetilde{Z}_1)^{-1} \widetilde{Z}_1' P_{\widetilde{X}} \widetilde{y}_1 \tag{7.6}$$

with $\text{var}(\widehat{\delta}_{1,\text{W2SLS}}) = \sigma^2_{\nu_{11}}(\widetilde{Z}_1' P_{\widetilde{X}} \widetilde{Z}_1)^{-1}$. This can be obtained using the Stata command (xtivreg, fe) specifying the endogenous variables Y_1 and the set of instruments X. Within 2SLS can also be obtained as GLS on

$$\widetilde{X}'\widetilde{y}_1 = \widetilde{X}'\widetilde{Z}_1\delta_1 + \widetilde{X}'\widetilde{u}_1 \tag{7.7}$$

see problem 7.1. Similarly, if we let $\bar{y}_1 = Py_1$ and $\bar{Z}_1 = PZ_1$, we can transform (7.1) by P and perform 2SLS with $\bar{X} = PX$ as the set of instruments. In this case, we get the Between 2SLS estimator of δ_1

$$\widehat{\delta}_{1,\text{B2SLS}} = (\bar{Z}_1' P_{\bar{X}} \bar{Z}_1)^{-1} \bar{Z}_1' P_{\bar{X}} \bar{y}_1 \tag{7.8}$$

with $\text{var}(\widehat{\delta}_{1,\text{B2SLS}}) = \sigma^2_{1_{11}}(\bar{Z}_1' P_{\bar{X}} \bar{Z}_1)^{-1}$ where $\sigma^2_{1_{11}} = T\sigma^2_{\mu_{11}} + \sigma^2_{\nu_{11}}$. This can also be obtained using the Stata command (xtivreg,be) specifying the endogenous variables Y_1 and the set of instruments X. Between 2SLS can also be obtained as GLS on

$$\bar{X}'\bar{y}_1 = \bar{X}'\bar{Z}_1\delta_1 + \bar{X}'\bar{u}_1 \tag{7.9}$$

Stacking these two transformed equations in (7.7) and (7.9) as a system, as in (2.28) and noting that δ_1 is the same for these two transformed equations, one gets

$$\begin{pmatrix} \widetilde{X}'\widetilde{y}_1 \\ \bar{X}'\bar{y}_1 \end{pmatrix} = \begin{pmatrix} \widetilde{X}'\widetilde{Z}_1 \\ \bar{X}'\bar{Z}_1 \end{pmatrix} \delta_1 + \begin{pmatrix} \widetilde{X}'\widetilde{u}_1 \\ \bar{X}'\bar{u}_1 \end{pmatrix} \tag{7.10}$$

where

$$E\begin{pmatrix} \widetilde{X}'\widetilde{u}_1 \\ \bar{X}'\bar{u}_1 \end{pmatrix} = 0 \quad \text{and} \quad \text{var}\begin{pmatrix} \widetilde{X}'\widetilde{u}_1 \\ \bar{X}'\bar{u}_1 \end{pmatrix} = \begin{bmatrix} \sigma^2_{\nu_{11}}\widetilde{X}'\widetilde{X} & 0 \\ 0 & \sigma^2_{1_{11}}\bar{X}'\bar{X} \end{bmatrix}$$

Performing GLS on (7.10) yields the error component two-stage least squares (EC2SLS) estimator of δ_1 derived by Baltagi (1981b):

$$\widehat{\delta}_{1,\text{EC2SLS}} = \left[\frac{\widetilde{Z}_1' P_{\widetilde{X}} \widetilde{Z}_1}{\sigma^2_{\nu_{11}}} + \frac{\bar{Z}_1' P_{\bar{X}} \bar{Z}_1}{\sigma^2_{1_{11}}} \right]^{-1} \left[\frac{\widetilde{Z}_1' P_{\widetilde{X}} \widetilde{y}_1}{\sigma^2_{\nu_{11}}} + \frac{\bar{Z}_1' P_{\bar{X}} \bar{y}_1}{\sigma^2_{1_{11}}} \right] \tag{7.11}$$

with $\text{var}(\widehat{\delta}_{1,\text{EC2SLS}})$ given by the first inverted bracket in (7.11), see problem 7.2. Note that $\widehat{\delta}_{1,\text{EC2SLS}}$ can also be written as a matrix-weighted average of $\widetilde{\delta}_{1,\text{W2SLS}}$ and $\widehat{\delta}_{1,\text{B2SLS}}$ with the weights depending on their respective variance–covariance matrices:

$$\widehat{\delta}_{1,\text{EC2SLS}} = W_1 \widehat{\delta}_{1,\text{W2SLS}} + W_2 \widehat{\delta}_{1,\text{B2SLS}} \tag{7.12}$$

with

$$W_1 = \left[\frac{\widetilde{Z}_1' P_{\widetilde{X}} \widetilde{Z}_1}{\sigma^2_{\nu_{11}}} + \frac{\bar{Z}_1' P_{\bar{X}} \bar{Z}_1}{\sigma^2_{1_{11}}} \right]^{-1} \left[\frac{\widetilde{Z}_1' P_{\widetilde{X}} \widetilde{Z}_1}{\sigma^2_{\nu_{11}}} \right]$$

and

$$W_2 = \left[\frac{\widetilde{Z}_1' P_{\widetilde{X}} \widetilde{Z}_1}{\sigma^2_{\nu_{11}}} + \frac{\bar{Z}_1' P_{\bar{X}} \bar{Z}_1}{\sigma^2_{1_{11}}}\right]^{-1} \left[\frac{\bar{Z}_1' P_{\bar{X}} \bar{Z}_1}{\sigma^2_{1_{11}}}\right]$$

Consistent estimates of $\sigma^2_{\nu_{11}}$ and $\sigma^2_{1_{11}}$ can be obtained from W2SLS and B2SLS residuals, respectively. In fact

$$\widehat{\sigma}^2_{\nu_{11}} = (y_1 - Z_1\widetilde{\delta}_{1,\text{W2SLS}})' Q (y_1 - Z_1\widetilde{\delta}_{1,\text{W2SLS}})/N(T-1) \tag{7.13}$$

$$\widehat{\sigma}^2_{1_{11}} = (y_1 - Z_1\widehat{\delta}_{1,\text{B2SLS}})' P (y_1 - Z_1\widehat{\delta}_{1,\text{B2SLS}})/N \tag{7.14}$$

Substituting these variance components estimates in (7.11) one gets a feasible estimate of EC2SLS. Note that unlike the usual 2SLS procedure, EC2SLS requires estimates of the variance components. One can correct for degrees of freedom in (7.13) and (7.14) especially for small samples, but the panel is assumed to have large N. Also, one should check that $\widehat{\sigma}^2_{\mu_{11}} = (\widehat{\sigma}^2_{1_{11}} - \widehat{\sigma}^2_{\nu_{11}})/T$ is positive.

Alternatively, one can premultiply (7.1) by $\Omega_{11}^{-1/2}$ where Ω_{11} is given in (7.4), to get

$$y_1^* = Z_1^*\delta_1 + u_1^* \tag{7.15}$$

with $y_1^* = \Omega_{11}^{-1/2} y_1$, $Z_1^* = \Omega_{11}^{-1/2} Z_1$ and $u_1^* = \Omega_{11}^{-1/2} u_1$. In this case, $\Omega_{11}^{-1/2}$ is given by (2.20) with the additional subscripts (1, 1) for the variance components, i.e.

$$\Omega_{11}^{-1/2} = (P/\sigma_{1_{11}}) + (Q/\sigma_{\nu_{11}}) \tag{7.16}$$

Therefore, the typical element of y_1^* is $y^*_{1_{it}} = (y_{1_{it}} - \theta_1 \bar{y}_{1_{i.}})/\sigma_{\nu_{11}}$ where $\theta_1 = 1 - (\sigma_{\nu_{11}}/\sigma_{1_{11}})$ and $\bar{y}_{1_{i.}} = \sum_{t=1}^T y_{1_{it}}/T$.

Given a set of instruments A, then 2SLS on (7.15) using A gives

$$\widehat{\delta}_{1,\text{2SLS}} = (Z_1^{*\prime} P_A Z_1^*)^{-1} Z_1^{*\prime} P_A y_1^* \tag{7.17}$$

where $P_A = A(A'A)^{-1}A'$. Using the results in White (1986), the optimal set of instrumental variables in (7.15) is

$$X^* = \Omega_{11}^{-1/2} X = \frac{QX}{\sigma_{\nu_{11}}} + \frac{PX}{\sigma_{1_{11}}} = \frac{\widetilde{X}}{\sigma_{\nu_{11}}} + \frac{\bar{X}}{\sigma_{1_{11}}}$$

Using $A = X^*$, one gets the Balestra and Varadharajan-Krishnakumar (1987) generalized two-stage least squares (G2SLS):

$$\widehat{\delta}_{1,\text{G2SLS}} = (Z_1^{*\prime} P_{X^*} Z_1^*)^{-1} Z_1^{*\prime} P_{X^*} y_1^* \tag{7.18}$$

Cornwell, Schmidt and Wyhowski (1992) showed that Baltagi's (1981b) EC2SLS can be obtained from (7.17), i.e. using a 2SLS package on the transformed equation (7.15) with the set of instruments $A = [QX, PX] = [\widetilde{X}, \bar{X}]$. In fact, QX is orthogonal to PX and $P_A = P_{\widetilde{X}} + P_{\bar{X}}$. This also means that

$$\begin{aligned} P_A Z_1^* &= (P_{\widetilde{X}} + P_{\bar{X}})[\Omega_{11}^{-1/2} Z_1] \qquad (7.19) \\ &= (P_{\widetilde{X}} + P_{\bar{X}})\left[\frac{Q}{\sigma_{\nu_{11}}} + \frac{P}{\sigma_{1_{11}}}\right] Z_1 = \frac{P_{\widetilde{X}}\widetilde{Z}_1}{\sigma_{\nu_{11}}} + \frac{P_{\bar{X}}\bar{Z}_1}{\sigma_{1_{11}}} \end{aligned}$$

with

$$Z_1^{*\prime} P_A Z_1^* = \left(\frac{\widetilde{Z}_1' P_{\widetilde{X}} \widetilde{Z}_1}{\sigma_{\nu_{11}}^2} + \frac{\bar{Z}_1' P_{\bar{X}} \bar{Z}_1}{\sigma_{1_{11}}^2} \right)$$

and

$$Z_1^{*\prime} P_A y_1^* = \left(\frac{\widetilde{Z}_1' P_{\widetilde{X}} \widetilde{y}_1}{\sigma_{\nu_{11}}^2} + \frac{\bar{Z}_1' P_{\bar{X}} \bar{y}_1}{\sigma_{1_{11}}^2} \right)$$

Therefore, $\widehat{\delta}_{1,\text{EC2SLS}}$ given by (7.11) is the same as (7.17) with $A = [\widetilde{X}, \bar{X}]$.

So, how is Baltagi's (1981b) EC2SLS given by (7.11) different from the Balestra and Varadharajan-Krishnakumar (1987) G2SLS given by (7.18)? It should be clear to the reader that the set of instruments used by Baltagi (1981b), i.e. $A = [\widetilde{X}, \bar{X}]$, spans the set of instruments used by Balestra and Varadharajan-Krishnakumar (1987), i.e. $X^* = [\widetilde{X}/\sigma_{\nu_{11}} + \bar{X}/\sigma_{1_{11}}]$. In fact, one can show that $A = [\widetilde{X}, \bar{X}]$, $B = [X^*, \widetilde{X}]$ and $C = [X^*, \bar{X}]$ yield the same projection, and therefore the same 2SLS estimator given by EC2SLS (see problem 7.3). Without going into proofs, we note that Baltagi and Li (1992c) showed that $\widehat{\delta}_{1,\text{G2SLS}}$ and $\widehat{\delta}_{1,\text{EC2SLS}}$ yield the same asymptotic variance–covariance matrix. Therefore, using White's (1986) terminology, $\widetilde{X}$ in B and $\bar{X}$ in C are *redundant* with respect to X^*. Redundant instruments can be interpreted loosely as additional sets of instruments that do not yield extra gains in asymptotic efficiency; see White (1986) for the strict definition and Baltagi and Li (1992c) for the proof in this context.

For applications, it is easy to obtain EC2SLS using a standard 2SLS package.

Step 1. Run W2SLS and B2SLS using a standard 2SLS package on (7.5) and (7.9), i.e., run 2SLS of $\widetilde{y}$ on $\widetilde{Z}$ using $\widetilde{X}$ as instruments and run 2SLS of $\bar{y}$ on $\bar{Z}$ using $\bar{X}$ as instruments. This yields (7.6) and (7.8), respectively.[2] Alternatively, this can be computed using the (xtivreg,fe) and (xtivreg,be) commands in Stata, specifying the endogenous variables and the set of instruments.

Step 2. Compute $\widehat{\sigma}_{\nu_{11}}^2$ and $\widehat{\sigma}_{1_{11}}^2$ from (7.13) and (7.14) and obtain y_1^*, Z_1^* and X^* as described below (7.17). This transforms (7.1) by $\widehat{\Omega}_{11}^{-1/2}$ as in (7.15).

Step 3. Run 2SLS on this transformed equation (7.15) using the instrument set $A = X^*$ or $A = [QX, PX]$ as suggested above, i.e., run 2SLS of y_1^* on Z_1^* using X^* as instruments to get G2SLS, or $[\widetilde{X}, \bar{X}]$ as instruments to get EC2SLS. This yields (7.18) and (7.11), respectively. These computations are easy. They involve simple transformations on the data and the application of 2SLS three times. Alternatively, this can be done with Stata using the (xtivreg,re) command to get G2SLS and (xtivreg,re ec2sls) to get EC2SLS.

7.2 EMPIRICAL EXAMPLE: CRIME IN NORTH CAROLINA

This section replicates the study by Cornwell and Trumbull (1994), hereafter (CT), who estimated an economic model of crime using panel data on 90 counties in North Carolina over the period 1981–87. It is based on Baltagi (2005). The empirical model relates the crime rate (which is an FBI index measuring the number of crimes divided by the county population) to a set of explanatory variables which include deterrent variables as well as variables measuring returns to legal opportunities. All variables are in logs except for the regional and time dummies. The explanatory variables consist of the probability of arrest (which is measured by the ratio of arrests to offenses), probability of conviction given arrest (which is measured by the ratio of convictions to arrests), probability of a prison sentence given a conviction (measured by the proportion of total convictions resulting in prison sentences),

Table 7.1 Economics of Crime Estimates for North Carolina, 1981–87 (standard errors in parentheses)

lcrmrte	Between	Fixed Effects	FE2SLS	BE2SLS	EC2SLS
lprbarr	−0.648	−0.355	−0.576	−0.503	−0.413
	(0.088)	(0.032)	(0.802)	(0.241)	(0.097)
lprbconv	−0.528	−0.282	−0.423	−0.525	−0.323
	(0.067)	(0.021)	(0.502)	(0.100)	(0.054)
lprbpris	0.297	−0.173	−0.250	0.187	−0.186
	(0.231)	(0.032)	(0.279)	(0.318)	(0.042)
lavgsen	−0.236	−0.002	0.009	−0.227	−0.010
	(0.174)	(0.026)	(0.049)	(0.179)	(0.027)
lpolpc	0.364	0.413	0.658	0.408	0.435
	(0.060)	(0.027)	(0.847)	(0.193)	(0.090)
ldensity	0.168	0.414	0.139	0.226	0.429
	(0.077)	(0.283)	(1.021)	(0.102)	(0.055)
lwcon	0.195	−0.038	−0.029	0.314	−0.007
	(0.210)	(0.039)	(0.054)	(0.259)	(0.040)
lwtuc	−0.196	0.046	0.039	−0.199	0.045
	(0.170)	(0.019)	(0.031)	(0.197)	(0.020)
lwtrd	0.129	−0.021	−0.018	0.054	−0.008
	(0.278)	(0.040)	(0.045)	(0.296)	(0.041)
lwfir	0.113	−0.004	−0.009	0.042	−0.004
	(0.220)	(0.028)	(0.037)	(0.306)	(0.029)
lwser	−0.106	0.009	0.019	−0.135	0.006
	(0.163)	(0.019)	(0.039)	(0.174)	(0.020)
lwmfg	−0.025	−0.360	−0.243	−0.042	−0.204
	(0.134)	(0.112)	(0.420)	(0.156)	(0.080)
lwfed	0.156	−0.309	−0.451	0.148	−0.164
	(0.287)	(0.176)	(0.527)	(0.326)	(0.159)
lwsta	−0.284	0.053	−0.019	−0.203	−0.054
	(0.256)	(0.114)	(0.281)	(0.298)	(0.106)
lwloc	0.010	0.182	0.263	0.044	0.163
	(0.463)	(0.118)	(0.312)	(0.494)	(0.120)
lpctmle	−0.095	0.627	0.351	−0.095	−0.108
	(0.158)	(0.364)	(1.011)	(0.192)	(0.140)
lpctmin	0.148	—	—	0.169	0.189
	(0.049)			(0.053)	(0.041)
west	−0.230	—	—	−0.205	−0.227
	(0.108)			(0.114)	(0.100)
central	−0.164	—	—	−0.173	−0.194
	(0.064)			(0.067)	(0.060)
urban	−0.035	—	—	−0.080	−0.225
	(0.132)			(0.144)	(0.116)
_cons	−2.097	—	—	−1.977	−0.954
	(2.822)			(4.001)	(1.284)

Time dummies were included except for Between and BE2SLS. The number of observations is 630. The F-statistic for significance of county dummies in fixed effects is $F(89,518) = 36.38$. The corresponding F-statistic using FE2SLS is 29.66. Both are significant. Hausman's test for (fixed effects – random effects) is χ^2 (22) = 49.4 with p-value of 0.0007. The corresponding Hausman test for (FE2SLS – EC2SLS) is χ^2 (22) = 19.5 with p-value of 0.614.
Source: Baltagi (2005). Reproduced by permission of John Wiley & Sons Ltd.

average prison sentence in days as a proxy for sanction severity, the number of police per capita as a measure of the county's ability to detect crime, the population density (which is the county population divided by county land area), a dummy variable indicating whether the county is in the SMSA with population larger than 50 000, percent minority (which is the proportion of the county's population that is minority or non-white), percent young male (which is the proportion of the county's population that is male and between the ages of 15 and 24), regional dummies for western and central counties. Opportunities in the legal sector are captured by the average weekly wage in the county by industry. These industries are: construction; transportation, utilities and communication; wholesale and retail trade; finance, insurance and real estate; services; manufacturing; and federal, state and local government.

Table 7.1 replicates the Between and fixed effects estimates of CT using Stata. Fixed effects results show that the probability of arrest, the probability of conviction given arrest and the probability of imprisonment given conviction all have a negative and significant effect on the crime rate with estimated elasticities of −0.355, −0.282 and −0.173, respectively. The sentence severity has a negative but insignificant effect on the crime rate. The greater the number of police per capita, the greater the number of reported crimes per capita. The estimated elasticity is 0.413 and it is significant. This could be explained by the fact that the larger the police force, the larger the reported crime. Alternatively, this could be an endogeneity problem with more crime resulting in the hiring of more police. The higher the density of the population the higher the crime rate, but this is insignificant. Returns to legal activity are insignificant except for wages in the manufacturing sector and wages in the transportation, utilities and communication sector. The manufacturing wage has a negative and significant effect on crime with an estimated elasticity of −0.36, while the transportation, utilities and communication sector wage has a positive and significant effect on crime with an estimated elasticity of 0.046. Percent young male is insignificant at the 5% level.

Cornwell and Trumbull (1994) argue that the Between estimates do not control for county effects and yield much higher deterrent effects than the fixed effects estimates. These Between estimates, as well as the random effects estimates are rejected as inconsistent by a Hausman (1978) test. In our replication, this statistic yields a value of 49.4 which is distributed as $\chi^2(22)$ and is significant with a p-value of 0.0007. CT worried about the endogeneity of police per capita and the probability of arrest. They used as instruments two additional variables. Offense mix is the ratio of crimes involving face-to-face contact (such as robbery, assault and rape) to those that do not. The rationale for using this variable is that arrest is facilitated by positive identification of the offender. The second instrument is per capita tax revenue. This is justified on the basis that counties with preferences for law enforcement will vote for higher taxes to fund a larger police force. The fixed effects 2SLS estimates are reported in Table 7.1. All deterrent variables had insignificant t-statistics. These include the probability of arrest, the probability of conviction given arrest as well as the probability of imprisonment given conviction. Also insignificant were sentence severity and police per capita. CT also report 2SLS estimates ignoring the heterogeneity in the county effects for comparison. However, they warn against using biased and inconsistent estimates that ignore county effects. In fact, county effects were always significant, see the F-statistics reported in Table 7.1.

Another way of dealing with the endogeneity problem is to run a random effects 2SLS estimator that allows for the endogeneity of police per capita and the probability of arrest. This estimator is a matrix-weighted average of Between 2SLS and fixed effects 2SLS and was denoted by EC2SLS in (7.11). The Stata output for EC2SLS is given in Table 7.2 using

Table 7.2 EC2SLS Estimates for the Crime Data

```
. xtivreg  lcrmrte lprbconv lprbpris lavgsen  ldensity lwcon lwtuc
< lwtrd lwfir lwser lwmfg lwfed lwsta lwloc lpctymle lpctmin west
< central urban d82 d83 d84 d85 d86 d87 (lprbarr lpolpc= ltaxpc
< lmix), ec2sls

EC2SLS Random-effects regression      Number of obs      =       630
Group variable: county                Number of groups   =        90

R-sq:  within  = 0.4521               Obs per group: min =         7
       between = 0.8158                              avg =       7.0
       overall = 0.7840                              max =         7

                                      Wald chi2(26)      =    575.74
corr(u_i, X)       = 0 (assumed)      Prob > chi2        =    0.0000

-------------------------------------------------------------------
  lcrmrte  |     Coef.   Std. Err.    z   P>|z|   [95% Conf. Interval]
-----------+-------------------------------------------------------
   lprbarr | -.4129261   .097402  -4.24  0.000  -.6038305  -.2220218
    lpolpc |  .4347492   .089695   4.85  0.000   .2589502   .6105482
  lprbconv | -.3228872  .0535517  -6.03  0.000  -.4278465  -.2179279
  lprbpris | -.1863195  .0419382  -4.44  0.000  -.2685169  -.1041222
   lavgsen | -.0101765  .0270231  -0.38  0.706  -.0631408   .0427877
  ldensity |  .4290282  .0548483   7.82  0.000   .3215275    .536529
     lwcon | -.0074751  .0395775  -0.19  0.850  -.0850455   .0700954
     lwtuc |   .045445  .0197926   2.30  0.022   .0066522   .0842379
     lwtrd | -.0081412  .0413828  -0.20  0.844  -.0892499   .0729676
     lwfir | -.0036395  .0289238  -0.13  0.900  -.0603292   .0530502
     lwser |  .0056098  .0201259   0.28  0.780  -.0338361   .0450557
     lwmfg | -.2041398  .0804393  -2.54  0.011   -.361798  -.0464816
     lwfed | -.1635108  .1594496  -1.03  0.305  -.4760263   .1490047
     lwsta | -.0540503  .1056769  -0.51  0.609  -.2611732   .1530727
     lwloc |  .1630523   .119638   1.36  0.173  -.0714339   .3975384
  lpctymle | -.1081057  .1396949  -0.77  0.439  -.3819026   .1656912
   lpctmin |   .189037  .0414988   4.56  0.000   .1077009   .2703731
      west | -.2268433  .0995913  -2.28  0.023  -.4220387  -.0316479
   central | -.1940428  .0598241  -3.24  0.001  -.3112958  -.0767898
     urban | -.2251539  .1156302  -1.95  0.052  -.4517851   .0014772
       d82 |  .0107452  .0257969   0.42  0.677  -.0398158   .0613062
       d83 | -.0837944  .0307088  -2.73  0.006  -.1439825  -.0236063
       d84 | -.1034997  .0370885  -2.79  0.005  -.1761918  -.0308076
       d85 | -.0957017  .0494502  -1.94  0.053  -.1926223   .0012189
       d86 | -.0688982  .0595956  -1.16  0.248  -.1857036   .0479071
       d87 | -.0314071  .0705197  -0.45  0.656  -.1696232   .1068091
     _cons | -.9538032  1.283966  -0.74  0.458  -3.470331   1.562725
-----------+-------------------------------------------------------
   sigma_u | .21455964
   sigma_e | .14923892
       rho | .67394413   (fraction of variance due to u_i)
-------------------------------------------------------------------
Instrumented:   lprbarr lpolpc
Instruments:    lprbconv lprbpris lavgsen ldensity lwcon lwtuc lwtrd
                lwfir lwser  lwmfg lwfed lwsta lwloc lpctymle
                lpctmin west central urban d82  d83 d84 d85 d86 d87
                ltaxpc lmix
```

Table 7.3 Random Effects 2SLS for Crime Data (G2SLS)

```
. xtivreg  lcrmrte lprbconv lprbpris lavgsen  ldensity lwcon lwtuc
< lwtrd lwfir lwser lwmfg lwfed lwsta lwloc lpctymle lpctmin west
< central urban d82 d83 d84 d85 d86 d87 (lprbarr lpolpc= ltaxpc
< lmix), re

G2SLS Random-effects regression       Number of obs      =       630
Group variable: county                Number of groups   =        90

R-sq:  within  = 0.4521               Obs per group: min =         7
       between = 0.8036                              avg =       7.0
       overall = 0.7725                              max =         7

                                      Wald chi2(26)      =    542.48
corr(u_i, X)       = 0 (assumed)      Prob > chi2        =    0.0000
--------------------------------------------------------------------
   lcrmrte |     Coef.  Std. Err.    z    P>|z|    [95% Conf. Interval]
-----------+--------------------------------------------------------
   lprbarr | -.4141383  .2210496  -1.87  0.061  -.8473875   .0191109
    lpolpc |  .5049461  .2277778   2.22  0.027   .0585098   .9513824
  lprbconv | -.3432506  .1324648  -2.59  0.010  -.6028768  -.0836244
  lprbpris | -.1900467  .0733392  -2.59  0.010   -.333789  -.0463045
   lavgsen | -.0064389  .0289407  -0.22  0.824  -.0631617   .0502838
  ldensity |  .4343449  .0711496   6.10  0.000   .2948943   .5737956
     lwcon | -.0042958  .0414226  -0.10  0.917  -.0854826   .0768911
     lwtuc |  .0444589  .0215448   2.06  0.039   .0022318   .0866859
     lwtrd | -.0085579  .0419829  -0.20  0.838  -.0908428    .073727
     lwfir | -.0040305  .0294569  -0.14  0.891  -.0617649   .0537038
     lwser |  .0105602  .0215823   0.49  0.625  -.0317403   .0528608
     lwmfg |  -.201802  .0839373  -2.40  0.016  -.3663161  -.0372878
     lwfed | -.2134579  .2151046  -0.99  0.321  -.6350551   .2081393
     lwsta | -.0601232  .1203149  -0.50  0.617   -.295936   .1756896
     lwloc |  .1835363  .1396775   1.31  0.189  -.0902265   .4572992
  lpctymle | -.1458703  .2268086  -0.64  0.520  -.5904071   .2986664
   lpctmin |  .1948763  .0459385   4.24  0.000   .1048384   .2849141
      west | -.2281821   .101026  -2.26  0.024  -.4261894  -.0301747
   central | -.1987703  .0607475  -3.27  0.001  -.3178332  -.0797075
     urban | -.2595451  .1499718  -1.73  0.084  -.5534844   .0343942
       d82 |  .0132147  .0299924   0.44  0.660  -.0455692   .0719987
       d83 | -.0847693   .032001  -2.65  0.008  -.1474901  -.0220485
       d84 | -.1062027  .0387893  -2.74  0.006  -.1822284  -.0301769
       d85 | -.0977457  .0511681  -1.91  0.056  -.1980334    .002542
       d86 | -.0719451  .0605819  -1.19  0.235  -.1906835   .0467933
       d87 | -.0396595  .0758531  -0.52  0.601  -.1883289   .1090099
      cons | -.4538501  1.702983  -0.27  0.790  -3.791636   2.883935
-----------+--------------------------------------------------------
   sigma_u |  .21455964
   sigma_e |  .14923892
     rho   |  .67394413   (fraction of variance due to u_i)
--------------------------------------------------------------------
Instrumented:   lprbarr lpolpc
Instruments:    lprbconv lprbpris lavgsen ldensity lwcon lwtuc lwtrd
                lwfir lwser  lwmfg lwfed lwsta lwloc lpctymle
                 lpctmin west
                central urban d82  d83 d84 d85 d86 d87 ltaxpc lmix
```

(xtreg,re ec2sls) and the results are summarized in Table 7.1. All the deterrent variables are significant with slightly higher elasticities in absolute value than those reported by fixed effects: -0.413 for the probability of arrest as compared to -0.355; -0.323 for the probability of conviction given arrest as compared to -0.282; -0.186 for the probability of imprisonment given conviction as compared to -0.173. The sentence severity variable is still insignificant and police per capita is still positive and significant. Manufacturing wage is negative and significant and percent minority is positive and significant. Obtaining an estimate of the last coefficient is an advantage of EC2SLS over the fixed effects estimators, because it allows us to recapture estimates of variables that were invariant across time and wiped out by the fixed effects transformation, see also Hausman and Taylor (1981) and section 7.4. Table 7.3 gives the random effects (G2SLS) estimator described in (7.18) using (xtreg,re). G2SLS gives basically the same results as EC2SLS but the standard errors are higher. Remember that EC2SLS uses more instruments than G2SLS. Problem 04.1.1 in *Econometric Theory* by Baltagi (2004) suggests a Hausman test based on the difference between fixed effects 2SLS and random effects 2SLS. For the crime data, this yields a Hausman statistic of 19.50 which is distributed as $\chi^2(22)$ and is insignificant with a p-value of 0.614. This does not reject the null hypothesis that EC2SLS yields a consistent estimator. This can be computed using the Hausman command after storing the EC2SLS and FE2SLS estimates. Recall that the random effects estimator was rejected by Cornwell and Trumbull (1994) based on the standard Hausman (1978) test. This was based on the contrast between fixed effects and random effects assuming that the endogeneity comes entirely from the correlation between the county effects and the explanatory variables. This does not account for the endogeneity of the conventional simultaneous equation type between police per capita and the probability of arrest and the crime rate. This alternative Hausman test based on the contrast between fixed effects 2SLS and EC2SLS failed to reject the null hypothesis. This result should be tempered by the fact that FE2SLS is imprecise for this application and its consistency depends on the legitimacy of the instruments chosen by CT. We also ran the first stage regressions to check for weak instruments. For the probability of arrest, the F-statistic of the fixed effects first-stage regression was 15.6 as compared to 4.62 for the between first stage regression. Similarly, for the police per capita, the F-statistic of the fixed effects first-stage regression was 9.27 as compared to 2.60 for the between first stage regression. This indicates that these instruments may be weaker in the between first stage regressions (for between 2SLS) than in the fixed effects first stage regressions (for fixed effects 2SLS).

7.3 SYSTEM ESTIMATION

Consider the system of identified equations:

$$y = Z\delta + u \tag{7.20}$$

where $y' = (y_1', \ldots, y_M')$, $Z = \text{diag}[Z_j]$, $\delta' = (\delta_1', \ldots, \delta_M')$ and $u' = (u_1', \ldots, u_M')$ with $Z_j = [Y_j, X_j]$ of dimension $NT \times (g_j + k_j)$, for $j = 1, \ldots, M$. In this case, there are g_j included right-hand side Y_j and k_j included right-hand side X_j. This differs from the SUR model only in the fact that there are right-hand side endogenous variables in the system of equations. For the one-way error component model, the disturbance of the jth equation u_j is given by (6.2) and

$\Omega_{jl} = E(u_j u_l')$ is given by (6.4) as in the SUR case. Once again, the covariance matrix between the disturbances of different equations has the same error component form. Except now, there are additional cross-equations variance components to be estimated. The variance–covariance matrix of the set of M structural equations $\Omega = E(uu')$ is given by (6.5) and $\Omega^{-1/2}$ is given by (6.8). Premultiplying (7.20) by $(I_M \otimes Q)$ yields

$$\widetilde{y} = \widetilde{Z}\delta + \widetilde{u} \tag{7.21}$$

where $\widetilde{y} = (I_M \otimes Q)y$, $\widetilde{Z} = (I_M \otimes Q)Z$ and $\widetilde{u} = (I_M \otimes Q)u$. Performing 3SLS on (7.21) with $(I_M \otimes \widetilde{X})$ as the set of instruments, where $\widetilde{X} = QX$, one gets the Within 3SLS estimator:

$$\widetilde{\delta}_{\text{W3SLS}} = [\widetilde{Z}'(\Sigma_\nu^{-1} \otimes P_{\widetilde{X}})\widetilde{Z}]^{-1}[\widetilde{Z}'(\Sigma_\nu^{-1} \otimes P_{\widetilde{X}})\widetilde{y}] \tag{7.22}$$

Similarly, transforming (7.20) by $(I_M \otimes P)$ yields

$$\bar{y} = \bar{Z}\delta + \bar{u} \tag{7.23}$$

where $\bar{y} = (I_M \otimes P)y$, $\bar{Z} = (I_M \otimes P)Z$ and $\bar{u} = (I_M \otimes P)u$. Performing 3SLS on the transformed system (7.23) using $(I_M \otimes \bar{X})$ as the set of instruments, where $\bar{X} = PX$, one gets the Between 3SLS estimator:

$$\widehat{\delta}_{\text{B3SLS}} = [\bar{Z}'(\Sigma_1^{-1} \otimes P_{\bar{X}})\bar{Z}]^{-1}[\bar{Z}'(\Sigma_1^{-1} \otimes P_{\bar{X}})\bar{y}] \tag{7.24}$$

Next, we stack the two transformed systems given in (7.21) and (7.23) after premultiplying by $(I_M \otimes \widetilde{X}')$ and $(I_M \otimes \bar{X}')$, respectively. Then we perform GLS noting that δ is the same for these two transformed systems (see problem 7.5). The resulting estimator of δ is the error components three-stage least squares (EC3SLS) given by Baltagi (1981b):

$$\begin{aligned}\widehat{\delta}_{\text{EC3SLS}} = {} & [\widetilde{Z}'(\Sigma_\nu^{-1} \otimes P_{\widetilde{X}})\widetilde{Z} + \bar{Z}'(\Sigma_1^{-1} \otimes P_{\bar{X}})\bar{Z}]^{-1} \\ & \times[\widetilde{Z}'(\Sigma_\nu^{-1} \otimes P_{\widetilde{X}})\widetilde{y} + \bar{Z}'(\Sigma_1^{-1} \otimes P_{\bar{X}})\bar{y}]\end{aligned} \tag{7.25}$$

Note that $\widehat{\delta}_{\text{EC3SLS}}$ can also be written as a matrix-weighted average of $\widehat{\delta}_{\text{W3SLS}}$ and $\widehat{\delta}_{\text{B3SLS}}$ as follows:

$$\widehat{\delta}_{\text{EC3SLS}} = W_1\widehat{\delta}_{\text{W3SLS}} + W_2\widehat{\delta}_{\text{B3SLS}} \tag{7.26}$$

with

$$W_1 = [\widetilde{Z}'(\Sigma_\nu^{-1} \otimes P_{\widetilde{X}})\widetilde{Z} + \bar{Z}'(\Sigma_1^{-1} \otimes P_{\bar{X}})\bar{Z}]^{-1}[\widetilde{Z}'(\Sigma_\nu^{-1} \otimes P_{\widetilde{X}})\widetilde{Z}]$$

and

$$W_2 = [\widetilde{Z}'(\Sigma_\nu^{-1} \otimes P_{\widetilde{X}})\widetilde{Z} + \bar{Z}'(\Sigma_1^{-1} \otimes P_{\bar{X}})\bar{Z}]^{-1}[\bar{Z}'(\Sigma_1^{-1} \otimes P_{\bar{X}})\bar{Z}]$$

Consistent estimates of Σ_ν and Σ_1 can be obtained as in (7.13) and (7.14) using W2SLS and B2SLS residuals with

$$\widehat{\sigma}^2_{\nu_{jl}} = (y_j - Z_j\widetilde{\delta}_{j,\text{W2SLS}})'Q(y_l - Z_l\widetilde{\delta}_{l,\text{W2SLS}})/N(T-1) \tag{7.27}$$

$$\widehat{\sigma}^2_{1_{jl}} = (y_j - Z_j\widehat{\delta}_{j,\text{B2SLS}})'P(y_l - Z_l\widehat{\delta}_{l,\text{B2SLS}})/N \tag{7.28}$$

One should check whether $\widehat{\Sigma}_\mu = (\widehat{\Sigma}_1 - \widehat{\Sigma}_\nu)/T$ is positive definite.

Using $\Omega^{-1/2}$ from (6.8), one can transform (7.20) to get

$$y^* = Z^*\delta + u^* \tag{7.29}$$

with $y^* = \Omega^{-1/2}y$, $Z^* = \Omega^{-1/2}Z$ and $u^* = \Omega^{-1/2}u$. For an arbitrary set of instruments A, the 3SLS estimator of (7.29) becomes

$$\widehat{\delta}_{\text{3SLS}} = (Z^{*\prime}P_A Z^*)^{-1}Z^{*\prime}P_A y^* \tag{7.30}$$

Using the results of White (1986), the optimal set of instruments is

$$X^* = \Omega^{-1/2}(I_M \otimes X) = (\Sigma_\nu^{-1/2} \otimes QX) + (\Sigma_1^{-1/2} \otimes PX)$$

Substituting $A = X^*$ in (7.30), one gets the efficient three-stage least squares (E3SLS) estimator:

$$\widehat{\delta}_{\text{E3SLS}} = (Z^{*\prime}P_{X^*}Z^*)^{-1}Z^{*\prime}P_{X^*}y^* \tag{7.31}$$

This is not the G3SLS estimator suggested by Balestra and Varadharajan-Krishnakumar (1987). In fact, Balestra and Varadharajan-Krishnakumar (1987) suggest using

$$\begin{aligned} A &= \Omega^{1/2}\,\text{diag}[\Omega_{jj}^{-1}](I_M \otimes X) \\ &= \Sigma_\nu^{1/2}\,\text{diag}\left(\frac{1}{\sigma^2_{\nu_{jj}}}\right) \otimes \widetilde{X} + \Sigma_1^{1/2}\,\text{diag}\left(\frac{1}{\sigma^2_{1_{jj}}}\right) \otimes \bar{X} \end{aligned} \tag{7.32}$$

Substituting this A in (7.30) yields the G3SLS estimator of δ. So, how are G3SLS, EC3SLS and E3SLS related? Baltagi and Li (1992c) show that Baltagi's (1981b) EC3SLS estimator can be obtained from (7.30) with $A = [I_M \otimes \widetilde{X}, I_M \otimes \bar{X}]$. From this it is clear that the set of instruments $[I_M \otimes \widetilde{X}, I_M \otimes \bar{X}]$ used by Baltagi (1981b) spans the set of instruments $[\Sigma_\nu^{-1/2} \otimes \widetilde{X} + \Sigma_1^{-1/2} \otimes \bar{X}]$ needed for E3SLS. In addition, we note without proof that Baltagi and Li (1992c) show that $\widehat{\delta}_{\text{EC3SLS}}$ and $\widehat{\delta}_{\text{E3SLS}}$ yield the same asymptotic variance–covariance matrix. Problem 7.6 shows that Baltagi's (1981b) EC3SLS estimator has redundant instruments with respect to those used by the E3SLS estimator. Therefore, using White's (1984) terminology, the extra instruments used by Baltagi (1981b) do not yield extra gains in asymptotic efficiency. However, Baltagi and Li (1992c) also show that both EC3SLS and E3SLS are asymptotically more efficient than the G3SLS estimator corresponding to the set of instruments given by (7.32). In applications, it is easy to obtain EC3SLS using a standard 3SLS package.

Step 1. Obtain W2SLS and B2SLS estimates of each structural equation as described in the first step of computing EC2SLS.

Step 2. Compute estimates of $\widehat{\Sigma}_1$ and $\widehat{\Sigma}_\nu$ as described in (7.27) and (7.28).

Step 3. Obtain the Cholesky decomposition of $\widehat{\Sigma}_1^{-1}$ and $\widehat{\Sigma}_\nu^{-1}$ and use those instead of $\widehat{\Sigma}_1^{-1/2}$ and $\widehat{\Sigma}_\nu^{-1/2}$ in the transformation described in (7.29), i.e., obtain y^*, Z^* and X^* as described below (7.30).

Step 4. Apply 3SLS to this transformed system (7.29) using as a set of instruments $A = X^*$ or $A = [I_M \otimes \widetilde{X}, I_M \otimes \bar{X}]$, i.e., run 3SLS of y^* on Z^* using as instruments X^* or $[I_M \otimes \widetilde{X}, I_M \otimes \bar{X}]$. These yield (7.31) and (7.25), respectively. The computations are again easy, requiring simple transformations and a 3SLS package.

Baltagi (1981b) shows that EC3SLS reduces to EC2SLS when the disturbances of the different structural equations are uncorrelated with each other, but not necessarily when all the structural equations are just identified. This is different from the analogous conditions between 2SLS and 3SLS in the classical simultaneous equations model (see problem 7.7).

Baltagi (1984) also performs Monte Carlo experiments on a two-equation simultaneous model with error components and demonstrates the efficiency gains in terms of mean squared error in performing EC2SLS and EC3SLS over the standard simultaneous equation counterparts, 2SLS and 3SLS. EC2SLS and EC3SLS also performed better than a two- or three-stage variance components method suggested by Maddala (1977) where right-hand side endogenous variables are replaced by their predicted values from the reduced form and the standard error component GLS is performed in the second step. Also, Baltagi (1984) demonstrates that better estimates of the variance components do not necessarily imply better estimates of the structural or reduced form parameters.[3] Mátyás and Lovrics (1990) performed Monte Carlo experiments on a just identified two-equation static model and compared OLS, Within-2SLS, true EC2SLS and a feasible EC2SLS for various generated exogenous variables and a variety of N and T. They recommend the panel data estimators as long as N and T are both larger than 15. Prucha (1985) derives the full information maximum likelihood (FIML) estimator of the simultaneous equation model with error components assuming normality of the disturbances. Prucha shows that this FIML estimator has an instrumental variable representation. The instrumental variable form of the normal equations of the FIML estimator is used to generate a wide class of instrumental variable estimators. Prucha also establishes the existence of wide asymptotic equivalence classes of full and limited information estimators of which Baltagi's EC2SLS and EC3SLS are members. Balestra and Varadharajan-Krishnakumar (1987) derive the limiting distributions of both the coefficient estimators and covariance estimators of the FIML method for the SEM with error components. Krishnakumar (1988) provides a useful summary of the simultaneous equations with error components literature, which is updated in her chapter in Mátyás and Sevestre (1996).

For an application of Within-2SLS to estimate regional supply and demand functions for the Southern Pine lumber industry, see Shim (1982). See Nguyen and Bernier (1988) for an application of Within-2SLS to a system of simultaneous equations which examines the influence of a firm's market power on its risk level using Tobin's q. See Baltagi and Blien (1998) for an application of Within-2SLS to the estimation of a wage curve for Germany using data for 142 labor market regions over the period 1981–90. Briefly, the wage curve describes the negative relationship between the local unemployment rate and the level of wages. Baltagi and Blien (1998) find that ignoring endogeneity of the local employment rate yields results in favor of the wage curve only for younger and less qualified workers. Accounting for endogeneity of the unemployment rate yields evidence in favor of the wage curve across all types of workers. In particular, the wages of less qualified workers are more responsive to local unemployment rates than the wages of more qualified workers. Also, the wages of men are slightly more responsive to local unemployment rates than the wages of women. Applications of EC2SLS and EC3SLS include: (i) an econometric rational-expectations macroeconomic model for developing countries with capital controls (see Haque, Lahiri and Montiel, 1993), and (ii) an econometric model measuring income and price elasticities of foreign trade for developing countries (see Kinal and Lahiri, 1993).

7.4 THE HAUSMAN AND TAYLOR ESTIMATOR

Let us reconsider the single equation estimation case but now focus on endogeneity occurring through the unobserved individual effects. Examples where μ_i and the explanatory variables may be correlated include an earnings equation, where the unobserved individual ability may be correlated with schooling and experience; also a production function, where managerial ability may be correlated with the inputs. Mundlak (1978) considered the one-way error component regression model in (2.5) but with the additional auxiliary regression

$$\mu_i = \bar{X}'_{i.}\pi + \epsilon_i \tag{7.33}$$

where $\epsilon_i \sim \text{IIN}(0, \sigma^2_\epsilon)$ and $\bar{X}'_{i.}$ is a $1 \times K$ vector of observations on the explanatory variables averaged over time. In other words, Mundlak assumed that the individual effects are a linear function of the averages of *all* the explanatory variables across time. These effects are uncorrelated with the explanatory variables if and only if $\pi = 0$. Mundlak (1978) assumed, without loss of generality, that the X are deviations from their sample mean. In vector form, one can write (7.33) as

$$\mu = Z'_\mu X\pi/T + \epsilon \tag{7.34}$$

where $\mu' = (\mu_1, \ldots, \mu_N)$, $Z_\mu = I_N \otimes \iota_T$ and $\epsilon' = (\epsilon_1, \ldots, \epsilon_N)$. Substituting (7.34) in (2.5), with no constant, one gets

$$y = X\beta + PX\pi + (Z_\mu\epsilon + \nu) \tag{7.35}$$

where $P = I_N \otimes \bar{J}_T$. Using the fact that the ϵ and the ν are uncorrelated, the new error in (7.35) has zero mean and variance–covariance matrix

$$V = E(Z_\mu\epsilon + \nu)(Z_\mu\epsilon + \nu)' = \sigma^2_\epsilon(I_N \otimes J_T) + \sigma^2_\nu I_{NT} \tag{7.36}$$

Using partitioned inverse, one can verify (see problem 7.8), that GLS on (7.35) yields

$$\widehat{\beta}_{\text{GLS}} = \widetilde{\beta}_{\text{Within}} = (X'QX)^{-1}X'Qy \tag{7.37}$$

and

$$\widehat{\pi}_{\text{GLS}} = \widehat{\beta}_{\text{Between}} - \widetilde{\beta}_{\text{Within}} = (X'PX)^{-1}X'Py - (X'QX)^{-1}X'Qy \tag{7.38}$$

with

$$\begin{aligned}\text{var}(\widehat{\pi}_{\text{GLS}}) &= \text{var}(\widehat{\beta}_{\text{Between}}) + \text{var}(\widetilde{\beta}_{\text{Within}}) \\ &= (T\sigma^2_\epsilon + \sigma^2_\nu)(X'PX)^{-1} + \sigma^2_\nu(X'QX)^{-1}\end{aligned} \tag{7.39}$$

Therefore, Mundlak (1978) showed that the best linear unbiased estimator of (2.5) becomes the fixed effects (Within) estimator once these individual effects are modeled as a linear function of *all* the X_{it} as in (7.33). The random effects estimator on the other hand is biased because it ignores (7.33). Note that Hausman's test based on the Between minus Within estimators is basically a test for $H_0 : \pi = 0$ and this turns out to be another natural derivation for the test considered in Chapter 4, namely,

$$\widehat{\pi}'_{\text{GLS}}(\text{var}(\widehat{\pi}'_{\text{GLS}}))^{-1}\widehat{\pi}_{\text{GLS}} \overset{H_0}{\to} \chi^2_K$$

Mundlak's (1978) formulation in (7.35) assumes that *all* the explanatory variables are related to the individual effects. The random effects model on the other hand assumes *no* correlation between the explanatory variables and the individual effects. The random effects model generates the GLS estimator, whereas Mundlak's formulation produces the Within estimator. Instead of this "*all* or *nothing*" correlation among the X and the μ_i, Hausman and Taylor (1981) consider a model where some of the explanatory variables are related to the μ_i. In particular, they consider the following model:

$$y_{it} = X_{it}\beta + Z_i\gamma + \mu_i + \nu_{it} \tag{7.40}$$

where the Z_i are cross-sectional time-invariant variables. Hausman and Taylor (1981), hereafter HT, split X and Z into two sets of variables: $X = [X_1; X_2]$ and $Z = [Z_1; Z_2]$ where X_1 is $n \times k_1$, X_2 is $n \times k_2$, Z_1 is $n \times g_1$, Z_2 is $n \times g_2$ and $n = NT$. X_1 and Z_1 are assumed exogenous in that they are not correlated with μ_i, and ν_{it} while X_2 and Z_2 are endogenous because they are correlated with the μ_i, but not the ν_{it}. The Within transformation would sweep the μ_i and remove the bias, but in the process it would also remove the Z_i and hence the Within estimator will not give an estimate of γ. To get around that, HT suggest premultiplying the model by $\Omega^{-1/2}$ and using the following set of instruments: $A_0 = [Q, X_1, Z_1]$, where $Q = I_{NT} - P$ and $P = (I_N \otimes \bar{J}_T)$. Breusch, Mizon and Schmidt (1989), hereafter BMS, show that this set of instruments yields the same projection and is therefore equivalent to another set, namely $A_{\text{HT}} = [QX_1, QX_2, PX_1, Z_1]$. The latter set of instruments A_{HT} is feasible, whereas A_0 is not.[4] The order condition for identification gives the result that k_1 the number of variables in X_1 must be at least as large as g_2 the number of variables in Z_2. Note that $\widetilde{X}_1 = QX_1$, $\widetilde{X}_2 = QX_2$, $\bar{X}_1 = PX_1$ and Z_1 are used as instruments. Therefore X_1 is used twice, once as averages and another time as deviations from these averages. This is an advantage of panel data allowing instruments from *within* the model. Note that the Within transformation wipes out the Z_i and does not allow the estimation of γ. In order to get consistent estimates of γ, HT propose obtaining the Within residuals and averaging them over time

$$\widehat{d}_i = \bar{y}_{i.} - \bar{X}_{i.}'\widetilde{\beta}_W \tag{7.41}$$

Then, (7.40) averaged over time can be estimated by running 2SLS of $\widehat{d}_i$ on Z_i with the set of instruments $A = [X_1, Z_1]$. This yields

$$\widehat{\gamma}_{\text{2SLS}} = (Z'P_A Z)^{-1} Z'P_A\widehat{d} \tag{7.42}$$

where $P_A = A(A'A)^{-1}A'$. It is clear that the order condition has to hold $(k_1 \geq g_2)$ for $(Z'P_A Z)$ to be nonsingular. Next, the variance components estimates are obtained as follows:

$$\widetilde{\sigma}_\nu^2 = \widetilde{y}'\bar{P}_{\widetilde{X}}\widetilde{y}/N(T-1) \tag{7.43}$$

where $\widetilde{y} = Qy$, $\widetilde{X} = QX$, $\bar{P}_A = I - P_A$ and

$$\widetilde{\sigma}_1^2 = \frac{(y_{it} - X_{it}\widetilde{\beta}_W - Z_i\widehat{\gamma}_{\text{2SLS}})'P(y_{it} - X_{it}\widetilde{\beta}_W - Z_i\widehat{\gamma}_{\text{2SLS}})}{N} \tag{7.44}$$

This last estimate is based upon an NT vector of residuals. Once the variance components estimates are obtained, the model in (7.40) is transformed using $\widehat{\Omega}^{-1/2}$ as follows:

$$\widehat{\Omega}^{-1/2}y_{it} = \widehat{\Omega}^{-1/2}X_{it}\beta + \widehat{\Omega}^{-1/2}Z_i\gamma + \widehat{\Omega}^{-1/2}u_{it} \tag{7.45}$$

The HT estimator is basically 2SLS on (7.45) using $A_{\text{HT}} = [\widetilde{X}, \bar{X}_1, Z_1]$ as a set of instruments.

(1) If $k_1 < g_2$, then the equation is under-identified. In this case $\widehat{\beta}_{\text{HT}} = \widetilde{\beta}_W$ and $\widehat{\gamma}_{\text{HT}}$ does not exist.
(2) If $k_1 = g_2$, then the equation is just-identified. In this case, $\widehat{\beta}_{\text{HT}} = \widetilde{\beta}_W$ and $\widehat{\gamma}_{\text{HT}} = \widehat{\gamma}_{\text{2SLS}}$ given by (7.42).

(3) If $k_1 > g_2$, then the equation is over-identified and the HT estimator obtained from (7.45) is more efficient than the Within estimator.

A test for over-identification is obtained by computing

$$\widehat{m} = \hat{q}'[\mathrm{var}(\widetilde{\beta}_W) - \mathrm{var}(\widehat{\beta}_{\mathrm{HT}})]^{-}\widehat{q} \tag{7.46}$$

with $\widehat{q} = \widehat{\beta}_{\mathrm{HT}} - \widetilde{\beta}_W$ and $\widehat{\sigma}_\nu^2 \widehat{m} \overset{H_0}{\to} \chi_l^2$ where $l = \min[k_1 - g_2, NT - K]$.

Note that $y^* = \widehat{\sigma}_\nu \widehat{\Omega}^{-1/2} y$ has a typical element $y_{it}^* = y_{it} - \widehat{\theta}\bar{y}_{i.}$ where $\widehat{\theta} = 1 - \widehat{\sigma}_\nu/\widehat{\sigma}_1$ and similar terms exist for X_{it}^* and Z_i^*. In this case 2SLS on (7.45) yields

$$\begin{pmatrix} \widehat{\beta} \\ \widehat{\gamma} \end{pmatrix} = \left[\begin{pmatrix} X^{*\prime} \\ Z^{*\prime} \end{pmatrix} P_A (X^*, Z^*) \right]^{-1} \begin{pmatrix} X^{*\prime} \\ Z^{*\prime} \end{pmatrix} P_A y^* \tag{7.47}$$

where P_A is the projection matrix on $A_{\mathrm{HT}} = [\widetilde{X}, \bar{X}_1, Z_1]$.

Amemiya and MaCurdy (1986), hereafter AM, suggest a more efficient set of instruments $A_{\mathrm{AM}} = [QX_1, QX_2, X_1^*, Z_1]$ where $X_1^* = X_1^0 \otimes \iota_T$ and

$$X_1^0 = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1T} \\ \vdots & \vdots & \dots & \vdots \\ X_{N1} & X_{N2} & \dots & X_{NT} \end{bmatrix} \tag{7.48}$$

is an $(N \times k_1 T)$ matrix. So X_1 is used $(T+1)$ times, once as $\widetilde{X}_1$ and T times as X_1^*. The order condition for identification is now more likely to be satisfied $(Tk_1 > g_2)$. However, this set of instruments requires a stronger exogeneity assumption than that of Hausman and Taylor (1981). The latter requires only uncorrelatedness of the mean of X_1 from the μ_i, i.e.

$$\operatorname{plim}\left(\frac{1}{N} \sum_{i=1}^{N} \bar{X}_{1i.}\mu_i \right) = 0$$

while Amemiya and MaCurdy (1986) require

$$\operatorname{plim}\left(\frac{1}{N} \sum_{i=1}^{N} X_{1it}\mu_i \right) = 0 \quad \text{for } t = 1, \dots, T$$

i.e. uncorrelatedness at each point in time. Breusch et al. (1989) suggest yet a more efficient set of instruments

$$A_{\mathrm{BMS}} = [QX_1, QX_2, PX_1, (QX_1)^*, (QX_2)^*, Z_1]$$

so that X_1 is used $(T+1)$ times and X_2 is used T times. This requires even more exogeneity assumptions, i.e. $\widetilde{X}_2 = QX_2$ should be uncorrelated with the μ_i effects. The BMS order condition becomes $Tk_1 + (T-1)k_2 \geq g_2$.

For the Hausman and Taylor (1981) model given in (7.40), Metcalf (1996) shows that using less instruments may lead to a more powerful Hausman specification test. Asymptotically, more instruments lead to more efficient estimators. However, the asymptotic bias of the less efficient estimator will also be greater as the null hypothesis of no correlation is violated. Metcalf argues that if the bias increases at the same rate as the variance (as the null is violated) for the less efficient estimator, then the power of the Hausman test will increase. This is due to the fact that the test statistic is linear in variance but quadratic in bias.

Computational Note

The number of instruments used by the AM and BMS procedures can increase rapidly as T and the number of variables in the equation get large. For large N panels, small T and reasonable k, this should not be a problem. However, even for $T = 7$, $k_1 = 4$ and $k_2 = 5$ as in the empirical illustration used in the next section, the number of additional instruments used by HT is 4 as compared to 28 for AM and 58 for BMS.[5]

7.5 EMPIRICAL EXAMPLE: EARNINGS EQUATION USING PSID DATA

Cornwell and Rupert (1988) apply these three instrumental variable (IV) methods to a returns to schooling example based on a panel of 595 individuals observed over the period 1976–82 and drawn from the Panel Study of Income Dynamics (PSID). A description of the data is given in Cornwell and Rupert (1988) and is put on the Wiley web site as Wage.xls. In particular, log wage is regressed on years of education (ED), weeks worked (WKS), years of full-time work experience (EXP), occupation (OCC = 1, if the individual is in a blue-collar occupation), residence (SOUTH = 1, SMSA = 1, if the individual resides in the South, or in a standard metropolitan statistical area), industry (IND = 1, if the individual works in a manufacturing industry), marital status (MS = 1, if the individual is married), sex and race (FEM = 1, BLK = 1, if the individual is female or black), union coverage (UNION = 1, if the individual's wage is set by a union contract) and time dummies to capture productivity and price level effects. Baltagi and Khanti-Akom (1990) replicate this study and some of their results in table II are reproduced in Table 7.4. The conventional GLS indicates that an additional year of schooling produces a 10% wage gain. But conventional GLS does not account for the possible correlation of the explanatory variables with the individual effects. The Within transformation eliminates the individual effects and all the Z_i variables, and the resulting Within estimator is consistent even if the individual effects are correlated with the explanatory variables. The Within estimates are quite different from those of GLS, and the Hausman test based on the difference between these two estimates yields $\chi^2_9 = 5075$ which is significant. This rejects the hypothesis of no correlation between the individual effects and the explanatory variables. This justifies the use of the IV methods represented as HT and AM in Table 7.4. We let $X_1 =$ (OCC, SOUTH, SMSA, IND), $X_2 =$ (EXP, EXP^2, WKS, MS, UNION), $Z_1 =$ (FEM, BLK) and $Z_2 =$ (ED). Table 7.5 reproduces the Hausman and Taylor (1981) estimates using the (xthtaylor) command in Stata. The coefficient of ED is estimated as 13.8%, 38% higher than the estimate obtained using GLS (10%). A Hausman test based on the difference between HT and the Within estimator yields $\chi^2_3 = 5.26$, which is not significant at the 5% level. There are three degrees of freedom since there are three over-identifying conditions (the number of X_1 variables minus the number of Z_2 variables).

Therefore, we cannot reject that the set of instruments X_1 and Z_1 chosen are legitimate. Table 7.6 reproduces the Amemiya and MaCurdy (1986) estimates using the (xthtaylor) command in Stata with the (amacurdy) option. These estimates are close to the HT estimates. The additional exogeneity assumptions needed for the AM estimator are not rejected using a Hausman test based on the difference between the HT and AM estimators. This yields $\chi^2_{13} = 14.67$, which is not significant at the 5% level. The BMS estimates (not reported here but available in Baltagi and Khanti-Akom (1990)) are similar to those of AM. Again, the additional exogeneity assumptions needed for the BMS estimator are not rejected using a Hausman test based on the

Table 7.4 Dependent Variable: Log Wage*

	GLS	Within	HT	AM
Constant	4.264	—	2.913	2.927
	(0.098)		(0.284)	(0.275)
WKS	0.0010	0.0008	0.0008	0.0008
	(0.0008)	(0.0006)	(0.0006)	(0.0006)
SOUTH	− 0.017	−0.002	0.007	0.007
	(0.027)	(0.034)	(0.032)	(0.032)
SMSA	− 0.014	−0.042	−0.042	−0.042
	(0.020)	(0.019)	(0.019)	(0.019)
MS	− 0.075	−0.030	−0.030	−0.030
	(0.023)	(0.019)	(0.019)	(0.019)
EXP	0.082	0.113	0.113	0.113
	(0.003)	(0.002)	(0.002)	(0.002)
EXP^2	−0.0008	−0.0004	−0.0004	−0.0004
	(0.00006)	(0.00005)	(0.00005)	(0.00005)
OCC	−0.050	−0.021	−0.021	−0.021
	(0.017)	(0.014)	(0.014)	(0.014)
IND	0.004	0.019	0.014	0.014
	(0.017)	(0.015)	(0.015)	(0.015)
UNION	0.063	0.033	0.033	0.032
	(0.017)	(0.015)	(0.015)	(0.015)
FEM	−0.339	—	−0.131	−0.132
	(0.051)		(0.127)	(0.127)
BLK	−0.210	—	−0.286	−0.286
	(0.058)		(0.156)	(0.155)
ED	0.100	—	0.138	0.137
	(0.006)		(0.021)	(0.021)
		$\chi^2_9 = 5075$	$\chi^2_3 = 5.26$	$\chi^2_{13} = 14.67$

* X_2 (OCC, SOUTH, SMSA, IND), Z_1 = (FEM, BLK).
Source: Baltagi and Khanti-Akom (1990). Reproduced by permission of John Wiley & Sons Ltd.

difference between the AM and BMS estimators. This yields $\chi^2_{13} = 9.59$, which is not significant at the 5% level. Bowden and Turkington (1984) argue that canonical correlations are a useful device for comparing different sets of instruments. In fact, as far as asymptotic efficiency is concerned, one should use instruments for which the canonical correlations with the regressors are maximized. Baltagi and Khanti-Akom (1990) compute the canonical correlations for these three sets of instruments. The geometric average of the canonical correlations (which is a measure of the squared correlations between the set of instruments and the regressors) gives an idea of the gains in asymptotic efficiency for this particular data set as one moves from A_{HT} to A_{AM} to A_{BMS}. These are 0.682 for HT, 0.740 for AM and 0.770 for BMS.

For another application of the HT, AM and BMS estimators to a study of the impact of health on wages, see Contoyannis and Rice (2001). This paper considers the effect of self-assessed general and psychological health on hourly wages using longitudinal data from the six waves of the British Household Panel Survey. Contoyannis and Rice show that reduced psychological health reduces the hourly wage for males, while excellent self-assessed health increases the hourly wage for females. Recently, Egger and Pfaffermayr (2004b) used a Hausman–Taylor SUR model to study the effects of distance as a common determinant of exports and foreign direct investment (FDI) in a three-factor New Trade Theory model. They used industry-level

Table 7.5 Hausman and Taylor Estimates of a Mincer Wage Equation

```
. xthtaylor lwage occ south smsa ind exp exp2 wks ms union fem blk
   ed, endog (exp exp2 wks ms union ed)

Hausman-Taylor estimation                    Number of obs      =      4165
Group variable (i): id                       Number of groups   =       595

                                             Obs per group: min =         7
                                                            avg =         7
                                                            max =         7

Random effects u_i ~ i.i.d.                  Wald chi2(12)      =   6891.87
                                             Prob > chi2        =    0.0000

-------------------------------------------------------------------
       lwage |  Coef.      Std. Err.  z     P>|z|  [95% Conf. Interval]
------------+------------------------------------------------------
TVexogenous  |
         occ |-.0207047  .0137809  -1.50  0.133  -.0477149 .0063055
       south | .0074398   .031955   0.23  0.816  -.0551908 .0700705
        smsa |-.0418334  .0189581  -2.21  0.027  -.0789906-.0046761
         ind | .0136039  .0152374   0.89  0.372  -.0162608 .0434686
TVendogenous |
         exp | .1131328   .002471  45.79  0.000   .1082898 .1179758
        exp2 |-.0004189  .0000546  -7.67  0.000  -.0005259-.0003119
         wks | .0008374  .0005997   1.40  0.163  -.0003381 .0020129
          ms |-.0298508    .01898  -1.57  0.116  -.0670508 .0073493
       union | .0327714  .0149084   2.20  0.028   .0035514 .0619914
TIexogenous  |
         fem |-.1309236   .126659  -1.03  0.301  -.3791707  .1173234
         blk |-.2857479  .1557019  -1.84  0.066  -.5909179  .0194221
TIendogenous |
          ed |  .137944  .0212485   6.49  0.000   .0962977  .1795902
             |
       _cons | 2.912726  .2836522  10.27  0.000   2.356778  3.468674
------------+------------------------------------------------------
     sigma_u | .94180304
     sigma_e | .15180273
         rho | .97467788   (fraction of variance due to u_i)
-------------------------------------------------------------------
note:  TV refers to time-varying; TI refers to time-invariant.
```

data of bilateral outward FDI stocks and exports of the USA and Germany to other countries between 1989 and 1999. They find that distance exerts a positive and significant impact on bilateral stocks of outward FDI of both the USA and Germany. However, the effect of distance on exports is much smaller in absolute size and significantly negative for the USA but insignificant for Germany.

7.6 EXTENSIONS

Cornwell et al. (1992) consider a simultaneous equation model with error components that distinguishes between two types of exogenous variables, namely *singly exogenous* and *doubly*

Table 7.6 Amemiya and MaCurdy Estimates of a Mincer Wage Equation

```
. xthtaylor lwage occ south smsa ind exp exp2 wks ms union fem blk
  ed, endog (exp exp2 wks ms union ed) amacurdy

Amemiya-MaCurdy estimation                Number of obs      =      4165
Group variable (i): id                    Number of groups   =       595

                                          Obs per group: min =         7
                                                         avg =         7
                                                         max =         7

Random effects u_i ~ i.i.d.               Wald chi2(12)      =   6879.20
                                          Prob > chi2        =    0.0000

--------------------------------------------------------------------------
       lwage |     Coef.   Std. Err.   z     P>|z|   [95% Conf. Interval]
-------------+------------------------------------------------------------
TVexogenous  |
         occ | -.0208498 .0137653   -1.51   0.130   -.0478292   .0061297
       south |  .0072818 .0319365    0.23   0.820   -.0553126   .0698761
        smsa | -.0419507 .0189471   -2.21   0.027   -.0790864  -.0048149
         ind |  .0136289  .015229    0.89   0.371   -.0162194   .0434771
TVendogenous |
         exp |  .1129704 .0024688   45.76   0.000    .1081316   .1178093
        exp2 | -.0004214 .0000546   -7.72   0.000   -.0005283  -.0003145
         wks |  .0008381 .0005995    1.40   0.162   -.0003368    .002013
          ms | -.0300894 .0189674   -1.59   0.113   -.0672649   .0070861
       union |  .0324752 .0148939    2.18   0.029    .0032837   .0616667
TIexogenous  |
         fem |  -.132008 .1266039   -1.04   0.297    -.380147   .1161311
         blk | -.2859004 .1554857   -1.84   0.066   -.5906468   .0188459
TIendogenous |
          ed |  .1372049 .0205695    6.67   0.000    .0968894   .1775205
             |
       _cons |  2.927338 .2751274   10.64   0.000    2.388098   3.466578
-------------+------------------------------------------------------------
     sigma_u |  .94180304
     sigma_e |  .15180273
         rho |  .97467788   (fraction of variance due to u_i)
--------------------------------------------------------------------------
note:  TV refers to time-varying; TI refers to time-invariant.
```

exogenous variables. A singly exogenous variable is correlated with the individual effects but not with the remainder noise. These are given the subscript (2). On the other hand, a doubly exogenous variable is uncorrelated with both the effects and the remainder disturbance term. These are given the subscript (1). Cornwell et al. extend the results of HT, AM and BMS by transforming each structural equation by its $\Omega^{-1/2}$ and applying 2SLS on the transformed equation using $A = [QX, PB]$ as the set of instruments in (7.47). B is defined as follows:

(1) $B_{\mathrm{HT}} = [X_{(1)}, Z_{(1)}]$ for the Hausman and Taylor (1981)-type estimator. This B_{HT} is the set of all *doubly* exogenous variables in the system.

(2) $B_{\text{AM}} = [X^*_{(1)}, Z_{(1)}]$ for the Amemiya and MaCurdy (1986)-type estimator. The (*) notation has been defined in (7.48).

(3) $B_{\text{BMS}} = [X^*_{(1)}, Z_{(1)}, (QX_{(2)})^*]$ for the Breusch et al. (1989)-type estimator. They also derive a similar set of instruments for the 3SLS analogue and give a generalized method of moments interpretation to these estimators. Finally, they consider the possibility of a different set of instruments for each equation, say $A_j = [QX, PB_j]$ for the jth equation, where for the HT-type estimator, B_j consists of all doubly exogenous variables of equation j (i.e. exogenous variables that are uncorrelated with the individual effects in equation j). Wyhowski (1994) extends the HT, AM and BMS approaches to the two-way error component model and gives the appropriate set of instruments. Revankar (1992) establishes conditions for exact equivalence of instrumental variables in a simultaneous two-way error component model.

Baltagi and Chang (2000) compare the performance of several single and system estimators of a two-equation simultaneous model with unbalanced panel data. The Monte Carlo design varies the degree of unbalancedness in the data and the variance components ratio due to the individual effects. Many of the results obtained for the simultaneous equation error component model with balanced data carry over to the unbalanced case. For example, both feasible EC2SLS estimators considered performed reasonably well and it is hard to choose between them. Simple ANOVA methods can still be used to obtain good estimates of the structural and reduced form parameters even in the unbalanced panel data case. Replacing negative estimates of the variance components by zero did not seriously affect the performance of the corresponding structural or reduced form estimates. Better estimates of the structural variance components do not necessarily imply better estimates of the structural coefficients. Finally, do not make the data balanced to simplify the computations. The loss in root mean squared error can be huge.

Most applied work in economics has made the choice between the RE and FE estimators based upon the standard Hausman (1978) test. This is based upon a contrast between the FE and RE estimators. If this standard Hausman test rejects the null hypothesis that the conditional mean of the disturbances given the regressors is zero, the applied researcher reports the FE estimator. Otherwise, the researcher reports the RE estimator, see the discussion in Chapter 4 and the two empirical applications by Owusu-Gyapong (1986) and Cardellichio (1990). Baltagi, Bresson and Pirotte (2003a) suggest an alternative *pre-test* estimator based on the Hausman and Taylor (1981) model. This pre-test estimator reverts to the RE estimator if the standard Hausman test based on the FE vs the RE estimators is not rejected. It reverts to the HT estimator if the choice of strictly exogenous regressors is not rejected by a second Hausman test based on the difference between the FE and HT estimators. Otherwise, this pre-test estimator reverts to the FE estimator. In other words, this pre-test alternative suggests that the researcher consider a Hausman–Taylor model where some of the variables, but not all, may be correlated with the individual effects. Monte Carlo experiments were performed to compare the performance of this pre-test estimator with the standard panel data estimators under various designs. The estimators considered were: OLS, fixed effects (FE), random effects (RE) and the Hausman–Taylor (HT) estimators. In one design, some regressors were correlated with the individual effects, i.e., a Hausman–Taylor world. In another design, the regressors were not allowed to be correlated with the individual effects, i.e., an RE world. Results showed that the pre-test estimator is a viable estimator and is always second best to the efficient estimator. It is second in RMSE performance to the RE estimator in an RE world and second to the HT estimator in an HT world. The FE estimator is a consistent estimator under both designs but its disadvantage is that it does not allow the estimation of the coefficients of the time-invariant regressors. When there is endogeneity among the regressors, Baltagi et al. (2003a) show that there is substantial bias in OLS and the RE estimators and both yield misleading inference. Even when there is no correlation between the individual effects and the regressors, i.e., in an RE world, inference based on OLS can be seriously misleading. This last result was emphasized by Moulton (1986).

NOTES

1. The analysis in this chapter can easily be extended to the two-way error component model; see the problems at the end of this chapter and Baltagi (1981b).
2. As in the classical regression case, the variances of W2SLS have to be adjusted by the factor $(NT - k_1 - g_1 + 1)/[N(T-1) - k_1 - g_1 + 1]$, whenever 2SLS is performed on the Within transformed equation (see Pliskin, 1991). Note also that the set of instruments is $\widetilde{X}$ and not X as emphasized in (7.6).
3. This is analogous to the result found in the single equation error component literature by Taylor (1980) and Baltagi (1981a).
4. Gardner (1998) shows how to modify the Hausman and Taylor (1981) instrumental variable estimator to allow for unbalanced panels. This utilizes the $\Omega^{-1/2}$ transformation derived for the unbalanced panel data model by Baltagi (1985), see equation (9.5), and the application of the IV interpretation of the HT estimator by Breusch et al. (1989) given above.
5. Im et al. (1999) point out that for panel data models, the exogeneity assumptions imply many more moment conditions than the standard random and fixed effects estimators use. Im et al. (1999) provide the assumptions under which the efficient GMM estimator based on the entire set of available moment conditions reduces to these simpler estimators. In other words, the efficiency of the simple estimators is established by showing the redundancy of the moment conditions that they do not use.

PROBLEMS

7.1 Verify that GLS on (7.7) yields (7.6) and GLS on (7.9) yields (7.8), the Within 2SLS and Between 2SLS estimators of δ_1, respectively.

7.2 Verify that GLS on (7.10) yields the EC2SLS estimator of δ_1 given in (7.11) (see Baltagi, 1981b).

7.3 Show that $A = [\widetilde{X}, \bar{X}]$, $B = [X^*, \widetilde{X}]$ and $C = [X^*, \bar{X}]$ yield the same projection, i.e. $P_A = P_B = P_C$ and hence the same EC2SLS estimator given by (7.11) (see Baltagi and Li, 1992c).

7.4 Verify that 3SLS on (7.21) with $(I_M \otimes \widetilde{X})$ as the set of instruments yields (7.22). Similarly, verify that 3SLS on (7.23) with $(I_M \otimes \bar{X})$ as the set of instruments yields (7.24). These are the Within 3SLS and Between 3SLS estimators of δ_1, respectively.

7.5 Verify that GLS on the stacked system (7.21) and (7.23) each premultiplied by $(I_M \otimes \widetilde{X}')$ and $(I_M \otimes \bar{X}')$, respectively, yields the EC3SLS estimator of δ given in (7.25) (see Baltagi, 1981b).

7.6 (a) Prove that $A = (I_M \otimes \widetilde{X}, I_M \otimes \bar{X})$ yields the same projection as $B = (H \otimes \widetilde{X}, G \otimes \bar{X})$ or $C = [(H \otimes \widetilde{X} + G \otimes \bar{X}), H \otimes \widetilde{X}]$ or $D = [H \otimes \widetilde{X} + G \otimes \bar{X}), G \otimes \bar{X}]$ where H and G are nonsingular $M \times M$ matrices (see Baltagi and Li, 1992c). Conclude that these sets of instruments yield the same EC3SLS estimator of δ given by (7.25).

(b) Let $H = \Sigma_\nu^{-1/2}$ and $G = \Sigma_1^{-1/2}$, and note that A is the set of instruments proposed by Baltagi (1981b) while B is the optimal set of instruments X^* defined below (7.30). Conclude that $H \otimes \widetilde{X}$ is redundant in C and $G \otimes \bar{X}$ is redundant in D with respect to the optimal set of instruments X^*.

7.7 (a) Consider a system of two structural equations with one-way error component disturbances. Show that if the disturbances between the two equations are uncorrelated, then EC3SLS is equivalent to EC2SLS (see Baltagi, 1981b).
(b) Show that if this system of two equations with one-way error component disturbances is just-identified, then EC3SLS does not necessarily reduce to EC2SLS (see Baltagi, 1981b).

7.8 (a) Using partitioned inverse, show that GLS on (7.35) yields $\widehat{\beta}_{\text{GLS}} = \widetilde{\beta}_{\text{Within}}$ and $\widehat{\pi}_{\text{GLS}} = \widehat{\beta}_{\text{Between}} - \widetilde{\beta}_{\text{Within}}$ as given in (7.37) and (7.38).
(b) Verify that $\text{var}(\widehat{\pi}_{\text{GLS}}) = \text{var}(\widehat{\beta}_{\text{Between}}) + \text{var}(\widetilde{\beta}_{\text{Within}})$ as given in (7.39).

7.9 For the two-way error component model given in (6.9) and the covariance matrix Ω_{jl} between the jth and lth equation disturbances given in (6.11):
(a) Derive the EC2SLS estimator for δ_1 in (7.1).
(b) Derive the EC3SLS estimator for δ in (7.20) (*Hint*: See Baltagi, 1981b).
(c) Repeat problem 7.7 parts (a) and (b) for the two-way error component EC2SLS and EC3SLS.

