Test (2007) 16: 1–22
DOI 10.1007/s11749-007-0046-x

INVITED PAPER

Panel data analysis—advantages and challenges

Cheng Hsiao

Published online: 16 March 2007

Abstract We explain the proliferation of panel data studies in terms of (i) data availability, (ii) the more heightened capacity for modeling the complexity of human behavior than a single cross-section or time series data can possibly allow, and (iii) challenging methodology. Advantages and issues of panel data modeling are also discussed.

Keywords Panel data · Longitudinal data · Unobserved heterogeneity · Random effects · Fixed effects

Mathematics Subject Classification (2000) 62-02

1 Introduction

Panel data or longitudinal data typically refer to data containing time series observations of a number of individuals. Therefore, observations in panel data involve at least two dimensions; a cross-sectional dimension, indicated by subscript i, and a time series dimension, indicated by subscript t. However, panel data could have a more complicated clustering or hierarchical structure. For instance, variable y may

This invited paper is discussed in the comments available at:
http://dx.doi.org/10.1007/s11749-007-0047-9, http://dx.doi.org/10.1007/s11749-007-0048-8,
http://dx.doi.org/10.1007/s11749-007-0049-7, http://dx.doi.org/10.1007/s11749-007-0050-1,
http://dx.doi.org/10.1007/s11749-007-0051-0, http://dx.doi.org/10.1007/s11749-007-0052-z,
http://dx.doi.org/10.1007/s11749-007-0053-y, http://dx.doi.org/10.1007/s11749-007-0054-x.

C. Hsiao (✉)
Department of Economics, University of Southern California, Los Angeles, USA
e-mail: chsiao@usc.edu

C. Hsiao
Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, China

be the measurement of the level of air pollution at station ℓ in city j of country i at time t (e.g., Antweiler 2001; Davis 2002). For ease of exposition, I shall confine my presentation to a balanced panel involving N cross-sectional units, $i = 1, \ldots, N$, over T time periods, $t = 1, \ldots, T$.

There is a proliferation of panel data studies, be it methodological or empirical. In 1986, when Hsiao's (1986) first edition of *Panel Data Analysis* was published, there were 29 studies listing the key words: "panel data or longitudinal data", according to Social Sciences Citation index. By 2004 there were 687, and by 2005 there were 773. The growth of applied studies and the methodological development of new econometric tools of panel data have been simply phenomenal since the seminal paper of Balestra and Nerlove (1966).

There are at least three factors contributing to the geometric growth of panel data studies: (i) data availability, (ii) greater capacity for modeling the complexity of human behavior than a single cross-section or time series data, and (iii) challenging methodology. In what follows we shall briefly elaborate each of these one by one. However, it is impossible to do justice to the vast literature on panel data. For further reference, see Arellano (2003), Baltagi (2001), Hsiao (2003), Mátyás and Sevestre (1996), and Nerlove (2002), etc.

2 Data availability

The collection of panel data is obviously much more costly than the collection of cross-sectional or time series data. However, panel data have become widely available in both developed and developing countries.

The two most prominent panel data sets in the US are the National Longitudinal Surveys of Labor Market Experience (NLS) and the University of Michigan's Panel Study of Income Dynamics (PSID). The NLS began in the mid 1960's. It contains five separate annual surveys covering distinct segments of the labor force with different spans: men whose ages were 45 to 59 in 1966, young men 14 to 24 in 1966, women 30 to 44 in 1967, young women 14 to 24 in 1968, and youth of both sexes 14 to 21 in 1979. In 1986 the NLS expanded to include annual surveys of the children born to women who participated in the National Longitudinal Survey of Youth 1979. The list of variables surveyed is running into the thousands, with emphasis on the supply side of market.

The PSID began with collection of annual economic information from a representative national sample of about 6,000 families and 15,000 individuals in 1968 and has continued to the present. The data set contains over 5,000 variables (Becketti et al. 1988). In addition to the NLS and PSID data sets, there are many other panel data sets that could be of interest to economists, see Juster (2000).

In Europe many countries have their annual national or more frequent surveys such as the Netherlands Socio-Economic Panel (SEP), the German Social Economics Panel (GSOEP), the Luxembourg Social Panel (PSELL), the British Household Panel Survey (BHS), etc. Starting in 1994, the National Data Collection Units (NDUS) of the Statistical Office of the European Committees have been coordinating and linking existing national panels with centrally designed multi-purpose annual longitudinal surveys. The European Community Household Panel (ECHP) is published in Eurostat's reference data base New Cronos in three domains: health, housing, and income and living conditions.

Panel data have also become increasingly available in developing countries. In these countries there may not have been a long tradition of statistical collection. It is of special importance to obtain original survey data to answer many significant and important questions. Many international agencies have sponsored and helped to design panel surveys. For instance, the Dutch non-government organization (NGO), ICS, Africa, collaborated with the Kenya Ministry of Health to carry out a Primary School Deworming Project (PDSP). The project took place in Busia district, a poor and densely-settled farming region in western Kenya. The 75 project schools included nearly all rural primary schools in this area, with over 30,000 enrolled pupils between the ages of six to eighteen from 1998–2001. Another example is the Development Research Institute of the Research Center for Rural Development of the State Council of China, in collaboration with the World Bank, which undertook an annual survey of 200 large Chinese township and village enterprises from 1984 to 1990.

3 Advantages of panel data

Panel data, by blending the inter-individual differences and intra-individual dynamics, have several advantages over cross-sectional or time-series data:

(i) More accurate inference of model parameters. Panel data usually contain more degrees of freedom and more sample variability than cross-sectional data which may be viewed as a panel with $T = 1$, or time series data which is a panel with $N = 1$, hence improving the efficiency of econometric estimates (e.g., Hsiao et al. 1995).

(ii) Greater capacity for capturing the complexity of human behavior than a single cross-section or time series data. These include:

- (ii.a) Constructing and testing more complicated behavioral hypotheses. For instance, consider the example of Ben-Porath (1973) that a cross-sectional sample of married women was found to have an average yearly labor-force participation rate of 50 percent. These could be the outcome of random draws from a homogeneous population or could be draws from heterogeneous populations in which 50% were from the population who always work and 50% who never work. If the sample was from the former, each woman would be expected to spend half of her married life in the labor force and half out of the labor force. The job turnover rate would be expected to be frequent and the average job duration would be about two years. If the sample was from the latter, there is no turnover. The current information about a woman's work status is a perfect predictor of her future work status. A cross-sectional data is not able to distinguish between these two possibilities, but panel data can because the sequential observations for a number of women contain information about their labor participation in different subintervals of their life cycle.

Another example is the evaluation of the effectiveness of social programs (e.g., Heckman et al. 1998; Hsiao et al. 2006; Rosenbaum and Rubin 1985). Evaluating the effectiveness of certain programs using cross-sectional sample typically suffers from the fact that those receiving treatment are different from those without. In other words, one does not simultaneously observe what happens to an individual when she receives treatment or when she does not. An individual is observed as either receiving treatment or not receiving treatment. Using the difference between the treatment group and the control group could suffer from two sources of biases: selection bias due to differences in observable factors between the treatment and control groups, and selection bias due to endogeneity of participation in treatment. For instance, Northern Territory (NT) in Australia decriminalized possession of small amount of marijuana in 1996. Evaluating the effects of decriminalization on marijuana smoking behavior by comparing the differences between NT and other states that were still non-decriminalized could suffer from either or both sorts of bias. If panel data over this time period were available, it would allow the possibility of observing the before- and affect-effects on individuals of decriminalization as well as the possibility of isolating the effects of treatment from other factors affecting the outcome.

(ii.b) Controlling the impact of omitted variables. It is frequently argued that the real reason one finds (or does not find) certain effects is due to ignoring the effects of certain variables in one's model specification which are correlated with the included explanatory variables. Panel data contain information on both the intertemporal dynamics and the individuality of the entities and may allow one to control the effects of missing or unobserved variables. For instance, MaCurdy's (1981) life-cycle labor supply model under certainty implies that, because the logarithm of a worker's hours worked is a linear function of the logarithm of her wage rate and the logarithm of worker's marginal utility of initial wealth, leaving out the logarithm of the worker's marginal utility of initial wealth from the regression of hours worked on wage rate because it is unobserved can lead to seriously biased inference on the wage elasticity on hours worked since initial wealth is likely to be correlated with wage rate. However, since a worker's marginal utility of initial wealth stays constant over time, if time series observations of an individual are available, one can take the difference of a worker's labor supply equation over time to eliminate the effect of marginal utility of initial wealth on hours worked. The rate of change of an individual's hours worked now depends only on the rate of change of her wage rate. It no longer depends on her marginal utility of initial wealth.

(ii.c) Uncovering dynamic relationships. "Economic behavior is inherently dynamic so that most econometrically interesting relationships are explicitly or implicitly dynamic" (Nerlove 2002). However, the estimation of time-adjustment pattern using time series data often has to rely on arbitrary prior restrictions such as Koyck or Almon distributed lag models because time series observations of current and lagged variables are likely to be highly collinear (e.g., Griliches 1967). With panel data we can rely on the inter-individual differences to reduce the collinearity between current and lag variables to estimate unrestricted time-adjustment patterns (e.g., Pakes and Griliches 1984).

