Economic Growth Centre
Working Paper Series

Longitudinal Data Analysis

by

Cheng HSIAO

Economic Growth Centre
Division of Economics
School of Humanities and Social Sciences
Nanyang Technological University
Nanyang Avenue
SINGAPORE 639798

Website: http://www.hss.ntu.edu.sg/egc/

Working Paper No: 2005/10

Copies of the working papers are available from the World Wide Web at:

Website: http://www.hss.ntu.edu.sg/egc/

Longitudinal Data Analysis

Cheng Hsiao
Nanyang Technological University, Singapore

and

University of Southern California
Department of Economics
KAP300
3620 S. Vermont Ave.
Los Angeles, CA 90089-0253
email: chsiao@usc.edu

July 28, 2005

1. Why Panel Data?

Panel data, or longitudinal data, refers to data set that contains observations of a number of individuals over time. In other words, it provides multiple observations for each individual in the sample. Compared to the cross-sectional data in which observations for a number of individuals are available only for a given time, or the time series data, in which a single entity is observed over time, panel data has the obvious advantages of having more degrees of freedom and less collinearity among explanatory variables, hence provides the possibility of obtaining more accurate parameter estimates. More importantly, panel data by blending inter-individual differences with intra-individual dynamics, allows the investigation of more complicated behavioral hypotheses than those that can be addressed using cross-sectional or time series data. For instance, standard assumption for the analysis of cross-sectional data is that conditional on certain variables, each woman is a random sample from a homogeneous population. Therefore, if a cross-sectional sample yields an average labor-participation rate of 50 percent for married women, it would imply that each woman has a 50 percent chance of being in the labor force at any given time, hence a married 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 would be frequent, and the average job duration would be expected just two years (Ben-Porath (1973)). However, the cross-sectional data could be drawn from a heterogeneous population in which 50 percent of the sample coming from the population that always work and 50 percent from the population that never work. In this situation, there is no turnover and current work status about a woman is a perfect predictor of her future work status. To discriminate between these two possibilities, we need information on individual labor-force histories in different subintervals of the life cycle, which can only be provided if information on intertemporal dynamics of individual entities are available. On the other hand, although time series data provide information on dynamic adjustment, variables over time tend to move collinearly, hence makes it difficult to identify microdynamic or macrodynamic effects. Often estimation of

distributed lag models has to rely on strong prior restrictions like Koyck or Almon lag with very little empirical justification. With panel data, the interindividual differences often can reduce or lessen the problem of multicollinearity and provide the possibility of estimating unrestricted time adjustment patterns (e.g. Pakes and Griliches (1984)).

By utilizing information on both the intertemporal dynamics and the individuality of the entities, panel data may also allow an investigator to control the effects of missing or unobserved variables. For instance, MaCurdy's (1981) life cycle labor supply of prime-age males under certainty model assumes that the logarithm of hours worked is a linear function of the real wage rate and the logarithm of the worker's marginal utility of initial wealth, which is unobserved. Since wage rate and marginal utility of initial wealth are correlated, any instrument that is correlated with the wage rate will be correlated with the marginal utility of initial wealth. There is no way one can obtain consistent estimate of the coefficient of the wage rate with cross-sectional data. But if panel data are available, one can transform the labor supply model by taking first difference to get rid of the marginal utility of initial wealth as an explanatory variable. The resulting regression can yeild consistent estimates of the coefficient of wage rate and other explanatory varibles.

Panel data may also provide micro foundations for aggregate data analysis. Aggregate data analysis often invokes the "representative agent" assumption. If micro units are heterogeneous, the time series properties of aggregate data may be very different from those of disaggregate data (e.g. Granger (1990), Lewbel (1992, 94), Pesaran (1999)) and policy evaluation based on aggregate data could also be grossly misleading (e.g. Hsiao, Shen and Fujiki (2004)). Panel data by providing time series observations for a number of individuals is ideal for the investigation of homogeneity issue.

Panel data involve observations of two or more dimensions. In normal circumstance, one would expect that the computation and inference of panel data models be more complicated than cross-section or time series data. However, in certain situations, the availability of panel data actually simplify inference. For instance, statistical inference for nonstation-

ary panel data can be complicated (e.g. Phillips (1986)). But, if observations are independently distributed across cross-sectional units, central limit theorems applied across cross-sectional units lead to asymptotically normally distributed statistics (e.g. Levin, Lin and Chu (2002), Pesaran, Shin and Smith (2002)).