7.10 Using the Monte Carlo set-up for a two-equation simultaneous model with error component disturbances, given in Baltagi (1984), compare EC2SLS and EC3SLS with the usual 2SLS and 3SLS estimators that ignore the error component structure.

7.11 Using the Cornwell and Trumbull (1994) panel data set described in the empirical example in section 7.1 and given on the Wiley web site as crime.dat, replicate Table 7.1 and the associated test statistics.

7.12 Using the Cornwell and Rupert (1988) panel data set described in the empirical example in section 7.4 and given on the Wiley web site as wage.xls, replicate Table 7.4 and the associated test statistics.

7.13 *A Hausman test based on the difference between fixed effects two-stage least squares and error components two-stage least squares*. This is based on problem 04.1.1 in *Econometric Theory* by Baltagi (2004). Consider the first structural equation of a simultaneous equation panel data model given in (7.1). Hausman (1978) suggests comparing the FE and RE estimators in the classic panel data regression. With endogenous right-hand side regressors like Y_1 this test can be generalized to test $H_0 : E(u_1 \mid Z_1) = 0$ based on $\widehat{q}_1 = \widetilde{\delta}_{1,\text{FE2SLS}} - \widehat{\delta}_{1,\text{EC2SLS}}$ where $\widetilde{\delta}_{1,\text{FE2SLS}}$ is defined in (7.6) and $\widehat{\delta}_{1,\text{EC2SLS}}$ is defined in (7.11).
(a) Show that under $H_0 : E(u_1 \mid Z_1) = 0$, plim $\widehat{q}_1 = 0$ and the asymptotic $\text{cov}(\widehat{q}_1, \widehat{\delta}_{1,\text{EC2SLS}}) = 0$.
(b) Conclude that $\text{var}(\widehat{q}_1) = \text{var}(\widetilde{\delta}_{1,\text{FE2SLS}}) - \text{var}(\widehat{\delta}_{1,\text{EC2SLS}})$, where var denotes the asymptotic variance. This is used in computing the Hausman test statistic given by $m_1 = \hat{q}_1'[\text{var}(\widehat{q}_1)]^{-1}\widehat{q}_1$. Under H_0, m_1 is asymptotically distributed as χ^2_r, where r denotes the dimension of the slope vector of the time-varying variables in Z_1. This can easily be implemented using Stata.
(c) Compute the usual Hausman test based on FE and RE and this alternative Hausman test based on FE2SLS and EC2SLS for the crime data considered in problem 7.12. What do you conclude?

8 Dynamic Panel Data Models

8.1 INTRODUCTION

Many economic relationships are dynamic in nature and one of the advantages of panel data is that they allow the researcher to better understand the dynamics of adjustment. See, for example, Balestra and Nerlove (1966) on dynamic demand for natural gas, Baltagi and Levin (1986) on dynamic demand for an addictive commodity like cigarettes, Holtz-Eakin (1988) on a dynamic wage equation, Arellano and Bond (1991) on a dynamic model of employment, Blundell et al. (1992) on a dynamic model of company investment, Islam (1995) on a dynamic model for growth convergence, and Ziliak (1997) on a dynamic lifecycle labor supply model. These dynamic relationships are characterized by the presence of a lagged dependent variable among the regressors, i.e.

$$y_{it} = \delta y_{i,t-1} + x'_{it}\beta + u_{it} \quad i = 1, \ldots, N;\ t = 1, \ldots, T \tag{8.1}$$

where δ is a scalar, x'_{it} is $1 \times K$ and β is $K \times 1$. We will assume that the u_{it} follow a one-way error component model

$$u_{it} = \mu_i + \nu_{it} \tag{8.2}$$

where $\mu_i \sim \text{IID}(0, \sigma^2_\mu)$ and $\nu_{it} \sim \text{IID}(0, \sigma^2_\nu)$ independent of each other and among themselves. The dynamic panel data regression described in (8.1) and (8.2) is characterized by two sources of persistence over time. Autocorrelation due to the presence of a lagged dependent variable among the regressors and individual effects characterizing the heterogeneity among the individuals. In this chapter, we review some of the recent econometric studies that propose new estimation and testing procedures for this model.

Let us start with some of the basic problems introduced by the inclusion of a lagged dependent variable. Since y_{it} is a function of μ_i, it immediately follows that $y_{i,t-1}$ is also a function of μ_i. Therefore, $y_{i,t-1}$, a right-hand regressor in (8.1), is correlated with the error term. This renders the OLS estimator biased and inconsistent even if the ν_{it} are not serially correlated. See Sevestre and Trognon (1985) for the magnitude of this asymptotic bias in dynamic error component models. For the fixed effects (FE) estimator, the Within transformation wipes out the μ_i (see Chapter 2), but $(y_{i,t-1} - \bar{y}_{i.-1})$ where $\bar{y}_{i.-1} = \sum_{t=2}^{T} y_{i,t-1}/(T-1)$ will still be correlated with $(\nu_{it} - \bar{\nu}_{i.})$ even if the ν_{it} are not serially correlated. This is because $y_{i,t-1}$ is correlated with $\bar{\nu}_{i.}$ by construction. The latter average contains $\nu_{i,t-1}$ which is obviously correlated with $y_{i,t-1}$. In fact, the Within estimator will be biased of $O(1/T)$ and its consistency will depend upon T being large; see Nickell (1981). More recently, Kiviet (1995) derives an approximation for the bias of the Within estimator in a dynamic panel data model with serially uncorrelated disturbances and strongly exogenous regressors. Kiviet (1995) proposed a corrected Within estimator that subtracts a consistent estimator of this bias from the original Within estimator.[1] Therefore, for the typical labor panel where N is large and T is fixed, the Within estimator is biased and inconsistent. It is worth emphasizing that only if $T \to \infty$ will the Within estimator of δ and β be consistent for the dynamic error component model. For macro panels, studying for example long-run growth, the data covers a large number of countries N over a moderate size T, see Islam (1995). In this case, T is not very small relative to N. Hence, some researchers may still favor the Within estimator arguing that its bias may not be large. Judson and Owen (1999) performed some Monte Carlo experiments for $N = 20$ or 100 and $T = 5, 10, 20$ and 30 and found that the bias in the Within estimator can be sizeable, even when $T = 30$. This bias increases with δ and decreases with T. But even for $T = 30$, this bias could be as much as 20% of the true value of the coefficient of interest.[2]

The random effects GLS estimator is also biased in a dynamic panel data model. In order to apply GLS, quasi-demeaning is performed (see Chapter 2), and $(y_{i,t-1} - \theta \bar{y}_{i.,-1})$ will be correlated with $(u_{i,t} - \theta \bar{u}_{i.,-1})$. An alternative transformation that wipes out the individual effects is the first difference (FD) transformation. In this case, correlation between the predetermined explanatory variables and the remainder error is easier to handle. In fact, Anderson and Hsiao (1981) suggested first differencing the model to get rid of the μ_i and then using $\Delta y_{i,t-2} = (y_{i,t-2} - y_{i,t-3})$ or simply $y_{i,t-2}$ as an instrument for $\Delta y_{i,t-1} = (y_{i,t-1} - y_{i,t-2})$. These instruments will not be correlated with $\Delta \nu_{it} = \nu_{i,t} - \nu_{i,t-1}$, as long as the ν_{it} themselves are not serially correlated. This instrumental variable (IV) estimation method leads to consistent but not necessarily efficient estimates of the parameters in the model because it does not make use of all the available moment conditions (see Ahn and Schmidt, 1995), and it does not take into account the differenced structure on the residual disturbances ($\Delta \nu_{it}$). Arellano (1989) finds that for simple dynamic error components models, the estimator that uses differences $\Delta y_{i,t-2}$ rather than levels $y_{i,t-2}$ for instruments has a singularity point and very large variances over a significant range of parameter values. In contrast, the estimator that uses instruments in levels, i.e. $y_{i,t-2}$, has no singularities and much smaller variances and is therefore recommended. Arellano and Bond (1991) proposed a generalized method of moments (GMM) procedure that is more efficient than the Anderson and Hsiao (1982) estimator, while Ahn and Schmidt (1995) derived additional nonlinear moment restrictions not exploited by the Arellano and Bond (1991) GMM estimator. This literature is generalized and extended by Arellano and Bover (1995) and Blundell and Bond (1998) to mention a few. In addition, an alternative method of estimation of the dynamic panel data model is proposed by Keane and Runkle (1992). This is based on the forward filtering idea in time-series analysis. We focus on these studies and describe their respective contributions to the estimation and testing of dynamic panel data models. This chapter concludes with recent developments and some applications.

8.2 THE ARELLANO AND BOND ESTIMATOR

Arellano and Bond (1991) argue that additional instruments can be obtained in a dynamic panel data model if one utilizes the orthogonality conditions that exist between lagged values of y_{it} and the disturbances ν_{it}. Let us illustrate this with the simple autoregressive model with no regressors:

$$y_{it} = \delta y_{i,t-1} + u_{it} \quad i = 1, \dots, N;\ t = 1, \dots, T \tag{8.3}$$

where $u_{it} = \mu_i + \nu_{it}$ with $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$ and $\nu_{it} \sim \text{IID}(0, \sigma_\nu^2)$, independent of each other and among themselves. In order to get a consistent estimate of δ as $N \to \infty$ with T fixed, we first difference (8.3) to eliminate the individual effects

$$y_{it} - y_{i,t-1} = \delta(y_{i,t-1} - y_{i,t-2}) + (\nu_{it} - \nu_{i,t-1}) \tag{8.4}$$

and note that $(\nu_{it} - \nu_{i,t-1})$ is MA(1) with unit root. For $t = 3$, the first period we observe this relationship, we have

$$y_{i3} - y_{i2} = \delta(y_{i2} - y_{i1}) + (\nu_{i3} - \nu_{i2})$$

In this case, y_{i1} is a valid instrument, since it is highly correlated with $(y_{i2} - y_{i1})$ and not correlated with $(\nu_{i3} - \nu_{i2})$ as long as the ν_{it} are not serially correlated. But note what happens for $t = 4$, the second period we observe (8.4):

$$y_{i4} - y_{i3} = \delta(y_{i3} - y_{i2}) + (\nu_{i4} - \nu_{i3})$$

In this case, y_{i2} as well as y_{i1} are valid instruments for $(y_{i3} - y_{i2})$, since both y_{i2} and y_{i1} are not correlated with $(\nu_{i4} - \nu_{i3})$. One can continue in this fashion, adding an extra valid instrument with each forward period, so that for period T, the set of valid instruments becomes $(y_{i1}, y_{i2}, \ldots, y_{i,T-2})$.

This instrumental variable procedure still does not account for the differenced error term in (8.4). In fact,

$$E(\Delta\nu_i\ \Delta\nu_i') = \sigma_\nu^2(I_N \otimes G) \tag{8.5}$$

where $\Delta\nu_i' = (\nu_{i3} - \nu_{i2}, \ldots, \nu_{iT} - \nu_{i,T-1})$ and

$$G = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 2 \end{pmatrix}$$

is $(T-2) \times (T-2)$, since $\Delta\nu_i$ is MA(1) with unit root. Define

$$W_i = \begin{bmatrix} [y_{i1}] & & & 0 \\ & [y_{i1}, y_{i2}] & & \\ & & \ddots & \\ 0 & & & [y_{i1}, \ldots, y_{i,T-2}] \end{bmatrix} \tag{8.6}$$

Then, the matrix of instruments is $W = [W_1', \ldots, W_N']'$ and the moment equations described above are given by $E(W_i'\Delta\nu_i) = 0$. These moment conditions have also been pointed out by Holtz-Eakin (1988), Holtz-Eakin, Newey and Rosen (1988) and Ahn and Schmidt (1995). Premultiplying the differenced equation (8.4) in vector form by W', one gets

$$W'\Delta y = W'(\Delta y_{-1})\delta + W'\Delta\nu \tag{8.7}$$

Performing GLS on (8.7) one gets the Arellano and Bond (1991) preliminary one-step consistent estimator

$$\begin{aligned}\widehat{\delta}_1 = {} & [(\Delta y_{-1})'W(W'(I_N \otimes G)W)^{-1}W'(\Delta y_{-1})]^{-1} \\ & \times[(\Delta y_{-1})'W(W'(I_N \otimes G)W)^{-1}W'(\Delta y)]\end{aligned} \tag{8.8}$$

The optimal GMM estimator of δ_1 à la Hansen (1982) for $N \to \infty$ and T fixed using only the above moment restrictions yields the same expression as in (8.8) except that

$$W'(I_N \otimes G)W = \sum_{i=1}^{N} W_i' G W_i$$

is replaced by

$$V_N = \sum_{i=1}^{N} W_i'(\Delta \nu_i)(\Delta \nu_i)' W_i$$

This GMM estimator requires no knowledge concerning the initial conditions or the distributions of ν_i and μ_i. To operationalize this estimator, $\Delta\nu$ is replaced by differenced residuals obtained from the preliminary consistent estimator $\widehat{\delta}_1$. The resulting estimator is the two-step Arellano and Bond (1991) GMM estimator:

$$\widehat{\delta}_2 = [(\Delta y_{-1})' W \widehat{V}_N^{-1} W'(\Delta y_{-1})]^{-1}[(\Delta y_{-1})' W \widehat{V}_N^{-1} W'(\Delta y)] \tag{8.9}$$

A consistent estimate of the asymptotic var$(\widehat{\delta}_2)$ is given by the first term in (8.9),

$$\widehat{\text{var}}(\widehat{\delta}_2) = [(\Delta y_{-1})' W \widehat{V}_N^{-1} W'(\Delta y_{-1})]^{-1} \tag{8.10}$$

Note that $\widehat{\delta}_1$ and $\widehat{\delta}_2$ are asymptotically equivalent if the ν_{it} are IID$(0, \sigma_\nu^2)$.

8.2.1 Testing for Individual Effects in Autoregressive Models

Holtz-Eakin (1988) derives a simple test for the presence of individual effects in dynamic panel data models. The basic idea of the test can be explained using the simple autoregressive model given in (8.3). Assume there are only three periods, i.e. $T = 3$. Then (8.3) can be estimated using the last two periods. Under the null hypothesis of no individual effects, the following orthogonality conditions hold:

$$E(y_{i,2}u_{i,3}) = 0 \qquad E(y_{i,1}u_{i,3}) = 0 \qquad E(y_{i,1}u_{i,2}) = 0$$

Three conditions to identify one parameter, the remaining two over-identifying restrictions can be used to test for individual effects. We can reformulate these orthogonality restrictions as follows:

$$E[(y_{i,1}(u_{i,3} - u_{i,2})] = 0 \tag{8.11a}$$

$$E(y_{i,1}u_{i,2}) = 0 \tag{8.11b}$$

$$E(y_{i,2}u_{i,3}) = 0 \tag{8.11c}$$

The first restriction can be used to identify δ even if there are individual effects in (8.3). The null hypothesis of no individual effects imposes only two additional restrictions (8.11b) and (8.11c) on the data. Intuitively, the test for individual effects is a test of whether the sample moments corresponding to these restrictions are sufficiently close to zero; contingent upon imposing (8.11a) to identify δ.

Stacking the following equations:

$$\begin{aligned}(y_3 - y_2) &= (y_2 - y_1)\delta + (u_3 - u_2)\\ y_3 &= y_2\delta + u_3\\ y_2 &= y_1\delta + u_2\end{aligned}$$

we can write

$$y^* = Y^*\delta + u^* \tag{8.12}$$

where $y^{*\prime} = (y_3' - y_2', y_3', y_2')$, $Y^{*\prime} = (y_2' - y_1', y_2', y_1')$ and $u^{*\prime} = (u_3' - u_2', u_3', u_2')$. Holtz-Eakin (1988) estimates this system of simultaneous equations with different instrumental variables for each equation. This is due to the dynamic nature of these equations. Variables which qualify for use as IVs in one period may not qualify in earlier periods. Let $W = \text{diag}[W_i]$ for $i = 1, 2, 3$ be the matrix of instruments such that $\text{plim}(W'u^*/N) = 0$ as $N \to \infty$. Perform GLS on (8.12) after premultiplying by W'. In this case, $\Omega = W'E(u^*u^{*\prime})W$ is estimated by $\widehat{\Omega} = (\sum_{i=1}^{N} \widehat{u}_{i,r}^* \widehat{u}_{i,s}^* W_{i,r}' W_{i,s})$ where $\widehat{u}_r^*$ denotes 2SLS residuals on each equation separately,

$$\widehat{\delta} = [Y^{*\prime} W \widehat{\Omega}^{-1} W' Y^*]^{-1} Y^{*\prime} W \widehat{\Omega}^{-1} W' y^*$$

Let SSQ be the weighted sum of the squared transformed residuals:

$$\text{SSQ} = (y^* - Y^*\widehat{\delta})' W \widehat{\Omega}^{-1} W' (y^* - Y^*\widehat{\delta})/N$$

This has χ^2 distribution with degrees of freedom equal to the number of over-identifying restrictions as N grows. Compute $L = \text{SSQ}_R - \text{SSW}$ where SSQ_R is the sum of squared residuals when imposing the full set of orthogonality conditions implied by the null hypothesis, SSW is the sum of squared residuals that impose only those restrictions needed for the first-differenced version. The same estimate of Ω should be used in both computations, and Ω should be estimated under the null. Holtz-Eakin generalizes this to an AR(p) where p is unknown and applies this test to a dynamic wage equation based on a subsample of 898 males from the Panel Study of Income Dynamics (PSID) observed over the years 1968–81. He finds evidence of individual effects and thus support for controlling heterogeneity in estimating a dynamic wage equation.

Recently, Jimenez-Martin (1998) performed Monte Carlo experiments to study the performance of the Holtz-Eakin (1988) test for the presence of individual heterogeneity effects in dynamic small T unbalanced panel data models. The design of the experiment included both endogenous and time-invariant regressors in addition to the lagged dependent variable. The test behaved correctly for a moderate autoregressive coefficient. However, when this autoregressive coefficient approached unity, the presence of an additional regressor sharply affected the power and the size of the test. The results of the Monte Carlo show that the power of this test is higher when the variance of the specific effects increases (they are easier to detect), when the sample size increases, when the data set is balanced (for a given number of cross-section units) and when the regressors are strictly exogenous.

8.2.2 Models with Exogenous Variables

If there are additional *strictly exogenous* regressors x_{it} as in (8.1) with $E(x_{it}\nu_{is}) = 0$ for all $t, s = 1, 2, \ldots, T$, but where all the x_{it} are correlated with μ_i, then all the x_{it} are valid instruments for the first-differenced equation of (8.1). Therefore, $[x_{i1}', x_{i2}', \ldots, x_{iT}']$ should be added to each diagonal element of W_i in (8.6). In this case, (8.7) becomes

$$W'\Delta y = W'(\Delta y_{-1})\delta + W'(\Delta X)\beta + W'\Delta\nu$$

where ΔX is the stacked $N(T-2) \times K$ matrix of observations on Δx_{it}. One- and two-step estimators of (δ, β') can be obtained from

$$\begin{pmatrix} \widehat{\delta} \\ \widehat{\beta} \end{pmatrix} = ([\Delta y_{-1}, \Delta X]' W \widehat{V}_N^{-1} W' [\Delta y_{-1}, \Delta X])^{-1} ([\Delta y_{-1}, \Delta X]' W \widehat{V}_N^{-1} W' \Delta y) \quad (8.13)$$

as in (8.8) and (8.9).

If x_{it} are *predetermined* rather than *strictly exogenous* with $E(x_{it}\nu_{is}) \neq 0$ for $s < t$, and zero otherwise, then only $[x'_{i1}, x'_{i2}, \ldots, x'_{i(s-1)}]$ are valid instruments for the differenced equation at period s. This can be illustrated as follows: for $t = 3$, the first-differenced equation of (8.1) becomes

$$y_{i3} - y_{i2} = \delta(y_{i2} - y_{i1}) + (x'_{i3} - x'_{i2})\beta + (\nu_{i3} - \nu_{i2})$$

For this equation, x'_{i1} and x'_{i2} are valid instruments, since both are not correlated with $(\nu_{i3} - \nu_{i2})$. For $t = 4$, the next period we observe this relationship,

$$y_{i4} - y_{i3} = \delta(y_{i3} - y_{i2}) + (x'_{i4} - x'_{i3})\beta + (\nu_{i4} - \nu_{i3})$$

and we have additional instruments since now x'_{i1}, x'_{i2} and x'_{i3} are not correlated with $(\nu_{i4} - \nu_{i3})$. Continuing in this fashion, we get

$$W_i = \begin{bmatrix} [y_{i1}, x'_{i1}, x'_{i2}] & & & 0 \\ & [y_{i1}, y_{i2}, x'_{i1}, x'_{i2}, x'_{i3}] & & \\ & & \ddots & \\ 0 & & & [y_{i1}, \ldots, y_{i,T-2}, x'_{i1}, \ldots, x'_{i,T-1}] \end{bmatrix} \quad (8.14)$$

and one- and two-step estimators are again given by (8.13) with this choice of W_i.

In empirical studies, a combination of both *predetermined* and *strictly exogenous* variables may occur rather than the above two extreme cases, and the researcher can adjust the matrix of instruments W accordingly. Also, not all the x_{it} have to be correlated with μ_i. As in Hausman and Taylor (1981), we can separate $x_{it} = [x_{1it}, x_{2it}]$ where x_{1it} is uncorrelated with μ_i, while x_{2it} is correlated with μ_i. For the predetermined x_{it} case, Arellano and Bond (1991) count T additional restrictions from the level equations (8.1), i.e. $E(u_{i2}x_{1i1}) = 0$ and $E(u_{it}x_{1it}) = 0$ for $t = 2, \ldots, T$. All additional linear restrictions from the level equations are redundant given those already exploited from the first-differenced equations. Define $u_i = (u_{i2}, \ldots, u_{iT})'$ and $\nu_i^+ = (\Delta\nu'_i, u'_i)'$, where we stack the differenced disturbances from period $t = 3$ to $t = T$ on top of the undifferenced disturbances from period $t = 2$ to $t = T$. Now, let

$$\nu^+ = y^+ - y^+_{-1}\delta - X^+\beta \quad (8.15)$$

with $\nu^+ = (\nu_1^{+\prime}, \ldots, \nu_N^{+\prime})'$ and y^+, y^+_{-1} and X^+ defined similarly. The optimal matrix of instruments becomes

$$W_i^+ = \begin{bmatrix} W_i & & & & 0 \\ & [x'_{1i1}, x'_{1i2}] & & & \\ & & x'_{1i3} & & \\ & & & \ddots & \\ 0 & & & & x'_{1iT} \end{bmatrix} \quad (8.16)$$

where W_i is given by (8.14). The two-step estimator is of the same form as (8.13) with y^+, y_{-1}^+, X^+ and W^+ replacing Δy, Δy_{-1}, ΔX and W, respectively.

If x_{1it} is *strictly exogenous*, the observations for all periods become valid instruments in the level equations. However, given those previously exploited in first differences we only have T extra restrictions which Arellano and Bond (1991) express as $E(\sum_{s=1}^{T} x_{1it}u_{is}/T) = 0$ for $t = 1, \ldots, T$. Thus, the two-step estimator would just combine the $(T-1)$ first-differenced equations and the average level equation.

Arellano and Bond (1991) propose a test for the hypothesis that there is no second-order serial correlation for the disturbances of the first-differenced equation. This test is important because the consistency of the GMM estimator relies upon the fact that $E[\Delta\nu_{it}\Delta\nu_{i,t-2}] = 0$. The test statistic is given in equation (8) of Arellano and Bond (1991, p. 282) and will not be reproduced here. This hypothesis is true if the ν_{it} are not serially correlated or follow a random walk. Under the latter situation, both OLS and GMM of the first-differenced version of (8.1) are consistent and Arellano and Bond (1991) suggest a Hausman-type test based on the difference between the two estimators.

Additionally, Arellano and Bond (1991) suggest Sargan's test of over-identifying restrictions given by

$$m = \Delta\widehat{\nu}' W \left[\sum_{i=1}^{N} W_i'(\Delta\widehat{\nu}_i)(\Delta\widehat{\nu}_i)' W_i\right]^{-1} W'(\Delta\widehat{\nu}) \sim \chi^2_{p-K-1}$$

where p refers to the number of columns of W and $\Delta\widehat{\nu}$ denotes the residuals from a two-step estimation given in (8.13).[3] Other tests suggested are Sargan's difference statistic to test nested hypotheses concerning serial correlation in a sequential way, or a Griliches and Hausman (1986)-type test based on the difference between the two-step GMM estimators assuming the disturbances in levels are MA(0) and MA(1), respectively. These are described in more detail in Arellano and Bond (1991, p. 283).

A limited Monte Carlo study was performed based on 100 replications from a simple autoregressive model with one regressor and no constant, i.e. $y_{it} = \delta y_{i,t-1} + \beta x_{it} + \mu_i + \nu_{it}$ with $N = 100$ and $T = 7$. The results showed that the GMM estimators have negligible finite sample biases and substantially smaller variances than those associated with simpler IV estimators à la Anderson and Hsiao (1981). However, the estimated standard error of the two-step GMM estimator was found to be downward biased. The tests proposed above also performed reasonably well. These estimation and testing methods were applied to a model of employment using a panel of 140 quoted UK companies for the period 1979–84. This is the benchmark data set used in Stata to obtain the one-step and two-step estimators described in (8.8) and (8.10) as well as the Sargan test for over-identification using the command (xtabond,twostep), see problem 8.9.

Windmeijer (2005) attributes the small sample downward bias of the estimated asymptotic standard errors of the two-step efficient GMM estimator to the estimation of the weight matrix W. He suggests a correction term based on a Taylor series expansion that accounts for the estimation of W. He shows that this correction term provides a more accurate approximation in finite samples when all the moment conditions are linear. These corrected standard errors are available using xtabond2 in Stata.

Using Monte Carlo experiments, Bowsher (2002) finds that the use of too many moment conditions causes the Sargan test for over-identifying restrictions to be undersized and have extremely low power. Fixing N at 100, and letting T increase over the range (5, 7, 9, 11, 13, 15), the performance of Sargan's test using the full set of Arellano–Bond moment conditions is examined for $\delta = 0.4$. For $T = 5$, the Monte Carlo mean of the Sargan χ^2_5 statistic is 5.12 when it should be 5, and its Monte Carlo variance is 9.84 when it should be 10. The size of the test is 0.052 at the 5% level and the power under the alternative is 0.742. For $T = 15$, the Sargan χ^2_{90} statistic has a Monte Carlo mean of 91.3 when its theoretical mean is 90. However, its Monte Carlo variance is 13.7 when it should be 180. This underestimation of the theoretical variance results in zero rejection rate under the null and alternative. In general, the Monte Carlo mean is a good estimator of the mean of the asymptotic χ^2 statistic. However, the Monte Carlo variance is much smaller than its asymptotic counterpart when T is large. The Sargan test never rejects when T is too large for a given N. Zero rejection rates under the null and alternative were also observed for the following (N, T) pairs: (125, 16), (85, 13), (70, 112), and (40, 10). This is attributed to poor estimates of the weighting matrix in GMM rather than to weak instruments.

Another application of the Arellano and Bond GMM estimator is given by Blundell et al. (1992), who used a panel of 532 UK manufacturing companies over the period 1975–86 to determine the importance of Tobin's Q in the determination of investment decisions. Tobin's Q was allowed to be endogenous and possibly correlated with the firm-specific effects. A GMM-type estimator was utilized using past variables as instruments, and Tobin's Q effect was found to be small but significant. These results were sensitive to the choice of dynamic specification, exogeneity assumptions and measurement error in Q. Similar findings using Tobin's Q model were reported by Hayashi and Inoue (1991) based on a panel of 687 quoted Japanese manufacturing firms over the period 1977–86.

8.3 THE ARELLANO AND BOVER ESTIMATOR

Arellano and Bover (1995) develop a unifying GMM framework for looking at efficient IV estimators for dynamic panel data models. They do that in the context of the Hausman and Taylor (1981) model given in (7.40), which in static form is reproduced here for convenience:

$$y_{it} = x'_{it}\beta + Z'_i\gamma + u_{it} \tag{8.17}$$

where β is $K \times 1$ and γ is $g \times 1$. The Z_i are time-invariant variables whereas the x_{it} vary over individuals and time. In vector form, (8.17) can be written as

$$y_i = W_i\eta + u_i \tag{8.18}$$

with the disturbances following a one-way error component model

$$u_i = \mu_i\iota_T + \nu_i \tag{8.19}$$

where $y_i = (y_{i1}, \ldots, y_{iT})'$, $u_i = (u_{i1}, \ldots, u_{iT})'$, $\eta' = (\beta', \gamma')$, $W_i = [X_i, \iota_T Z'_i]$, $X_i = (x_{i1}, \ldots, x_{iT})'$ and ι_T is a vector of ones of dimension T. In general, $E(u_iu'_i/w_i)$ will be unrestricted depending on $w_i = (x'_i, Z'_i)'$ where $x_i = (x'_{i1}, \ldots, x'_{iT})'$. However, the literature emphasizes two cases with cross-sectional homoskedasticity:

Case 1. $E(u_iu'_i) = \Omega$ independent of w_i, but general to allow for arbitrary Ω as long as it is the same across individuals, i.e. Ω is the same for $i = 1, \ldots, N$

Case 2. the traditional error component model where $\Omega = \sigma^2_\nu I_T + \sigma^2_\mu \iota_T\iota'_T$.

Arellano and Bover transform the system of T equations in (8.18) using the nonsingular transformation

$$H = \begin{bmatrix} C \\ \iota_T'/T \end{bmatrix} \tag{8.20}$$

where C is any $(T-1) \times T$ matrix of rank $(T-1)$ such that $C\iota_T = 0$. For example, C could be the first $(T-1)$ rows of the Within group operator or the first difference operator.[4] Note that the transformed disturbances

$$u_i^+ = Hu_i = \begin{bmatrix} Cu_i \\ \bar{u}_i \end{bmatrix} \tag{8.21}$$

have the first $(T-1)$ transformed errors free of μ_i. Hence, all exogenous variables are valid instruments for these first $(T-1)$ equations. Let m_i denote the subset of variables of w_i assumed to be uncorrelated in levels with μ_i and such that the dimension of m_i is greater than or equal to the dimension of η. In the Hausman and Taylor study, $X = [X_1, X_2]$ and $Z = [Z_1, Z_2]$ where X_1 and Z_1 are exogenous of dimension $NT \times k_1$ and $N \times g_1$. X_2 and Z_2 are correlated with the individual effects and are of dimension $NT \times k_2$ and $N \times g_2$. In this case, m_i includes the set of X_1 and Z_1 variables and m_i would be based on $(Z_{1,i}', x_{1,i1}', \ldots, x_{1iT}')'$. Therefore, a valid IV matrix for the complete transformed system is

$$M_i = \begin{bmatrix} w_i' & & & 0 \\ & \ddots & & \\ & & w_i' & \\ 0 & & & m_i' \end{bmatrix} \tag{8.22}$$

and the moment conditions are given by

$$E(M_i' H u_i) = 0 \tag{8.23}$$

Defining $W = (W_1', \ldots, W_N')'$, $y = (y_1', \ldots, y_N')'$, $M = (M_1', \ldots, M_N')'$, $\bar{H} = I_N \otimes H$ and $\bar{\Omega} = I_N \otimes \Omega$, and premultiplying (8.18) in vector form by $M'\bar{H}$ one gets

$$M'\bar{H}y = M'\bar{H}W\eta + M'\bar{H}u \tag{8.24}$$

Performing GLS on (8.24) one gets the Arellano and Bover (1995) estimator

$$\widehat{\eta} = [W'\bar{H}'M(M'\bar{H}\bar{\Omega}\bar{H}'M)^{-1}M'\bar{H}W]^{-1}W'\bar{H}'M(M'\bar{H}\bar{\Omega}\bar{H}'M)^{-1}M'\bar{H}y \tag{8.25}$$

In practice, the covariance matrix of the transformed system $\Omega^+ = H\Omega H'$ is replaced by a consistent estimator, usually

$$\widehat{\Omega}^+ = \sum_{i=1}^{N} \widehat{u}_i^+ \widehat{u}_i^{+\prime}/N \tag{8.26}$$

where $\widehat{u}_i^+$ are residuals based on consistent preliminary estimates. The resulting $\widehat{\eta}$ is the optimal GMM estimator of η with constant Ω based on the above moment restrictions. Further efficiency can be achieved using Chamberlain's (1982) or Hansen's (1982) GMM-type estimator which replaces $(\sum_i M_i'\Omega^+ M_i)$ in (8.25) by $(\sum_i M_i'\widehat{u}_i^+\widehat{u}_i^{+\prime}M_i)$. For the error component model, $\widetilde{\Omega}^+ = H\widetilde{\Omega}H'$ with $\widetilde{\Omega} = \widetilde{\sigma}_\nu^2 I_T + \widetilde{\sigma}_\mu^2 \iota_T\iota_T'$, where $\widetilde{\sigma}_\nu^2$ and $\widetilde{\sigma}_\mu^2$ denote consistent estimates σ_ν^2 and σ_μ^2.

The Hausman and Taylor (1981) (HT) estimator, given in section 7.3, is $\widehat{\eta}$ with $\widetilde{\Omega}^+$ and $m_i = (Z_{1,i}', \bar{x}_{1,i}')'$ where $\bar{x}_i' = \iota_T' X_i/T = (\bar{x}_{1,i}', \bar{x}_{2,i}')$. The Amemiya and MaCurdy (1986) (AM) estimator is $\widehat{\eta}$ with $\widetilde{\Omega}^{+}$ and $m_i = (Z'_{1i}, x'_{1,i1}, \ldots, x'_{1,iT})'$. The Breusch et al. (1989) (BMS) estimator exploits the additional moment restrictions that the correlation between $x_{2,it}$, and μ_i is constant over time. In this case, $\widetilde{x}_{2,it} = x_{2,it} - \bar{x}_{2,i}$ are valid instruments for the last equation of the transformed system. Hence, BMS is $\widehat{\eta}$ with $\widetilde{\Omega}^{+}$ and $m_i = (Z'_{1,i}, x'_{1,i1}, \ldots, x'_{1,iT}, \widetilde{x}'_{2,i1}, \ldots, \widetilde{x}'_{2,iT})'$.

Because the set of instruments M_i is block-diagonal, Arellano and Bover show that $\widehat{\eta}$ is invariant to the choice of C. Another advantage of their representation is that the form of $\Omega^{-1/2}$ need not be known. Hence, this approach generalizes the HT, AM, BMS-type estimators to a more general form of Ω than that of error components, and it easily extends to the dynamic panel data case as can be seen next.[5]

Let us now introduce a lagged dependent variable into the right-hand side of (8.17):

$$y_{it} = \delta y_{i,t-1} + x'_{it}\beta + Z'_i\gamma + u_{it} \tag{8.27}$$

Assuming that $t = 0$ is observed, we redefine $\eta' = (\delta, \beta', \gamma')$ and $W_i = [y_{i(-1)}, X_i, \iota_T Z'_i]$ with $y_{i(-1)} = (y_{i,0}, \ldots, y_{i,T-1})'$. Provided there are enough valid instruments to ensure identification, the GMM estimator defined in (8.25) remains consistent for this model. The matrix of instruments M_i is the same as before, adjusting for the fact that $t = 0$ is now the first period observed, so that $w_i = [x'_{i0}, \ldots, x'_{iT}, Z'_i]'$. In this case $y_{i(-1)}$ is excluded despite its presence in W_i. The same range of choices for m_i are available, for example, $m_i = (Z'_{1i}, x'_{1i}, \widetilde{x}'_{2,i1}, \ldots \widetilde{x}'_{2,iT})$ is the BMS-type estimator. However, for this choice of m_i the rows of CX_i are linear combinations of m_i. This means that the same instrument set is valid for all equations and we can use $M_i = I_T \otimes m'_i$ without altering the estimator. The consequence is that the transformation is unnecessary and the estimator can be obtained by applying 3SLS to the original system of equations using m_i as the vector of instruments for all equations:

$$\begin{aligned}\widehat{\eta} = &\left[\sum_i (W_i \otimes m_i)' \left(\widehat{\Omega} \otimes \sum_i m_i m'_i\right)^{-1} \sum_i (W_i \otimes m_i)\right]^{-1} \sum_i (W_i \otimes m_i)' \\ &\times \left(\widehat{\Omega} \otimes \sum_i m_i m'_i\right)^{-1} \sum_i (y_i \otimes m_i)\end{aligned} \tag{8.28}$$

Arellano and Bover (1995) prove that this 3SLS estimator is asymptotically equivalent to the limited information maximum likelihood procedure with Ω unrestricted developed by Bhargava and Sargan (1983).

Regardless of the existence of individual effects, the previous model assumes *unrestricted* serial correlation in the ν_{it} implying that $y_{i,t-1}$ is an endogenous variable. If the ν_{it} are *not* serially correlated, additional orthogonality restrictions can easily be incorporated in estimating (8.27) provided that the transformation C is now upper triangular in addition to the previous requirements. In this case, the transformed error in the equation for period t is independent of μ_i and $(\nu_{i1}, \ldots, \nu_{i,t-1})$ so that $(y_{i0}, y_{i1}, \ldots, y_{i,t-1})$ are additional valid instruments for this equation (see section 8.2). Therefore, the matrix of instruments M_i becomes

$$M_i = \begin{bmatrix} (w'_i, y_{i0}) & & & & 0 \\ & (w'_i, y_{i0}, y_{i1}) & & & \\ & & \ddots & & \\ & & & (w'_i, y_{i0}, \ldots, y_{i,T-2}) & \\ 0 & & & & m'_i \end{bmatrix} \tag{8.29}$$

Once again, Arellano and Bover (1995) show that the GMM estimator (8.25) that uses (8.29) as the matrix of instruments is invariant to the choice of C provided C satisfies the above required conditions.

8.4 THE AHN AND SCHMIDT MOMENT CONDITIONS

Ahn and Schmidt (1995) show that under the standard assumptions used in a dynamic panel data model, there are additional moment conditions that are ignored by the IV estimators suggested by Anderson and Hsiao (1981), Holtz-Eakin et al. (1988) and Arellano and Bond (1991). In this section, we explain how these additional restrictions arise for the simple dynamic model and show how they can be utilized in a GMM framework.

Consider the simple dynamic model with no regressors given in (8.3) and assume that $y_{i0}, \ldots, y_{iT}$ are observable. In vector form, this is given by

$$y_i = \delta y_{i-1} + u_i \tag{8.30}$$

where $y_i' = (y_{i1}, \ldots, y_{iT})$, $y_{i-1}' = (y_{i0}, \ldots, y_{i,T-1})$ and $u_i' = (u_{i1}, \ldots, u_{iT})$. The standard assumptions on the dynamic model (8.30) are that:

(A.1) For all i, ν_{it} is uncorrelated with y_{i0} for all t.
(A.2) For all i, ν_{it} is uncorrelated with μ_i for all t.
(A.3) For all i, the ν_{it} are mutually uncorrelated.

Ahn and Schmidt (1995) argue that these assumptions on the initial value y_{i0} are weaker than those often made in the literature (see Bhargava and Sargan, 1983 and Blundell and Smith, 1991).

Under these assumptions, one obtains the following $T(T-1)/2$ moment conditions:

$$E(y_{is}\Delta u_{it}) = 0 \quad t = 2, \ldots, T;\ s = 0, \ldots, t-2 \tag{8.31}$$

These are the same moment restrictions given below (8.6) and exploited by Arellano and Bond (1991). However, Ahn and Schmidt (1995) find $T-2$ additional moment conditions not implied by (8.31). These are given by

$$E(u_{iT}\Delta u_{it}) = 0 \quad t = 2, \ldots, T-1 \tag{8.32}$$

Therefore, (8.31) and (8.32) imply a set of $T(T-1)/2 + (T-2)$ moment conditions which represent all of the moment conditions implied by the assumptions that the ν_{it} are mutually uncorrelated among themselves and with μ_i and y_{i0}. More formally, the standard assumptions impose restrictions on the following covariance matrix:

$$\Sigma = \text{cov}\begin{bmatrix} \nu_{i1} \\ \nu_{i1} \\ \vdots \\ \nu_{iT} \\ y_{i0} \\ \mu_i \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1T} & \sigma_{10} & \sigma_{1\mu} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2T} & \sigma_{20} & \sigma_{2\mu} \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma_{TT} & \sigma_{T0} & \sigma_{T\mu} \\ \sigma_{01} & \sigma_{02} & \cdots & \sigma_{0T} & \sigma_{00} & \sigma_{0\mu} \\ \sigma_{\mu 1} & \sigma_{\mu 2} & \cdots & \sigma_{\mu T} & \sigma_{\mu 0} & \sigma_{\mu\mu} \end{bmatrix} \tag{8.33}$$

But, we do not observe μ_i and ν_{it}, only their sum $u_{it} = \mu_i + \nu_{it}$ which can be written in terms of the data and δ. Hence to get observable moment restrictions, we have to look at the following covariance matrix:

$$\Lambda = \mathrm{cov}\begin{bmatrix} \mu_i + \nu_{i1} \\ \mu_i + \nu_{i2} \\ \vdots \\ \mu_i + \nu_{iT} \\ y_{i0} \end{bmatrix} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1T} & \lambda_{10} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2T} & \lambda_{20} \\ \vdots & \vdots & & \vdots & \vdots \\ \lambda_{T1} & \lambda_{T2} & \dots & \lambda_{TT} & \lambda_{T0} \\ \lambda_{01} & \lambda_{02} & \dots & \lambda_{0T} & \lambda_{00} \end{bmatrix}$$

$$= \begin{bmatrix} (\sigma_{\mu\mu} + \sigma_{11} + 2\sigma_{\mu 1}) & (\sigma_{\mu\mu} + \sigma_{12} + \sigma_{\mu 1} + \sigma_{\mu 2}) \\ (\sigma_{\mu\mu} + \sigma_{12} + \sigma_{\mu 1} + \sigma_{\mu 2}) & (\sigma_{\mu\mu} + \sigma_{22} + 2\sigma_{\mu 2}) \\ \vdots & \vdots \\ (\sigma_{\mu\mu} + \sigma_{1T} + \sigma_{\mu 1} + \sigma_{\mu T}) & (\sigma_{\mu\mu} + \sigma_{2T} + \sigma_{\mu 2} + \sigma_{\mu T}) \\ (\sigma_{0\mu} + \sigma_{01}) & (\sigma_{0\mu} + \mu_{02}) \end{bmatrix} \tag{8.34}$$

$$\begin{bmatrix} \dots & (\sigma_{\mu\mu} + \sigma_{1T} + \sigma_{\mu 1} + \sigma_{\mu T}) & (\sigma_{0\mu} + \sigma_{01}) \\ \dots & (\sigma_{\mu\mu} + \sigma_{2T} + \sigma_{\mu 2} + \sigma_{\mu T}) & (\sigma_{0\mu} + \sigma_{02}) \\ \dots & \vdots & \vdots \\ \dots & (\sigma_{\mu\mu} + \sigma_{TT} + 2\sigma_{\mu T}) & (\sigma_{0\mu} + \sigma_{0T}) \\ \dots & (\sigma_{0\mu} + \sigma_{0T}) & \sigma_{00} \end{bmatrix}$$

Under the standard assumptions (A.1)–(A.3), we have $\sigma_{ts} = 0$ for all $t \neq s$, and $\sigma_{\mu t} = \sigma_{0t} = 0$ for all t. Then Λ simplifies as follows:

$$\Delta = \begin{bmatrix} (\sigma_{\mu\mu} + \sigma_{11}) & \sigma_{\mu\mu} & \dots & \sigma_{\mu\mu} & \sigma_{0\mu} \\ \sigma_{\mu\mu} & (\sigma_{\mu\mu} + \sigma_{22}) & \dots & \sigma_{\mu\mu} & \sigma_{0\mu} \\ \vdots & \vdots & & \vdots & \vdots \\ \sigma_{\mu\mu} & \sigma_{\mu\mu} & \dots & (\sigma_{\mu\mu} + \sigma_{TT}) & \sigma_{0\mu} \\ \sigma_{0\mu} & \sigma_{0\mu} & \dots & \sigma_{0\mu} & \sigma_{00} \end{bmatrix} \tag{8.35}$$

There are $T - 1$ restrictions, that $\lambda_{0t} = E(y_{i0}u_{it})$ is the same for $t = 1, \dots, T$; and $[T(T - 1)/2] - 1$ restrictions, that $\lambda_{ts} = E(u_{is}u_{it})$ is the same for $t, s = 1, \dots, T, t \neq s$. Adding the number of restrictions, we get $T(T - 1)/2 + (T - 2)$.

In order to see how these additional moment restrictions are utilized, consider our simple dynamic model in differenced form along with the last period's observation in levels:

$$\Delta y_{it} = \delta \Delta y_{i,t-1} + \Delta u_{it} \quad t = 2, 3, \dots, T \tag{8.36}$$

$$y_{iT} = \delta y_{i,T-1} + u_{iT} \tag{8.37}$$

The usual IV estimator, utilizing the restrictions in (8.31), amounts to estimating the first-differenced equations (8.36) by three-stage least squares, imposing the restriction that δ is the same in every equation, where the instrument set is y_{i0} for $t = 2; (y_{i0}, y_{i1})$ for $t = 3; \dots; (y_{i0}, \dots, y_{i,T-2})$ for $t = T$ (see section 8.2). Even though there are no legitimate observable instruments for the levels equation (8.37), Ahn and Schmidt argue that (8.37) is still useful in estimation because of the additional covariance restrictions implied by (8.32), i.e. that u_{iT} is uncorrelated with Δu_{it} for $t = 2, \dots, T - 1$. Ahn and Schmidt show that any additional covariance restrictions besides (8.32) are redundant and implied by the basic moment conditions given by (8.31). Ahn and Schmidt also point out that the moment conditions

(8.31) and (8.32) hold under weaker conditions than those implied by the standard assumptions (A.1)–(A.3). In fact, one only needs:

(B.1) $\text{cov}(\nu_{it}, y_{i0})$ is the same for all i and t instead of $\text{cov}(\nu_{it}, y_{i0}) = 0$, as in (A.1).
(B.2) $\text{cov}(\nu_{it}, \mu_i)$ is the same for all i and t instead of $\text{cov}(\nu_{it}, \mu_i) = 0$, as in (A.2).
(B.3) $\text{cov}(\nu_{it}, \nu_{is})$ is the same for all i and $t \neq s$, instead of $\text{cov}(\nu_{it}, \nu_{is}) = 0$, as in (A.3).

Problem 8.7 asks the reader to verify this claim in the same way as described above. Ahn and Schmidt (1995) show that GMM based on (8.31) and (8.32) is asymptotically equivalent to Chamberlain's (1982, 1984) optimal minimum distance estimator, and that it reaches the semiparametric efficiency bound. Ahn and Schmidt also explore additional moment restrictions obtained from assuming the ν_{it} homoskedastic for all i and t and the stationarity assumption of Arellano and Bover (1995) that $E(y_{it}\mu_i)$ is the same for all t. The reader is referred to their paper for more details. For specific parameter values, Ahn and Schmidt compute asymptotic covariance matrices and show that the extra moment conditions lead to substantial gains in asymptotic efficiency.

Ahn and Schmidt also consider the dynamic version of the Hausman and Taylor (1981) model studied in section 8.3 and show how one can make efficient use of exogenous variables as instruments. In particular, they show that the strong exogeneity assumption implies more orthogonality conditions which lie in the deviations from mean space. These are irrelevant in the static Hausman–Taylor model but are relevant for the dynamic version of that model. For more details on these conditions, see Schmidt, Ahn and Wyhowski (1992) and Ahn and Schmidt (1995).

8.5 THE BLUNDELL AND BOND SYSTEM GMM ESTIMATOR

Blundell and Bond (1998) revisit the importance of exploiting the initial condition in generating efficient estimators of the dynamic panel data model when T is small. They consider a simple autoregressive panel data model with no exogenous regressors,

$$y_{it} = \delta y_{i,t-1} + \mu_i + \nu_{it} \tag{8.38}$$

with $E(\mu_i) = 0$, $E(\nu_{it}) = 0$ and $E(\mu_i\nu_{it}) = 0$ for $i = 1, 2, \ldots, N$; $t = 1, 2, \ldots, T$. Blundell and Bond (1998) focus on the case where $T = 3$ and therefore there is only one orthogonality condition given by $E(y_{i1}\Delta\nu_{i3}) = 0$, so that δ is just-identified. In this case, the first-stage IV regression is obtained by running Δy_{i2} on y_{i1}. Note that this regression can be obtained from (8.38) evaluated at $t = 2$ by subtracting y_{i1} from both sides of this equation, i.e.

$$\Delta y_{i2} = (\delta - 1)y_{i,1} + \mu_i + \nu_{i2} \tag{8.39}$$

Since we expect $E(y_{i1}\mu_i) > 0$, $(\delta - 1)$ will be biased upwards with

$$\text{plim}(\widehat{\delta} - 1) = (\delta - 1)\frac{c}{c + (\sigma_\mu^2/\sigma_u^2)} \tag{8.40}$$

where $c = (1 - \delta)/(1 + \delta)$. The bias term effectively scales the estimated coefficient on the instrumental variable y_{i1} towards zero. They also find that the F-statistic of the first-stage IV regression converges to χ_1^2 with noncentrality parameter

$$\tau = \frac{(\sigma_u^2 c)^2}{\sigma_\mu^2 + \sigma_u^2 c} \rightarrow 0 \quad \text{as} \quad \delta \rightarrow 1 \tag{8.41}$$

As $\tau \to 0$, the instrumental variable estimator performs poorly. Hence, Blundell and Bond attribute the bias and the poor precision of the first-difference GMM estimator to the problem of weak instruments and characterize this by its concentration parameter τ.[6]

Next, Blundell and Bond (1998) show that an additional mild stationarity restriction on the initial conditions process allows the use of an extended system GMM estimator that uses lagged differences of y_{it} as instruments for equations in levels, in addition to lagged levels of y_{it} as instruments for equations in first differences, see Arellano and Bover (1995). The system GMM estimator is shown to have dramatic efficiency gains over the basic first-difference GMM as $\delta \to 1$ and $(\sigma_\mu^2/\sigma_u^2)$ increases. In fact, for $T = 4$ and $(\sigma_\mu^2/\sigma_u^2) = 1$, the asymptotic variance ratio of the first-difference GMM estimator to this system GMM estimator is 1.75 for $\delta = 0$ and increases to 3.26 for $\delta = 0.5$ and 55.4 for $\delta = 0.9$. This clearly demonstrates that the levels restrictions suggested by Arellano and Bover (1995) remain informative in cases where first-differenced instruments become weak. Things improve for first-difference GMM as T increases. However, with short T and persistent series, the Blundell and Bond findings support the use of the extra moment conditions. These results are reviewed and corroborated in Blundell and Bond (2000) and Blundell, Bond and Windmeijer (2000). Using Monte Carlo experiments, Blundell et al. (2000) find that simulations that include the weakly exogenous covariates exhibit large finite sample bias and very low precision for the standard first-differenced estimator. However, the system GMM estimator not only improves the precision but also reduces the finite sample bias. Blundell and Bond (2000) revisit the estimates of the capital and labor coefficients in a Cobb–Douglas production function considered by Griliches and Mairesse (1998). Using data on 509 R&D performing US manufacturing companies observed over 8 years (1982–89), the standard GMM estimator that uses moment conditions on the first-differenced model finds a low estimate of the capital coefficient and low precision for all coefficients estimated. However, the system GMM estimator gives reasonable and more precise estimates of the capital coefficient and constant returns to scale is not rejected. Blundell et al. conclude that "a careful examination of the original series and consideration of the system GMM estimator can usefully overcome many of the disappointing features of the standard GMM estimator for dynamic panel models".

Hahn (1999) examined the role of the initial condition imposed by the Blundell and Bond (1998) estimator. This was done by numerically comparing the semiparametric information bounds for the case that incorporates the stationarity of the initial condition and the case that does not. Hahn (1999) finds that the efficiency gain can be substantial.

Bond and Windmeijer (2002) project the unobserved individual effects on the vector of observations of the lagged dependent variable. This approach yields the Arellano and Bond (1991) estimator when no restrction is imposed on the initial conditions except for the assumption that they are uncorrelated with later shocks of the autoregressive process. It yields the Blundell and Bond (1998) estimator when the initial conditions satisfy mean stationarity. Bond and Windmeijer suggest a simple linear estimator for the case where the initial conditions satisfy covariance stationarity.

8.6 THE KEANE AND RUNKLE ESTIMATOR

Let $y = X\beta + u$ be our panel data model with X containing a lagged dependent variable. We consider the case where $E(u_{it}/X_{it}) \neq 0$, and there exists a set of predetermined instruments W such that $E(u_{it}/W_{is}) = 0$ for $s \leq t$, but $E(u_{it}/W_{is}) \neq 0$ for $s > t$. In other words, W may contain lagged values of y_{it}. For this model, the 2SLS estimator will provide a consistent estimator for β. Now consider the random effects model or any other kind of serial correlation which is invariant across individuals, $\Omega_{TS} = E(uu') = I_N \otimes \Sigma_{TS}$. In this case, 2SLS will not be efficient. Keane and Runkle (1992), henceforth KR, suggest an alternative more efficient algorithm that takes into account this more general variance–covariance structure for the disturbances based on the forward filtering idea from the time-series literature. This method of estimation eliminates the general serial correlation pattern in the data, while preserving the use of predetermined instruments in obtaining consistent parameter estimates. First, one gets a consistent estimate of Σ_{TS}^{-1} and its corresponding Cholesky's decomposition $\widehat{P}_{TS}$. Next, one premultiplies the model by $\widehat{Q}_{TS} = (I_N \otimes \widehat{P}_{TS})$ and estimates the model by 2SLS using the original instruments. In this case

$$\widehat{\beta}_{KR} = [X'\widehat{Q}'_{TS} P_W \widehat{Q}_{TS} X]^{-1} X'\widehat{Q}'_{TS} P_W \widehat{Q}_{TS} y \tag{8.42}$$

where $P_W = W(W'W)^{-1}W'$ is the projection matrix for the set of instruments W. Note that this allows for a general covariance matrix Σ_{TS} and its distinct elements $T(T+1)/2$ have to be much smaller than N. This is usually the case for large consumer or labor panels where N is very large and T is very small. Using the consistent 2SLS residuals, say $\widehat{u}_i$ for the ith individual, where $\widehat{u}_i$ is of dimension $T \times 1$, one can form

$$\widehat{\Sigma}_{TS} = \widehat{U}'\widehat{U}/N = \sum_{i=1}^{N} \widehat{u}_i \widehat{u}'_i / N$$

where $\widehat{U}' = [\widehat{u}_1, \widehat{u}_2, \ldots, \widehat{u}_N]$ is of dimension $(T \times N)$.[7]

First-differencing is also used in dynamic panel data models to get rid of individual specific effects. The resulting first-differenced errors are serially correlated of an MA(1) type with unit root if the original ν_{it} are classical errors. In this case, there will be gain in efficiency in performing the KR procedure on the first-differenced (FD) model. Get $\widehat{\Sigma}_{\mathrm{FD}}$ from FD-2SLS residuals and obtain $\widehat{Q}_{\mathrm{FD}} = I_N \otimes \widehat{P}_{\mathrm{FD}}$, then estimate the transformed equation by 2SLS using the original instruments.

Underlying this estimation procedure are two important hypotheses that are testable. The first is H_A : the set of instruments W are *strictly exogenous*. In order to test H_A, KR propose a test based on the difference between fixed effects 2SLS (FE-2SLS) and first-difference 2SLS (FD-2SLS). FE-2SLS is consistent only if H_A is true. In fact if the W are predetermined rather than strictly exogenous, then $E(W_{it}\bar{\nu}_{i.}) \neq 0$ and our estimator would not be consistent. In contrast, FD-2SLS is consistent whether H_A is true or not, i.e. $E(W_{it}\Delta\nu_{it}) = 0$ rain or shine. An example of this is when $y_{i,t-2}$ is a member of $W_{it,}$ then $y_{i,t-2}$ is predetermined and not correlated with $\Delta\nu_{it}$ as long as the ν_{it} are not serially correlated. However, $y_{i,t-2}$ is correlated with $\bar{\nu}_{i.}$ because this last average contains $\nu_{i,t-2}$. If H_A is not rejected, one should check whether the individual effects are correlated with the set of instruments. In this case, the usual Hausman and Taylor (1981) test applies. This is based on the difference between the FE and GLS estimator of the regression model. The FE estimator would be consistent rain or shine since it wipes out the individual effects. However, the GLS estimator would be consistent and efficient only if $E(\mu_i/W_{it}) = 0$, and inconsistent otherwise. If H_A is rejected, the instruments are predetermined and the Hausman–Taylor test is inappropriate. The test for H_B : $E(\mu_i/W_{it}) = 0$ will now be based on the difference between FD-2SLS and 2SLS. Under H_B, both estimators are consistent, but if H_B is not true, FD-2SLS remains consistent while 2SLS does not.

These tests are Hausman (1978)-type tests except that

$$\begin{aligned}\mathrm{var}(\widehat{\beta}_{\mathrm{FE-2SLS}} - \widehat{\beta}_{\mathrm{FD-2SLS}}) = {} & (\widetilde{X}' P_W \widetilde{X})^{-1}(\widetilde{X}' P_W \widetilde{\Omega}_{\mathrm{FE-2SLS}} P_W \widetilde{X})(\widetilde{X}' P_W \widetilde{X})^{-1} \\ & -(\widetilde{X}' P_W \widetilde{X})^{-1}(\widetilde{X}' P_W \widetilde{\Omega}_{\mathrm{FEFD}} P_W X_{\mathrm{FD}})(X'_{\mathrm{FD}} P_W X_{\mathrm{FD}})^{-1} \\ & -(X'_{\mathrm{FD}} P_W X_{\mathrm{FD}})^{-1}(X'_{\mathrm{FD}} P_W \widetilde{\Omega}_{\mathrm{FEFD}} P_W \widetilde{X})(\widetilde{X}' P_W \widetilde{X})^{-1} \\ & +(X'_{\mathrm{FD}} P_W X_{\mathrm{FD}})^{-1}(X'_{\mathrm{FD}} P_W \widehat{\Omega}_{\mathrm{FD-2SLS}} P_W X_{\mathrm{FD}})(X'_{\mathrm{FD}} P_W X_{\mathrm{FD}})^{-1}\end{aligned} \tag{8.43}$$

where $\widetilde{\Sigma}_{\mathrm{FE-2SLS}} = \widetilde{U}'_{\mathrm{FE}}\widetilde{U}_{\mathrm{FE}}/N$, $\widehat{\Sigma}_{\mathrm{FD-2SLS}} = \widehat{U}'_{\mathrm{FD}}\widehat{U}_{\mathrm{FD}}/N$ and $\widehat{\Sigma}_{\mathrm{FEFD}} = \widetilde{U}'_{\mathrm{FE}}\widehat{U}_{\mathrm{FD}}/N$. As described above, $\widetilde{U}'_{\mathrm{FE}} = [\widetilde{u}_1, \ldots, \widetilde{u}_N]_{\mathrm{FE}}$ denotes the FE-2SLS residuals and $\widetilde{U}'_{\mathrm{FD}} = [\widetilde{u}_1, \ldots, \widetilde{u}_N]_{\mathrm{FD}}$ denotes the FD-2SLS residuals. Recall that for the Keane–Runkle approach, $\Omega = I_N \otimes \Sigma$.

Similarly, the $\mathrm{var}(\widehat{\beta}_{\mathrm{2SLS}} - \widehat{\beta}_{\mathrm{FD-2SLS}})$ is computed as above with $\widetilde{X}$ being replaced by X, $\widetilde{\Omega}_{\mathrm{FE-2SLS}}$ by $\widehat{\Omega}_{\mathrm{2SLS}}$ and $\widetilde{\Omega}_{\mathrm{FEFD}}$ by $\widehat{\Omega}_{\mathrm{2SLSFD}}$. Also, $\widehat{\Sigma}_{\mathrm{2SLS}} = \widehat{U}'_{\mathrm{2SLS}}\widehat{U}_{\mathrm{2SLS}}/N$ and $\widehat{\Sigma}_{\mathrm{2SLSFD}} = \widehat{U}'_{\mathrm{2SLS}}\widehat{U}_{\mathrm{FD}}/N$.

The variances are complicated because KR do not use the efficient estimator under the null as required by a Hausman-type test (see Schmidt et al. 1992). Keane and Runkle (1992) apply their testing and estimation procedures to a simple version of the rational expectations lifecycle consumption model. Based on a sample of 627 households surveyed between 1972 and 1982 by the Michigan Panel Study on Income Dynamics (PSID), KR reject the strong exogeneity of the instruments. This means that the Within estimator is inconsistent and the standard Hausman test based on the difference between the standard Within and GLS estimators is inappropriate. In fact, for this consumption example the Hausman test leads to the wrong conclusion that the Within estimator is appropriate. KR also fail to reject the null hypothesis of no correlation between the individual effects and the instruments. This means that there is no need to first-difference to get rid of the individual effects. Based on the KR-2SLS estimates, the authors cannot reject the simple lifecycle model. However, they show that if one uses the inconsistent Within estimates for inference one would get misleading evidence against the lifecycle model.

8.7 FURTHER DEVELOPMENTS

The literature on dynamic panel data models continues to exhibit phenomenal growth. This is understandable given that most of our economic models are implicitly or explicitly dynamic in nature. This section summarizes some of the findings of these recent studies. In section 8.4, we pointed out that Ahn and Schmidt (1995) gave a complete count of the set of orthogonality conditions corresponding to a variety of assumptions imposed on the disturbances and the initial conditions of the dynamic panel data model. Many of these moment conditions were nonlinear in the parameters. More recently, Ahn and Schmidt (1997) propose a linearized GMM estimator that is asymptotically as efficient as the nonlinear GMM estimator. They also provide simple moment tests of the validity of these nonlinear restrictions. In addition, they investigate the circumstances under which the optimal GMM estimator is equivalent to a linear instrumental variable estimator. They find that these circumstances are quite restrictive and go beyond uncorrelatedness and homoskedasticity of the errors. Ahn and Schmidt (1995) provide some evidence on the efficiency gains from the nonlinear moment conditions which in turn provide support for their use in practice. By employing all these conditions, the resulting GMM estimator is asymptotically efficient and has the same asymptotic variance as the MLE under normality. In fact, Hahn (1997) showed that GMM based on an increasing set of instruments as $N \to \infty$ would achieve the semiparametric efficiency bound.

Hahn (1997) considers the asymptotic efficient estimation of the dynamic panel data model with sequential moment restrictions in an environment with i.i.d. observations. Hahn shows that the GMM estimator with an increasing set of instruments as the sample size grows attains the semiparametric efficiency bound of the model. He also explains how Fourier series or polynomials may be used as the set of instruments for efficient estimation. In a limited Monte Carlo comparison, Hahn finds that this estimator has similar finite sample properties as the Keane and Runkle (1992) and/or Schmidt et al. (1992) estimators when the latter estimators are efficient. In cases where the latter estimators are not efficient, the Hahn efficient estimator outperforms both estimators in finite samples.

Wansbeek and Bekker (1996) consider a simple dynamic panel data model with no exogenous regressors and disturbances u_{it} and random effects μ_i that are independent and normally distributed. They derive an expression for the optimal instrumental variable estimator, i.e., one with minimal asymptotic variance. A striking result is the difference in efficiency between the IV and ML estimators. They find that for regions of the autoregressive parameter δ which are likely in practice, ML is superior. The gap between IV (or GMM) and ML can be narrowed down by adding moment restrictions of the type considered by Ahn and Schmidt (1995). Hence, Wansbeek and Bekker (1996) find support for adding these nonlinear moment restrictions and warn against the loss in efficiency as compared with MLE by ignoring them.

Ziliak (1997) asks the question whether the bias/efficiency tradeoff for the GMM estimator considered by Tauchen (1986) for the time series case is still binding in panel data where the sample size is normally larger than 500. For time series data, Tauchen (1986) shows that even for $T = 50$ or 75 there is a bias/efficiency tradeoff as the number of moment conditions increases. Therefore, Tauchen recommends the use of suboptimal instruments in small samples. This result was also corroborated by Andersen and Sørensen (1996) who argue that GMM using too few moment conditions is just as bad as GMM using too many moment conditions. This problem becomes more pronounced with panel data since the number of moment conditions increases dramatically as the number of strictly exogenous variables and the number of time series observations increase. Even though it is desirable from an asymptotic efficiency point of view to include as many moment conditions as possible, it may be infeasible or impractical to do so in many cases. For example, for $T = 10$ and five strictly exogenous regressors, this generates 500 moment conditions for GMM. Ziliak (1997) performs an extensive set of Monte Carlo experiments for a dynamic panel data model and finds that the same tradeoff between bias and efficiency exists for GMM as the number of moment conditions increases, and that one is better off with suboptimal instruments. In fact, Ziliak finds that GMM performs well with suboptimal instruments, but is not recommended for panel data applications when all the moments are exploited for estimation. Ziliak estimates a lifecycle labor supply model under uncertainty based on 532 men observed over 10 years of data (1978–87) from the panel study of income dynamics. The sample was restricted to continuously married, continuously working prime age men aged 22–51 in 1978. These men were paid an hourly wage or salaried and could not be piece-rate workers or self-employed. Ziliak finds that the downward bias of GMM is quite severe as the number of moment conditions expands, outweighing the gains in efficiency. Ziliak reports estimates of the intertemporal substitution elasticity which is the focal point of interest in the labor supply literature. This measures the intertemporal changes in hours of work due to an anticipated change in the real wage. For GMM, this estimate changes from 0.519 to 0.093 when the number of moment conditions used in GMM is increased from 9 to 212. The standard error of this estimate drops from 0.36 to 0.07. Ziliak attributes this bias to the correlation between the sample moments used in estimation and the estimated weight matrix. Interestingly, Ziliak finds that the forward filter 2SLS estimator proposed by Keane and Runkle (1992) performs best in terms of the bias/efficiency tradeoff and is recommended. Forward filtering eliminates all forms of serial correlation while still maintaining orthogonality with the initial instrument set. Schmidt et al. (1992) argued that filtering is irrelevant if one exploits all sample moments during estimation. However, in practice, the number of moment conditions increases with the number of time periods T and the number of regressors K and can become computationally intractable. In fact for $T = 15$ and $K = 10$, the number of moment conditions for Schmidt et al. (1992) is $T(T-1)K/2$ which is 1040 restrictions, highlighting the computational burden of this approach. In addition, Ziliak argues that the over-identifying restrictions are less likely to be satisfied possibly due to the weak correlation between the instruments and the endogenous regressors. In this case, the forward filter 2SLS estimator is desirable yielding less bias than GMM and sizeable gains in efficiency. In fact, for the lifecycle labor example, the forward filter 2SLS estimate of the intertemporal substitution elasticity was 0.135 for 9 moment conditions compared to 0.296 for 212 moment conditions. The standard error of these estimates dropped from 0.32 to 0.09.

The practical problem of not being able to use more moment conditions as well as the statistical problem of the tradeoff between small sample bias and efficiency prompted Ahn and Schmidt (1999a) to pose the following questions: "Under what conditions can we use a smaller set of moment conditions without incurring any loss of asymptotic efficiency? In other words, under what conditions are some moment conditions redundant in the sense that utilizing them does not improve efficiency?" These questions were first dealt with by Im et al. (1999) who considered panel data models with strictly exogenous explanatory variables. They argued that, for example, with ten strictly exogenous time-varying variables and six time periods, the moment conditions available for the random effects (RE) model is 360 and this reduces to 300 moment conditions for the FE model. GMM utilizing all these moment conditions leads to an efficient estimator. However, these moment conditions exceed what the simple RE and FE estimators use. Im et al. (1999) provide the assumptions under which this efficient GMM estimator reduces to the simpler FE or RE estimator. In other words, Im et al. (1999) show the redundancy of the moment conditions that these simple estimators do not use. Ahn and Schmidt (1999a) provide a more systematic method by which redundant instruments can be found and generalize this result to models with time-varying individual effects. However, both papers deal only with strictly exogenous regressors. In a related paper, Ahn and Schmidt (1999b) consider the cases of strictly and weakly exogenous regressors. They show that the GMM estimator takes the form of an instrumental variables estimator if the assumption of no conditional heteroskedasticity (NCH) holds. Under this assumption, the efficiency of standard estimators can often be established showing that the moment conditions not utilized by these estimators are redundant. However, Ahn and Schmidt (1999b) conclude that the NCH assumption necessarily fails if the full set of moment conditions for the dynamic panel data model is used. In this case, there is clearly a need to find modified versions of GMM, with reduced sets of moment conditions that lead to estimates with reasonable finite sample properties.

Crépon, Kramarz and Trognon (1997) argue that for the dynamic panel data model, when one considers a set of orthogonal conditions, the parameters can be divided into parameters of interest (like δ) and nuisance parameters (like the second-order terms in the autoregressive error component model). They show that the elimination of such nuisance parameters using their empirical counterparts does not entail an efficiency loss when only the parameters of interest are estimated. In fact, Sevestre and Trognon in chapter 6 of Mátyás and Sevestre (1996) argue that if one is only interested in δ, then one can reduce the number of orthogonality restrictions without loss in efficiency as far as δ is concerned. However, the estimates of the other nuisance parameters are not generally as efficient as those obtained from the full set of orthogonality conditions.

The Alonso-Borrego and Arellano (1999) paper is also motivated by the finite sample bias in panel data instrumental variable estimators when the instruments are weak. The dynamic panel model generates many over-identifying restrictions even for moderate values of T. Also, the number of instruments increases with T, but the quality of these instruments is often poor because they tend to be only weakly correlated with first-differenced endogenous variables that appear in the equation. Limited information maximum likelihood (LIML) is strongly preferred to 2SLS if the number of instruments gets large as the sample size tends to infinity. Hillier (1990) showed that the alternative normalization rules adopted by LIML and 2SLS are at the root of their different sampling behavior. Hillier (1990) also showed that a symmetrically normalized 2SLS estimator has properties similar to those of LIML. Following Hillier (1990), Alonso-Borrego and Arellano (1999) derive a symmetrically normalized GMM (SNM) and compare it with ordinary GMM and LIML analogues by means of simulations. Monte Carlo and empirical results show that GMM can exhibit large biases when the instruments are poor, while LIML and SNM remain essentially unbiased. However, LIML and SNM always had a larger interquartile range than GMM. For $T = 4$, $N = 100$, $\sigma_\mu^2 = 0.2$ and $\sigma_\nu^2 = 1$, the bias for $\delta = 0.5$ was 6.9% for GMM, 1.7% for SNM and 1.7% for LIML. This bias increases to 17.8% for GMM, 3.7% for SNM and 4.1% for LIML for $\delta = 0.8$.

Alvarez and Arellano (2003) studied the asymptotic properties of FE, one-step GMM and nonrobust LIML for a first-order autoregressive model when both N and T tend to infinity with $(N/T) \to c$ for $0 \leq c < 2$. For this autoregressive model, the FE estimator is inconsistent for T fixed and N large, but becomes consistent as T gets large. GMM is consistent for fixed T, but the number of orthogonality conditions increases with T. The common conclusion among the studies cited above is that GMM estimators that use the full set of moments available can be severely biased, especially when the instruments are weak and the number of moment conditions is large relative to N. Alvarez and Arellano show that for $T < N$, GMM bias is always smaller than FE bias and LIML bias is smaller than the other two. In a fixed T framework, GMM and LIML are asymptotically equivalent, but as T increases, LIML has a smaller asymptotic bias than GMM. These results provide some theoretical support for LIML over GMM.[8] Alvarez and Arellano (2003) derive the asymptotic properties of the FE, GMM and LIML estimators of a dynamic model with random effects. When both T and $N \to \infty$, GMM and LIML are *consistent* and asymptotically equivalent to the FE estimator. When $T/N \to 0$, the fixed T results for GMM and LIML remain valid, but FE, although consistent, still exhibits an asymptotic bias term in its asymptotic distribution. When $T/N \to c$, where $0 < c \leq 2$, all three estimators are consistent. The basic intuition behind this result is that, contrary to the structural equation setting where too many instruments produce over-fitting and undesirable closeness to OLS; here, a larger number of instruments is associated with larger values of T and closeness to FE is desirable since the endogeneity bias $\to 0$ as $T \to \infty$. Nevertheless, FE, GMM and LIML exhibit a bias term in their asymptotic distributions; the biases are of order $1/T$, $1/N$ and $1/(2N - T)$, respectively. Provided $T < N$, the asymptotic bias of GMM is always smaller than the FE bias, and the LIML bias is smaller than the other two. When $T = N$, the asymptotic bias is the same for all three estimators.

Alvarez and Arellano (2003) also consider a random effects MLE that leaves the mean and variance of the initial conditions unrestricted but enforces time-series homoskedasticity. This estimator has no asymptotic bias because it does not entail incidental parameters in the

N and T dimensions, and it becomes robust to heteroskedasticity as $T \to \infty$. For the simple autoregressive model in (8.38) with $|\ \delta\ | < 1$, ν_{it} being iid across time and individuals and independent of μ_i and y_{i0}, Alvarez and Arellano (2003) find that as $T \to \infty$, regardless of whether N is fixed or tends to ∞, provided $N/T^3 \ \to 0$,

$$\sqrt{NT}\left[\widetilde{\delta}_{\text{FE}} - \left(\delta - \frac{1}{T}(1+\delta)\right)\right] \to N(0, 1-\delta^2) \tag{8.44}$$

Also, as N, $T \to \infty$ such that $(\log T^2)/N \to 0$, $\widehat{\delta}_{\text{GMM}} \to \delta$. Moreover, provided $T/N \to c$, $0 < c < \infty$,

$$\sqrt{NT}\left[\widehat{\delta}_{\text{GMM}} - \left(\delta - \frac{1}{N}(1+\delta)\right)\right] \to N(0, 1-\delta^2) \tag{8.45}$$

when $T \to \infty$, the number of GMM orthogonality conditions $T(T-1)/2 \to \infty$. In spite of this fact, $\widehat{\delta}_{\text{GMM}} \to \delta$. Also, as N, $T \to \infty$ provided $T/N \to c$, $0 \leq c \leq 2$, $\widehat{\delta}_{\text{LIML}} \to \delta$. Moreover,

$$\sqrt{NT}\left[\widehat{\delta}_{\text{LIML}} - \left(\delta - \frac{1}{2N-T}(1+\delta)\right)\right] \to N(0, 1-\delta^2) \tag{8.46}$$

LIML like GMM is consistent for δ despite $T \to \infty$ and $T/N \to c$. Provided $T < N$, the bias of LIML < bias of GMM < bias of FE. In fact, for $\delta = 0.2$, $T = 11$, $N = 100$, the median of 1000 Monte Carlo replications yields an estimate for δ of 0.063 for FE, 0.188 for GMM and 0.196 for LIML. For $\delta = 0.8$, $T = 11$, $N = 100$, the median of 1000 Monte Carlo replications yields an estimate for δ of 0.554 for FE, 0.763 for GMM and 0.792 for LIML. When we increase T to 51, $N = 100$ and $\delta = 0.8$, the median of 1000 Monte Carlo replications yields an estimate for δ of 0.760 for FE, 0.779 for GMM and 0.789 for LIML.

Wansbeek and Knapp (1999) consider a simple dynamic panel data model with heterogeneous coefficients on the lagged dependent variable and the time trend, i.e.

$$y_{it} = \delta_i y_{i,t-1} + \xi_i t + \mu_i + u_{it} \tag{8.47}$$

This model results from Islam's (1995) version of Solow's model on growth convergence among countries. Wansbeek and Knapp (1999) show that double-differencing gets rid of the individual country effects (μ_i) on the first round of differencing and the heterogeneous coefficient on the time trend (ξ_i) on the second round of differencing. Modified OLS, IV and GMM methods are adapted to this model and LIML is suggested as a viable alternative to GMM to guard against the small sample bias of GMM. Simulations show that LIML is the superior estimator for $T \geq 10$ and $N \geq 50$. Macroeconomic data are subject to measurement error and Wansbeek and Knapp (1999) show how these estimators can be modified to account for measurement error that is white noise. For example, GMM is modified so that it discards the orthogonality conditions that rely on the absence of measurement error.