(ii.d) Generating more accurate predictions for individual outcomes by pooling the data rather than generating predictions of individual outcomes using the data on the individual in question. If individual behaviors are similar, conditional on certain variables, panel data provide the possibility of learning an individual's behavior by observing the behavior of others. Thus, it is possible to obtain a more accurate description of an individual's behavior by supplementing observations of the individual in question with data on other individuals (e.g., Hsiao et al. 1993, 1989).

(ii.e) Providing micro foundations for aggregate data analysis. Aggregate data analysis often invokes the "representative agent" assumption. However, if micro units are heterogeneous, not only can the time series properties of aggregate data be very different from those of disaggregate data (e.g., Granger 1990; Lewbel 1994; Pesaran 2003), but policy evaluation based on aggregate data may be grossly misleading. Furthermore, the prediction of aggregate outcomes using aggregate data can be less accurate than the prediction based on micro-equations (e.g., Hsiao et al. 2005). Panel data containing time series observations for a number of individuals is ideal for investigating the "homogeneity" versus "heterogeneity" issue.

(iii) Simplifying computation and statistical inference. Panel data involve at least two dimensions: a cross-sectional dimension and a time series dimension. Under normal circumstances one would expect that the computation of panel data estimator or inference would be more complicated than cross-sectional or time series data. However, in certain cases, the availability of panel data actually simplifies computation and inference. For instance:

(iii.a) Analysis of nonstationary time series. When time series data are not stationary, the large sample approximation of the distributions of the least-squares or maximum likelihood estimators are no longer normally distributed, (e.g., Anderson 1959; Dickey and Fuller 1979; Dickey and Fuller 1981; Phillips and Durlauf 1986). But if panel data are available, and observations among cross-sectional units are independent, then one can invoke the central limit theorem across cross-sectional units to show that the limiting distributions of many estimators remain asymptotically normal (e.g., Binder et al. 2005; Levin et al. 2002; Im et al. 2003; Phillips and Moon 1999).

(iii.b) Measurement errors. Measurement errors can lead to under-identification of an econometric model (e.g., Aigner et al. 1984). The availability of multiple observations for a given individual or at a given time may allow a researcher to make different transformations to induce different and deducible changes in the estimators; hence, to identify an otherwise unidentified model (e.g., Biørn 1992; Griliches and Hausman 1986; Wansbeek and Koning 1989).

(iii.c) Dynamic Tobit models. When a variable is truncated or censored, the actual realized value is unobserved. If an outcome variable depends on previous realized value and the previous realized value is unobserved, one has to take integration over the truncated range to obtain the likelihood of the observables. In a dynamic framework with multiple missing values, multiple integration is computationally unfeasible. With panel data the problem can be simplified by only focusing on the subsample in which previous realized values are observed (e.g., Arellano et al. 1999).

4 Methodology

Standard statistical methodology is based on the assumption that the outcomes, say $\underset{\sim}{y}$, conditional on certain variables, say $\underset{\sim}{x}$, are random draws from a probability distribution that is characterized by a fixed dimensional parameter vector, $\underset{\sim}{\theta}$, $f(\underset{\sim}{y} \mid \underset{\sim}{x}; \underset{\sim}{\theta})$. For instance, the standard linear regression model assumes that $f(\underset{\sim}{y} \mid \underset{\sim}{x}; \underset{\sim}{\theta})$ takes the form that

$$E(y \mid \underset{\sim}{x}) = \alpha + \underset{\sim}{\beta}' \underset{\sim}{x}, \tag{1}$$

and

$$\mathrm{Var}(y \mid \underset{\sim}{x}) = \sigma^2, \tag{2}$$

where $\underset{\sim}{\theta}' = (\alpha, \underset{\sim}{\beta}', \sigma^2)$. Typical panel data focuses on individual outcomes. Factors affecting individual outcomes are numerous. It is rare to be able to assume a common conditional probability density function of y conditional on $\underset{\sim}{x}$ for all cross-sectional units, i, at all time, t. For instance, suppose that in addition to $\underset{\sim}{x}$, individual outcomes are also affected by unobserved individual abilities (or marginal utility of initial wealth as in MaCurdy (1981) labor supply model discussed in (ii.b) on Sect. 3), represented by α_i, so that the observed $(y_{it}, \underset{\sim}{x}_{it})$, $i = 1, \ldots, N$, $t = 1, \ldots, T$, are actually generated by

$$y_{it} = \alpha_i + \underset{\sim}{\beta}' \underset{\sim}{x}_{it} + u_{it}, \quad \begin{matrix} i = 1, \ldots, N, \\ t = 1, \ldots, T, \end{matrix} \tag{3}$$

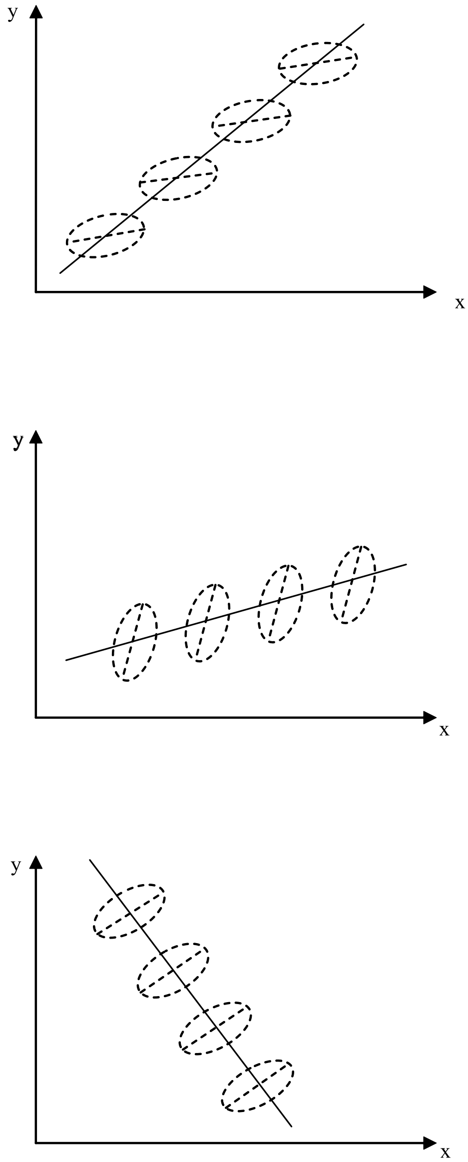
as depicted by Fig. 1 in which the broken-line ellipses represent the point scatter of individual observations around their respective mean, represented by the broken straight lines. If an investigator mistakenly were to impose the homogeneity assumption (1), (2), the solid lines in those figures would represent the estimated relationships between y and $\underset{\sim}{x}$, which can be grossly misleading.

If the conditional density of y given $\underset{\sim}{x}$ varies across i and over t, the fundamental theorems for statistical inference, the laws of large numbers and central limit theorems, will be difficult to implement. One way to restore homogeneity across i and/or over t is to add more conditional variables, say $\underset{\sim}{z}$,

$$f(y_{it} \mid \underset{\sim}{x}_{it}, \underset{\sim}{z}_{it}; \underset{\sim}{\theta}). \tag{4}$$

However, the dimension of $\underset{\sim}{z}$ can be large. A model is a simplification of reality, not a mimic of reality. The inclusion of $\underset{\sim}{z}$ may confuse the fundamental relationship between y and $\underset{\sim}{x}$, in particular, when there is a shortage of degrees of freedom or multicollinearity, etc. Moreover, $\underset{\sim}{z}$ may not be observable. If an investigator is only

Fig. 1 Scatter diagrams of $(y(i, t), x(i, t))$



interested in the relationship between y and $\underset{\sim}{x}$, one approach to characterize the heterogeneity not captured by $\underset{\sim}{x}$ is to assume that the parameter vector varies across i and over t, $\underset{\sim}{\theta}_{it}$, so that the conditional density of y given $\underset{\sim}{x}$ takes the form $f(y_{it} \mid \underset{\sim}{x}_{it}; \underset{\sim}{\theta}_{it})$. However, without a structure being imposed on $\underset{\sim}{\theta}_{it}$, such a model only has descriptive value. It is not possible to draw any inference about $\underset{\sim}{\theta}_{it}$.

The methodological literature on panel data is to suggest possible structures on $\underset{\sim}{\theta}_{it}$ (e.g. Hsiao 2003). One way to impose some structure on $\underset{\sim}{\theta}_{it}$ is to decompose $\underset{\sim}{\theta}_{it}$ into $(\underset{\sim}{\beta}, \underset{\sim}{\gamma}_{it})$, where $\underset{\sim}{\beta}$ is the same across i and over t, referred to as *structural parameters*, and $\underset{\sim}{\gamma}_{it}$, referred to as *incidental parameters*, because when cross-section units, N, and/or time series observations, T, increases, so does the dimension of $\underset{\sim}{\gamma}_{it}$. The focus of panel data literature is to make inference on $\underset{\sim}{\beta}$ after controlling the impact of $\underset{\sim}{\gamma}_{it}$.