Andrews and Lu (2001) develop consistent model and moment selection criteria and downward testing procedures for GMM estimation that are able to select the correct model and moments with probability that goes to one as the sample size goes to infinity. This is applied to dynamic panel data models with unobserved individual effects. The selection criteria can be used to select the lag length for the lagged dependent variables, to determine the exogeneity of the regressors, and/or to determine the existence of correlation between some regressors and the individual effects. Monte Carlo experiments are performed to study the small sample performance of the selection criteria and the testing procedures and their impact on parameter estimation.

Hahn and Kuersteiner (2002) consider the simple autoregressive model given in (8.38) with $\nu_{it} \sim N(0, \Omega)$ iid across i, $0 < \lim(N/T) = c < \infty$, $|\ \delta\ | < 1$ and $\sum_{i=1}^{N} y_{i0}^2/N = O(1)$ and $\sum_{i=1}^{N} \mu_i^2/N = O(1)$. The MLE of δ is the FE estimator which is inconsistent for fixed T and $N \to \infty$. For large T, large N, as in cross-country studies, such that $\lim(N/T) = c$ is finite, Hahn and Kuersteiner derive a bias-corrected estimator which reduces to

$$\widehat{\delta}_c = \left(\frac{T+1}{T}\right)\widetilde{\delta}_{\text{FE}} + \frac{1}{T}$$

with $\sqrt{NT}(\widehat{\delta}_c - \delta) \to N(0, 1-\delta^2)$. Under the assumption of normality of the disturbances, $\widehat{\delta}_c$ is assymptotically efficient as N, $T \to \infty$ at the same rate. Monte Carlo results for $T = 5$, 10, 20 and $N = 100$, 200 show that this bias-corrected MLE has comparable bias properties to the Arellano and Bover (1995) GMM estimator and often dominates in terms of RMSE for $T = 10, 20$ and $N = 100, 200$. Kiviet (1995) showed that a bias-corrected MLE (knowing δ) has much more desirable finite sample properties than various instrumental variable estimators. However, in order to make this estimator feasible, an initial instrumental variable for δ is used and its asymptotic properties are not derived. In contrast, Hahn and Kuersteiner's (2002) correction does not require a preliminary estimate of δ and its asymptotic properties are well derived. They also showed that this bias-corrected MLE is not expected to be asymptotically unbiased under a unit root ($\delta = 1$).

Hahn, Hausman and Kuersteiner (2003) consider the simple autoregressive panel data model in (8.38) with the following strong assumptions: (i) $\nu_{it} \sim \text{IIN}(0, \sigma_\nu^2)$ over i and t, (ii) stationarity conditions $(y_{i0}/\mu_i) \sim N(\frac{\mu_i}{1-\delta}, \frac{\sigma_\nu^2}{1-\delta^2})$ and $\mu_i \sim N(0, \sigma_\mu^2)$. They show that the Arellano and Bover (1995) GMM estimator, based on the forward demeaning transformation described in problem 8.4, can be represented as a linear combination of 2SLS estimators and therefore may be subject to a substantial finite sample bias. Based on 5000 Monte Carlo replications, they show that this indeed is the case for $T = 5$, 10, $N = 100$, 500 and $\delta = 0.1, 0.3, 0.5, 0.8$ and 0.9. For example, for $T = 5$, $N = 100$ and $\delta = 0.1$, the %bias of the GMM estimator is -16%, for $\delta = 0.8$, this %bias is -28% and for $\delta = 0.9$, this %bias is -51%. Hahn et al. attempt to eliminate this bias using two different approaches. The first is a second-order Taylor series-type approximation and the second is a long-difference estimator. The Monte Carlo results show that the second-order Taylor series-type approximation does a reasonably good job except when δ is close to 1 and N is small. Based on this, the bias-corrected (second-order theory) should be relatively free of bias. Monte Carlo results show that this is the case unless δ is close 1. For $T = 5$, $N = 100$ and $\delta = 0.1, 0.8, 0.9$ the %bias for this bias-corrected estimator is 0.25%, -11% and -42%, respectively.

The second-order asymptotics fails to be a good approximation around $\delta = 1$. This is due to the *weak instrument* problem, see Blundell and Bond (1998) in section 8.5. In fact, the latter paper argued that the weak IV problem can be alleviated by assuming stationarity on the initial observation y_{i0}. The stationarity condition turns out to be a predominant source of information around $\delta = 1$, as noted by Hahn (1999). The stationarity condition may or may not be appropriate for particular applications, and substantial finite sample biases due to inconsistency will result under violation of stationarity. Hahn et al. turn to the long-difference estimator to deal with weak IV around the unit circle avoiding the stationarity assumption:

$$y_{it} - y_{i1} = \delta(y_{it} - y_{i0}) + \nu_{it} - \nu_{i1}$$

Here y_{i0} is a valid instrument. The residuals $(y_{i,T-1} - \delta y_{i,T-2}), \ldots, (y_{i,2} - \delta y_{i,1})$ are also valid instruments. To make it operational, they suggest using the Arellano and Bover estimator for the first step and iterating using the long-difference estimator. The bias of the 2SLS (GMM) estimator depends on four factors, the sample size, the number of instruments, the covariance between the stochastic disturbance of the structural equation and the reduced form equation and the explained variance of the first-stage reduced form. The long-difference estimator increases the R^2, but decreases the covariance between the stochastic disturbance of the structural equation and the reduced form equation. This alleviates the weak instruments problem. Further, the number of instruments is smaller for the long-difference specification than for the first-difference GMM and therefore one should expect smaller bias. The actual properties of the long-difference estimator turn out to be much better than those predicted by higher-order theory, especially around the unit circle. Monte Carlo results show that the long-difference estimator does better than the other estimators for large δ and not significantly different for moderate δ.

Hahn et al. analyze the class of GMM estimators that exploit the Ahn and Schmidt (1997) complete set of moment conditions and show that a strict subset of the full set of moment restrictions should be used in estimation in order to minimize bias. They show that the long-difference estimator is a good approximation to the bias minimal procedure. They report the numerical values of the biases of the Arellano and Bond, Arellano and Bover and Ahn and Schmidt estimators under near unit root asymptotics and compare them with biases for the long-difference estimator as well as the bias minimal estimator. Despite the fact that the long-difference estimator does not achieve small bias reduction, as the fully optimal estimator it has significantly less bias than the more commonly used implementations of the GMM estimator.

8.8 EMPIRICAL EXAMPLE: DYNAMIC DEMAND FOR CIGARETTES

Baltagi and Levin (1992) estimate a dynamic demand model for cigarettes based on panel data from 46 American states. This data, updated from 1963–92, is available on the Wiley web site as cigar.txt. The estimated equation is

$$\ln C_{it} = \alpha + \beta_1 \ln C_{i,t-1} + \beta_2 \ln P_{i,t} + \beta_3 \ln Y_{it} + \beta_4 \ln Pn_{it} + u_{it} \tag{8.48}$$

where the subscript i denotes the ith state ($i = 1, \ldots, 46$) and the subscript t denotes the tth year ($t = 1, \ldots, 30$). C_{it} is real per capita sales of cigarettes by persons of smoking age (14 years and older). This is measured in packs of cigarettes per head. P_{it} is the average retail price of a pack of cigarettes measured in real terms. Y_{it} is real per capita disposable income. Pn_{it} denotes the minimum real price of cigarettes in any neighboring state. This last variable is a proxy for the casual smuggling effect across state borders. It acts as a substitute price attracting consumers from high-tax states like Massachusetts with 26¢ per pack to cross over to New Hampshire where the tax is only 12¢ per pack. The disturbance term is specified as a two-way error component model:

$$u_{it} = \mu_i + \lambda_t + \nu_{it} \quad i = 1, \ldots, 46; \quad t = 1, \ldots, 30 \tag{8.49}$$

where μ_i denotes a state-specific effect, and λ_t denotes a year-specific effect. The time-period effects (the λ_t) are assumed fixed parameters to be estimated as coefficients of time dummies for each year in the sample. This can be justified given the numerous policy interventions as well as health warnings and Surgeon General's reports. For example:

(1) The imposition of warning labels by the Federal Trade Commission effective January 1965.
(2) The application of the Fairness Doctrine Act to cigarette advertising in June 1967, which subsidized antismoking messages from 1968 to 1970.
(3) The Congressional ban on broadcast advertising of cigarettes effective January 1971.

The μ_i are state-specific effects which can represent any state-specific characteristic including the following:

(1) States with Indian reservations like Montana, New Mexico and Arizona are among the biggest losers in tax revenues from non-Indians purchasing tax-exempt cigarettes from the reservations.
(2) Florida, Texas, Washington and Georgia are among the biggest losers of revenues due to the purchasing of cigarettes from tax-exempt military bases in these states.
(3) Utah, which has a high percentage of Mormon population (a religion which forbids smoking), has a per capita sales of cigarettes in 1988 of 55 packs, a little less than half the national average of 113 packs.
(4) Nevada, which is a highly touristic state, has a per capita sales of cigarettes of 142 packs in 1988, 29 more packs than the national average.

These state-specific effects may be assumed fixed, in which case one includes state dummy variables in equation (8.48). The resulting estimator is the Within estimator reported in Table 8.1. Note that OLS, which ignores the state and time effects, yields a low short-run price elasticity of −0.09. However, the coefficient of lagged consumption is 0.97 which implies a high long-run price elasticity of −2.98. The Within estimator with both state and time

Table 8.1 Pooled Estimation Results.* Cigarette Demand Equation 1963–92

	$\ln C_{i,t-1}$	$\ln P_{it}$	$\ln Pn_{it}$	$\ln Y_{it}$
OLS	0.97	−0.090	0.024	−0.03
	(157.7)	(6.2)	(1.8)	(5.1)
Within	0.83	−0.299	0.034	0.10
	(66.3)	(12.7)	(1.2)	(4.2)
2SLS	0.85	−0.205	0.052	−0.02
	(25.3)	(5.8)	(3.1)	(2.2)
2SLS-KR	0.71	−0.311	0.071	−0.02
	(22.7)	(13.9)	(3.7)	(1.5)
Within-2SLS	0.60	−0.496	−0.016	0.19
	(17.0)	(13.0)	(0.5)	(6.4)
FD-2SLS	0.51	−0.348	0.112	0.10
	(9.5)	(12.3)	(3.5)	(2.9)
FD-2SLS-KR	0.49	−0.348	0.095	0.13
	(18.7)	(18.0)	(4.7)	(9.0)
GMM-one-step	0.84	−0.377	−0.016	0.14
	(52.0)	(11.7)	(0.4)	(3.8)
GMM-two-step	0.80	−0.379	−0.020	0.24
	(3.7)	(8.0)	(0.4)	(0.9)

* Numbers in parentheses are t-statistics. All regressions except OLS and 2SLS include time dummies.
Source: Some of the results in this table are reported in Baltagi, Griffin and Xiong (2000).

effects yields a higher short-run price elasticity of -0.30, but a lower long-run price elasticity of -1.79. Both state and time dummies were jointly significant with an observed F-statistic of 7.39 and a p-value of 0.0001. The observed F-statistic for the significance of state dummies (given the existence of time dummies) is 4.16 with a p-value of 0.0001. The observed F-statistic for the significance of time dummies (given the existence of state dummies) is 16.05 with a p-value of 0.0001. These results emphasize the importance of including state and time effects in the cigarette demand equation. This is a dynamic equation and the OLS and Within estimators do not take into account the endogeneity of the lagged dependent variable. Hence, we report 2SLS and Within-2SLS using as instruments the lagged exogenous regressors. These give lower estimates of lagged consumption and higher estimates of own price elasticities. The Hausman-type test based on the difference between Within-2SLS and FD-2SLS and discussed in section 8.6 yields a χ^2_4 statistic $= 118.6$. This rejects the consistency of the Within-2SLS estimator. The Hausman-type test based on the difference between 2SLS and FD-2SLS yields a χ^2_4 statistic $= 96.6$. This rejects the consistency of 2SLS. The FD-2SLS-KR estimator yields the lowest coefficient estimate of lagged consumption (0.49). The own price elasticity is -0.35 and significant. The income effect is very small (0.13) but significant and the bootlegging effect is small (0.095) and significant. The last two rows give the Arellano and Bond (1991) GMM one-step and two-step estimators. The lagged consumption coefficient estimate is 0.80 while the own price elasticity is -0.38 and significant. Table 8.2 gives the Stata output replicating the two-step estimator using (xtabond,twostep). Note that the two-step Sargan test for over-identification does not reject the null, but this could be due to the bad power of this test for $N = 46$ and $T = 28$. The test for first-order serial correlation rejects the null of no first-order serial correlation, but it does not reject the null that there is no second-order serial correlation. This is what one expects in a first-differenced equation with the original untransformed disturbances assumed to be not serially correlated.

8.9 FURTHER READING

Hsiao (2003) has an extensive discussion of the dynamic panel data model under the various assumptions on the initial values; see also Anderson and Hsiao (1981, 1982) and Bhargava and Sargan (1983). In particular, Hsiao (2003) shows that for the random effects dynamic model the consistency property of MLE and GLS depends upon various assumptions on the initial observations and on the way in which N and T tend to infinity. Read also the Arellano and Honoré (2001) chapter in the *Handbook of Econometrics*. The latter chapter pays careful attention to the implications of strict exogeneity for identification of the regression parameters controlling for unobserved heterogeneity and contrasts those with the case of predetermined regressors. Arellano's (2003) recent book has an excellent discussion on dynamic panel data models.

For applications of the dynamic error component model, see Becker, Grossman and Murphy (1994) who estimate a rational addiction model for cigarettes using a panel of 50 states (and the District of Columbia) over the period 1955–85. They apply fixed effects 2SLS to estimate a second-order difference equation in consumption of cigarettes, finding support for forward-looking consumers and rejecting myopic behavior. Their long-run price elasticity estimate is -0.78 as compared to -0.44 for the short-run. Baltagi and Griffin (2001) apply the FD-2SLS, FE-2SLS and GMM dynamic panel estimation methods studied in this chapter to the Becker et al. rational addiction model for cigarettes. Although the results are in general supportive of rational addiction, the estimates of the implied discount rate are not precise. Baltagi and Griffin (1995) estimate a dynamic demand for liquor across 43 states over the period 1960–82.

Table 8.2 Arellano and Bond Estimates of Cigarette Demand

```
. xtabond lnc lnrp lnrpn lnrdi dum3-dum29, lag(1) twostep

Arellano-Bond dynamic panel data      Number of obs      =       1288
Group variable (i) : state            Number of groups   =         46

                                      Wald chi2(31)      =   19321.43

Time variable (t) : yr                min number of obs  =         28
                                      max number of obs  =         28
                                      mean number of obs =         28

Two-step results
-------------------------------------------------------------------------------
lnc         |    Coef.  Std. Err.    z    P > |z|   [95% Conf. Interval]
------------+------------------------------------------------------------------
lnc         |
         LD |  .8036647 .2200745   3.65  0.000    .3723267  1.235003
lnrp        |
         D1 | -.3786939 .0471325  -8.03  0.000   -.4710719 -.2863159
lnrpn       |
         D1 | -.0197172 .0495158  -0.40  0.690   -.1167663  .0773319
lnrdi       |
         D1 |   .239147 .2778919   0.86  0.389   -.3055111   .783805

-------------------------------------------------------------------------------
The time dummies are not shown here to save space.

Sargan test of over-identifying restrictions:
         chi2(405) =    15.40     Prob > chi2 = 1.0000

Arellano-Bond test that average autocovariance in residuals of
 order 1 is 0:
          H0: no autocorrelation   z =  -3.53   Pr > z = 0.0004
Arellano-Bond test that average autocovariance in residuals of
 order 2 is 0:
          H0: no autocorrelation   z =   1.63   Pr > z = 0.1028
```

Fixed effects 2SLS as well as FD-2SLS-KR are performed. A short-run price elasticity of -0.20 and a long-run price elasticity of -0.69 are reported. Their findings support strong habit persistence, a small positive income elasticity and weak evidence of bootlegging from adjoining states.

Alternative estimation methods of a static and dynamic panel data model with arbitrary error structure are considered by Chamberlain (1982, 1984). Chamberlain (1984) considers the panel data model as a multivariate regression of T equations subject to restrictions and derives an efficient minimum distance estimator that is robust to residual autocorrelation of arbitrary form. Chamberlain (1984) also first-differences these equations to get rid of the individual effects and derives an asymptotically equivalent estimator to his efficient minimum distance estimator based on 3SLS of the $(T-2)$ differenced equations. Building on Chamberlain's work, Arellano (1990) develops minimum chi-square tests for various covariance restrictions. These tests are based on 3SLS residuals of the dynamic error component model and can be calculated from a generalized linear regression involving the sample autocovariance and dummy variables. The asymptotic distribution of the unrestricted autocovariance estimates is derived without imposing the normality assumption. In particular, Arellano (1990) considers testing covariance restrictions for error components or first-difference structures with white noise, moving average or autoregressive schemes. If these covariance restrictions are true, 3SLS is inefficient and Arellano (1990) proposes a GLS estimator which achieves asymptotic efficiency in the sense that it has the same limiting distribution as the optimal minimum distance estimator. Meghir and Windmeijer (1999) argue that it is important to model the higher-order moments of the dynamic process using panel data. For example, in a model for income dynamics and uncertainty, it is likely that persons at different levels of the income distribution face a different variance of their time–income profile. Meghir and Windmeijer model the dynamic variance process as an ARCH-type variance with multiplicative individual effects. They derive orthogonality conditions for estimating the coefficients of the conditional variance using GMM. This is done for nonautocorrelated errors, moving average errors and for models allowing for time-varying individual effects. Monte Carlo results show that large sample sizes are needed for estimating this conditional variance function with precision.

Li and Stengos (1992) propose a Hausman specification test based on $\sqrt{N}$-consistent semiparametric estimators. They apply it in the context of a dynamic panel data model of the form

$$y_{it} = \delta y_{i,t-1} + g(x_{it}) + u_{it} \quad i = 1, \ldots, N; t = 1, \ldots, T \tag{8.50}$$

where the function $g(.)$ is unknown, but satisfies certain moment and differentiability conditions. The x_{it} observations are IID with finite fourth moments and the disturbances u_{it} are IID$(0, \sigma^2)$ under the null hypothesis. Under the alternative, the disturbances u_{it} are IID in the i subscript but are serially correlated in the t subscript. Li and Stengos base the Hausman test for $H_0 : E(u_{it}|y_{i,t-1}) = 0$ on the difference between two $\sqrt{N}$-consistent instrumental variables estimators for δ, under the null and the alternative respectively. In other papers, Li and Stengos (1996) derived a $\sqrt{N}$-consistent instrumental variable estimator for a semiparametric dynamic panel data model, while Li and Stengos (1995) proposed a nonnested test for parametric vs semiparametric dynamic panel data models. Baltagi and Li (2002) proposed new semiparametric instrumental variable (IV) estimators that avoid the weak instrument problem which the Li and Stengos (1996) estimator may suffer from. Using Monte Carlo experiments, they show that these estimators yield substantial gains in efficiency over the estimators suggested by Li and Stengos (1996) and Li and Ullah (1998).

Kniesner and Li (2002) considered a semiparametric dynamic panel data model

$$y_{it} = \gamma z_{it} + f(y_{i,t-1}, x_{it}) + u_{it}$$

where the functional form of $f(.)$ is unknown to the researcher. They considered the common case of N large and T small, and proposed a two-step semiparametric $\sqrt{N}$-consistent estimation procedure for this model. Kniesner and Li (2002) also used labor panel data to illustrate the advantages of their semiparametric approach, vs OLS or IV approaches, which treat the parameters as constants. They argued that when the regression function is unknown, imposing a false parametric functional form may not only lead to inconsistent parameter estimation, but may aggravate the problem of individual heterogeneity. For a survey of nonparametric and semiparametric panel data models, see Ullah and Roy (1998).

Holtz-Eakin et al. (1988) formulate a coherent set of procedures for estimating and testing vector autoregressions (VAR) with panel data. The model builds upon Chamberlain's (1984) study and allows for nonstationary individual effects. It is applied to the study of dynamic relationships between wages and hours worked in two samples of American males. The data are based on a sample of 898 males from the PSID covering the period 1968–81. Two variables are considered for each individual, log of annual average hourly earnings and log of annual hours of work. Some of the results are checked using data from the National Longitudinal Survey of Men 45–59. Tests for parameter stationarity, minimum lag length and causality are performed. Holtz-Eakin et al. (1988) emphasize the importance of testing for the appropriate lag length before testing for causality, especially in short panels. Otherwise, misleading results on causality can be obtained. They suggest a simple method of estimating VAR equations with panel data that has a straightforward GLS interpretation. This is based on applying instrumental variables to the quasi-differenced autoregressive equations. They demonstrate how inappropriate methods that deal with individual effects in a VAR context can yield misleading results. Another application of these VAR methods with panel data is Holtz-Eakin, Newey and Rosen (1989) who study the dynamic relationships between local government revenues and expenditures. The data are based on 171 municipal governments over the period 1972–80. It is drawn from the Annual Survey of Governments between 1973 and 1980 and the Census of Governments conducted in 1972 and 1977. The main findings include the following:

(1) Lags of one or two years are sufficient to summarize the dynamic interrelationships in local public finance.
(2) There are important intertemporal linkages among expenditures, taxes and grants.
(3) Results of the stationarity test cast doubt over the stability of parameters over time.
(4) Contrary to previous studies, this study finds that past revenues help predict current expenditures, but past expenditures do not alter the future path of revenues.

NOTES

1. This corrected Within estimator performed well in simulations when compared with eight other consistent instrumental variable or GMM estimators discussed later in this chapter. Kiviet (1999) later extends this derivation of the bias to the case of weakly exogenous variables and examines to what degree this order of approximation is determined by the initial conditions of the dynamic panel data model.
2. Judson and Owen (1999) recommended the corrected Within estimator proposed by Kiviet (1995) as the best choice, followed by GMM as the second best choice. For long panels, they recommended the computationally simpler Anderson and Hsiao (1982) estimator.
3. Arellano and Bond (1991) warn about circumstances where their proposed serial correlation test is not defined, but where Sargan's over-identification test can still be computed. This is evident for $T = 4$ where no differenced residuals two periods apart are available to compute the serial correlation test. However, for the simple autoregressive model given in (8.3), Sargan's statistic tests two linear combinations of the three moment restrictions available, i.e. $E[(\nu_{i3} - \nu_{i2})y_{i1}] = E[(\nu_{i4} - \nu_{i3})y_{i1}] = E[(\nu_{i4} - \nu_{i3})y_{i2}] = 0$.
4. Arellano and Bover (1995) also discuss a forward orthogonal deviations operator as another example of C which is useful in the context of models with predetermined variables. This transformation essentially subtracts the mean of future observations available in the sample from the first $(T - 1)$ observations, see problem 8.4.
5. Arellano and Bover (1995) derive the Fisher information bound for η in order to assess the efficiency of the GMM estimators proposed in this section.
6. See the growing literature on weak instruments by Angrist and Kreuger (1995) and Staiger and Stock (1997) to mention a few.

7. It may be worth emphasizing that if $T > N$, this procedure will fail since Σ_{TS} will be singular with rank N. Also, the estimation of an unrestricted P_{TS} matrix will be difficult with missing data.
8. An alternative one-step method that achieves the same asymptotic efficiency as robust GMM or LIML estimators is the maximum empirical likelihood estimation method, see Imbens (1997). This maximizes a multinomial pseudo-likelihood function subject to the orthogonality restrictions. These are invariant to normalization because they are maximum likelihood estimators. See also Newey and Smith (2004) who give general analytical bias corrected versions of GMM and generalized empirical likelihood estimators.

PROBLEMS

8.1 For the simple autoregressive model with no regressors given in (8.3):
- (a) Write the first-differenced form of this equation for $t = 5$ and $t = 6$ and list the set of valid instruments for these two periods.
- (b) Show that the variance–covariance matrix of the first-difference disturbances is given by (8.5).
- (c) Verify that (8.8) is the GLS estimator of (8.7).

8.2 Consider the Monte Carlo set-up given in Arellano and Bond (1991, p. 283) for a simple autoregressive equation with one regressor with $N = 100$ and $T = 7$.
- (a) Compute the bias and mean-squared error based on 100 replications of the following estimators: OLS, Within, one-step and two-step Arellano and Bond GMM estimators, two Anderson and Hsiao-type estimators that use $\Delta y_{i,t-2}$ and $y_{i,t-2}$ as an instrument for $\Delta y_{i,t-1}$, respectively. Compare with table 1, p. 284 of Arellano and Bond (1991).
- (b) Compute Sargan's test of over-identifying restrictions given below (8.16) and count the number of rejections out of 100 replications. Compare with table 2 of Arellano and Bond (1991).

8.3 For $T = 5$, list the moment restrictions available for the simple autoregressive model given in (8.3). What over-identifying restrictions are being tested by Sargan's statistic given below (8.16)?

8.4 Consider three $(T-1) \times T$ matrices defined in (8.20) as follows: $C_1 =$ the first $(T-1)$ rows of $(I_T - \bar{J}_T)$, $C_2 =$ the first-difference operator, $C_3 =$ the forward orthogonal deviations operator which subtracts the mean of future observations from the first $(T-1)$ observations. This last matrix is given by Arellano and Bover (1995) as

$$C_3 = \operatorname{diag}\left[\frac{T-1}{T}, \ldots, \frac{1}{2}\right]^{1/2}$$

$$\times \begin{bmatrix} 1 & -\frac{1}{(T-1)} & -\frac{1}{(T-1)} & \ldots & -\frac{1}{(T-1)} & -\frac{1}{(T-1)} & -\frac{1}{(T-1)} \\ 0 & 1 & -\frac{1}{(T-2)} & \ldots & -\frac{1}{(T-2)} & -\frac{1}{(T-2)} & -\frac{1}{(T-2)} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & \ldots & 0 & 1 & -1 \end{bmatrix}$$

Verify that each one of these C matrices satisfies:
- (a) $C_j \iota_T = 0$ for $j = 1, 2, 3$.
- (b) $C_j'(C_j C_j')^{-1} C_j = I_T - \bar{J}_T$, the Within transformation, for $j = 1, 2, 3$.
- (c) For C_3, show that $C_3 C_3' = I_{T-1}$ and $C_3' C_3 = I_T - \bar{J}_T$. Hence $C_3 = (C'C)^{-1/2} C$ for any upper triangular C such that $C\iota_T = 0$.

8.5 (a) Verify that GLS on (8.24) yields (8.25).

(b) For the error component model with $\widetilde{\Omega} = \widetilde{\sigma}_v^2 I_T + \widetilde{\sigma}_\mu^2 J_T$ and $\widetilde{\sigma}_v^2$ and $\widetilde{\sigma}_\mu^2$ denoting consistent estimates of σ_v^2 and σ_μ^2, respectively, show that $\widehat{\eta}$ in (8.25) can be written as

$$\widehat{\eta} = \left[\sum_{i=1}^{N} W_i'(I_T - \bar{J}_T)W_i + \widetilde{\theta}^2 T \sum_{i=1}^{N} \bar{w}_i m_i' \left(\sum_{i=1}^{N} m_i m_i'\right)^{-1} \sum_{i=1}^{N} m_i \bar{w}_i'\right]^{-1}$$
$$\times \left[\sum_{i=1}^{N} W_i'(I_T - \bar{J}_T)y_i + \widetilde{\theta}^2 T \sum_{i=1}^{N} \bar{w}_i m_i' \left(\sum_{i=1}^{N} m_i m_i'\right)^{-1} \sum_{i=1}^{N} m_i \bar{y}_i\right]$$

where $\bar{w}_i = W_i'\iota_T/T$ and $\widetilde{\theta}^2 = \widetilde{\sigma}_v^2/(T\widetilde{\sigma}_\mu^2 + \widetilde{\sigma}_v^2)$. These are the familiar expressions for the HT, AM and BMS estimators for the corresponding choices of m_i. (*Hint*: See the proof in the appendix of Arellano and Bover (1995)).

8.6 For $T = 4$ and the simple autoregressive model considered in (8.3):

(a) What are the moment restrictions given by (8.31)? Compare with problem 8.3.

(b) What are the additional moment restrictions given by (8.32)?

(c) Write down the system of equations to be estimated by 3SLS using these additional restrictions and list the matrix of instruments for each equation.

8.7 Using the notation in (8.33)–(8.35), show that (8.31) and (8.32) hold under the weaker conditions (B.1)–(B.3) than those implied by assumptions (A.1)−(A.3).

8.8 Consider the Baltagi and Levin (1992) cigarette demand example for 46 states described in section 8.8. This data, updated from 1963–92, is available on the Wiley web site as cigar.txt.

(a) Estimate equation (8.48) using 2SLS, FD-2SLS and their Keane and Runkle (1992) version. (Assume only $\ln C_{i,t-1}$ is endogenous.)

(b) Estimate question (8.48) using the Within and FE-2SLS and perform the Hausman-type test based on FE-2SLS vs FD-2SLS.

(c) Perform the Hausman-type test based on 2SLS vs FD-2SLS.

(d) Perform the Anderson and Hsiao (1981) estimator for equation (8.48).

(e) Perform the Arellano and Bond (1991) GMM estimator for equation (8.48).

Hint: Some of the results are available in table 1 of Baltagi et al. (2000).

8.9 Consider the Arellano and Bond (1991) employment equation for 140 UK companies over the period 1979–84. Replicate all the estimation results in table 4 of Arellano and Bond (1991, p. 290).

9

Unbalanced Panel Data Models

9.1 INTRODUCTION

So far we have dealt only with "complete panels" or "balanced panels", i.e. cases where the individuals are observed over the entire sample period. Incomplete panels are more likely to be the norm in typical economic empirical settings. For example, in collecting data on US airlines over time, a researcher may find that some firms have dropped out of the market while new entrants emerged over the sample period observed. Similarly, while using labor or consumer panels on households, one may find that some households moved and can no longer be included in the panel. Additionally, if one is collecting data on a set of countries over time, a researcher may find some countries can be traced back longer than others. These typical scenarios lead to "unbalanced" or "incomplete" panels. This chapter deals with the econometric problems associated with these incomplete panels and how they differ from the complete panel data case. Throughout this chapter the panel data are assumed to be incomplete due to randomly missing observations. Nonrandomly missing data and rotating panels will be considered in Chapter 10.[1] Section 9.2 starts with the simple one-way error component model case with unbalanced data and surveys the estimation methods proposed in the literature. Section 9.4 treats the more complicated two-way error component model with unbalanced data. Section 9.5 looks at how some of the tests introduced earlier in the book are affected by the unbalanced panel, while section 9.6 gives some extensions of these unbalanced panel data methods to the nested error component model.

9.2 THE UNBALANCED ONE-WAY ERROR COMPONENT MODEL

To simplify the presentation, we analyze the case of two cross-sections with an unequal number of time-series observations and then generalize the analysis to the case of N cross-sections. Let n_1 be the shorter time series observed for the first cross-section ($i = 1$), and n_2 be the extra time-series observations available for the second cross-section ($i = 2$).[2] Stacking the n_1 observations for the first individual on top of the $(n_1 + n_2)$ observations on the second individual, we get

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \beta + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \tag{9.1}$$

where y_1 and y_2 are vectors of dimensions n_1 and $n_1 + n_2$, respectively. X_1 and X_2 are matrices of dimensions $n_1 \times K$ and $(n_1 + n_2) \times K$, respectively. In this case, $u_1' = (u_{11}, \ldots, u_{1,n_1})$, $u_2' = (u_{21}, \ldots, u_{2,n_1}, \ldots, u_{2,n_1+n_2})$ and the variance–covariance matrix is given by

$$\Omega = \begin{bmatrix} \sigma_\nu^2 I_{n_1} + \sigma_\mu^2 J_{n_1 n_1} & 0 & 0 \\ 0 & \sigma_\nu^2 I_{n_1} + \sigma_\mu^2 J_{n_1 n_1} & \sigma_\mu^2 J_{n_1 n_2} \\ 0 & \sigma_\mu^2 J_{n_2 n_1} & \sigma_\nu^2 I_{n_2} + \sigma_\mu^2 J_{n_2 n_2} \end{bmatrix} \tag{9.2}$$

where $u' = (u_1', u_2')$, I_{n_i} denotes an identity matrix of order n_i and $J_{n_i n_j}$ denotes a matrix of ones of dimension $n_i \times n_j$. Note that all the nonzero off-diagonal elements of Ω are equal to σ_μ^2. Therefore, if we let $T_j = \sum_{i=1}^{j} n_i$ for $j = 1, 2$, then Ω is clearly block-diagonal, with the jth block

$$\Omega_j = (T_j \sigma_\mu^2 + \sigma_\nu^2)\bar{J}_{T_j} + \sigma_\nu^2 E_{T_j} \tag{9.3}$$

where $\bar{J}_{T_j} = J_{T_j}/T_j$, $E_{T_j} = I_{T_j} - \bar{J}_{T_j}$ and there is no need for the double subscript anymore. Using the Wansbeek and Kapteyn (1982b) trick extended to the unbalanced case, Baltagi (1985) derived

$$\Omega_j^r = (T_j \sigma_\mu^2 + \sigma_\nu^2)^r \bar{J}_{T_j} + (\sigma_\nu^2)^r E_{T_j} \tag{9.4}$$

where r is any scalar. Let $w_j^2 = T_j \sigma_\mu^2 + \sigma_\nu^2$, then the Fuller and Battese (1974) transformation for the unbalanced case is the following:

$$\sigma_\nu \Omega_j^{-1/2} = (\sigma_\nu / w_j)\bar{J}_{T_j} + E_{T_j} = I_{T_j} - \theta_j \bar{J}_{T_j} \tag{9.5}$$

where $\theta_j = 1 - \sigma_\nu / w_j$, and $\sigma_\nu \Omega_j^{-1/2} y_j$ has a typical element $(y_{jt} - \theta_j \bar{y}_{j.})$ with $\bar{y}_{j.} = \sum_{t=1}^{T_j} y_{jt}/T_j$. Note that θ_j varies for each cross-sectional unit j depending on T_j. Hence GLS can be obtained as a simple weighted least squares (WLS) as in the complete panel data case. The basic difference, however, is that in the incomplete panel data case, the weights are crucially dependent on the lengths of the time series available for each cross-section.

The above results generalize in two directions: (i) the same analysis applies no matter how the observations for the two firms overlap; (ii) the results extend from the two cross-sections to the N cross-sections case. The proof is simple. Since the off-diagonal elements of the covariance matrix are zero for observations belonging to different firms, Ω remains block-diagonal as long as the observations are ordered by firms. Also, the nonzero off-diagonal elements are all equal to σ_μ^2. Hence $\Omega_j^{-1/2}$ can be derived along the same lines shown above.

In general, the regression model with unbalanced one-way error component disturbances is given by

$$\begin{aligned} y_{it} &= \alpha + X_{it}'\beta + u_{it} \quad i = 1, \ldots, N;\ t = 1, \ldots, T_i \\ u_{it} &= \mu_i + \nu_{it} \end{aligned} \tag{9.6}$$

where X_{it} is a $(K-1) \times 1$ vector of regressors, $\mu_i \sim \text{IIN}(0, \sigma_\mu^2)$ and independent of $\nu_{it} \sim \text{IIN}(0, \sigma_\nu^2)$. This model is unbalanced in the sense that there are N individuals observed over varying time-period length (T_i for $i = 1, \ldots, N$). Writing this equation in vector form, we have

$$\begin{aligned} y &= \alpha\iota_n + X\beta + u = Z\delta + u \\ u &= Z_\mu \mu + \nu \end{aligned} \tag{9.7}$$

where y and Z are of dimensions $n \times 1$ and $n \times K$, respectively, $Z = (\iota_n, X)$, $\delta' = (\alpha', \beta')$, $n = \sum T_i$, $Z_\mu = \text{diag}(\iota_{T_i})$ and ι_{T_i} is a vector of ones of dimension T_i. $\mu = (\mu_1, \mu_2, \ldots, \mu_N)'$ and $\nu = (\nu_{11}, \ldots, \nu_{1T_1}, \ldots, \nu_{N1}, \ldots, \nu_{NT_N})'$.

The ordinary least squares (OLS) on the unbalanced data is given by

$$\widehat{\delta}_{\text{OLS}} = (Z'Z)^{-1} Z'y \tag{9.8}$$

OLS is the best linear unbiased estimator when the variance component σ_μ^2 is equal to zero. Even when σ_μ^2 is positive, OLS is still unbiased and consistent, but its standard errors are biased (see Moulton, 1986). The OLS residuals are denoted by $\widehat{u}_{\text{OLS}} = y - Z\widehat{\delta}_{\text{OLS}}$.

The Within estimator can be obtained by first transforming the dependent variable y and X, the exogenous regressors excluding the intercept, using the matrix $Q = \text{diag}(E_{T_i})$, and then applying OLS to the transformed data:

$$\widetilde{\beta} = (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\widetilde{y} \tag{9.9}$$

where $\widetilde{X} = QX$, $\widetilde{y} = Qy$. The estimate of the intercept can be retrieved as follows: $\widetilde{\alpha} = (\bar{y}_{..} - \bar{X}_{..}\widetilde{\beta})$ where the dot indicates summation and the bar indicates averaging, for example, $\bar{y}_{..} = \sum\sum y_{it}/n$. Following Amemiya (1971), the Within residuals $\widetilde{u}$ for the unbalanced panel are given by

$$\widetilde{u} = y - \widetilde{\alpha}\iota_n - X\widetilde{\beta} \tag{9.10}$$

The Between estimator $\widehat{\delta}_{\text{Between}}$ is obtained as follows:

$$\widehat{\delta}_{\text{Between}} = (Z'PZ)^{-1}Z'Py \tag{9.11}$$

where $P = \text{diag}[\bar{J}_{T_i}]$, and the Between residuals are denoted by $\hat{u}^b = y - Z\widehat{\delta}_{\text{Between}}$.

GLS using the true variance components is obtained as follows:

$$\widehat{\delta}_{\text{GLS}} = (Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}y \tag{9.12}$$

where $\Omega = \sigma_\nu^2\Sigma = E(uu')$ with

$$\Sigma = I_n + \rho Z_\mu Z_\mu' = \text{diag}(E_{T_i}) + \text{diag}[(1 + \rho T_i)\bar{J}_{T_i}] \tag{9.13}$$

and $\rho = \sigma_\mu^2/\sigma_\nu^2$. Note that $(1 + \rho T_i) = (w_i^2/\sigma_\nu^2)$ where $w_i^2 = (T_i\sigma_\mu^2 + \sigma_\nu^2)$ was defined in (9.4). Therefore, GLS can be obtained by applying OLS on the transformed variables y^* and Z^*, i.e.

$$\widehat{\delta} = (Z^{*\prime}Z^*)^{-1}Z^{*\prime}y^*$$

where $Z^* = \sigma_\nu\Omega^{-1/2}Z$, $y^* = \sigma_\nu\Omega^{-1/2}y$ and

$$\sigma_\nu\Omega^{-1/2} = \text{diag}(E_{T_i}) + \text{diag}[(\sigma_\nu/w_i)\bar{J}_{T_i}] \tag{9.14}$$

as described in (9.5).

We now focus on methods of estimating the variance components, which are described in more detail in Baltagi and Chang (1994).

9.2.1 ANOVA Methods

The ANOVA method is one of the most popular methods in the estimation of variance components. The ANOVA estimators are method of moments-type estimators, which equate quadratic sums of squares to their expectations and solve the resulting linear system of equations. For the balanced model, ANOVA estimators are best quadratic unbiased (BQU) estimators of the variance components (see Searle, 1971). Under normality of the disturbances, these ANOVA estimators are minimum variance unbiased. For the unbalanced one-way model, BQU estimators of the variance components are a function of the variance components themselves (see Townsend and Searle, 1971). Still, unbalanced ANOVA methods are available (see Searle, 1987), but optimal properties beyond unbiasedness are lost. In what follows, we generalize some of the ANOVA methods described in Chapter 2 to the unbalanced case. In particular, we consider the two quadratic forms defining the Within and Between sums of squares:

$$q_1 = u'Qu \quad \text{and} \quad q_2 = u'Pu \tag{9.15}$$

where $Q = \text{diag}[E_{T_i}]$ and $P = \text{diag}[\bar{J}_{T_i}]$. Since the true disturbances are not known, we follow the Wallace and Hussain (1969) suggestion by substituting OLS residuals $\widehat{u}_{\text{OLS}}$ for u in (9.15). Upon taking expectations, we get

$$\begin{aligned} E(\widehat{q}_1) &= E(\hat{u}'_{\text{OLS}} Q \hat{u}_{\text{OLS}}) = \delta_{11}\sigma_\mu^2 + \delta_{12}\sigma_\nu^2 \\ E(\widehat{q}_2) &= E(\hat{u}'_{\text{OLS}} P \hat{u}_{\text{OLS}}) = \delta_{21}\sigma_\mu^2 + \delta_{22}\sigma_\nu^2 \end{aligned} \tag{9.16}$$

where $\delta_{11}, \delta_{12}, \delta_{21}, \delta_{22}$ are given by

$$\begin{aligned} \delta_{11} &= \text{tr}((Z'Z)^{-1}Z'Z_\mu Z'_\mu Z) - \text{tr}((Z'Z)^{-1}Z'PZ(Z'Z)^{-1}Z'Z_\mu Z'_\mu Z) \\ \delta_{12} &= n - N - K + \text{tr}((Z'Z)^{-1}Z'PZ) \\ \delta_{21} &= n - 2\text{tr}((Z'Z)^{-1}Z'Z_\mu Z'_\mu Z) + \text{tr}((Z'Z)^{-1}Z'PZ(Z'Z)^{-1}Z'Z_\mu Z'_\mu Z) \\ \delta_{22} &= N - \text{tr}((Z'Z)^{-1}Z'PZ) \end{aligned}$$

Equating $\widehat{q}_i$ to its expected value $E(\widehat{q}_i)$ in (9.16) and solving the system of equations, we get the Wallace and Hussain (WH)-type estimators of the variance components.

Alternatively, we can substitute Within residuals in the quadratic forms given in (9.15) to get $\widetilde{q}_1 = \widetilde{u}'Q\widetilde{u}$ and $\widetilde{q}_2 = \widetilde{u}'P\widetilde{u}$ as suggested by Amemiya (1971) for the balanced case. The expected values of $\widetilde{q}_1$ and $\widetilde{q}_2$ are given by

$$\begin{aligned} E(\widetilde{q}_1) &= (n - N - K + 1)\sigma_\nu^2 \\ E(\widetilde{q}_2) &= (N - 1 + \text{tr}[(X'QX)^{-1}X'PX] - \text{tr}[(X'QX)^{-1}X'\bar{J}_n X])\sigma_\nu^2 \\ &\quad + \left[n - \left(\sum_{i=1}^N T_i^2/n\right)\right]\sigma_\mu^2 \end{aligned} \tag{9.17}$$

Equating $\widetilde{q}_i$ to its expected value $E(\widetilde{q}_i)$ in (9.17), we get the Amemiya-type estimators of the variance components

$$\begin{aligned} \widehat{\sigma}_\nu^2 &= \widetilde{u}'Q\widetilde{u}/(n - N - K + 1) \\ \widehat{\sigma}_\mu^2 &= \frac{\widetilde{u}'P\widetilde{u} - \{N - 1 + \text{tr}[(X'QX)^{-1}X'PX] - \text{tr}[(X'QX)^{-1}X'\bar{J}_n X]\}\widehat{\sigma}_\nu^2}{n - \sum_{i=1}^N T_i^2/n} \end{aligned} \tag{9.18}$$

Next, we follow the Swamy and Arora (1972) suggestion of using the Between and Within regression mean square errors to estimate the variance components. In fact, their method amounts to substituting Within residuals in q_1 and Between residuals in q_2, to get $\widetilde{q}_1 = \widetilde{u}'Q\widetilde{u}$ and $\hat{q}_2^b = \hat{u}^{b\prime}P\hat{u}^b$. Since $\widetilde{q}_1$ is exactly the same as that for the Amemiya method, the Swamy and Arora (SA)-type estimator of $\widehat{\sigma}_\nu^2$ is the same as that given in equation (9.18). The expected value of $\hat{q}_2^b$ is given by

$$E(\hat{q}_2^b) = [n - \text{tr}((Z'PZ)^{-1}Z'Z_\mu Z'_\mu Z)]\sigma_\mu^2 + (N - K)\sigma_\nu^2 \tag{9.19}$$

Equating $E(\hat{q}_2^b)$ to $\hat{q}_2^b$ one gets the following estimator of σ_μ^2:

$$\widehat{\sigma}_\mu^2 = \frac{\hat{u}^{b\prime} P \hat{u}^b - (N-K)\widehat{\sigma}_\nu^2}{n - \operatorname{tr}((Z'PZ)^{-1} Z' Z_\mu Z_\mu' Z)} \tag{9.20}$$

Note that $\hat{u}^{b'} P \hat{u}^b$ can be obtained as the OLS residual sum of squares from the regression involving $\sqrt{T_i}\,\bar{y}_{i.}$ on $\sqrt{T_i}\,\bar{Z}_{i.}$.

Finally, we consider Henderson's method III (see Fuller and Battese, 1974) which will be denoted by HFB. This method utilizes the fitting constants method described in Searle (1971, p. 489). Let

$$R(\mu) = y' Z_\mu (Z_\mu' Z_\mu)^{-1} Z_\mu' y = \sum_{i=1}^{N} (y_{i.}^2 / T_i);\ R(\delta \mid \mu) = \tilde{y}' \widetilde{X} (\widetilde{X}' \widetilde{X})^{-1} \widetilde{X}' \tilde{y}$$

$$R(\delta) = y' Z (Z'Z)^{-1} Z' y \qquad \text{and} \qquad R(\mu \mid \delta) = R(\delta \mid \mu) + R(\mu) - R(\delta)$$

Then Henderson's (1953) method III estimators are given by

$$\begin{aligned} \widehat{\sigma}_\nu^2 &= \frac{y'y - R(\delta \mid \mu) - R(\mu)}{n - K - N + 1} \\ \widehat{\sigma}_\mu^2 &= \frac{R(\mu \mid \delta) - (N-1)\widehat{\sigma}_\nu^2}{n - \operatorname{tr}(Z_\mu' Z (Z'Z)^{-1} Z' Z_\mu)} \end{aligned} \tag{9.21}$$

9.2.2 Maximum Likelihood Estimators

Maximum likelihood (ML) estimates of the variance components and regression coefficients are obtained by maximizing the following loglikelihood function:

$$\begin{aligned} \log L = &-(n/2)\log(2\pi) - (n/2)\log \sigma_\nu^2 - \tfrac{1}{2}\log|\Sigma| \\ &-(y - Z\delta)' \Sigma^{-1} (y - Z\delta)/2\sigma_\nu^2 \end{aligned} \tag{9.22}$$

where ρ and Σ are given in (9.13). The first-order conditions give closed form solutions for $\widehat{\delta}$ and $\widehat{\sigma}_\nu^2$ conditional on $\widehat{\rho}$:

$$\begin{aligned} \widehat{\delta} &= (Z' \widehat{\Sigma}^{-1} Z)^{-1} Z' \widehat{\Sigma}^{-1} y \\ \widehat{\sigma}_\nu^2 &= (y - Z\delta)' \widehat{\Sigma}^{-1} (y - Z\delta)/n \end{aligned} \tag{9.23}$$

However, the first-order condition based on ρ is nonlinear in ρ even for known values of δ and σ_ν^2:

$$0 = \frac{\partial \log L}{\partial \rho} = \frac{1}{2}\operatorname{tr}(Z' \Sigma^{-1} Z) + \frac{1}{2\sigma_\nu^2}(y - Z\delta)' \Sigma^{-1} Z_\mu Z_\mu' \Sigma^{-1} (y - Z\delta) \tag{9.24}$$

A numerical solution by means of iteration is necessary for $\widehat{\rho}$. The second derivative of $\log L$ with respect to ρ is given by

$$\begin{aligned} \frac{\partial^2 \log L}{\partial \rho \partial \rho} = &\ \frac{1}{2}\operatorname{tr}\{(Z_\mu' \Sigma^{-1} Z_\mu)(Z_\mu' \Sigma^{-1} Z_\mu)\} \\ &- \frac{1}{\sigma_\nu^2}\{(y - Z\delta)' \Sigma^{-1} Z_\mu (Z_\mu' \Sigma^{-1} Z_\mu) Z_\mu' \Sigma^{-1} (y - Z\delta)\} \end{aligned} \tag{9.25}$$

Starting with an initial value of ρ_0, one obtains $\widehat{\Sigma}_0$ from (9.13) and $\widehat{\delta}_0$ and $\widehat{\sigma}^2_{\nu 0}$ from (9.23). The updated value ρ_1 is given from the following formula:

$$\rho_1 = \rho_0 - s\left[\frac{\partial^2 \log L}{\partial\rho\partial\rho}\right]_0^{-1}\left[\frac{\partial \log L}{\partial\rho}\right]_0 \tag{9.26}$$

where the subscript zero means evaluated at $\widehat{\Sigma}_0, \widehat{\delta}_0$ and $\widehat{\sigma}^2_{\nu 0}$ and s is a step size which is adjusted by step halving.[3] For the computational advantage of this algorithm as well as other algorithms like the Fisher scoring algorithm, see Jennrich and Sampson (1976) and Harville (1977). Maximum likelihood estimators are functions of sufficient statistics and are consistent and asymptotically efficient; see Harville (1977) for a review of the properties, advantages and disadvantages of ML estimators.

The ML approach has been criticized on grounds that it does not take into account the loss of degrees of freedom due to the regression coefficients in estimating the variance components. Patterson and Thompson (1971) remedy this by partitioning the likelihood function into two parts, one part depending only on the variance components and free of the regression coefficients. Maximizing this part yields the restricted maximum likelihood estimator (REML). REML estimators of the variance components are asymptotically equivalent to the ML estimators, however, little is known about their finite sample properties, and they reduce to the ANOVA estimators under several balanced data cases. For details, see Corbeil and Searle (1976a,b).

9.2.3 Minimum Norm and Minimum Variance Quadratic Unbiased Estimators (MINQUE and MIVQUE)

Under normality of the disturbances, Rao's (1971a) MINQUE and MIVQUE procedures for estimating the variance components are identical. Since we assume normality, we will focus on MIVQUE. Basically, the MIVQUE of a linear combination of the variance components, $p_\mu\sigma^2_\mu + p_\nu\sigma^2_\nu$, is obtained by finding a symmetric matrix G such that $\mathrm{var}(y'Gy) = 2\,\mathrm{tr}\{\sigma^2_\mu(GZ_\mu Z'_\mu) + \sigma^2_\nu G\}^2$ is minimized subject to the conditions that $y'Gy$ is an unbiased estimator of $(p_\mu\sigma^2_\mu + p_\nu\sigma^2_\nu)$ and is invariant to any translation of the δ parameter. These yield the following constraints:

1. $GZ = 0$.
2. $\mathrm{tr}(GZ_\mu Z'_\mu) = p_\mu$ and $\mathrm{tr}(G) = p_\nu$.

Rao (1971b) showed that the MIVQUE estimates of the variance components are given by

$$\begin{bmatrix}\hat{\sigma}^2_\mu\\ \hat{\sigma}^2_\nu\end{bmatrix} = \begin{bmatrix}\gamma_{11} & \gamma_{12}\\ \gamma_{12} & \gamma_{22}\end{bmatrix}^{-1}\begin{bmatrix}\delta_1\\ \delta_2\end{bmatrix}$$

where $\gamma_{11} = \mathrm{tr}(Z_\mu Z'_\mu R Z_\mu Z'_\mu R)$, $\gamma_{12} = \mathrm{tr}(Z_\mu Z'_\mu RR)$, $\gamma_{22} = \mathrm{tr}(RR)$, $\delta_1 = y'RZ_\mu Z'_\mu Ry$, $\delta_2 = y'RRy$ and $R = (\Sigma^{-1} - \Sigma^{-1}Z(Z'\Sigma^{-1}Z)^{-1}Z'\Sigma^{-1})/\sigma^2_\nu$. It is clear that MIVQUE requires *a priori* values of the variance components, and the resulting estimators possess the minimum variance property only if these *a priori* values coincide with the true values. Therefore, MIVQUE are only "locally best" and "locally minimum variance". If one iterates on the initial values of the variance components, the iterative estimators (IMIVQUE) become biased after the first iteration and MINQUE properties are not preserved. Two priors for the MINQUE estimator used in practice are: (i) the identity matrix, denoted by (MQ0) and (ii) the ANOVA estimator of Swamy and Arora, denoted by (MQA). Under normality, if one iterates until convergence, IMINQUE, IMIVQUE and REML will be identical (see Hocking and Kutner, 1975; Swallow and Monahan, 1984).[4]

9.2.4 Monte Carlo Results

Baltagi and Chang (1994) performed an extensive Monte Carlo study using a simple as well as a multiple regression with unbalanced one-way error component disturbances. The degree of unbalance in the sample as well as the variance component ratio ρ were varied across the experiments. The total number of observations as well as the total variance were fixed across the experiments to allow comparison of MSE for the various estimators considered. Some of the basic results of the Monte Carlo study suggest the following:

(1) As far as the estimation of regression coefficients is concerned, the simple ANOVA-type feasible GLS estimators compare well with the more complicated estimators such as ML, REML and MQA and are never more than 4% above the MSE of true GLS. However, MQ0 is not recommended for large ρ and unbalanced designs.
(2) For the estimation of the remainder variance component σ_ν^2 the ANOVA, MIVQUE(A), ML and REML estimators show little difference in relative MSE performance. However, for the individual specific variance component estimation, σ_μ^2, the ANOVA-type estimators perform poorly relative to ML, REML and MQA when the variance component ratio $\rho > 1$ and the pattern is severely unbalanced. MQ0 gives an extremely poor performance for severely unbalanced patterns and large ρ and is not recommended for these cases.
(3) Better estimates of the variance components, in the MSE sense, do not necessarily imply better estimates of the regression coefficients. This echoes similar findings for the balanced panel data case.
(4) Negative estimates of the variance components occurred when the true value of σ_μ^2 was zero or close to zero. In these cases, replacing these negative estimates by zero did not lead to much loss in efficiency.
(5) Extracting a balanced panel out of an unbalanced panel by either maximizing the number of households observed or the total number of observations in the balanced panel leads in both cases to an enormous loss in efficiency and is not recommended.[5]

9.3 EMPIRICAL EXAMPLE: HEDONIC HOUSING

Baltagi and Chang (1994) apply the various unbalanced variance components methods to the data set collected by Harrison and Rubinfeld (1978) for a study of hedonic housing prices and the willingness to pay for clean air. This data is available on the Wiley web site as Hedonic.xls. The total number of observations is 506 census tracts in the Boston area in 1970 and the number of variables is 14. Belsley, Kuh and Welsch (1980) identify 92 towns, consisting of 15 within Boston and 77 in its surrounding area. Thus, it is possible to group these data and analyze them as an unbalanced one-way model with random group effects. The group sizes range from one to 30 observations. The dependent variable is the median value (MV) of owner-occupied homes. The regressors include two structural variables, RM the average number of rooms, and

AGE representing the proportion of owner units built prior to 1940. In addition there are eight neighborhood variables: B, the proportion of blacks in the population; LSTAT, the proportion of population that is lower status; CRIM, the crime rate; ZN, the proportion of 25 000 square feet residential lots; INDUS, the proportion of nonretail business acres; TAX, the full value property tax rate ($/$10 000); PTRATIO, the pupil–teacher ratio; and CHAS, the dummy variable for Charles River = 1 if a tract bounds the Charles. There are also two accessibility variables, DIS the weighted distances to five employment centers in the Boston region and RAD the index of accessibility to radial highways. One more regressor is an air pollution variable NOX, the annual average nitrogen oxide concentration in parts per hundred million.[6] Moulton (1987) performed the Breusch and Pagan (1980) Lagrange multiplier (LM) test on this data set and found compelling evidence against the exclusion of random group effects.[7]

Table 9.1 shows the OLS, Within, ANOVA, ML, REML and MIVQUE-type estimates using the entire data set of 506 observations for 92 towns. Unlike the drastic difference between OLS and the Within estimators which were analyzed in Moulton (1987), the various ANOVA, MLE

Table 9.1 One-way Unbalanced Variance Components Estimates for the Harrison–Rubinfeld Hedonic Housing Equation. Dependent Variable: MV

	OLS	Within	SA	WH	HFB	ML	REML	MQ0	MQA
Intercept	9.76	—	9.68	9.68	9.67	9.68	9.67	9.68	9.67
	(0.15)		(0.21)	(0.21)	(0.21)	(0.21)	(0.21)	(0.21)	(0.21)
CRIM	−0.12	−0.63	−0.72	−0.74	−0.72	−0.72	−0.71	−0.73	−0.71
	(0.12)	(0.10)	(0.10)	(0.11)	(0.10)	(0.10)	(0.10)	(0.11)	(0.10)
ZN	0.08	—	0.04	0.07	0.02	0.03	0.01	0.06	0.01
	(0.51)		(0.69)	(0.68)	(0.70)	(0.69)	(0.71)	(0.69)	(0.71)
INDUS	0.02	—	0.21	0.16	0.24	0.22	0.24	0.18	0.24
	(0.24)		(0.43)	(0.43)	(0.45)	(0.44)	(0.46)	(0.43)	(0.45)
CHAS	0.91	−0.45	−0.01	−0.06	−0.13	−0.12	−0.14	−0.08	−0.14
	(0.33)	(0.30)	(0.29)	(0.30)	(0.29)	(0.29)	(0.29)	(0.30)	(0.29)
NOX	−0.64	−0.56	−0.59	−0.58	−0.59	−0.59	−0.59	−0.59	−0.59
	(0.11)	(0.14)	(0.12)	(0.13)	(0.12)	(0.12)	(0.12)	(0.13)	(0.12)
RM	0.63	0.93	0.92	0.91	0.92	0.92	0.92	0.91	0.92
	(0.13)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)
AGE	0.09	−1.41	−0.93	−0.87	−0.96	−0.94	−0.97	−0.90	−0.96
	(0.53)	(0.49)	(0.46)	(0.49)	(0.46)	(0.46)	(0.46)	(0.48)	(0.46)
DIS	−1.91	0.80	−1.33	−1.42	−1.27	−1.30	−1.25	−1.38	−1.26
	(0.33)	(0.71)	(0.46)	(0.46)	(0.46)	(0.47)	(0.47)	(0.46)	(0.47)
RAD	0.96	—	0.97	0.96	0.97	0.97	0.98	0.96	0.97
	(0.19)		(0.28)	(0.28)	(0.29)	(0.28)	(0.30)	(0.28)	(0.29)
TAX	−0.42	—	−0.37	−0.38	−0.37	−0.37	−0.37	−0.38	−0.37
	(0.12)		(0.19)	(0.19)	(0.19)	(0.19)	(0.20)	(0.19)	(0.20)
PTRATIO	−3.11	—	−2.97	−2.95	−2.99	−2.98	−2.99	−2.96	−2.99
	(0.50)		(0.98)	(0.96)	(1.01)	(0.98)	(1.02)	(0.97)	(1.02)
B	0.36	0.66	0.58	0.57	0.58	0.58	0.58	0.57	0.58
	(0.10)	(0.10)	(0.10)	(0.11)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
LSTAT	−3.71	−2.45	−2.85	−2.90	−2.82	−2.84	−2.82	−2.88	−2.82
	(0.25)	(0.26)	(0.24)	(0.25)	(0.24)	(0.24)	(0.24)	(0.25)	(0.24)
$\hat{\sigma}_\nu^2$	—	—	0.017	0.020	0.017	0.017	0.017	0.019	0.017
$\hat{\sigma}_\mu^2$	—	—	0.017	0.016	0.019	0.018	0.020	0.017	0.020

* Approximate standard errors are given in parentheses. $n = 506$ observations for $N = 92$ towns.
Source: Baltagi and Chang (1994). Reproduced by permission of Elsevier Science Publishers B.V. (North-Holland).

and MIVQUE-type estimators reported in Table 9.1 give similar estimates. Exceptions are ZN, INDUS and CHAS estimates which vary across methods, but are all statistically insignificant. For the statistically significant variables, AGE varies from -0.87 for WH to -0.97 for REML, and DIS varies from -1.25 for REML to -1.42 for WH.[8] These results were verified using Stata and TSP. The higher the crime rate, air pollution, tax rate, pupil–teacher ratio, proportion of the population in lower status, the older the home and the greater the distance from employment centers in Boston, the lower is the median value of the house. Similarly, the more rooms a house has and the more accessible to radial highways the higher is the median value of that home. Table 9.2 produces the maximum likelihood estimates using Stata. The likelihood ratio for $H_0 : \sigma_\mu^2 = 0$ is 172.7. This is asymptotically distributed as χ_1^2 and is significant. The Breusch–Pagan LM test for H_0 is 240.8. This is asymptotically distributed as χ_1^2 and is also significant. The Hausman specification test based on the contrast between the fixed effects and random effects estimators in Stata yields a χ_8^2 statistic of 66.1 which is statistically significant. Table 9.3 reproduces the Swamy and Arora estimator using Stata.

Table 9.2 Hedonic Housing Equation: Maximum Likelihood Estimator

```
Random-effects ML regression           Number of obs      =        506
Group variable (i) : town              Number of groups   =         92

Random effects u_i ~ Gaussian         Obs per group: min  =          1
                                                     avg  =        5.5
                                                     max  =         30

                                       LR chi2(13)        =     604.11
Log likelihood  =  236.26918           Prob > chi2        =     0.0000

------------------------------------------------------------------------------
      mv |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
    crim |  -.0071948   .0010277    -7.00   0.000     -.009209   -.0051806
      zn |   .0000286   .0006894     0.04   0.967    -.0013226    .0013799
   indus |   .0022167   .0043906     0.50   0.614    -.0063887    .0108221
    chas |  -.0119739   .028971     -0.41   0.679    -.0687561    .0448083
     nox |  -.0058672   .0012282    -4.78   0.000    -.0082744     -.00346
      rm |   .0092024   .0011643     7.90   0.000     .0069204    .0114843
     age |   -.000943   .0004614    -2.04   0.041    -.0018473   -.0000387
     dis |  -.1298569   .0469261    -2.77   0.006    -.2218304   -.0378834
     rad |   .0971024   .0284233     3.42   0.001     .0413938     .152811
     tax |  -.0003741   .0001895    -1.97   0.048    -.0007456   -2.59e-06
      pt |  -.0297989   .0097987    -3.04   0.002    -.0490041   -.0105938
       b |   .5778527   .0999609     5.78   0.000      .381933    .7737724
     lst |  -.2837924     .02405   -11.80   0.000    -.3309295   -.2366552
   _cons |   9.675679   .2069417    46.76   0.000     9.270081    10.08128
------------------------------------------------------------------------------
/sigma_u  |   .1337509   .0132895    10.06   0.000      .107704    .1597979
/sigma_e  |   .1304801   .0045557    28.64   0.000     .1215512    .1394091
------------------------------------------------------------------------------
     rho |   .5123767   .0546929                      .4060176    .6178576
------------------------------------------------------------------------------
Likelihood ratio test of sigma_u=0: chibar2(01)=  172.71
Prob=chibar2 = 0.000
```

Table 9.3 Hedonic Housing Equation: Swamy–Arora Estimator

```
Random-effects GLS regression          Number of obs      =       506
Group variable (i) : town              Number of groups   =        92

R-sq:  within  = 0.6682                Obs per group: min =         1
       between = 0.8088                               avg =       5.5
       overall = 0.7875                               max =        30

Random effects u_i ~ Gaussian          Wald chi2(13)      =   1169.62
corr(u_i, X)       = 0 (assumed)       Prob > chi2        =    0.0000
                      theta
-----------------------------------------------
  min      5%       median       95%      max
0.2915   0.2915     0.5514     0.7697   0.8197

------------------------------------------------------------------------------
      mv |     Coef.   Std. Err.     z    P>|z|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
    crim |  -.0072338  .0010346   -6.99   0.000   -.0092616   -.0052061
      zn |   .0000396  .0006878    0.06   0.954   -.0013084    .0013876
   indus |   .0020794  .0043403    0.48   0.632   -.0064273    .0105861
    chas |  -.0105913  .0289598   -0.37   0.715   -.0673515     .046169
     nox |   -.005863  .0012455   -4.71   0.000   -.0083041   -.0034219
      rm |   .0091774  .0011792    7.78   0.000    .0068662    .0114885
     age |  -.0009272  .0004647   -2.00   0.046   -.0018379   -.0000164
     dis |  -.1328825  .0456826   -2.91   0.004   -.2224186   -.0433463
     rad |   .0968634  .0283495    3.42   0.001    .0412994    .1524274
     tax |  -.0003747  .000189    -1.98   0.047   -.0007452   -4.25e-06
      pt |   -.029723  .0097538   -3.05   0.002   -.0488402   -.0106059
       b |   .5750649   .101031    5.69   0.000    .3770479     .773082
     lst |  -.2851401  .0238546  -11.95   0.000   -.3318942   -.2383859
   _cons |   9.677802  .2071417   46.72   0.000    9.271811    10.08379
---------+--------------------------------------------------------------------
 sigma_u | .12973801
 sigma_e | .13024876
     rho | .49803548   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

In conclusion, for the regression coefficients, both the Monte Carlo and the empirical illustration indicate that the computationally simple ANOVA estimates compare favorably with the computationally demanding ML, REML and MQA-type estimators. For the variance components, the ANOVA methods are recommended except when ρ is large and the unbalancedness of the data is severe. For these cases, ML, REML or MQA are recommended. As a check for misspecification, one should perform at least one of the ANOVA methods and one of the ML methods to see if the estimates differ widely. This is the Maddala and Mount (1973) suggestion for the balanced data case and applies as well for the unbalanced data case.

In another application studying the damage associated with proximity to a hazardous waste site, Mendelsohn et al. (1992) use panel data on repeated single family home sales in the harbor area surrounding New Bedford, MA over the period 1969–88. Note that one observes the dependent variable, in this case the value of the house, only when an actual sale occurs.

Therefore, these data are "unbalanced" with different time-period intervals between sales, and different numbers of repeated sales for each single family house over the period observed. These comprised 780 properties and 1916 sales. Mendelsohn et al. (1992) used first-differenced and fixed effects estimation methods to control for specific individual housing characteristics. Information on the latter variables are rarely available or complete. They find a significant reduction in housing values, between 7000 and 10 000 (1989 dollars), as a result of these houses' proximity to hazardous waste sites.[9]

9.4 THE UNBALANCED TWO-WAY ERROR COMPONENT MODEL

Wansbeek and Kapteyn (1989), henceforth WK, consider the regression model with unbalanced two-way error component disturbances:

$$\begin{aligned} y_{it} &= X_{it}'\beta + u_{it} \quad i = 1, \ldots, N_t; \quad t = 1, \ldots, T \\ u_{it} &= \mu_i + \lambda_t + \nu_{it} \end{aligned} \tag{9.27}$$

where $N_t (N_t \leq N)$ denotes the number of individuals observed in year t, with $n = \sum_t N_t$. Let D_t be the $N_t \times N$ matrix obtained from I_N by omitting the rows corresponding to individuals not observed in year t. Define

$$\Delta = (\Delta_1, \Delta_2) \equiv \begin{bmatrix} D_1 & D_1\iota_N & & \\ \vdots & & \ddots & \\ D_T & & & D_T\iota_N \end{bmatrix} \tag{9.28}$$

where $\Delta_1 = (D_1', \ldots, D_T')'$ is $n \times N$ and $\Delta_2 = \text{diag}[D_t\iota_N] = \text{diag}[\iota_{N_t}]$ is $n \times T$. The matrix Δ gives the dummy-variable structure for the incomplete data model. Note that WK order the data on the N individuals in T consecutive sets, so that t runs slowly and i runs fast. This is exactly the opposite ordering that has been used so far in the text. For complete panels, $\Delta_1 = (\iota_T \otimes I_N)$ and $\Delta_2 = I_T \otimes \iota_N$.

9.4.1 The Fixed Effects Model

If the μ_i and λ_t are fixed, one has to run the regression given in (9.27) with the matrix of dummies given in (9.28). Most likely, this will be infeasible for large panels with many households or individuals and we need the familiar Within transformation. This was easy for the balanced case and extended easily to the unbalanced one-way case. However, for the unbalanced two-way case, WK showed that this transformation is a little complicated but nevertheless manageable. To see this, we need some more matrix results.

Note that $\Delta_N \equiv \Delta_1'\Delta_1 = \text{diag}[T_i]$ where T_i is the number of years individual i is observed in the panel. Also, $\Delta_T \equiv \Delta_2'\Delta_2 = \text{diag}[N_t]$ and $\Delta_{TN} \equiv \Delta_2'\Delta_1$ is the $T \times N$ matrix of zeros and ones indicating the absence or presence of a household in a certain year. For complete panels, $\Delta_N = TI_N$, $\Delta_T = NI_T$ and $\Delta_{TN} = \iota_T\iota_N' = J_{TN}$. Define $P_{[\Delta]} = \Delta(\Delta'\Delta)^-\Delta'$, then the Within transformation is $Q_{[\Delta]} = I_n - P_{[\Delta]}$. For the two-way unbalanced model with $\Delta = (\Delta_1, \Delta_2)$ given by (9.28), WK show that

$$P_{[\Delta]} = P_{\Delta_1} + P_{[Q_{[\Delta_1]}\Delta_2]} \tag{9.29}$$

The proof is sketched out in problem 9.6. Therefore,

$$Q_{[\Delta]} = Q_{[\Delta_1]} - Q_{[\Delta_1]}\Delta_2(\Delta_2' Q_{[\Delta_1]}\Delta_2)^{-}\Delta_2' Q_{[\Delta_1]} \tag{9.30}$$

Davis (2001) generalizes the WK Within transformation to the three-way, four-way and higher-order error component models. Davis shows that the Within transformation can be applied in stages to the variables in the regression, just like in (9.30). This reduces the computational burden considerably. For example, consider a three-way error component model, representing products sold at certain locations and observed over some time period. These fixed effects are captured by three dummy variables matrices $\Delta = [\Delta_1, \Delta_2, \Delta_3]$. In order to get the Within transformation, Davis (2001) applies (9.29) twice and obtains $Q_{[\Delta]} = Q_{[A]} - P_{[B]} - P_{[C]}$ where $A = \Delta_1$, $B = Q_{[A]}\Delta_2$ and $C = Q_{[B]}Q_{[A]}\Delta_3$, see problem 9.27. This idea generalizes readily to higher-order fixed effects error components models.

9.4.2 The Random Effects Model

In vector form, the incomplete two-way random effects model can be written as

$$u = \Delta_1\mu + \Delta_2\lambda + \nu \tag{9.31}$$

where $\mu' = (\mu_1, \ldots, \mu_N)$, $\lambda' = (\lambda_1, \ldots, \lambda_T)$ and ν are random variables described exactly as in the two-way error component model considered in Chapter 3. μ, λ and ν are independent of each other and among themselves with zero means and variances σ_μ^2, σ_λ^2 and σ_ν^2, respectively. In this case,

$$\begin{aligned}\Omega = E(uu') &= \sigma_\nu^2 I_n + \sigma_\mu^2\Delta_1\Delta_1' + \sigma_\lambda^2\Delta_2\Delta_2' \\ &= \sigma_\nu^2(I_n + \phi_1\Delta_1\Delta_1' + \phi_2\Delta_2\Delta_2') = \sigma_\nu^2\Sigma\end{aligned} \tag{9.32}$$

with $\phi_1 = \sigma_\mu^2/\sigma_\nu^2$ and $\phi_2 = \sigma_\lambda^2/\sigma_\nu^2$. Using the general expression for the inverse of $(I + XX')$, see problem 9.8, Wansbeek and Kapteyn (1989) obtain the inverse of Σ as

$$\Sigma^{-1} = V - V\Delta_2\widetilde{P}^{-1}\Delta_2' V \tag{9.33}$$

where

$$\begin{array}{ll} V = I_n - \Delta_1\widetilde{\Delta}_N^{-1}\Delta_1' & (n \times n) \\ \widetilde{P} = \widetilde{\Delta}_T - \Delta_{TN}\widetilde{\Delta}_N^{-1}\Delta_{TN}' & (T \times T) \\ \widetilde{\Delta}_N = \Delta_N + (\sigma_\nu^2/\sigma_\mu^2)I_N & (N \times N) \\ \widetilde{\Delta}_T = \Delta_T + (\sigma_\nu^2/\sigma_\lambda^2)I_T & (T \times T) \end{array}$$

Note that we can no longer obtain the simple Fuller and Battese (1973) transformation for the unbalanced two-way model. The expression for Σ^{-1} is messy and asymmetric in individuals and time, but it reduces computational time considerably relative to inverting Σ numerically. Davis (2001) shows that the WK results can be generalized to an arbitrary number of random error components. In fact, for a three-way random error component, like the one considered in problem 9.7, the added random error component η adds an extra $\sigma_\eta^2\Delta_3\Delta_3'$ term to the variance–covariance given in (9.32). Therefore, Σ remains of the $(I + XX')$ form and its inverse can be obtained by repeated application of this inversion formula. This idea generalizes readily to higher-order unbalanced random error component models. WK suggest an ANOVA-type quadratic unbiased estimator (QUE) of the variance components based on the Within residuals.

In fact, the MSE of the Within regression is unbiased for σ_ν^2 even under the random effects specification. Let $e = y - X\widetilde{\beta}$ where $\widetilde{\beta}$ denotes the Within estimates and define

$$q_W = e'Q_{[\Delta]}e \tag{9.34}$$

$$q_N = e'\Delta_2\Delta_T^{-1}\Delta_2'e = e'P_{[\Delta_2]}e \tag{9.35}$$

$$q_T = e'\Delta_1\Delta_N^{-1}\Delta_1'e = e'P_{[\Delta_1]}e \tag{9.36}$$

By equating q_W, q_N and q_T to their expected values and solving these three equations one gets QUE of σ_ν^2, σ_μ^2 and σ_λ^2. WK also derive the ML iterative first-order conditions as well as the information matrix under normality of the disturbances. These will not be reproduced here and the reader is referred to the WK article for details. A limited Monte Carlo experiment was performed using 50 replications and three kinds of data designs: complete panel data, 20% random attrition and a rotating panel. This was done using a simple regression with a Nerlove type X for fixed $\sigma_\mu^2 = 400$, $\sigma_\lambda^2 = 25$ and $\sigma_\nu^2 = 25$. The regression coefficients were fixed at $\alpha = 25$ and $\beta = 2$, and the number of individuals and time periods were $N = 100$ and $T = 5$, respectively. The results imply that the QUE of the variance components are in all cases at least as close to the true value as the MLE so that iteration on these values does not seem to pay off. Also, GLS gives nearly identical results to MLE as far as the regression coefficient estimates are concerned. Therefore, WK recommend GLS over MLE in view of the large difference in computational cost.

Baltagi, Song and Jung (2002a) reconsider the unbalanced two-way error component given in (9.27) and (9.28) and provide alternative analysis of variance (ANOVA), minimum norm quadratic unbiased (MINQUE) and restricted maximum likelihood (REML) estimation procedures. These are similar to the methods studied in section 9.2 for the unbalanced one-way error component model. The mean squared error performance of these estimators is compared using Monte Carlo experiments. Focusing on the estimates of the variance components, the computationally more demanding MLE, REML, MIVQUE estimators are recommended especially if the unbalanced pattern is severe. However, focusing on the regression coefficients estimates, the simple ANOVA methods perform just as well as the computationally demanding MLE, REML and MIVQUE methods and are recommended.

9.5 TESTING FOR INDIVIDUAL AND TIME EFFECTS USING UNBALANCED PANEL DATA

In Chapter 4, we derived the Breusch and Pagan (1980) LM test for $H_0 : \sigma_\mu^2 = \sigma_\lambda^2 = 0$ in a complete panel data model with two-way error component disturbances. Baltagi and Li (1990) derived the corresponding LM test for the unbalanced two-way error component model. This model is given by (9.27) and the variance–covariance matrix of the disturbances is given by (9.32). Following the same derivations as given in section 4.2 (see problem 9.8), one can show that under normality of the disturbances

$$\partial\Omega/\partial\sigma_\mu^2 = \Delta_1\Delta_1', \quad \partial\Omega/\partial\sigma_\lambda^2 = \Delta_2\Delta_2' \quad \text{and} \quad \partial\Omega/\partial\sigma_\nu^2 = I_n \tag{9.37}$$

with

$$\text{tr}(\Delta_2\Delta_2') = \text{tr}(\Delta_2'\Delta_2) = \text{tr}(\text{diag}[N_t]) = \sum_{t=1}^{T} N_t = n \tag{9.38}$$

and

$$\operatorname{tr}(\Delta_1'\Delta_1) = \operatorname{tr}(\operatorname{diag}[T_i]) = \sum_{i=1}^{N} T_i = n \tag{9.39}$$

Substituting these results in (4.17) and noting that under H_0, $\widetilde{\Omega}^{-1} = (1/\widetilde{\sigma}_\nu^2)I_n$, where $\widehat{\sigma}_\nu^2 = \widetilde{u}'\widetilde{u}/NT$ and $\widetilde{u}$ denote the OLS residuals, one gets

$$\widetilde{D} = (\partial L/\partial\theta)\mid_{\theta=\widetilde{\theta}} = (n/2\widetilde{\sigma}_\nu^2)\begin{bmatrix} A_1 \\ A_2 \\ 0 \end{bmatrix} \tag{9.40}$$

where $\theta' = (\sigma_\mu^2, \sigma_\lambda^2, \sigma_\nu^2)$ and $\widetilde{\theta}$ denotes the restricted MLE of θ under H_0. Also, $A_r = [(\widetilde{u}'\Delta_r\Delta_r'\widetilde{u}/\widetilde{u}'\widetilde{u}) - 1]$ for $r = 1, 2$. Similarly, one can use (4.19) to obtain the information matrix

$$\widetilde{J} = (n/2\widetilde{\sigma}_\nu^4)\begin{bmatrix} M_{11}/n & 1 & 1 \\ 1 & M_{22}/n & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{9.41}$$

where $M_{11} = \sum_{i=1}^{N} T_i^2$ and $M_{22} = \sum_{t=1}^{T} N_t^2$. This makes use of the fact that

$$\operatorname{tr}(\Delta_2\Delta_2')^2 = \sum_{t=1}^{T} N_t^2, \quad \operatorname{tr}(\Delta_1\Delta_1')^2 = \sum_{i=1}^{N} T_i^2 \tag{9.42}$$

and

$$\operatorname{tr}[(\Delta_1\Delta_1')(\Delta_2\Delta_2')] = \sum_{t=1}^{T} \operatorname{tr}[(D_t D_t')J_{N_t}] = \sum_{t=1}^{T} \operatorname{tr}(J_{N_t}) = \sum_{t=1}^{T} N_t = n$$

Using (9.40) and (9.41) one gets the LM statistic

$$\mathrm{LM} = \widetilde{D}'\widetilde{J}^{-1}\widetilde{D} = (\tfrac{1}{2})n^2[A_1^2/(M_{11} - n) + A_2^2/(M_{22} - n)] \tag{9.43}$$

which is asymptotically distributed as χ_2^2 under the null hypothesis. For computational purposes, one need not form the Δ_r matrices to compute $A_r(r = 1, 2)$. In fact,

$$\widetilde{u}\Delta_1\Delta_1'\widetilde{u} = \sum_{i=1}^{N} \widetilde{u}_{i.}^2 \quad \text{where} \quad \widetilde{u}_{i.} = \sum_{t=1}^{T_i} \widetilde{u}_{it} \tag{9.44}$$

and

$$\widetilde{u}'\Delta_2\Delta_2'\widetilde{u} = \sum_{t=1}^{T} \widetilde{u}_{.t}^2 \quad \text{where} \quad \widetilde{u}_{.t} = \sum_{i=1}^{N_t} \widetilde{u}_{it} \tag{9.45}$$

(9.45) is obvious, since $\Delta_2 = \operatorname{diag}[\iota_{N_t}]$, and (9.44) can be similarly obtained, by restacking the residuals such that the faster index is t. The LM statistic given in (9.43) is easily computed using least squares residuals, and retains a similar form to that of the complete panel data case. In fact, when $N_t = N$, (9.43) reverts back to the LM statistic derived in Breusch and Pagan (1980). Also, (9.43) retains the additive property exhibited in the complete panel data case, i.e. if $H_0 : \sigma_\mu^2 = 0$, the LM test reduces to the first term of (9.43), whereas if $H_0 : \sigma_\lambda^2 = 0$, the LM test reduces to the second term of (9.43). Both test statistics are asymptotically distributed as χ_1^2 under the respective null hypotheses.

These variance components cannot be negative and therefore $H_0 : \sigma_\mu^2 = 0$ has to be against a one-sided alternative $H_1 : \sigma_\mu^2 > 0$. Moulton and Randolph (1989) derived the one-sided LM_1 statistic

$$\text{LM}_1 = n[2(M_{11} - n)]^{-1/2} A_1 \tag{9.46}$$

which is the square root of the first term in (9.43). Under weak conditions as $n \to \infty$ and $N \to \infty$ the LM_1 statistic has an asymptotic standard normal distribution under H_0. However, Moulton and Randolph (1989) showed that this asymptotic $N(0, 1)$ approximation can be poor even in large samples. This occurs when the number of regressors is large or the intraclass correlation of some of the regressors is high. They suggest an alternative standardized Lagrange multiplier SLM given by

$$\text{SLM} = \frac{\text{LM}_1 - E(\text{LM}_1)}{\sqrt{\text{var}(\text{LM}_1)}} = \frac{d - E(d)}{\sqrt{\text{var}(d)}} \tag{9.47}$$

where $d = (\tilde{u}' D \tilde{u})/\tilde{u}'\tilde{u}$ and $D = \Delta_1 \Delta_1'$. Using the results on moments of quadratic forms in regression residuals (see, for example, Evans and King, 1985), we get

$$E(d) = \text{tr}(D\bar{P}_Z)/p$$

where $p = [n - (K + 1)]$ and

$$\text{var}(d) = 2\{p\ \text{tr}(D\bar{P}_Z)^2 - [\text{tr}(D\bar{P}_Z)]^2\}/p^2(p + 2)$$

Under H_0, this SLM has the same asymptotic $N(0, 1)$ distribution as the LM_1 statistic. However, the asymptotic critical values for the SLM are generally closer to the exact critical values than those of the LM_1 statistic. Similarly, for $H_0 : \sigma_\lambda^2 = 0$, the one-sided LM test statistic is the square root of the second term in (9.43), i.e.

$$\text{LM}_2 = n[2(M_{22} - n)]^{-1/2} A_2 \tag{9.48}$$

Honda's (1985) "handy" one-sided test for the two-way model with unbalanced data is simply

$$\text{HO} = (\text{LM}_1 + \text{LM}_2)/\sqrt{2}$$

It is also easy to show, see Baltagi, Chang and Li (1998), that the locally mean most powerful (LMMP) one-sided test suggested by King and Wu (1997) for the unbalanced two-way error component model is given by

$$\text{KW} = \frac{\sqrt{M_{11} - n}}{\sqrt{M_{11} + M_{22} - 2n}} \text{LM}_1 + \frac{\sqrt{M_{22} - n}}{\sqrt{M_{11} + M_{22} - 2n}} \text{LM}_2 \tag{9.49}$$

where LM_1 and LM_2 are given by (9.46) and (9.48), respectively. Both HO and KW are asymptotically distributed as $N(0, 1)$ under H_0. These test statistics can be standardized and the resulting SLM given by $\{d - E(d)\}/\sqrt{\text{var}(d)}$ where $d = \tilde{u}' D \tilde{u}/\tilde{u}'\tilde{u}$ with

$$D = \frac{1}{2} \frac{n}{\sqrt{M_{11} - n}} (\Delta_1 \Delta_1') + \frac{1}{2} \frac{n}{\sqrt{M_{22} - n}} (\Delta_2 \Delta_2') \tag{9.50}$$

for Honda's (1985) version, and

$$D = \frac{n}{\sqrt{2}\sqrt{M_{11} + M_{22} - 2n}} [(\Delta_1 \Delta_1') + (\Delta_2 \Delta_2')] \tag{9.51}$$

for the King and Wu (1997) version of this test. $E(d)$ and $\mathrm{var}(d)$ are obtained from the same formulas shown below (9.47) using the appropriate D matrices.

Since LM_1 and LM_2 can be negative for a specific application, especially when one or both variance components are small and close to zero, one can use the Gourieroux et al. (1982) (GHM) test which is given by

$$\chi_m^2 = \begin{cases} \mathrm{LM}_1^2 + \mathrm{LM}_2^2 & \text{if } \mathrm{LM}_1 > 0, \mathrm{LM}_2 > 0 \\ \mathrm{LM}_1^2 & \text{if } \mathrm{LM}_1 > 0, \mathrm{LM}_2 \leq 0 \\ \mathrm{LM}_2^2 & \text{if } \mathrm{LM}_1 \leq 0, \mathrm{LM}_2 > 0 \\ 0 & \text{if } \mathrm{LM}_1 \leq 0, \mathrm{LM}_2 \leq 0 \end{cases} \tag{9.52}$$

χ_m^2 denotes the mixed χ^2 distribution. Under the null hypothesis,

$$\chi_m^2 \sim \left(\tfrac{1}{4}\right)\chi^2(0) + \left(\tfrac{1}{2}\right)\chi^2(1) + \left(\tfrac{1}{4}\right)\chi^2(2)$$

where $\chi^2(0)$ equals zero with probability one.[11] The weights $\frac{1}{4}$, $\frac{1}{2}$ and $\frac{1}{4}$ follow from the fact that LM_1 and LM_2 are asymptotically independent of each other and the results in Gourieroux et al. (1982). This proposed test has the advantage over the Honda and KW tests in that it is immune to the possible negative values of LM_1 and LM_2.

Baltagi et al. (1998) compare the performance of these tests using Monte Carlo experiments for an unbalanced two-way error component model. The results of the Monte Carlo experiments show that the nominal sizes of the Honda and King–Wu tests based on asymptotic critical values are inaccurate for all unbalanced patterns considered. However, the nominal size of the standardized version of these tests is closer to the true significance value and is recommended. This confirms similar results for the unbalanced one-way error component model by Moulton and Randolph (1989). In cases where at least one of the variance components is close to zero, the Gourieroux et al. (1982) test is found to perform well in Monte Carlo experiments and is recommended. All the tests considered have larger power as the number of individuals N in the panel and/or the variance components increase. In fact, for typical labor or consumer panels with large N, the Monte Carlo results show that the power of these tests is one except for cases where the variance components comprise less than 10% of the total variance.[12]

9.6 THE UNBALANCED NESTED ERROR COMPONENT MODEL

Baltagi, Song and Jung (2001) extend the ANOVA, MINQUE and MLE estimation procedures described in section 9.2 to the unbalanced nested error component regression model. For this model, the incomplete panel data exhibits a natural nested grouping. For example, data on firms may be grouped by industry, data on states by region, data on individuals by profession and data on students by schools.[13] The unbalanced panel data regression model is given by

$$y_{ijt} = x'_{ijt}\beta + u_{ijt} \quad i = 1, \ldots, M;\ j = 1, \ldots, N_i;\ t = 1, \ldots, T_i \tag{9.53}$$

where y_{ijt} could denote the output of the jth firm in the ith industry for the tth time period. x_{ijt} denotes a vector of k nonstochastic inputs. The disturbances are given by

$$u_{ijt} = \mu_i + \nu_{ij} + \varepsilon_{ijt} \quad i = 1, \ldots, M;\ j = 1, \ldots, N_i;\ t = 1, \ldots, T_i \tag{9.54}$$

where μ_i denotes the ith unobservable industry-specific effect which is assumed to be $\mathrm{IID}(0, \sigma_\mu^2)$, ν_{ij} denotes the nested effect of the jth firm within the ith industry which is assumed to be $\mathrm{IID}(0, \sigma_\nu^2)$, and ε_{ijt} denotes the remainder disturbance which is also assumed to be IID(0, σ_ε^2). The μ_i's, ν_{ij}'s and ε_{ijt}'s are independent of each other and among themselves. This is a nested classification in that each successive component of the error term is imbedded or "nested" within the preceding component, see Graybill (1961, p. 350). This model allows for unequal number of firms in each industry as well as different number of observed time periods across industries. Detailed derivation of the variance–covariance matrix Ω, the Fuller and Battese (1973) transformation, as well as ANOVA, MINQUE and MLE methods are given in Baltagi et al. (2001) and will not be reproduced here. Baltagi et al. (2001) compared the performance of these estimators using Monte Carlo experiments. While the MLE and MIVQUE methods perform the best in estimating the variance components and the standard errors of the regression coefficients, the simple ANOVA methods perform just as well in estimating the regression coefficients. These estimation methods are also used to investigate the productivity of public capital in private production. In a companion paper, Baltagi, Song and Jung (2002b) extend the Lagrange multiplier tests described in section 9.4 to the unbalanced nested error component model. Later, Baltagi, Song and Jung (2002c) derived the Lagrange multiplier tests for the unbalanced nested error component model with serially correlated disturbances.

9.6.1 Empirical Example

In Chapter 2, example 3, we estimated a Cobb–Douglas production function investigating the productivity of public capital in each state's private output. This was based on a panel of 48 states over the period 1970–86. The data was provided by Munnell (1990). Here we group these states into nine geographical regions with the Middle Atlantic region for example containing three states: New York, New Jersey and Pennsylvania and the Mountain region containing eight states: Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah and Nevada. In this case, the primary group would be the regions, the nested group would be the states and these are observed over 17 years. The dependent variable y is the gross state product and the regressors include the private capital stock (K) computed by apportioning the Bureau of Economic Analysis (BEA) national estimates. The public capital stock is measured by its components: highways and streets (KH), water and sewer facilities (KW), and other public buildings and structures (KO), all based on the BEA national series. Labor (L) is measured by the employment in nonagricultural payrolls. The state unemployment rate is included to capture the business cycle in a given state. All variables except the unemployment rate are expressed in natural logarithms:

$$y_{ijt} = \alpha + \beta_1 \mathrm{K}_{ijt} + \beta_2 \mathrm{KH}_{ijt} + \beta_3 \mathrm{KW}_{ijt} + \beta_4 \mathrm{KO}_{ijt} + \beta_5 \mathrm{L}_{ijt} + \beta_6 \mathrm{Unemp}_{ijt} + u_{it} \quad (9.55)$$

where $i = 1, 2, \ldots, 9$ regions, $j = 1, \ldots, N_i$ with N_i equaling three for the Middle Atlantic region and eight for the Mountain region and $t = 1, 2, \ldots, 17$. The data is unbalanced only in the differing number of states in each region. The disturbances follow the nested error component specification given by (9.54).

Table 9.4 gives the OLS, Within, ANOVA, MLE, REML and MIVQUE-type estimates using this unbalanced nested error component model. The OLS estimates show that the highways and streets and water and sewer components of public capital have a positive and significant effect upon private output whereas that of other public buildings and structures is not significant. Because OLS ignores the state and region effects, the corresponding standard errors and t-statistics are biased, see Moulton (1986). The Within estimator shows that the effects of

Table 9.4 Cobb–Douglas Production Function Estimates with Unbalanced Nested Error Components 1970–86, 9 Regions, 48 States[a]

Variable	OLS	Within	WH	WK	SA	HFB	MLE	REML	MV1	MV2	MV3
Intercept	1.926	—	2.082	2.131	2.089	2.084	2.129	2.127	2.083	2.114	2.127
	(0.053)		(0.152)	(0.160)	(0.144)	(0.150)	(0.154)	(0.157)	(0.152)	(0.154)	(0.156)
K	0.312	0.235	0.273	0.264	0.274	0.272	0.267	0.266	0.272	0.269	0.267
	(0.011)	(0.026)	(0.021)	(0.022)	(0.020)	(0.021)	(0.021)	(0.022)	(0.021)	(0.021)	(0.021)
L	0.550	0.801	0.742	0.758	0.740	0.743	0.754	0.756	0.742	0.750	0.755
	(0.016)	(0.030)	(0.026)	(0.027)	(0.025)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
KH	0.059	0.077	0.075	0.072	0.073	0.075	0.071	0.072	0.075	0.072	0.072
	(0.015)	(0.031)	(0.023)	(0.024)	(0.022)	(0.022)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
KW	0.119	0.079	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076
	(0.012)	(0.015)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
KO	0.009	−0.115	−0.095	−0.102	−0.094	−0.096	−0.100	−0.101	−0.095	−0.098	−0.100
	(0.012)	(0.018)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Unemp	−0.007	−0.005	−0.006	−0.006	−0.006	−0.006	−0.006	−0.006	−0.006	−0.006	−0.006
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
σ_ε^2	0.0073	0.0013	0.0014	0.0014	0.0014	0.0014	0.0013	0.0014	0.0014	0.0014	0.0014
σ_μ^2	—	—	0.0027	0.0022	0.0015	0.0029	0.0015	0.0019	0.0027	0.0017	0.0017
σ_ν^2	—	—	0.0045	0.0069	0.0043	0.0044	0.0063	0.0064	0.0046	0.0056	0.0063

[a] The dependent variable is log of gross state product. Standard errors are given in parentheses.
Source: Baltagi et al. (2001). Reproduced by permission of Elsevier Science Publishers B.V. (North-Holland).

KH and KW are insignificant whereas that of KO is negative and significant. The primary region and nested state effects are significant using several LM tests developed in Baltagi et al. (2002b). This justifies the application of the feasible GLS, MLE and MIVQUE methods. For the variance components estimates, there are no differences in the estimate of σ_ε^2, but estimates of σ_μ^2 and σ_ν^2 vary. $\widehat{\sigma}_\mu^2$ is as low as 0.0015 for SA and MLE and as high as 0.0029 for HFB. Similarly, $\widehat{\sigma}_\nu^2$ is as low as 0.0043 for SA and as high as 0.0069 for WK. This variation had little effect on estimates of the regression coefficients or their standard errors. For all estimators of the random effects model, the highways and streets and water and sewer components of public capital had a positive and significant effect, while the other public buildings and structures had a negative and significant effect upon private output. These results were verified using TSP.

Other empirical applications of the nested error component model include Montmarquette and Mahseredjian (1989) who study whether schooling matters in educational achievements in Montreal's Francophone public elementary schools. Also, Antweiler (2001) who derives the maximum likelihood estimator for an unbalanced nested three-way error component model. This is applied to the problem of explaining the determinants of pollution concentration (measured by the log of atmospheric sulfuric dioxide) at 293 observation stations located in 44 countries over the time period 1971–96. This data is highly unbalanced in that out of a total of 2621 observations, about a third of these are from stations in one country, the United States. Also, the time period of observation is not necessarily continuous. Comparing the results of maximum likelihood for a nested vs a simple (nonnested) unbalanced error component model, Antweiler (2001) finds that the scale elasticity coefficient estimate for the nested model is less than half that for the nonnested model. Scale elasticity is the coefficient of log of economic intensity as measured by GDP per square kilometer. This is also true for the estimate of the income effect which is negative and much lower in absolute value for the nested model than the nonnested model. Finally, the estimate of the composition effect which is the coefficient of the log of the country's capital abundance is higher for the nested model than for the nonnested model.

Davis (2001) applies OLS, Within, MIVQUE and MLE procedures to a three-way unbalanced error component model using data on film revenues for six movie theaters near New Haven, CT, observed over a six-week period in 1998. Some of the reasons for unbalancedness in the data occur because (i) not all films are shown at all locations, (ii) films start and stop being shown at theaters during the observation period, and (iii) data on revenues are missing due to nonresponse. The estimates obtained reveal a complex set of asymmetric cross-theater price elasticities of demand. These estimates are useful for the analysis of the impact of mergers on pricing, and for determining the appropriate extent of geographic market definition in these markets.

NOTES

1. Other methods of dealing with missing data include: (i) inputing the missing values and analyzing the filled-in data by complete panel data methods; (ii) discarding the nonrespondents and weighting the respondents to compensate for the loss of cases; see Little (1988) and the section on nonresponse adjustments in Kasprzyk et al. (1989).
2. This analysis assumes that the observations of the individual with the shortest time series are nested in a specific manner within the observations of the other individual. However, the same derivations apply for different types of overlapping observations.
3. Note that if the updated value is negative, it is replaced by zero and the iteration continues until the convergence criterion is satisfied.

4. It is important to note that ML and restricted ML estimates of the variance components are by definition nonnegative. However, ANOVA and MINQUE methods can produce negative estimates of the variance component σ_μ^2. In these cases, the negative variance estimates are replaced by zero. This means that the resulting variance component estimator is $\widetilde{\sigma}_\mu^2 = \max(\widehat{\sigma}_\mu^2, 0)$ which is no longer unbiased.
5. Problem 90.2.3 in *Econometric Theory* by Baltagi and Li (1990) demonstrated analytically that for a random error component model, one can construct a simple unbiased estimator of the variance components using the entire unbalanced panel that is more efficient than the BQU estimator using only the subbalanced pattern (see problem 9.5). Also, Chowdhury (1991) showed that for the fixed effects error component model, the Within estimator based on the entire unbalanced panel is efficient relative to any Within estimator based on a subbalanced pattern. Mátyás and Lovrics (1991) performed some Monte Carlo experiments to compare the loss in efficiency of Within and GLS based on the entire incomplete panel data and complete subpanel. They found the loss in efficiency is negligible if $NT > 250$, but serious for $NT < 150$.
6. The variable descriptions are from table IV of Harrison and Rubinfeld (1978). See Belsley et al. (1980) for a listing of the data and further diagnostic analysis of these data. Moulton (1986) used these data to show the inappropriate use of OLS in the presence of random group effects and Moulton (1987) applied a battery of diagnostic tools to this data set.
7. Later, Moulton and Randolph (1989) found that asymptotic critical values of the one-sided LM test can be very poor, and suggested a standardized LM test whose asymptotic critical value approximations are likely to be much better than those of the LM statistic. They applied it to this data set and rejected the null hypothesis of no random group effect using an exact critical value.
8. Note that the Amemiya-type estimator is not calculated for this data set since there are some regressors without Within variation.
9. Another application of unbalanced panels include the construction of a number of quality-adjusted price indexes for personal computers in the USA over the period 1989–92, see Berndt, Griliches and Rappaport (1995).
10. If the data were arranged differently, one would get the generalized inverse of an $N \times N$ matrix rather than that of a $T \times T$ one as in P. Since $N > T$ in most cases, this choice is most favorable from the point of view of computations.
11. Critical values for the mixed χ_m^2 are 7.289, 4.321 and 2.952 for $\alpha = 0.01, 0.05$ and 0.1, respectively.
12. A Gauss program for testing individual and time effects using unbalanced panel data is given in the appendix of Baltagi et al. (1998, pp. 16–19).
13. See problem 3.14 for an introduction to the balanced nested error component model.

PROBLEMS

9.1 (a) Show that the variance–covariance matrix of the disturbances in (9.1) is given by (9.2).
(b) Show that the two nonzero block matrices in (9.2) can be written as in (9.3).
(c) Show that $\sigma_\nu \Omega_j^{-1/2} y_j$ has a typical element $(y_{jt} - \theta_j \bar{y}_{j.})$, where $\theta_j = 1 - \sigma_\nu/\omega_j$ and $\omega_j^2 = T_j \sigma_\mu^2 + \sigma_\nu^2$.

9.2 (a) Verify the $E(\widehat{q}_1)$ and $E(\widehat{q}_2)$ equations given in (9.16).
(b) Verify $E(\widetilde{q}_1)$ and $E(\widetilde{q}_2)$ given in (9.17).
(c) Verify $E(\hat{q}_2^b)$ given in (9.19).

9.3 Using the Monte Carlo set-up for the unbalanced one-way error component model considered by Baltagi and Chang (1994), compare the various estimators of the variance components and the regression coefficients considered in section 9.2.4.

9.4 Using the Harrison and Rubinfeld (1978) data published in Belsley et al. (1980) and provided on the Wiley web site as Hedonic.xls, reproduce Table 9.1. Perform the Hausman test based on the fixed effects and the random effects contrast. Perform the LM test for $H_0 : \sigma_\mu^2 = 0$.

9.5 This exercise is based on problem 90.2.3 in *Econometric Theory* by Baltagi and Li (1990). Consider the following unbalanced one-way analysis of variance model:

$$y_{it} = \mu_i + \nu_{it} \quad i = 1, \ldots, N; \quad t = 1, 2, \ldots, T_i$$

where for simplicity's sake no explanatory variables are included. y_{it} could be the output of firm i at time period t and μ_i could be the managerial ability of firm i, whereas ν_{it} is a remainder disturbance term. Assume that $\mu_i \sim \text{IIN}(0, \sigma_\mu^2)$ and $\nu_{it} \sim\text{IIN}(0, \sigma_\nu^2)$ independent of each other. Let T be the maximum overlapping period over which a complete panel could be established ($T \leq T_i$ for all i). In this case the corresponding vector of balanced observations on y_{it} is denoted by y_b and is of dimension NT. One could estimate the variance components using this complete panel as follows:

$$\widehat{\sigma}_\nu^2 = y_b'(I_N \otimes E_T)y_b/N(T-1)$$

and

$$\sigma_\mu^2 = [y_b'(I_N \otimes \bar{J}_T)y_b/NT] - (\widehat{\sigma}_\nu^2/T)$$

where $E_T = I_T - \bar{J}_T$, $\bar{J}_T = J_T/T$ and J_T is a matrix of ones of dimension T. $\widehat{\sigma}_\nu^2$ and $\widehat{\sigma}_\mu^2$ are the best quadratic unbiased estimators (BQUE) of the variance components based on the complete panel. Alternatively, one could estimate the variance components from the entire unbalanced panel as follows:

$$\widetilde{\sigma}_\nu^2 = y'\text{diag}(E_{T_i})y/(n-N)$$

where $n = \sum_{i=1}^N T_i$ and $E_{T_i} = I_{T_i} - \bar{J}_{T_i}$. Also, $\sigma_i^2 = (T_i\sigma_\mu^2 + \sigma_\nu^2)$ can be estimated by $\widetilde{\sigma}_i^2 = y_i'\bar{J}_{T_i}y_i$, where y_i denotes the vector of T_i observations on the ith individual. Therefore, there are N estimators of σ_μ^2 obtained from $(\widetilde{\sigma}_i^2 - \widetilde{\sigma}_\nu^2)/T_i$ for $i = 1, \ldots, N$. One simple way of combining them is to take the average

$$\widetilde{\sigma}_\mu^2 = \sum_{i=1}^N [(\widetilde{\sigma}_i^2 - \widetilde{\sigma}_\nu^2)/T_i]/N = \left\{ y'\text{diag}[\bar{J}_{T_i}/T_i]y - \sum_{i=1}^N \widetilde{\sigma}_\nu^2/T_i \right\}/N$$

(a) Show that $\widetilde{\sigma}_\nu^2$ and $\widetilde{\sigma}_\mu^2$ are unbiased estimators σ_ν^2 and σ_μ^2.
(b) Show that $\text{var}(\widetilde{\sigma}_\nu^2) \leq \text{var}(\widehat{\sigma}_\nu^2)$ and $\text{var}(\widetilde{\sigma}_\mu^2) \leq \text{var}(\widehat{\sigma}_\mu^2)$. (*Hint:* See solution 90.2.3 in *Econometric Theory* by Koning (1991).)

9.6 For $X = (X_1, X_2)$, the generalized inverse of $(X'X)$ is given by

$$(X'X)^- = \begin{bmatrix} (X_1'X_1)^- & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -(X_1'X_1)^- X_1'X_2 \\ I \end{bmatrix} (X_2'Q_{[X_1]}X_2)^-[-X_2'X_1(X_1'X_1)^- I]$$

see Davis (2001), appendix A. Use this result to show that $P_{[X]} = P_{[X_1]} + P_{[Q_{[X_1]}X_2]}$. (*Hint*: Premultiply this expression by X, and postmultiply by X'.) This verifies (9.29).

9.7 Consider the three-way error component model described in problem 3.15. The panel data can be unbalanced and the matrices of dummy variables are $\Delta = [\Delta_1, \Delta_2, \Delta_3]$ with

$$u = \Delta_1\mu + \Delta_2\lambda + \Delta_3\eta + \nu$$

where μ, λ and ν are random variables defined below (9.31) and the added random error η has mean zero and variance σ_η^2. All random errors are independent among themselves and with each other. Show that $P_{[\Delta]} = P_{[A]} + P_{[B]} + P_{[C]}$ where $A = \Delta_1$, $B = Q_{[A]}\Delta_2$

and $C = Q_{[B]}Q_{[A]}\Delta_3$. This is Corollary 1 of Davis (2001). (*Hint*: Apply (9.29) twice. Let $X_1 = \Delta_1$ and $X_2 = (\Delta_2, \Delta_3)$. Using problem 9.6, we get $P_{[X]} = P_{[\Delta_1]} + P_{[Q_{[\Delta_1]}X_2]}$. Now, $Q_{[\Delta_1]}X_2 = Q_{[\Delta_1]}(\Delta_2, \Delta_3) = [B, Q_{[A]}\Delta_3]$. Applying (9.29) again we get $P_{[B, Q_{[A]}\Delta_3]} = P_{[B]} + P_{[Q_{[B]}Q_{[A]}\Delta_3]}$.)

9.8 (a) For Δ_1 and Δ_2 defined in (9.28), verify that $\Delta_N \equiv \Delta_1'\Delta_1 = \text{diag}[T_i]$ and $\Delta_T \equiv \Delta_2'\Delta_2 = \text{diag}[N_t]$. Show that for the complete panel data case $\Delta_1 = \iota_T \otimes I_N$, $\Delta_2 = I_T \otimes \iota_N$, $\Delta_N = TI_N$ and $\Delta_T = NI_T$.
(b) Under the complete panel data case, verify that $\Delta_{TN} \equiv \Delta_2'\Delta_1$ is J_{TN} and $Q = E_T \otimes E_N$, see Chapter 3, equation (3.3) and problem 3.1.
(c) Let $X = (X_1, X_2)$ with $|I + XX'| \neq 0$. Using the result that $[I_n + XX']^{-1} = I_n - X(I + X'X)^{-1}X'$, apply the partitioned inverse formula for matrices to show that $(I + XX')^{-1} = \widetilde{Q}_{[X_2]} - \widetilde{Q}_{[X_2]}X_1S^{-1}X_1'\widetilde{Q}_{[X_2]}$ where $\widetilde{Q}_{[X_2]} = I - X_2(I + X_2'X_2)^{-1}X_2' = (I + X_2X_2')^{-1}$ and $S = I + X_1'\widetilde{Q}_{[X_2]}X_1$. This is lemma 2 of Davis (2001).
(d) Apply the results in part (c) using $X = (\frac{\sigma_\mu}{\sigma_\nu}\Delta_1, \frac{\sigma_\lambda}{\sigma_\nu}\Delta_2)$ to verify Σ^{-1} given in (9.33). *Hint*: Show that $V = \widetilde{Q}_{\Delta_1}$ and $S = \phi_2 P^*$.
(e) Derive $E(q_W)$, $E(q_N)$ and $E(q_T)$ given in (9.34), (9.35) and (9.36).

9.9 Using the Monte Carlo set-up for the unbalanced two-way error component model considered by Wansbeek and Kapteyn (1989), compare the MSE performance of the variance components and the regression coefficients estimates.

9.10 Assuming normality on the disturbances, verify (9.37), (9.40) and (9.41).

9.11 Verify that the King and Wu (1997) test for the unbalanced two-way error component model is given by (9.49).

9.12 Verify that the SLM version of the KW and HO tests are given by (9.47) with D defined in (9.50) and (9.51).

10
Special Topics

10.1 MEASUREMENT ERROR AND PANEL DATA

Micro panel data on households, individuals and firms are highly likely to exhibit measurement error. In Chapter 1, we cited Duncan and Hill (1985) who found serious measurement error in average hourly earnings in the Panel Study of Income Dynamics (PSID). This got worse for a two-year recall as compared to a one-year recall. Bound et al. (1990) use two validation data sets to study the extent of measurement error in labor market variables. The first data set is the Panel Study of Income Dynamics Validation Study (PSIDVS) which uses a two-wave panel survey taken in 1983 and 1987 from a single large manufacturing company. The second data set matches panel data on earnings from the 1977 and 1978 waves of the US Current Population Survey (CPS) to Social Security earnings records for those same individuals. They find that biases from measurement errors could be very serious for hourly wages and unemployment spells, but not severe for annual earnings.[1] In analyzing data from household budget surveys, total expenditure and income are known to contain measurement error. Aasness, Biorn and Skjerpen (1993) estimate a system of consumer expenditure functions from a Norwegian panel of households over the years 1975–77. The hypothesis of no measurement error in total expenditure is soundly rejected and substantial biases in Engle function elasticities are found when measurement error in total expenditure is ignored. Altonji and Siow (1987) find that measurement error in micro panel data sets has a strong influence on the relationship between consumption and income. Based on data from the 1968–81 PSID individuals tape, they show that ignoring the measurement error in the income process, a Keynesian model of consumption cannot be rejected. However, when one accounts for this measurement error, the results are supportive of the rational expectations lifecycle model of consumption and reject the Keynesian model.

Econometric textbooks emphasize that measurement error in the explanatory variables results in bias and inconsistency of the OLS estimates, and the solution typically involves the existence of *extraneous* instrumental variables or additional assumptions to identify the model parameters (see Maddala, 1977). Using panel data, Griliches and Hausman (1986) showed that one can identify and estimate a variety of errors in variables models *without* the use of external instruments. Let us illustrate their approach with a simple regression with random individual effects:

$$y_{it} = \alpha + \beta x_{it}^* + u_{it} \quad i = 1, \ldots, N; \; t = 1, \ldots, T \tag{10.1}$$

where the error follows a one-way error component model

$$u_{it} = \mu_i + \nu_{it} \tag{10.2}$$

and the x_{it}^* are observed only with error

$$x_{it} = x_{it}^* + \eta_{it} \tag{10.3}$$

In this case, $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$, $\nu_{it} \sim \text{IID}(0, \sigma_\nu^2)$ and $\eta_{it} \sim \text{IID}(0, \sigma_\eta^2)$ are all independent of each other. Additionally, x_{it}^* is independent of u_{it} and η_{it}. In terms of observable variables, the model becomes

$$y_{it} = \alpha + \beta x_{it} + \epsilon_{it} \tag{10.4}$$

where

$$\epsilon_{it} = \mu_i + \nu_{it} - \beta\eta_{it} \tag{10.5}$$

It is clear that OLS on (10.4) is inconsistent, since x_{it} is correlated with η_{it} and therefore ϵ_{it}. We follow Wansbeek and Koning (1991) by assuming that the variance–covariance matrix of x denoted by $\Sigma_x(T \times T)$ is the same across individuals, but otherwise of general form over time. In vector form, the model becomes

$$y = \alpha\iota_{NT} + x\beta + \epsilon \tag{10.6}$$

with

$$\begin{aligned} \epsilon &= (\iota_T \otimes \mu) + \nu - \beta\eta; \quad \mu' = (\mu_1, \ldots, \mu_N) \\ \nu &= (\nu_{11}, \ldots, \nu_{N1}, \ldots, \nu_{1T}, \ldots, \nu_{NT}) \end{aligned}$$

and

$$\eta' = (\eta_{11}, \ldots, \eta_{N1}, \ldots, \eta_{1T}, \ldots, \eta_{NT})$$

Note that the data are ordered such that the faster index is over individuals. Now consider any matrix P that wipes out the individual effects. P must satisfy $P\iota_T = 0$ and let $Q = P'P$ For example, $P = I_T - (\iota_T\iota_T'/T)$ is one such matrix, and the resulting estimator is the Within estimator. In general, for any Q, the estimator of β is given by

$$\begin{aligned} \widehat{\beta} &= x'(Q \otimes I_N)y/x'(Q \otimes I_N)x \\ &= \beta + x'(Q \otimes I_N)(\nu - \beta\eta)/x'(Q \otimes I_N)x \end{aligned} \tag{10.7}$$

For a fixed T, taking probability limits as the limit of expectations of the numerator and denominator as $N \to \infty$, we get

$$\begin{aligned} \frac{1}{N}E[x'(Q \otimes I_N)(\nu - \beta\eta)] &= -\frac{1}{N}\beta \operatorname{tr}[(Q \otimes I_N)E(\eta\eta')] = -\beta\sigma_\eta^2 \operatorname{tr}Q \\ \frac{1}{N}E[x'(Q \otimes I_N)x] &= \frac{1}{N} \operatorname{tr}[(Q \otimes I_N)(\Sigma_x \otimes I_N)] = \operatorname{tr}Q\Sigma_x \end{aligned}$$

and

$$\begin{aligned} \operatorname{plim}\widehat{\beta} &= \beta - \beta\sigma_\eta^2(\operatorname{tr}Q/\operatorname{tr}Q\Sigma_x) \\ &= \beta(1 - \sigma_\eta^2\phi) \end{aligned} \tag{10.8}$$

where $\phi \equiv (\operatorname{tr}Q/\operatorname{tr}Q\Sigma_x) > 0$. Griliches and Hausman (1986) used various Q transformations like the Within estimator and difference estimators to show that although these transformations wipe out the individual effect, they may aggravate the measurement error bias. Also, consistent estimators of β and σ_η^2 can be obtained by combining these inconsistent estimators. There are actually $\frac{1}{2}T(T-1) - 1$ linearly independent Q transformations. Let Q_1 and Q_2 be two choices for Q and $\phi_i = \operatorname{tr}(Q_i)/\operatorname{tr}(Q_i\Sigma_x)$ be the corresponding choices for ϕ, for $i = 1, 2$. Then

$\text{plim}\widehat{\beta}_i = \beta(1-\sigma_\eta^2\phi_i)$, and by replacing $\text{plim}\widehat{\beta}_i$ by $\widehat{\beta}_i$ itself, one can solve these two equations in two unknowns to get

$$\widehat{\beta} = \frac{\phi_1\widehat{\beta}_2 - \phi_2\widehat{\beta}_1}{\phi_1 - \phi_2} \tag{10.9}$$

and

$$\widehat{\sigma}_\eta^2 = \frac{\widehat{\beta}_2 - \widehat{\beta}_1}{\phi_1\widehat{\beta}_2 - \phi_2\widehat{\beta}_1} \tag{10.10}$$

In order to make these estimators operational, ϕ_i is replaced by $\widehat{\phi}_i$, where $\widehat{\phi}_i = \text{tr}(Q_i)/\text{tr}(Q_i\widehat{\Sigma}_x)$. Note that $P = I_T - (\iota_T\iota_T')/T$ yields the Within estimator, while $P = L'$, where L' is the $(T-1)\times T$ matrix defined in Chapter 8, yields the first-difference estimator. Other P matrices suggested by Griliches and Hausman (1986) are based on differencing the data j periods apart, $(y_{it} - y_{i,t-j})$, thus generating "different lengths" difference estimators. The remaining question is how to combine these consistent estimators of β into an efficient estimator of β. The generalized method of moments (GMM) approach can be used and this is based upon fourth-order moments of the data. Alternatively, under normality one can derive the asymptotic covariance matrix of the $\widehat{\beta}_i$ which can be consistently estimated by second-order moments of the data. Using the latter approach, Wansbeek and Koning (1991) showed that for m different consistent estimators of β given by $b = (\widehat{\beta}_1, \ldots, \widehat{\beta}_m)'$ based on m different Q_i

$$\sqrt{N}[b - \beta(\iota_m - \sigma_\eta^2\phi)] \sim N(0, V)$$

where

$$\begin{aligned} \phi &= (\phi_1, \ldots, \phi_m)' \\ V &= F'\{\sigma_v^2\Sigma_x \otimes I_T + \beta^2\sigma_\eta^2(\Sigma_x + \sigma_\eta^2 I_N)\otimes I_T\}F \end{aligned} \tag{10.11}$$

and F is the $(T^2 \times m)$ matrix with ith column $f_i = \text{vec}Q_i/(\text{tr}Q_i\Sigma_x)$. By minimizing $[b - \beta(\iota_m - \sigma_\eta^2\phi)]'V^{-1}[b - \beta(\iota_m - \sigma_\eta^2\phi)]$ one gets the asymptotically efficient estimators (as far as they are based on b) of β and σ_v^2 given by

$$\widehat{\beta} = \left\{\frac{\phi'\widehat{V}^{-1}b}{\phi'\widehat{V}^{-1}\phi} - \frac{\iota'\widehat{V}^{-1}b}{\iota'\widehat{V}^{-1}\phi}\right\}\bigg/\left\{\frac{\phi'\widehat{V}^{-1}\iota}{\phi'\widehat{V}^{-1}\phi} - \frac{\iota'\widehat{V}^{-1}\iota}{\iota'\widehat{V}^{-1}\phi}\right\} \tag{10.12}$$

and

$$\widehat{\sigma}_v^2 = \left\{\frac{\phi'\widehat{V}^{-1}\iota}{\phi'\widehat{V}^{-1}b} - \frac{\iota'\widehat{V}^{-1}\iota}{\iota'\widehat{V}^{-1}b}\right\}\bigg/\left\{\frac{\phi'\widehat{V}^{-1}\phi}{\phi'\widehat{V}^{-1}b} - \frac{\iota'\widehat{V}^{-1}\phi}{\iota'\widehat{V}^{-1}b}\right\} \tag{10.13}$$

with $\sqrt{N}(\widehat{\beta} - \beta, \widehat{\sigma}_v^2 - \sigma_v^2)$ asymptotically distributed as $N(0, W)$ and

$$W = \frac{1}{\Delta}\begin{bmatrix} \beta^2\phi'V^{-1}\phi & \beta(\iota_m - \sigma_\eta^2\phi)'V^{-1}\phi \\ & (\iota_m - \sigma_\eta^2\phi)'V^{-1}(\iota_m - \sigma_\eta^2\phi) \end{bmatrix} \tag{10.14}$$

where

$$\Delta = \beta^2(\iota_m - \sigma_\eta^2\phi)'V^{-1}(\iota_m - \sigma_\eta^2\phi)(\phi'V^{-1}\phi) - \beta^2[\phi'V^{-1}(\iota_m - \sigma_\eta^2\phi)]^2 \tag{10.15}$$

Griliches and Hausman (1986) argue that their results can be extended to the case of several independent variables provided that the measurement errors in the explanatory variables are mutually uncorrelated, or correlated with a known correlation structure. Under some stringent assumptions these results can be extended to the case of serially correlated η_{it}. Griliches and Hausman (1986) illustrate their approach by estimating a labor demand relationship using data on $N = 1242$ US manufacturing firms over six years (1972–77) drawn from the National Bureau of Economic Research R&D panel. For some applications of measurement error in panel data, see Hamermesh (1989) for the case of academic salaries in the USA and Björklund (1989) for the case of job mobility in Sweden and Abowd and Card (1989) on the covariance structure of earnings and hours changes. Extensions of this model to the case where the measurement error itself follows an error component structure are given by Biorn (1992). Biorn (1996) also gives an extensive treatment for the case where the model disturbances u_{it} in equation (10.1) are white noise, i.e. without any error component, and the case where η_{it}, the measurement error, is autocorrelated over time. For all cases considered, Biorn derives the asymptotic bias of the Within, Between, various difference estimators and the GLS estimator as either N or T tend to ∞. Biorn shows how the different panel data transformations implied by these estimators affect measurement error differently. Biorn and Klette (1998) consider GMM estimation of a simple static panel data regression with errors in variables, as described in (10.4). Assuming μ_i to be fixed effects and the measurement error to be not autocorrelated, Biorn and Klette show that only the one-period and a few two-period differences are essential, i.e., relevant for GMM estimation. The total number of orthogonality conditions is $T(T-1)(T-2)/2$ while the essential set of orthogonality conditions is only a fraction $2/(T-1)$ of the complete set, i.e., $T(T-2)$. Among these essential conditions, $(T-1)(T-2)$ are based on one-period differences and $(T-2)$ on two-period differences. Exploiting only the nonredundant moment conditions reduces the computational burden considerably. For a moderate size panel with $T = 9$, the essential moment conditions are one-fourth of the complete set of orthogonality conditions and involve inverting a 63×63 matrix rather than a 252×252 matrix to compute GMM. Biorn (2000) proposes GMM estimators that use either (A) equations in differences with level values as instruments or (B) equations in levels with differenced values as instruments. The conditions needed for the consistency of the (B) procedures under individual heterogeneity are stronger than for the (A) procedures. These procedures are illustrated for a simple regression of log of gross production on log of material input for the manufacture of textiles. The data uses $N = 215$ firms observed over $T = 8$ years 1983–90 and obtained from the annual Norwegian manufacturing census. For this empirical illustration, Biorn shows that adding the essential two-period difference orthogonality conditions to the one-period conditions in the GMM algorithm may significantly increase estimation efficiency. However, redundant orthogonality conditions are of little practical use. Overall, the GMM estimates based on the level equations are more precise than those based on differenced equations. Recently, Wansbeek (2001) presented a simple approach to derive moment conditions for the panel data model with a single mismeasured variable under a variety of assumptions. For other extensions, see Kao and Schnell (1987a, b) for the fixed effects logit model and the random effects probit model using panel data with measurement error, and Hsiao (1991) for identification conditions of binary choice errors in variables models as well as conditions for consistency and asymptotic normality of the maximum likelihood estimators when the explanatory variables are unbounded.

10.2 ROTATING PANELS

Biorn (1981) considers the case of rotating panels, where in order to keep the *same* number of households in the survey, the fraction of households that drops from the sample in the second period is replaced by an equal number of new households that are freshly surveyed. This is a necessity in survey panels where the same household may not want to be interviewed again and again. In the study by Biorn and Jansen (1983) based on data from the Norwegian household budget surveys, half the sample is rotated in each period. In other words, half the households surveyed drop from the sample each period and are replaced by new households.[2] To illustrate the basics of rotating panels, let us assume that $T = 2$ and that half the sample is being rotated each period. In this case, without loss of generality, households $1, 2, \ldots, (N/2)$ are replaced by households $N + 1$, $N + 2, \ldots, N + (N/2)$ in period 2. It is clear that only households $(N/2) + 1$, $(N/2) + 2, \ldots, N$ are observed over two periods.[3] In this case there are $3N/2$ distinct households, only $N/2$ households of which are observed for two periods. In our case, the first and last $N/2$ households surveyed are only observed for one period. Now consider the usual one-way error component model

$$u_{it} = \mu_i + \nu_{it}$$

with $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$ and $\nu_{it} \sim \text{IID}(0, \sigma_\nu^2)$ independent of each other and the x_{it}. Order the observations such that the faster index is that of households and the slower index is that of time. This is different from the ordering we used in Chapter 2. In this case, $u' = (u_{11}, u_{21}, \ldots, u_{N1}, u_{N/2+1,2}, \ldots, u_{3N/2,2})$ and

$$E(uu') = \Omega = \begin{bmatrix} \sigma^2 I_{N/2} & 0 & 0 & 0 \\ 0 & \sigma^2 I_{N/2} & \sigma_\mu^2 I_{N/2} & 0 \\ 0 & \sigma_\mu^2 I_{N/2} & \sigma^2 I_{N/2} & 0 \\ 0 & 0 & 0 & \sigma^2 I_{N/2} \end{bmatrix} \tag{10.16}$$

where $\sigma^2 = \sigma_\mu^2 + \sigma_\nu^2$. It is easy to see that Ω is block-diagonal and that the middle block has the usual error component model form $\sigma_\mu^2(J_2 \otimes I_{N/2}) + \sigma_\nu^2(I_2 \otimes I_{N/2})$. Therefore

$$\Omega^{-1/2} = \begin{bmatrix} \frac{1}{\sigma} I_{N/2} & 0 & 0 \\ 0 & \left(\frac{1}{\sigma_1^*}\bar{J}_2 + \frac{1}{\sigma_\nu}E_2\right) \otimes I_{N/2} & 0 \\ 0 & 0 & \frac{1}{\sigma} I_{N/2} \end{bmatrix} \tag{10.17}$$

where $E_2 = I_2 - \bar{J}_2$, $\bar{J}_2 = J_2/2$ and $\sigma_1^{*2} = 2\sigma_\mu^2 + \sigma_\nu^2$. By premultiplying the regression model by $\Omega^{-1/2}$ and performing OLS one gets the GLS estimator of the rotating panel. In this case, one divides the first and last $N/2$ observations by σ. For the middle N observations, with $i = (N/2) + 1, \ldots, N$ and $t = 1, 2$, quasi-demeaning similar to the usual error component transformation is performed, i.e. $(y_{it} - \theta^* \bar{y}_{i.})/\sigma_\nu$ with $\theta^* = 1 - (\sigma_\nu/\sigma_1^*)$ and $\bar{y}_{i.} = (y_{i1} + y_{i2})/2$. A similar transformation is also performed on the regressors. In order to make this GLS estimator feasible, we need estimates of σ_μ^2 and σ_ν^2. One consistent estimator of σ_ν^2 can be obtained from the middle N observations or simply the households that are observed over two periods.

For these observations, σ_ν^2 is estimated consistently from the Within residuals

$$\widetilde{\sigma}_\nu^2 = \sum_{t=1}^{2} \sum_{i=N/2+1}^{N} [(y_{it} - \bar{y}_{i.}) - (x_{it} - \bar{x}_{i.})'\widetilde{\beta}_{\text{Within}}]^2/N \tag{10.18}$$

whereas the total variance can be estimated consistently from the least squares mean square error over the entire sample

$$\widetilde{\sigma}^2 = \widetilde{\sigma}_\nu^2 + \widetilde{\sigma}_\mu^2 = \sum_{t=1}^{2} \sum_{i=1}^{3N/2} (y_{it} - x_{it}'\widehat{\beta}_{\text{OLS}})^2/(3N/2) \tag{10.19}$$

Note that we could have reordered the data such that households observed over one period are stacked on top of households observed over two time periods. This way the rotating panel problem becomes an unbalanced panel problem with N households observed over one period and $N/2$ households observed for two periods. In fact, except for this different way of ordering the observations, one can handle the estimation as in Chapter 9.

This feasible GLS estimation can easily be derived for other rotating schemes. In fact, the reader is asked to do that for $T = 3$ with $N/2$ households rotated every period, and $T = 3$ with $N/3$ households rotated every period (see problem 10.2). For the estimation of more general rotation schemes as well as maximum likelihood estimation under normality, see Biorn (1981). The analysis of rotating panels can also easily be extended to a set of seemingly unrelated regressions, simultaneous equations or a dynamic model. Biorn and Jansen (1983) consider a rotating panel of 418 Norwegian households, one half of which are observed in 1975 and 1976 and the other half in 1976 and 1977. They estimate a complete system of consumer demand functions using maximum likelihood procedures.

Rotating panels allow the researcher to test for the existence of "time-in-sample" bias effects mentioned in Chapter 1. These correspond to a significant change in response between the initial interview and a subsequent interview when one would expect the same response.[4] With rotating panels, the fresh group of individuals that are added to the panel with each wave provide a means of testing for time-in-sample bias effects. Provided that all other survey conditions remain constant for all rotation groups at a particular wave, one can compare these various rotation groups (for that wave) to measure the extent of rotation group bias. This has been done for various labor force characteristics in the Current Population Survey. For example, several studies have found that the first rotation reported an unemployment rate that is 10% higher than that of the full sample (see Bailar, 1975). While the findings indicate a pervasive effect of rotation group bias in panel surveys, the survey conditions do not remain the same in practice and hence it is hard to disentangle the effects of time-in-sample bias from other effects.

10.3 PSEUDO-PANELS

For some countries, panel data may not exist. Instead the researcher may find annual household surveys based on a large random sample of the population. Examples of some of these cross-sectional consumer expenditure surveys include: the UK Family Expenditure Survey which surveys about 7000 households annually, and also a number of household surveys from less developed countries like the World Bank's poverty net inventory of household surveys. This is available at http://www.world.bank.org/poverty/data/index.htm. Examples of repeated cross-section surveys in the USA include the Current Population Survey, the National Health

Interview Survey, the Consumer Expenditure Survey, the National Crime Survey, the Monthly Retail Trade Survey and the Survey of Manufacturers' Shipments, Inventories and Orders. See Bailar (1989) for the corresponding data sources. Also, the adult education and life-long learning surveys and the early childhood program participation surveys available from the National Center for Education Statistics at http://nces.ed.gov/surveys/, the general social survey available from the National Opinion Research Center at http://www.norc.uchicago.edu/gss/homepage.htm and the survey of small business finances from the Federal Reserve Board at http://www.federalreserve.gov/ssbf/, to mention a few. For these repeated cross-section surveys, it may be impossible to track the same household over time as required in a genuine panel. Instead, Deaton (1985) suggests tracking cohorts and estimating economic relationships based on cohort means rather than individual observations. One cohort could be the set of all males born between 1945 and 1950. This birth cohort is well-defined, and can easily be identified from the data. Deaton (1985) argued that these pseudo-panels do not suffer the attrition problem that plagues genuine panels, and may be available over longer time periods compared to genuine panels.[5] In order to illustrate the basic ideas involved in constructing a pseudo-panel, we start with the set of T independent cross-sections given by

$$y_{it} = x'_{it}\beta + \mu_i + \nu_{it} \quad t = 1, \ldots, T \tag{10.20}$$

Note that the individual subscript i corresponds to a new and most likely different set of individuals in each period. This is why it is denoted by $i(t)$ to denote that each period different individuals are sampled, making these individuals time-dependent. For ease of exposition, we continue the use of the subscript i and assume that the same number of households N is randomly surveyed each period. Define a set of C cohorts, each with a fixed membership that remains the same throughout the entire period of observation. Each individual observed in the survey belongs to exactly one cohort. Averaging the observations over individuals in each cohort, one gets

$$\bar{y}_{ct} = \bar{x}'_{ct}\beta + \bar{\mu}_{ct} + \bar{\nu}_{ct} \quad c = 1, \ldots, C;\ t = 1, \ldots, T \tag{10.21}$$

where $\bar{y}_{ct}$ is the average of y_{it} over all individuals belonging to cohort c at time t. Since the economic relationship for the individual includes an individual fixed effect, the corresponding relationship for the cohort will also include a fixed cohort effect. However, $\bar{\mu}_{ct}$ now varies with t, because it is averaged over a different number of individuals belonging to cohort c at time t. These $\bar{\mu}_{ct}$ are most likely correlated with the x_{it} and a random effect specification will lead to inconsistent estimates. On the other hand, treating the $\bar{\mu}_{ct}$ as fixed effects leads to an identification problem, unless $\bar{\mu}_{ct} = \bar{\mu}_c$ and is invariant over time. The latter assumption is plausible if the number of observations in each cohort is very large. In this case,

$$\bar{y}_{ct} = \bar{x}'_{ct}\beta + \bar{\mu}_c + \bar{\nu}_{ct} \quad c = 1, \ldots, C;\ t = 1, \ldots, T \tag{10.22}$$

For this pseudo-panel with T observations on C cohorts, the fixed effects estimator $\widetilde{\beta}_W$, based on the Within cohort transformation $\widetilde{y}_{ct} = \bar{y}_{ct} - \bar{y}_c$, is a natural candidate for estimating β. Note that the cohort population means are genuine panels in that, at the population level, the groups contain the same individuals over time. However, as Deaton (1985) argued, the sample-based averages of the cohort means, $\bar{y}_{ct}$, can only estimate the unobserved population cohort means with measurement error. Therefore, one has to correct the Within estimator for measurement error using estimates of the errors in the measurement variance–covariance matrix obtained from the individual survey data. Details are given in Deaton (1985), Verbeek

(1996) and Verbeek and Nijman (1993). Deaton (1985) shows that his proposed measurement error-corrected within groups estimator for the static model with individual effects is consistent for a fixed number of observations per cohort. Verbeek and Nijman (1993) modify Deaton's estimator to achieve consistency for a fixed number of time periods and a fixed number of individuals per cohort. If the number of individuals in each cohort is large, so that the average cohort size $n_c = N/C$ tends to infinity, then the measurement errors as well as their estimates tend to zero and the Within cohort estimator of β is asymptotically identical to Deaton's (1985) estimator of β, denoted by $\widetilde{\beta}_D$. In fact, when n_c is large, most applied researchers ignore the measurement error problem and compute the Within cohort estimator of β (see Browning, Deaton and Irish, 1985).

There is an obvious tradeoff in the construction of a pseudo-panel. The larger the number of cohorts, the smaller is the number of individuals per cohort. In this case, C is large and the pseudo-panel is based on a large number of observations. However, the fact that n_c is not large implies that the sample cohort averages are not precise estimates of the population cohort means. In this case, we have a large number C of imprecise observations. In contrast, a pseudo-panel constructed with a smaller number of cohorts (C) and therefore more individuals per cohort (n_c) is trading a large pseudo-panel with imprecise observations for a smaller pseudo-panel with more precise observations. Verbeek and Nijman (1992b) study the consistency properties of the above two estimators as the number of cohorts C, the number of individuals per cohort n_c, and N and T are fixed or tend to infinity. They find that $n_c \to \infty$ is a crucial condition for the consistency of the Within estimator. On the other hand, Deaton's estimator is consistent for β, for finite n_c when either C or T tend to infinity. Verbeek and Nijman (1992b) also find that the bias in the Within estimator may be substantial even for large n_c. They also emphasize the importance of choosing the cohorts under study very carefully. For example, in order to minimize the measurement error variance, the individuals in each cohort should be as homogeneous as possible. Additionally, to maximize the variation in the pseudo-panel, and get precise estimates, the different cohorts should be as heterogeneous as possible.

Moffitt (1993) extends Deaton's (1985) analysis to the estimation of dynamic models with repeated cross-sections. By imposing certain restrictions, Moffitt shows that linear and nonlinear models, with and without fixed effects, can be identified and consistently estimated with pseudo-panels. Moffitt (1993) gives an instrumental variable interpretation for the Within estimator based on the pseudo-panel using cohort dummies, and a set of time dummies interacted with the cohort dummies. Because n_c is assumed to tend to ∞, the measurement error problem is ignored. Since different individuals are sampled in each period, the lagged dependent variable is not observed. Moffitt suggests replacing the unknown $y_{i,t-1}$ by a fitted value obtained from observed data at time $t-1$. Moffitt (1993) illustrates his estimation method for the linear fixed effects lifecycle model of labor supply using repeated cross-sections from the US Current Population Survey (CPS). The sample included white males, ages 20–59, drawn from 21 waves over the period 1968–88. In order to keep the estimation problem manageable, the data were randomly subsampled to include a total of 15 500 observations. Moffitt concludes that there is a considerable amount of parsimony achieved in the specification of age and cohort effects. Also, individual characteristics are considerably more important than either age, cohort or year effects. Blundell, Meghir and Neves (1990) use the annual UK Family Expenditure Survey covering the period 1970–84 to study the intertemporal labor supply and consumption of married women. The total number of households considered was 43 671. These were allocated to ten different cohorts depending on the year of birth. The average number of observations per cohort was 364. Their findings indicate reasonably sized intertemporal labor supply elasticities.

Collado (1997) proposes measurement error-corrected estimators for dynamic models with individual effects using time series of independent cross-sections. A GMM estimator corrected for measurement error is proposed that is consistent as the number of cohorts tends to infinity for a fixed T and a fixed number of individuals per cohort. In addition, a measurement error-corrected within groups estimator is proposed which is consistent as T tends to infinity. Monte Carlo simulations are performed to study the small sample properties of the estimators proposed. Some of the main results indicate that the measurement error correction is important, and that corrected estimators reduce the bias obtained. Also, for small T, GMM estimators are better than within groups estimators.

Verbeek and Vella (2004) review the identification conditions for consistent estimation of a linear dynamic model from repeated cross-sections. They show that Moffitt's (1993) estimator is inconsistent, unless the exogenous variables are either time-invariant or exhibit no autocorrelation. They propose an alternative instrumental variable estimator, corresponding to the Within estimator applied to the pseudo-panel of cohort averages. This estimator is consistent under the same conditions as those suggested by Collado (1997). However, Verbeek and Vella argue that those conditions are not trivially satisfied in applied work.

Girma (2000) suggests an alternative GMM method of estimating linear dynamic models from a time series of independent cross-sections. Unlike the Deaton (1985) approach of averaging across individuals in a cohort, Girma suggests a quasi-differencing transformation across pairs of individuals that belong to the same group. The asymptotic properties of the proposed GMM estimators are based upon having a large number of individuals per group/time cell. This is in contrast to the Deaton-type estimator which requires the number of group/time periods to grow without limit. Some of the other advantages of this method include the fact that no aggregation is involved, the dynamic response parameters can freely vary across groups, and the presence of unobserved individual specific heterogeneity is explicitly allowed for.

McKenzie (2001) considers the problem of estimating dynamic models with unequally spaced pseudo-panel data. Surveys in developing countries are often taken at unequally spaced intervals and this unequal spacing, in turn, imposes nonlinear restrictions on the parameters.[6] Nonlinear least squares, minimum distance and one-step estimators are suggested that are consistent and asymptotically normal for finite T as the number of individuals per cohort is allowed to pass to infinity. In another paper, McKenzie (2004) allows for parameter heterogeneity amongst cohorts, and argues that in many practical applications, it is important to investigate whether there are systematic differences between cohorts. McKenzie (2004) develops an asymptotic theory for pseudo-panels using sequential and diagonal path limit techniques following the work of Phillips and Moon (1999) for nonstationary panels. McKenzie uses 20 years of household survey data (1976–96) from the Taiwanese personal income distribution survey, to quantify the degree of intercohort parameter heterogeneity. He finds that younger cohorts experienced faster consumption growth over the sample period than older cohorts.

10.4 ALTERNATIVE METHODS OF POOLING TIME SERIES OF CROSS-SECTION DATA

This book has focused on the error component model as a popular method in economics for pooling time series of cross-section data. Another alternative method for pooling these data is described in Kmenta (1986) using timewise autocorrelated and cross-sectionally heteroskedastic disturbances. The basic idea is to allow for first-order autoregressive disturbances

$$u_{it} = \rho_i u_{i,t-1} + \epsilon_{it} \quad i = 1, \ldots, N;\, t = 1, \ldots, T \tag{10.23}$$

where the autoregressive parameter can vary across cross-sections with $|\rho_i| < 1$. Also, the remainder error ϵ_{it} is assumed to be normal with zero mean and a general variance–covariance matrix that allows for possible heteroskedasticity as well as correlation across cross-sections, i.e.

$$E(\epsilon\epsilon') = \Sigma \otimes I_T \quad \text{where} \quad \epsilon' = (\epsilon_{11}, \ldots, \epsilon_{1T}, \ldots, \epsilon_{N1}, \ldots, \epsilon_{NT}) \tag{10.24}$$

and Σ is $N \times N$. The initial values are assumed to have the following properties:

$$u_{i0} \sim N\left(0, \frac{\sigma_{ii}}{1-\rho_i^2}\right) \quad \text{and} \quad E(u_{i0}u_{j0}) = \frac{\sigma_{ij}}{1-\rho_i\rho_j} \qquad i, j = 1, 2, \ldots, N$$

Kmenta (1986) describes how to obtain feasible GLS estimators of the regression coefficients.[7] In the first step, OLS residuals are used to get consistent estimates of the ρ_i. Next, a Prais–Winsten transformation is applied using the estimated $\widehat{\rho}_i$ to get a consistent estimate of Σ from the resulting residuals. In the last step, GLS is applied to the Prais–Winsten transformed model using the consistent estimate of Σ. This can be done using the xtgls command in Stata. This may be a suitable pooling method for N small and T very large, but for typical labor or consumer panels where N is large and T is small it may be infeasible. In fact, for $N > T$, the estimate of Σ will be singular. Note that the number of extra parameters to be estimated for this model is $N(N+1)/2$, corresponding to the elements of Σ plus N distinct ρ_i. This is in contrast to the simple one-way error component model with N extra parameters to estimate for the fixed effects model or two extra variance components to estimate for the random effects model. For example, even for a small $N = 50$, the number of extra parameters to estimate for the Kmenta technique is 1325 compared to 50 for fixed effects and two for the random effects model. Baltagi (1986) discusses the advantages and disadvantages of the Kmenta and the error components methods and compares their performance using Monte Carlo experiments. For typical panels with N large and T small, the error component model is parsimonious in its estimation of variance–covariance parameters compared to the timewise autocorrelated, cross-sectionally heteroskedastic specification and is found to be more robust to misspecification.

Some economic applications of the Kmenta method include: (1) van der Gaag et al. (1977) who applied the timewise autocorrelated cross-sectionally heteroskedastic technique to estimate a dynamic model of demand for specialist medical care in the Netherlands. The panel data covered 11 provinces ($N = 11$) in the Netherlands collected over the period 1960–72 ($T = 13$). The disturbances were allowed to be timewise autocorrelated with differing ρ_i cross-sectionally heteroskedastic and correlated across regions. (2) Wolpin (1980) used annual observations on robberies covering the period 1955–71 for three countries: Japan, England and the USA (represented by California). Per capita robbery rate in country i at time t was modeled as a loglinear function of a set of deterrence variables, a set of environment variables and a time-invariant "culture" variable which is related to the propensity to commit robbery. Country-specific dummy variables were used to capture these cultural effects. The remainder error was assumed cross-sectionally heteroskedastic and timewise autocorrelated with a different ρ_i for each country. In addition, the disturbances were allowed to be correlated across countries. (3) Griffin (1982) applied the Kmenta technique to estimate the demand for a long-distance telephone service. The panel data consisted of seasonally adjusted quarterly data for the period 1964–78, for five southwestern states. Per capita intrastate long-distance messages were modeled as a distributed lag of real per capita income, advertising exposure and the real price of message telecommunication service. Additionally, population and population squared were used to measure the effect of market size, since the quantity of long-distance service depends on the number of possible calling combinations. A dummy variable was included for each state with the remainder error assumed to be first-order autocorrelated with a different ρ for each state, heteroskedastic, and not correlated across states. Griffin (1982) found a long-run price elasticity of -0.6 and a statistically significant effect of advertising. This price elasticity, coupled with marginal costs between one-fourth and one-half price, indicated a large welfare loss in the long-distance market.

In this context, Baltagi, Chang and Li (1992a) considered the case where all cross-sections in (10.23) have the same autoregressive parameter ρ, with $|\rho| < 1$ and $\epsilon_{it} \sim \text{IIN}(0, \sigma_\epsilon^2)$, but where the initial disturbance $u_{i1} \sim \text{IIN}(0, \sigma_\epsilon^2/\tau)$ where τ is an arbitrary positive number. They show that the resulting disturbances are heteroskedastic unless $\tau = (1 - \rho^2)$ or the process started a long time ago. With panel data, no matter when the process started, one can translate this starting date into an "effective" initial variance assumption. This initial variance can be estimated and tested for departures from homoskedasticity. Using Monte Carlo experiments, the authors show that for short time series ($T = 10, 20$), if $\tau \neq (1 - \rho^2)$, the conventional MLE which assumes $\tau = 1 - \rho^2$ performs poorly relative to the MLE that estimates the arbitrary τ.

Finally, Larson and Watters (1993) suggested a joint test of functional form and non-spherical disturbances for the Kmenta model with fixed effects using the Box–Cox transformation and an artificial linear regression approach. They apply this test to a model of intrastate long-distance demand for Southwestern Bell's five-state region observed quarterly over the period 1979–88. Their results reject the logarithmic transformation on both the dependent and independent variables and is in favor of correcting for serial correlation and heteroskedasticity.

10.5 SPATIAL PANELS

In randomly drawn samples at the individual level, one does not usually worry about cross-section correlation. However, when one starts looking at a cross-section of countries, regions, states, counties, etc., these aggregate units are likely to exhibit cross-sectional correlation that has to be dealt with. There is an extensive literature using spatial statistics that deals with this type of correlation. These spatial dependence models are popular in regional science and urban economics. More specifically, these models deal with spatial interaction (spatial autocorrelation) and spatial structure (spatial heterogeneity) primarily in cross-section data, see Anselin (1988, 2001) for a nice introduction to this literature. Spatial dependence models may use a metric of economic distance which provides cross-sectional data with a structure similar to that provided by the time index in time series. With the increasing availability of micro as well as macro level panel data, spatial panel data models are becoming increasingly attractive in empirical economic research. See Case (1991), Holtz-Eakin (1994), Driscoll and Kraay (1998), Bell and Bockstael (2000) and Baltagi and Li (2004) for a few applications. For example, in explaining per capita R&D expenditures and spillover effects across countries, one can model the spatial correlation as well as the heterogeneity across countries using a spatial error component regression model:

$$y_{ti} = X'_{ti}\beta + u_{ti} \quad i = 1, \dots, N;\ t = 1, \dots, T \tag{10.25}$$

where y_{ti} is the observation on the ith country for the tth time period, X_{ti} denotes the $k \times 1$ vector of observations on the nonstochastic regressors and u_{ti} is the regression disturbance. In vector form, the disturbance vector of (10.25) is assumed to have random country effects as well as spatially autocorrelated remainder disturbances, see Anselin (1988):