Without imposing a structure for $\underset{\sim}{\gamma}_{it}$, again it is difficult to make any inference on $\underset{\sim}{\beta}$ because estimation of $\underset{\sim}{\beta}$ can depend on γ_{it}, and the estimation of the unknown $\underset{\sim}{\gamma}_{it}$ will probably exhaust all available sample information. Assuming that the impacts of the observable variables, $\underset{\sim}{x}$, are the same across i and over t, represented by the structure parameters, $\underset{\sim}{\beta}$, the incidental parameters, $\underset{\sim}{\gamma}_{it}$, represent the heterogeneity across i and over t that are not captured by $\underset{\sim}{x}_{it}$. They can be considered composed of the effects of omitted individual time-invariant, α_i, period individual-invariant, λ_t, and individual time-varying variables, δ_{it}. The individual time-invariant variables are variables that are the same for a given cross-sectional unit through time but vary across cross-sectional units such as individual-firm management, ability, gender, and socio-economic background variables. The period individual-invariant variables are variables that are the same for all cross-sectional units at a given time but vary through time such as prices, interest rates, and wide spread optimism or pessimism. The individual time-varying variables are variables that vary across cross-sectional units at a given point in time and also exhibit variations through time such as firm profits, sales and capital stock. The effects of unobserved heterogeneity can either be assumed as random variables, referred to as the *random effects* model, as fixed parameters, referred to as the *fixed effects* model, or as a mixture of both, refereed to as the *mixed effects* model.

The challenge of panel methodology is to control the impact of unobserved heterogeneity, represented by the incidental parameters, γ_{it}, to obtain valid inference on the structural parameters $\underset{\sim}{\beta}$. A general principle of obtaining valid inference of $\underset{\sim}{\beta}$ in the presence of incidental parameters $\underset{\sim}{\gamma}_{it}$ is to find proper transformation to eliminate $\underset{\sim}{\gamma}_{it}$ from the specification or to integrate out the effects of $\underset{\sim}{\gamma}_{it}$. Since proper transformations depend on the model one is interested, as illustrations, I shall try to demonstrate the fundamental issues from the perspective of linear static models, dynamic models, nonlinear models, models with cross-sectional dependencies and models with large N and large T.

For ease of exposition, I shall assume, for the most of time, that there are no time-specific effects, λ_t, and that the individual time-varying effects, δ_{it}, can be represented by a random variable u_{it}, that is treated as the error of an equation. In other words, only individual-specific effects, α_i, are present. The individual-specific effects, α_i, can be assumed as either random or fixed. The standard assumption for the random effects specification is that these effects are randomly distributed with a common mean and are independent of a fixed $\underset{\sim}{x}_{it}$.

The advantages of the random effects (RE) specification are: (a) The number of parameters stays constant when sample size increases. (b) It allows the derivation of efficient estimators that make use of both within and between (group) variation. (c) It allows the estimation of the impact of time-invariant variables. The disadvantage is that one has to specify a conditional density of α_i given $\underset{\sim}{x}_i' = (\underset{\sim}{x}_{it}, \ldots, \underset{\sim}{x}_{iT})$, $f(\alpha_i \mid \underset{\sim}{x}_i)$, while α_i are unobservable. A common assumption is that $f(\alpha_i \mid \underset{\sim}{x}_i)$ is identical to the marginal density $f(\alpha_i)$. However, if the effects are correlated with $\underset{\sim}{x}_{it}$ or if there is a fundamental difference among individual units, i.e., conditional on

$\underset{\sim}{x}_{it}$, y_{it} cannot be viewed as a random draw from a common distribution, common RE model is misspecified and the resulting estimator is biased.

The advantages of the fixed effects (FE) specification are that it can allow the individual-and/or time specific effects to be correlated with explanatory variables $\underset{\sim}{x}_{it}$. Neither does it require an investigator to model their correlation patterns. The disadvantages of the FE specification are: (a′) The number of unknown parameters increases with the number of sample observations. In the case when T (or N for λ_t) is finite, it introduces the classical incidental parameter problem (e.g., Neyman and Scott 1948). (b′) The FE estimator does not allow the estimation of the coefficients that are time-invariant.

In order words, the advantages of the RE specification are the disadvantages of the FE specification and the disadvantages of the RE specification are the advantages of the FE specification. To choose between the two specifications, Hausman (1978) notes that the FE estimator (or GMM), $\underset{\sim}{\theta}_{\mathrm{FE}}$, is consistent, whenever α_i is fixed or random, and the commonly used RE estimator (or GLS), $\underset{\sim}{\theta}_{\mathrm{RE}}$, is consistent and efficient only when α_i is indeed uncorrelated with $\underset{\sim}{x}_{it}$ and is inconsistent if α_i is correlated with $\underset{\sim}{x}_{it}$. Therefore, he suggests using the statistic

$$(\underset{\sim}{\theta}_{\mathrm{FE}} - \underset{\sim}{\theta}_{\mathrm{RE}})'\big[\mathrm{Cov}\,(\underset{\sim}{\theta}_{\mathrm{FE}}) - \mathrm{Cov}\,(\underset{\sim}{\theta}_{\mathrm{RE}})\big]^{-}(\underset{\sim}{\theta}_{\mathrm{FE}} - \underset{\sim}{\theta}_{\mathrm{RE}}) \tag{5}$$

to test RE vs FE specification. The statistic (5) is asymptotically chi-square distributed with the degrees of freedom equal to the rank of $[\mathrm{Cov}(\underset{\sim}{\theta}_{\mathrm{FE}}) - \mathrm{Cov}(\underset{\sim}{\theta}_{\mathrm{RE}})]$.

4.1 Linear static models

A widely used panel data model is to assume that the effects of observed explanatory variables, $\underset{\sim}{x}$, are identical across cross-sectional units, i, and over time, t, while the effects of omitted variables can be decomposed into the individual-specific effects, α_i, time-specific effects, λ_t, and individual time-varying effects, $\delta_{it} = u_{it}$, as follows:

$$y_{it} = \underset{\sim}{\beta}'\underset{\sim}{x}_{it} + \alpha_i + \lambda_t + u_{it}, \quad \begin{array}{l} i = 1, \ldots, N, \\ t = 1, \ldots, T. \end{array} \tag{6}$$

In a single equation framework, individual-time effects, u, are assumed random and uncorrelated with $\underset{\sim}{x}$, while α_i and λ_t may or may not be correlated with $\underset{\sim}{x}$. When α_i and λ_t are treated as fixed constants, as coefficients of dummy explanatory variables, $d_{it} = 1$, if the observation corresponds to the ith individual at time t, and 0 otherwise; whether they are correlated with $\underset{\sim}{x}$ is not an issue. On the other hand, when α_i and λ_t are treated as random, they become part of the error term and are typically assumed to be uncorrelated with $\underset{\sim}{x}_{it}$.

For ease of exposition, we shall assume that there are no time-specific effects, i.e., $\lambda_t = 0$ for all t, and u_{it} are independent identically distributed (i.i.d.) across i and over t. Stacking an individual's T time series observations of $(y_{it}, \underset{\sim}{x}'_{it})$ into a vector and a matrix, (6) may alternatively be written as

$$\underset{\sim}{y}_i = X_i\underset{\sim}{\beta} + \underset{\sim}{e}\alpha_i + \underset{\sim}{u}_i, \quad i = 1, \ldots, N, \tag{7}$$

where $\underset{\sim}{y}_i = (y_{i1}, \ldots, y_{iT})'$, $X_i = (\underset{\sim}{x}_{i1}, \ldots, \underset{\sim}{x}_{iT})'$, $\underset{\sim}{u}_i = (u_{i1}, \ldots, u_{iT})'$, and $\underset{\sim}{e}$ is a $T \times 1$ vector of 1's.

Let Q be a $T \times T$ matrix satisfying the condition that $Q\underset{\sim}{e} = \underset{\sim}{0}$. Multiplying (7) by Q yields

$$Q\underset{\sim}{y}_i = QX_i\underset{\sim}{\beta} + Q\underset{\sim}{u}_i, \quad i = 1 \ldots, N. \tag{8}$$

Equation (8) no longer involves α_i. The issue of whether α_i is correlated with $\underset{\sim}{x}_{it}$ or whether α_i should be treated as fixed or random is no longer relevant for (8). Moreover, since X_i is exogenous, $E(QX_i\underset{\sim}{u}_i'Q') = QE(X_i\underset{\sim}{u}_i')Q' = \underset{\sim}{0}$ and $EQ\underset{\sim}{u}_i\underset{\sim}{u}_i'Q' = \sigma_u^2 QQ'$. An efficient estimator of $\underset{\sim}{\beta}$ is the generalized least squares estimator (GLS),

$$\underset{\sim}{\beta} = \left[\sum_{i=1}^{N} X_i'(Q'Q)^- X_i\right]^{-1}\left[\sum_{i=1}^{N} X_i'(Q'Q)^- \underset{\sim}{y}_i\right], \tag{9}$$

where $(Q'Q)^-$ denotes the Moore–Penrose generalized inverse (e.g., Rao 1973).