$$u_t = \mu + \epsilon_t \tag{10.26}$$

with

$$\epsilon_t = \lambda W \epsilon_t + \nu_t \tag{10.27}$$

where $\mu' = (\mu_1, \dots, \mu_N)$ denotes the vector of random country effects which is assumed to be IIN(0, σ^2_μ). λ is the scalar spatial autoregressive coefficient with $|\lambda| < 1$. W is a known $N \times N$ spatial weight matrix whose diagonal elements are zero. W also satisfies the condition that $(I_N - \lambda W)$ is nonsingular. $\nu'_t = (\nu_{t1}, \dots, \nu_{tN})$, where ν_{ti} is assumed to be IIN(0, σ^2_ν) and also independent of μ_i. One can rewrite (10.27) as

$$\epsilon_t = (I_N - \lambda W)^{-1}\nu_t = B^{-1}\nu_t \tag{10.28}$$

where $B = I_N - \lambda W$ and I_N is an identity matrix of dimension N. The model (10.25) can be rewritten in matrix notation as

$$y = X\beta + u \tag{10.29}$$

where y is now of dimension $NT \times 1$, X is $NT \times k$, β is $k \times 1$ and u is $NT \times 1$. X is assumed to be of full column rank and its elements are assumed to be bounded in absolute value. Equation (10.26) can be written in vector form as

$$u = (\iota_T \otimes I_N)\mu + (I_T \otimes B^{-1})\nu \tag{10.30}$$

where $\nu' = (\nu'_1, \dots, \nu'_T)$. Under these assumptions, the variance–covariance matrix for u is given by

$$\Omega = \sigma^2_\mu (J_T \otimes I_N) + \sigma^2_\nu (I_T \otimes (B'B)^{-1}) \tag{10.31}$$

This matrix can be rewritten as

$$\Omega = \sigma^2_\nu \Big[\bar{J}_T \otimes (T\phi I_N + (B'B)^{-1}) + E_T \otimes (B'B)^{-1}\Big] = \sigma^2_\nu \Sigma \tag{10.32}$$

where $\phi = \sigma^2_\mu/\sigma^2_\nu$, $\bar{J}_T = J_T/T$ and $E_T = I_T - \bar{J}_T$. Using results in Wansbeek and Kapteyn (1983), Σ^{-1} is given by

$$\Sigma^{-1} = \bar{J}_T \otimes (T\phi I_N + (B'B)^{-1})^{-1} + E_T \otimes B'B \tag{10.33}$$

Also, $|\Sigma| = |T\phi I_N + (B'B)^{-1}| \cdot |(B'B)^{-1}|^{T-1}$. Under the assumption of normality, the loglikelihood function for this model was derived by Anselin (1988, p. 154) as

$$
\begin{aligned}
L &= -\frac{NT}{2}\ln 2\pi\sigma_\nu^2 - \frac{1}{2}\ln|\Sigma| - \frac{1}{2\sigma_\nu^2}u'\Sigma^{-1}u \\
&= -\frac{NT}{2}\ln 2\pi\sigma_\nu^2 - \frac{1}{2}\ln[|T\phi I_N + (B'B)^{-1}|] + \frac{(T-1)}{2}\ln|B'B| \\
&\quad -\frac{1}{2\sigma_\nu^2}u'\Sigma^{-1}u
\end{aligned}
\tag{10.34}
$$

with $u = y - X\beta$. For a derivation of the first-order conditions of MLE as well as the LM test for $\lambda = 0$ for this model, see Anselin (1988). As an extension to this work, Baltagi, Song and Koh (2003) derived the joint LM test for spatial error correlation as well as random country effects. Additionally, they derived conditional LM tests, which test for random country effects given the presence of spatial error correlation. Also, spatial error correlation given the presence of random country effects. These conditional LM tests are an alternative to the one-directional LM tests that test for random country effects ignoring the presence of spatial error correlation or the one-directional LM tests for spatial error correlation ignoring the presence of random country effects. Extensive Monte Carlo experiments are conducted to study the performance of these LM tests as well as the corresponding likelihood ratio tests.

More recently, generalized method of moments has been proposed for spatial cross-section models by Conley (1999) and Kelejian and Prucha (1999) and an application of the latter method to housing data is given in Bell and Bockstael (2000). Frees (1995) derives a distribution-free test for spatial correlation in panels. This is based on Spearman rank correlation across pairs of cross-section disturbances. Driscoll and Kraay (1998) show through Monte Carlo simulations that the presence of even modest spatial dependence can impart large bias to OLS standard errors when N is large. They present conditions under which a simple modification of the standard nonparametric time series covariance matrix estimator yields estimates of the standard errors that are robust to general forms of spatial and temporal dependence as $T \to \infty$. However, if T is small, they conclude that the problem of consistent nonparametric covariance matrix estimation is much less tractable. Parametric corrections for spatial correlation are possible only if one places strong restrictions on their form, i.e., knowing W. For typical micropanels with N much larger than T, estimating this correlation is impossible without imposing restrictions, since the number of spatial correlations increases at the rate N^2, while the number of observations grows at rate N. Even for macro panels where $N = 100$ countries observed over $T = 20$ to 30 years, N is still larger than T and prior restrictions on the form of spatial correlation are still needed.

Baltagi and Li (2004) derive the best linear unbiased predictor for the random error component model with spatial correlation using a simple demand equation for cigarettes based on a panel of 46 states over the period 1963–92. They compare the performance of several predictors of the states demand for cigarettes for one year and five years ahead. The estimators whose predictions are compared include OLS, fixed effects ignoring spatial correlation, fixed effects with spatial correlation, random effects GLS estimator ignoring spatial correlation and random effects estimator accounting for the spatial correlation. Based on the RMSE criteria, the fixed effects and the random effects spatial estimators gave the best out-of-sample forecast performance. For the estimation and testing of spatial autoregressive panel models as well as an extensive set of references on spatial studies, read Anselin (1988, 2001).

ML estimation, even in its simplest form, entails substantial computational problems when the number of cross-sectional units N is large. Kelejian and Prucha (1999) suggested a generalized moments (GM) estimation method which is computationally feasible even when N is large. Kapoor, Kelejian and Prucha (2004) generalized this GM procedure from cross-section to panel data and derived its large sample properties when T is fixed and $N \to \infty$.

The basic regression model is the same as in (10.29), however the disturbance term u follows the first-order spatial autoregressive process

$$u = \lambda(I_T \otimes W)u + \epsilon \tag{10.35}$$

with

$$\epsilon = (\iota_T \otimes I_N)\mu + \nu \tag{10.36}$$

where μ, ν and W were defined earlier. This is different from the specification described in (10.26) and (10.27) since it also allows the individual country effects μ to be spatially correlated.

Defining $\bar{u} = (I_T \otimes W)u$, $\bar{\bar{u}} = (I_T \otimes W)\bar{u}$ and $\bar{\epsilon} = (I_T \otimes W)\epsilon$, Kapoor et al. (2004) suggest a GM estimator based on the following six moment conditions:

$$\begin{aligned}
&E[\epsilon' Q\epsilon/N(T-1)] = \sigma_\nu^2\\
&E[\bar{\epsilon}' Q\bar{\epsilon}/N(T-1)] = \sigma_\nu^2 \operatorname{tr}(W'W)/N\\
&E[\bar{\epsilon}' Q\epsilon/N(T-1)] = 0\\
&E(\epsilon' P\epsilon/N) = T\sigma_\mu^2 + \sigma_\nu^2 = \sigma_1^2\\
&E(\bar{\epsilon}' P\bar{\epsilon}/N) = \sigma_1^2 \operatorname{tr}(W'W)/N\\
&E(\bar{\epsilon}' P\epsilon/N) = 0
\end{aligned} \tag{10.37}$$

From (10.35), $\epsilon = u - \lambda\bar{u}$ and $\bar{\epsilon} = \bar{u} - \lambda\bar{\bar{u}}$, substituting these expressions in (10.37) we obtain a system of six equations involving the second moments of u, $\bar{u}$ and $\bar{\bar{u}}$. Under the random effects specification considered, the OLS estimator of β is consistent. Using $\widehat{\beta}_{\text{OLS}}$ one gets a consistent estimator of the disturbances $\widehat{u} = y - X\widehat{\beta}_{\text{OLS}}$. The GM estimator of σ_1^2, σ_ν^2 and λ is the solution of the sample counterpart of the six equations in (10.37).

Kapoor et al. (2004) suggest three GM estimators. The first involves only the first three moments in (10.37) which do not involve σ_1^2 and yield estimates of λ and σ_ν^2. The fourth moment condition is then used to solve for σ_1^2 given estimates of λ and σ_ν^2. Kapoor et al. (2004) give the conditions needed for the consistency of this estimator as $N \to \infty$. The second GM estimator is based upon weighing the moment equations by the inverse of a properly normalized variance–covariance matrix of the sample moments evaluated at the true parameter values. A simple version of this weighting matrix is derived under normality of the disturbances. The third GM estimator is motivated by computational considerations and replaces a component of the weighting matrix for the second GM estimator by an identity matrix. Kapoor et al. (2004) perform Monte Carlo experiments comparing MLE and these three GM estimation methods. They find that on average, the RMSE of ML and their weighted GM estimators are quite similar. However, the first unweighted GM estimator has a RMSE that is 17% to 14% larger than that of the weighted GM estimators.

10.6 SHORT-RUN VS LONG-RUN ESTIMATES IN POOLED MODELS

Applied studies using panel data find that the Between estimator (which is based on the cross-sectional component of the data) tends to give long-run estimates while the Within estimator (which is based on the time-series component of the data) tends to give short-run estimates.

This agrees with the folk wisdom that cross-sectional studies tend to yield long-run responses while time-series studies tend to yield short-run responses (see Kuh, 1959; Houthakker, 1965). Both are consistent estimates of the same regression coefficients as long as the disturbances are uncorrelated with the explanatory variables. In fact, Hausman's specification test is based on the difference between these estimators (see Chapter 4). Rejection of the null implies that the random individual effects are correlated with the explanatory variables. This means that the Between estimator is inconsistent while the Within estimator is consistent since it sweeps away the individual effects. In these cases, the applied researcher settles on the Within estimator rather than the Between or GLS estimators. (See Mundlak, 1978 for additional support of the Within estimator.) Baltagi and Griffin (1984) argue that in panel data models, the difference between the Within and Between estimators is due to dynamic misspecification. The basic idea is that even with a rich panel data set, long-lived lag effects coupled with the shortness of the time series is a recipe for dynamic underspecification. This is illustrated using Monte Carlo experiments. In this context, Pirotte (1999) showed that the probability limit of the Between estimator for a static panel data regression converges to the long-run effect. This occurs despite the fact that the true model is a *dynamic* error components model. The only requirements are that the number of individuals tend to infinity with the time periods held fixed and the coefficients of the model are homogeneous among individual units. Egger and Pfaffermayr (2004a) show that the asymptotic bias of the Within and Between estimators as estimates of short-run and long-run effects depend upon the memory of the data generating process, the length of the time series and the importance of the cross-sectional variation in the explanatory variables. Griliches and Hausman (1986) attribute the difference between the Within and Between estimators to measurement error in panel data (see section 10.1). Mairesse (1990) tries to explain why these two estimators differ in economic applications using three samples of large manufacturing firms in France, Japan and the USA over the period 1967–79, and a Cobb–Douglas production function. Mairesse (1990) compares OLS, Between and Within estimators using levels and first-differenced regressions with and without constant returns to scale. Assuming constant returns to scale, he finds that the Between estimates of the elasticity of capital are of the order of 0.31 for France, 0.47 for Japan and 0.22 for the USA, whereas the Within estimates are lower, varying from 0.20 for France to 0.28 for Japan and 0.21 for the USA. Mairesse (1990) argues that if the remainder error ν_{it} is correlated with the explanatory variables, then the Within estimator will be inconsistent, while the Between estimator is much less affected by these correlations because the ν_{it} are averaged and practically wiped out for large enough T. This is also the case when measurement error in the explanatory variables is present. In fact, if these measurement errors are not serially correlated from one year to the next, the Between estimator tends to minimize their importance by averaging. In contrast, the Within estimator magnifies the variability of these measurement errors and increases the resulting bias. (For additional arguments in favor of the Between estimator, see Griliches and Mairesse, 1984.)

10.7 HETEROGENEOUS PANELS

For panel data studies with large N and small T, it is usual to pool the observations, assuming homogeneity of the slope coefficients. The latter is a testable assumption which is quite often rejected, see Chapter 4. Moreover, with the increasing time dimension of panel data sets, some researchers including Robertson and Symons (1992) and Pesaran and Smith (1995) have questioned the poolability of the data across heterogeneous units. Instead, they argue in favor of heterogeneous estimates that can be combined to obtain homogeneous estimates if the need arises. To make this point, Robertson and Symons (1992) studied the properties of some panel data estimators when the regression coefficients vary across individuals, i.e., they are *heterogeneous* but are assumed *homogeneous* in estimation. This is done for both stationary and nonstationary regressors. The basic conclusion is that severe biases can occur in dynamic estimation even for relatively small parameter variation. They consider the case of say two countries ($N = 2$), where the asymptotics depend on $T \to \infty$. Their true model is a simple *heterogeneous static* regression model with one regressor:

$$y_{it} = \beta_i x_{it} + \nu_{it} \quad i = 1, 2; t = 1, \ldots, T \tag{10.38}$$

where ν_{it} is independent for $i = 1, 2$, and β_i varies across $i = 1, 2$. However, their estimated model is *dynamic and homogeneous* with $\beta_1 = \beta_2 = \beta$ and assumes an identity covariance matrix for the disturbances:

$$y_{it} = \lambda y_{i,t-1} + \beta x_{it} + w_{it} \quad i = 1, 2 \tag{10.39}$$

The regressors are assumed to follow a stationary process $x_{it} = \rho x_{i,t-1} + \epsilon_{it}$ with $|\rho| < 1$ but different variances σ_i^2 for $i = 1, 2$. Seemingly unrelated regression estimation with the equality restriction imposed and an identity covariance matrix reduces to OLS on this system of two equations. Robertson and Symons (1992) obtain the probability limits of the resulting $\widehat{\lambda}$ and $\widehat{\beta}$ as $T \to \infty$. They find that the coefficient λ of $y_{i,t-1}$ is overstated, while the mean effect of the regressors (the x_{it}) is underestimated. In case the regressors are random walks ($\rho = 1$), then plim $\widehat{\lambda} = 1$ and plim $\widehat{\beta} = 0$. Therefore, false imposition of parameter homogeneity, and dynamic estimation of a static model when the regressors follow a random walk lead to perverse results. Using Monte Carlo experiments they show that the dynamics become misleading even for T as small as 40, which corresponds to the annual postwar data period. Even though these results are derived for $N = 2$, one regressor and no lagged dependent variable in the true model, Robertson and Symons (1992) show that the same phenomenon occurs for an empirical example of a real wage equation for a panel of 13 OECD countries observed over the period 1958–86. Parameter homogeneity across countries is rejected and the true relationship appears dynamic. Imposing false equality restriction biases the coefficient of the lagged wage upwards and the coefficient of the capital–labor ratio downwards.

For typical labor or consumer panels where N is large but T is fixed, Robertson and Symons (1992) assume that the true model is given by (10.35) with $\beta_i \sim \text{IID}(\beta, \sigma_\beta^2)$ for $i = 1, \ldots, N$, and $\nu_{it} \sim \text{IID}(0, 1)$. In addition, x_{it} is AR(1) with innovations $\epsilon_{it} \sim \text{IID}(0, 1)$ and $x_{i0} = \nu_{i0} = 0$. The estimated model is dynamic as given by (10.36), with known variance–covariance matrix I, and $\beta_i = \beta$ imposed for $i = 1, \ldots, N$. For fixed T, and random walk regressors, plim $\widehat{\lambda} > 0$ and plim $\widehat{\beta} < \beta$ as $N \to \infty$, so that the coefficient of $y_{i,t-1}$ is overestimated and the mean effect of the β_i is underestimated. As $T \to \infty$, one gets the same result obtained previously, plim $\widehat{\lambda} = 1$ and plim $\widehat{\beta} = 0$. If the regressor x_{it} is white noise, no biases arise. These results are confirmed with Monte Carlo experiments for $T = 5$ and $N = 50$, 100 and 200. The dynamics are overstated even for $N = 50$ and $T = 5$, but they disappear as the regressor approaches white noise, and remain important for autoregressive regressors with $\rho = 0.5$. Finally, Robertson and Symons (1992) reconsider the Anderson and Hsiao (1982) estimator of a dynamic panel data model that gets rid of the individual effects by first-differencing and uses lagged first-differences of the regressors as instruments. Imposing false equality restrictions renders these instruments invalid unless x_{it} is white noise or follows a random walk. Only the second case is potentially important because many economic variables are well approximated by random walks. However, Robertson and Symons (1992) show that if x_{it} is a random walk, the instrument is orthogonal to the instrumented variable and the resulting estimator has infinite asymptotic variance, a result obtained in the stationary case by Arellano (1989). Using levels ($y_{i,t-2}$) as instruments as suggested by Arellano (1989) will not help when x_{it} is a random walk, since the correlation between the stationary variable ($y_{i,t-1} - y_{i,t-2}$) and the $I(1)$ variable $y_{i,t}$ will be asymptotically zero. Using Monte Carlo experiments, with $T = 5$ and $N = 50$, Robertson and Symons (1992) conclude that the Anderson and Hsiao (1982) estimator is useful only when x_{it} is white noise or a random walk. Otherwise, severe biases occur when x_{it} is stationary and autocorrelated.

Pesaran and Smith (1995) consider the problem of estimating a dynamic panel data model when the parameters are individually heterogeneous and illustrate their results by estimating industry-specific UK labor demand functions. In this case the model is given by

$$y_{it} = \lambda_i y_{i,t-1} + \beta_i x_{it} + u_{it} \quad i = 1, \ldots, N; \; t = 1, \ldots, T \tag{10.40}$$

where λ_i is IID$(\lambda, \sigma_\lambda^2)$ and β_i is IID(β, σ_β^2). Further, λ_i and β_i are independent of y_{is}, x_{is} and u_{is} for all s. The objective in this case is to obtain consistent estimates of the mean values of λ_i and β_i. Pesaran and Smith (1995) present four different estimation procedures:

(1) aggregate time series regressions of group averages;
(2) cross-section regressions of averages over time;
(3) pooled regressions allowing for fixed or random intercepts; or
(4) separate regressions for each group, where coefficients estimates are averaged over these groups.

They show that when T is small (even if N is large), all the procedures yield inconsistent estimators. The difficulty in obtaining consistent estimates for λ and β can be explained by rewriting (10.37) as

$$y_{it} = \lambda y_{i,t-1} + \beta x_{it} + \nu_{it} \tag{10.41}$$

where $\nu_{it} = u_{it} + (\lambda_i - \lambda) y_{i,t-1} + (\beta_i - \beta) x_{it}$. By continuous substitution of $y_{i,t-s}$ it is easy to see that ν_{it} is correlated with all present and past values of $y_{i,t-1-s}$ and x_{it-s} for $s \geqslant 0$. The fact that ν_{it} is correlated with the regressors of (10.38) renders the OLS estimator inconsistent, and the fact that ν_{it} is correlated with $(y_{i,t-1-s}, x_{i,t-s})$ for $s > 0$ rules out the possibility of choosing any lagged value of y_{it} and x_{it} as legitimate instruments. When both N and T are large, Pesaran and Smith (1995) show that the cross-section regression procedure will yield consistent estimates of the mean values of λ and β. Intuitively, when T is large, the individual parameters λ_i and β_i can be consistently estimated using T observations of each individual i, say $\widehat{\lambda}_i$ and $\widehat{\beta}_i$, then averaging these individual estimators, $\sum_{i=1}^N \widehat{\lambda}_i / N$ and $\sum_{i=1}^N \widehat{\beta}_i / N$, will lead to consistent estimators of the mean values of λ and β.

Maddala et al. (1997) on the other hand argued that the heterogeneous time series estimates yield inaccurate estimates and even wrong signs for the coefficients, while the panel data estimates are not valid when one rejects the hypothesis of homogeneity of the coefficients. They argued that shrinkage estimators are superior to either heterogeneous or homogeneous parameter estimates, especially for prediction purposes. In fact, Maddala et al. (1997) considered the problem of estimating short-run and long-run elasticities of residential demand for electricity and natural gas for each of 49 states over the period 1970–90. They conclude that individual heterogeneous state estimates were hard to interpret and had the wrong signs. Pooled data regressions were not valid because the hypothesis of homogeneity of the coefficients was rejected. They recommend shrinkage estimators if one is interested in obtaining elasticity estimates for each state since these give more reliable results.

In the context of dynamic demand for gasoline across 18 OECD countries over the period 1960–90, Baltagi and Griffin (1997) argued for pooling the data as the best approach for obtaining reliable price and income elasticities. They also pointed out that pure cross-section studies cannot control for unobservable country effects, whereas pure time-series studies cannot control for unobservable oil shocks or behavioral changes occurring over time. Baltagi and Griffin (1997) compared the homogeneous and heterogeneous estimates in the context of gasoline demand based on the plausibility of the price and income elasticities as well as the speed of adjustment path to the long-run equilibrium. They found considerable variability in the parameter estimates among the heterogeneous estimators, some implausible estimates, while the homogeneous estimators gave similar plausible short-run estimates that differed only in estimating the long-run effects. Baltagi and Griffin (1997) also compared the forecast performance of these homogeneous and heterogeneous estimators over one-, five- and ten-year horizons. Their findings show that the homogeneous estimators outperformed their heterogeneous counterparts based on mean squared forecast error. This result was replicated using a panel data set of 21 French regions over the period 1973–98 by Baltagi, Bresson, Griffin and Pirotte (2003). Unlike the international OECD gasoline data set, the focus on the interregional differences in gasoline prices and income within France posed a different type of data set for the heterogeneity vs homogeneity debate. The variations in these prices and income were much smaller than international price and income differentials. This in turn reduces the efficiency gains from pooling and favors the heterogeneous estimators, especially given the differences between the Paris region and the rural areas of France. Baltagi et al. (2003) showed that the time series estimates for each region are highly variable, unstable and offer the worst out-of-sample forecasts. Despite the fact that the shrinkage estimators proposed by Maddala et al. (1997) outperformed these individual heterogeneous estimates, they still had a wide range and were outperformed by the homogeneous estimators in out-of-sample forecasts. Baltagi et al. (2000) carried out this comparison for a dynamic demand for cigarettes across 46 US states over 30 years (1963–92). Once again the results showed that the homogeneous panel data estimators beat the heterogeneous and shrinkage-type estimators in RMSE performance for out-of-sample forecasts. In another application, Driver et al. (2004) utilize the Confederation of British Industry's (CBI) survey data to measure the impact of uncertainty on UK investment authorizations. The panel consists of 48 industries observed over 85 quarters 1978(Q1) to 1999(Q1). The uncertainty measure is based on the dispersion of beliefs across survey respondents about the general business situation in their industry. The heterogeneous estimators considered are OLS and 2SLS at the industry level, as well as the unrestricted SUR estimation method. Fixed effects, random effects, pooled 2SLS and restricted SUR are the homogeneous estimators considered. The panel estimates find that uncertainty has a negative, nonnegligible effect on investment, while the heterogeneous estimates vary considerably across industries. Forecast performance for 12 out-of-sample quarters 1996(Q2) to 1999(Q1) are compared. The pooled homogeneous estimators outperform their heterogeneous counterparts in terms of RMSE.

Baltagi et al. (2002) reconsidered the two US panel data sets on residential electricity and natural gas demand used by Maddala et al. (1997) and compared the out-of-sample forecast performance of the homogeneous, heterogeneous and shrinkage estimators. Once again the results show that when the data is used to estimate heterogeneous models across states, individual estimates offer the worst out-of-sample forecasts. Despite the fact that shrinkage estimators outperform these individual estimates, they are outperformed by simple homogeneous panel data estimates in out-of-sample forecasts. Admittedly, these are additional case studies, but they do add to the evidence that simplicity and parsimony in model estimation offered by the homogeneous estimators yield better forecasts than the more parameter consuming heterogeneous estimators.

Hsiao and Tahmiscioglu (1997) use a panel of 561 US firms over the period 1971–92 to study the influence of financial constraints on company investment. They find substantial differences across firms in terms of their investment behavior. When a homogeneous pooled model is assumed, the impact of liquidity on firm investment is seriously underestimated. The authors recommend a mixed fixed and random coefficients framework based on the recursive predictive density criteria.

Pesaran, Smith and Im (1996) investigated the small sample properties of various estimators of the long-run coefficients for a dynamic heterogeneous panel data model using Monte Carlo experiments. Their findings indicate that the mean group estimator performs reasonably well for large T. However, when T is small, the mean group estimator could be seriously biased, particularly when N is large relative to T. Pesaran and Zhao (1999) examine the effectiveness of alternative bias correction procedures in reducing the small sample bias of these estimators using Monte Carlo experiments. An interesting finding is that when the coefficient of the lagged dependent variable is greater than or equal to 0.8, none of the bias correction procedures seems to work. Hsiao, Pesaran and Tahmiscioglu (1999) suggest a Bayesian approach for estimating the mean parameters of a dynamic heterogeneous panel data model. The coefficients are assumed to be normally distributed across cross-sectional units and the Bayes estimator is implemented using Markov chain Monte Carlo methods. Hsiao et al. (1999) argue that Bayesian methods can be a viable alternative in the estimation of mean coefficients in dynamic panel data models, even when the initial observations are treated as fixed constants. They establish the asymptotic equivalence of this Bayes estimator and the mean group estimator proposed by Pesaran and Smith (1995). The asymptotics are carried out for both N and $T \to \infty$ with $\sqrt{N}/T \to 0$. Monte Carlo experiments show that this Bayes estimator has better sampling properties than other estimators for both small and moderate size T. Hsiao et al. also caution against the use of the mean group estimator unless T is sufficiently large relative to N. The bias in the mean coefficient of the lagged dependent variable appears to be serious when T is small and the true value of this coefficient is larger than 0.6. Hsiao et al. apply their methods to estimate the q-investment model using a panel of 273 US firms over the period 1972–93. Baltagi et al. (2004) reconsider the Tobin q-investment model studied by Hsiao et al. (1999) using a slightly different panel of 337 US firms over the period 1982–98. They contrast the out-of-sample forecast performance of 9 homogeneous panel data estimators and 11 heterogeneous and shrinkage Bayes estimators over a five-year horizon. Results show that the average heterogeneous estimators perform the worst in terms of mean squared error, while the hierarchical Bayes estimator suggested by Hsiao et al. (1999) performs the best. Homogeneous panel estimators and iterative Bayes estimators are a close second. In conclusion, while the performance of various estimators and their corresponding forecasts may vary in ranking from one empirical example to another, the consistent finding in all these studies is that homogeneous panel data estimators perform well in forecast performance mostly due to their simplicity, their parsimonious representation and the stability of the parameter estimates. Average heterogeneous estimators perform badly due to parameter estimate instability caused by the estimation of several parameters with short time series. Shrinkage estimators did well for some applications, especially iterative Bayes and iterative empirical Bayes.

NOTES

1. Peracchi and Welch (1995) use the Current Population Survey (CPS) to illustrate some problems that arise from analyzing panel data constructed by matching person records across files of rotating cross-section surveys. In particular, the matched CPS is studied to understand the process of attrition from the sample and the nature of measurement error.
2. Estimation of the consumer price index in the USA is based on a complex rotating panel survey, with 20% of the sample being replaced by rotation each year (see Valliant, 1991).
3. In general, for any T, as long as the fraction of the sample being rotated is greater than or equal to half, then no individual will be observed more than twice.
4. The terms "panel conditioning", "reinterview effect" and "rotation group bias" are also used in the literature synonymously with "time-in-sample bias" effects.
5. Blundell and Meghir (1990) also argue that pseudo-panels allow the estimation of lifecycle models which are free from aggregation bias. In addition, Moffitt (1993) explains that many researchers in the USA prefer to use pseudo-panels like the CPS because it has larger, more representative samples and the questions asked are more consistently defined over time than the available US panels.
6. Table 1 of McKenzie (2001) provides examples of unequally spaced surveys and their sources.
7. Two special cases of this general specification are also considered. The first assumes that Σ is diagonal, with no correlation across different cross-sections but allowing for heteroskedasticity. The second special case uses the additional restriction that all the ρ_i are equal for $i = 1, 2, \ldots, N$. Since OLS is still unbiased and consistent under this model, it can be used to estimate the ρ_i's and Σ. Also, it can be used to obtain robust estimates of the variance–covariance of the OLS estimator, see Beck and Katz (1995). This can be done using the xtpcse command in Stata.

PROBLEMS

10.1 This problem is based upon Griliches and Hausman (1986). Using the simple regression given in (10.1)–(10.3):

(a) Show that for the first-difference (FD) estimator of β, the expression in (10.8) reduces to

$$\text{plim}\,\widehat{\beta}_{\text{FD}} = \beta\left(1 - \frac{2\sigma_\eta^2}{\text{var}(\Delta x)}\right)$$

where $\Delta x_{it} = x_{it} - x_{i,t-1}$.

(b) Also show that (10.8) reduces to

$$\text{plim}\,\widetilde{\beta}_W = \beta\left(1 - \frac{T-1}{T}\frac{\sigma_\eta^2}{\text{var}(\widetilde{x})}\right)$$

where $\widetilde{\beta}_W$ denotes the Within estimator and $\widetilde{x}_{it} = x_{it} - \bar{x}_i$.

(c) For most economic series, the x_{it}^* are positively serially correlated exhibiting a declining correlogram, with

$$\text{var}(\Delta x) < \frac{2T}{T-1}\text{var}(\widetilde{x}) \quad \text{for } T > 2$$

Using this result, conclude that

$$|\text{ bias }\widehat{\beta}_{\text{FD}}\ | > |\text{ bias }\widetilde{\beta}_W\ |$$

(d) Solve the expressions in parts (a) and (b) for β and σ_η^2 and verify that the expressions in (10.9) and (10.10) reduce to

$$\widehat{\beta} = \frac{[2\widetilde{\beta}_W/\text{var}(\Delta x) - (T-1)\widehat{\beta}_{\text{FD}}/T\ \text{var}(\widetilde{x})]}{[2/\text{var}(\Delta x) - (T-1)/T\ \text{var}(\widetilde{x})]}$$

$$\sigma_\eta^2 = (\widehat{\beta} - \widehat{\beta}_{\text{FD}})\ \text{var}(\Delta x)/2\widehat{\beta}$$

(e) For $T = 2$, the Within estimator is the same as the first-difference estimator since $\frac{1}{2}\Delta x_{it} = \widetilde{x}_{it}$. Verify that the expressions in part (a) and (b) are also the same.

10.2 For the rotating panel considered in section 10.2, assume that $T = 3$ and that the number of households being replaced each period is equal to $N/2$.
(a) Derive the variance–covariance of the disturbances Ω.
(b) Derive $\Omega^{-1/2}$ and describe the transformation needed to make GLS a weighted least squares regression.
(c) How would you consistently estimate the variance components σ_μ^2 and σ_ν^2?
(d) Repeat this exercise for the case where the number of households being replaced each period is $N/3$. How about $2N/3$?

10.3 Using the Grunfeld data, perform
(a) the common ρ and
(b) the varying ρ estimation methods, described in Baltagi (1986). Compare with the error component estimates obtained in Chapter 2.

10.4 Using the gasoline data, perform
(a) the common ρ and
(b) the varying ρ estimation methods, described in Baltagi (1986). Compare with the error component estimates obtained in Chapter 2.

10.5 Using the Monte Carlo set-up of Baltagi (1986), compare the timewise autocorrelated, cross-sectionally heteroskedastic estimation method with the error component method and observe which method is more robust to misspecification.

10.6 *Prediction in the spatially autocorrelated error component model*. This is based on problem 99.2.4 in *Econometric Theory* by Baltagi and Li (1999). Consider the panel data regression model described in (10.25) with random country effects and spatially autocorrelated remainder disturbances described by (10.26) and (10.27). Using the Goldberger (1962) best linear unbiased prediction results discussed in section 2.5, equation (2.37), derive the BLUP of $y_{i,T+S}$ for the ith country at period $T+S$ for this spatial panel model. *Hint*: See solution 99.2.4 in *Econometric Theory* by Song and Jung (2000).

10.7 Download the Maddala et al. (1997) data set on residential natural gas and electricity consumption for 49 states over 21 years (1970–90) from the *Journal of Business and Economic Statistics* web site www.amstat.org/publications/jbes/ftp.html.
(a) Using this data set, replicate the individual OLS state regressions for electricity, given in table 6 and natural gas, given in table 8 of Maddala et al. (1997).
(b) Replicate the shrinkage estimates for electricity and natural gas given in tables 7 and 9, respectively.
(c) Replicate the fixed effects estimator given in column 1 of table 2 and the pooled OLS model given in column 2 of that table.
(d) Replicate the average OLS, the average shrinkage estimator and the average Stein rule estimator in table 2.
(e) Redo parts (c) and (d) for the natural gas equation as given in table 4 in that paper.

11 Limited Dependent Variables and Panel Data

In many economic studies, the dependent variable is discrete, indicating for example that a household purchased a car or that an individual is unemployed or that he or she joined a labor union or defaulted on a loan or was denied credit. This dependent variable is usually represented by a binary choice variable $y_{it} = 1$ if the event happens and 0 if it does not for individual i at time t. In fact, if p_{it} is the probability that individual i participated in the labor force at time t, then $E(y_{it}) = 1 \cdot p_{it} + 0 \cdot (1 - p_{it}) = p_{it}$, and this is usually modeled as a function of some explanatory variables

$$p_{it} = \Pr[y_{it} = 1] = E(y_{it}/x_{it}) = F(x'_{it}\beta) \tag{11.1}$$

For the linear probability model, $F(x'_{it}\beta) = x'_{it}\beta$ and the usual panel data methods apply except that $\widehat{y}_{it}$ is not guaranteed to lie in the unit interval. The standard solution has been to use the logistic or normal cumulative distribution functions that constrain $F(x'_{it}\beta)$ to be between zero and one. These probability functions are known in the literature as *logit* and *probit*, corresponding to the logistic and normal distributions, respectively.[1] For example, a worker participates in the labor force if his offered wage exceeds his unobserved reservation wage. This threshold can be described as

$$\begin{aligned} y_{it} &= 1 \quad \text{if } y^*_{it} > 0 \\ &= 0 \quad \text{if } y^*_{it} \leq 0 \end{aligned} \tag{11.2}$$

where $y^*_{it} = x'_{it}\beta + u_{it}$. So that

$$\Pr[y_{it} = 1] = \Pr[y^*_{it} > 0] = \Pr[u_{it} > -x'_{it}\beta] = F(x'_{it}\beta) \tag{11.3}$$

where the last equality holds as long as the density function describing F is symmetric around zero. This is true for the logistic and normal density functions.

11.1 FIXED AND RANDOM LOGIT AND PROBIT MODELS

For panel data, the presence of individual effects complicates matters significantly. To see this, consider the fixed effects panel data model, $y^*_{it} = x'_{it}\beta + \mu_i + \nu_{it}$ with

$$\Pr[y_{it} = 1] = \Pr[y^*_{it} > 0] = \Pr[\nu_{it} > -x'_{it}\beta - \mu_i] = F(x'_{it}\beta + \mu_i) \tag{11.4}$$

where the last equality holds as long as the density function describing F is symmetric around zero. In this case, μ_i and β are unknown parameters and as $N \to \infty$, for a fixed T, the number of parameters μ_i increases with N. This means that μ_i cannot be consistently estimated for a fixed T. This is known as the *incidental parameters problem* in statistics, which is discussed by Neyman and Scott (1948) and reviewed more recently by Lancaster (2000). For the linear panel data regression model, when T is fixed, only β was estimated consistently by first getting rid of the μ_i using the Within transformation.[2] This was possible for the linear case because the MLE of β and μ_i are asymptotically independent (see Hsiao, 2003). This is no longer the case for a qualitative limited dependent variable model with fixed T as demonstrated by Chamberlain (1980). For a simple illustration of how the inconsistency of the MLE of μ_i is transmitted into inconsistency for $\widehat{\beta}_{mle}$, see Hsiao (2003). This is done in the context of a logit model with one regressor x_{it} that is observed over two periods, with $x_{i1} = 0$ and $x_{i2} = 1$. Hsiao shows that as $N \to \infty$ with $T = 2$, plim $\widehat{\beta}_{mle} = 2\beta$, see also problem 11.4. Greene (2004a) shows that despite the large number of incidental parameters, one can still perform maximum likelihood for the fixed effects model by brute force, i.e., including a large number of dummy variables. Using Monte Carlo experiments, he shows that the fixed effects MLE is biased even when T is large. For $N = 1000$, $T = 2$ and 200 replications, this bias is 100%, confirming the results derived by Hsiao (2003). However, this bias improves as T increases. For example, when $N = 1000$ and $T = 10$ this bias is 16% and when $N = 1000$ and $T = 20$ this bias is 6.9%.

The usual solution around this incidental parameters problem is to find a minimal sufficient statistic for μ_i. For the logit model, Chamberlain (1980) finds that $\sum_{t=1}^{T} y_{it}$ is a minimum sufficient statistic for μ_i. Therefore, Chamberlain suggests maximizing the *conditional* likelihood function

$$L_c = \prod_{i=1}^{N} \Pr\left(y_{i1}, \ldots, y_{iT} / \sum_{t=1}^{T} y_{it} \right) \tag{11.5}$$

to obtain the *conditional logit* estimates for β. By definition of a sufficient statistic, the distribution of the data given this sufficient statistic will not depend on μ_i. For the fixed effects logit model, this approach results in a computationally convenient estimator and the basic idea can be illustrated for $T = 2$. The observations over the two periods and for all individuals are independent and the *unconditional* likelihood is given by

$$L = \prod_{i=1}^{N} \Pr(y_{i1}) \Pr(y_{i2}) \tag{11.6}$$

The sum $(y_{i1} + y_{i2})$ can be 0, 1 or 2. If it is 0, both y_{i1} and y_{i2} are 0 and

$$\Pr[y_{i1} = 0, y_{i2} = 0 / y_{i1} + y_{i2} = 0] = 1 \tag{11.7}$$

Similarly, if the sum is 2, both y_{i1} and y_{i2} are 1 and

$$\Pr[y_{i1} = 1, y_{i2} = 1 / y_{i1} + y_{i2} = 2] = 1 \tag{11.8}$$

These terms add nothing to the conditional loglikelihood since $\log 1 = 0$. Only the observations for which $y_{i1} + y_{i2} = 1$ matter in $\log L_c$ and these are given by

$$\Pr[y_{i1} = 0, y_{i2} = 1 / y_{i1} + y_{i2} = 1] \quad \text{and} \quad \Pr[y_{i1} = 1, y_{i2} = 0 / y_{i1} + y_{i2} = 1]$$

The latter can be calculated as $\Pr[y_{i1} = 1, y_{i2} = 0] / \Pr[y_{i1} + y_{i2} = 1]$ with

$$\Pr[y_{i1} + y_{i2} = 1] = \Pr[y_{i1} = 0, y_{i2} = 1] + \Pr[y_{i1} = 1, y_{i2} = 0]$$

since the latter two events are mutually exclusive. From (11.4), the logit model yields

$$\Pr[y_{it} = 1] = \frac{e^{\mu_i + x'_{it}\beta}}{1 + e^{\mu_i + x'_{it}\beta}} \tag{11.9}$$

Therefore

$$\Pr[y_{i1} = 1, y_{i2} = 0 \mid y_{i1} + y_{i2} = 1] = \frac{e^{x'_{i1}\beta}}{e^{x'_{i1}\beta} + e^{x'_{i2}\beta}} = \frac{1}{1 + e^{(x_{i2} - x_{i1})'\beta}} \tag{11.10}$$

Similarly

$$\Pr[y_{i1} = 0, y_{i2} = 1 \mid y_{i1} + y_{i2} = 1] = \frac{e^{x'_{i2}\beta}}{e^{x'_{i1}\beta} + e^{x'_{i2}\beta}} = \frac{e^{(x_{i2} - x_{i1})'\beta}}{1 + e^{(x_{i2} - x_{i1})'\beta}} \tag{11.11}$$

and neither probability involves the μ_i. Therefore, by conditioning on $y_{i1} + y_{i2}$, we swept away the μ_i. The product of terms such as these with $y_{i1} + y_{i2} = 1$ gives the conditional likelihood function which can be maximized with respect to β using conventional maximum likelihood logit programs. In this case, only the observations for individuals who *switched* status are used in the estimation. A standard logit package can be used with $x'_{i2} - x'_{i1}$ as explanatory variables and the dependent variable taking the value one if y_{it} switches from 0 to 1, and zero if y_{it} switches from 1 to 0. This procedure can easily be generalized for $T > 2$ (see problem 11.1).

In order to test for fixed individual effects one can perform a Hausman-type test based on the difference between Chamberlain's conditional MLE and the usual logit MLE ignoring the individual effects. The latter estimator is consistent and efficient only under the null of no individual effects and inconsistent under the alternative. Chamberlain's estimator is consistent whether H_0 is true or not, but it is inefficient under H_0 because it may not use all the data. Both estimators can easily be obtained from the usual logit ML routines. The constant is dropped and estimates of the asymptotic variances are used to form Hausman's χ^2 statistic. This will be distributed as χ^2_K under H_0. For an application of Chamberlain's conditional MLE see Björklund (1985) who studied the linkage between unemployment and mental health problems in Sweden using the Swedish Level of Living Surveys. The data was based on a random sample of 6500 individuals between the ages of 15 and 75 surveyed in 1968, 1974 and 1981. Another application is Winkelmann and Winkelmann (1998) who applied the conditional logit approach to study the effect of unemployment on the level of satisfaction. Using data from the first six waves of the GSOEP over the period 1984–89, the authors showed that unemployment had a large detrimental effect on satisfaction. This effect became even larger after controlling for individual-specific effects. The dependent variable was based on the response to the question "How satisfied are you at present with your life as a whole?" An ordinal scale from 0 to 10 is recorded, where 0 meant "completely dissatisfied" and 10 meant "completely satisfied". Winkelmann and Winkelmann constructed a binary variable taking the value 1 if this score was above 7 and 0 otherwise. They justified this on the basis that average satisfaction was between 7 and 8 and this was equivalent to classifying individuals into those who reported above and those who reported below average satisfaction. The explanatory variables included a set of dummy variables indicating current labor market status (unemployed out of the labor force) with employed as the reference category. A good health variable defined as the absence of any chronic condition or handicap. Age, age-squared, marital status and the duration of unemployment and its square. Since unemployment reduces income which in turn may reduce satisfaction, household income was included as a control variable to measure the nonpecuniary effect of unemployment holding income constant. Of particular concern with the measurement of life satisfaction is that individuals "anchor" their scale at different levels, rendering interpersonal comparisons of responses meaningless. This problem bears a close resemblance to the issue of cardinal vs ordinal utility. Any statistic that is calculated from a cross-section of individuals, for instance an average satisfaction, requires cardinality of the measurement scale. This problem is closely related to the unobserved individual-specific effects. Hence anchoring causes the estimator to be biased as long as it is not random but correlated with the explanatory variables. Panel data help if the metric used by individuals is time-invariant. Fixed effects makes inference based on intra- rather than interpersonal comparisons of satisfaction. This avoids not only the potential bias caused by anchoring, but also bias caused by other unobserved individual-specific factors. Hausman's test based on the difference between a standard logit and a fixed effects logit yielded a significant χ^2 variable. After controlling for individual-specific effects, this study found that unemployment had a significant and substantial negative impact on satisfaction. The nonpecuniary costs of unemployment by far exceeded the pecuniary costs associated with loss of income while unemployed.

In cases where the conditional likelihood function is not feasible, Manski (1987) shows that it is possible to relax the logistic assumption in (11.9). Manski allows for a strictly increasing distribution function which differs across individuals, but not over time for a given individual. For $T = 2$, the identification of β is based on the fact that, under certain regularity conditions on the distribution of the exogenous variables,

$$\text{sgn}\,[\Pr(y_{i2} = 1/x_{i1}, x_{i2}, \mu_i) - \Pr(y_{i1} = 1/x_{i1}, x_{i2}, \mu_i)] = \text{sgn}\,[(x_{i2} - x_{i1})'\beta]$$

This prompted Manski (1987) to suggest a conditional version of his *maximum score estimator* which can be applied to the first differences of the data in the subsample for which $y_{i1} \neq y_{i2}$. This estimator leaves the distribution of the errors unspecified but it requires these disturbances to be stationary conditional on the sequence of explanatory variables. Unlike the conditional logit approach, Manski's estimator is not root-N consistent nor asymptotically normal.[3] It is consistent as $N \to \infty$ if the conditional distribution of the disturbances u_{it} given μ_i, x_{it} and $x_{i,t-1}$ is identical to the conditional distribution of $u_{i,t-1}$ conditional on μ_i, x_{it} and $x_{i,t-1}$. This does not allow for the presence of lagged dependent variables among the regressors. Manski's approach is semiparametric and does not require the specification of the distribution of the disturbances. Hence, unlike standard MLE, it is robust to misspecification of the likelihood. However, this semiparametric approach cannot be used to generate predicted probabilities conditional on the regressors as in the parametric approach. Chamberlain (1993) showed that even if the distribution of the disturbances is known, the logit model is the only version of (11.4) for which β can be estimated at rate root-N. Honoré and Lewbel (2002) show that Chamberlain's negative result can be overturned as long as there is one explanatory variable which is independent of the fixed effects and the ν_{it}'s, conditional on the other explanatory variables and on a set of instruments. This assumption allows the root-N consistent estimation of the parameters of the binary choice model with individual-specific effects which are valid even when the explanatory variables are predetermined as opposed to strictly exogenous. See also Magnac (2004) who shows that in a two-period, two-state model, the sum of two binary variables is a sufficient statistic for the individual effect, under necessary and sufficient conditions that are much less restrictive than the conditional logit approach. He also shows that if the covariates are unbounded, then consistent estimation at a $\sqrt{N}$ rate is possible if and only if the sum of the binary variables is a sufficient statistic.

In contrast to the fixed effects logit model, the conditional likelihood approach does not yield computational simplifications for the fixed effects *probit* model. But the probit specification has been popular for the random effects model. In this case, $u_{it} = \mu_i + \nu_{it}$ where $\mu_i \sim \text{IIN}(0, \sigma_\mu^2)$ and $\nu_{it} \sim \text{IIN}(0, \sigma_\nu^2)$ independent of each other and the x_{it}. Since $E(u_{it}u_{is}) = \sigma_\mu^2$ for $t \neq s$, the joint likelihood of $(y_{1t}, \ldots, y_{Nt})$ can no longer be written as the product of the marginal likelihoods of the y_{it}. This complicates the derivation of maximum likelihood which will now involve T-dimensional integrals.[4] This gets to be infeasible if T is big. However, by conditioning on the individual effects, this T-dimensional integral problem reduces to a single integral involving the product of a standard normal density and T differences of two normal cumulative density functions. This can be evaluated using the Gaussian quadrature procedure suggested by Butler and Moffitt (1982). This approach has the advantage of being computationally feasible even for fairly large T. The accuracy of this quadrature procedure increases with the number of evaluation points. For an application of the *random effects probit* model, see Sickles and Taubman (1986) who estimated a two-equation structural model of the health and retirement decisions of the elderly using five biennial panels of males drawn from the Retirement History Survey. Both the health and retirement variables were limited dependent variables and MLE using the Butler and Moffitt (1982) Gaussian quadrature procedure was implemented. Sickles and Taubman found that retirement decisions were strongly affected by health status, and that workers not yet eligible for social security were less likely to retire.

Heckman (1981b) performed some limited Monte Carlo experiments on a probit model with a single regressor and a Nerlove (1971a)-type x_{it}. For $N = 100, T = 8, \sigma_\nu^2 = 1$ and $\sigma_\mu^2 = 0.5$, 1 and 3, Heckman computed the bias of the fixed effects MLE of β using 25 replications. He found at most 10% bias for $\beta = 1$ which was always *towards* zero. Replicating Heckman's design and using 100 replications, Greene (2004a) finds that the bias of the fixed effects MLE of β is of the order of 10–24% always *away* from zero. Another Monte Carlo study by Guilkey and Murphy (1993) showed that ignoring the random effects and performing a standard probit analysis results in misleading inference since the coefficient standard errors are badly biased. However, a probit estimator with a corrected asymptotic covariance matrix performed as well as MLE for almost all parametric configurations. LIMDEP and Stata provide basic routines for the random and fixed effects logit and probit model. In fact, in Stata these are the (xtprobit and xtlogit) commands with the (fe and re) options.

Underlying the random effects probit model is the equicorrelation assumption between successive disturbances belonging to the same individual. Avery, Hansen and Hotz (1983) suggest a method of moments estimator that allows for a general type of serial correlation among the disturbances. They apply their "orthogonality condition" estimators to the study of labor force participation of married women. They reject the hypothesis of equicorrelation across the disturbances. However, these random effects probit methods assume that the μ_i and x_{it} are uncorrelated and this may be a serious limitation. If exogeneity is rejected, then one needs to put more structure on the type of correlation between μ_i and the x_{it} to estimate this model. This is what we will turn to next.

Chamberlain (1980, 1984) assumes that μ_i is correlated with x_{it} as follows:

$$\mu_i = x_i' a + \epsilon_i \tag{11.12}$$

where $a' = (a_1', \ldots, a_T')$, $x_i' = (x_{i1}', \ldots, x_{iT}')$ and $\epsilon_i \sim \text{IID}(0, \sigma_\epsilon^2)$ independent of ν_{it}. In this case,

$$y_{it} = 1 \quad \text{if } (x_{it}'\beta + x_i' a + \epsilon_i + \nu_{it}) > 0 \tag{11.13}$$

and the distribution of y_{it} conditional on x_{it} but marginal on μ_i has the probit form

$$\Pr[y_{it} = 1] = \Phi[(1 + \sigma_\epsilon^2)^{-1/2}(x_{it}'\beta + x_{it}' a)] \tag{11.14}$$

where Φ denotes the cumulative normal distribution function. Once again, MLE involves numerical integration, but a computationally simpler approach suggested by Chamberlain is to run simple probit on this equation to get $\widehat{\Pi}$. In this case, Π satisfies the restriction

$$\Pi = (1+\sigma_\epsilon^2)^{-1/2}(I_T \otimes \beta' + \iota_T a') \tag{11.15}$$

Therefore, Chamberlain suggests a minimum distance estimator based on $(\widehat{\pi} - \pi)$, where $\pi = \text{vec}(\Pi')$, that imposes this restriction. For details, see Chamberlain (1984).

Chamberlain (1984) applies both his *fixed effects logit* estimator and his *minimum distance random effects probit* estimator to a study of the labor force participation of 924 married women drawn from the PSID. These estimation methods give different results, especially with regard to the effect of the presence of young children on labor force participation. These different results could be attributed to the misspecification of the relationship between μ_i and the x_{it} in the random effects specification or a misspecification of the fixed effects logit model in its omission of leads and lags of the x_{it} from the structural equation.

For another application of Chamberlain's (1984) approach to panel data probit estimation, see Laisney, Lechner and Pohlmeier (1992) who studied the process innovation behavior of 1325 West German exporting firms observed over the 1984–88 period. Also, Lechner (1995) who suggests several specification tests for the panel data probit model. These are generalized score and Wald tests employed to detect omitted variables, neglected dynamics, heteroskedasticity, nonnormality and random coefficient variations. The performance of these tests in small samples is investigated using Monte Carlo experiments. In addition, an empirical example is given on the probability of self-employment in West Germany using a random sample of 1926 working men selected from the German Socio-Economic Panel and observed over the period 1984–89. Extensive Monte Carlo simulations comparing the performance of several simple GMM estimators for the panel data probit model are given by Lechner and Breitung (1996), Bertschek and Lechner (1998) and Breitung and Lechner (1999). Their results show that the efficiency loss for GMM when compared to maximum likelihood is small. In addition, these GMM estimators are easy to compute and are robust to serial correlation in the error. Asymptotically optimal GMM estimators based on the conditional mean function are obtained by using both parametric and nonparametric methods. The Monte Carlo results indicate that the nonparametric method is superior in small samples. Bertschek and Lechner (1998) apply these GMM procedures to the product innovation decisions of 1270 German firms observed over the period 1984–88. Greene (2004b) reconsiders the binomial panel data probit model studied by Bertschek and Lechner (1998) and argues that full maximum likelihood estimation for their data set is feasible. Although GMM and MLE are based on different assumptions, Greene (2004b) argues that the estimation of the full covariance matrix is revealing about the structure of the model in a way that would not be evident from the Bertschek and Lechner GMM approach. Greene additionally applies two alternative panel-based discrete choice models to this data set. These are the random parameters models and the latent class model. The computation of the latent class model is simple. However, the computation of the random parameters model is intensive and requires the maximization of a simulated loglikelihood function, see section 11.2.

Bover and Arellano (1997) provide extensions of the random effects probit model of Chamberlain (1984) which has applications in the analysis of binary choice, linear regression subject to censoring and other models with endogenous selectivity. They propose a simple two-step Within estimator for limited dependent variable models, which may include lags of the dependent variable, other exogenous variables and unobservable individual effects. This estimator is based on reduced form predictions of the latent endogenous variables. It can be regarded as a member of Chamberlain's class of random effects minimum distance estimators, and as such it is consistent and asymptotically normal for fixed T. However, this Within estimator is not asymptotically efficient within the minimum distance class, since it uses a nonoptimal weighting matrix. Therefore, Bover and Arellano (1997) show how one can obtain in one more step a chi-squared test statistic for over-identifying restrictions and linear GMM estimators that are asymptotically efficient. The drawbacks of this approach are the same as those for the Chamberlain probit model. Both require the availability of strictly exogenous variables, and the specification of the conditional distribution of the effects. Labeaga (1999) applies the Bover and Arellano (1997) method to estimate a double-hurdle rational addiction model for tobacco consumption using an unbalanced panel of households drawn from the Spanish Permanent Survey of Consumption (SPSC). This is a panel collected by the Spanish Statistical Office for approximately 2000 households between 1977 and 1983.

11.2 SIMULATION ESTIMATION OF LIMITED DEPENDENT VARIABLE MODELS WITH PANEL DATA

Keane (1994) derived a computationally practical simulation estimator for the panel data probit model. The basic idea of simulation estimation methods is to replace intractable integrals by unbiased Monte Carlo probability simulators. This is ideal for limited dependent variable models where for a multinominal probit model, the choice probabilities involve multivariate integrals.[5] In fact, for cross-section data, McFadden's (1989) method of simulated moments (MSM) involves an $M - 1$ integration problem, where M is the number of possible choices facing the individual. For panel data, things get more complicated, because there are M choices facing any individual at each period. This means that there are M^T possible choice sequences facing each individual over the panel. Hence the MSM estimator becomes infeasible as T gets large. Keane (1994) sidesteps this problem of having to simulate M^T possible choice sequences by factorizing the method of simulated moments first-order conditions into transition probabilities. The latter are simulated using highly accurate importance sampling techniques (see Keane, 1993, 1994 for details). This method of simulating probabilities is referred to as the Geweke, Hajivassiliou and Keane (GHK) simulator because it was independently developed by these authors. Keane (1994) performs Monte Carlo experiments and finds that even for large T and small simulation sizes, the bias in the MSM estimator is negligible. When maximum likelihood methods are feasible, Keane (1994) finds that the MSM estimator performs well relative to quadrature-based maximum likelihood methods even where the latter are based on a large number of quadrature points. When maximum likelihood is not feasible, the MSM estimator outperforms the simulated MLE even when the highly accurate GHK probability simulator is used. Keane (1994) argues that MSM has three advantages over the other practical nonmaximum likelihood estimators considered above, i.e. Chamberlain's (1984) minimum distance estimator and the Avery et al. (1983) orthogonality condition estimator. First, MSM is asymptotically as efficient as maximum likelihood (in simulation size) while the other estimators are not. Second, MSM can easily be extended to handle multinominal probit situations whereas the extension of the other estimators is computationally burdensome. Third, MSM can be extended to handle nonlinear systems of equations which are intractable with maximum likelihood. Keane (1994) also finds that MSM can estimate random effects models with autoregressive moving average error in about the same time necessary for estimating a simple random effects model using maximum likelihood quadrature. The extension of limited dependent variable models to allow for a general pattern of serial correlation is now possible using MSM and could prove useful for out-of-sample predictions. An example of the MSM estimator is given by Keane (1993), who estimates probit employment equations using data from the National Longitudinal Survey of Youth (NLSY). This is a sample of 5225 males aged 14–24 and interviewed 12 times over the period 1966–81. For this example, Keane (1993) concludes that relaxing the equicorrelation assumption by including an MA(1) or AR(1) component to the error term had little effect on the parameter estimates. Keane (1993) discusses simulation estimation of models more complex than probit models. He argues that it is difficult to put panel data selection models and Tobit models in an MSM framework and that the method of simulated scores (MSS) may be a preferable way to go. Keane (1993) applies the MSS estimator to the same data set used by Keane, Moffitt and Runkle (1988) to study the cyclical behavior of real wages. He finds that the Keane et al. conclusion of a weakly procyclical movement in the real wage appears to be robust to relaxation of the equicorrelation assumption. For another application, see Hajivassiliou (1994) who reconsiders the problem of external debt crisis of 93 developing countries observed over the period 1970–88. Using several simulation estimation methods, Hajivassiliou concludes that allowing for flexible correlation patterns changes the estimates substantially and raises doubts over previous studies that assumed restrictive correlation structures.

More recently, Zhang and Lee (2004) argue that the statistical performance of the GHK simulator may be adequate for panels with small T, but this performance deteriorates when T is larger than 50 (for a moderate amount of simulation draws). In fact, the bias of the SML estimator may become larger than its standard deviation. Zhang and Lee suggest applying the accelerated importance sampling (AIS) procedure to SML estimation of dynamic discrete choice models with long panels. Using Monte Carlo experiments, they show that this can improve upon the GHK sampler when T is large and they illustrate their method using an application on firm's dividend decisions. They collect data on quarterly dividends and earnings per share from COMPUSTAT tapes. The sample period is 54 quarters (1987:1 to 2002:2). Two quarters were used for getting the initial value for each firm, so $T = 52$. The final sample used included $N = 150$ large US industrial firms and the total number of observations $NT = 7800$. The results confirm that the AIS improves the performance of the GHK sampler.

11.3 DYNAMIC PANEL DATA LIMITED DEPENDENT VARIABLE MODELS

So far the model is static implying that, for example, the probability of buying a car is independent of the individual's past history of car purchases. If the probability of buying a car is more likely if the individual has bought a car in the past than if he or she has not, then a dynamic model that takes into account the individual's past experience is more appropriate. Heckman (1981a,b,c) gives an extensive treatment of these dynamic models and the consequences of various assumptions on the initial values on the resulting estimators. Heckman (1981c) also emphasizes the importance of distinguishing between *true state dependence* and *spurious state dependence*. In the "true" case, once an individual experiences an event like unemployment, his preferences change and he or she will behave differently in the future as compared with an identical individual that has not experienced this event in the past. In fact, it is observed that individuals with a long history of unemployment are less likely to leave unemployment. They may be less attractive for employers to hire or may become discouraged in looking for a job.

In the "spurious" case, past experience has no effect on the probability of experiencing the event in the future. It is the individual's characteristics that make him or her less likely to leave unemployment. However, one cannot properly control for all the variables that distinguish one individual's decision from another's. In this case, past experience which is a good proxy for these omitted variables shows up as a significant determinant of the future probability of occurrence of this event. Testing for true vs spurious state dependence is therefore important in these studies, but it is complicated by the presence of the individual effects or heterogeneity. In fact, even if there is no state dependence, $\Pr[y_{it}/x_{it}, y_{i,t-l}] \neq \Pr[y_{it}/x_{it}]$ as long as there are random individual effects present in the model. If in addition to the absence of the state dependence, there is also no heterogeneity, then $\Pr[y_{it}/x_{it}, y_{i,t-l}] = \Pr[y_{it}/x_{it}]$. A test for this equality can be based on a test for $\gamma = 0$ in the model

$$\Pr[y_{it} = 1/x_{it}, y_{it-1}] = F(x_{it}'\beta + \gamma y_{i,t-1})$$

using standard maximum likelihood techniques. If $\gamma = 0$ is not rejected, we ignore the heterogeneity issue and proceed as in conventional limited dependent variable models not worrying about the panel nature of the data. However, rejecting the null does not necessarily imply that there is heterogeneity since γ can be different from zero due to serial correlation in the remainder error or due to state dependence. In order to test for time dependence one has to condition on the individual effects, i.e. test $\Pr[y_{it}/y_{i,t-l}, x_{it}, \mu_i] = \Pr[y_{it}/x_{it}, \mu_i]$. This can be implemented following the work of Lee (1987). In fact, if $\gamma = 0$ is rejected, Hsiao (2003) suggests testing for time dependence against heterogeneity. If heterogeneity is rejected, the model is misspecified. If heterogeneity is not rejected then one estimates the model correcting for heterogeneity. See Heckman (1981c) for an application to married women's employment decisions based on a three-year sample from the PSID. One of the main findings of this study is that neglecting heterogeneity in dynamic models overstates the effect of past experience on labor market participation. Das and van Soest (1999) use the October waves of 1984 till 1989 from the Dutch Socio-Economic Panel to study household subjective expectations about future income changes. Ignoring attrition and sample selection problems which could be serious, the authors estimate a static random effects probit model and a fixed effects conditional logit model as discussed in section 11.1 and extend them to the case of ordered response. Using Heckman's (1981b) procedure, they also estimate a dynamic random effects model which includes a measure of permanent and transitory income. They find that income change expectations depend strongly on realized income changes in the past. In particular, those whose income fell were more pessimistic than others, while those whose income rose were more optimistic. The paper rejects rational expectations finding that households whose income has decreased in the past underestimate their future income growth. Other applications dealing with heterogeneity and state dependence include Heckman and Willis (1977), Heckman and Borjas (1980), Vella and Verbeek (1999) and Hyslop (1999). In marketing research, one can attribute consumers' repeated purchases of the same brands to either state dependence or heterogeneity, see Keane (1997). For a recent application using household-level scanner panel data on six frequently purchased packaged products: ketchup, peanut butter, liquid detergent, tissue, tuna and sugar, see Erdem and Sun (2001). The authors find evidence of state dependence for all product categories except sugar.

Chamberlain's fixed effects conditional logit approach can be generalized to include lags of the dependent variable, provided there are *no* explanatory variables and $T \geq 4$, see Chamberlain (1985). Assuming the initial period y_{i0} is observed but its probability is unspecified, the model is given by

$$\Pr[y_{i0} = 1/\mu_i] = p_0(\mu_i)$$
$$\Pr[y_{it} = 1/\mu_i, y_{i0}, y_{i1}, \ldots, y_{i,t-1}] = \frac{e^{\gamma y_{i,t-1}+\mu_i}}{1+e^{\gamma y_{i,t-1}+\mu_i}} \quad t = 1, \ldots, T \tag{11.16}$$

where $p_0(\mu_i)$ is *unknown* but the logit specification is imposed from period 1 to T. Consider the two events

$$A = \{y_{i0} = d_0, y_{i1} = 0, y_{i2} = 1, y_{i3} = d_3\} \tag{11.17}$$

$$B = \{y_{i0} = d_0, y_{i1} = 1, y_{i2} = 0, y_{i3} = d_3\} \tag{11.18}$$

where d_0 and d_3 are either 0 or 1. If $T = 3$, inference on γ is based upon the fact that $\Pr[A/y_{i1} + y_{i2} = 1, \mu_i]$ and $\Pr[B/y_{i1} + y_{i2} = 1, \mu_i]$ do not depend upon μ_i, see problem 11.2. Honoré and Kyriazidou (2000b) consider the identification and estimation of panel data discrete choice models with lags of the dependent variable and strictly exogenous variables that allow for unobservable heterogeneity. In particular, they extend Chamberlain's (1985) fixed effects logit model in (11.16) to include *strictly exogenous* variables $x_i' = (x_{i1}, \ldots, x_{iT})$, i.e.,

$$\Pr[y_{i0} = 1/x_i', \mu_i] = p_0(x_i', \mu_i) \tag{11.19}$$
$$\Pr[y_{it} = 1/x_i', \mu_i, y_{i0}, \ldots, y_{i,t-1}] = \frac{e^{x_{it}'\beta+\gamma y_{i,t-1}+\mu_i}}{1+e^{x_{it}'\beta+\gamma y_{i,t-1}+\mu_i}} \quad t = 1, \ldots, T$$

The crucial assumption is that the errors in the threshold-crossing model leading to (11.19) are IID over time with logistic distributions and independent of (x_i', μ_i, y_{i0}) at all time periods. Honoré and Kyriazidou (2000b) show that $\Pr(A/x_i', \mu_i, A\cup B)$ and $\Pr(B/x_i', \mu_i, A\cup B)$ will still depend upon μ_i. This means that a conditional likelihood approach will not eliminate the fixed effects. However, if $x_{i2}' = x_{i3}'$, then the conditional probabilities

$$\Pr(A/x_i', \mu_i, A\cup B, x_{i2}' = x_{i3}') = \frac{1}{1+e^{(x_{i1}-x_{i2})'\beta+\gamma(d_0-d_3)}} \tag{11.20}$$

$$\Pr(B/x_i', \mu_i, A\cup B, x_{i2}' = x_{i3}') = \frac{e^{(x_{i1}-x_{i2})'\beta+\gamma(d_0-d_3)}}{1+e^{(x_{i1}-x_{i2})'\beta+\gamma(d_0-d_3)}}$$

do *not* depend on μ_i, see problem 11.3. If all the explanatory variables are discrete and $\Pr[x_{i2}' = x_{i3}'] > 0$, Honoré and Kyriazidou (2000b) suggest maximizing a weighted likelihood function based upon (11.20) for observations that satisfy $x_{i2}' = x_{i3}'$ and $y_{i1} + y_{i2} = 1$. The weakness of this approach is its reliance on observations for which $x_{i2}' = x_{i3}'$ which may not be useful for many economic applications. However, Honoré and Kyriazidou suggest weighing the likelihood function with weights that depend inversely on $x_{i2}' - x_{i3}'$, giving more weight to observations for which x_{i2}' is close to x_{i3}'. This is done using a kernel density $K(x_{i2}' - x_{i3}'/h_N)$where h_N is a bandwidth that shrinks as N increases. The resulting estimators are consistent and asymptotically normal under standard assumptions. However, their rate of convergence will be slower than $\sqrt{N}$ and will depend upon the number of continuous covariates in x_{it}'. The results of a small Monte Carlo study suggest that this estimator performs well and that the asymptotics provide a reasonable approximation to the finite sample behavior of the estimator. Honoré and Kyriazidou also consider identification in the semiparametric case where the logit assumption is relaxed. A conditional maximum score estimator à la Manski (1987) is proposed which is shown to be consistent.[6] In addition, Honoré and Kyriazidou discuss an extension of the identification result to multinomial discrete choice models and to the case where the dependent variable is lagged twice.

Chintagunta, Kyriazidou and Perktold (2001) apply the Honoré and Kyriazidou (2000b) method to study yogurt brand loyalty in South Dakota. They use household panel data with at least two purchases of Yoplait and Nordica yogurt brands over approximately a two-year period. They control for household effects, difference in price and whether the brand was featured in an advertisement that week or displayed in the store. They find that a previous purchase of a brand increases the probability of purchasing that brand in the next period. They also find that if one ignores household heterogeneity, this previous purchase effect is overstated.

Contoyannis, Jones and Rice (2004) utilize seven waves (1991–97) of the British Household Panel Survey (BHPS) to analyze the dynamics of individual health and to decompose the persistence in health outcomes in the BHPS data into components due to state dependence, serial correlation and unobserved heterogeneity. The indicator of health is defined by a binary response to the question: “Does your health in any way limit your daily activities compared to most people of your age?” A sample of 6106 individuals resulting in 42 742 panel observations is used to estimate static and dynamic panel probit models by maximum simulated likelihood using the GHK simulator with antithetic acceleration. The dynamic models show strong positive state dependence.

Hahn (2001) considers two simple dynamic panel logit models with fixed effects for $T = 3$. The first model has only the lagged dependent variable as an explanatory variable. Hahn shows that even though the conditional MLE is $\sqrt{N}$-consistent for this model, its asymptotic variance is strictly larger than the semiparametric asymptotic variance bound. In the second model, time dummies are added and the semiparametric information bound is shown to be singular.

Arellano and Carrasco (2003) consider a binary choice panel data model with predetermined variables. A semiparametric random effects specification is suggested as a compromise to the fixed effects specification that leaves the distribution of the individual effects unrestricted. Dependence is allowed through a nonparametric specification of the conditional expectation of the effects given the predetermined variables. The paper proposes a GMM estimator which is shown to be consistent and asymptotically normal for fixed T and large N. This method is used to estimate a female labor force participation equation with predetermined children using PSID data.

11.4 SELECTION BIAS IN PANEL DATA

In Chapter 9, we studied incomplete panels that had randomly missing data. In section 10.2 we studied rotating panels where, by the design of the survey, households that drop from the sample in one period are intentionally replaced in the next period. However, in many surveys, nonrandomly missing data may occur due to a variety of self-selection rules. One such self-selection rule is the problem of nonresponse of the economic agent. Nonresponse occurs, for example, when the individual refuses to participate in the survey, or refuses to answer particular questions. This problem occurs in cross-section studies, but it becomes aggravated in panel surveys. After all, panel surveys are repeated cross-sectional interviews. So, in addition to the above kinds of nonresponse, one may encounter individuals that refuse to participate in subsequent interviews or simply move or die. Individuals leaving the survey cause *attrition* in the panel. This distorts the random design of the survey and questions the representativeness of the observed sample in drawing inference about the population we are studying. Inference based on the balanced subpanel is inefficient even in randomly missing data since it is throwing away data. In nonrandomly missing data, this inference is misleading because it is no longer representative of the population. Verbeek and Nijman (1996) survey the reasons for nonresponse and distinguish between *ignorable* and *nonignorable* selection rules. This is important because, if the selection rule is ignorable for the parameters of interest, one can use the standard panel data methods for consistent estimation. If the selection rule is nonignorable, then one has to take into account the mechanism that causes the missing observations in order to obtain consistent estimates of the parameters of interest. In order to reduce the effects of attrition, *refreshment samples* are used which replace individuals who dropped from the panel by new individuals randomly sampled from the population. With these refreshment samples, it may be possible to test whether the missing data is *ignorable* or *nonignorable,* see Hirano et al. (2001).

For the one-way error component regression model

$$y_{it} = x'_{it}\beta + \mu_i + \nu_{it} \tag{11.21}$$

where $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$ and $\nu_{it} \sim \text{IIN}(0, \sigma_\nu^2)$ independent of each other and the x_{it}. Observations on y_{it} (and possibly x_{it}) are missing if a selection variable $r_{it} = 0$ and not missing if $r_{it} = 1$. The missing data mechanism is ignorable of order one for β if $E(\mu + \nu_i/r_i) = 0$ for $i = 1, \ldots, N$, where $\mu' = (\mu_1, \ldots, \mu_N)$, $\nu_i' = (\nu_{i1}, \ldots, \nu_{iT})$ and $r_i' = (r_{i1}, \ldots, r_{iT})$. In this case, both GLS on the unbalanced panel and the balanced subpanel are consistent if $N \to \infty$. The Within estimator is consistent for both the unbalanced and balanced subpanel as $N \to \infty$ if $E(\widetilde{\nu}_i/r_i) = 0$ where $\widetilde{\nu}_i' = (\widetilde{\nu}_{i1}, \ldots, \widetilde{\nu}_{iT})$ and $\widetilde{\nu}_{it} = \nu_{it} - \overline{\nu}_i$.[7]

We now consider a simple model of nonresponse in panel data. Following the work of Hausman and Wise (1979), Ridder (1990, 1992) and Verbeek and Nijman (1996), we assume that y_{it} is observed, i.e. $r_{it} = 1$, if a latent variable $r_{it}^* \geq 0$. This latent variable is given by

$$r_{it}^* = z'_{it}\gamma + \epsilon_i + \eta_{it} \tag{11.22}$$

where z_{it} is a set of explanatory variables possibly including some of the x_{it}.[8] The one-way error component structure allows for heterogeneity in the selection process. The errors are assumed to be normally distributed $\epsilon_i \sim \text{IIN}(0, \sigma_\varepsilon^2)$ and $\eta_{it} \sim \text{IIN}(0, \sigma_\eta^2)$ with the only nonzero covariances being $\text{cov}(\epsilon_i, \mu_i) = \sigma_{\mu\epsilon}$ and $\text{cov}(\eta_{it}, \nu_{it}) = \sigma_{\eta\nu}$. In order to get a consistent estimator for β, a generalization of Heckman's (1979) selectivity bias correction procedure from the cross-sectional to the panel data case can be employed. The conditional expectation of u_{it} given selection now involves two terms. Therefore, instead of one selectivity bias correction term, there are now two terms corresponding to the two covariances $\sigma_{\mu\epsilon}$ and $\sigma_{\eta\nu}$. However, unlike the cross-sectional case, these correction terms cannot be computed from simple probit regressions and require numerical integration. Fortunately, this is only a one-dimensional integration problem because of the error component structure. Once the correction terms are estimated, they are included in the regression equation as in the cross-sectional case and OLS or GLS can be run on the resulting augmented model. For details, see Verbeek and Nijman (1996) who also warn about heteroskedasticity and serial correlation in the second step regression if the selection rule is nonignorable. Verbeek and Nijman (1996) also discuss MLE for this random effect probit model with selection bias. The computations require two-dimensional numerical integration for all individuals with $r_{it} = 0$ for at least one t. Verbeek (1990) also considers the estimation of a fixed effects model with selection bias using the marginal maximum likelihood principle. As in the random effects case, the computation is reduced to a two-dimensional numerical integration problem but it is simpler in this case because the two variables over which we are integrating are independent. Verbeek's (1990) model is a hybrid of a fixed individual effect (μ_i) in the behavioral equation, and a random individual effect (ϵ_i) in the selectivity equation. Zabel (1992) argues that if μ_i and x_{it} are correlated then it is highly likely (though not necessary) that ϵ_i and z_{it} are correlated. If the latter is true, the estimates of γ and β will be inconsistent. Zabel suggests modeling the ϵ_i as a function of the $\overline{z}_i$, arguing that this will reduce the inconsistency. Zabel also criticizes Verbeek's specification because it excludes economic models that have the same individual effect in both the behavioral and selectivity equation. For these models both effects should be either fixed or random.