When $Q = I_T - \frac{1}{T}\underset{\sim}{e}\underset{\sim}{e}'$, Q is idempotent, the Moore–Penrose generalized inverse of $(Q'Q)^-$ is just $Q = I_T - \frac{1}{T}\underset{\sim}{e}\underset{\sim}{e}'$ itself. Multiplying (8) by Q is equivalent to transforming (6) into a model

$$(y_{it} - y_i) = \underset{\sim}{\beta}'(\underset{\sim}{x}_{it} - \underset{\sim}{x}_i) + (u_{it} - u_i), \quad \begin{array}{l} i = 1, \ldots, N, \\ t = 1, \ldots, T, \end{array} \tag{10}$$

where $y_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, $\underset{\sim}{x}_i = \frac{1}{T}\sum_{t=1}^{T} \underset{\sim}{x}_{it}$ and $u_i = \frac{1}{T}\sum_{t=1}^{T} u_{it}$. This transformation is called the *covariance transformation*. The least squares estimator (LS) (or a generalized least squares estimator (GLS)) of (10),

$$\underset{\sim}{\beta}_{cv} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T}(\underset{\sim}{x}_{it} - \underset{\sim}{x}_i)(\underset{\sim}{x}_{it} - \underset{\sim}{x}_i)'\right]^{-1}\left[\sum_{t=1}^{N}\sum_{t=1}^{T}(\underset{\sim}{x}_{it} - \underset{\sim}{x}_i)(y_{it} - y_i)\right], \tag{11}$$

is called the *covariance* estimator or *within* estimator because the estimation of $\underset{\sim}{\beta}$ only makes use of within (group) variation of y_{it} and $\underset{\sim}{x}_{it}$. The covariance estimator of $\underset{\sim}{\beta}$ turns out to be also the least squares estimator of (10). It is the best linear unbiased estimator of $\underset{\sim}{\beta}$ if α_i is treated as fixed and u_{it} is i.i.d.

If α_i is random, transforming (7) into (8) transforms T independent equations (or observations) into $(T - 1)$ independent equations, hence, the covariance estimator is not as efficient as the efficient generalized least squares estimator if $E\alpha_i\underset{\sim}{x}_{it}' = \underset{\sim}{0}'$. When α_i is independent of $\underset{\sim}{x}_{it}$ and is independently, identically distributed across i with mean $\underset{\sim}{0}$ and variance σ_α^2, the best linear unbiased estimator (BLUE) of $\underset{\sim}{\beta}$ is GLS,

$$\underset{\sim}{\beta} = \left[\sum_{i=1}^{N} X_i'V^{-1}X_i\right]^{-1}\left[\sum_{i=1}^{N} X_i'V^{-1}\underset{\sim}{y}_i\right], \tag{12}$$

where $V = \sigma_u^2 I_T + \sigma_\alpha^2\underset{\sim}{e}\underset{\sim}{e}'$, $V^{-1} = \frac{1}{\sigma_u^2}\left[I_T - \frac{\sigma_\alpha^2}{\sigma_u^2 + T\sigma_\alpha^2}\underset{\sim}{e}\underset{\sim}{e}'\right]$. Let $\psi = \frac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}$. The GLS is equivalent to first transforming the data by subtracting a fraction $(1 - \psi^{1/2})$ of individual means y_i and $\underset{\sim}{x}_i$ from their corresponding y_{it} and $\underset{\sim}{x}_{it}$, then regressing $[y_{it} - (1-\psi^{1/2})y_i]$ on $[\underset{\sim}{x}_{it} - (1-\psi^{1/2})\underset{\sim}{x}_i]$ (for details, see Baltagi 2001; Hsiao 2003).

If a variable is time-invariant, like gender dummy $x_{kit} = x_{kis} = x_{ki}$, the covariance transformation eliminates the corresponding variable from the specification. Hence, the coefficients of time-invariant variables cannot be estimated. On the other hand, if α_i is random and uncorrelated with $\underset{\sim}{x}_i$, $\psi \neq 1$, the GLS can still estimate the coefficients of those time-invariant variables.

4.2 Dynamic models

When the regressors of a linear model contains lagged dependent variables, say, of the form (e.g., Balestra and Nerlove 1966)

$$\underset{\sim}{y}_i = \underset{\sim}{y}_{i,-1}\gamma + X_i\underset{\sim}{\beta} + \underset{\sim}{e}\alpha_i + \underset{\sim}{u}_i = Z_i\underset{\sim}{\theta} + \underset{\sim}{e}\alpha_i + \underset{\sim}{u}_i, \quad i = 1,\ldots,N, \tag{13}$$

where $\underset{\sim}{y}_{i,-1} = (y_{i0},\ldots,y_{i,T-1})'$, $Z_i = (\underset{\sim}{y}_{i,-1}, X_i)$ and $\underset{\sim}{\theta} = (\gamma, \underset{\sim}{\beta}')'$. For ease of notation, we assume that y_{i0} are observable. Technically, we can still eliminate the individual-specific effects by multiplying (13) by the transformation matrix Q ($Q\underset{\sim}{e} = \underset{\sim}{0}$),

$$Q\underset{\sim}{y}_i = QZ_i\underset{\sim}{\theta} + Q\underset{\sim}{u}_i. \tag{14}$$

However, because of the presence of lagged dependent variables, $EQZ_i\underset{\sim}{u}_i'Q' \neq 0$ even with the assumption that u_{it} are independent, identically distributed across i and over t. For instance, the covariance transformation matrix $Q = I_T - \frac{1}{T}\underset{\sim}{e}\underset{\sim}{e}'$ transforms (13) into the form

$$(y_{it} - \bar{y}_i) = (y_{i,t-1} - \bar{y}_{i,-1})\gamma + (\underset{\sim}{x}_{it} - \bar{\underset{\sim}{x}}_i)'\underset{\sim}{\beta} + (u_{it} - \bar{u}_i), \quad \begin{array}{l} i = 1,\ldots,N, \\ t = 1,\ldots,T, \end{array} \tag{15}$$

where $\bar{y}_i = \frac{1}{T}\sum_{t=1}^T y_{it}$, $\bar{y}_{i,-1} = \frac{1}{T}\sum_{t=1}^T y_{i,t-1}$ and $\bar{u}_i = \frac{1}{T}\sum_{t=1}^T u_{it}$. Although, $y_{i,t-1}$ and u_{it} are uncorrelated under the assumption of serial independence of u_{it}, the covariance between $\bar{y}_{i,-1}$ and u_{it}, or between $y_{i,t-1}$ and $\bar{u}_i$ is of order $(1/T)$ if $|\gamma| < 1$. Therefore, the covariance estimator of $\underset{\sim}{\theta}$ creates a bias of order $(1/T)$, when $N \to \infty$ (Anderson and Hsiao 1981, 1982; Nickell 1981). Since most panel data contain large N but small T, the magnitude of the bias can not be ignored (e.g., with $T = 10$ and $\gamma = 0.5$, the asymptotic bias is -0.167).

When $EQZ_i\underset{\sim}{u}_i'Q' \neq \underset{\sim}{0}$, one way to obtain a consistent estimator for $\underset{\sim}{\theta}$ is to find instruments W_i that satisfy

$$EW_i\underset{\sim}{u}_i'Q' = \underset{\sim}{0}, \tag{16}$$

and

$$\mathrm{rank}(W_iQZ_i) = k, \tag{17}$$

where k denotes the dimension of $(\gamma, \underset{\sim}{\beta}')'$, then apply the generalized instrumental variable or generalized method of moments estimator (GMM) by minimizing the objective function

$$\left[\sum_{i=1}^{N} W_i(Q\underset{\sim}{y}_i - QZ_i\underset{\sim}{\theta})\right]'\left[\sum_{i=1}^{N} W_i Q\underset{\sim}{u}_i\underset{\sim}{u}_i' Q' W_i'\right]^{-1}\left[\sum_{i=1}^{N} W_i(Q\underset{\sim}{y}_i - Q\underset{\sim}{Z}_i\underset{\sim}{\theta})\right] \tag{18}$$

with respect to $\underset{\sim}{\theta}$ (e.g., Arellano 2003; Ahn and Schmidt 1995; Arellano and Bond 1991; Arellano and Bover 1995). For instance, one may let Q be a $(T-1)\times T$ matrix of the form

$$D = \begin{bmatrix} -1 & 1 & 0 & \cdot & \cdot \\ 0 & -1 & 1 & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & -1 & 1 \end{bmatrix}, \tag{19}$$

then the transformation (14) is equivalent to taking the first difference of (13) over time to eliminate α_i for $t = 2, \ldots, T$,

$$\Delta y_{it} = \Delta y_{i,t-1}\gamma + \Delta\underset{\sim}{x}_{it}'\underset{\sim}{\beta} + \Delta u_{it}, \quad \begin{array}{l} i = 1, \ldots, N, \\ t = 2, \ldots, T, \end{array} \tag{20}$$

where $\Delta = (1 - L)$ and L denotes the lag operator, $Ly_t = y_{t-1}$. Since $\Delta u_{it} = (u_{it} - u_{i,t-1})$ is uncorrelated with $y_{i,t-j}$, for $j \geq 2$, and with $\underset{\sim}{x}_{is}$, for all s, when u_{it} is independently distributed over time and $\underset{\sim}{x}_{it}$ is exogenous, one can let W_i be a $T(T-1)[K+\frac{1}{2}]\times(T-1)$ matrix of the form

$$W_i = \begin{bmatrix} \underset{\sim}{q}_{i2} & \underset{\sim}{0} & \cdot & \cdot \\ \underset{\sim}{0} & \underset{\sim}{q}_{is} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \underset{\sim}{q}_{iT} \end{bmatrix}, \tag{21}$$

where $\underset{\sim}{q}_{it} = (y_{i0}, y_{i1}, \ldots, y_{i,t-2}, \underset{\sim}{x}_i')'$, $\underset{\sim}{x}_i = (\underset{\sim}{x}_{i1}', \ldots, \underset{\sim}{x}_{iT}')'$, and $K = k-1$. Under the assumption that $(\underset{\sim}{y}_i', \underset{\sim}{x}_i')$ are independent, identically distributed across i, the Arellano and Bover (1995) GMM estimator takes the form

$$\begin{aligned}\underset{\sim}{\theta}_{\text{AB,GMM}} = &\left\{\left[\sum_{i=1}^{N} Z_i'D'W_i'\right]\left[\sum_{i=1}^{N} W_iAW_i'\right]^{-1}\left[\sum_{i=1}^{N} W_iDZ_i\right]\right\}^{-1} \\ &\times\left\{\left[\sum_{i=1}^{N} Z_i'D'W_i'\right]\left[\sum_{i=1}^{N} W_iAW_i'\right]^{-1}\left[\sum_{i=1}^{N} W_iD\underset{\sim}{y}_i\right]\right\},\end{aligned} \tag{22}$$

where A is a $(T-1)\times(T-1)$ matrix with 2 on the diagonal elements, -1 on the elements right above and below the diagonal elements and 0 elsewhere.

The GMM estimator has the advantage that it is consistent and asymptotically normally distributed whether α_i is treated as fixed or random because it eliminates α_i from the specification. However, the number of moment conditions increases at the order of T^2 which can create severe downward bias in a finite sample (Ziliak 1997). An alternative is to use a (quasi-)likelihood approach which has the advantage of having a fixed number of orthogonality conditions independent of the sample size. It also has the advantage of making use of all the available sample, hence, may yield a more efficient estimator than (22) (e.g., Hsiao et al. 2002; Binder et al. 2005). However, the likelihood approach has to include the specification of the joint likelihood function of $(y_{i0}, y_{i1}, \ldots, y_{iT})$ (or the conditional likelihood function $(y_{i1}, \ldots, y_{iT} \mid y_{i0})$). Since there is no reason to assume that the data generating process of initial observations, y_{i0}, is different from the rest of y_{it}, the initial y_{i0} depends on previous values of $\underset{\sim}{x}_{i,-j}$ and α_i which are unavailable. Bhargava and Sargan (1983) suggest to circumvent this missing data problem by conditioning y_{i0} on $\underset{\sim}{x}_i$ and α_i if α_i is treated as random. If α_i is treated as a fixed constant, Hsiao et al. (2002) proposes conditioning $(y_{i1} - y_{i0})$ on the first difference of $\underset{\sim}{x}_i$.

4.3 Nonlinear models

When the unobserved individual specific effects, α_i (and/or time-specific effects, λ_t) linearly affect the outcome, y_{it}, one can avoid the consideration of random versus fixed effects specification by eliminating them from the specification through some linear transformation such as the covariance transformation (8) or the first difference transformation (20). However, if α_i affects y_{it} nonlinearly, it is not easy to find a transformation that can eliminate α_i. For instance, consider the following binary choice model where the observed y_{it} takes the value of 1 or 0, depending on the latent response function

$$y_{it}^* = \underset{\sim}{\beta}'\underset{\sim}{x}_{it} + \alpha_i + u_{it}, \tag{23}$$

and

$$y_{it} = \begin{cases} 1, & \text{if } y_{it}^* > 0, \\ 0, & \text{if } y_{it}^* \le 0, \end{cases} \tag{24}$$

where u_{it} are independent, identically distributed with the density function $f(u_{it})$. Let

$$y_{it} = E(y_{it} \mid \underset{\sim}{x}_{it}, \alpha_i) + \epsilon_{it}, \tag{25}$$

then

$$E(y_{it} \mid \underset{\sim}{x}_{it}, \alpha_i) = \int_{-(\underset{\sim}{\beta}'\underset{\sim}{x}_{it}+\alpha_i)}^{\infty} f(u)\, du = \big[1 - F(-\underset{\sim}{\beta}'\underset{\sim}{x}_{it} - \alpha_i)\big]. \tag{26}$$

Since α_i affects $E(y_{it} \mid \underset{\sim}{x}_{it}, \alpha_i)$ nonlinearly, α_i remains after taking successive difference of y_{it},

$$y_{it} - y_{i,t-1} = \big[1 - F(-\underset{\sim}{\beta}'\underset{\sim}{x}_{it} - \alpha_i)\big] - \big[1 - F(-\underset{\sim}{\beta}'\underset{\sim}{x}_{i,t-1} - \alpha_i)\big] + (\epsilon_{it} - \epsilon_{i,t-1}). \tag{27}$$

The likelihood function, conditional on $\underset{\sim}{x}_i$ and α_i, takes the form

$$\Pi_{i=1}^{N} \Pi_{t=1}^{T} \big[F(-\underset{\sim}{\beta}'\underset{\sim}{x}_{it} - \alpha_i)\big]^{1-y_{it}} \big[1 - F(-\underset{\sim}{\beta}'\underset{\sim}{x}_{it} - \alpha_i)\big]^{y_{it}}. \tag{28}$$

If T is large, consistent estimators of $\underset{\sim}{\beta}$ and α_i can be obtained by maximizing (28). If T is finite, there is only limited information about α_i no matter how large is N. The presence of incidental parameters, α_i, violates the regularity conditions for the consistency of the maximum likelihood estimator of $\underset{\sim}{\beta}$.

If $f(\alpha_i \mid \underset{\sim}{x}_i)$ is known and is characterized by a fixed dimensional parameter vector, a consistent estimator of $\underset{\sim}{\beta}$ can be obtained by maximizing the marginal likelihood function

$$\Pi_{i=1}^{N} \int \Pi_{t=1}^{T} \left[F(-\underset{\sim}{\beta}' \underset{\sim}{x}_{it} - \alpha_i)\right]^{1-y_{it}} \left[1 - F(-\underset{\sim}{\beta}' \underset{\sim}{x}_{it} - \alpha_i)\right]^{y_{it}} f(\alpha_i \mid \underset{\sim}{x}_i)\, d\alpha_i. \quad (29)$$

However, maximizing (29) involves T-dimensional integration. Butler and Moffitt (1982); Chamberlain (1984); Heckman (1981), etc., have suggested methods to simplify the computation.

The advantage of the RE specification is that there is no incidental parameter problem. The problem is that $f(\alpha_i \mid \underset{\sim}{x}_i)$ is, in general, unknown. If a wrong $f(\alpha_i \mid \underset{\sim}{x}_i)$ is postulated, maximizing the wrong likelihood function will not yield a consistent estimator of $\underset{\sim}{\beta}$. Moreover, the derivation of the marginal likelihood through multiple integration may be computationally infeasible. The advantage of the FE specification is that there is no need to specify $f(\alpha_i \mid \underset{\sim}{x}_i)$. The likelihood function will be the product of individual likelihood (e.g., (28)) if the errors are i.i.d. The disadvantage is that it introduces incidental parameters.

A general approach of estimating a model involving incidental parameters is to find transformations of the original model into a model that does not involve incidental parameters. Unfortunately, there is no general rule available for nonlinear models. One has to explore the specific structure of a nonlinear model to find such a transformation. For instance, if $f(u)$ in (23) is logistic, then

$$\text{Prob}(y_{it} = 1 \mid \underset{\sim}{x}_{it}, \alpha_i) = \frac{e^{\underset{\sim}{\beta}' \underset{\sim}{x}_{it} + \alpha_i}}{1 + e^{\underset{\sim}{\beta}' \underset{\sim}{x}_{it} + \alpha_i}}. \quad (30)$$

Since, in a logit model, the denominators of $\text{Prob}(y_{it} = 1 \mid \underset{\sim}{x}_{it}, \alpha_i)$ and $\text{Prob}(y_{it} = 0 \mid \underset{\sim}{x}_{it}, \alpha_i)$ are identical, and the numerator of any sequence $\{y_{i1}, \ldots, y_{iT}\}$ with $\sum_{t=1}^{T} y_{it} = s$ is always equal to $\exp(\alpha_i s) \cdot \exp\left\{\sum_{t=1}^{T} (\underset{\sim}{\beta}' \underset{\sim}{x}_{it}) y_{it}\right\}$, the conditional likelihood function, conditional on $\sum_{t=1}^{T} y_{it} = s$, will not involve the incidental parameters α_i. For instance, consider the simple case that $T = 2$, then

$$\text{Prob}(y_{i1} = 1, y_{i2} = 0 \mid y_{i1} + y_{i2} = 1) = \frac{e^{\underset{\sim}{\beta}' \underset{\sim}{x}_{i1}}}{e^{\underset{\sim}{\beta}' \underset{\sim}{x}_{i1}} + e^{\underset{\sim}{\beta}' \underset{\sim}{x}_{i2}}} = \frac{1}{1 + e^{\underset{\sim}{\beta}' \Delta \underset{\sim}{x}_{i2}}}, \quad (31)$$

and

$$\text{Prob}(y_{i1} = 0, y_{i2} = 1 \mid y_{i1} + y_{i2} = 1) = \frac{e^{\underset{\sim}{\beta}' \Delta \underset{\sim}{x}_{i2}}}{1 + e^{\underset{\sim}{\beta}' \Delta \underset{\sim}{x}_{i2}}} \quad (32)$$

(Chamberlain 1980; Hsiao 2003).