Before one embarks on these complicated estimation procedures one should first test whether the selection rule is ignorable. Verbeek and Nijman (1992a) consider a Lagrange multiplier (LM) test for $H_0 : \sigma_{\nu\eta} = \sigma_{\mu\epsilon} = 0$. The null hypothesis is a sufficient condition for the selection rule to be ignorable for the random effects model. Unfortunately, this also requires numerical integration over a maximum of two dimensions and is cumbersome to use in applied work. In addition, the LM test is highly dependent on the specification of the selectivity equation and the distributional assumptions. Alternatively, Verbeek and Nijman (1992a) suggest some simple Hausman-type tests based on GLS and Within estimators for the unbalanced panel and the balanced subpanel.[9] All four estimators are consistent under the null hypothesis that the selection rule is ignorable and all four estimators are inconsistent under the alternative. This is different from the usual Hausman-type test where one estimator is consistent under both the null and alternative hypotheses, whereas the other estimator is efficient under the null, but inconsistent under the alternative. As a consequence, these tests may have low power, especially if under the alternative these estimators have close asymptotic biases. On the other hand, the advantages of these tests are that they are computationally simple and do not require the specification of a selection rule to derive these tests. Let $\widehat{\delta} = (\widetilde{\beta}_W(B), \widetilde{\beta}_W(U), \widehat{\beta}_{\mathrm{GLS}}(B), \widehat{\beta}_{\mathrm{GLS}}(U))$ where $\widetilde{\beta}_W$ denotes the Within estimator and $\widehat{\beta}_{\mathrm{GLS}}$ denotes the GLS estimator, $\widehat{\beta}(B)$ corresponds to an estimator of β from the balanced subpanel and $\widehat{\beta}(U)$ corresponds to an estimator of β from the unbalanced panel. Verbeek and Nijman (1992a) show that the variance–covariance matrix of $\widehat{\delta}$ is given by

$$\mathrm{var}(\widehat{\delta}) = \begin{bmatrix} V_{11} & V_{22} & V_{33} & V_{44} \\ & V_{22} & V_{22}V_{11}^{-1}V_{13} & V_{44} \\ & & V_{33} & V_{44} \\ & & & V_{44} \end{bmatrix} \tag{11.23}$$

where $V_{11} = \mathrm{var}(\widetilde{\beta}_W(B)$, $V_{22} = \mathrm{var}(\widetilde{\beta}_W(U))$, $V_{33} = \mathrm{var}(\widehat{\beta}_{\mathrm{GLS}}(B)$ and $V_{44} = \mathrm{var}(\widehat{\beta}_{\mathrm{GLS}}(U))$. Therefore an estimate of $\mathrm{var}(\widehat{\delta})$ can be obtained from the estimated variance–covariance matrices of the four estimation procedures. Hausman-type tests can now be performed on say $H_0 : R\delta = 0$, where R is a known matrix, as follows:

$$m = N\widehat{\delta}'R'[R\ \mathrm{var}(\widehat{\delta})R']^{-}R\widehat{\delta} \tag{11.24}$$

and this is distributed as χ^2 under the null with degrees of freedom equal to the rank of $[R\ \mathrm{var}(\widehat{\delta})R']$. Natural candidates for R are $R_1 = [I, 0, -I, 0]$, $R_2 = [0, I, 0, -I]$, $R_3 = [I, -I, 0, 0]$ and $R_4 = [0, 0, I - I]$. The first two are the standard Hausman tests based on the difference between the Within and GLS estimators for the balanced subpanel (R_1) and the unbalanced panel (R_2). The third is based on the difference between the Within estimators from the balanced and unbalanced panels (R_3), while the last is based on the difference between the GLS estimators from the balanced and unbalanced panels (R_4). For all four cases considered, the variance of the difference is the difference between the two variances and hence it is easy to compute. Verbeek and Nijman (1992a) perform some Monte Carlo experiments verifying the poor power of these tests in some cases, but also illustrating their usefulness in other cases. In practice, they recommend performing the tests based on R_2 and R_4.

Wooldridge (1995) derives some simple variable addition tests of selection bias as well as easy-to-apply estimation techniques that correct for selection bias in linear fixed effects panel data models. The auxiliary regressors are either Tobit residuals or inverse Mills ratios and the disturbances are allowed to be arbitrarily serially correlated and unconditionally heteroskedastic. Wooldridge (1995) considers the fixed effects model where the μ_i's are correlated with x_{it}. However, the remainder disturbances ν_{it} are allowed to display arbitrary serial correlation and unconditional heteroskedasticity. The panel is unbalanced with the selection indicator vector for each individual i denoted by $s_i' = (s_{i1}, s_{i2}, \ldots, s_{it})$. When $s_{it} = 1$, it is assumed that (x_{it}', y_{it}) is observed. The fixed effects estimator is given by

$$\widetilde{\beta} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\widetilde{x}_{it}\widetilde{x}_{it}'\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\widetilde{x}_{it}\widetilde{y}_{it}\right) \tag{11.25}$$

where $\widetilde{x}_{it}' = x_{it}' - \left(\sum_{r=1}^{T} s_{ir}x_{ir}'/T_i\right)$, $\widetilde{y}_{it} = y_{it} - \left(\sum_{r=1}^{T} s_{ir}y_{ir}/T_i\right)$ and $T_i = \sum_{i=1}^{T} s_{it}$. A sufficient condition for the fixed estimator to be consistent and asymptotically normal, as $N \to \infty$, is that $E(\nu_{it}/\mu_i,\ x_i', s_i') = 0$ for $t = 1, 2, \ldots, T$. Recall that $x_i' = (x_{i1}', \ldots, x_{iT}')$. Under this assumption, the selection process is strictly exogenous conditional on μ_i and x_i'.

Wooldridge (1995) considers two cases. The first is when the latent variable determining selection is partially observed. Define a latent variable

$$h_{it}^{*} = \delta_{t0} + x_{i1}'\delta_{t1} + \ldots + x_{iT}'\delta_{tT} + \epsilon_{it} \tag{11.26}$$

where ϵ_{it} is independent of (μ_i, x_i'), δ_{tr} is a $K \times 1$ vector of unknown parameters for $r = 1, 2, \ldots, T$ and $\epsilon_{it} \sim N(0, \sigma_t^2)$.

The binary selection indicator is defined as $s_{it} = 1$ if $h_{it}^{*} > 0$. For this case, the censored variable $h_{it} = \max(0, h_{it}^{*})$ is observed. For example, this could be a wage equation, and selection depends on whether or not individuals are working. If a person is working, the working hours h_{it} are recorded, and selection is determined by nonzero hours worked. This is what is meant by partial observability of the selection variable.

Because s_i is a function of (x_i', ϵ_i') where $\epsilon_i' = (\epsilon_{i1}, \ldots, \epsilon_{iT})$, a sufficient condition for the fixed effects estimator to be consistent and asymptotically normal as $N \to \infty$ is $E(\nu_{it}/\mu_i, x_i', \epsilon_i') = 0$ for $t = 1, 2, \ldots, T$. The simplest alternative that implies selectivity bias is $E(\nu_{it}/\mu_i, x_i', \epsilon_i') = E(\nu_{it}/\epsilon_{it}) = \gamma\epsilon_{it}$ for $t = 1, 2, \ldots, T$, with γ being an unknown scalar. Therefore,

$$E(y_{it}/\mu_i, x_i', \epsilon_i', s_i') = E(y_{it}/\mu_i, x_i', \epsilon_i') = \mu_i + x_{it}'\beta + \gamma\epsilon_{it} \tag{11.27}$$

It follows that, if we could observe ϵ_{it} when $s_{it} = 1$, then we could test for selectivity bias by including the ϵ_{it} as an additional regressor in fixed effects estimation and testing $H_0 : \gamma = 0$ using standard methods. While ϵ_{it} cannot be observed, it can be estimated whenever $s_{it} = 1$ because ϵ_{it} is simply the error of a Tobit model.

When h_{it} is observed, Wooldridge's (1995) test for selection bias is as follows.

Step 1. For each $t = 1, 2, \ldots, T$, estimate the equation

$$h_{it} = \max(0, x_i'\delta_t + \epsilon_{it}) \tag{11.28}$$

by standard Tobit, where $\delta_t' = (\delta_{t0}, \delta_{t1}', \ldots, \delta_{tT}')$ and x_i now has unity as its first element. For $s_{it} = 1$, let $\widehat{\epsilon}_{it} = h_{it} - x_i'\widehat{\delta}_t$ denote the Tobit residuals.

Step 2. Estimate the equation

$$\widetilde{y}_{it} = \widetilde{x}_{it}'\beta + \gamma\widetilde{\epsilon}_{it} + \text{residuals} \tag{11.29}$$

by pooled OLS using those observations for which $s_{it} = 1$. $\widetilde{x}_{it}$ and $\widetilde{y}_{it}$ were defined above, and

$$\widetilde{\epsilon}_{it} = \widehat{\epsilon}_{it} - \left(\sum_{r=1}^{T} s_{ir}\widehat{\epsilon}_{ir}/T\right) \tag{11.30}$$

Step 3. Test $H_0 : \gamma = 0$ using the t-statistic for $\widehat{\gamma}$. A serial correlation and heteroskedasticity-robust standard error should be used unless $E[\nu_i\nu_i'/\mu_i, x_i', s_i] = \sigma_\nu^2 I_T$. This robust standard error is given in the appendix to Wooldridge's (1995) paper.

The second case considered by Wooldridge is when h_{it} is not observed. In this case, one conditions on s_i rather than ϵ_i. Using iterated expectations, this gives

$$\begin{aligned} E(y_{it}/\mu_i, x_i', s_i') &= \mu_i + x_{it}'\beta + \gamma E(\epsilon_{it}/\mu_i, x_i', s_i') \\ &= \mu_i + x_{it}'\beta + \gamma E(\epsilon_{it}/x_i', s_i') \end{aligned} \tag{11.31}$$

If the ϵ_{it} were independent across t, then $E(\epsilon_{it}/x_i', s_i') = E(\epsilon_{it}/x_i', s_{it})$. The conditional expectation we need to estimate is $E[\epsilon_{it}/x_i', s_{it} = 1] = E[\epsilon_{it}/x_i', \epsilon_{it} > -x_i'\delta_t]$. Assuming that the var$(\epsilon_{it}) = 1$, we get $E[\epsilon_{it}/x_i', \epsilon_{it} > -x_i'\delta_t] = \lambda(x_i'\delta_t)$ where $\lambda(.)$ denotes the inverse Mills ratio.

When h_{it} is not observed, Wooldridge's (1995) test for selection bias is as follows.

Step 1. For each $t = 1, 2, \ldots, T$, estimate the equation

$$\Pr[s_{it} = 1/x_i'] = \Phi(x_i'\delta_t) \tag{11.32}$$

using standard probit. For $s_{it} = 1$, compute $\widehat{\lambda}_{it} = \lambda(x_i'\widehat{\delta}_t)$.

Step 2. Estimate the equation

$$\widetilde{y}_{it} = \widetilde{x}_{it}'\beta + \gamma\widetilde{\lambda}_{it} + \text{residuals} \tag{11.33}$$

by pooled OLS using those observations for which $s_{it} = 1$. $\widetilde{x}_{it}$ and $\widetilde{y}_{it}$ were defined above, and

$$\widetilde{\lambda}_{it} = \widehat{\lambda}_{it} - \left(\sum_{r=1}^{T} s_{ir}\widehat{\lambda}_{ir}/T_i\right)$$

Step 3. Test $H_0 : \gamma = 0$ using the t-statistic for $\gamma = 0$. Again, a serial correlation and heteroskedasticity-robust standard error is warranted unless

$$E(\nu_i\nu_i'/\mu_i, x_i', s_i) = \sigma^2 I_T \quad \text{under } H_0.$$

Both tests proposed by Wooldridge (1995) are computationally simple, involving variable addition tests. These require either Tobit residuals or inverse Mills ratios obtained from probit estimation for each time period. This is followed by fixed effects estimation.

For the random effects model, Verbeek and Nijman (1992a) suggest including three simple variables in the regression to check for the presence of selection bias. These are (i) the number of waves the ith individual participates in the panel, T_i; (ii) a binary variable taking the value 1 if and only if the ith individual is observed over the entire sample, $\prod_{r=1}^{T} s_{ir}$; and (iii) $s_{i,t-1}$ indicating whether the individual was present in the last period. Intuitively, testing the significance of these variables checks whether the pattern of missing observations affects the underlying regression. Wooldridge (1995) argues that the first two variables have no time variation and cannot be implemented in a fixed effects model. He suggested other variables to be used in place of $\widehat{\lambda}_{it}$ in a variable addition test during fixed effects estimation. These are $\sum_{r\neq t}^{T} s_{ir}$ and $\prod_{r\neq t}^{T} s_{ir}$. Such tests have the computational simplicity advantage and the need to only observe x_{it} when $s_{it} = 1$.[10]

Das (2004) considers a multiperiod random effects panel model with attrition in one period followed by reappearance in a future period. Das suggests a nonparametric two-step estimator that corrects for attrition bias by including estimates of retention probabilities, one for each period that experiences attrition. Asymptotic normality of a class of functionals of this model is derived, including as special cases the linear and partially linear versions of this model.

11.5 CENSORED AND TRUNCATED PANEL DATA MODELS

So far, we have studied economic relationships, say labor supply, based on a random sample of individuals where the dependent variable is 1 if the individual is employed and 0 if the individual is unemployed. However, for these random samples, one may observe the number of hours worked if the individual is employed. This sample is censored in that the hours worked are reported as zero if the individual does not work and the regression model is known as the Tobit model (see Maddala, 1983).[11] Heckman and MaCurdy (1980) consider a fixed effects Tobit model to estimate a lifecycle model of female labor supply. They argue that the individual effects have a specific meaning in a lifecycle model and therefore cannot be assumed independent of the x_{it}. Hence, a fixed effects rather than a random effects specification is appropriate. For this fixed effects Tobit model:

$$y_{it}^* = x_{it}'\beta + \mu_i + \nu_{it} \tag{11.34}$$

with $\nu_{it} \sim \text{IIN}(0, \sigma_\nu^2)$ and

$$\begin{aligned} y_{it} &= y_{it}^* \quad \text{if } y_{it}^* > 0 \\ &= 0 \quad\ \ \text{otherwise} \end{aligned} \tag{11.35}$$

where y_{it} could be the expenditures on a car or a house, or the number of hours worked. This will be zero if the individual does not buy a car or a house or if the individual is unemployed.[12] As in the fixed effects probit model, the μ_i cannot be swept away and as a result β and σ_ν^2 cannot be estimated consistently for T fixed, since the inconsistency in the μ_i is transmitted to β and σ_ν^2. Heckman and MaCurdy (1980) suggest estimating the loglikelihood using iterative methods. Using Monte Carlo experiments with $N = 1000$, $T = 2, 3, 5, 8, 10$ and 20, Greene (2004a) finds that the MLE for the Tobit model with fixed effects exhibits almost no bias even though in each data set in the design, roughly 40–50% of the observations were censored. For the truncated panel data regression model, Greene finds some downward bias in the estimates towards 0. He also finds that the estimated standard deviations are biased downwards in all cases.

Honoré (1992) suggested trimmed least absolute deviations and trimmed least squares estimators for truncated and censored regression models with fixed effects defined in (11.34). These are semiparametric estimators with no distributional assumptions necessary on the error term. The main assumption is that the remainder error ν_{it} is independent and identically distributed conditional on the x_{it} and the μ_i for $t = 1, \ldots, T$. Honoré (1992) exploits the symmetry in the distribution of the latent variables and finds that when the true values of the parameters are known, trimming can transmit the same symmetry in distribution to the observed variables. This generates orthogonality conditions which must hold at the true value of the parameters. Therefore, the resulting GMM estimator is consistent provided the orthogonality conditions are satisfied at a unique point in the parameter space. Honoré (1992) shows that these estimators are consistent and asymptotically normal. Monte Carlo results show that as long as $N \geq 200$, the asymptotic distribution is a good approximation of the small sample distribution. However, if N is small, the small sample distribution of these estimators is skewed.[13] Kang and Lee (2003) use the Korean Household Panel Survey for 1996–98 to study the determinants of private transfers. The results are based on two balanced panels for (1996–97) and (1997–98) containing 3692 and 3674 households, respectively. Applying the Honoré (1992) fixed effects censored estimator, the paper finds that private transfers are altruistically motivated. Also, that there is a strong almost dollar-for-dollar crowding effect of public transfers on private transfers.

Honoré and Kyriazidou (2000a) review recent estimators for censored regression and sample selection panel data models with unobserved individual-specific effects and show how they can easily be extended to other Tobit-type models. The proposed estimators are semiparametric and do not require the parametrization of the distribution of the unobservables. However, they do require that the explanatory variables be strictly exogenous. This rules out lags of the dependent variables among the regressors. The general approach exploits stationarity and exchangeability assumptions on the models' transitory error terms in order to construct moment conditions that do not depend on the individual-specific effects.

Kyriazidou (1997) studies the panel data sample selection model, also known as the Type 2 Tobit model, with

$$y^*_{1it} = x'_{1it}\beta_1 + \mu_{1i} + \nu_{1it} \tag{11.36}$$

$$y^*_{2it} = x'_{2it}\beta_2 + \mu_{2i} + \nu_{2it} \tag{11.37}$$

where

$$\begin{aligned} y_{1it} &= 1 \quad \text{if} \quad y^*_{1it} > 0 \\ &= 0 \quad \text{otherwise} \end{aligned}$$

and

$$\begin{aligned} y_{2it} &= y^*_{2it} \quad \text{if} \quad y_{1it} = 1 \\ &= 0 \qquad \text{otherwise} \end{aligned}$$

Kyriazidou suggests estimating β_1 by one of the estimation methods for discrete choice models with individual effects that were discussed in section 11.1. Next, μ_{2i} is eliminated by first-differencing the data for which y^*_{2it} is observed. With this sample selection, Kyriazidou (1997) focuses on individuals for whom $x'_{1it}\beta_1 = x'_{1is}\beta_1$. For these individuals, the same first-differencing that will eliminate the fixed effects will also eliminate the sample selection. This suggests a two-step Heckman procedure where β_1 is estimated in the first step and then β_2 is estimated by applying OLS to the first differences but giving more weight to observations for which $(x_{1it} - x_{1is})'\widehat{\beta}_1$ is close to zero. This weighting can be done using a kernel whose bandwidth h_N shrinks to zero as the sample size increases. The resulting estimator is $\sqrt{Nh_N}$-consistent and asymptotically normal. Monte Carlo results for $N = 250$, 1000 and 4000 and $T = 2$ indicate that this estimator works well for sufficiently large data sets. However, it is quite sensitive to the choice of the bandwidth parameters.

Charlier, Melenberg and van Soest (2001) apply the methods proposed by Kyriazidou (1997) to a model of expenditure on housing for owners and renters using an endogenous switching regression. The data is based on three waves of the Dutch Socio-Economic Panel from 1987–89. The share of housing in total expenditure is modeled using a household-specific effect, family characteristics, constant quality prices and total expenditure, where the latter is allowed to be endogenous. Estimates from a random effects model are compared to estimates from a linear panel data model in which selection only enters through the fixed effects, and a Kyriazidou-type estimator allowing for fixed effects and a more general type of selectivity. Hausman-type tests reject the random effects and linear panel data models as too restrictive. However, the over-identification restrictions of the more general semiparametric fixed effects model of Kyriazidou (1997) were rejected, suggesting possible misspecification.

Honoré (1993) also considers the dynamic Tobit model with fixed effects, i.e.

$$y_{it}^* = x_{it}'\beta + \lambda y_{i,t-1} + \mu_i + \nu_{it} \tag{11.38}$$

with $y_{it} = \max\{0, y_{it}^*\}$ for $i = 1, \ldots, N$; $t = 1, \ldots, T$. The basic assumption is that ν_{it} is IID$(0, \sigma_\nu^2)$ for $t = 1, \ldots, T$, conditional on y_{i0}, x_{it} and μ_i. Honoré (1993) shows how to trim the observations from a dynamic Tobit model so that the symmetry conditions are preserved for the observed variables at the true values of the parameters. These symmetry restrictions are free of the individual effects and no assumption is needed on the distribution of the μ_i or their relationship with the explanatory variables. These restrictions generate orthogonality conditions which are satisfied at the true value of the parameters. The orthogonality conditions can be used in turn to construct method of moments estimators. Honoré (1993) does not prove that the true values of the parameters are the only values in the parameter space where the orthogonality conditions are satisfied. This means that the resulting GMM estimator is not necessarily consistent. Using Monte Carlo experiments, Honoré (1993) shows that MLE for a dynamic Tobit model with fixed effects performs poorly, whereas the GMM estimator performs quite well, when λ is the only parameter of interest. The assumption that the ν_{it} are IID is too restrictive, especially for a dynamic model. Honoré (1993) relaxes this assumption to the case of stationary ν_{it} for $t = 1, \ldots, T$ conditional on the x_{it} and the μ_i. Still, this assumption is likely to be violated by many interesting economic models. In another paper, Honoré and Hu (2004) provide regularity conditions under which a related set of moment conditions is uniquely satisfied at the true parameter values. They prove that the GMM estimator is consistent and asymptotically normal in this case. Kyriazidou (2001), on the other hand, derives a kernel-weighted GMM estimator for the dynamic sample selection model and shows that this estimator is $\sqrt{Nh_N}$-consistent and asymptotically normal.

Hu (2002) proposes a method for estimating a censored dynamic panel data model with individual fixed effects and lagged *latent* dependent variables. Censoring destroys a certain symmetry between the latent variables. Hu shows that one can artificially truncate the observations in such a way that the symmetry is restored. Based on the restored symmetry, orthogonality conditions are constructed and GMM estimation can be implemented. Although it is hard to prove identification for nonlinear GMM, Hu shows that based on the moment conditions, one can still construct valid asymptotic confidence intervals for the parameters of interest. This is applied to matched data from the 1973 and 1978 March CPS and social security administration earnings records to estimate a dynamic earnings model for a sample of men living in the South during 1957–73, by race. The results suggest that white men's earnings' process appears to be more persistent than that of black men (conditional on individual heterogeneity).

Arellano, Bover and Labeaga (1999) consider a linear autoregressive model for a latent variable which is only partly observed due to a selection mechanism:

$$y_{it}^* = \alpha y_{i,t-1}^* + \mu_i + \nu_{it} \tag{11.39}$$

with $|\alpha| < 1$ and $E(\nu_{it}/y_{i1}^*, \ldots, y_{i,t-1}^*) = 0$. The variable y_{it}^* is observed subject to endogenous selection. Arellano et al. (1999) show that the intractability of this dynamic model subject to censoring using a single time series can be handled successfully using panel data by noting that individuals without censored past observations are exogenously selected. They propose an asymptotic least squares method to estimate features of the distribution of the censored endogenous variable conditional on its past. They apply these methods to a study of female labor supply and wages using two different samples from the PSID covering the periods 1970–76 and 1978–84.

Vella and Verbeek (1999) suggest two-step estimators for a wide range of parametric panel data models with censored endogenous variables and sample selection bias. This generalizes the treatment of sample selection models by Ridder (1990) and Nijman and Verbeek (1992) to a wide range of selection rules. This also generalizes the panel data dummy endogenous regressor model in Vella and Verbeek (1998) by allowing for other forms of censored endogenous regressors. In addition, this analysis shows how Wooldridge's (1995) estimation procedures for sample selection can be applied to more general specifications. The two-step procedure derives estimates of the unobserved heterogeneity responsible for the endogeneity/selection bias in the first step. These in turn are included as additional regressors in the primary equation. This is computationally simple compared to maximum likelihood procedures, since it requires only one-dimensional numerical integration. The panel nature of the data allows adjustment, and testing, for two forms of endogeneity and/or sample selection bias. Furthermore, it allows for dynamics and state dependence in the reduced form. This procedure is applied to the problem of estimating the impact of weekly hours worked on the offered hourly wage rate:

$$\begin{aligned}
w_{it} &= x_{1,it}'\beta_1 + x_{2,it}'\beta_2 + m(\text{hours}_{it}; \beta_3) + \mu_i + \eta_{it} \qquad (11.40)\\
\text{hours}_{it}^* &= x_{3,it}'\theta_1 + \text{hours}_{i,t-1}\theta_2 + \alpha_i + \nu_{it}\\
\text{hours}_{it} &= \text{hours}_{it}^* \qquad \text{if hours}_{it}^* > 0\\
\text{hours}_{it} &= 0 \quad w_{it} \text{ not observed if hours}_{it}^* \leq 0
\end{aligned}$$

Here, w_{it} represents log of the hourly wage for individual i at time t; $x_{1,it}$ and $x_{3,it}$ are variables representing individual characteristics, $x_{2,it}$ are workplace characteristics for individual i; hours_{it}^* and hours_{it} represent desired and observed number of hours worked; m denotes a polynomial of known length with unknown coefficients β_3. This is estimated using data for young females from the NLSY for the period 1980–87. This included a total of 18 400 observations of which 12 039 observations report positive hours of work in a given period.

Lee (2001) proposes a semiparametric first-difference estimator for panel censored selection models where the selection equation is of the Tobit type. This estimator minimizes a convex function and does not require any smoothing. This estimator is compared with that of

Wooldridge (1995) and Honoré and Kyriazidou (2000a) using Monte Carlo experiments. The results show that all three estimators are quite robust to model assumption violation.

11.6 EMPIRICAL APPLICATIONS

There are many empirical applications illustrating the effects of attrition bias; see Hausman and Wise (1979) for a study of the Gary Income Maintenance Experiment. For this experimental panel study of labor supply response, the treatment effect is an income guarantee/tax rate combination. People who benefit from this experiment are more likely to remain in the sample. Therefore, the selection rule is nonignorable, and attrition can overestimate the treatment effect on labor supply. For the Gary Income Maintenance Experiment, Hausman and Wise (1979) found little effect of attrition bias on the experimental labor supply response. Similar results were obtained by Robins and West (1986) for the Seattle and Denver Income Maintenance Experiments. For the latter sample, attrition was modest (11% for married men and 7% for married women and single heads during the period studied) and its effect was not serious enough to warrant extensive correction procedures.

Ridder (1992) studied the determinants of the total number of trips using the first seven waves of the Dutch Transportation Panel (DTP). This panel was commissioned by the Department of Transportation in the Netherlands to evaluate the effect of price increases on the use of public transportation. The first wave of interviews was conducted in March 1984. There is heavy attrition in the DTP, with only 38% of the original sample participating in all seven waves of the panel. Ridder (1992) found that nonrandom attrition from the DTP did not bias time-constant regression coefficients. However, it did bias the time-varying coefficients. Ridder (1992) also found that the restrictions imposed by the standard Hausman and Wise (1979) model for nonrandom attrition on the correlations between individual effects and random shocks may even prevent the detection of nonrandom attrition.

Keane et al. (1988) studied the movement of real wages over the business cycle for panel data drawn from the National Longitudinal Survey of Youth (NLSY) over the period 1966–81. They showed that failure to account for self-selection biased the behavior of real wages in a procyclical direction.

Nijman and Verbeek (1992) studied the effects of nonresponse on the estimates of a simple lifecycle consumption function using a Dutch panel of households interviewed over the period April 1984–March 1987. Several tests for attrition bias were performed, and the model was estimated using (i) one wave of the panel, (ii) the balanced subpanel and (iii) the unbalanced panel. For this application, attrition bias was not serious. The balanced subpanel estimates had implausible signs, while the one-wave estimates and the unbalanced panel estimates gave reasonably close estimates with the latter having lower standard errors.

Ziliak and Kniesner (1998) examine the importance of sample attrition in a lifecycle labor supply using both a Wald test comparing attriters to nonattriters and variable addition tests based on formal models of attrition. Estimates using waves I–XXII of the PSID (interview years 1968–89) show that nonrandom attrition is of little concern when estimating prime age male labor supply because the effect of attrition is absorbed into fixed effects in labor supply.

Dionne, Gagné and Vanasse (1998) estimate a cost model based on an incomplete panel of Ontario trucking firms. The data consists of 445 yearly observations of general freight carriers in Ontario observed over the period 1981–88. It includes 163 firms for which information is available for 2.7 years on average. The cost-input demand system is jointly estimated with a bivariate probit selection model of entry and exit from the sample. A test for selectivity bias reveals potential bias related to exit but not entry from the sample.

Vella and Verbeek (1998) estimate the union premium for young men over a period of declining unionization (1980–87). The panel data is taken from the NLSY and includes 545 full-time working males who completed their schooling by 1980. The probability of union membership is estimated using a dynamic random effects probit model. The coefficient of lagged union status is estimated at 0.61 with a standard error of 0.07, indicating a positive and statistically significant estimate of state dependence. OLS estimates of the wage equation yield a union wage effect of 15–18% depending on whether occupational status dummies are included or not. These estimates are contaminated by endogeneity. The corresponding fixed effects estimates are much lower, yielding 7.9–8.0%. These estimates eliminate only the endogeneity operating through the individual-specific effects. Thus, any time-varying endogeneity continues to contaminate these estimates. Including correction terms based on the estimated union model yields negative significant coefficients and reveals selection bias. This indicates that workers who receive lower wages, after conditioning on their characteristics and in the absence of unions, are most likely to be in the union. This is consistent with the findings that minority groups who are lower paid for discriminatory reasons have a greater tendency to seek union employment than whites. Vella and Verbeek conclude that the union effect is approximately 21% over the period studied. However, the return to unobserved heterogeneity operating through union status is substantial, making the union premium highly variable among individuals. Moreover, this union premium is sensitive to the pattern of sorting into union employment allowed in the estimation.

11.7 EMPIRICAL EXAMPLE: NURSES' LABOR SUPPLY

Shortage of nurses is a problem in several countries. It is an unsettled question whether increasing wages constitute a viable policy for extracting more labor supply from nurses. Askildsen, Baltagi and Holmås (2003) use a unique matched panel data set of Norwegian nurses covering the period 1993–98 to estimate wage elasticities. The data set collected from different official data registers and Statistics Norway includes detailed information on 19 638 individuals over six years totalling 69 122 observations. Female nurses younger than 62 years of age who were registered with a complete nursing qualification and employed by municipalities or counties were included in the sample. For the sample of female nurses considered, the average age was 37 years, with 35% of the nurses being single. The majority of these nurses worked in somatic hospitals (62%) or nursing homes (20%), with the remaining nurses engaged in home nursing (10%), at psychiatric institutions (5%), in health services (1%) and others (3%). Senior nurses comprised only 2% of the sample, while 16% were ward nurses, 20% were nursing specialists and the remaining majority (62%) worked as staff nurses. The average years of experience during the sample period was 12.5 years, and the average number of children below 18 years of age was 1.2. Nurses with children below the age of 3 comprised 22% of the sample, while those with children between the ages of 3 and 7 comprised 29% of the sample.

Verbeek and Nijman (1992a) proposed simple tests for sample selection in panel data models. One test is to include variables measuring whether the individual is observed in the previous period, whether the individual is observed in all periods and the total number of periods the individual is observed, see section 11.4. The null hypothesis says that these variables should not be significant in our model if there are no sample selection problems. Another test, a Hausman-type test, compares the fixed effects estimator from the balanced sample as opposed to an unbalanced sample. Both tests rejected the null hypothesis of no sample selection.

Table 11.1 reproduces the conditional logit model estimates as the first step of the Kyriazidou (1997) estimator. A number of variables were used that characterized the regions

Table 11.1 Participation Equation. Conditional Logit

Educated as nursing specialist	0.6155** (0.0685)
Age	0.1125** (0.034)
Age^2	−0.0035** (0.0004)
Single	−0.1256* (0.055)
Number of children	−0.2640** (0.0457)
Children < 3	−0.1424** (0.0424)
Children 3–7	0.0725 (0.0385)
Children > 7	−0.0619 (0.0369)
Disable	−1.2678** (0.224)
Hospital in municipality	0.6463** (0.0617)
Availability kindergarten	0.3303 (0.2511)
Participation rate	0.0345** (0.0073)
East Norway	0.5275** (0.1198)
South Norway	0.8874** (0.1448)
West Norway	−0.8383** (0.1318)
Mid Norway	1.4610** (0.1429)
Municipality size	−0.0082** (0.0003)
Centrality level 1	−0.0471 (0.1615)
Centrality level 2	−0.5202** (0.1809)
Centrality level 3	−0.5408** (0.1387)
Loglikelihood	−22287.461
Number of observations	61464

Standard errors in parentheses. ** and * is statistically different from zero at 1% and 5% significance level, respectively.
Source: Askildsen et al. (2003). Reproduced by permission of John Wiley & Sons, Ltd.

and municipalities where the individuals live (centrality, female work participation rates, availability of kindergarten and whether there is a hospital in the municipality). These variables were closely connected to the participation decision, and conditional on this are assumed not to affect hours of work. Job-related variables were excluded since they were not observed for those who did not participate. The conditional logit estimates were then used to construct kernel weights with the bandwidth set to $h = 1$. A Hausman test based on the weighted and unweighted estimates gave a value of the test statistic ($\chi^2_{23} = 821.27$) that clearly rejected the null hypothesis of no selection. As instruments for the wage of nurses, the authors used the financial situation of the municipality, measured by lagged net financial surplus in the preceding period. Also, the lagged mean wage of auxiliary nurses working in the same municipality as the nurse, and each nurse's work experience. These variables are assumed to affect wages of nurses but not their hours of work. The instruments pass the Hausman test of over-identifying restrictions. The results of the Kyriazidou instrumental variable estimator are given in Table 11.2. Age had a significant negative effect. Nurses worked shorter hours as they became older but to a diminishing degree. The effect of family variables was as expected. Being single had a positive and significant effect on hours of work. The presence of children in the home had a negative impact on hours of work. Nurses working in psychiatric institutions worked longer hours compared to the base category somatic hospitals, whereas shorter hours were supplied by nurses engaged in home nursing, as well as in nursing homes. Labor supply was highest in the less densely populated Northern Norway (the base category). This may reflect the fact that hours of work were not allowed to vary as much in these areas. Compared to a staff nurse, which served as the base work type category, nursing specialists, ward nurses and senior nurses all worked longer hours. The estimated wage elasticity after controlling for individual heterogeneity, sample selection and instrumenting for possible endogeneity was 0.21. Individual and institutional features were statistically significant and important for working hours. Contractual arrangements as represented by shift work were also important for hours of work, and omitting information about this common phenomenon will underestimate the wage effect.

11.8 FURTHER READING

One should read the related nonlinear panel data model literature, see for example Abrevaya (1999) who proposes a leapfrog estimator for the monotonic transformation panel data model of the type

$$h_t(y_{it}) = x'_{it}\beta + \mu_i + \nu_{it}$$

where $h_t(.)$ is assumed to be strictly increasing. The trick here is to difference across pairs of individuals at a given time period, rather than across time periods. This semiparametric estimator is shown to be $\sqrt{N}$-consistent and asymptotically normal. Examples of this model include the multiple-spell proportional hazards model and dependent variable transformation models with fixed effects. Abrevaya (2000) introduces a class of rank estimators that consistently estimate the coefficients of a generalized fixed effects regression model. This model allows for censoring, places no parametric assumptions on the error disturbances and allows the fixed effects to be correlated with the covariates. The maximum score estimator for the binary choice fixed effects model proposed by Manski (1987) is a member of this class of estimators. The class of rank estimators converge at less than $\sqrt{N}$ rate, while smoothed versions of these estimators converge at rates approaching the $\sqrt{N}$ rate. Some of the limitations of this

Table 11.2 Nurses' Labor Supply

Ln wage	0.2078*
	(0.0942)
Shift work	−0.0111**
	(0.0007)
Shift work 2	−0.00000
	(0.00001)
Hour_35.5	−0.0397**
	(0.0053)
Disable	−0.2581**
	(0.0261)
Age	−0.0098*
	(0.0048)
Age^2	0.0003**
	(0.00003)
Single	0.0205**
	(0.0035)
Number of children	−0.0991**
	(0.0035)
Children $<$ 3	−0.0495**
	(0.0028)
Children 3–7	−0.0177**
	(0.0024)
Children $>$ 7	−0.0307**
	(0.0021)
Psychiatric	0.0466**
	(0.0092)
Home nursing	−0.0206**
	(0.0067)
Health service	−0.0567**
	(0.0148)
Nursing home	−0.0177**
	(0.0059)
Other	0.0024
	(0.0078)
Nursing specialist	0.0144*
	(0.0067)
Ward nurse	−0.0004
	(0.0076)
Senior nurse	0.0057
	(0.0123)
East Norway	−0.0622**
	(0.0131)
South Norway	−0.0802**
	(0.017)

Table 11.2 (*Continued*)

West Norway	−0.1157** (0.0218)
Mid Norway	−0.1011** (0.0188)
Municipality size	0.0002* (0.00007)
Constant	−0.0068** (0.0014)
Number of observations	121 622

Standard errors in parentheses. ** and * is statistically different from zero at 1% and 5% significance level, respectively.
Source: Askildsen et al. (2003). Reproduced by permission of John Wiley & Sons, Ltd.

approach are that the estimation relies on a strict exogeneity assumption and does not deal with the inclusion of lagged dependent variables. Also, the coefficients of time-invariant covariates are not identified by this class of estimators.

Wooldridge (1997) considers the estimation of multiplicative, unobserved components panel data models without imposing a strict exogeneity assumption on the conditioning variables. A robust method of moments estimator is proposed which requires only a conditional mean assumption. This applies to binary choice models with multiplicative unobserved effects, and models containing parametric nonlinear transformations of the endogenous variables. This model is particularly suited to nonnegative explained variables, including count variables. In addition, it can also be applied to certain nonlinear Euler equations. Wooldridge (1999) offers some distribution-free estimators for multiplicative unobserved components panel data models. Requiring only the correct specification of the conditional mean, the multinomial quasi-conditional MLE is shown to be consistent and asymptotically normal. This estimation method is popular for estimating fixed effects count models, see Hausman et al. (1984). Wooldridge's results show that it can be used to obtain consistent estimates even when the dependent variable y_{it} is not a vector of counts. In fact, y_{it} can be a binary response variable, a proportion, a nonnegative continuously distributed random variable, or it can have discrete and continuous characteristics. Neither the distribution of y_{it} nor its temporal dependence are restricted. Additional orthogonality conditions can be used in a GMM framework to improve the efficiency of the estimator. Finally, Wooldridge (2000) proposes a method of estimating very general, nonlinear, dynamic, unobserved effects panel data models with feedback. Wooldridge shows how to construct the likelihood function for the conditional maximum likelihood estimator in dynamic, unobserved effects models where not all conditioning variables are strictly exogenous. A useful innovation is the treatment of the initial conditions which offers a flexible, relatively simple alternative to existing methods.

Hansen (1999) considers the estimation of threshold panel regressions with individual-specific effects. This is useful for situations where the regression function is not identical across all observations in the sample. In fact, the observations are divided into two regimes depending on whether a threshold variable q_{it} is smaller or larger than the threshold γ:

$$y_{it} = \mu_i + \beta_1' x_{it} 1(q_{it} \leq \gamma) + \beta_2' x_{it} 1(q_{it} > \gamma) + \nu_{it}$$

where 1(.) is the indicator function. The regimes are distinguished by differing slopes β_1 and β_2. Hansen (1999) proposes a least squares procedure to estimate the threshold and regression slopes using fixed effects transformations. Nonstandard asymptotic theory with T fixed and $N \to \infty$ is developed to allow the construction of confidence intervals and tests of hypotheses. This method is applied to a panel of 565 US firms observed over the period 1973–87 to test whether financial constraints affect investment decisions. Hansen finds overwhelming evidence of a double threshold effect which separates the firms based on their debt-to-asset ratio. The weakness of this approach is that it does not allow for heteroskedasticity, lagged dependent variables, endogenous variables and random effects.

Hahn and Newey (2004) consider two approaches to reducing the bias from fixed effects estimators in nonlinear models as T gets large. The first is a panel jacknife that uses the variation in the fixed effects estimators as each time period is dropped, one at a time, to form a bias corrected estimator. The second is an analytic bias correction using the bias formula obtained from an asymptotic expansion as T grows. They show that if T grows at the same rate as N, the fixed effects estimator is asymptotically biased, so that the asymptotic confidence intervals are incorrect. However, these are correct for the panel jacknife. If T grows faster than $N^{1/3}$, the analytical bias correction yields an estimator that is asymptotically normal and centered at the truth.

NOTES

1. For the probit model

$$F(x_{it}'\beta) = \Phi(x_{it}'\beta) = \int_{-\infty}^{x_{it}'\beta} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \mathrm{d}u$$

   and for the logit model

$$F(x_{it}'\beta) = \frac{e^{x_{it}'\beta}}{1 + e^{x_{it}'\beta}}$$

2. Note that for this nonlinear panel data model, it is not possible to get rid of the μ_i by taking differences or performing the Within transformation as in Chapter 2.
3. Charlier, Melenberg and van Soest (1995) provide a smoothed maximum score estimator for the binary choice panel data model which has an asymptotic normal distribution but a convergence rate that is slower than root-N. Lee (1999) proposes a $\sqrt{N}$-consistent semiparametric estimator which does not depend on a smoothing parameter and is asymptotically normal.
4. On the other hand, if there are no random individual effects, the joint likelihood will be the product of the marginals and one can proceed as in the usual cross-sectional limited dependent variable case, see Maddala (1983).
5. For good surveys of simulation methods, see Hajivassiliou and Ruud (1994) for limited dependent variable models and Gourieroux and Monfort (1993) with special reference to panel data. The methods surveyed include simulation of the likelihood, simulation of the moment functions and simulation of the score. For the use of the Gibbs sampling method to estimate panel data models, see Chib (1996).
6. In both the logistic and semiparametric case, the main limitations of the Honoré and Kyriazidou (2000b) approach are (i) the assumption that the errors in the underlying threshold-crossing model are independent over time and (ii) the assumption that $x_{i2}' - x_{i3}'$ has support in a neighborhood of 0. The latter restriction rules out time dummies.
7. Other sufficient conditions for consistency of these estimators are given by Verbeek and Nijman (1996). These are derived for specific selection rules. One interesting and practical sufficient condition that emerges is that the Within estimator is consistent and free of selectivity bias if the probability

of being in the sample is constant over time. In this case, the correction for selectivity bias is time-invariant and hence is absorbed in the individual effect term.

8. If the selection rule is unknown, identification problems arise regarding the parameters of interest (see Verbeek and Nijman, 1996). Also, for a more comprehensive analysis of the attrition problem in panel data studies with an arbitrary number of waves, see Ridder (1990).
9. Verbeek and Nijman (1992a) show that under nonresponse, the conditions for consistency of the Within estimator are weaker than those for the random effects GLS estimator. This means that the Within estimator is more robust to nonresponse bias than GLS.
10. It is important to point out that both Verbeek and Nijman (1992a) as well as Wooldridge (1995) assume that the unobservable effects and the idiosyncratic errors in the selection process are normally distributed. Kyriazidou's (1997) treatment of sample selection leaves the distributions of all unobservables unspecified.
11. Alternatively, one could condition on the set of continuously working individuals, i.e. use only the sample with positive hours of work. In this case the sample is considered truncated (see Maddala, 1983).
12. Researchers may also be interested in panel data economic relationships where the dependent variable is a count of some individual actions or events. For example, the number of patents filed, the number of drugs introduced, the number of hospital visits or the number of jobs held. These models can be estimated using Poisson panel data regressions (see Hausman, Hall and Griliches, 1984) and the monographs on count data by Cameron and Trivedi (1998) and Winkelmann (2000).
13. For the special case of only one regressor and two panels ($T = 2$), Campbell and Honoré (1993) show that the semiparametric estimator derived by Honoré (1992) is median unbiased in finite samples under (basically) the same conditions that are used to derive its asymptotic distribution.

PROBLEMS

11.1 In section 11.1 we considered the fixed effects logit model with $T = 2$.

(a) In this problem, we look at $T = 3$ and ask the reader to compute the conditional probabilities that would get rid of the individual effects by conditioning on $\sum_{t=1}^{3} y_{it}$. Note that this sum can now be 0, 1, 2 or 3. (*Hint*: First show that terms in the conditional likelihood function, which are conditioned upon $\sum_{t=1}^{3} y_{it} = 0$ or 3 add nothing to the likelihood. Then focus on terms that condition on $\sum_{t=1}^{3} y_{it} = 1$ or 2.)

(b) Show that for $T = 10$, one has to condition on the sum being $1, 2, \ldots, 9$. One can see that the number of probability computations is increasing. To convince yourself, write down the probabilities conditioning on $\sum_{t=1}^{10} y_{it} = 1$.

11.2 Consider the Chamberlain (1985) fixed effects conditional logit model with a lagged dependent variable given in (11.16). Show that for $T = 3$, $\Pr[A/y_{il} + y_{i2} = 1, \mu_i]$ and therefore $\Pr[B/y_{i1} + y_{i2} = 1, \mu_i]$ do not depend on μ_i. Note that A and B are defined in (11.17) and (11.18), respectively.

11.3 Consider the Honoré and Kyriazidou (2000b) fixed effects logit model given in (11.19).

(a) Show that for $T = 3$, $\Pr[A/x_i', \mu_i, A \cup B]$ and $\Pr[B/x_i', \mu_i, A \cup B]$ both depend on μ_i. This means that the conditional likelihood approach will *not* eliminate the fixed effect μ_i.

(b) If $x_{i2}' = x_{i3}'$, show that $\Pr[A/x_i', \mu_i, A \cup B, x_{i2}' = x_{i3}']$ and $\Pr[B/x_i', \mu_i, A \cup B, x_{i2}' = x_{i3}']$ do *not* depend on μ_i.

11.4 *Fixed effects logit model*. This is based on Abrevaya (1997). Consider the fixed effects logit model given in (11.4) with $T = 2$. In (11.10) and (11.11) we showed the conditional maximum likelihood of β, call it $\widehat{\beta}_{\text{CML}}$, can be obtained by running a logit estimator of the dependent variable $1(\Delta y = 1)$ on the independent variables Δx for the subsample of observations satisfying $y_{i1} + y_{i2} = 1$. Here $1(\Delta y = 1)$ is an indicator function taking the value one if $\Delta y = 1$. Therefore, $\widehat{\beta}_{\text{CML}}$ maximizes the loglikelihood

$$\ln L_c(\beta) = \sum_{i\epsilon\vartheta} [1(\Delta y = 1)\ln F(\Delta x\beta) + 1(\Delta y = -1)\ln(1 - F(\Delta x\beta))]$$

where $\vartheta = \{i : y_{i1} + y_{i2} = 1\}$.

(a) Maximize the unconditional loglikelihood for (11.4) given by

$$\ln L(\beta, \mu_i) = \sum_{i=1}^{N}\sum_{t=1}^{2}[y_{it}\ln F(x_{it}'\beta + \mu_i) + (1 - y_{it})\ln(1 - F(x_{it}'\beta + \mu_i))]$$

with respect to μ_i and show that i

$$\widehat{\mu}_i = \begin{cases} -\infty & \text{if} \quad y_{i1} + y_{i2} = 0 \\ -(x_{i1} + x_{i2})'\beta/2 & \text{if} \quad y_{i1} + y_{i2} = 1 \\ +\infty & \text{if} \quad y_{i1} + y_{i2} = 2 \end{cases}$$

(b) Concentrate the likelihood by plugging $\widehat{\mu}_i$ in the unconditional likelihood and show that

$$\ln L(\beta, \widehat{\mu}_i) = \sum_{i\epsilon\vartheta} 2[1(\Delta y = 1)\ln F(\Delta x'\beta/2) + 1(\Delta y = -1)\ln(1 - F(\Delta x'\beta/2))]$$

Hint: Use the symmetry of F and the fact that

$$1(\Delta y = 1) = y_{i2} = 1 - y_{i1} \quad \text{and} \quad 1(\Delta y = -1) = y_{i1} = 1 - y_{i2} \quad \text{for } i\epsilon\vartheta.$$

(c) Conclude that $\ln L(\beta, \widehat{\mu}_i) = 2\ln L_c(\beta/2)$. This shows that a scale adjusted maximum likelihood estimator is equivalent to the conditional maximum likelihood estimator, i.e., $\widehat{\beta}_{\text{ML}} = 2\widehat{\beta}_{\text{CML}}$. Whether a similar result holds for $T > 2$ remains an open question.

11.5 *Binary response model regression* (BRMR). This is based on problem 95.5.4 in *Econometric Theory* by Baltagi (1995). Davidson and MacKinnon (1993) derive an artificial regression for testing hypotheses in a binary response model. For the fixed effects model described in (11.4), the reader is asked to derive the BRMR to test $H_o : \mu_i = 0$, for $i = 1, 2, \ldots, N$. Show that if $F(.)$ is the logistic (or normal) cumulative distribution function, this BRMR is simply a weighted least squares regression of logit (or probit) residuals, ignoring the fixed effects, on the matrix of regressors X and the matrix of individual dummies. The test statistic in this case is the explained sum of squares from this BRMR. See solution 95.5.4 in *Econometric Theory* by Gurmu (1996).

11.6 Using the Vella and Verbeek (1998) data set posted on the *Journal of Applied Econometrics* web site:

(a) Replicate their descriptive statistics given in table I and confirm that the unconditional union premium is around 15%.

(b) Replicate their random effects probit estimates of union membership given in table II.

(c) Replicate the wage regressions with union effects given in table III.

(d) Replicate the wage regressions under unrestricted sorting given in table V.

12 Nonstationary Panels

12.1 INTRODUCTION

With the growing use of cross-country data over time to study purchasing power parity, growth convergence and international R&D spillovers, the focus of panel data econometrics has shifted towards studying the asymptotics of macro panels with large N (number of countries) and large T (length of the time series) rather than the usual asymptotics of micro panels with large N and small T. The limiting distribution of double indexed integrated processes has been studied extensively by Phillips and Moon (1999, 2000). The fact that T is allowed to increase to infinity in macro panel data generated two strands of ideas. The first rejected the homogeneity of the regression parameters implicit in the use of a pooled regression model in favor of heterogeneous regressions, i.e., one for each country, see Pesaran and Smith (1995), Im, Pesaran and Shin (2003), Lee, Pesaran and Smith (1997), Pesaran, Shin and Smith (1999) and Pesaran and Zhao (1999), to mention a few. This literature critically relies on T being large to estimate each country's regression separately. This literature warns against the use of standard pooled estimators such as FE to estimate the dynamic panel data model, arguing that they are subject to large potential bias when the parameters are heterogeneous across countries and the regressors are serially correlated. Another strand of literature applied time series procedures to panels, worrying about nonstationarity, spurious regressions and cointegration.[1] Consider, for example, the Penn World Tables which have been used to study growth convergence among various countries, see /www.nber.org/. Phillips and Moon (2000) argue that the time series components of the variables used in these tables, like per capita GDP growth, have strong nonstationarity, a feature which we have paid no attention to in the previous chapters. This is understandable given that micro panels deal with large N and small T. With large N, large T macro panels, nonstationarity deserves more attention. In particular, time series fully modified estimation techniques that account for endogeneity of the regressors and correlation and heteroskedasticity of the residuals can now be combined with fixed and random effects panel estimation methods. Some of the distinctive results that are obtained with nonstationary panels are that many test statistics and estimators of interest have normal limiting distributions. This is in contrast to the nonstationary time series literature where the limiting distributions are complicated functionals of Weiner processes. Several unit root tests applied in the time series literature have been extended to panel data. When the panel data are both heterogeneous and nonstationary, issues of combining individual unit root tests applied on each time series are tackled by Im et al. (2003), Maddala and Wu (1999) and Choi (2001). Using panel data, one can avoid the problem of spurious regression, see Kao (1999) and Phillips and Moon (1999). Unlike the single time series spurious regression literature, the panel data spurious regression estimates give a consistent estimate of the true value of the parameter as both N and T tend to ∞. This is because, the panel estimator averages across individuals and the information in the independent cross-section data in the panel leads to a stronger overall signal than the pure time series case. Of course, letting both N and T tend to ∞ brings in a new host of issues dealing with how to do asymptotic analysis. This is studied by Phillips and Moon (1999, 2000) and Kauppi (2000).

One can find numerous applications of time series methods applied to panels in recent years, especially panel unit root tests, panel cointegration tests and the estimation of long-run average relations. Examples from the purchasing power parity literature and real exchange rate stationarity include Frankel and Rose (1996), Jorion and Sweeney (1996), MacDonald (1996), Oh (1996), Wu (1996), Coakley and Fuertes (1997), Papell (1997), O'Connell (1998), Groen and Kleibergen (2003), Choi (2001), Canzoneri, Cumby and Diba (1999), Groen (2000), Pedroni (2001) and Smith et al. (2004), to mention a few. On interest rates, see Wu and Chen (2001); on real wage stationarity, see Fleissig and Strauss (1997). On the inflation rate, see Culver and Papell (1997); on the current account balance, see Wu (2000); on the consumption–income ratio stationarity, see Sarantis and Stewart (1999). On health care expenditures, see McCoskey and Selden (1998); on growth and convergence, see Islam (1995), Bernard and Jones (1996), Evans and Karras (1996), Sala-i-Martin (1996), Lee et al. (1997) and Nerlove (2000). On international R&D spillovers, see Kao, Chiang and Chen (1999). On savings and investment models, see Coakely, Kulasi and Smith (1996) and Moon and Phillips (1999).

However, the use of such panel data methods is not without their critics, see Maddala et al. (2000) who argue that panel data unit root tests do not rescue purchasing power parity (PPP). In fact, the results on PPP with panels are mixed depending on the group of countries studied, the period of study and the type of unit root test used. More damaging is the argument by Maddala et al. (2000) that for PPP, panel data tests are the wrong answer to the low power of unit root tests in single time series. After all, the null hypothesis of a single unit root is different from the null hypothesis of a panel unit root for the PPP hypothesis. Using the same line of criticism, Maddala (1999) argued that panel unit root tests did not help settle the question of growth convergence among countries. However, it was useful in spurring much needed research into dynamic panel data models. Also, Quah (1996) argued that the basic issues of whether poor countries catch up with the rich can never be answered by the use of traditional panels. Instead, Quah suggested formulating and estimating models of income dynamics. Recently, Smith (2000) warned about the mechanical application of panel unit root or cointegration tests, arguing that the application of these tests requires that the hypotheses involved be interesting in the context of the substantive application. The latter is a question of theory rather than statistical technique. See also Banerjee et al. (2004) in the empirical section of this chapter, who criticize existing panel unit root tests for assuming that cross-unit cointegrating relationships among the countries are not present. They warn that the empirical size of these tests is substantially higher than their nominal size. Panel unit root tests have also been criticized because they assume cross-sectional independence. This is restrictive as macro time series exhibit significant cross-sectional correlation among the countries in the panel. This correlation has been modeled using a dynamic factor model by Bai and Ng (2004), Bai (2004), Moon and Perron (2004a,b), Moon, Perron and Phillips (2003) and Phillips and Sul (2003). Alternative panel unit root tests that account for cross-section dependence include Choi (2002), Chang (2002, 2004) and Pesaran (2003).

Early surveys on nonstationary panels include Phillips and Moon (2000) on multi-indexed processes, Banerjee (1999), Baltagi and Kao (2000) and Smith (2000) on panel unit roots and cointegration tests. This chapter studies panel unit root tests assuming cross-sectional independence in section 12.2, while section 12.3 discusses panel unit root tests allowing for cross-sectional dependence. Section 12.4 studies the spurious regression in panel models, while section 12.5 considers various panel cointegration tests. Section 12.6 discusses estimation and inference in panel cointegration models, while section 12.7 illustrates the panel unit root tests using a purchasing power parity example. Section 12.8 gives some additional readings.

12.2 PANEL UNIT ROOTS TESTS ASSUMING CROSS-SECTIONAL INDEPENDENCE

Testing for unit roots in time series studies is now common practice among applied researchers and has become an integral part of econometric courses. However, testing for unit roots in panels is recent, see Levin, Lin and Chu (2002), Im et al. (2003), Harris and Tzavalis (1999), Maddala and Wu (1999), Choi (2001) and Hadri (2000).[2] Exceptions are Bhargava et al. (1982), Boumahdi and Thomas (1991), Breitung and Meyer (1994) and Quah (1994). Bhargava et al. (1982) proposed a test for random walk residuals in a dynamic model with fixed effects. They suggested a modified Durbin–Watson (DW) statistic based on fixed effects residuals and two other test statistics based on differenced OLS residuals. In typical micro panels with $N \to \infty$, they recommended their modified DW statistic. Boumahdi and Thomas (1991) proposed a generalization of the Dickey–Fuller (DF) test for unit roots in panel data to assess the efficiency of the French capital market using 140 French stock prices over the period January 1973 to February 1986. Breitung and Meyer (1994) applied various modified DF test statistics to test for unit roots in a panel of contracted wages negotiated at the firm and industry level for Western Germany over the period 1972–87. Quah (1994) suggested a test for unit root in a panel data model without fixed effects where both N and T go to infinity at the same rate such that N/T is constant. Levin et al. (2002), hereafter LLC, generalized this model to allow for fixed effects, individual deterministic trends and heterogeneous serially correlated errors. They assumed that both N and T tend to infinity. However, T increases at a faster rate than N with $N/T \to 0$. Even though this literature grew from time series and panel data, the way in which N, the number of cross-section units, and T, the length of the time series, tend to infinity is crucial for determining asymptotic properties of estimators and tests proposed for nonstationary panels, see Phillips and Moon (1999). Several approaches are possible, including (i) sequential limits where one index, say N, is fixed and T is allowed to increase to infinity, giving an intermediate limit. Then by letting N tend to infinity subsequently, a sequential limit theory is obtained. Phillips and Moon (2000) argued that these sequential limits are easy to derive and are helpful in extracting quick asymptotics. However, Phillips and Moon provided a simple example that illustrates how sequential limits can sometimes give misleading asymptotic results. (ii) A second approach, used by Quah (1994) and Levin et al. (2002), is to allow the two indexes, N and T, to pass to infinity along a specific diagonal path in the two-dimensional array. This path can be determined by a monotonically increasing functional relation of the type $T = T(N)$ which applies as the index $N \to \infty$. Phillips and Moon (2000) showed that the limit theory obtained by this approach is dependent on the specific functional relation $T = T(N)$ and the assumed expansion path may not provide an appropriate approximation for a given (T, N) situation. (iii) A third approach is a joint limit theory allowing both N and T to pass to infinity simultaneously without placing specific diagonal path restrictions on the divergence. Some control over the relative rate of expansion may have to be exercised in order to get definitive results. Phillips and Moon argued that, in general, joint limit theory is more robust than either sequential limit or diagonal path limit. However, it is usually more difficult to derive and requires stronger conditions such as the existence of higher moments that will allow for uniformity in the convergence arguments. The multi-index asymptotic theory in Phillips and Moon (1999, 2000) is applied to joint limits in which $T, N \to \infty$ and $(T/N) \to \infty$, i.e., to situations where the time series sample is large relative to the cross-section sample. However, the general approach given there is also applicable to situations in which $(T/N) \to 0$, although different limit results will generally obtain in that case.

12.2.1 Levin, Lin and Chu Test

LLC argued that individual unit root tests have limited power against alternative hypotheses with highly persistent deviations from equilibrium. This is particularly severe in small samples. LLC suggest a more powerful panel unit root test than performing individual unit root tests for each cross-section. The null hypothesis is that each individual time series contains a unit root against the alternative that each time series is stationary.

The maintained hypothesis is that

$$\Delta y_{it} = \rho y_{i,t-1} + \sum_{L=1}^{p_i} \theta_{iL} \Delta y_{it-L} + \alpha_{mi} d_{mt} + \varepsilon_{it} \quad m = 1, 2, 3 \tag{12.1}$$

with d_{mt} indicating the vector of deterministic variables and α_{mi} the corresponding vector of coefficients for model $m = 1, 2, 3$. In particular, $d_{1t} = \{\text{empty set}\}$, $d_{2t} = \{1\}$ and $d_{3t} = \{1, t\}$. Since the lag order p_i is unknown, LLC suggest a three-step procedure to implement their test.

Step 1. Perform separate augmented Dickey–Fuller (ADF) regressions for each cross-section:

$$\Delta y_{it} = \rho_i y_{i,t-1} + \sum_{L=1}^{p_i} \theta_{iL} \Delta y_{i,t-L} + \alpha_{mi} d_{mt} + \varepsilon_{it} \quad m = 1, 2, 3 \tag{12.2}$$

The lag order p_i is permitted to vary across individuals.

For a given T, choose a maximum lag order $p_{\max}$ and then use the t-statistic of $\widehat{\theta}_{iL}$ to determine if a smaller lag order is preferred. (These t-statistics are distributed $N(0, 1)$ under the null hypothesis ($\theta_{iL} = 0$), both when $\rho_i = 0$ and when $\rho_i < 0$.)

Once p_i is determined, two auxiliary regressions are run to get *orthogonalized residuals*:

Run	Δy_{it}	on	$\Delta y_{i,t-L} (L = 1, \ldots, p_i)$	and	d_{mt}	to get residuals	$\widehat{e}_{it}$
Run	$y_{i,t-1}$	on	$\Delta y_{i,t-L} (L = 1, \ldots, p_i)$	and	d_{mt}	to get residuals	$\widehat{\nu}_{i,t-1}$

Standardize these residuals to control for different variances across i

$$\widetilde{e}_{it} = \widehat{e}_{it} / \widehat{\sigma}_{\varepsilon i} \quad \text{and} \quad \widetilde{\nu}_{i,t-1} = \widehat{\nu}_{it} / \widehat{\sigma}_{\varepsilon i}$$

where $\widehat{\sigma}_{\varepsilon i}$ = standard error from each ADF regression, for $i = 1, \ldots, N$.

Step 2. Estimate the ratio of long-run to short-run standard deviations. Under the null hypothesis of a unit root, the long-run variance of (12.1) can be estimated by

$$\widehat{\sigma}_{yi}^2 = \frac{1}{T-1} \sum_{t=2}^{T} \Delta y_{it}^2 + 2 \sum_{L=1}^{\bar{K}} w_{\bar{K}L} \left[\frac{1}{T-1} \sum_{t=2+L}^{T} \Delta y_{it} \Delta y_{i,t-L} \right] \tag{12.3}$$

where $\bar{K}$ is a truncation lag that can be data-dependent. $\bar{K}$ must be obtained in a manner that ensures the consistency of $\widehat{\sigma}_{yi}^2$. For a Bartlett kernel, $w_{\bar{K}L} = 1 - (L/(\bar{K} + 1))$. For each cross-section i, the ratio of the long-run standard deviation to the innovation standard deviation is estimated by $\widehat{s}_i = \widehat{\sigma}_{yi} / \widehat{\sigma}_{\varepsilon i}$. The average standard deviation is estimated by $\widehat{S}_N = \frac{1}{N} \sum_{i=1}^{N} \widehat{s}_i$.

Step 3. Compute the panel test statistics. Run the pooled regression

$$\widetilde{e}_{it} = \rho \widetilde{\nu}_{i,t-1} + \widetilde{\varepsilon}_{it}$$

based on $N\widetilde{T}$ observations where $\widetilde{T} = T - \bar{p} - 1$. $\widetilde{T}$ is the average number of observations per individual in the panel with $\bar{p} = \sum_{i=1}^{N} p_i/N$. $\bar{p}$ is the average lag order of individual ADF regressions. The conventional t-statistic for $H_0\colon \rho = 0$ is $t_\rho = \frac{\widehat{\rho}}{\widehat{\sigma}(\widehat{\rho})}$, where

$$\widehat{\rho} = \sum_{i=1}^{N} \sum_{t=2+p_i}^{T} \widetilde{v}_{i,t-1}\widetilde{e}_{it} \Bigg/ \sum_{i=1}^{N} \sum_{t=2+p_i}^{T} \widetilde{v}_{i,t-1}^2$$

$$\widehat{\sigma}(\widehat{\rho}) = \widehat{\sigma}_{\widetilde{\varepsilon}} \Bigg/ \left[\sum_{i=1}^{N} \sum_{t=2+p_i}^{T_i} \widetilde{v}_{i,t-1}^2\right]^{1/2}$$

and

$$\widehat{\sigma}_{\widetilde{\varepsilon}}^2 = \frac{1}{N\widetilde{T}} \sum_{i=1}^{N} \sum_{t=2+p_i}^{T} (\widetilde{e}_{it} - \widehat{\rho}\widetilde{v}_{i,t-1})^2$$

is the estimated variance of $\widetilde{\varepsilon}_{it}$.

Compute the adjusted t-statistic

$$t_\rho^* = \frac{t_\rho - N\widetilde{T}\widehat{S}_N\widehat{\sigma}_{\widetilde{\varepsilon}}^{-2}\widehat{\sigma}(\widehat{\rho})\mu_{m\widetilde{T}}^*}{\sigma_{m\widetilde{T}}^*} \tag{12.4}$$

where $\mu_{m\widetilde{T}}^*$ and $\sigma_{m\widetilde{T}}^*$ are the mean and standard deviation adjustments provided by table 2 of LLC. This table also includes suggestions for the truncation lag parameter $\bar{K}$ for each time series $\widetilde{T}$. LLC show that t_ρ^* is asymptotically distributed as $N(0, 1)$.

The asymptotics require $\sqrt{N_T}/T \to 0$ where N_T emphasizes that the cross-sectional dimension N is an arbitrary monotonically increasing function of T. LLC argue that this is relevant for micro panel data where T is allowed to grow slower than N_T. Other divergence speeds such as $N_T/T \to 0$ and $N_T/T \to$ constant are sufficient, but not necessary.

Computationally, the LLC method requires a specification of the *number of lags* used in each cross-section ADF regression (p_i), as well as *kernel choices* used in the computation of S_N. In addition, you must specify the exogenous variables used in the test equations. You may elect to include no exogenous regressors, or to include individual constant terms (fixed effects), or to employ constants and trends.

LLC suggest using their panel unit root test for panels of moderate size with N between 10 and 250 and T between 25 and 250. They argue that the standard panel procedures may not be computationally feasible or sufficiently powerful for panels of this size. However, for very large T, they argue that individual unit root time series tests will be sufficiently powerful to apply for each cross-section. Also, for very large N and very small T, they recommend the usual panel data procedures. The Monte Carlo simulations performed by LLC indicate that the normal distribution provides a good approximation to the empirical distribution of the test statistic, even in relatively small samples. Also, that the panel unit root test provides dramatic improvements in power over separate unit root tests for each cross-section.

The proposed LLC test has its limitations. The test crucially depends upon the *independence* assumption across cross-sections and is not applicable if cross-sectional correlation is present. Second, the assumption that *all* cross-sections have or do not have a unit root is restrictive.

Harris and Tzavalis (1999) also derived unit root tests for (12.1) with $d_{mt} = \{\text{empty set}\}, \{1\}$ or $\{1, t\}$ when the time dimension of the panel T is *fixed*. This is the typical case for micro panel studies. The main results are:

$$\begin{array}{ll} d_{mt} & \widehat{\rho} \\ \{\text{empty set}\} & \sqrt{N}\,(\widehat{\rho} - 1) \Rightarrow N\left(0, \frac{2}{T(T-1)}\right) \\ \{1\} & \sqrt{N}\left(\widehat{\rho} - 1 + \frac{3}{T+1}\right) \Rightarrow N\left(0, \frac{3(17T^2-20T+17)}{5(T-1)(T+1)^3}\right) \\ \{1, t\} & \sqrt{N}\left(\widehat{\rho} - 1 + \frac{15}{2(T+2)}\right) \Rightarrow N\left(0, \frac{15(193T^2-728T+1147)}{112(T+2)^3(T-2)}\right) \end{array} \tag{12.5}$$

Harris and Tzavalis (1999) also showed that the assumption that T tends to infinity at a faster rate than N, as in Levin and Lin, rather than T fixed, as in the case of micro panels, yields tests which are substantially undersized and have low power especially when T is small.

Frankel and Rose (1996), Oh (1996) and Lothian (1996) tested the PPP hypothesis using panel data. All of these articles use Levin and Lin tests and some of them report evidence supporting the PPP hypothesis. O'Connell (1998), however, showed that the Levin and Lin tests suffered from significant size distortion in the presence of correlation among contemporaneous cross-sectional error terms. O'Connell highlighted the importance of controlling for cross-sectional dependence when testing for a unit root in panels of real exchange rates. He showed that, controlling for cross-sectional dependence, no evidence against the null of a random walk can be found in panels of up to 64 real exchange rates.

Most of the recent literature on panel unit root testing deals with cross-sectional dependence and this will be surveyed later on in this chapter.

12.2.2 Im, Pesaran and Shin Test

The Levin, Lin and Chu test is restrictive in the sense that it requires ρ to be homogeneous across i. As Maddala (1999) pointed out, the null may be fine for testing convergence in growth among countries, but the alternative restricts every country to converge at the same rate. Im et al. (2003) (IPS) allow for a heterogeneous coefficient of y_{it-1} and propose an alternative testing procedure based on averaging individual unit root test statistics. IPS suggest an average of the ADF tests when u_{it} is serially correlated with different serial correlation properties across cross-sectional units, i.e., the model given in (12.2). The null hypothesis is that each series in the panel contains a unit root, i.e., $H_0\colon \rho_i = 0$ for all i and the alternative hypothesis allows for some (but not all) of the individual series to have unit roots, i.e.,

$$H_1\colon \begin{cases} \rho_i < 0 & \text{for} \quad i = 1, 2, \ldots, N_1 \\ \rho_i = 0 & \text{for} \quad i = N_1 + 1, \ldots, N \end{cases} \tag{12.6}$$

Formally, it requires the fraction of the individual time series that are stationary to be nonzero, i.e., $\lim_{N\to\infty}(N_1/N) = \delta$ where $0 < \delta \leq 1$. This condition is necessary for the consistency of the panel unit root test. The IPS t-bar statistic is defined as the average of the individual ADF statistics as

$$\bar{t} = \frac{1}{N}\sum_{i=1}^{N} t_{\rho_i} \tag{12.7}$$

where t_{ρ_i} is the individual t-statistic for testing $H_0\colon \rho_i = 0$ for all i in (12.6). In case the lag order is always zero ($p_i = 0$ for all i), IPS provide simulated critical values for $\bar{t}$ for different number of cross-sections N, series length T and Dickey–Fuller regressions containing intercepts only or intercepts and linear trends. In the general case where the lag order p_i may be nonzero for some cross-sections, IPS show that a properly standardized $\bar{t}$ has an asymptotic $N(0, 1)$ distribution. Starting from the well-known result in time series that for a fixed N

$$t_{\rho_i} \Rightarrow \frac{\int_0^1 W_{iZ} \mathrm{d}W_{iZ}}{\left[\int_0^1 W_{iZ}^2\right]^{1/2}} = t_{iT} \tag{12.8}$$

as $T \to \infty$, where $\int W(r)\,\mathrm{d}r$ denotes a Weiner integral with the argument r suppressed in (12.8), IPS assume that t_{iT} are IID and have finite mean and variance. Then

$$\frac{\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} t_{iT} - \frac{1}{N}\sum_{i=1}^{N} E\left[t_{iT}|\rho_i = 0\right]\right)}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} \operatorname{var}\left[t_{iT}|\rho_i = 0\right]}} \Rightarrow N(0, 1) \tag{12.9}$$

as $N \to \infty$ by the Lindeberg–Levy central limit theorem. Hence

$$t_{\mathrm{IPS}} = \frac{\sqrt{N}\left(\bar{t} - \frac{1}{N}\sum_{i=1}^{N} E\left[t_{iT}|\rho_i = 0\right]\right)}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} \operatorname{var}\left[t_{iT}|\rho_i = 0\right]}} \Rightarrow N(0, 1) \tag{12.10}$$

as $T \to \infty$ followed by $N \to \infty$ sequentially. The values of $E\left[t_{iT}|\rho_i = 0\right]$ and $\operatorname{var}[t_{iT}|\rho_i = 0]$ have been computed by IPS via simulations for different values of T and p_i's. In Monte Carlo experiments, they show that if a large enough lag order is selected for the underlying ADF regressions, then the small sample performance of the t-bar test is reasonably satisfactory and generally better than the LLC test.

12.2.3 Breitung's Test

The LLC and IPS tests require $N \to \infty$ such that $N/T \to 0$, i.e., N should be small enough relative to T. This means that both tests may not keep nominal size well when either N is small or N is large relative to T. In fact, the simulation results of Im et al. (2003) show that both IPS and LLC have size distortions as N gets large relative to T. Breitung (2000) studies the local power of LLC and IPS test statistics against a sequence of local alternatives. Breitung finds that the LLC and IPS tests suffer from a dramatic loss of power if individual-specific trends are included. This is due to the bias correction that also removes the mean under the sequence of local alternatives. Breitung suggests a test statistic that does not employ a bias adjustment whose power is substantially higher than that of LLC or the IPS tests using Monte Carlo experiments. The simulation results indicate that the power of LLC and IPS tests is very sensitive to the specification of the deterministic terms.

Breitung's (2000) test statistic without bias adjustment is obtained as follows. Step 1 is the same as LLC but only $\Delta y_{i,t-L}$ is used in obtaining the residuals $\widehat{e}_{it}$ and $\widehat{\nu}_{i,t-1}$. The residuals are then adjusted (as in LLC) to correct for individual-specific variances. Step 2, the residuals $\widehat{e}_{it}$ are transformed using the forward orthogonalization transformation employed by Arellano and Bover (1995):

$$e_{it}^{*} = \sqrt{\frac{T-t}{(T-t+1)}}\left(\widetilde{e}_{it} - \frac{\widetilde{e}_{i,t+1} + \ldots + \widetilde{e}_{i,T}}{T-t}\right)$$

Also,

$$\begin{aligned}
\nu^*_{i,t-1} &= \widetilde{\nu}_{i,t-1} - \widetilde{\nu}_{i,1} - \frac{t-1}{T}\widetilde{\nu}_{iT} \text{ with intercept and trend} \\
&= \widetilde{\nu}_{i,t-1} - \widetilde{\nu}_{i,1} \text{ with intercept, no trend} \\
&= \widetilde{\nu}_{i,t-1} \text{ with no intercept or trend}
\end{aligned}$$

The last step is to run the pooled regression

$$e^*_{it} = \rho \nu^*_{i,t-1} + \varepsilon^*_{it}$$

and obtain the t-statistic for $H_0 : \rho = 0$ which has in the limit a standard $N(0, 1)$ distribution. Note that no kernel computations are required.

McCoskey and Selden (1998) applied the IPS test for testing unit root for per capita national health care expenditures (HE) and gross domestic product (GDP) for a panel of 20 OECD countries. McCoskey and Selden rejected the null hypothesis that these two series contain unit roots. Gerdtham and Löthgren (2000) claimed that the stationarity found by McCoskey and Selden is driven by the omission of time trends in their ADF regression in (12.6). Using the IPS test with a time trend, Gerdtham and Löthgren found that both HE and GDP are nonstationary. They concluded that HE and GDP are cointegrated around linear trends. Wu (2000) applied the IPS test to the current account balances for 10 OECD countries over the period 1977Q1 to 1997Q4. Current account balances measure changes in national net indebtedness. Persistent deficits could have serious effects. Wu does not reject current account stationarity, which in turn is consistent with the sustainability of external debts among the industrial countries considered.

12.2.4 Combining p-Value Tests

Let G_{iT_i} be a unit root test statistic for the ith group in (12.1) and assume that as the time series observations for the ith group $T_i \to \infty$, $G_{iT_i} \Rightarrow G_i$ where G_i is a nondegenerate random variable. Let p_i be the asymptotic p-value of a unit root test for cross-section i, i.e., $p_i = F\left(G_{iT_i}\right)$, where $F(\cdot)$ is the distribution function of the random variable G_i. Maddala and Wu (1999) and Choi (2001) proposed a Fisher-type test

$$P = -2\sum_{i=1}^{N} \ln p_i \tag{12.11}$$