This approach works because of the logit structure. In the case when $f(u)$ is unknown, Manski (1987) exploits the latent linear structure of (23) by noting that for a given i,

$$\underset{\sim}{\beta}' \underset{\sim}{x}_{it} \gtreqless \underset{\sim}{\beta}' \underset{\sim}{x}_{i,t-1} \quad \Longleftrightarrow \quad E(y_{it} \mid \underset{\sim}{x}_{it}, \alpha_i) \gtreqless E(y_{i,t-1} \mid \underset{\sim}{x}_{i,t-1}, \alpha_i), \quad (33)$$

and suggests maximizing the objective function

$$H_N(b) = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\operatorname{sgn}(\underset{\sim}{b}'\Delta\underset{\sim}{x}_{it})\Delta y_{it}, \tag{34}$$

where $\operatorname{sgn}(w) = 1$ if $w > 0$, $\operatorname{sgn}(w) = 0$ if $w = 0$, and $\operatorname{sgn}(w) = -1$ if $w < 0$. The advantage of the Manski (1987) maximum score estimator is that it is consistent without the knowledge of $f(u)$. The disadvantage is that (33) holds for any $c\underset{\sim}{\beta}$ where $c > 0$. Only the relative magnitude of the coefficients can be estimated with some normalization rule, say $\parallel \underset{\sim}{\beta} \parallel = 1$. Moreover, the speed of convergence is considerably slower ($N^{1/3}$) and the limiting distribution is quite complicated. Horowitz (1992) and Lee (1999) have proposed modified estimators that improve the speed of convergence and are asymptotically normally distributed.

Other examples of exploiting specific structure of nonlinear models to eliminate the effects of incidental parameters α_i include dynamic discrete choice models (Chamberlain 1993; Honoré and Kyriazidou 2000; Hsiao et al. 2006), symmetrically trimmed least squares estimator for truncated and censored data (Tobit models) (Honoré 1992), sample selection models (or type II Tobit models) (Kyriazidou 1997), etc. However, they often impose very severe restrictions on the data such that not much information of the data can be utilized to obtain parameter estimates. Moreover, there are models such that they do not appear to possess consistent estimators when T is finite.

An alternative to consistent estimators is to consider bias reduced estimators. The advantage of such an approach is that the bias reduced estimators may still allow the use of all the sample information so that, from a mean square error point of view, the bias reduced estimator may still dominate consistent estimators because the latter often have to throw away a lot from a sample, thus, tending to have large variances.

Following an idea of Cox and Reid (1987), Arellano (2001) and Carro (2005) propose deriving the modified MLE by maximizing the modified log-likelihood function

$$L^*(\underset{\sim}{\beta}) = \sum_{i=1}^{N}\ell_i^*\big(\underset{\sim}{\beta}, \hat{\alpha}_i(\underset{\sim}{\beta})\big) - \frac{1}{2}\,\log \ell_{i,d_i d_i}^*\big(\underset{\sim}{\beta}, \hat{\alpha}_i(\underset{\sim}{\beta})\big), \tag{35}$$

where $\ell_i^*(\underset{\sim}{\beta}, \hat{\alpha}_i(\underset{\sim}{\beta}))$ denotes the concentrated log-likelihood function of $\underset{\sim}{y}_i$ after substituting the MLE of α_i in terms of $\underset{\sim}{\beta}$, $\hat{\alpha}_i(\underset{\sim}{\beta})$ (i.e., the solution of $\frac{\partial \log L}{\partial \alpha_i} = 0$ in terms of $\underset{\sim}{\beta}, i = 1, \ldots, N$), into the log-likelihood function, and $\ell_{i,\alpha_i\alpha_i}^*(\underset{\sim}{\beta}, \hat{\alpha}_i(\underset{\sim}{\beta}))$ denotes the second derivative of ℓ_i^* with respect to α_i. The bias correction term is derived by noting that, up to the order of $(1/T)$, the first derivative of ℓ_i^* with respect to $\underset{\sim}{\beta}$ converges to $\frac{1}{2}\frac{E[\ell_{i,\beta\alpha_i\alpha_i}^*(\underset{\sim}{\beta},\alpha_i)]}{E[\ell_{i,\alpha_i\alpha_i}^*(\underset{\sim}{\beta},\alpha_i)]}$. By subtracting the order $(1/T)$ bias from the likelihood function, the modified MLE is biased only up to the order of $(1/T^2)$, without increasing the asymptotic variance.

Monte Carlo experiments conducted by Carro (2005) have shown that, when $T = 8$, the bias of modified MLEs for dynamic probit and logit models are negligible. Another advantage of the Arellano–Carro approach is its generality. For instance, a dynamic logit model with time dummy explanatory variable can not meet the Honoré and Kyriazidou (2000) conditions for generating consistent estimator but can still be estimated by the modified MLE with good finite sample properties.

4.4 Modeling cross-sectional dependence

Most panel studies assume that apart from the possible presence of individual invariant but period varying time specific effects, λ_t, the effects of omitted variables are independently distributed across cross-sectional units. However, economic theory often predicts that agents take actions leading to interdependence among themselves. For example, the prediction that risk averse agents will make insurance contracts allowing them to smooth idiosyncratic shocks implies dependence in consumption across individuals. Ignoring cross-sectional dependence can lead to inconsistent estimators, in particular, when T is finite (e.g., Hsiao and Tahmiscioglu 2005). Unfortunately, contrary to the time series data in which the time label gives a natural ordering and structure, general forms of dependence for cross-sectional dimension are difficult to formulate. Therefore, econometricians have relied on strong parametric assumptions to model cross-sectional dependence. Two approaches have been proposed to model cross-sectional dependence: economic distance or spatial approach and factor approach.

In regional science correlation across cross-section units is assumed to follow a certain spatial ordering, i.e., dependence among cross-sectional units is related to location and distance, in a geographic or more general economic or social network space (e.g., Anselin 1988; Anselin and Griffith 1988; Anselin et al. 2006). A known spatial weights matrix, $W = (w_{ij})$, an $N \times N$ positive matrix in which the rows and columns correspond to the cross-sectional units, is specified to express the prior strength of the interaction between individual (location) i (in the row of the matrix) and individual (location) j (column), w_{ij}. By convention, the diagonal elements, $w_{ii} = 0$. The weights are often standardized so that the sum of each row, $\sum_{j=1}^{N} w_{ij} = 1$.

The spatial weight matrix, W, is often included into a model specification of the dependent variable, of the explanatory variables, or of the error term. For instance, a *spatial lag* model for the $NT \times 1$ variable $\underset{\sim}{y} = (\underset{\sim}{y}'_1, \ldots, \underset{\sim}{y}'_N)'$, $\underset{\sim}{y}_i = (y_{i1}, \ldots, y_{iT})'$, may take the form

$$\underset{\sim}{y} = \rho(W \otimes I_T)\underset{\sim}{y} + X\underset{\sim}{\beta} + \underset{\sim}{u}, \tag{36}$$

where X and $\underset{\sim}{u}$ denote the $NT \times K$ explanatory variables and $NT \times 1$ vector of error terms, respectively, and $\otimes$ denotes the Kronecker product. A *spatial error* model may take the form

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{v}, \tag{37}$$

where $\underset{\sim}{v}$ may be specified as a *spatial autoregressive* form,

$$\underset{\sim}{v} = \theta(W \otimes I_T)\underset{\sim}{v} + \underset{\sim}{u}, \tag{38}$$

or as a *spatial moving average* form,

$$\underset{\sim}{v} = \gamma(W \otimes I_T)\underset{\sim}{u} + \underset{\sim}{u}. \tag{39}$$

The spatial model can be estimated by the instrumental variables (generalized method of moments estimator) or the maximum likelihood method. However, the approach of defining cross-sectional dependence in terms of "economic distance" measure requires that the econometricians have information regarding this "economic distance" (e.g., Conley 1999). Another approach to model cross-sectional dependence is to assume that the error of a model, say, model (37) follows a linear factor model

$$v_{it} = \sum_{j=1}^{r} b_{ij} f_{jt} + u_{it}, \tag{40}$$

where $\underset{\sim}{f}_t = (f_{1t}, \ldots, f_{rt})'$ is a $r \times 1$ vector of random factors, $\underset{\sim}{b}_i' = (b_{i1}, \ldots, b_{ir})$ is a $r \times 1$ nonrandom factor loading coefficients, u_{it} represents the effects of idiosyncratic shocks which is independent of $\underset{\sim}{f}_t$ and is independently distributed across i (e.g., Bai and Ng 2002; Moon and Perron 2004; Pesaran 2004). The conventional time-specific effects model is a special case of (40) when $r = 1$ and $b_i = b_\ell$ for all i and ℓ.

The factor approach requires considerably less prior information than the economic distance approach. Moreover, the number of time-varying factors, r, and factor loading matrix $B = (b_{ij})$ can be empirically identified if both N and T are large. The estimation of a factor loading matrix when N is large may not be computationally feasible. Pesaran (2004) has, therefore, suggested to add cross-sectional means $\bar{y}_t = \frac{1}{N}\sum_{i=1}^{N} y_{it}$, $\bar{\underset{\sim}{x}}_t = \frac{1}{N}\sum_{i=1}^{N} \underset{\sim}{x}_{it}$ as additional regressors with individual-specific coefficients to (37) to filter out cross-sectional dependence. This approach is very appealing because of its simplicity. However, it is not clear how it will perform if N is neither small nor large. Neither is it clear how it can be generalized to nonlinear models.