which combines the p-values from unit root tests for each cross-section i to test for unit root in panel data. Note that $-2 \ln p_i$ has a χ^2 distribution with 2 degrees of freedom. This means that P is distributed as χ^2 with $2N$ degrees of freedom as $T_i \to \infty$ for finite N. Maddala and Wu (1999) argued that the IPS and Fisher tests relax the restrictive assumption of the LLC test that ρ_i is the same under the alternative. Both the IPS and Fisher tests combine information based on individual unit root tests. However, the Fisher test has the advantage over the IPS test in that it does not require a balanced panel. Also, the Fisher test can use different lag lengths in the individual ADF regressions and can be applied to any other unit root tests. The disadvantage is that the p-values have to be derived by Monte Carlo simulations. Maddala and Wu (1999) find that the Fisher test with bootstrap-based critical values performs the best and is the preferred choice for testing nonstationary as the null and also in testing for cointegration in panels.

Choi (2001) proposes two other test statistics besides Fisher's inverse chi-square test statistic P. The first is the inverse normal test $Z = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Phi^{-1}(p_i)$, where Φ is the standard normal cumulative distribution function. Since $0 \le p_i \le 1$, $\Phi^{-1}(p_i)$ is a $N(0, 1)$ random variable and as $T_i \to \infty$ for all i, $Z \Rightarrow N(0, 1)$. The second is the logit test $L = \sum_{i=1}^{N} \ln(\frac{p_i}{1-p_i})$ where $\ln(\frac{p_i}{1-p_i})$ has the logistic distribution with mean 0 and variance $\pi^2/3$. As $T_i \to \infty$ for all i, $\sqrt{m}L \Rightarrow t_{5N+4}$ where $m = \frac{3(5N+4)}{\pi^2 N(5N+2)}$. Choi (2001) echoes similar advantages for these three combining p-value tests: (1) the cross-sectional dimension, N, can be either finite or infinite; (2) each group can have different types of nonstochastic and stochastic components; (3) the time series dimension, T, can be different for each i; and (4) the alternative hypothesis would allow some groups to have unit roots while others may not.

When N is large, Choi (2001) proposed a modified P test,

$$P_m = \frac{1}{2\sqrt{N}} \sum_{i=1}^{N} (-2 \ln p_i - 2) \tag{12.12}$$

since $E[-2 \ln p_i] = 2$ and $\text{var}[-2 \ln p_i] = 4$. Applying the Lindeberg–Lévy central limit theorem to (12.12) we get $P_m \Rightarrow N(0, 1)$ as $T_i \to \infty$ followed by $N \to \infty$.[3] The distribution of the Z statistic is invariant to infinite N, and $Z \Rightarrow N(0, 1)$ as $T_i \to \infty$ and then $N \to \infty$. Also, the distribution of $\sqrt{m}L \approx \frac{1}{\sqrt{\pi^2 N/3}} \sum_{i=1}^{N} \ln(\frac{p_i}{1-p_i}) \Rightarrow N(0, 1)$ by the Lindeberg–Lévy central limit theorem as $T_i \to \infty$ and then $N \to \infty$. Therefore, Z and $\sqrt{m}L$ can be used without modification for infinite N. Simulation results for $N = 5, 10, 25, 50$ and 100, and $T = 50$ and 100 show that the empirical size of all the tests is reasonably close to the 0.05 nominal size when N is small. P and P_m show mild size distortions at $N = 100$, while Z and IPS show the most stable size. All tests become more powerful as N increases. The combined p-value tests have superior size-adjusted power to the IPS test. In fact, the power of the Z test is in some cases more than three times that of the IPS test. Overall, the Z test seems to outperform the other tests and is recommended.

Choi (2001) applied the combining p-value tests and the IPS test given in (12.7) to panel data of monthly US real exchange rates sampled from 1973:3 to 1996:3. The combining p-value tests provided evidence in favor of the PPP hypothesis, while the IPS test did not. Choi claimed that this is due to the improved finite sample power of the combination tests. Maddala and Wu (1999) and Maddala et al. (2000) find that the Fisher test is superior to the IPS test, which in turn is more powerful than the LLC test. They argue that these panel unit root tests still do not rescue the PPP hypothesis. When allowance is made for the deficiency in the panel data unit root tests and panel estimation methods, support for PPP turns out to be weak.

Choi (2002) considers four instrumental variable estimators of an error component model with stationary and nearly nonstationary regressors. The remainder disturbances follow an autoregressive process whose order as well as parameters vary across individuals. The IV estimators considered include the Within–IV, Within–IV–OLS, Within–IV–GLS and IV–GLS estimators. Using sequential and joint limit theories, Choi shows that, under proper conditions, all the estimators have normal distributions in the limit as N and $T \to \infty$. Simulation results show that the efficiency rankings of the estimators depend crucially on the type of regressor and the number of instruments. The Wald tests for coefficient restrictions keep reasonable nominal size as $N \to \infty$ and its power depends upon the number of instruments and the degree of serial correlation and heterogeneity in the errors.

12.2.5 Residual-Based LM Test

Hadri (2000) derives a residual-based Lagrange multiplier (LM) test where the null hypothesis is that there is no unit root in any of the series in the panel against the alternative of a unit root in the panel. This is a generalization of the KPSS test from time series to panel data. It is based on OLS residuals of y_{it} on a constant, or on a constant and a trend. In particular, Hadri (2000) considers the following two models:

$$y_{it} = r_{it} + \varepsilon_{it} \quad i = 1, \ldots, N; \quad t = 1, \ldots, T$$

and

$$y_{it} = r_{it} + \beta_i t + \varepsilon_{it} \tag{12.13}$$

where $r_{it} = r_{i,t-1} + u_{it}$ is a random walk. $\varepsilon_{it} \sim \text{IIN}(0, \sigma_\varepsilon^2)$ and $u_{it} \sim \text{IIN}(0, \sigma_\mu^2)$ are mutually independent normals that are IID across i and over t. Using back substitution, model (12.13) becomes

$$y_{it} = r_{io} + \beta_i t + \sum_{s=1}^{t} u_{is} + \varepsilon_{it} = r_{io} + \beta_i t + \nu_{it} \tag{12.14}$$

where $\nu_{it} = \sum_{s=1}^{t} u_{is} + \varepsilon_{it}$. The stationarity hypothesis is simply $H_0 : \sigma_u^2 = 0$, in which case $\nu_{it} = \varepsilon_{it}$. The LM statistic is given by

$$\text{LM}_1 = \frac{1}{N}\left(\sum_{i=1}^{N} \frac{1}{T^2} \sum_{t=1}^{T} S_{it}^2\right) / \widehat{\sigma}_\varepsilon^2$$

where $S_{it} = \sum_{s=1}^{t} \widehat{\varepsilon}_{is}$ are the partial sum of OLS residuals $\widehat{\varepsilon}_{is}$ from (12.14) and $\widehat{\sigma}_\varepsilon^2$ is a consistent estimate of σ_ε^2 under the null hypothesis H_0. A possible candidate is $\widehat{\sigma}_\varepsilon^2 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \widehat{\varepsilon}_{it}^2$.

Hadri (2000) suggested an alternative LM test that allows for heteroskedasticity across i, say $\sigma_{\varepsilon i}^2$. This is in fact

$$\text{LM}_2 = \frac{1}{N}\left(\sum_{i=1}^{N}\left(\frac{1}{T^2} \sum_{t=1}^{T} S_{it}^2 / \widehat{\sigma}_{\varepsilon i}^2\right)\right)$$

The test statistic is given by $Z = \sqrt{N}(LM - \xi_1)/\zeta$ and is asymptotically distributed as $N(0, 1)$, where $\xi = \frac{1}{6}$ and $\zeta = \frac{1}{45}$ if the model only includes a constant, and $\xi = \frac{1}{15}$ and $\zeta = \frac{11}{6300}$, otherwise. EViews computes both test statistics. Hadri (2000) shows, using Monte Carlo experiments, that the empirical size of the test is close to its nominal 5% level for sufficiently large N and T.

Yin and Wu (2000) propose stationarity tests for a heterogeneous panel data model. The authors consider the case of serially correlated errors in the level and trend stationary models. The proposed panel tests utilize the Kwiatkowski et al. (1992) test and the Leybourne and McCabe (1994) test from the time series literature. Two different ways of pooling information from the independent tests are used. In particular, the group mean and the Fisher-type tests are used to develop the panel stationarity tests. Monte Carlo experiments are performed that reveal good small sample performance in terms of size and power.

Extensive simulations have been conducted to explore the finite sample performance of panel unit root tests. Choi (2001), for example, studied the small sample properties of the IPS t-bar test in (12.7) and Fisher's test in (12.11). Choi's major findings were the following.

(1) The empirical sizes of the IPS and the Fisher test are reasonably close to their nominal size 0.05 when N is small. But the Fisher test shows mild size distortions at $N = 100$, which is expected from the asymptotic theory. Overall, the IPS t-bar test has the most stable size.
(2) In terms of the size-adjusted power, the Fisher test seems to be superior to the IPS t-bar test.
(3) When a linear time trend is included in the model, the power of all tests decreases considerably.

Karlsson and Löthgren (2000) compare the LLC and IPS tests for various size panels. They warn that for large T, panel unit root tests have high power and there is the potential risk of concluding that the *whole* panel is stationary even when there is only a *small proportion* of stationary series in the panel. For small T, panel unit root tests have low power and there is the potential risk of concluding that the whole panel is nonstationary even when there is a large proportion of stationary series in the panel. They suggest careful analysis of both the individual and panel unit root test results to fully assess the stationarity properties of the panel.

12.3 PANEL UNIT ROOTS TESTS ALLOWING FOR CROSS-SECTIONAL DEPENDENCE

Pesaran (2004) suggests a simple test of error cross-section dependence (CD) that is applicable to a variety of panel models including stationary and unit root dynamic heterogeneous panels with short T and large N. The proposed test is based on an average of pairwise correlation coefficients of OLS residuals from the individual regressions in the panel rather than their squares as in the Breusch–Pagan LM test:

$$\mathrm{CD} = \sqrt{\frac{2T}{N(N-1)}}\left(\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\widehat{\rho}_{ij}\right)$$

where $\widehat{\rho}_{ij} = \sum_{t=1}^{T} e_{it}e_{jt}/(\sum_{t=1}^{T} e_{it}^2)^{1/2}(\sum_{t=1}^{T} e_{jt}^2)^{1/2}$, with e_{it} denoting the OLS residuals based on T observations for each $i = 1, \ldots, N$. Monte Carlo experiments show that the standard Breusch–Pagan LM test performs badly for $N > T$ panels, whereas Pesaran's CD test performs well even for small T and large N.

Dynamic factor models have been used to capture cross-section correlation.[3] Moon and Perron (2004c) consider the following model:

$$\begin{aligned} y_{it} &= \alpha_i + y_{it}^0 \\ y_{it}^0 &= \rho_i y_{i,t-1}^0 + \epsilon_{it} \end{aligned}$$

where ϵ_{it} are unobservable error terms with a factor structure and α_i are fixed effects. ϵ_{it} is generated by M unobservable random factors f_t and idiosyncratic shocks e_{it} as follows:

$$\epsilon_{it} = \Lambda_i' f_t + e_{it}$$

where Λ_i are nonrandom factor loading coefficient vectors and the number of factors M is unknown. Each ϵ_{it} contains the common random factor f_t, generating the correlation among the cross-sectional units of ϵ_{it} and y_{it}. The extent of the correlation is determined by the factor loading coefficients Λ_i, i.e., $E(y_{it}y_{jt}) = \Lambda_i' E(f_t f_t')\Lambda_j$. Moon and Perron treat the factors as nuisance parameters and suggest pooling defactored data to construct a unit root test. Let Q_Λ

be the matrix projecting onto the space orthogonal to the factor loadings. The defactored data is YQ_Λ and the defactored residuals eQ_Λ no longer have cross-sectional dependence, where Y is a $T \times N$ matrix whose ith column contains the observations for cross-sectional unit i.

Let $\sigma_{e,i}^2$ be the variance of e_{it}, $w_{e,i}^2$ be the long-run variance of e_{it} and $\lambda_{e,i}$ be the one-sided long run variance of e_{it}. Also, σ_e^2, w_e^2 and λ_e be their cross-sectional averages, and ϕ_e^4 be the cross-sectional average of $w_{e,i}^4$. The pooled bias-correlated estimate of ρ is

$$\widehat{\rho}_{pool}^{+} = \frac{\mathrm{tr}(Y_{-1}Q_\Lambda Y') - NT\lambda_e^N}{\mathrm{tr}(Y_{-1}Q_\Lambda Y_{-1}')}$$

where Y_{-1} is the matrix of lagged data. Moon and Perron suggest two statistics to test $H_0: \rho_i = 1$ for all $i = 1, \ldots, M$ against the alternative hypothesis $H_A: \rho_i < 1$ for some i. These are

$$t_a = \frac{\sqrt{N}T(\widehat{\rho}_{pool}^{+} - 1)}{\sqrt{\frac{2\phi_e^4}{w_e^4}}}$$

and

$$t_b = \sqrt{N}T(\widehat{\rho}_{pool}^{+} - 1)\sqrt{\frac{1}{NT^2}\,\mathrm{tr}\,(Y_{-1}Q_\Lambda Y_{-1}')\frac{w_e^2}{\phi_e^4}}$$

These tests have a standard $N(0, 1)$ limiting distribution where N and T tend to infinity such that $N/T \to 0$. Moon and Perron also show that estimating the factors by principal components and replacing w_e^2 and ϕ_e^4 by consistent estimates leads to feasible statistics with the same limiting distribution.

Phillips and Sul (2003) consider the following common time factor model on the disturbances that can impact individual series differently:

$$u_{it} = \delta_i\theta_t + \varepsilon_{it}$$

where $\theta_t \sim \mathrm{IIN}(0, 1)$ across time. δ_i are "idiosyncratic share" parameters that measure the impact of the common time effects on series i. $\varepsilon_{it} \sim \mathrm{IIN}(0, \sigma_i^2)$ over t, with ε_{it} independent of ε_{js} and θ_s for all $i \neq j$ and for all s, t. This model is in effect a one-factor model which is independently distributed over time. $E(u_{it}u_{js}) = \delta_i\delta_j$ and there is no cross-sectional correlation if $\delta_i = 0$ for all i, and identical cross-section correlation when $\delta_i = \delta_j = \delta_0$ for all i, j. Phillips and Sul propose an orthogonalization procedure based on iterated method of moments estimation to eliminate the common factor which is different from principal components. They suggest a series of unit root tests based on these orthogonalized data. The statistic that performs best in their simulation is a combination of p-values of individual unit root tests as in Choi (2001), i.e., $Z = \frac{1}{\sqrt{N}}\sum_{i=1}^{N-1}\Phi^{-1}(p_i)$. The sum is over $N - 1$ components, since the orthogonalization they propose reduces the cross-sectional dimension by 1. The null hypothesis is rejected for large values of the Z statistic.

Bai and Ng (2004) consider the following dynamic factor model:

$$\begin{aligned} y_{it} &= \alpha_i + \Lambda_i' f_t + y_{it}^0 \\ y_{it}^0 &= \rho_i y_{i,t-1}^0 + \varepsilon_{it} \end{aligned}$$

They test separately the stationarity of the factors and the idiosyncratic component. To do so, they obtain consistent estimates of the factors regardless of whether residuals are stationary or not. They accomplish this by estimating factors on first-differenced data and cumulating these estimated factors. Bai and Ng suggest pooling results from individual ADF tests on the estimated defactored data by combining p-values as in Maddala and Wu (1999) and Choi (2001):

$$P_{\widehat{e}}^{c} = \frac{-2\sum_{i=1}^{N} \ln p_{\widehat{e}}^{c}(i) - 2N}{\sqrt{4N}} \xrightarrow{d} N(0, 1)$$

where $p_{\widehat{e}}^{c}(i)$ is the p-value of the ADF test (without any deterministic component) on the estimated idiosyncratic shock for cross-section i.

Choi (2002) uses an error component model given by

$$\begin{aligned} y_{it} &= \alpha_i + f_t + y_{it}^0 \\ y_{it}^0 &= \rho_i y_{i,t-1}^0 + \varepsilon_{it} \end{aligned}$$

This is a restricted factor model where the cross-sections respond homogeneously to the single common factor f_t in contrast to the factor models considered above. Choi suggests demeaning the data by GLS as in Elliott, Rothenberg and Stock (1996) and taking cross-sectional means to obtain a new variable $\widetilde{y}_{it} \simeq y_{it}^0 - y_{i1}^0$ which is independent in the cross-sectional dimension as both N and T tend to infinity. Choi combines p-values from individual ADF tests as in Choi (2001). The resulting tests have a standard $N(0, 1)$ distribution. In addition, Choi suggests using an ADF test for the hypothesis that the common component f_t is nonstationary. To do so, he proposes using the cross-sectional means (at each t) of the residuals from the GLS regression used to demean the data, i.e.,

$$\widehat{f}_t = \frac{1}{N}\sum_{i=1}^{N}(y_{it} - \widehat{\alpha}_i)$$

Pesaran (2003) suggests a simpler way of geting rid of cross-sectional dependence than estimating the factor loading. His method is based on augmenting the usual ADF regression with the lagged cross-sectional mean and its first difference to capture the cross-sectional dependence that arises through a single factor model. This is called the cross-sectionally augmented Dickey–Fuller (CADF) test. This simple CADF regression is

$$\Delta y_{it} = \alpha_i + \rho_i^* y_{i,t-1} + d_0\overline{y}_{t-1} + d_1\Delta\overline{y}_t + \varepsilon_{it}$$

where $\overline{y}_t$ is the average at time t of all N observations. The presence of the lagged cross-sectional average and its first difference accounts for the cross-sectional dependence through a factor structure. If there is serial correlation in the error term or the factor, the regression must be augmented as usual in the univariate case, but lagged first-differences of both y_{it} and $\overline{y}_t$ must be added, which leads to

$$\Delta y_{it} = \alpha_i + \rho_i^* y_{i,t-1} + d_0\overline{y}_{t-1} + \sum_{j=0}^{p} d_{j+1}\Delta\overline{y}_{t-j} + \sum_{k=1}^{p} c_k\Delta y_{i,t-k} + \varepsilon_{it}$$

where the degree of augmentation can be chosen by an information criterion or sequential testing. After running this CADF regression for each unit i in the panel, Pesaran averages the

t-statistics on the lagged value (called CADF_i) to obtain the CIPS statistic

$$\text{CIPS} = \frac{1}{N}\sum_{i=1}^{N}\text{CADF}_i$$

The joint asymptotic limit of the CIPS statistic is nonstandard and critical values are provided for various choices of N and T. The t-tests based on this regression should be devoid of $\Lambda_i' f_t$ in the limit and therefore free of cross-sectional dependence. The limiting distribution of these tests is different from the Dickey–Fuller distribution due to the presence of the cross-sectional average of the lagged level. Pesaran uses a truncated version of the IPS test that avoids the problem of moment calculation. In addition, the t-tests are used to formulate a combination test based on the inverse normal principle. Experimental results show that these tests perform well.

12.4 SPURIOUS REGRESSION IN PANEL DATA

Entorf (1997) studied spurious fixed effects regressions when the true model involves independent random walks with and without drifts. Entorf found that for $T \to \infty$ and N finite, the nonsense regression phenomenon holds for spurious fixed effects models and inference based on t-values can be highly misleading. Kao (1999) and Phillips and Moon (1999) derived the asymptotic distributions of the least squares dummy variable estimator and various conventional statistics from the spurious regression in panel data.

Suppose that y_t and X_t are unit root nonstationary time series variables with long-run variance matrix

$$\Omega = \begin{pmatrix} \Omega_{yy} & \Omega_{yx} \\ \Omega_{xy} & \Omega_{xx} \end{pmatrix}$$

Then $\beta = \Omega_{yx}\Omega_{xx}^{-1}$ can be interpreted as a classical long-run regression coefficient relating the two nonstationary variables y_t and X_t. When Ω has deficient rank, β is a cointegrating coefficient because $y_t - \beta X_t$ is stationary. Even in the absence of time series cointegration, β is a measure of a statistical long-run correlation between y_t and X_t. Phillips and Moon (1999) extend this concept to panel regressions with nonstationary data. In this case, heterogeneity across individuals i can be characterized by heterogeneous long-run covariance matrices Ω_i. Then Ω_i are randomly drawn from a population with mean $\Omega = E(\Omega_i)$. In this case

$$\beta = E(\Omega_{y_i x_i})E(\Omega_{x_i x_i})^{-1} = \Omega_{yx}\Omega_{xx}^{-1}$$

is the regression coefficient corresponding to the average long-run covariance matrix Ω.

Phillips and Moon (1999) studied various regressions between two panel vectors that may or may not have cointegrating relations, and present a fundamental framework for studying sequential and joint limit theories in nonstationary panel data. The panel models considered allow for four cases: (i) panel spurious regression, where there is no time series cointegration; (ii) heterogeneous panel cointegration, where each individual has its own specific cointegration relation; (iii) homogeneous panel cointegration, where individuals have the same cointegration relation; and (iv) near-homogeneous panel cointegration, where individuals have slightly different cointegration relations determined by the value of a localizing parameter. Phillips and Moon (1999) investigated these four models and developed panel asymptotics for regression coefficients and tests using both sequential and joint limit arguments. In all cases considered the pooled estimator is consistent and has a normal limiting distribution. In fact, for the spurious panel regression, Phillips and Moon (1999) showed that under quite weak regularity conditions, the pooled least squares estimator of the slope coefficient β is $\sqrt{N}$-consistent for the long-run average relation parameter β and has a limiting normal distribution. Also, Moon and Phillips (1998) showed that a limiting cross-section regression with time-averaged data is also $\sqrt{N}$-consistent for β and has a limiting normal distribution. This is different from the pure time series spurious regression where the limit of the OLS estimator of β is a nondegenerate random variate that is a functional of Brownian motions and is therefore not consistent for β. The idea in Phillips and Moon (1999) is that independent cross-section data in the panel adds information and this leads to a stronger overall signal than the pure time series case. Pesaran and Smith (1995) studied limiting cross-section regressions with time-averaged data and established consistency with restrictive assumptions on the heterogeneous panel model. This differs from Phillips and Moon (1999) in that the former use an average of the cointegrating coefficients which is different from the long-run average regression coefficient. This requires the existence of cointegrating time series relations, whereas the long-run average regression coefficient β is defined irrespective of the existence of individual cointegrating relations and relies only on the long-run average variance matrix of the panel. Phillips and Moon (1999) also showed that for the homogeneous and near homogeneous cointegration cases, a consistent estimator of the long-run regression coefficient can be constructed which they call a pooled FM estimator. They showed that this estimator has a faster convergence rate than the simple cross-section and time series estimators. Pedroni (2000) and Kao and Chiang (2000) also investigated limit theories for various estimators of the homogeneous panel cointegration regression model. See also Phillips and Moon (2000) for a concise review. In fact, the latter paper also shows how to extend the above ideas to models with individual effects in the data generating process. For the panel spurious regression with individual-specific deterministic trends, estimates of the trend coefficients are obtained in the first step and the detrended data is pooled and used in least squares regression to estimate β in the second step. Two different detrending procedures are used based on OLS and GLS regressions. OLS detrending leads to an asymptotically more efficient estimator of the long-run average coefficient β in pooled regression than GLS detrending. Phillips and Moon (2000) explain that "the residuals after time series GLS detrending have more cross-section variation than they do after OLS detrending and this produces great variation in the limit distribution of the pooled regression estimator of the long run average coefficient".

Moon and Phillips (1999) investigate the asymptotic properties of the Gaussian MLE of the localizing parameter in local to unity dynamic panel regression models with deterministic and stochastic trends. Moon and Phillips find that for the homogeneous trend model, the Gaussian MLE of the common localizing parameter is $\sqrt{N}$-consistent, while for the heterogeneous trends model, it is inconsistent. The latter inconsistency is due to the presence of an infinite number of incidental parameters (as $N \to \infty$) for the individual trends. Unlike the fixed effects dynamic panel data model where this inconsistency due to the incidental parameter problem disappears as $T \to \infty$, the inconsistency of the localizing parameter in the Moon and Phillips model persists even when both N and T go to infinity. Moon and Phillips (2000) show that the local to unity parameter in a simple panel near-integrated regression model can be estimated consistently using pooled OLS. When deterministic trends are present, pooled panel estimators of the localizing parameter are asymptotically biased. Some techniques are developed to obtain consistent estimates of this localizing parameter but only in the region where it is negative. These methods are used to show how to perform efficient trend extraction for panel data. They are also used to deliver consistent estimates of distancing parameters in nonstationary panel models where the initial conditions are in the distant past. The joint asymptotics in the paper rely on $N/T \to 0$, so that the results are most relevant in panels where T is large relative to N.

Consider the nonstationary dynamic panel data model

$$\begin{aligned} y_{it} &= \alpha_{i,0} + \alpha_{i,1}t + y_{it}^0 \\ y_{it}^0 &= \beta y_{i,t-1}^0 + u_{i,t} \end{aligned}$$

with $\beta = \exp(c/T)$. Moon and Phillips (2000) focused on estimating the localizing parameter c in β which characterizes the local behavior of the unit root process. Information about c is useful for the analysis of the power properties of unit root tests, cointegration tests, the construction of confidence intervals for the long-run autoregressive coefficient, the development of efficient detrending methods and the construction of point optimal invariant tests for a unit root and cointegrating rank. Moon and Phillips (2000) show that when $c \leq 0$, it is possible to estimate this local parameter consistently using panel data. In turn, they show how to extract the deterministic trend efficiently using this consistent estimate of c.

12.5 PANEL COINTEGRATION TESTS

Like the panel unit root tests, panel cointegration tests can be motivated by the search for more powerful tests than those obtained by applying individual time series cointegration tests. The latter tests are known to have low power, especially for short T and short span of the data, which is often limited to post-war annual data. In the case of purchasing power parity and convergence in growth, economists pool data on similar countries, like G7, OECD or Euro countries in the hopes of adding cross-sectional variation to the data that will increase the power of unit root tests or panel cointegration tests.

12.5.1 Residual-Based DF and ADF Tests (Kao Tests)

Consider the panel regression model

$$y_{it} = x_{it}'\beta + z_{it}'\gamma + e_{it} \tag{12.15}$$

where y_{it} and x_{it} are $I(1)$ and noncointegrated. For $z_{it} = \{\mu_i\}$, Kao (1999) proposed DF and ADF-type unit root tests for e_{it} as a test for the null of no cointegration. The DF-type tests can be calculated from the fixed effects residuals

$$\widehat{e}_{it} = \rho\widehat{e}_{it-1} + \nu_{it} \tag{12.16}$$

where $\widehat{e}_{it} = \widetilde{y}_{it} - \widetilde{x}_{it}'\widehat{\beta}$ and $\widetilde{y}_{it} = y_{it} - \overline{y}_{i.}$. In order to test the null hypothesis of no cointegration, the null can be written as $H_0: \ \rho = 1$. The OLS estimate of ρ and the t-statistic are given as

$$\widehat{\rho} = \frac{\sum_{i=1}^N \sum_{t=2}^T \widehat{e}_{it}\widehat{e}_{it-1}}{\sum_{i=1}^N \sum_{t=2}^T \widehat{e}_{it}^2}$$

and

$$t_\rho = \frac{(\widehat{\rho} - 1)\sqrt{\sum_{i=1}^N \sum_{t=2}^T \widehat{e}_{it-1}^2}}{s_e} \tag{12.17}$$

where $s_e^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}(\widehat{e}_{it} - \widehat{\rho}\widehat{e}_{it-1})^2$. Kao proposed the following four DF-type tests:

$$\begin{aligned}
\mathrm{DF}_\rho &= \frac{\sqrt{N}T(\widehat{\rho}-1)+3\sqrt{N}}{\sqrt{10.2}} \\
\mathrm{DF}_t &= \sqrt{1.25}t_\rho + \sqrt{1.875N} \\
\mathrm{DF}_\rho^* &= \frac{\sqrt{N}T(\widehat{\rho}-1)+\frac{3\sqrt{N}\widehat{\sigma}_\nu^2}{\widehat{\sigma}_{0\nu}^2}}{\sqrt{3+\frac{36\widehat{\sigma}_\nu^4}{5\widehat{\sigma}_{0\nu}^4}}}
\end{aligned}$$

and

$$\mathrm{DF}_t^* = \frac{t_\rho + \frac{\sqrt{6N}\widehat{\sigma}_\nu}{2\widehat{\sigma}_{0\nu}}}{\sqrt{\frac{\widehat{\sigma}_{0\nu}^2}{2\widehat{\sigma}_\nu^2}+\frac{3\widehat{\sigma}_\nu^2}{10\widehat{\sigma}_{0\nu}^2}}}$$

where $\widehat{\sigma}_\nu^2 = \widehat{\sum}_{yy} - \widehat{\sum}_{yx}\widehat{\sum}_{xx}^{-1}$ and $\widehat{\sigma}_{0\nu}^2 = \widehat{\Omega}_{yy} - \widehat{\Omega}_{yx}\widehat{\Omega}_{xx}^{-1}$. While DF_ρ and DF_t are based on the strong exogeneity of the regressors and errors, DF_ρ^* and DF_t^* are for the cointegration with endogenous relationship between regressors and errors. For the ADF test, we can run the following regression:

$$\widehat{e}_{it} = \rho\widehat{e}_{it-1} + \sum_{j=1}^{p}\vartheta_j\Delta\widehat{e}_{it-j} + \nu_{itp} \tag{12.18}$$

With the null hypothesis of no cointegration, the ADF test statistic can be constructed as:

$$\mathrm{ADF} = \frac{t_{\mathrm{ADF}} + \frac{\sqrt{6N}\widehat{\sigma}_\nu}{2\widehat{\sigma}_{0\nu}}}{\sqrt{\frac{\widehat{\sigma}_{0\nu}^2}{2\widehat{\sigma}_\nu^2}+\frac{3\widehat{\sigma}_\nu^2}{10\widehat{\sigma}_{0\nu}^2}}} \tag{12.19}$$

where t_{ADF} is the t-statistic of ρ in (12.18). The asymptotic distributions of DF_ρ, DF_t, DF_ρ^*, DF_t^* and ADF converge to a standard normal distribution $N(0, 1)$ by sequential limit theory.

12.5.2 Residual-Based LM Test

McCoskey and Kao (1998) derived a residual-based test for the null of cointegration rather than the null of no cointegration in panels. This test is an extension of the LM test and the locally best invariant (LBI) test for an MA unit root in the time series literature. Under the null, the asymptotics no longer depend on the asymptotic properties of the estimating spurious regression, rather the asymptotics of the estimation of a cointegrated relationship are needed. For models which allow the cointegrating vector to change across the cross-sectional observations, the asymptotics depend merely on the time series results as each cross-section is estimated independently. For models with common slopes, the estimation is done jointly and therefore the asymptotic theory is based on the joint estimation of a cointegrated relationship in panel data.

For the residual based test of the null of cointegration, it is necessary to use an efficient estimation technique of cointegrated variables. In the time series literature a variety of methods have been shown to be efficient asymptotically. These include the fully modified (FM)-OLS estimator of Phillips and Hansen (1990) and the dynamic least squares (DOLS) estimator proposed by Saikkonen (1991) and Stock and Watson (1993). For panel data, Kao and Chiang (2000) showed that both the FM and DOLS methods can produce estimators which are asymptotically normally distributed with zero means.

The model presented allows for varying slopes and intercepts:

$$y_{it} = \alpha_i + x'_{it}\beta_i + e_{it} \tag{12.20}$$

$$x_{it} = x_{it-1} + \varepsilon_{it} \tag{12.21}$$

$$e_{it} = \gamma_{it} + u_{it} \tag{12.22}$$

and

$$\gamma_{it} = \gamma_{it-1} + \theta u_{it}$$

where u_{it} are $\mathrm{IID}(0, \sigma_u^2)$. The null of hypothesis of cointegration is equivalent to $\theta = 0$.

The test statistic proposed by McCoskey and Kao (1998) is defined as follows:

$$\mathrm{LM} = \frac{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}S_{it}^2}{\widehat{\sigma}_e^2} \tag{12.23}$$

where S_{it} is partial sum process of the residuals, $S_{it} = \sum_{j=1}^{t}\widehat{e}_{ij}$ and $\widehat{\sigma}_e^2$ is defined in McCoskey and Kao. The asymptotic result for the test is

$$\sqrt{N}(\mathrm{LM} - \mu_v) \Rightarrow N(0, \sigma_v^2) \tag{12.24}$$

The moments, μ_v and σ_v^2, can be found through Monte Carlo simulation. The limiting distribution of LM is then free of nuisance parameters and robust to heteroskedasticity.

Urban economists have long sought to explain the relationship between urbanization levels and output. McCoskey and Kao (1999) revisited this question and test the long-run stability of a production function including urbanization using nonstationary panel data techniques. McCoskey and Kao applied the IPS test and LM in (12.23) and showed that a long-run relationship between urbanization, output per worker and capital per worker cannot be rejected for the sample of 30 developing countries or the sample of 22 developed countries over the period 1965–89. They do find, however, that the sign and magnitude of the impact of urbanization varies considerably across the countries. These results offer new insights and potential for dynamic urban models rather than the simple cross-section approach.

12.5.3 Pedroni Tests

Pedroni (2000, 2004) also proposed several tests for the null hypothesis of cointegration in a panel data model that allows for considerable heterogeneity. His tests can be classified into two categories. The first set is similar to the tests discussed above, and involves averaging test statistics for cointegration in the time series across cross-sections. For the second set, the averaging is done in pieces so that the limiting distributions are based on limits of piecewise numerator and denominator terms.

The first set of statistics includes a form of the average of the Phillips and Ouliaris (1990) statistic:

$$\tilde{Z}_\rho = \sum_{i=1}^{N}\frac{\sum_{t=1}^{T}(\hat{e}_{it-1}\Delta\hat{e}_{it} - \hat{\lambda}_i)}{(\sum_{t=1}^{T}\hat{e}_{it-1}^2)} \tag{12.25}$$

where $\hat{e}_{it}$ is estimated from (12.15) and $\hat{\lambda}_i = \frac{1}{2}\left(\widehat{\sigma}_i^2 - \widehat{s}_i^2\right)$, for which $\widehat{\sigma}_i^2$ and $\widehat{s}_i^2$ are individual long-run and contemporaneous variances of the residual $\hat{e}_{it}$. For his second set of statistics, Pedroni defines four panel variance ratio statistics. Let $\hat{\Omega}_i$ be a consistent estimate of Ω_i, the long-run variance–covariance matrix. Define $\hat{L}_i$ to be the lower triangular Cholesky composition of $\hat{\Omega}_i$ such that in the scalar case $\hat{L}_{22i} = \hat{\sigma}_\varepsilon$ and $\hat{L}_{11i} = \hat{\sigma}_u^2 - \hat{\sigma}_{u\varepsilon}^2/\hat{\sigma}_\varepsilon^2$ is the long-run conditional variance. Here we consider only one of these statistics:

$$Z_{t_{\hat{\rho}_{NT}}} = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}\hat{L}_{11i}^{-2}(\hat{e}_{it-1}\Delta\hat{e}_{it} - \hat{\lambda}_i)}{\sqrt{\tilde{\sigma}_{NT}^2(\sum_{i=1}^{N}\sum_{t=2}^{T}\hat{L}_{11i}^{-2}\hat{e}_{it-1}^2)}} \tag{12.26}$$

where $\tilde{\sigma}_{NT} = \frac{1}{N}\sum_{i=1}^{N}\frac{\hat{\sigma}_i^2}{\hat{L}_{11i}^2}$.

It should be noted that Pedroni bases his test on the average of the numerator and denominator terms, respectively, rather than the average for the statistic as a whole. Using results on convergence of functionals of Brownian motion, Pedroni finds the following result:

$$Z_{t_{\hat{\rho}_{NT}}} + 1.73\sqrt{N} \Rightarrow N(0, 0.93)$$

Note that this distribution applies to the model including an intercept and not including a time trend. Asymptotic results for other model specifications can be found in Pedroni (2000). The intuition on these tests with varying slopes is not straightforward. The convergence in distribution is based on individual convergence of the numerator and denominator terms. What is the intuition of rejection of the null hypothesis? Using the average of the overall test statistic allows more ease in interpretation: rejection of the null hypothesis means that enough of the individual cross-sections have statistics "far away" from the means predicted by theory were they to be generated under the null.

Pedroni (1999) derived asymptotic distributions and critical values for several residual-based tests of the null of no cointegration in panels where there are multiple regressors. The model includes regressions with individual-specific fixed effects and time trends. Considerable heterogeneity is allowed across individual members of the panel with regard to the associated cointegrating vectors and the dynamics of the underlying error process. By comparing results from individual countries and the panel as a whole, Pedroni (2001) rejects the strong PPP hypothesis and finds that no degree of cross-sectional dependency would be sufficient to overturn the rejection of strong PPP.

12.5.4 Likelihood-Based Cointegration Test

Larsson, Lyhagen and Löthgren (2001) presented a likelihood-based (LR) panel test of cointegrating rank in heterogeneous panel models based on the average of the individual rank trace statistics developed by Johansen (1995). The proposed LR-bar statistic is very similar to the IPS t-bar statistic in (12.7)–(12.10). In Monte Carlo simulation, Larsson et al. investigated the small sample properties of the standardized LR-bar statistic. They found that the proposed test requires a large time series dimension. Even if the panel has a large cross-sectional dimension, the size of the test will be severely distorted.

Groen and Kleibergen (2003) proposed a likelihood-based framework for cointegrating analysis in panels of a fixed number of vector error correction models. This improves on Larsson et al. (2001) since it allows cross-sectional correlation. Maximum likelihood estimators of the cointegrating vectors are constructed using iterated generalized method of moments (GMM) estimators. Using these estimators, Groen and Kleibergen construct likelihood ratio statistics to test for a common cointegration rank across the individual vector error correction models, both with heterogeneous and homogeneous cointegrating vectors. Groen and Kleibergen (2003) applied this likelihood ratio test to a data set of exchange rates and appropriate monetary fundamentals. They found strong evidence for the validity of the monetary exchange rate model within a panel of vector correction models for three major European countries, whereas the results based on individual vector error correction models for each of these countries separately are less supportive.

12.5.5 Finite Sample Properties

McCoskey and Kao (1999) conducted Monte Carlo experiments to compare the size and power of different residual-based tests for cointegration in heterogeneous panel data: varying slopes and varying intercepts. Two of the tests are constructed under the null hypothesis of no cointegration. These tests are based on the average ADF test and Pedroni's pooled tests in (12.25)–(12.26). The third test is based on the null hypothesis of cointegration which is based on the McCoskey and Kao LM test in (12.23). Wu and Yin (1999) performed a similar comparison for panel tests in which they consider only tests for which the null hypothesis is that of no cointegration. Wu and Yin compared ADF statistics with maximum eigenvalue statistics in pooling information on means and p-values, respectively. They found that the average ADF performs better with respect to power and their maximum eigenvalue-based p-value performs better with regard to size.

The test of the null hypothesis was originally proposed in response to the low power of the tests of the null of no cointegration, especially in the time series case. Further, in cases where economic theory predicted a long-run steady state relationship, it seemed that a test of the null of cointegration rather than the null of no cointegration would be appropriate. The results from the Monte Carlo study showed that the McCoskey and Kao LM test outperforms the other two tests.

Of the two reasons for the introduction of the test of the null hypothesis of cointegration, low power and attractiveness of the null, the introduction of the cross-section dimension of the panel solves one: all of the tests show decent power when used with panel data. For those applications where the null of cointegration is more logical than the null of no cointegration, McCoskey and Kao (1999), at a minimum, conclude that using the McCoskey and Kao LM test does not compromise the ability of the researcher in determining the underlying nature of the data.

Gutierrez (2003) performed Monte Carlo experiments and compared some of the panel cointegration tests proposed by Kao (1999), Pedroni (2000) and Larsson et al. (2001). The Kao and Pedroni tests assume that either *all* the relationships are not cointegrated or *all* the relationships are cointegrated, while the Larsson et al. (2001) test assume that *all* N cross-sections have at most r cointegrating relationships against the alternative of a higher rank. Gutierrez (2003) finds that for a large T panel, the *whole* panel may be erroneously modeled as cointegrated when only a *small fraction* of the relationships are actually cointegrated. For $N = 10, 25, 100$; $T = 10, 50, 100$ and the proportion of cointegrated relationships varying between $0, 0.1, 0.2, \ldots, 1$, Gutierrez (2003) finds that for small $T = 10$, and as N increases, Kao's tests show higher power than the Pedroni tests. But, this power is still fairly low even when $N = 100$. As T gets large, the Pedroni tests have higher power than the Kao tests. Both tests performed better than the Larsson et al. (2001) LR-bar test.

12.6 ESTIMATION AND INFERENCE IN PANEL COINTEGRATION MODELS

For panel cointegrated regression models, the asymptotic properties of the estimators of the regression coefficients and the associated statistical tests are different from those of the time series cointegration regression models. Some of these differences have become apparent in recent works by Kao and Chiang (2000), Phillips and Moon (1999) and Pedroni (2000, 2004). The panel cointegration models are directed at studying questions that surround long-run economic relationships typically encountered in macroeconomic and financial data. Such a long-run relationship is often predicted by economic theory and it is then of central interest to estimate the regression coefficients and test whether they satisfy theoretical restrictions. Chen, McCoskey and Kao (1999) investigated the finite sample proprieties of the OLS estimator the t-statistic, the bias-corrected OLS estimator, and the bias-corrected t-statistic. They found that the bias-corrected OLS estimator does not improve over the OLS estimator in general. The results of Chen et al. suggested that alternatives, such as the fully modified (FM) estimator or dynamic OLS (DOLS) estimator, may be more promising in cointegrated panel regressions. Phillips and Moon (1999) and Pedroni (2000) proposed an FM estimator, which can be seen as a generalization of Phillips and Hansen (1990). Recently, Kao and Chiang (2000) proposed an alternative approach based on a panel dynamic least squares (DOLS) estimator, which builds upon the work of Saikkonen (1991) and Stock and Watson (1993).

Kao and Chiang (2000) consider the following panel regression:

$$y_{it} = x'_{it}\beta + z'_{it}\gamma + u_{it} \tag{12.27}$$

where $\{y_{it}\}$ are 1×1, β is a $k \times 1$ vector of the slope parameters, z_{it} is the deterministic component and $\{u_{it}\}$ are the stationary disturbance terms. $\{x_{it}\}$ are $k \times 1$ integrated processes of order one for all i, where

$$x_{it} = x_{it-1} + \varepsilon_{it}$$

The assumption of cross-sectional independence is maintained. Under these specifications, (12.27) describes a system of cointegrated regressions, i.e., y_{it} is cointegrated with x_{it}. The OLS estimator of β is

$$\widehat{\beta}_{\text{OLS}} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{x}'_{it}\right]^{-1}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{y}_{it}\right] \tag{12.28}$$

It is easy to show that

$$\frac{1}{N}\sum\nolimits_{i=1}^{N}\frac{1}{T^2}\sum\nolimits_{t=1}^{T}\widetilde{x}_{it}\widetilde{x}'_{it} \quad \overset{p}{\to} \quad \lim_{N\to\infty}\frac{1}{N}\sum\nolimits_{i=1}^{N}E\left[\zeta_{2i}\right] \tag{12.29}$$

and

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{u}_{it} \Rightarrow \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E\left[\zeta_{1i}\right] \tag{12.30}$$

using sequential limit theory, where

$$\begin{array}{lll} z_{it} & E\left[\zeta_{1i}\right] & E\left[\zeta_{2i}\right] \\ 0 & 0 & \frac{1}{2} \\ 1 & 0 & 0 \\ \mu_i & -\frac{1}{2}\Omega_{\varepsilon ui} + \Delta_{\varepsilon ui} & \frac{1}{6}\Omega_{\varepsilon i} \\ (\mu_i, t) & -\frac{1}{2}\Omega_{\varepsilon ui} + \Delta_{\varepsilon ui} & \frac{1}{15}\Omega_{\varepsilon i} \end{array} \tag{12.31}$$

and

$$\Omega_i = \begin{bmatrix} \Omega_{ui} & \Omega_{u\varepsilon i} \\ \Omega_{\varepsilon ui} & \Omega_{\varepsilon i} \end{bmatrix}$$

is the long-run covariance matrix of $\left(u_{it}, \varepsilon_{it}'\right)'$, also $\Delta_i = \begin{bmatrix} \Delta_{ui} & \Delta_{u\varepsilon i} \\ \Delta_{\varepsilon ui} & \Delta_{\varepsilon i} \end{bmatrix}$ is the one-sided long-run covariance. For example, when $z_{it} = \{\mu_i\}$, we get

$$\sqrt{N}T\left(\widehat{\beta}_{\text{OLS}} - \beta\right) - \sqrt{N}\delta_{NT} \Rightarrow N\left(0, 6\Omega_{\varepsilon}^{-1}\left(\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\Omega_{u.\varepsilon i}\Omega_{\varepsilon i}\right)\Omega_{\varepsilon}^{-1}\right) \tag{12.32}$$

where $\Omega_\varepsilon = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\Omega_{\varepsilon i}$ and

$$\delta_{NT} = \left[\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}\left(x_{it} - \overline{x}_i\right)\left(x_{it} - \overline{x}_i\right)'\right]^{-1}\frac{1}{N}\left[\sum_{i=1}^{N}\Omega_{\varepsilon i}^{1/2}\left(\int \widetilde{W}_i \mathrm{d}W_i'\right)\Omega_{\varepsilon i}^{-1/2}\Omega_{\varepsilon ui} + \Delta_{\varepsilon ui}\right] \tag{12.33}$$

This shows that $\widehat{\beta}_{\text{OLS}}$ is inconsistent using panel data. This is in sharp contrast with the consistency of $\widehat{\beta}_{\text{OLS}}$ in time series under similar circumstances. Kao and Chiang (2000) suggest a fully modified (FM) and DOLS estimators in a cointegrated regression and show that their limiting distribution is normal. Phillips and Moon (1999) and Pedroni (2000) also obtained similar results for the FM estimator. The reader is referred to the cited papers for further details. Kao and Chiang also investigated the finite sample properties of the OLS, FM and DOLS estimators. They found that (i) the OLS estimator has a nonnegligible bias in finite samples, (ii) the FM estimator does not improve over the OLS estimator in general, and (iii) the DOLS estimator may be more promising than OLS or FM estimators in estimating the cointegrated panel regressions.

Kao et al. (1999) apply the asymptotic theory of panel cointegration developed by Kao and Chiang (2000) to the Coe and Helpman (1995) international R&D spillover regression. Using a sample of 21 OECD countries and Israel, they re-examine the effects of domestic and foreign R&D capital stocks on total factor productivity of these countries. They find that OLS with bias correction, the fully modified (FM) and the dynamic OLS (DOLS) estimators produce different predictions about the impact of foreign R&D on total factor productivity (TFP), although all the estimators support the result that domestic R&D is related to TFP. Kao et al.'s empirical results indicate that the estimated coefficients in the Coe and Helpman's regressions are subject to estimation bias. Given the superiority of the DOLS over FM as suggested by Kao and Chiang (2000), Kao et al. leaned towards rejecting the Coe and Helpman hypothesis that international R&D spillovers are trade-related.

Choi (2002) studied instrumental variable estimation for an error component model with stationary and nearly nonstationary regressors. In contrast to the time series literature, Choi (2002) shows that IV estimation can be used for panel data with endogenous and nearly nonstationary regressors. To illustrate, consider the simple panel regression

$$y_{it} = \alpha + \beta x_{it} + u_{it}$$

where x_{it} is nearly nonstationary, u_{it} is $I(0)$ and z_t is an instrumental variable yielding the panel IV (Within) estimator

$$\widehat{\beta}_{\mathrm{IV}} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \overline{x}_{i.})(z_{it} - \overline{z}_{i.})\right]^{-1}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - \overline{y}_{i.}\right)(z_{it} - \overline{z}_{i.})\right]$$

Choi (2002) shows $\sqrt{N}T(\widehat{\beta}_{\mathrm{IV}} - \beta)$ has the weak limit as $T \to \infty$ of a standardized sum (over $i = 1, \ldots, N$) of zero mean random variables divided by a standardized sum of random variables. Thus when N is large, and proper conditions hold, the central limit theorem can be applied which leads to the asymptotic normality result for the panel estimator. In time series, standard hypothesis testing cannot be performed based on the corresponding IV estimator for β. The same intuition holds for Within–IV–OLS, IV–GLS and Within–IV–GLS estimators discussed in Choi (2002). For panel regressions that allow for cross-section correlation, one can use the SUR approach in the panel unit root test for fixed N, see Mark, Ogaki and Sul (2000) and Moon and Perron (2004b) who adopt this approach and show that the dynamic GLS estimator is most efficient.

Kauppi (2000) developed a new joint limit theory where the panel data may be cross-sectionally heterogeneous in a general way. This limit theory builds upon the concepts of joint convergence in probability and in distribution for double indexed processes by Phillips and Moon (1999) and develops new versions of the law of large numbers and the central limit theorem that apply in panels where the data may be cross-sectionally heterogeneous in a fairly general way. Kauppi demonstrates how this joint limit theory can be applied to derive asymptotics for a panel regression where the regressors are generated by a local to unit root with heterogeneous localizing coefficients across cross-sections. Kauppi discusses issues that arise in the estimation and inference of panel cointegrated regressions with near integrated regressors. Kauppi shows that a bias corrected pooled OLS for a common cointegrating parameter has an asymptotic normal distribution centered on the true value irrespective of whether the regressor has near or exact unit root. However, if the regression model contains individual effects and/or deterministic trends, then Kauppi's bias corrected pooled OLS still produces asymptotic bias. Kauppi also shows that the panel FM estimator is subject to asymptotic bias regardless of how individual effects and/or deterministic trends are contained if the regressors are nearly rather than exactly integrated. This indicates that much care should be taken in interpreting empirical results achieved by the recent panel cointegration methods that assume exact unit roots when near unit roots are equally plausible.

12.7 EMPIRICAL EXAMPLE: PURCHASING POWER PARITY

Banerjee et al. (2005) survey the empirical literature on the validity of purchasing power parity. The strong version of PPP tests whether the real exchange rate is stationary. A common finding is that PPP holds when tested in panel data, but not when tested on a country by country basis. The usual explanation is that panel tests for unit roots are more powerful than their univariate counterparts. Banerjee et al. (2005) offer an alternative explanation. Their results indicate that this mismatch may be due simply to the over-sizing that is present when cointegrating relationships link the countries of the panel together. Existing panel unit root tests assume that cross-unit cointegrating relationships among the countries are not present. Banerjee et al. (2005) show through simulations that when this assumption is violated, the empirical size of the tests is substantially higher than the nominal level and the null hypothesis of a unit root is rejected too often when it is true. They demonstrate this using quarterly data on real exchange rates for the period 1975:1–2002:4 for 18 OECD countries. Computing the ADF test on a country by country basis using both the USA and Germany in turn as numeraire, Banerjee et al. (2005) fail to reject the null hypothesis of a unit root for each country at any choice of lag length except for France and Korea, when Germany is the numeraire. The panel unit roots (assuming no cross-country cointegration), on the other hand, reject the null of unit root in 13 out of 16 cases with the USA as numeraire. These 16 cases correspond to four tests and four different lag-length choices. The four tests include two versions of the IPS test on ($\bar{t}$ and $\overline{\text{LM}}$), the LLC test and the Maddala and Wu (1999) Fisher test. If Germany is the numeraire, the corresponding rejections are in 12 out of 16 cases. Using critical values adjusted for the presence of cross-country cointegration, these rejections decrease. For example, with 14 bivariate cointegrating relationships, the unit root hypothesis is rejected in only 2 out of 16 cases with the USA as the numeraire and never with Germany as the numeraire. The authors conclude that this finding warns against the "automatic" use of panel methods for testing for unit roots in macroeconomic time series.

Table 12.1 performs panel unit root tests on the Banerjee et al. (2005) data on real exchange rates with Germany as the numeraire. This is done using EViews. This data was kindly provided by Chiara Osbat. The EViews options allow for the choice of exogenous variables, in this case, the inclusion of individual effects. Also, the automatic selection of maximum lags, or the choice of a user-specified lag. In fact, Table 12.2 performs these panel unit root tests with a user-specified lag of 1. Note that EViews performs the LLC, Breitung, IPS and Fisher-type tests of Maddala and Wu (1999) and Choi (2001) using ADF and Phillips–Perron type individual unit root tests. Both Tables 12.1 and 12.2 confirm the results in Banerjee et al. (2005), i.e., that all panel unit root tests including individual effects reject the null hypothesis of a common unit root. EViews also computes the Hadri's (2000) residual-based LM test which reverses the null hypothesis. In this case, it rejects the null hypothesis of no unit root in any of the series in the panel in favor of a common unit root in the panel. Problem 12.3 asks the reader to replicate these results and check their sensitivity to user-specified lags as well as the choice of Germany or USA as the numeraire.

Banerjee et al. (2004) also show that both univariate and multivariate panel cointegration tests can be substantially over-sized in the presence of cross-unit cointegration. They argue that the panel cointegration literature assume a unique cointegrating vector in each unit, either homogeneous (Kao, 1999) or heterogeneous (Pedroni, 1999) across the units of the panel. Groen and Kleibergen (2003) and Larsson et al. (2001) have developed techniques, à la Johansen's maximum likelihood method, that allow for multiple cointegrating vectors in each unit. These models allow for cross-unit dependence through the effects of the dynamics of short-run, but no account is taken of the possibility of long-run cross-unit dependence induced by the existence of cross-unit cointegrating relationships. Banerjee et al. (2004) show through Monte Carlo simulations that the consequences of using panel cointegrated methods when the restriction of the no cross-unit cointegration is violated are dramatic. They also confirm the gains in efficiency when the use of the panel approach is justified. Hence, they suggest testing

Table 12.1 Panel Unit Root Test (Automatic Lag) for Real Exchange Rates: Germany as Numeraire

Pool unit root test: Summary
Sample: 1975Q1 2002Q4
Series: RER_AUSTRIA, RER_BELGIUM, RER_CANADA,
RER_DENMARK, RER_FINLAND, RER_FRANCE,
RER_GREECE, RER_ITALY, RER_JAPAN, RER_KOREA,
RER_NETHERLANDS, RER_NORWAY, RER_PORTUGAL,
RER_SPAIN, RER_SWEDEN, RER_SWITZ, RER_UK, RER_US
Exogenous variables: Individual effects
Automatic selection of maximum lags
Automatic selection of lags based on SIC: 0 to 8
Newey–West bandwidth selection using Bartlett kernel

Method	Statistic	Prob.**	Cross-sections	Obs.
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t^*	−1.83839	0.0330	18	1970
Breitung t-stat	−3.06048	0.0011	18	1952
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	−3.42675	0.0003	18	1970
ADF-Fisher chi-square	63.6336	0.0030	18	1970
PP-Fisher chi-square	58.1178	0.0112	18	1998
Null: No unit root (assumes common unit root process)				
Hadri Z-stat	9.43149	0.0000	18	2016

**Probabilities for Fisher tests are computed using an asympotic chi-square distribution. All other tests assume asymptotic normality.

for the validity of the no cross-unit cointegration hypothesis prior to applying panel cointegration methods. Specifically, they recommend the extraction of the common trends from each unit using the Johansen ML method, and then testing for cointegration among these trends to rule out the existence of cross-unit cointegration. Their simulation results show that this procedure works well in practice.

12.8 FURTHER READING

Cermeño (1999) extends Andrews' (1993) median-unbiased estimation for autoregressive/unit root time series to panel data dynamic fixed effects models. This estimator is robust to heteroskedasticity and serial correlation in the individual dimension. However, this method is justified only for a purely autoregressive model. This estimator is used to evaluate conditional convergence among 48 US states, 13 OECD countries and two wider samples from the Penn World Tables with 57 and 100 countries. Support for conditional convergence is found only among US states and the 13 OECD countries. Phillips and Sul (2003) extend Cermeño's study by developing a class of panel median-unbiased estimators that address a more general case of cross-section dependence. This allows one to test the homogeneity restrictions on the dynamics, including the important case of unit root homogeneity.

Pesaran et al. (1999) derived the asymptotics of a pooled mean group (PMG) estimator. The PMG estimation constrains the long-run coefficients to be identical, but allows the short-run

Table 12.2 Panel Unit Root Test (Lag = 1) for Real Exchange Rates: Germany as Numeraire

Pool unit root test: Summary
Sample: 1975Q1 2002Q4
Series: RER_AUSTRIA, RER_BELGIUM, RER_CANADA,
RER_DENMARK, RER_FINLAND, RER_FRANCE,
RER_GREECE, RER_ITALY, RER_JAPAN, RER_KOREA,
RER_NETHERLANDS, RER_NORWAY, RER_PORTUGAL,
RER_SPAIN, RER_SWEDEN, RER_SWITZ, RER_UK, RER_US
Exogenous variables: Individual effects
User specified lags at: 1
Newey–West bandwidth selection using Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs.
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t^*	−1.71432	0.0432	18	1980
Breitung t-stat	−2.86966	0.0021	18	1962
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	−3.04702	0.0012	18	1980
ADF-Fisher chi-square	59.3350	0.0085	18	1980
PP-Fisher chi-square	58.1178	0.0112	18	1998
Null: No unit root (assumes common unit root process)				
Hadri Z-stat	9.43149	0.0000	18	2016

**Probabilities for Fisher tests are computed using an asympotic chi-square distribution. All other tests assume asymptotic normality.

and adjustment coefficients as the error variances to differ across the cross-sectional dimension. Binder, Hsiao and Pesaran (2002) considered estimation and inference in panel vector autoregressions (PVARS) with fixed effects when T is finite and N is large. A quasi-maximum likelihood estimator as well as unit root and cointegration tests are proposed based on a transformed likelihood function. This QMLE is shown to be consistent and asymptotically normally distributed irrespective of the unit root and cointegrating properties of the PVAR model. The tests proposed are based on standard chi-square and normal distributed statistics. Binder et al. also show that the conventional GMM estimators based on standard orthogonality conditions break down if the underlying time series contain unit roots. Monte Carlo evidence is provided which favors MLE over GMM in small samples.

Granger and Hyung (1999) consider the problem of estimating a dynamic panel regression model when the variables in the model are strongly correlated with individual-specific size factors. For a large N cross-country panel with small T, the size variable could be country-specific like its area or time-varying like population or total income. They show that if the size is not explicitly taken into account, one gets a spurious regression. In particular, they show that implementing unit root tests is likely to lead to the wrong decision. Moreover, if the size variable is slightly varying over time or its distribution has thick tails (such as a panel of countries including Luxembourg and Cyprus as well as China and India), post-sample predictions are biased. A pooling model appears to fit well in-sample, but forecast poorly out-of-sample if the individual-specific size factor has a fat-tailed distribution. A panel model with individual-specific effects could be problematic if the panel series has a very short time dimension. Since individual constant terms are estimated poorly, the forecasts based on them are poor. These problems may be more serious if the individual-specific factor is not constant but time-varying.

Hall, Lazarova and Urga (1999) proposed an approach based on principal components analysis to test for the number of common stochastic trends driving the nonstationary series in a panel data set. The test is consistent even if there is a mixture of $I(0)$ and $I(1)$ series in the sample. This makes it unnecessary to pretest the panel for unit root. It also has the advantage of solving the problem of dimensionality encountered in large panel data sets.

Hecq, Palm and Urbain (2000) extend the concept of serial correlation common features analysis to nonstationary panel data models. This analysis is motivated both by the need to study and test for common structures and comovements in panel data with autocorrelation present and by an increase in efficiency due to pooling. The authors propose sequential testing procedures and test their performance using a small-scale Monte Carlo. Concentrating upon the fixed effects model, they define homogeneous panel common feature models and give a series of steps to implement these tests. These tests are used to investigate the liquidity constraints model for 22 OECD and G7 countries. The presence of a panel common feature vector is rejected at the 5% nominal level.

Murray and Papell (2000) propose a panel unit root test in the presence of structural change. In particular, they propose a unit root test for nontrending data in the presence of a one-time change in the mean for a heterogeneous panel. The date of the break is endogenously determined. The resultant test allows for both serial and contemporaneous correlation, both of which are often found to be important in the panel unit roots context. Murray and Papell conduct two power experiments for panels of nontrending, stationary series with a one time change in means and find that conventional panel unit root tests generally have very low power. Then they conduct the same experiment using methods that test for unit roots in the presence of structural change and find that the power of the test is much improved.

NOTES

1. See the survey by Baltagi and Kao (2000). Chiang and Kao (2001) have put together a fairly comprehensive set of subroutines, for studying nonstationary panel data. These can be downloaded from http://web.syr.edu/~cdkao.
2. The Levin et al. (2002) paper has its origins in a Levin and Lin working paper in 1992, and most early applications in economics were based on the latter paper. In fact, this panel unit root test was commonly cited as the Levin–Lin test.
3. Other tests allowing for cross-section correlation based on a SUR model were suggested by O'Connell (1998) and Taylor and Sarno (1998).

PROBLEMS

12.1 *A simple linear trend model with error components.* This is based on problem 97.2.1 in *Econometric Theory* by Baltagi and Krämer (1997). Consider the following simple linear trend model

$$y_{it} = \alpha + \beta t + u_{it} \quad i = 1, 2, \ldots, N;\ t = 1, 2, \ldots, T$$

where y_{it} denotes the gross domestic product of country i at time t. The disturbances follow the one-way error component model given by

$$u_{it} = \mu_i + \nu_{it}$$

where $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$ denote the random country (time-invariant) effects and $\nu_{it} \sim \text{IID}(0, \sigma_\nu^2)$ denote the remainder effects. These error components are assumed to be independent of each other and among themselves. Our interest is focused on the estimates of the trend coefficient β, and the estimators to be considered are ordinary least squares (OLS), first difference (FD), the fixed effects (FE) estimator, assuming the μ_i's are fixed effects, and the generalized least squares estimator (GLS), knowing the true variance components, which is the best linear unbiased estimator in this case.

(a) Show that the OLS, GLS and FE estimators of β are identical and given by $\widehat{\beta}_{\text{GLS}} = \widehat{\beta}_{\text{OLS}} = \widetilde{\beta}_{\text{FE}} = \sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}(t-\bar{t})/N\sum_{t=1}^{T}(t-\bar{t})^2$ where $\bar{t} = \sum_{t=1}^{T} t/T$.

(b) Show that the variance of the OLS, GLS and FE estimators of β is given by $\text{var}(\widehat{\beta}_{\text{GLS}}) = \text{var}(\widehat{\beta}_{\text{OLS}}) = \text{var}(\widetilde{\beta}_{\text{FE}}) = 12\sigma_\nu^2/NT(T^2-1)$ and is therefore $O(N^{-1}T^{-3})$.

(c) Show that this simple linear trend model satisfies the necessary and sufficient condition for OLS to be equivalent to GLS.

(d) Show that the FD estimator of β is given by $\widehat{\beta}_{\text{FD}} = \sum_{i=1}^{N}(y_{iT} - y_{i1})/N(T-1)$ with $\text{var}(\widehat{\beta}_{\text{FD}}) = 2\sigma_\nu^2/N(T-1)^2$ of $O(N^{-1}T^{-2})$.

(e) What do you conclude about the asymptotic relative efficiency of FD with respect to the other estimators of β as $T \to \infty$? *Hint*: See solution 97.2.1 in *Econometric Theory* by Song and Jung (1998). Also, use the fact that $\sum_{t=1}^{T} t^2 = T(T+1)(2T+1)/6$ and $\sum_{t=1}^{T} t = T(T+1)/2$.

12.2 Download the International R&D spillovers panel data set used by Kao et al. (1999) along with the GAUSS subroutines from http://web.syr.edu/~cdkao. Using this data set, replicate the following results.

(a) Perform the Harris and Tzavalis (1999) panel unit root tests on total factor productivity, domestic R&D and foreign R&D capital stocks. Show that the null hypothesis of nonstationarity is not rejected for all three variables.

(b) Perform the Kao (1999) and Pedroni (2000) panel cointegration tests on the regression relating total factor productivity to domestic and foreign R&D stocks. Show that the null hypothesis of no cointegration is rejected.

(c) Estimate the cointegrating relationship using the Kao and Chiang (2000) procedure. *Note:* This example is used by Chiang and Kao (2001).

12.3 Using the Banerjee et al. (2005) quarterly data set on real exchange rate for 18 OECD countries over the period 1975:1–2002:4:

(a) Replicate the panel unit root test in Table 12.1 with Germany as the numeraire. Check the sensitivity of these results to a user-specified lag of 1, 2, 3 and 4. Compare with table 8 of Banerjee et al. (2005).

(b) Perform the panel unit root test as in Table 12.1 but now with the USA as the numeraire. Check the sensitivity of these results to a user-specified lag of 1, 2, 3 and 4. Compare with table 8 of Banerjee et al. (2005).

(c) Perform the individual ADF unit root tests on a country by country basis for both parts (a) and (b). Compare with table 7 of Banerjee et al. (2005). What do you conclude?

(d) Check the sensitivity of the results in parts (a) and (b) when both individual effects and individual linear trends are included.

12.4 Using the EViews G7 countries work file containing the GDP of Canada, France, Germany, Italy, Japan, UK and USA:

(a) Perform the panel unit root tests using individual effects in the deterministic variables.

(b) Check the sensitivity of these results to a user-specified lag of 1, 2, 3 and 4. Show that all tests are in agreement about the possibility of a common unit root for all series.

(c) Check the sensitivity of the results in parts (a) and (b) when both individual effects and individual linear trends are included.

References

Aasness, J., E. Biorn and T. Skjerpen, 1993, Engle functions, panel data, and latent variables, *Econometrica* **61**, 1395–1422.

Abbring, J.H. and G.J. Van den Berg, 2004, Analyzing the effect of dynamically assigned treatments using duration models, binary treatment models, and panel data models, *Empirical Economics* **29**, 5–20.

Abowd, J.M. and D. Card, 1989, On the covariance structure of earnings and hours changes, *Econometrica* **57**, 411–445.

Abrevaya, J., 1997, The equivalence of two estimators of the fixed-effects logit model, *Economics Letters* **55**, 41–43.

Abrevaya, J., 1999, Leapfrog estimation of a fixed-effects model with unknown transformation of the dependent variable, *Journal of Econometrics* **93**, 203–228.

Abrevaya, J., 2000, Rank estimation of a generalized fixed-effects regression model, *Journal of Econometrics* **95**, 1–23.

Ahn, S.C. and S. Low, 1996, A reformulation of the Hausman test for regression models with pooled cross-section time-series data, *Journal of Econometrics* **71**, 309–319.

Ahn, S.C. and P. Schmidt, 1995, Efficient estimation of models for dynamic panel data, *Journal of Econometrics* **68**, 5–27.

Ahn, S.C. and P. Schmidt, 1997, Efficient estimation of dynamic panel data models: Alternative assumptions and simplified estimation, *Journal of Econometrics* **76**, 309–321.

Ahn, S.C. and P. Schmidt, 1999a, Modified generalized instrumental variables estimation of panel data models with strictly exogenous instrumental variables, Chapter 7 in C. Hsiao, K. Lahiri, L.F. Lee and M.H. Pesaran, eds., *Analysis of Panels and Limited Dependent Variable Models* (Cambridge University Press, Cambridge), 171–198.

Ahn, S.C. and P. Schmidt, 1999b, Estimation of linear panel data models using GMM, Chapter 8 in L. Mátyás, ed., *Generalized Method of Moments Estimation* (Cambridge University Press, Cambridge), 211–247.

Alessie, R., A. Kapteyn and B. Melenberg, 1989, The effects of liquidity constraints on consumption: Estimation from household panel data, *European Economic Review* **33**, 547–555.

Alonso-Borrego, C. and M. Arellano, 1999, Symmetrically normalized instrumental variable estimation using panel data, *Journal of Business and Economic Statistics* **17**, 36–49.

Altonji, J.G. and A. Siow, 1987, Testing the response of consumption to income changes with (noisy) panel data, *Quarterly Journal of Economics* **102**, 293–328.

Alvarez, J. and M. Arellano, 2003, The time series and cross-section asymptotics of dynamic panel data estimators, *Econometrica* **71**, 1121–1159.

Amemiya, T., 1971, The estimation of the variances in a variance-components model, *International Economic Review* **12**, 1–13.

Amemiya, T. and T.E. MaCurdy, 1986, Instrumental-variable estimation of an error components model, *Econometrica* **54**, 869–881.

Andersen, T.G. and R.E. Sørensen, 1996, GMM estimation of a stochastic volatility model: A Monte Carlo study, *Journal of Business and Economic Statistics* **14**, 328–352.

Anderson, T.W. and C. Hsiao, 1981, Estimation of dynamic models with error components, *Journal of the American Statistical Association* **76**, 598–606.

Anderson, T.W. and C. Hsiao, 1982, Formulation and estimation of dynamic models using panel data, *Journal of Econometrics* **18**, 47–82.

Andrews, D.W.K., 1993, Exactly median-unbiased estimation of first order autoregressive/unit root models, *Econometrica* **61**, 139–165.

Andrews, D.W.K. and B. Lu, 2001, Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models, *Journal of Econometrics* **101**, 123–164.

Angrist, J.D. and A.B. Krueger, 1995, Split sample instrumental variable estimates of return to schooling, *Journal of Business and Economic Statistics* **13**, 225–235.

Angrist, J.D. and W.K. Newey, 1991, Over-identification tests in earnings functions with fixed effects, *Journal of Business and Economic Statistics* **9**, 317–323.

Anselin, L., 1988, *Spatial Econometrics: Methods and Models* (Kluwer Academic Publishers, Dordrecht).

Anselin, L., 2001, Spatial econometrics, Chapter 14 in B. Baltagi, ed., *A Companion to Theoretical Econometrics* (Blackwell Publishers, Massachusetts), 310–330.

Antweiler, W., 2001, Nested random effects estimation in unbalanced panel data, *Journal of Econometrics* **101**, 295–313.

Arellano, M., 1987, Computing robust standard errors for within-groups estimators, *Oxford Bulletin of Economics and Statistics* **49**, 431–434.

Arellano, M., 1989, A note on the Anderson–Hsiao estimator for panel data, *Economics Letters* **31**, 337–341.

Arellano, M., 1990, Some testing for autocorrelation in dynamic random effects models, *Review of Economic Studies* **57**, 127–134.

Arellano, M., 1993, On the testing of correlated effects with panel data, *Journal of Econometrics* **59**, 87–97.

Arellano, M., 2003, *Panel Data Econometrics* (Oxford University Press, Oxford).

Arellano, M. and S. Bond, 1991, Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *Review of Economic Studies* **58**, 277–297.

Arellano, M. and O. Bover, 1995, Another look at the instrumental variables estimation of error-component models, *Journal of Econometrics* **68**, 29–51.

Arellano, M. and R. Carrasco, 2003, Binary choice panel data models with predetermined variables, *Journal of Econometrics* **115**, 125–157.

Arellano, M. and B. Honoré, 2001, Panel data models: Some recent developments, Chapter 53 in J. Heckman and E. Leamer, eds., *Handbook of Econometrics* (North-Holland, Amsterdam), 3229–3296.

Arellano, M., O. Bover and J.M. Labeaga, 1999, Autoregressive models with sample selectivity for panel data, Chapter 2 in C. Hsiao, K. Lahiri, L.F. Lee and H. Pesaran, eds., *Analysis of Panels and Limited Dependent Variable Models* (Cambridge University Press, Cambridge), 23–48.

Askildsen, J.E., B.H. Baltagi and T.H. Holmås, 2003, Wage policy in the health care sector: A panel data analysis of nurses' labour supply, *Health Economics* **12**, 705–719.

Avery, R.B., 1977, Error components and seemingly unrelated regressions, *Econometrica* **45**, 199–209.

Avery, R.B., L.P. Hansen and V.J. Hotz, 1983, Multiperiod probit models and orthogonality condition estimation, *International Economic Review* **24**, 21–35.

Bai, J., 2004, Estimating cross-section common stochastic trends in nonstationary panel data, *Journal of Econometrics* **122**, 137–183.

Bai, J. and S. Ng, 2004, A PANIC attack on unit roots and cointegration, *Econometrica* **72**, 1127–1177.

Bailar, B.A., 1975, The effects of rotation group bias on estimates from panel surveys, *Journal of the American Statistical Association* **70**, 23–30.

Bailar, B.A., 1989, Information needs, surveys, and measurement errors, in D. Kasprzyk, G.J. Duncan, G. Kalton and M.P. Singh, eds., *Panel Surveys* (John Wiley, New York), 1–24.

Baillie, R.T. and B.H. Baltagi, 1999, Prediction from the regression model with one-way error components, Chapter 10 in C. Hsiao, K. Lahiri, L.F. Lee and H. Pesaran, eds., *Analysis of Panels and Limited Dependent Variable Models* (Cambridge University Press, Cambridge), 255–267.

Balestra, P., 1980, A note on the exact transformation associated with the first-order moving average process, *Journal of Econometrics* **14**, 381–394.

Balestra, P. and M. Nerlove, 1966, Pooling cross-section and time-series data in the estimation of a dynamic model: The demand for natural gas, *Econometrica* **34**, 585–612.

Balestra, P. and J. Varadharajan-Krishnakumar, 1987, Full information estimations of a system of simultaneous equations with error components structure, *Econometric Theory* **3**, 223–246.

Baltagi, B.H., 1980, On seemingly unrelated regressions with error components, *Econometrica* **48**, 1547–1551.

Baltagi, B.H., 1981a, Pooling: An experimental study of alternative testing and estimation procedures in a two-way error components model, *Journal of Econometrics* **17**, 21–49.

Baltagi, B.H., 1981b, Simultaneous equations with error components, *Journal of Econometrics* **17**, 189–200.

Baltagi, B.H., 1984, A Monte Carlo study for pooling time-series of cross-section data in the simultaneous equations model, *International Economic Review* **25**, 603–624.

Baltagi, B.H., 1985, Pooling cross-sections with unequal time-series lengths, *Economics Letters* **18**, 133–136.

Baltagi, B.H., 1986, Pooling under misspecification: Some Monte Carlo evidence on the Kmenta and the error components techniques, *Econometric Theory* **2**, 429–440.

Baltagi, B.H., 1987, On estimating from a more general time-series cum cross-section data structure, *The American Economist* **31**, 69–71.

Baltagi, B.H., 1989, Applications of a necessary and sufficient condition for OLS to be BLUE, *Statistics and Probability Letters* **8**, 457–461.

Baltagi, B.H., 1993, Useful matrix transformations for panel data analysis: A survey, *Statistical Papers* **34**, 281–301.

Baltagi, B.H., 1995, *Econometric Analysis of Panel Data* (John Wiley, Chichester).

Baltagi, B.H., 1996, Testing for random individual and time effects using a Gauss–Newton regression, *Economics Letters* **50**, 189–192.

Baltagi, B.H., 1997, Testing for linear and loglinear error component regression against Box–Cox alternatives, *Statistics and Probability Letters* **33**, 63–68.

Baltagi, B.H., 1999, Specification tests in panel data models using artificial regressions, *Annales D'Économie et de Statistique* **55–56**, 277–297.

Baltagi, B.H., ed., 2002, *Recent Developments in the Econometrics of Panel Data*, Volumes I and II (Edward Elgar Publishing, Cheltenham).

Baltagi, B.H., ed., 2004, *Panel Data: Theory and Applications*, (Physica-Verlag, Heidelberg).

Baltagi, B.H., 2005, Estimating an economic model of crime using panel data from North Carolina, *Journal of Applied Econometrics*, forthcoming.

Baltagi, B.H. and U. Blien, 1998, The German wage curve: Evidence from the IAB employment sample, *Economics Letters* **61**, 135–142.

Baltagi, B.H. and Y.J. Chang, 1994, Incomplete panels: A comparative study of alternative estimators for the unbalanced one-way error component regression model, *Journal of Econometrics* **62**, 67–89.

Baltagi, B.H. and Y.J. Chang, 1996, Testing for random individual effects using recursive residuals, *Econometric Reviews* **15**, 331–338.

Baltagi, B.H. and Y.J. Chang, 2000, Simultaneous equations with incomplete panels, *Econometric Theory* **16**, 269–279.

Baltagi, B.H. and J.M. Griffin, 1983, Gasoline demand in the OECD: An application of pooling and testing procedures, *European Economic Review* **22**, 117–137.

Baltagi, B.H. and J.M. Griffin, 1984, Short and long run effects in pooled models, *International Economic Review* **25**, 631–645.

Baltagi, B.H. and J.M. Griffin, 1988a, A generalized error component model with heteroscedastic disturbances, *International Economic Review* **29**, 745–753.

Baltagi, B.H. and J.M. Griffin, 1988b, A general index of technical change, *Journal of Political Economy* **96**, 20–41.

Baltagi, B.H. and J.M. Griffin, 1995, A dynamic demand model for liquor: The case for pooling, *Review of Economics and Statistics* **77**, 545–553.

Baltagi, B.H. and J.M. Griffin, 1997, Pooled estimators vs. their heterogeneous counterparts in the context of dynamic demand for gasoline, *Journal of Econometrics* **77**, 303–327.

Baltagi, B.H. and J.M. Griffin, 2001, The econometrics of rational addiction: The case of cigarettes, *Journal of Business and Economic Statistics* **19**, 449–454.

Baltagi, B.H. and C. Kao, 2000, Nonstationary panels, cointegration in panels and dynamic panels: A survey, *Advances in Econometrics* **15**, 7–51.

Baltagi, B.H. and S. Khanti-Akom, 1990, On efficient estimation with panel data: An empirical comparison of instrumental variables estimators, *Journal of Applied Econometrics* **5**, 401–406.

Baltagi, B.H. and W. Krämer, 1994, Consistency, asymptotic unbiasedness and bounds on the bias of s^2 in the linear regression model with error component disturbances, *Statistical Papers* **35**, 323–328.

Baltagi, B.H. and D. Levin, 1986, Estimating dynamic demand for cigarettes using panel data: The effects of bootlegging, taxation, and advertising reconsidered, *Review of Economics and Statistics* **68**, 148–155.

Baltagi, B.H. and D. Levin, 1992, Cigarette taxation: Raising revenues and reducing consumption, *Structural Change and Economic Dynamics* **3**, 321–335.

Baltagi, B.H. and D. Li, 2004, Prediction in the panel data model with spatial correlation, Chapter 13 in L. Anselin, R.J.G.M. Florax and S.J. Rey, eds., *Advances in Spatial Econometrics: Methodology, Tools and Applications* (Springer, Berlin), 283–295.

Baltagi, B.H. and Q. Li, 1990, A Lagrange multiplier test for the error components model with incomplete panels, *Econometric Reviews* **9**, 103–107.

Baltagi, B.H. and Q. Li, 1991a, A transformation that will circumvent the problem of autocorrelation in an error component model, *Journal of Econometrics* **48**, 385–393.