4.5 Large-N and large-T panels

Our discussion has been mostly focusing on panels with large N and finite T. There are panel data sets, like the Penn–World tables, covering different individuals, industries, and countries over long periods. In general, if an estimator is consistent in the fixed-T, large-N case, it will remain consistent if both N and T tend to infinity. Moreover, even in the case when an estimator is inconsistent for fixed T and large N (say, the MLE of dynamic model (13) or fixed effects probit or logit models (26)), it can become consistent if T also tends to infinity. The probability limit of an estimator, in general, is identical irrespective of how N and T tend to infinity. However, the properly scaled limiting distribution may depend on how the two indexes, N and T, tend to infinity.

There are several approaches for deriving the limits of large-N, large-T panels:

(a) Sequential limits—first, fix one index, say N, and allow the other, say T, to go to infinity, giving an intermediate limit, then let N go to infinity.
(b) Diagonal-path limits—let the two indexes, N and T, pass to infinity along a specific diagonal path, say $T = T(N)$, as $N \longrightarrow \infty$.
(c) Joint limits—let N and T pass to infinity simultaneously without placing specific diagonal path restrictions on the divergence.

In many applications sequential limits are easy to derive. However, sometimes sequential limits can give misleading asymptotic results. A joint limit will give a more robust result than either a sequential limit or a diagonal-path limit, but will also be substantially more difficult to derive and will be applicable only under stronger conditions, such as the existence of higher moments. Phillips and Moon (1999) have given a set of sufficient conditions that ensures that sequential limits are equivalent to joint limits.

When T is large, there is a need to consider serial correlations more generally, including both short-memory and persistent components. For instance, if unit roots are present in y and x (i.e., both are integrated of order 1) but are not cointegrated, Phillips and Moon (1999) show that, if N is fixed but $T \longrightarrow \infty$, the least squares regression of y on x is a nondegenerate random variables that is a functional of Brownian motion that does not converge to the long-run average relation between y and x, but it does if N also tends to infinity. In other words, the issue of spurious regression will not arise in a panel with large N (e.g., Kao 1999).

Both theoretical and applied researchers have paid a great deal of attention to unit root and cointegration properties of variables. When N is finite and T is large, standard time series techniques can be used to derive the statistical properties of panel data estimators. When N is large and cross-sectional units are independently distributed across i, central limit theorems can be invoked along the cross-sectional dimension. Asymptotically normal estimators and test statistics (with suitable adjustment for finite T bias) for unit roots and cointegration have been proposed (e.g., Baltagi and Kao 2000; Im et al. 2003; Levin et al. 2002). They, in general, gain statistical power over their standard time series counterparts (e.g., Choi 2001).

When both N and T are large and cross-sectional units are not independent, a factor analytic framework of the form (40) has been proposed to model cross-sectional dependency and variants of unit root tests are proposed (e.g., Moon and Perron 2004). However, the implementation of those panel unit root tests is quite complicated. When $N \longrightarrow \infty$ and $\frac{1}{N}\sum_{i=1}^{N} u_{it} \longrightarrow 0$, (40) implies that $\bar{v}_t = \bar{\underset{\sim}{b}}' \underset{\sim}{f}_t$, where $\bar{\underset{\sim}{b}}'$ is the cross-sectional average of $\underset{\sim}{b}_i' = (b_{i1}, \ldots, b_{ir})$ and $\underset{\sim}{f}_t = (f_{1t}, \ldots, f_{rt})$. Pesaran (2004, 2005) suggests a simple approach to filter out the cross-sectional dependency by augmenting the cross-sectional means $\bar{y}_t$ and $\bar{\underset{\sim}{x}}_t$ to the regression model (37), i.e.,

$$y_{it} = \underset{\sim}{x}_{it}' \underset{\sim}{\beta} + \alpha_i + \bar{y}_t c_i + \bar{\underset{\sim}{x}}_t' d_i + e_{it}, \tag{41}$$

or $\bar{y}_t$, $\Delta\bar{y}_{t-j}$ to the Dickey and Fuller (1979) type regression model, i.e.,

$$\Delta y_{it} = \alpha_i + \delta_i t + \gamma_i y_{i,t-1} + \sum_{\ell=1}^{p_i} \phi_{i\ell} \Delta y_{i,t-\ell} + c_i \bar{y}_{t-1} + \sum_{\ell=1}^{p_i} d_{i\ell} \Delta \bar{y}_{t-\ell} + e_{it}, \tag{42}$$

for testing of unit root, where $\bar{y}_t = \frac{1}{N}\sum_{i=1}^{N} y_{it}$, $\bar{\underset{\sim}{x}}_t = \frac{1}{N}\sum_{i=1}^{N} \underset{\sim}{x}_{it}$, $\Delta\bar{y}_{t-j} = \frac{1}{N}\sum_{i=1}^{N} \Delta y_{i,t-j}$, $\Delta = (1-L)$, and L denotes the lag operator. The resulting pooled estimator will again be asymptotically normally distributed.

When cross-sectional dependency is of unknown form, Chang (2002) suggests using nonlinear transformations $F(y_{i,t-1})$ of the lagged level variable $y_{i,t-1}$ as instruments (IV) for the usual augmented Dickey and Fuller (1979) type regression.

The test static for the unit root hypothesis is simply defined as a standardized sum of individual IV t-ratios. As long as $F(\cdot)$ is regularly integrable, say $F(y_{i,t-1}) = y_{i,t-1}e^{-c_i|y_{i,t-1}|}$, where c_i is a positive constant, the products of the nonlinear instruments $F(y_{i,t-1})$ and $F(y_{j,t-1})$ from different cross-sectional units i and j are asymptotically uncorrelated, even if the variables $y_{i,t-1}$ and $y_{j,t-1}$, generating the instruments, are correlated. Hence, the usual central limit theorems can be invoked and the standardized sum of individual IV t-ratios is asymptotically normally distributed.

For further review of the literature on unit roots and cointegration in panels, see Breitung and Pesaran (2006) and Choi (2006). However, a more fundamental issue of panel modeling with large N and large T is whether the standard approach of formulating unobserved heterogeneity for the data with finite T remains a good approximation to the true data generating process with large T.

5 Concluding remarks

In this paper we have tried to provide a summary of advantages of using panel data and the fundamental issues of panel data analysis. Assuming that the heterogeneity across cross-sectional units and over time that is not captured by the observed variables can be captured by period-invariant individual specific and/or individual-invariant time specific effects, we surveyed the fundamental methods for the analysis of linear static and dynamic models. We have also discussed difficulties of analyzing nonlinear models and modeling cross-sectional dependence. There are many important issues such as modeling of joint dependence, simultaneous equations models, varying parameter models (e.g., Hsiao 1992, 2003; Hsiao and Pesaran 2006), unbalanced panel, measurement errors (e.g., Griliches and Hausman 1986; Wansbeek and Koning 1989), nonparametric or semiparametric approach, repeated cross-section data, etc., that are not discussed here but are of no less importance.

Although panel data offer many advantages, they are not panacea. The power of panel data to isolate the effects of specific actions, treatments or more general policies depends critically on the compatibility of the assumptions of statistical tools with the data generating process. In choosing a proper method for exploiting the richness and unique properties of the panel, it might be helpful to keep the following factors in mind. First, what advantages do panel data offer us in investigating economic issues over data sets consisting of a single cross section or time series? Second, what are the limitations of panel data and the econometric methods that have been proposed for analyzing such data? Third, what are the assumptions underlying the statistical inference procedures and are they compatible with the data-generating process? Fourth, when using panel data, how can we increase the efficiency of parameter estimates and reliability of statistical inference?

Acknowledgements I would like to thank Irene C. Hsiao for helpful discussion and editorial assistance and Kannika Damrongplasit for drawing the figures. Some of the arguments presented here also appear in Hsiao (2005, 2006).

References

Ahn SC, Schmidt P (1995) Efficient estimation of models for dynamic panel data. J Econom 68:5–27
Aigner DJ, Hsiao C, Kapteyn A, Wansbeek T (1984) Latent variable models in econometrics. In: Griliches Z, Intriligator MD (eds) Handbook of econometrics, vol 2. North-Holland, Amsterdam, pp 1322–1393
Anderson TW (1959) On asymptotic distributions of estimates of parameters of stochastic difference equations. Ann Math Stat 30:676–687
Anderson TW, Hsiao C (1981) Estimation of dynamic models with error components. J Am Stat Assoc 767:598–606
Anderson TW, Hsiao C (1982) Formulation and estimation of dynamic models using panel data. J Econom 18:47–82
Anselin L (1988) Spatial econometrics: methods and models. Kluwer Academic, Dordrecht
Anselin L, Griffith DA (1988) Do spatial effects really matter in regression analysis? Pap Reg Sci Assoc 65:11–34
Anselin L, Le Gallo J, Jayet H (2006) Spatial panel econometrics. In: Mátyás L, Sevestre P (eds) The econometrics of panel data: fundamentals and recent developments in theory, 3rd edn. Kluwer Academic, Dordrecht, Chap 18
Antweiler W (2001) Nested random effects estimation in unbalanced panel data. J Econom 101:295–313
Arellano M (2001) Discrete choice with panel data. Working paper 0101, CEMFI, Madrid
Arellano M (2003) Panel data econometrics. Oxford University Press, Oxford
Arellano M, Bond SR (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Rev Econ Stud 58:277–297
Arellano M, Bover O (1995) Another look at the instrumental variable estimation of error-components models. J Econom 68:29–51
Arellano M, Bover O, Labeaga J (1999) Autoregressive models with sample selectivity for panel data. In: Hsiao C, Lahiri K, Lee LF, Pesaran MH (eds) Analysis of panels and limited dependent variable models. Cambridge University Press, Cambridge, pp 23–48
Bai J, Ng S (2002) Determining the number of factors in approximate factor models. Econometrica 70:91–121
Balestra P, Nerlove M (1966) Pooling cross-section and time series data in the estimation of a dynamic model: the demand for natural gas. Econometrica 34:585–612
Baltagi BH (2001) Econometric analysis of panel data, 2nd edn. Wiley, New York
Baltagi BH, Kao C (2000) Nonstationary panels, cointegration in panels and dynamic panel, a survey. In: Baltagi BH (ed) Nonstationary panels, panel cointegration, and dynamic panels. Advances in econometrics, vol 15. JAI Press, Amsterdam, pp 7–51
Becketti S, Gould W, Lillard L, Welch F (1988) The panel study of income dynamics after fourteen years: an evaluation. J Labor Econ 6:472–492
Ben-Porath Y (1973) Labor force participation rates and the supply of labor. J Political Econ 81:697–704
Bhargava A, Sargan JD (1983) Estimating dynamic random effects models from panel data covering short time periods. Econometrica 51:1635–1659
Binder M, Hsiao C, Pesaran MH (2005) Estimation and inference in short panel vector autoregressions with unit roots and cointegration. Econom Theory 21:795–837
Biørn, E (1992) Econometrics of panel data with measurement errors. In: Mátyás L, Sevestre P (eds) The econometrics of panel data: theory and applications, 1st edn. Kluwer Academic, Dordrecht, pp 152–195
Breitung J, Pesaran MH (2006) Unit roots and cointegration in panels. In: Mátyás L, Sevestre, P (eds) The econometrics of panel data: fundamentals and recent developments in theory, 3rd edn. Kluwer Academic, Dordrecht, Chap 8
Butler JS, Moffitt R (1982) A computationally efficient quadrature procedure for the one factor multinominal probit model. Econometrica 50:761–764
Carro JM (2005) Estimating dynamic panel data discrete choice models with fixed effects. J Econom (in press)
Chamberlain G (1980) Analysis of covariance with qualitative data. Rev Econ Stud 47:225–238
Chamberlain G (1984) Panel data. In: Griliches Z, Intriligator M (eds) Handbook of econometrics, vol 2. North-Holland, Amsterdam, pp 1247–1318
Chamberlain G (1993) Feedback in panel data models. Discussion paper, Department of Economics, Harvard University
Chang Y (2002) Nonlinear IV unit root tests in panels with cross-sectional dependency. J Econom 110:261–292

Choi I (2001) Unit root tests for panel data. J Int Money Financ 20:249–272
Choi I (2006) Nonstationary panels. In: Patterson K, Mills TC (eds) Palgrave handbooks of econometrics, vol 1. MacMillan, New York, pp 511–539
Conley TG (1999) GMM estimation with cross-sectional dependence. J Econom 92:1–45
Cox DR, Reid N (1987) Parameter orthogonality and approximate conditional inference. J Roy Stat Soc Ser B 49:1–39
Davis P (2002) Estimating multi-way error components models with unbalanced panel data structure. J Econom 106(1):67–95. Reprinted in: Baltagi BH, Sumney G (eds) Recent developments in the econometrics of panel data. Edward Elgar, UK
Dickey DA, Fuller WA (1979) Distribution of the estimators for autoregressive time series with a unit root. J Am Stat Assoc 74:427–431
Dickey DA, Fuller WA (1981) Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica 49:1057–1072
Granger CWJ (1990) Aggregation of time-series variables: a survey. In: Barker T, Pesaran MH (eds) Disaggregation in econometric modeling. Routledge, London
Griliches Z (1967) Distributed lags: a survey. Econometrica 35:16–49
Griliches Z, Hausman JA (1986) Errors-in-variables in panel data. J Econom 31:93–118
Hausman JA (1978) Specification tests in econometrics. Econometrica 46:1251–1271
Heckman JJ (1981) Statistical models for discrete panel data. In: Manski CF, McFadden D (eds) Structural analysis of discrete data with econometric applications. MIT Press, Cambridge, pp 114–178
Heckman JJ, Ichimura H, Smith J, Todd P (1998) Characterizing selection bias using experimental data. Econometrica 66:1017–1098
Honoré B (1992) Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects. Econometrica 60:533–567
Honoré B, Kyriazidou E (2000) Panel data discrete choice models with lagged dependent variables. Econometrica 68:839–874
Horowitz JL (1992) A smoothed maximum score estimator for the binary response model. Econometrica 60:505–531
Hsiao C (1986) Analysis of panel data. Econometric society monographs, vol 11. Cambridge University Press, New York
Hsiao C (1992) Random coefficient models. In: Mátyás L, Sevestre P (eds) The econometrics of panel data, 1st edn. Kluwer Academic, Dordrecht, pp 223–241. Reprinted in 2nd edn (1996), pp 410–428
Hsiao C (2003) Analysis of panel data, 2nd edn. Econometric society monographs, vol 34. Cambridge University Press, Cambridge
Hsiao C (2005) Why panel data? Singap Econ Rev 50(2):1–12
Hsiao C (2006) Longitudinal data analysis. In: The new Palgrave dictionary of economics. MacMillan (in press)
Hsiao C, Pesaran MH (2006) Random coefficients models. In: Mátyás L, Sevestre P (eds) The econometrics of panel data: fundamentals and recent developments in theory, 3rd edn. Kluwer Academic, Dordrecht, Chap 5
Hsiao C, Tahmiscioglu T (2005) Estimation of dynamic panel data models with both individual and time specific effects
Hsiao C, Luke Chan MW, Mountain DC, Tsui, KY (1989) Modeling Ontario regional electricity system demand using a mixed fixed and random coefficients approach. Reg Sci Urban Econ 19:567–587
Hsiao C, Appelbe TW, Dineen CR (1993) A general framework for panel data analysis—with an application to Canadian customer dialed long distance service. J Econom 59:63–86
Hsiao C, Mountain DC, Ho-Illman K (1995) Bayesian integration of end-use metering and conditional demand analysis. J Bus Econ Stat 13:315–326
Hsiao C, Pesaran MH, Tahmiscioglu AK (2002) Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. J Econom 109:107–150
Hsiao C, Shen Y, Fujiki H (2005) Aggregate vs disaggregate data analysis—a paradox in the estimation of money demand function of Japan under the low interest rate policy. J Appl Econom 20:579–601
Hsiao C, Shen Y, Wang B, Weeks G (2006) Evaluating the effectiveness of Washington state repeated job search services on the employment rate of prime-age female welfare recipients. Working paper
Im K, Pesaran MH, Shin Y (2003) Testing for unit roots in heterogeneous panels. J Econom 115:53–74
Juster T (2000) Economics/micro data. In: International encyclopedia of social sciences
Kao C (1999) Spurious regression and residual-based tests for cointegration in panel data. J Econom 90:1–44
Kyriazidou E (1997) Estimation of a panel data sample selection model. Econometrica 65:1335–1364

Lee MJ (1999) A root-n-consistent semiparametric estimator for related effects binary response panel data. Econometrica 67:427–433

Levin A, Lin C, Chu J (2002) Unit root tests in panel data: asymptotic and finite-sample properties. J Econom 108:1–24

Lewbel A (1994) Aggregation and simple dynamics. Am Econ Rev 84:905–918

MaCurdy TE (1981) An empirical model of labor supply in a life cycle setting. J Political Econ 89:1059–1085

Manski CF (1987) Semiparametric analysis of random effects linear models from binary panel data. Econometrica 55:357–362

Mátyás L, Sevestre P (1996) The econometrics of panel data—handbook of theory and applications, 2nd edn. Kluwer Academic, Dordrecht

Moon HR, Perron B (2004) Testing for a unit roots in panels with dynamic factors. J Econom 122:81–126

Nerlove M (2002) Essays in panel data econometrics. Cambridge University Press, Cambridge

Neyman J, Scott EL (1948) Consistent estimates based on partially consistent observations. Econometrica 16:1–32

Nickell S (1981) Biases in dynamic models with fixed effects. Econometrica 49:1399–1416

Pakes A, Griliches Z (1984) Estimating distributed lags in short panels with an application to the specification of depreciation patterns and capital stock constructs. Rev Econ Stud 51:243–262

Pesaran MH (2003) On aggregation of linear dynamic models: an application to life-cycle consumption models under habit formation. Econ Model 20:227–435

Pesaran MH (2004) Estimation and inference in large heterogeneous panels with cross-sectional dependence

Pesaran MH (2005) A simple panel unit root test in the presence of cross-section dependence. Working paper 0346, Cambridge University DAE

Phillips PCB, Durlauf SN (1986) Multiple time series regression with integrated processes. Rev Econ Stud 53:473–495

Phillips PCB, Moon HR (1999) Linear regression limit theory for nonstationary panel data. Econometrica 67:1057–1111

Rao CR (1973) Linear statistical inference and its applications, 2nd edn. Wiley, New York

Rosenbaum P, Rubin D (1985) Reducing bias in observational studies using subclassification on the propensity score. J Am Stat Assoc 79:516–524

Wansbeek TJ, Koning RH (1989) Measurement error and panel data. Stat Neerl 45:85–92

Ziliak JP (1997) Efficient estimation with panel data when instruments are predetermined: an empirical comparison of moment-condition estimators. J Bus Econ Stat 15:419–431