Baltagi, B.H. and Q. Li, 1991b, A joint test for serial correlation and random individual effects, *Statistics and Probability Letters* **11**, 277–280.

Baltagi, B.H. and Q. Li, 1992a, A monotonic property for iterative GLS in the two-way random effects model, *Journal of Econometrics* **53**, 45–51.

Baltagi, B.H. and Q. Li, 1992b, Prediction in the one-way error component model with serial correlation, *Journal of Forecasting* **11**, 561–567.

Baltagi, B.H. and Q. Li, 1992c, A note on the estimation of simultaneous equations with error components, *Econometric Theory* **8**, 113–119.

Baltagi, B.H. and Q. Li, 1994, Estimating error component models with general MA(q) disturbances, *Econometric Theory* **10**, 396–408.

Baltagi, B.H. and Q. Li, 1995, Testing AR(1) against MA(1) disturbances in an error component model, *Journal of Econometrics* **68**, 133–151.

Baltagi, B.H. and Q. Li, 1997, Monte Carlo results on pure and pretest estimators of an error component model with autocorrelated disturbances, *Annales D'Économie et de Statistique* **48**, 69–82.

Baltagi, B.H. and Q. Li, 2002, On instrumental variable estimation of semiparametric dynamic panel data models, *Economics Letters* **76**, 1–9.

Baltagi, B.H. and N. Pinnoi, 1995, Public capital stock and state productivity growth: Further evidence from an error components model, *Empirical Economics* **20**, 351–359.

Baltagi, B.H. and P.X. Wu, 1999, Unequally spaced panel data regressions with AR(1) disturbances, *Econometric Theory* **15**, 814–823.

Baltagi, B.H., G. Bresson and A. Pirotte, 2002, Comparison of forecast performance for homogeneous, heterogeneous and shrinkage estimators: Some empirical evidence from US electricity and natural-gas consumption, *Economics Letters* **76**, 375–382.

Baltagi, B.H., G. Bresson and A. Pirotte, 2003a, Fixed effects, random effects or Hausman–Taylor? A pretest estimator, *Economics* Letters **79**, 361–369.

Baltagi, B.H., G. Bresson and A. Pirotte, 2003b, A comparative study of pure and pretest estimators for a possibly misspecified two-way error component model, *Advances in Econometrics* **17**, 1–27.

Baltagi, B.H., G. Bresson and A. Pirotte, 2004, Tobin q: forecast performance for hierarchical Bayes, shrinkage, heterogeneous and homogeneous panel data estimators, *Empirical Economics* **29**, 107–113.

Baltagi, B.H., G. Bresson and A. Pirotte, 2005a, Adaptive estimation of heteroskedastic error component models, *Econometric Reviews* **24**, 39–58.

Baltagi, B.H., G. Bresson and A. Pirotte, 2005b, Joint LM test for heteroskedasticity in a one-way error component model, Working Paper, Texas A&M University, College Station, TX.

Baltagi, B.H., Y.J. Chang and Q. Li, 1992a, Monte Carlo evidence on panel data regressions with AR(1) disturbances and an arbitrary variance on the initial observations, *Journal of Econometrics* **52**, 371–380.

Baltagi, B.H., Y.J. Chang and Q. Li, 1992b, Monte Carlo results on several new and existing tests for the error component model, *Journal of Econometrics* **54**, 95–120.

Baltagi, B.H., Y.J. Chang and Q. Li, 1998, Testing for random individual and time effects using unbalanced panel data, *Advances in Econometrics* **13**, 1–20.

Baltagi, B.H., P. Egger and M. Pfaffermayr, 2003, A generalized design for bilateral trade flow models, *Economics Letters* **80**, 391–397.

Baltagi, B.H., T.B. Fomby and R.C. Hill, eds., 2000, Nonstationary panels, panel cointegration, and dynamic panels, *Advances in Econometrics*, Vol. 15 (Elsevier Science, Amsterdam).

Baltagi, B.H., J.M. Griffin and D. Rich, 1995, Airline deregulation: The cost pieces of the puzzle, *International Economic Review* **36**, 245–258.

Baltagi, B.H., J.M. Griffin and W. Xiong, 2000, To pool or not to pool: Homogeneous versus heterogeneous estimators applied to cigarette demand, *Review of Economics and Statistics* **82**, 117–126.

Baltagi, B.H., J. Hidalgo and Q. Li, 1996, A non-parametric test for poolability using panel data, *Journal of Econometrics* **75**, 345–367.

Baltagi, B.H., S.H. Song and B.C. Jung, 2001, The unbalanced nested error component regression model, *Journal of Econometrics* **101**, 357–381.

Baltagi, B.H., S.H. Song and B.C. Jung, 2002a, A comparative study of alternative estimators for the unbalanced two-way error component regression model, *Econometrics Journal* **5**, 480–493.

Baltagi, B.H., S.H. Song and B.C. Jung, 2002b, Simple LM tests for the unbalanced nested error component regression model, *Econometric Reviews* **21**, 167–187.

Baltagi, B.H., S.H. Song and B.C. Jung, 2002c, LM tests for the unbalanced nested panel data regression model with serially correlated errors, *Annales D'Économie et de Statistique* **65**, 219–268.

Baltagi, B.H., S.H. Song and W. Koh, 2003, Testing panel data regression models with spatial error correlation, *Journal of Econometrics* **117**, 123–150.

Baltagi, B.H., G. Bresson, J.M. Griffin and A. Pirotte, 2003, Homogeneous, heterogeneous or shrinkage estimators? Some empirical evidence from French regional gasoline consumption, *Empirical Economics* **28**, 795–811.

Banerjee, A., 1999, Panel data unit roots and cointegration: An overview, *Oxford Bulletin of Economics and Statistics* **61**, 607–629.

Banerjee, A., M. Marcellino and C. Osbat, 2004, Some cautions on the use of panel methods for integrated series of macro-economic data, *Econometrics Journal*, **7**, 322–340.

Banerjee, A., M. Marcellino and C. Osbat, 2005, Testing for PPP: Should we use panel methods?, *Empirical Economics*, forthcoming.

Battese, G.E. and T.J. Coelli, 1988, Prediction of firm level technical efficiencies with a generalized frontier production function and panel data, *Journal of Econometrics* **38**, 387–399.

Beck, N. and J. Katz, 1995, What to do (and not to do) with time-series-cross-section data in comparative politics, *American Political Science Review* **89**, 634–647.

Becker, G.S., M. Grossman and K.M. Murphy, 1994, An empirical analysis of cigarette addiction, *American Economic Review* **84**, 396–418.

Becketti, S., W. Gould, L. Lillard and F. Welch, 1988, The panel study of income dynamics after fourteen years: An evaluation, *Journal of Labor Economics* **6**, 472–492.

Behrman, J.R. and A.B. Deolalikar, 1990, The intrahousehold demand for nutrients in rural south India: Individual estimates, fixed effects and permanent income, *Journal of Human Resources* **25**, 665–696.

Beierlein, J.G., J.W. Dunn and J.C. McConnon, Jr., 1981, The demand for electricity and natural gas in the Northeastern United States, *Review of Economics and Statistics* **63**, 403–408.

Bell, K.P. and N.R. Bockstael, 2000, Applying the generalized-moments estimation approach to spatial problems involving microlevel data, *Review of Economics and Statistics* **82**, 72–82.

Bellmann, L., J. Breitung and J. Wagner, 1989, Bias correction and bootstrapping of error component models for panel data: Theory and applications, *Empirical Economics* **14**, 329–342.

Belsley, D.A., E. Kuh and R.E. Welsch, 1980, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity* (John Wiley, New York).

Bera, A.K., W. Sosa-Escudero and M. Yoon, 2001, Tests for the error component model in the presence of local misspecification, *Journal of Econometrics* **101**, 1–23.

Bernard, A. and C. Jones, 1996, Productivity across industries and countries: Time series theory and evidence, *Review of Economics and Statistics* **78**, 135–146.

Berndt, E.R., Z. Griliches and N. Rappaport, 1995, Econometric estimates of prices indexes for personal computers in the 1990's, *Journal of Econometrics* **68**, 243–268.

Berry, S., P. Gottschalk and D. Wissoker, 1988, An error components model of the impact of plant closing on earnings, *Review of Economics and Statistics* **70**, 701–707.

Bertrand, M., E. Duflo and S. Mullainathan, 2004, How much should we trust differences-in-differences estimates?, *Quarterly Journal of Economics* **119**, 249–275.

Bertschek, I. and M. Lechner, 1998, Convenient estimators for the panel probit model, *Journal of Econometrics* **87**, 329–371.

Berzeg, K., 1979, The error component model: Conditions for the existence of the maximum likelihood estimates, *Journal of Econometrics* **10**, 99–102.

Berzeg, K., 1982, Demand for motor gasoline, a generalized error components model, *Southern Economic Journal* **49**, 462–471.

Bhargava, A. and J.D. Sargan, 1983, Estimating dynamic random effects models from panel data covering short time periods, *Econometrica* **51**, 1635–1659.

Bhargava, A., L. Franzini and W. Narendranathan, 1982, Serial correlation and fixed effects model, *Review of Economic Studies* **49**, 533–549.

Binder, M., C. Hsiao and M.H. Pesaran, 2002, Estimation and inference in short panel vector autoregressions with unit roots and cointegration, Working Paper, Trinity College, Cambridge.

Biorn, E., 1981, Estimating economic relations from incomplete cross-section/time-series data, *Journal of Econometrics* **16**, 221–236.

Biorn, E., 1992, The bias of some estimators for panel data models with measurement errors, *Empirical Economics* **17**, 51–66.

Biorn, E., 1996, Panel data with measurement errors, Chapter 10 in L. Mátyás and P. Sevestre, eds., *The Econometrics of Panel Data: A Handbook of the Theory With Applications* (Kluwer Academic Publishers, Dordrecht), 236–279.

Biorn, E., 2000, Panel data with measurement errors: Instrumental variables and GMM procedures combining levels and differences, *Econometric Reviews* **19**, 391–424.

Biorn, E., 2004, Regression systems for unbalanced panel data: A stepwise maximum likelihood procedure, *Journal of Econometrics* **122**, 281–291.

Biorn, E. and E.S. Jansen, 1983, Individual effects in a system of demand functions, *Scandinavian Journal of Economics* **85**, 461–483.

Biorn, E. and T.J. Klette, 1998, Panel data with errors-in-variables: Essential and redundant orthogonality conditions in GMM-estimation, *Economics Letters* **59**, 275–282.

Björklund, A., 1985, Unemployment and mental health: Some evidence from panel data, *Journal of Human Resources* **20**, 469–483.

Björklund, A., 1989, Potentials and pitfalls of panel data: The case of job mobility, *European Economic Review* **33**, 537–546.

Blanchard, P., 1996, Software review, Chapter 33 in L. Mátyás and P. Sevestre, eds., *The Econometrics of Panel Data: A Handbook of the Theory With Applications* (Kluwer Academic Publishers, Dordrecht), 879–913.

Blanchard, P. and L. Mátyás, 1996, Robustness of tests for error components models to non-normality, *Economics Letters* **51**, 161–167.

Blundell, R., 1988, Consumer behaviour: Theory and empirical evidence – A survey, *The Economic Journal* **98**, 16–65.

Blundell, R. and S. Bond, 1998, Initial conditions and moment restrictions in dynamic panel data models, *Journal of Econometrics* **87**, 115–143.

Blundell, R. and S. Bond, 2000, GMM estimation with persistent panel data: An application to production functions, *Econometric Reviews* **19**, 321–340.

Blundell, R. and C. Meghir, 1990, Panel data and life-cycle models, in J. Hartog, G. Ridder and J. Theeuwes, eds., *Panel Data and Labor Market Studies* (North-Holland, Amsterdam), 231–252.

Blundell, R. and R.S. Smith, 1991, Conditions initiales et estimation efficace dans les modèles dynamiques sur données de panel, *Annales D'Économie et de Statistique* **20–21**, 109–123.

Blundell, R., S. Bond and F. Windmeijer, 2000, Estimation in dynamic panel data models: Improving on the performance of the standard GMM estimator, *Advances in Econometrics* **15**, 53–91.

Blundell, R., C. Meghir and P. Neves, 1990, Labor supply and intertemporal substitution, *Journal of Econometrics* **59**, 137–160.

Blundell, R., S. Bond, M. Devereux and F. Schiantarelli, 1992, Investment and Tobin's *q*: Evidence from company panel data, *Journal of Econometrics* **51**, 233–257.

Boehmer, E. and W.L. Megginson, 1990, Determinants of secondary market prices for developing country syndicated loans, *The Journal of Finance* **45**, 1517–1540.

Bond, S. and F. Windmeijer, 2002, Projection estimators for autoregressive panel data models, *Econometrics Journal* **5**, 457–479.

Boumahdi, R. and A. Thomas, 1991, Testing for unit roots using panel data, *Economics Letters* **37**, 77–79.

Bound, L., C. Brown, G.J. Duncan and W.L. Rodgers, 1990, Measurement error in cross-sectional and longitudinal labor market surveys: Validation study evidence, in J. Hartog, G. Ridder and T. Theeuwes, eds., *Panel Data and Labor Market Studies* (North-Holland, Amsterdam), 1–19.

Bover, O. and M. Arellano, 1997, Estimating dynamic limited variable models from panel data, *Investigaciones Economicas* **21**, 141–165.

Bowden, R.J. and D.A. Turkington, 1984, *Instrumental Variables* (Cambridge University Press, Cambridge).

Bowsher, C.G., 2002, On testing overidentifying restrictions in dynamic panel data models, *Economics Letters* **77**, 211–220.

Breitung, J., 2000, The local power of some unit root tests for panel data, *Advances in Econometrics* **15**, 161–177.

Breitung, J. and M. Lechner, 1999, Alternative GMM methods for nonlinear panel data models, Chapter 9 in L. Mátyás, ed., *Generalized Method of Moments Estimation* (Cambridge University Press, Cambridge), 248–274.

Breitung, J. and W. Meyer, 1994, Testing for unit roots in panel data: Are wages on different bargaining levels cointegrated?, *Applied Economics* **26**, 353–361.

Breusch, T.S., 1987, Maximum likelihood estimation of random effects models, *Journal of Econometrics* **36**, 383–389.

Breusch, T.S. and L.G. Godfrey, 1981, A review of recent work on testing for autocorrelation in dynamic simultaneous models, in D.A. Currie, R. Nobay and D. Peel, eds., *Macroeconomic Analysis, Essays in Macroeconomics and Economics* (Croom Helm, London), 63–100.

Breusch, T.S. and A.R. Pagan, 1979, A simple test for heteroskedasticity and random coefficient variation, *Econometrica* **47**, 1287–1294.

Breusch, T.S. and A.R. Pagan, 1980, The Lagrange multiplier test and its applications to model specification in econometrics, *Review of Economic Studies* **47**, 239–253.

Breusch, T.S., G.E. Mizon and P. Schmidt, 1989, Efficient estimation using panel data, *Econometrica* **57**, 695–700.

Brown, J.N. and A. Light, 1992, Interpreting panel data on job tenure, *Journal of Labor Economics* **10**, 219–257

Brown, P., A.W. Kleidon and T.A. Marsh, 1983, New evidence on the nature of size-related anomalies in stock prices, *Journal of Financial Economics* **12**, 33–56.

Brown, R.L., J. Durbin and J.M. Evans, 1975, Techniques for testing the constancy of regression relationships over time, *Journal of the Royal Statistical Society* **37**, 149–192.

Browning, M., A. Deaton and M. Irish, 1985, A profitable approach to labor supply and commodity demands over the life cycle, *Econometrica* **53**, 503–543.

Bun, M.J.G., 2004, Testing poolability in a system of dynamic regressions with nonspherical disturbances, *Empirical Economics* **29**, 89–106.

Burke, S.P., L.G. Godfrey and A.R. Termayne, 1990, Testing AR(1) against MA(1) disturbances in the linear regression model: An alternative procedure, *Review of Economic Studies* **57**, 135–145.

Butler, J.S. and R. Moffitt, 1982, A computationally efficient quadrature procedure for the one factor multinominal probit model, *Econometrica* **50**, 761–764.

Cameron, C. and P. Trivedi, 1998, *Regression Analysis of Count Data* (Cambridge University Press, New York).

Campbell, J.R. and B.E. Honoré, 1993, Median unbiasedness of estimators of panel data censored regression models, *Econometric Theory* **9**, 499–503.

Canzoneri, M.B., R.E. Cumby and B. Diba, 1999, Relative labor productivity and the real exchange rate in the long run: Evidence for a panel of OECD countries, *Journal of International Economics* **47**, 245–266.

Cardellichio, P.A., 1990, Estimation of production behavior using pooled microdata, *Review of Economics and Statistics* **72**, 11–18.

Carpenter, R.E., S.M. Fazzari and B.C. Petersen, 1998, Financing constraints and inventory investment: A comparative study with high-frequency panel data, *Review of Economics and Statistics* **80**, 513–519.

Case, A.C., 1991, Spatial patterns in household demand, *Econometrica* **59**, 953–965.

Cashel-Cordo, P. and S.G. Craig, 1990, The public sector impact of international resource transfers, *Journal of Development Economics* **32**, 17–42.

Cermeño, R., 1999, Median-unbiased estimation in fixed-effects dynamic panels, *Annales D'Économie et de Statistique* **55–56**, 351–368.

Chamberlain, G., 1980, Analysis of covariance with qualitative data, *Review of Economic Studies* **47**, 225–238.

Chamberlain, G., 1982, Multivariate regression models for panel data, *Journal of Econometrics* **18**, 5–46.

Chamberlain, G., 1984, Panel data, Chapter 22 in Z. Griliches and M. Intrilligator, eds., *Handbook of Econometrics* (North-Holland, Amsterdam), 1247–1318.

Chamberlain, G., 1985, Heterogeneity, omitted variable bias and duration dependence, Chapter 1 in J.J. Heckman and B. Singer, eds., *Longitudinal Analysis of Labor Market Data* (Cambridge University Press, Cambridge), 3–38.

Chamberlain, G., 1993, Feedback in panel data models, Working Paper (Department of Economics, Harvard University).

Chamberlain, G. and K. Hirano, 1999, Predictive distributions based on longitudinal earnings data, *Annales D'Économie et de Statistique* **55–56**, 211–242.

Chang, Y., 2002, Nonlinear IV unit root tests in panels with cross sectional dependency, *Journal of Econometrics* **110**, 261–292.

Chang, Y., 2004, Bootstrap unit root tests in panels with cross sectional dependency, *Journal of Econometrics* **120**, 263–293.

Charlier, E., B. Melenberg and A. van Soest, 1995, A smoothed maximum score estimator for the binary choice panel data model with an application to labour force participation, *Statistica Neerlandica* **49**, 324–342.

Charlier, E., B. Melenberg and A. van Soest, 2001, An analysis of housing expenditure using semiparametric models and panel data, *Journal of Econometrics* **101**, 71–107.

Chen, B., S. McCoskey and C. Kao, 1999, Estimation and inference of a cointegrated regression in panel data: A Monte Carlo study, *American Journal of Mathematical and Management Sciences* **19**, 75–114.

Chesher, A., 1984, Testing for neglected heterogeneity, *Econometrica* **52**, 865–872.

Chiang, M.H. and C. Kao, 2001, Nonstationary panel time series using NPT 1.2 – A user guide (Center for Policy Research, Syracuse University).

Chib, S., 1996, Inference in panel data models via Gibbs sampling, Chapter 24 in L. Mátyás and P. Sevestre, eds., *The Econometrics of Panel Data: A Handbook of the Theory With Applications* (Kluwer Academic Publishers, Dordrecht), 639–651.

Chintagunta, P., E. Kyriazidou and J. Perktold, 2001, Panel data analysis of household brand choices, *Journal of Econometrics* **103**, 111–153.

Choi, I., 2001, Unit root tests for panel data, *Journal of International Money and Finance* **20**, 249–272.

Choi, I., 2002, Instrumental variables estimation of a nearly nonstationary, heterogeneous error component model, *Journal of Econometrics* **109**, 1–32.

Chow, G.C., 1960, Tests of equality between sets of coefficients in two linear regressions, *Econometrica* **28**, 591–605.

Chowdhury, G., 1991, A comparison of covariance estimators for complete and incomplete panel data models, *Oxford Bulletin of Economics and Statistics* **53**, 88–93.

Coakley, J. and A.M. Fuertes, 1997, New panel unit root tests of PPP, *Economics Letters* **57**, 17–22.

Coakely, J., F. Kulasi and R. Smith, 1996, Current account solvency and the Feldstein–Horioka puzzle, *Economic Journal* **106**, 620–627.

Coe, D. and E. Helpman, 1995, International R&D spillovers, *European Economic Review* **39**, 859–887.

Collado, M.D., 1997, Estimating dynamic models from time series of independent cross-sections, *Journal of Econometrics* **82**, 37–62.

Conley, T.G., 1999, GMM estimation with cross sectional dependence, *Journal of Econometrics* **92**, 1–45.

Contoyannis, P. and N. Rice, 2001, The impact of health on wages: Evidence from the British Household Panel Survey, *Empirical Economics*, **26**, 599–622.

Contoyannis, P., A.M. Jones and N. Rice, 2004, simulation-based inference in dynamic panel probit models: An application to health, *Empirical Economics* **29**, 49–77.

Conway, K.S. and T.J. Kniesner, 1992, How fragile are male labor supply function estimates, *Empirical Economics* **17**, 170–182.

Corbeil, R.R. and S.R. Searle, 1976a, A comparison of variance component estimators, *Biometrics* **32**, 779–791.

Corbeil, R.R. and S.R. Searle, 1976b, Restricted maximum likelihood (REML), estimation of variance components in the mixed model, *Technometrics* **18**, 31–38.

Cornwell, C. and P. Rupert, 1988, Efficient estimation with panel data: An empirical comparison of instrumental variables estimators, *Journal of Applied Econometrics* **3**, 149–155.

Cornwell, C. and P. Rupert, 1997, Unobservable individual effects, marriage, and then the earnings of young men, *Economic Inquiry* **35**, 285–294.

Cornwell, C. and W.N. Trumbull, 1994, Estimating the economic model of crime with panel data, *Review of Economics and Statistics* **76**, 360–366.

Cornwell, C., P. Schmidt and R.C. Sickles, 1990, Production frontiers with cross-sectional and time-series variation in efficiency levels, *Journal of Econometrics* **46**, 185–200.

Cornwell, C., P. Schmidt and D. Wyhowski, 1992, Simultaneous equations and panel data, *Journal of Econometrics* **51**, 151–181.

Crépon, B. and J. Mairesse, 1996, The Chamberlain approach, Chapter 14 in L. Mátyás and P. Sevestre, eds., *The Econometrics of Panel Data: A Handbook of the Theory With Applications* (Kluwer Academic Publishers, Dordrecht), 323–390.

Crépon, B., F. Kramarz and A. Trognon, 1997, Parameters of interest, nuisance parameters and orthogonality conditions: An application to autoregressive error components models, *Journal of Econometrics* **82**, 135–156.

Culver, S.E. and D.H. Papell, 1997, Is there a unit root in the inflation rate? Evidence from sequential break and panel data model, *Journal of Applied Econometrics* **12**, 435–444.

Das, M., 2004, Simple estimators for nonparametric panel data models with sample attrition, *Journal of Econometrics*, **120**, 159–180.

Das, M. and A. van Soest, 1999, A panel data model for subjective information on household income growth, *Journal of Economic Behavior & Organization* **40**, 409–426.

Davidson, R. and J.G. MacKinnon, 1993, *Estimation and Inference in Econometrics* (Oxford University Press, New York).

Davis, P., 2001, Estimating multi-way error components models with unbalanced data structures using instrumental variables, *Journal of Econometrics*, **106**, 67–95.

Deaton, A., 1985, Panel data from time series of cross-sections, *Journal of Econometrics* **30**, 109–126.

Deaton, A., 1995, Data and econometric tools for development analysis, Chapter 33 in J. Behrman and T.N. Srinivasan, eds., *Handbook of Development Economics* (Elsevier Science, Amsterdam), 1785–1882.

Deschamps, P., 1991, On the estimated variances of regression coefficients in misspecified error components models, *Econometric Theory* **7**, 369–384.

Dionne, G., R. Gagné and C. Vanasse, 1998, Inferring technological parameters from incomplete panel data, *Journal of Econometrics* **87**, 303–327.

Driscoll, J.C. and A.C. Kraay, 1998, Consistent covariance matrix estimation with spatially dependent panel data, *Review of Economics and Statistics* **80**, 549–560.

Driver, C., K. Imai, P. Temple and A. Urga, 2004, The effect of uncertainty on UK investment authorisation: homogeneous vs. heterogeneous estimators, *Empirical Economics* **29**, 115–128.

Duncan, G.J. and D.H. Hill, 1985, An investigation of the extent and consequences of measurement error in labor economic survey data, *Journal of Labor Economics* **3**, 508–532.

Egger, P. and M. Pfaffermayr, 2004a, Estimating long and short run effects in static panel models, *Econometric Reviews* **23**, 199–214.

Egger, P. and M. Pfaffermayr, 2004b, Distance, trade and FDI: A Hausman–Taylor SUR approach, *Journal of Applied Econometrics* **19**, 227–246.

Elliott, G., T.J. Rothenberg and J.H. Stock, 1996, Efficient tests for an autoregressive unit root, *Econometrica* **64**, 813–836.

England, P., G. Farkas, B.S. Kilbourne and T. Dou, 1988, Explaining occupational sex segregation and wages: Findings from a model with fixed effects, *American Sociological Review* **53**, 544–558.

Entorf, H., 1997, Random walks with drifts: Nonsense regression and spurious fixed-effect estimation, *Journal of Econometrics* **80**, 287–296.

Erdem, T., 1996, A dynamic analysis of market structure based on panel data, *Marketing Science* **15**, 359–378.

Erdem, T. and B. Sun, 2001, Testing for choice dynamics in panel data, *Journal of Business and Economic Statistics* **19**, 142–152.

Evans, M.A. and M.L. King, 1985, Critical value approximations for tests of linear regression disturbances, *Australian Journal of Statistics* **27**, 68–83.

Evans, P. and G. Karras, 1996, Convergence revisited, *Journal of Monetary Economics* **37**, 249–265.

Fiebig, D., R. Bartels and W. Krämer, 1996, The Frisch–Waugh theorem and generalized least squares, *Econometric Reviews* **15**, 431–443.

Fisher, F.M., 1970, Tests of equality between sets of coefficients in two linear regressions: An expository note, *Econometrica* **38**, 361–366.

Fleissig, A.R. and J. Strauss, 1997, Unit root tests on real wage panel data for the G7, *Economics Letters* **56**, 149–155.

Florens, J.P., D. Forgére and M. Monchart, 1996, Duration models, Chapter 19 in L. Mátyás and P. Sevestre, eds., *The Econometrics of Panel Data: A Handbook of the Theory With Applications* (Kluwer Academic Publishers, Dordrecht), 491–536.

Frankel, J.A. and A.K. Rose, 1996, A panel project on purchasing power parity: Mean reversion within and between countries, *Journal of International Economics* **40**, 209–224.

Frees, E.W., 1995, Assessing cross-sectional correlation in panel data, *Journal of Econometrics* **69**, 393–414.

Frees, E.W., V.R. Young and Y. Luo, 1999, A longitudinal data analysis interpretation of credibility models, *Insurance: Mathematics and Economics* **24**, 229–247.

Fuller, W.A. and G.E. Battese, 1973, Transformations for estimation of linear models with nested error structure, *Journal of the American Statistical Association* **68**, 626–632.

Fuller, W.A. and G.E. Battese, 1974, Estimation of linear models with cross-error structure, *Journal of Econometrics* **2**, 67–78.

Galbraith, J.W. and V. Zinde-Walsh, 1995, Transforming the error-component model for estimation with general ARMA disturbances, *Journal of Econometrics* **66**, 349–355.

Gardner, R., 1998, Unobservable individual effects in unbalanced panel data, *Economics Letters* **58**, 39–42.

Gerdtham, U.G. and M. Löthgren, 2000, On stationarity and cointegration of international health expenditure and GDP, *Journal of Health Economics* **19**, 461–475.

Ghosh, S.K., 1976, Estimating from a more general time-series cum cross-section data structure, *The American Economist* **20**, 15–21.

Girma, S., 2000, A quasi-differencing approach to dynamic modelling from a time series of independent cross-sections, *Journal of Econometrics* **98**, 365–383.

Glick, R. and A.K. Rose, 2002, Does a currency union affect trade? The time series evidence, *European Economic Review* **46**, 1125–1151.

Goldberger, A.S., 1962, Best linear unbiased prediction in the generalized linear regression model, *Journal of the American Statistical Association* **57**, 369–375.

Gourieroux, C. and A. Monfort, 1993, Simulation-based inference: A survey with special reference to panel data models, *Journal of Econometrics* **59**, 5–33.

Gourieroux, C., A. Holly and A. Monfort, 1982, Likelihood ratio test, Wald test, and Kuhn–Tucker test in linear models with inequality constraints on the regression parameters, *Econometrica* **50**, 63–80.

Granger, C.W.J. and N. Hyung, 1999, Spurious stochastics in a short time-series panel data, *Annales D'Économie et de Statistique* **55–56**, 299–315.

Graybill, F.A., 1961, *An Introduction to Linear Statistical Models* (McGraw-Hill, New York).

Greene, W.H., 2003, *Econometric Analysis* (Prentice Hall, New Jersey).

Greene, W.H., 2004a, The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects, *Econometrics Journal* **7**, 98–119.

Greene, W.H., 2004b, Convenient estimators for the panel probit models: Further results, *Empirical Economics* **29**, 21–47.

Griffin, J.M., 1982, The welfare implications of externalities and price elasticities for telecommunications pricing, *Review of Economics and Statistics* **64**, 59–66.

Griliches, Z., 1986, Economic data issues, Chapter 25 in Z. Griliches and M.D. Intrilligator, eds., *Handbook of Econometrics* (North-Holland, Amsterdam), 1466–1514.

Griliches, Z. and J.A. Hausman, 1986, Errors in variables in panel data, *Journal of Econometrics* **31**, 93–118.

Griliches, Z. and J. Mairesse, 1984, Productivity and R&D at the firm level, in Z. Griliches, ed., *R&D, Patents and Productivity* (University of Chicago Press, Chicago), 339–374.

Griliches, Z. and J. Mairesse, 1998, Production functions: The search for identification, in S. Strom, ed., *Essays in Honour of Ragnar Frisch*, Econometric Society Monograph Series (Cambridge University Press, Cambridge).

Groen, J.J.J., 2000, The monetary exchange rate model as a long-run phenomenon, *Journal of International Economics* **52**, 299–319.

Groen, J.J.J. and F. Kleibergen, 2003, Likelihood-based cointegration analysis in panels of vector error correction models, *Journal of Business and Economic Statistics* **21**, 295–318.

Grunfeld, Y., 1958, The determinants of corporate investment, unpublished Ph.D. dissertation (University of Chicago, Chicago).

Gutierrez, L., 2003, On the power of panel cointegration tests: A Monte Carlo comparison, *Economics Letters* **80**, 105–111.

Guilkey, D.K. and J.L. Murphy, 1993, Estimation and testing in the random effects probit model, *Journal of Econometrics* **59**, 301–317.

Haavelmo, T., 1944, The probability approach in econometrics, *Supplement to Econometrica* **12**, 1–118.

Hadri, K., 2000, Testing for stationarity in heterogeneous panel data, *Econometrics Journal* **3**, 148–161.

Häggström, E., 2002, Properties of Honda's test for random individual effects in non-linear regressions, *Statistical Papers,* **43**, 177–196.

Hahn, J., 1997, Efficient estimation of panel data models with sequential moment restrictions, *Journal of Econometrics* **79**, 1–21.

Hahn, J., 1999, How informative is the initial condition in the dynamic panel model with fixed effects?, *Journal of Econometrics* **93**, 309–326.

Hahn, J., 2001, The information bound of a dynamic panel logit model with fixed effects, *Econometric Theory* **17**, 913–932.

Hahn, J. and G. Kuersteiner, 2002, Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large, *Econometrica* **70**, 1639–1657.

Hahn, J. and W. Newey, 2004, Jacknife and analytical bias reduction for nonlinear panel models, *Econometrica* **72**, 1295–1319.

Hahn, J., J. Hausman and G. Kuersteiner, 2003, Bias corrected instrumental variables estimation for dynamic panel models with fixed effects, Working Paper (MIT, Massachusetts).

Hajivassiliou, V.A., 1987, The external debt repayments problems of LDC's: An econometric model based on panel data, *Journal of Econometrics* **36**, 205–230.

Hajivassiliou, V.A., 1994, Estimation by simulation of the external debt repayment problems, *Journal of Applied Econometrics* **9**, 109–132.

Hajivassiliou, V.A. and P.A. Ruud, 1994, Classical estimation methods for LDV models using simulation, Chapter 40 in R.F. Engle and D.L. McFadden, eds., *Handbook of Econometrics* (North-Holland, Amsterdam), 2383–2441.

Hall, S., S. Lazarova and G. Urga, 1999, A principal components analysis of common stochastic trends in heterogeneous panel data: Some Monte Carlo evidence, *Oxford Bulletin of Economics and Statistics* **61**, 749–767.

Hamerle, A., 1990, On a simple test for neglected heterogeneity in panel studies, *Biometrics* **46**, 193–199.

Hamermesh, D.S., 1989, Why do individual-effects models perform so poorly? The case of academic salaries, *Southern Economic Journal* **56**, 39–45.

Han, A.K. and D. Park, 1989, Testing for structural change in panel data: Application to a study of U.S. foreign trade in manufacturing goods, *Review of Economics and Statistics* **71**, 135–142.

Hansen, B.E., 1999, Threshold effects in non-dynamic panels: Estimation, testing and inference, *Journal of Econometrics* **93**, 345–368.

Hansen, L.P., 1982, Large sample properties of generalized method of moments estimators, *Econometrica* **50**, 1029–1054.

Haque, N.U., K. Lahiri and P. Montiel, 1993, Estimation of a macroeconomic model with rational expectations and capital controls for developing countries, *Journal of Development Economics* **42**, 337–356.

Harris, R.D.F. and E. Tzavalis, 1999, Inference for unit roots in dynamic panels where the time dimension is fixed, *Journal of Econometrics* **91**, 201–226.

Harrison, D. and D.L. Rubinfeld, 1978, Hedonic housing prices and the demand for clean air, *Journal of Environmental Economics and Management* **5**, 81–102.

Hartley, H.O. and J.N.K. Rao, 1967, Maximum likelihood estimation for the mixed analysis of variance model, *Biometrika* **54**, 93–108.

Harville, D.A., 1977, Maximum likelihood approaches to variance component estimation and to related problems, *Journal of the American Statistical Association* **72**, 320–340.

Hausman, J.A., 1978, Specification tests in econometrics, *Econometrica* **46**, 1251–1271.

Hausman, J.A. and W.E. Taylor, 1981, Panel data and unobservable individual effects, *Econometrica* **49**, 1377–1398.

Hausman, J.A. and D. Wise, 1979, Attrition bias in experimental and panel data: The Gary income maintenance experiment, *Econometrica* **47**, 455–473.

Hausman, J.A., B.H. Hall and Z. Griliches, 1984, Econometric models for count data with an application to the patents–R&D relationship, *Econometrica* **52**, 909–938.

Hayashi, F. and T. Inoue, 1991, The relation between firm growth and q with multiple capital goods: Theory and evidence from Japanese panel data, *Econometrica* **59**, 731–753.

Heckman, J.J., 1979, Sample selection bias as a specification error, *Econometrica* **47**, 153–161.

Heckman, J.J., 1981a, Statistical models for discrete panel data, in C.F. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications* (MIT Press, Cambridge), 114–178.

Heckman, J.J., 1981b, The incidental parameters problem and the problem of initial conditions in estimating a discrete time–discrete data stochastic process, in C.F. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications* (MIT Press, Cambridge), 179–195.

Heckman, J.J., 1981c, Heterogeneity and state dependence, in S. Rosen, ed., *Studies in Labor Markets* (Chicago University Press, Chicago), 91–139.

Heckman, J.J. and G.J. Borjas, 1980, Does unemployment cause future unemployment? Definitions, questions and answers from a continuous time model of heterogeneity and state dependence, *Economica* **47**, 247–283.

Heckman, J.J. and T.E. MaCurdy, 1980, A life-cycle model of female labor supply, *Review of Economic Studies* **47**, 47–74.

Heckman, J.J. and B. Singer, 1985, *Longitudinal Analysis of Labor Market Data* (Cambridge University Press, Cambridge).

Heckman, J.J. and R. Willis, 1975, Estimation of a stochastic model of reproduction: An econometric approach, in N. Terleckyi, ed., *Household Production and Consumption* (National Bureau of Economic Research, New York), 99–138.

Heckman, J.J. and R. Willis, 1977, A beta-logistic model for the analysis of sequential labor force participation by married women, *Journal of Political Economy* **85**, 27–58.

Heckman, J.J., H. Ichimura and P. Todd, 1998, Matching as an econometric evaluations estimator, *Review of Economic Studies* **65**, 261–294.

Hecq, A., F.C. Palm and J.P. Urbain, 2000, Testing for common cyclical features in non-stationary panel data models, *Advances in Econometrics* **15**, 131–160.

Hemmerle, W.J. and H.O. Hartley, 1973, Computing maximum likelihood estimates for the mixed A.O.V. model using the W-transformation, *Technometrics* **15**, 819–831.

Henderson, C.R., 1953, Estimation of variance components, *Biometrics* **9**, 226–252.

Herriot, R.A. and E.F. Spiers, 1975, Measuring the impact of income statistics of reporting differences between the current population survey and administrative sources, *Proceedings of the Social Statistics Section* (American Statistical Association), 147–158.

Hillier, G.H., 1990, On the normalization of structural equations: Properties of direction estimators, *Econometrica* **58**, 1181–1194.

Hirano, K., G.W. Imbens, G. Ridder and D. Rubin, 2001, Combining panel data sets with attrition and refreshment samples, *Econometrica* **69**, 1645–1660.

Hoch, L., 1962, Estimation of production function parameters combining time-series and cross-section data, *Econometrica* **30**, 34–53.

Hocking, R.R. and M.H. Kutner, 1975, Some analytical and numerical comparisons of estimators for the mixed A.O.V. model, *Biometrics* **31**, 19–28.

Holly, A., 1982, A remark on Hausman's specification test, *Econometrica* **50**, 749–759.

Holly, A. and L. Gardiol, 2000, A score test for individual heteroscedasticity in a one-way error components model, Chapter 10 in J. Krishnakumar and E. Ronchetti, eds., *Panel Data Econometrics: Future Directions* (North-Holland, Amsterdam), 199–211.

Holtz-Eakin, D., 1988, Testing for individual effects in autoregressive models, *Journal of Econometrics* **39**, 297–307.
Holtz-Eakin, D., 1994, Public-sector capital and the productivity puzzle, *Review of Economics and Statistics* **76**, 12–21.
Holtz-Eakin, D., W. Newey and H.S. Rosen, 1988, Estimating vector autoregressions with panel data, *Econometrica* **56**, 1371–1395.
Holtz-Eakin, D., W. Newey and H.S. Rosen, 1989, The revenues–expenditures nexus: Evidence from local government data, *International Economic Review* **30**, 415–429.
Honda, Y., 1985, Testing the error components model with non-normal disturbances, *Review of Economic Studies* **52**, 681–690.
Honda, Y., 1991, A standardized test for the error components model with the two-way layout, *Economics Letters* **37**, 125–128.
Hong, Y. and C.D. Kao, 2004, Wavelet-based testing for serial correlation of unknown form in panel models, *Econometrica*, forthcoming.
Honoré, B.E., 1992, Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects, *Econometrica* **60**, 533–565.
Honoré, B.E., 1993, Orthogonality conditions for Tobit models with fixed effects and lagged dependent variables, *Journal of Econometrics* **59**, 35–61.
Honoré, B.E. and L. Hu, 2004, Estimation of cross sectional and panel data censored regression models with endogeneity, *Journal of Econometrics* **122**, 293–316.
Honoré, B.E. and E. Kyriazidou, 2000a, Estimation of Tobit-type models with individual specific effects, *Econometric Reviews* **19**, 341–366.
Honoré, B.E. and E. Kyriazidou, 2000b, Panel data discrete choice models with lagged dependent variables, *Econometrica* **68**, 839–874.
Honoré, B.E. and A. Lewbel, 2002, Semiparametric binary choice panel data models without strictly exogenous regressors, *Econometrica*, **70**, 2053–2063.
Horowitz, J.L. and S. Lee, 2004, Semiparametric estimation of a panel data proportional hazards model with fixed effects, *Journal of Econometrics* **119**, 155–198.
Horowitz, J.L. and C.F. Manski, 1998, Censoring of outcomes and regressors due to survey nonresponse: Identification and estimation using weights and imputations, *Journal of Econometrics* **84**, 37–58.
Horowitz, J.L. and M. Markatou, 1996, Semiparametric estimation of regression models for panel data, *Review of Economic Studies* **63**, 145–168.
Houthakker, H.S., 1965, New evidence on demand elasticities, *Econometrica* **33**, 277–288.
Howrey, E.P. and H.R. Varian, 1984, Estimating the distributional impact of time-of-day pricing of electricity, *Journal of Econometrics* **26**, 65–82.
Hsiao, C., 1991, Identification and estimation of dichotomous latent variables models using panel data, *Review of Economic Studies* **58**, 717–731.
Hsiao, C., 1996, Logit and probit models, Chapter 16 in L. Mátyás and P. Sevestre, eds., *The Econometrics of Panel Data: A Handbook of the Theory With Applications* (Kluwer Academic Publishers, Dordrecht), 410–428.
Hsiao, C., 2001, Panel data models, Chapter 16 in B.H. Baltagi, ed., *A Companion to Theoretical Econometrics* (Blackwell Publishers, Massachusetts), 349–365.
Hsiao, C., 2003, *Analysis of Panel Data* (Cambridge University Press, Cambridge).
Hsiao, C. and B. Sun, 2000, To pool or not to pool panel data, Chapter 9 in J. Krishnakumar and E. Ronchetti, eds., *Panel Data Econometrics: Future Directions* (North-Holland, Amsterdam), 181–198.
Hsiao, C. and A.K. Tahmiscioglu, 1997, A panel analysis of liquidity constraints and firm investment, *Journal of the American Statistical Association* **92**, 455–465.
Hsiao, C., M.H. Pesaran and A.K. Tahmiscioglu, 1999, Bayes estimation of short run coefficients in dynamic panel data models, Chapter 11 in C. Hsiao, K. Lahiri, L.F. Lee and M.H. Pesaran, eds., *Analysis of Panels and Limited Dependent Variable Models* (Cambridge University Press, Cambridge), 268–296.
Hsiao, C., M.H. Pesaran and A.K. Tahmiscioglu, 2002, Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods, *Journal of Econometrics* **109**, 107–150.
Hsiao, C., K. Lahiri, L.F. Lee and M.H. Pesaran, eds., 1999, *Analysis of Panels and Limited Dependent Variable Models* (Cambridge University Press, Cambridge).
Hu, L., 2002, Estimation of a censored dynamic panel data model with an application to earnings dynamics, *Econometrica* **70**, 2499–2517.

Hujer, R. and H. Schneider, 1989, The analysis of labor market mobility using panel data, *European Economic Review* **33**, 530–536.
Hyslop, D., 1999, State dependence, serial correlation and heterogeneity in labor force participation of married women, *Econometrica* **67**, 1255–1294.
Im, K.S., S.C. Ahn, P. Schmidt and J.M. Wooldridge, 1999, Efficient estimation of panel data models with strictly exogenous explanatory variables, *Journal of Econometrics* **93**, 177–201.
Im, K.S., M.H. Pesaran and Y. Shin, 2003, Testing for unit roots in heterogeneous panels, *Journal of Econometrics* **115**, 53–74.
Imbens, G., 1997, One-step estimators for over-identified generalized method of moments models, *Review of Economic Studies* **64**, 359–383.
Islam, N., 1995, Growth empirics: A panel data approach, *Quarterly Journal of Economics* **110**, 1127–1170.
Jennrich, R.I. and P.F. Sampson, 1976, Newton–Raphson and related algorithms for maximum likelihood variance component estimation, *Technometrics* **18**, 11–17.
Jimenez-Martin, S., 1998, On the testing of heterogeneity effects in dynamic unbalanced panel data models, *Economics Letters* **58**, 157–163.
Johansen, S., 1995, *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models* (Oxford University Press, Oxford).
Johnson, S.C. and K. Lahiri, 1992, A panel data analysis of productive efficiency in freestanding health clinics, *Empirical Economics* **17**, 141–151.
Jorion, P. and R. Sweeney, 1996, Mean reversion in real exchange rates: Evidence and implications for forecasting, *Journal of International Money and Finance* **15**, 535–550.
Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lutkepohl and T.C. Lee, 1985, *The Theory and Practice of Econometrics* (John Wiley, New York).
Judson, R.A. and A.L. Owen, 1999, Estimating dynamic panel data models: A guide for macroeconomists, *Economics Letters* **65**, 9–15.
Kalton, G., D. Kasprzyk and D. McMillen, 1989, Nonsampling errors in panel surveys, in D. Kasprzyk, G.J. Duncan, G. Kalton and M.P. Singh, eds., *Panel Surveys* (John Wiley, New York), 249–270.
Kang, S., 1985, A note on the equivalence of specification tests in the two-factor multivariate variance components model, *Journal of Econometrics* **28**, 193–203.
Kang, S.J. and M.J. Lee, 2003, Analysis of private transfers with panel fixed effect censored model estimator, *Economics Letters* **80**, 233–237.
Kao, C., 1999, Spurious regression and residual-based tests for cointegration in panel data, *Journal of Econometrics* **90**, 1–44.
Kao, C. and M.H. Chiang, 2000, On the estimation and inference of a cointegrated regression in panel data, *Advances in Econometrics* **15**, 179–222.
Kao, C. and J.F. Schnell, 1987a, Errors in variables in panel data with binary dependent variable, *Economics Letters* **24**, 45–49.
Kao, C. and J.F. Schnell, 1987b, Errors in variables in a random effects probit model for panel data, *Economics Letters* **24**, 339–342.
Kao, C., M.H. Chiang and B. Chen, 1999, International R&D spillovers: An application of estimation and inference in panel cointegration, *Oxford Bulletin of Economics and Statistics* **61**, 691–709.
Kapoor, M., H.H. Kelejian and I.R. Prucha, 2004, Panel data models with spatially correlated error components, *Journal of Econometrics*, forthcoming.
Karlsson, S. and M. Löthgren, 2000, On the power and interpretation of panel unit root tests, *Economics Letters* **66**, 249–255.
Karlsson, S. and J. Skoglund, 2004, Maximum-likelihood based inference in the two-way random effects model with serially correlated time effects, *Empirical Economics* **29**, 79–88.
Kasprzyk, D., G.J. Duncan, G. Kalton and M.P. Singh, 1989, *Panel Surveys* (John Wiley, New York).
Kauppi, H., 2000, Panel data limit theory and asymptotic analysis of a panel regression with near integrated regressors, *Advances in Econometrics* **15**, 239–274.
Keane, M., 1993, Simulation estimation for panel data models with limited dependent variables, in G.S. Maddala, C.R. Rao and H.D. Vinod, eds., *Handbook of Statistics* (North-Holland, Amsterdam).
Keane, M., 1994, A computationally practical simulation estimator for panel data, *Econometrica* **62**, 95–116.

Keane, M., 1997, Modeling heterogeneity and static dependence in consumer choice behavior, *Journal of Business and Economic Statistics* **15**, 310–327.

Keane, M.P. and D.E. Runkle, 1992, On the estimation of panel-data models with serial correlation when instruments are not strictly exogenous, *Journal of Business and Economic Statistics* **10**, 1–9.

Keane, M.P., R. Moffitt and D.E. Runkle, 1988, Real wages over the business cycle: Estimating the impact of heterogeneity with micro data, *Journal of Political Economy* **96**, 1232–1266.

Kelejian, H.H. and I. Prucha, 1999, A generalized moments estimator for the autoregressive parameter in a spatial model, *International Economic Review* **40**, 509–533.

Kiefer, N.M., 1980, Estimation of fixed effects models for time series of cross sections with arbitrary intertemporal covariance, *Journal of Econometrics* **14**, 195–202.

Kim, B.S. and G.S. Maddala, 1992, Estimation and specification analysis of models of dividend behavior based on censored panel data, *Empirical Economics* **17**, 111–124.

Kinal, T. and K. Lahiri, 1990, A computational algorithm for multiple equation models with panel data, *Economics Letters* **34**, 143–146.

Kinal, T. and K. Lahiri, 1993, A simplified algorithm for estimation of simultaneous equations error components models with an application to a model of developing country foreign trade, *Journal of Applied Econometrics* **8**, 81–92.

King, M.L., 1983, Testing for autoregressive against moving average errors in the linear regression model, *Journal of Econometrics* **21**, 35–51.

King, M.L. and M. McAleer, 1987, Further results on testing AR(1) against MA(1) disturbances in the linear regression model, *Review of Economic Studies* **54**, 649–663.

King, M.L. and P.X. Wu, 1997, Locally optimal one-sided tests for multiparameter hypotheses, *Econometric Reviews* **16**, 131–156.

Kiviet, J.F., 1995, On bias, inconsistency and efficiency of various estimators in dynamic panel data models, *Journal of Econometrics* **68**, 53–78.

Kiviet, J.F., 1999, Expectations of expansions for estimators in a dynamic panel data model: Some results for weakly exogenous regressors, Chapter 8 in C. Hsiao, K. Lahiri, L.F. Lee and M.H. Pesaran, eds., *Analysis of Panels and Limited Dependent Variable Models* (Cambridge University Press, Cambridge), 199–225.

Kiviet, J.F. and W. Krämer, 1992, Bias of s^2 in the linear regression model with correlated errors, *Empirical Economics* **16**, 375–377.

Klevmarken, N.A., 1989, Panel studies: What can we learn from them? Introduction, *European Economic Review* **33**, 523–529.

Kmenta, J., 1986, *Elements of Econometrics* (MacMillan, New York).

Kniesner, T. and Q. Li, 2002, Semiparametric panel data models with heterogeneous dynamic adjustment: Theoretical consideration and an application to labor supply, *Empirical Economics*, **27**, 131–148.

Koop, G. and M.F.J. Steel, 2001, Bayesian analysis of stochastic frontier models, Chapter 24 in B.H. Baltagi, ed., *A Companion to Theoretical Econometrics* (Blackwell Publishers, Massachusetts), 520–537.

Krishnakumar, J., 1988, *Estimation of Simultaneous Equation Models with Error Components Structure* (Springer-Verlag, Berlin).

Krishnakumar, J. and E. Ronchetti, 2000, *Panel Data Econometrics: Future Directions* (North-Holland, Amsterdam).

Kuh, E., 1959, The validity of cross-sectionally estimated behavior equations in time series applications, *Econometrica* **27**, 197–214.

Kumbhakar, S.C. and C.A.K. Lovell, 2000, *Stochastic Frontier Analysis* (Cambridge University Press, Cambridge).

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin, 1992, Testing the null hypothesis of stationarity against the alternative of a unit root, *Journal of Econometrics* **54**, 159–178.

Kyriazidou, E., 1997, Estimation of a panel data sample selection model, *Econometrica* **65**, 1335–1364.

Kyriazidou, E., 2001, Estimation of dynamic panel data sample selection models, *Review of Economic Studies* **68**, 543–572.

Labeaga, J.M., 1999, A double-hurdle rational addiction model with heterogeneity: Estimating the demand for tobacco, *Journal of Econometrics* **93**, 49–72.

Laisney, F., M. Lechner and W. Pohlmeier, 1992, Innovation activity and firm heterogeneity: Empirical evidence from West Germany, *Structural Change and Economic Dynamics* **3**, 301–320.

Lancaster, T., 2000, The incidental parameter problem since 1948, *Journal of Econometrics* **95**, 391–413.
Larson, A.C. and J.S. Watters, 1993, A convenient test of functional form for pooled econometric models, *Empirical Economics* **18**, 271–280.
Larsson, R., J. Lyhagen and M. Löthgren, 2001, Likelihood-based cointegration tests in heterogeneous panels, *Econometrics Journal*, **4**, 109–142.
Lechner, M., 1995, Some specification tests for probit models estimated on panel data, *Journal of Business and Economic Statistics* **13**, 475–488.
Lechner, M. and J. Breitung, 1996, Some GMM estimation methods and specification tests for nonlinear models, Chapter 22 in L. Mátyás and P. Sevestre, eds., *The Econometrics of Panel Data: A Handbook of the Theory With Applications* (Kluwer Academic Publishers, Dordrecht), 583–611.
Lee, K., M.H. Pesaran and R. Smith, 1997, Growth and convergence in a multi-country empirical stochastic Solow model, *Journal of Applied Econometrics* **12**, 357–392.
Lee, L.F., 1987, Non-parametric testing of discrete panel data models, *Journal of Econometrics* **34**, 147–177.
Lee, M.J., 1999, A root-N-consistent semiparametric estimator for related effect binary response panel data, *Econometrica* **67**, 427–434.
Lee, M.J., 2001, First-difference estimator for panel censored-selection models, *Economics Letters* **70**, 43–49.
Lejeune B., 1996, A full heteroscedastic one-way error components model for incomplete panel: maximum likelihood estimation and Lagrange multiplier testing, CORE Discussion Paper 9606, Universite Catholique de Louvain, 1–28.
Levin, A., C.F. Lin and C. Chu, 2002, Unit root test in panel data: Asymptotic and finite sample properties, *Journal of Econometrics* **108**, 1–25.
Leybourne, S.J. and B.P.M. McCabe, 1994, A consistent test for a unit root, *Journal of Business and Economic Statistics* **12**, 157–166.
Li, Q. and C. Hsiao, 1998, Testing serial correlation in semiparametric panel data models, *Journal of Econometrics* **87**, 207–237.
Li, Q. and T. Stengos, 1992, A Hausman specification test based on root N consistent semiparametric estimators, *Economics Letters* **40**, 141–146.
Li, Q. and T. Stengos, 1994, Adaptive estimation in the panel data error component model with heteroskedasticity of unknown form, *International Economic Review* **35**, 981–1000.
Li, Q. and T. Stengos, 1995, A semi-parametric non-nested test in a dynamic panel data model, *Economics Letters* **49**, 1–6.
Li, Q. and T. Stengos, 1996, Semiparametric estimation of partially linear panel data models, *Journal of Econometrics* **71**, 389–397.
Li, Q. and A. Ullah, 1998, Estimating partially linear models with one-way error components, *Econometric Reviews* **17**, 145–166.
Lillard, L.A. and Y. Weiss, 1979, Components of variation in panel earnings data: American scientists 1960–1970, *Econometrica* **47**, 437–454.
Lillard, L.A. and R.J. Willis, 1978, Dynamic aspects of earning mobility, *Econometrica* **46**, 985–1012.
Little, R.J.A., 1988, Missing-data adjustments in large surveys, *Journal of Business and Economic Statistics* **6**, 287–297.
Lothian, J.R., 1996, Multi-country evidence on the behavior of purchasing power parity under the current float, *Journal of International Money and Finance* **16**, 19–35.
Lundberg, S. and E. Rose, 2002, The effects of sons and daughters on men's labor supply and wages, *The Review of Economics and Statistics* **84**, 251–268.
MaCurdy, T.A., 1982, The use of time series processes to model the error structure of earnings in a longitudinal data analysis, *Journal of Econometrics* **18**, 83–114.
MacDonald, R., 1996, Panel unit root tests and real exchange rates, *Economics Letters* **50**, 7–11.
Maddala, G.S., 1971, The use of variance components models in pooling cross section and time series data, *Econometrica* **39**, 341–358.
Maddala, G.S., 1977, *Econometrics* (McGraw-Hill, New York).
Maddala, G.S., 1983, *Limited Dependent and Qualitative Variables in Econometrics* (Cambridge University Press, Cambridge).
Maddala, G.S., 1991, To pool or not to pool: That is the question, *Journal of Quantitative Economics* **7**, 255–264.

Maddala, G.S., ed., 1993, *The Econometrics of Panel Data*, Vols I and II (Edward Elgar Publishing, Cheltenham).

Maddala, G.S., 1999, On the use of panel data methods with cross country data, *Annales D'Économie et de Statistique* **55–56**, 429–448.

Maddala, G.S. and T.D. Mount, 1973, A comparative study of alternative estimators for variance components models used in econometric applications, *Journal of the American Statistical Association* **68**, 324–328.

Maddala, G.S. and S. Wu, 1999, A comparative study of unit root tests with panel data and a new simple test, *Oxford Bulletin of Economics and Statistics* **61**, 631–652.

Maddala, G.S., S. Wu and P.C. Liu, 2000, Do panel data rescue purchasing power parity (PPP) theory?, Chapter 2 in J. Krishnakumar and E. Ronchetti, eds., *Panel Data Econometrics: Future Directions* (North-Holland, Amsterdam), 35–51.

Maddala, G.S., R.P. Trost, H. Li and F. Joutz, 1997, Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage estimators, *Journal of Business and Economic Statistics* **15**, 90–100.

Magnac, T., 2004, Panel binary variables and sufficiency: generalized conditional logit, *Econometrica* **72**, 1859–1876.

Magnus, J.R., 1982, Multivariate error components analysis of linear and nonlinear regression models by maximum likelihood, *Journal of Econometrics* **19**, 239–285.

Magnus, J.R. and A.D. Woodland, 1988, On the maximum likelihood estimation of multivariate regression models containing serially correlated error components, *International Economic Review* **29**, 707–725.

Mairesse, J., 1990, Time-series and cross-sectional estimates on panel data: Why are they different and why should they be equal?, in J. Hartog, G. Ridder and J. Theeuwes, eds., *Panel Data and Labor Market Studies* (North-Holland, Amsterdam), 81–95.

Manski, C.F., 1987, Semiparametric analysis of random effects linear models from binary panel data, *Econometrica* **55**, 357–362.

Mark, N., M. Ogaki and D. Sul, 2000, Dynamic seemingly unrelated cointegrating regression, Working Paper, University of Auckland, Auckland, New Zealand.

Mátyás, L. and L. Lovrics, 1990, Small sample properties of simultaneous error components models, *Economics Letters* **32**, 25–34.

Mátyás, L. and L. Lovrics, 1991, Missing observations and panel data: A Monte Carlo analysis, *Economics Letters* **37**, 39–44.

Mátyás, L. and P. Sevestre, eds., 1996, *The Econometrics of Panel Data*: *A Handbook of the Theory With Applications* (Kluwer Academic Publishers, Dordrecht).

Mazodier, P. and A. Trognon, 1978, Heteroskedasticity and stratification in error components models, *Annales de l'INSEE* **30–31**, 451–482.

McCoskey, S. and C. Kao, 1998, A residual-based test of the null of cointegration in panel data, *Econometric Reviews* **17**, 57–84.

McCoskey, S. and C. Kao, 1999, Testing the stability of a production function with urbanization as a shift factor: An application of non-stationary panel data techniques, *Oxford Bulletin of Economics and Statistics* **61**, 671–690.

McCoskey, S. and T. Selden, 1998, Health care expenditures and GDP: Panel data unit root test results, *Journal of Health Economics* **17**, 369–376.

McElroy, M.B., 1977, Weaker MSE criteria and tests for linear restrictions in regression models with non-spherical disturbances, *Journal of Econometrics* **6**, 389–394.

McFadden, D.J., 1989, A method of simulated moments for estimation of discrete response models without numerical integration, *Econometrica* **57**, 995–1026.

McKenzie, D.J., 2001, Estimation of AR(1) models with unequally spaced pseudo-panels, *Econometrics Journal*, **4**, 89–108.

McKenzie, D.J., 2004, Asymptotic theory for heterogeneous dynamic pseudo-panels, *Journal of Econometrics* **120**, 235–262.

Meghir, C. and F. Windmeijer, 1999, Moment conditions for dynamic panel data models with multiplicative individual effects in the conditional variance, *Annales D'Économie et de Statistique* **55–56**, 317–330.

Mendelsohn, R., D. Hellerstein, M. Huguenin, R. Unsworth and R. Brazee, 1992, Measuring hazardous waste damages with panel models, *Journal of Environmental Economics and Management* **22**, 259–271.

Metcalf, G.E., 1996, Specification testing in panel data with instrumental variables, *Journal of Econometrics* **71**, 291–307.

Moffitt, R., 1993, Identification and estimation of dynamic models with a time series of repeated cross-sections, *Journal of Econometrics* **59**, 99–123.

Moffitt, R., J. Fitzgerald and P. Gottschalk, 1999, Sample attrition in panel data: The role of selection on observables, *Annales D'Économie et de Statistique* **55–56**, 129–152.

Montmarquette, C. and S. Mahseredjian, 1989, Does school matter for educational achievement? A two-way nested-error components analysis, *Journal of Applied Econometrics* **4**, 181–193.

Moon, H.R. and B. Perron, 2004a, Testing for unit root in panels with dynamic factors, *Journal of Econometrics* **122**, 81–126.

Moon, H.R. and B. Perron, 2004b, An empirical analysis of nonstationarity in panels of exchange rates and interest rates with factors, Working Paper, University of Southern California, Los Angeles, California.

Moon, H.R. and B. Perron, 2004c, Efficient estimation of the SUR cointegration regression model and testing for purchasing power parity, *Econometric Reviews*, forthcoming.

Moon, H.R. and P.C.B. Phillips, 1999, Maximum likelihood estimation in panels with incidental trends, *Oxford Bulletin of Economics and Statistics* **61**, 711–747.

Moon, H.R. and P.C.B. Phillips, 2000, Estimation of autoregressive roots near unity using panel data, *Econometric Theory* **16**, 927–997.

Moon, H.R. and P.C.B. Phillips, 2004, GMM estimation of autoregressive roots near unity with panel data, *Econometrica* **72**, 467–522.

Moon, H.R., B. Perron and P.C.B. Phillips, 2003, Incidental trends and the power of panel unit root tests, Working Paper, University of Southern California, Los Angeles, California.

Moulton, B.R., 1986, Random group effects and the precision of regression estimates, *Journal of Econometrics* **32**, 385–397.

Moulton, B.R., 1987, Diagnostics for group effects in regression analysis, *Journal of Business and Economic Statistics* **5**, 275–282.

Moulton, B.R. and W.C. Randolph, 1989, Alternative tests of the error components model, *Econometrica* **57**, 685–693.

Mundlak, Y., 1961, Empirical production function free of management bias, *Journal of Farm Economics* **43**, 44–56.

Mundlak, Y., 1978, On the pooling of time series and cross-section data, *Econometrica* **46**, 69–85.

Munnell, A., 1990, Why has productivity growth declined? Productivity and public investment, *New England Economic Review*, 3–22.

Murray, C.J. and D.H. Papell, 2000, Testing for unit roots in panels in the presence of structural change with an application to OECD unemployment, *Advances in Econometrics* **15**, 223–238.

Nerlove, M., 1971a, Further evidence on the estimation of dynamic economic relations from a time-series of cross-sections, *Econometrica* **39**, 359–382.

Nerlove, M., 1971b, A note on error components models, *Econometrica* **39**, 383–396.

Nerlove, M., 2000, Growth rate convergence, fact or artifact? An essay on panel data econometrics, Chapter 1 in J. Krishnakumar and E. Ronchetti, eds., *Panel Data Econometrics: Future Directions* (North-Holland, Amsterdam), 3–33.

Nerlove, M., 2002, *Essays in Panel Data Econometrics* (Cambridge University Press, Cambridge).

Nerlove, M. and P. Balestra, 1996, Formulation and estimation of econometric models for panel data, Chapter 1 in L. Mátyás and P. Sevestre, eds., *The Econometrics of Panel Data: A Handbook of the Theory With Applications* (Kluwer Academic Publishers, Dordrecht), 3–22.

Newey, W.K. and R.J. Smith, 2004, Higher order properties of GMM and generalized empirical likelihood estimators, *Econometrica* **72**, 214–253.

Neyman, J. and E.L. Scott, 1948, Consistent estimation from partially consistent observations, *Econometrica* **16**, 1–32.

Nguyen, T.H. and C. Bernier, 1988, Beta and q in a simultaneous framework with pooled data, *Review of Economics and Statistics* **70**, 520–523.

Nickell, S., 1981, Biases in dynamic models with fixed effects, *Econometrica* **49**, 1417–1426.

Nijman, Th.E. and M. Verbeek, 1992, Nonresponse in panel data: The impact of estimates of a life cycle consumption function, *Journal of Applied Econometrics* **7**, 243–257.

O'Connell, P.G.J., 1998, The overvaluation of purchasing power parity, *Journal of International Economics* **44**, 1–19.

Oh, K.Y., 1996, Purchasing power parity and unit roots tests using panel data, *Journal of International Money and Finance* **15**, 405–418.

Orme, C., 1993, A comment on "a simple test for neglected heterogeneity in panel studies", *Biometrics* **49**, 665–667.

Owusu-Gyapong, A., 1986, Alternative estimating techniques for panel data on strike activity, *Review of Economics and Statistics* **68**, 526–531.

Papell, D., 1997, Searching for stationarity: Purchasing power parity under the current float, *Journal of International Economics* **43**, 313–332.

Patterson, H.D. and R. Thompson, 1971, Recovery of inter-block information when block sizes are unequal, *Biometrika* **58**, 545–554.

Pedroni, P., 1999, Critical values for cointegration tests in heterogeneous panels with multiple regressors, *Oxford Bulletin of Economics and Statistics* **61**, 653–678.

Pedroni, P., 2000, Fully modified OLS for heterogeneous cointegrated panels, *Advances in Econometrics* **15**, 93–130.

Pedroni, P., 2001, Purchasing power parity tests in cointegrated panels, *Review of Economics and Statistics* **83**, 727–731.

Pedroni, P., 2004, Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis, *Econometric Theory* **20**, 597–625.

Peracchi, F., 2002, The European Community Household Panel: A review, *Empirical Economics* **27**, 63–90.

Peracchi, F. and F. Welch, 1995, How representative are matched cross-sections? Evidence from the current population survey, *Journal of Econometrics* **68**, 153–179.

Pesaran, M.H., 2003, A simple panel unit root test in the presence of cross section dependence, Working Paper, Trinity College, Cambridge.

Pesaran, M.H., 2004, General diagnostic tests for cross-section dependence in panels, Working Paper, Trinity College, Cambridge.

Pesaran, M.H. and R. Smith, 1995, Estimating long-run relationships from dynamic heterogenous panels, *Journal of Econometrics* **68**, 79–113.

Pesaran, M.H. and Z. Zhao, 1999, Bias reduction in estimating long-run relationships from dynamic heterogeneous panels, Chapter 12 in C. Hsiao, K. Lahiri, L.F. Lee and M.H. Persaran, eds., *Analysis of Panels and Limited Dependent Variable Models* (Cambridge University Press, Cambridge), 297–322.

Pesaran, M.H., Y. Shin and R. Smith, 1999, Pooled mean group estimation of dynamic heterogeneous panels, *Journal of the American Statistical Association* **94**, 621–634.

Pesaran, M.H., R. Smith and K.S. Im, 1996, Dynamic linear models for heterogenous panels, Chapter 8 in L. Mátyás and P. Sevestre, eds., *The Econometrics of Panel Data: A Handbook of the Theory With Applications* (Kluwer Academic Publishers, Dordrecht), 145–195.

Phillips, P.C.B. and B.E. Hansen, 1990, Statistical inference in instrumental variables regression with $I(1)$ processes, *Review of Economic Studies* **57**, 99–125.

Phillips, P.C.B. and H. Moon, 1999, Linear regression limit theory for nonstationary panel data, *Econometrica* **67**, 1057–1111.

Phillips, P.C.B. and H. Moon, 2000, Nonstationary panel data analysis: An overview of some recent developments, *Econometric Reviews* **19**, 263–286.

Phillips, P.C.B. and S. Ouliaris, 1990, Asymptotic properties of residual based tests for cointegration, *Econometrica* **58**, 165–193.

Phillips, P.C.B. and D. Sul, 2003, Dynamic panel estimation and homogeneity testing under cross section dependence, *Econometrics Journal* **6**, 217–259.

Phillips, R.L. 2003, Estimation of a stratified error components model, *International Economic Review* **44**, 501–521.

Pirotte, A., 1999, Convergence of the static estimation toward long run effects of dynamic panel data models, *Economics Letters* **53**, 151–158.

Pliskin, L., 1991, The covariance transformation and the instrumental variables estimator of the fixed effects model, *Oxford Bulletin of Economics and Statistics* **53**, 95–98.

Prucha, I.R., 1984, On the asymptotic efficiency of feasible Aitken estimators for seemingly unrelated regression models with error components, *Econometrica* **52**, 203–207.

Prucha, I.R., 1985, Maximum likelihood and instrumental variable estimation in simultaneous equation systems with error components, *International Economic Review* **26**, 491–506.

Quah, D., 1994, Exploiting cross section variation for unit root inference in dynamic data, *Economics Letters* **44**, 9–19.

Quah, D., 1996, Empirics for economic growth and convergence, *European Economic Review* **40**, 1353–1375.

Randolph, W.C., 1988, A transformation for heteroscedastic error components regression models, *Economics Letters* **27**, 349–354.

Rao, C.R., 1970, Estimation of heteroscedastic variances in linear models, *Journal of the American Statistical Association* **65**, 161–172.

Rao, C.R., 1971a, Estimation of variance and covariance components – MINQUE theory, *Journal of Multivariate Analysis* **1**, 257–275.

Rao, C.R., 1971b, Minimum variance quadratic unbiased estimation of variance components, *Journal of Multivariate Analysis* **1**, 445–456.

Rao, C.R., 1972, Estimation variance and covariance components in linear models, *Journal of the American Statistical Association* **67**, 112–115.

Revankar, N.S., 1979, Error component models with serial correlated time effects, *Journal of the Indian Statistical Association* **17**, 137–160.

Revankar, N.S., 1992, Exact equivalence of instrumental variable estimators in an error component structural system, *Empirical Economics* **17**, 77–84.

Ridder, G., 1990, Attrition in multi-wave panel data, in J. Hartog, G. Ridder and J. Theeuwes, eds., *Panel Data and Labor Market Studies* (North-Holland, Amsterdam), 45–67.

Ridder, G., 1992, An empirical evaluation of some models for non-random attrition in panel data, *Structural Change and Economic Dynamics* **3**, 337–355.

Ridder, G. and T.J. Wansbeek, 1990, Dynamic models for panel data, in R. van der Ploeg, ed., *Advanced Lectures in Quantitative Economics* (Academic Press, New York), 557–582.

Robertson, D. and J. Symons, 1992, Some strange properties of panel data estimators, *Journal of Applied Econometrics* **7**, 175–189.

Robins, P.K. and R.W. West, 1986, Sample attrition and labor supply response in experimental panel data: A study of alternative correction procedures, *Journal of Business and Economic Statistics* **4**, 329–338.

Roy, N., 2002, Is adaptive estimation useful for panel models with heteroscedasticity in the individual specific error component? Some Monte Carlo evidence, *Econometric Reviews* **21**, 189–203.

Roy, S.N., 1957, *Some Aspects of Multivariate Analysis* (John Wiley, New York).

Saikkonen, P., 1991, Asymptotically efficient estimation of cointegrating regressions, *Econometric Theory* **7**, 1–21.

Sala-i-Martin, X., 1996, The classical approach to convergence analysis, *Economic Journal* **106**, 1019–1036.

Sarantis, N. and C. Stewart, 1999, Is the consumption–income ratio stationary? Evidence from panel unit root tests, *Economics Letters* **64**, 309–314.

Schmidt, P., 1983, A note on a fixed effect model with arbitrary interpersonal covariance, *Journal of Econometrics* **22**, 391–393.

Schmidt, P. and R.C. Sickles, 1984, Production frontiers and panel data, *Journal of Business and Economic Statistics* **2**, 367–374.

Schmidt, P., S.C. Ahn and D. Wyhowski, 1992, Comment, *Journal of Business and Economic Statistics* **10**, 10–14.

Searle, S.R., 1971, *Linear Models* (John Wiley, New York).

Searle, S.R., 1987, *Linear Models for Unbalanced Data* (John Wiley, New York).

Searle, S.R. and H.V. Henderson, 1979, Dispersion matrices for variance components models, *Journal of the American Statistical Association* **74**, 465–470.

Sevestre, P., 1999, 1977–1997: Changes and continuities in panel data econometrics, *Annales D'Économie et de Statistique* **55–56**, 15–25.

Sevestre, P. and A. Trognon, 1985, A note on autoregressive error component models, *Journal of Econometrics* **28**, 231–245.

Shim, J.K., 1982, Pooling cross section and time series data in the estimation of regional demand and supply functions, *Journal of Urban Economics* **11**, 229–241.

Sickles, R.C., 1985, A nonlinear multivariate error components analysis of technology and specific factor productivity growth with an application to U.S. airlines, *Journal of Econometrics* **27**, 61–78.

Sickles, R.C. and P. Taubman, 1986, A multivariate error components analysis of the health and retirement study of the elderly, *Econometrica* **54**, 1339–1356.

Smith, R.P., 2000, Estimation and inference with non-stationary panel time-series data, Working Paper (Department of Economics, Birkbeck College, London).

Smith, L.V., S. Leybourne, T.H. Kim and P. Newbold, 2004, More powerful panel data unit root tests with an application to mean reversion in real exchange rates, *Journal of Applied Econometrics* **19**, 147–170.

Staiger, D. and J.H. Stock, 1997, Instrumental variables regression with weak instruments, *Econometrica* **65**, 557–586.

Stock, J. and M. Watson, 1993, A simple estimator of cointegrating vectors in higher order integrated systems, *Econometrica* **61**, 783–820.

Suits, D., 1984, Dummy variables: Mechanics vs. interpretation, *Review of Economics and Statistics* **66**, 177–180.

Swallow, W.H. and L.F. Monahan, 1984, Monte Carlo comparison of ANOVA, MIVQUE, REML, and ML estimators of variance components, *Technometrics* **26**, 47–57.

Swamy, P.A.V.B., 1971, *Statistical Inference in Random Coefficient Regression Models* (Springer-Verlag, New York).

Swamy, P.A.V.B. and S.S. Arora, 1972, The exact finite sample properties of the estimators of coefficients in the error components regression models, *Econometrica* **40**, 261–275.

Swamy, P.A.V.B. and G.S. Tavlas, 2001, Random coefficient models, Chapter 19 in B.H. Baltagi, ed., *A Companion to Theoretical Econometrics* (Blackwell Publishers, Massachusetts), 410–428.

Taub, A.J., 1979, Prediction in the context of the variance-components model, *Journal of Econometrics* **10**, 103–108.

Tauchen, G., 1986, Statistical properties of generalized method of moments estimators of structural parameters obtained from financial market data, *Journal of Business and Economic Statistics* **4**, 397–416.

Taylor, M.P. and L. Sarno, 1998, The behavior of real exchange rates during the post-Bretton Woods period, *Journal of International Economics* **46**, 281–312.

Taylor, W.E., 1980, Small sample considerations in estimation from panel data, *Journal of Econometrics* **13**, 203–223.

Toro-Vizcarrondo, C. and T.D. Wallace, 1968, A test of the mean square error criterion for restrictions in linear regression, *Journal of the American Statistical Association* **63**, 558–572.

Townsend, E.C. and S.R. Searle, 1971, Best quadratic unbiased estimation of variance components from unbalanced data in the one-way classification, *Biometrics* **27**, 643–657.

Toyoda, T., 1974, Use of the Chow test under heteroscedasticity, *Econometrica* **42**, 601–608.

Ullah, A. and N. Roy, 1998, Nonparametric and semiparametric econometrics of panel data, Chapter 17 in A. Ullah and D.E.A. Giles, eds., *Handbook on Applied Economic Statistics* (Marcel Dekker, New York), 579–604.

Valliant, R., 1991, Variance estimation for price indexes from a two-stage sample with rotating panels, *Journal of Business and Economic Statistics* **9**, 409–422.

van der Gaag, J., B.M.S. van Praag, F.F.H. Rutten and W. van de Ven, 1977, Aggregated dynamic demand equations for specialistic-outpatient medical care (estimated from a time-series of cross-sections), *Empirical Economics* **2**, 213–223.

Vella, F. and M. Verbeek, 1998, Whose wages do unions raise? A dynamic model of unionism and wage determination for young men, *Journal of Applied Econometrics* **13**, 163–168.

Vella, F. and M. Verbeek, 1999, Two-step estimation of panel data models with censored endogenous variables and selection bias, *Journal of Econometrics* **90**, 239–263.

Verbeek, M., 1990, On the estimation of a fixed effects model with selectivity bias, *Economics Letters* **34**, 267–270.

Verbeek, M., 1996, Pseudo panel data, Chapter 11 in L. Mátyás and P. Sevestre, eds., *The Econometrics of Panel Data: A Handbook of the Theory With Applications* (Kluwer Academic Publishers, Dordrecht), 280–292.

Verbeek, M. and Th.E. Nijman, 1992a, Testing for selectivity bias in panel data models, *International Economic Review* **33**, 681–703.

Verbeek, M. and Th.E. Nijman, 1992b, Can cohort data be treated as genuine panel data?, *Empirical Economics* **17**, 9–23.

Verbeek, M. and Th.E. Nijman, 1993, Minimum MSE estimation of a regression model with fixed effects and a series of cross-sections, *Journal of Econometrics* **59**, 125–136.

Verbeek, M. and Th.E. Nijman, 1996, Incomplete panels and selection bias, Chapter 18 in L. Mátyás and P. Sevestre, eds., *The Econometrics of Panel Data: A Handbook of the Theory With Applications* (Kluwer Academic Publishers, Dordrecht), 449–490.

Verbeek, M. and F. Vella, 2004, Estimating dynamic models from repeated cross-sections, *Journal of Econometrics*, forthcoming.

Verbon, H.A.A., 1980, Testing for heteroscedasticity in a model of seemingly unrelated regression equations with variance components (SUREVC), *Economics Letters* **5**, 149–153.

Wagner, G.G., R.V. Burkhauser and F. Behringer, 1993, The English public use file of the German socio-economic panel, *The Journal of Human Resources* **28**, 429–433.

Wallace, T.D., 1972, Weaker criteria and tests for linear restrictions in regression, *Econometrica* **40**, 689–698.

Wallace, T.D. and A. Hussain, 1969, The use of error components models in combining cross-section and time-series data, *Econometrica* **37**, 55–72.

Wan, G.H., W.E. Griffiths and J.R. Anderson, 1992, Using panel data to estimate risk effects in seemingly unrelated production functions, *Empirical Economics* **17**, 35–49.

Wansbeek, T.J., 1992, Transformations for panel data when the disturbances are autocorrelated, *Structural Change and Economic Dynamics* **3**, 375–384.

Wansbeek, T.J., 2001, GMM estimation in panel data models with measurement error, *Journal of Econometrics* **104**, 259–268.

Wansbeek, T.J. and P. Bekker, 1996, On IV, GMM and ML in a dynamic panel data model, *Economics Letters* **51**, 145–152.

Wansbeek, T.J. and A. Kapteyn, 1982a, A class of decompositions of the variance–covariance matrix of a generalized error components model, *Econometrica* **50**, 713–724.

Wansbeek, T.J. and A. Kapteyn, 1982b, A simple way to obtain the spectral decomposition of variance components models for balanced data, *Communications in Statistics* **A11**, 2105–2112.

Wansbeek, T.J. and A. Kapteyn, 1983, A note on spectral decomposition and maximum likelihood estimation of ANOVA models with balanced data, *Statistics and Probability Letters* **1**, 213–215.

Wansbeek, T.J. and A. Kapteyn, 1989, Estimation of the error components model with incomplete panels, *Journal of Econometrics* **41**, 341–361.

Wansbeek, T.J. and T. Knapp, 1999, Estimating a dynamic panel data model with heterogeneous trends, *Annales D'Économie et de Statistique* **55–56**, 331–349.

Wansbeek, T.J. and R.H. Koning, 1991, Measurement error and panel data, *Statistica Neerlandica* **45**, 85–92.

White, H., 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica* **48**, 817–838.

White, H., 1984, *Asymptotic Theory for Econometricians* (Academic Press, New York).

White, H., 1986, Instrumental variables analogs of generalized least squares estimators, in R.S. Mariano, ed., *Advances in Statistical Analysis and Statistical Computing*, Vol. 1 (JAI Press, New York), 173–277.

Windmeijer, F., 2005, A finite sample correction for the variance of linear efficient two-step GMM estimators, *Journal of Econometrics* **126**, 25–51.

Winkelmann, L. and R. Winkelmann, 1998, Why are the unemployed so unhappy? Evidence from panel data, *Economica* **65**, 1–15.

Winkelmann, R., 2000, *Econometric Analysis of Count Data* (Springer-Verlag, Berlin).

Wolpin, K.I., 1980, A time series–cross section analysis of international variation in crime and punishment, *Review of Economics and Statistics* **62**, 417–421.

Wooldridge, J.M., 1995, Selection corrections for panel data models under conditional mean independence assumptions, *Journal of Econometrics* **68**, 115–132.

Wooldridge, J.M., 1997, Multiplicative panel data models without the strict exogeneity assumption, *Econometric Theory* **13**, 667–678.

Wooldridge, J.M., 1999, Distribution-free estimation of some nonlinear panel data models, *Journal of Econometrics* **90**, 77–97.

Wooldridge, J.M., 2000, A framework for estimating dynamic, unobserved effects panel data models with possible feedback to future explanatory variables, *Economics Letters* **68**, 245–250.

Wooldridge, J.M., 2002, *Econometric Analysis of Cross-Section and Panel Data* (MIT Press, Massachusetts).

Wu, J.L., 2000, Mean reversion of the current account: Evidence from the panel data unit-root test, *Economics Letters* **66**, 215–222.

Wu, J. and S. Chen, 2001, Mean reversion of interest rates in the eurocurrency market, *Oxford Bulletin of Economics and Statistics* **63**, 459–473.

Wu, S. and Y. Yin, 1999, Tests for cointegration in heterogeneous panel: A Monte Carlo study, Working Paper, Department of Economics, State University of New York at Buffalo.

Wu, Y., 1996, Are real exchange rates nonstationary? Evidence from a panel data set, *Journal of Money, Credit and Banking* **28**, 54–63.

Wyhowski, D.J., 1994, Estimation of a panel data model in the presence of correlation between regressors and a two-way error component, *Econometric Theory* **10**, 130–139.

Yin, Y. and S. Wu, 2000, Stationarity tests in heterogeneous panels, *Advances in Econometrics* **15**, 275–296.

Zabel, J., 1992, Estimating fixed and random effects models with selectivity, *Economics Letters* **40**, 269–272.

Zhang, W. and L.F. Lee, 2004, Simulation estimation of dynamic discrete choice panel models with accelerated importance samplers, *Econometrics Journal* **7**, 120–142.

Zellner, A., 1962, An efficient method of estimating seemingly unrelated regression and tests for aggregation bias, *Journal of the American Statistical Association* **57**, 348–368.

Ziemer, R.F. and M.E. Wetzstein, 1983, A Stein-rule method for pooling data, *Economics Letters* **11**, 137–143.

Ziliak, J.P., 1997, Efficient estimation with panel data when instruments are predetermined: An empirical comparison of moment-condition estimators, *Journal of Business and Economic Statistics* **15**, 419–431.

Ziliak, J.P. and T.J. Kniesner, 1998, The importance of sample attrition in life cycle labor supply estimation, *Journal of Human Resources* **33**, 507–530.

Index