2. Issues of Panel Data Analysis

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

$$E(y \mid \underset{\sim}{x}) = \alpha + \underset{\sim}{\beta}' \underset{\sim}{x}. \tag{2.1}$$

and

$$\mathrm{Var}(y \mid \underset{\sim}{x}) = \sigma^2, \tag{2.2}$$

where $\underset{\sim}{\theta}' = (\alpha, \underset{\sim}{\beta}', \sigma^2)$. Panel data, by its nature, focus 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. 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. Blindly imposing a homogeneity assumption of $f(y \mid \underset{\sim}{x}; \underset{\sim}{\theta})$ across i and over t can lead to severely biased inference. For instance, suppose that the data is generated by

$$y_{it} = \alpha_i + \underset{\sim}{\beta}' \underset{\sim}{x}_{it} + v_{it}, \quad \begin{matrix} i = 1, \ldots, N, \\ t = 1, \ldots, T, \end{matrix} \tag{2.3}$$

as depicted by Figure 1 in which the broken-time ellipses represent the point scatter of individual observation around the mean, represented by the broken straight line. If an investigator mistakenly estimate a model of the form

$$y_{it} = \alpha + \underset{\sim}{\beta}' \underset{\sim}{x} + v^*_{it}. \tag{2.4}$$

The solid line in Figure 1 would depict the pooled least squares regression result which could be completely contradict the individual relation between y and $\underset{\sim}{x}$.

One way to restore homogeneity across i and/or over t is to add more conditional variables, say $\underset{\sim}{z}$,

$$f(y_{it} \mid \underset{\sim}{x}_{it}, \underset{\sim}{z}_{it}; \underset{\sim}{\theta}). \tag{2.3}$$

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 interested in the relationship between y and $\underset{\sim}{x}$, a common 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.

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 $\underset{\sim}{\gamma}_{it}$ as *incidental parameters* because when cross-units, N and/or time series observations, T increases, so is 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 structure for $\underset{\sim}{\gamma}_{it}$, again it is not possible to make any inference on $\underset{\sim}{\beta}$ because the unknown $\underset{\sim}{\gamma}_{it}$ will exhaust all available sample information. Assuming that the impacts of 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 as composed of the effects of omitted individual time-invariant, α_i, period individual-invariant, λ_t, and individual time-varying variables, u_{it}. The individuals time-invariant variables are variables that are the same for a given cross-sectional unit through time but that 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 that vary though 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. In a single equation frmaework, it is a common practice to assume that the effects of omitted individual time-varying variables, u_{it} as random and uncorrelated with $\underset{\sim}{x}$. The individual-specific effects, α_i and time specific effects, λ_t can either be assumed as random variables-referred to as the *random effects* model, or fixed parameters-referred to as the *fixed effects* model.

3. 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, u_{it}, as follows:

$$y_{it} = \underset{\sim}{\beta}' \underset{\sim}{x}_{it} + \alpha_i + \lambda_t + u_{it}, \begin{matrix} i = 1, \ldots, N, \\ t = 1, \ldots, T. \end{matrix} \tag{3.1}$$

In a single equation framework. individual time effects, u, are assumed to be uncorrelated with $\underset{\sim}{x}$, while α_i and λ_t may or may not correlated with $\underset{\sim}{x}$. When α_i and λ_t are treated as fixed constants, they are parameters to be estimated so 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 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 independently, identically distributed (i.i.d) across i and over t. Stack an individuals T time series observations of $(y_{it}, \underset{\sim}{x}'_{it})$ into a vector and a matrix, (3.1) may alternatively be written as

$$\underset{\sim}{y}_i = X_i \underset{\sim}{\beta} + \underset{\sim}{e} \alpha_i + \underset{\sim}{u}_i, i = 1, \ldots, N, \tag{3.2}$$

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 $Q\underset{\sim}{e} = \underset{\sim}{0}$. Premultiplying (3.2) by Q yields

$$Q\underset{\sim}{y}_i = QX_i\underset{\sim}{\beta} + Q\underset{\sim}{u}_i, \quad i = 1. \ldots, N. \tag{3.3}$$

Equation (3.3) 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 (3.3). 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),

$$\hat{\underset{\sim}{\beta}} = \left[\sum_{i=1}^{N} X_i'(Q'Q)^- X_i\right]^{-1} \left[\sum_{i=1}^{N} X_i'(Q'Q)^- \underset{\sim}{y}_i\right], \tag{3.4}$$

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. Premultiplying (3.3) by Q is equivalent to transforming (3.1) into a model

$$(y_{it} - \bar{y}_i) = \underset{\sim}{\beta}'(\underset{\sim}{x}_{it} - \bar{\underset{\sim}{x}}_i) + (u_{it} - \bar{u}_i), \begin{matrix} i = 1, \ldots, N, \\ t = 1, \ldots, T, \end{matrix} \tag{3.5}$$

where $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \bar{\underset{\sim}{x}}_i = \frac{1}{T}\sum_{t=1}^{T} \underset{\sim}{x}_{it}$ and $\bar{u}_i = \frac{1}{T}\sum_{t=1}^{T} u_{it}$. The transformation is called *covariance transformation*. The least squares estimator (LS) (or a generalized least squares estimator (GLS)) of (3.5),

$$\hat{\underset{\sim}{\beta}}_{cv} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T}(\underset{\sim}{x}_{it} - \bar{\underset{\sim}{x}}_i)(\underset{\sim}{x}_{it} - \bar{\underset{\sim}{x}}_i)'\right]^{-1} \left[\sum_{t=1}^{N}\sum_{t=1}^{T}(\underset{\sim}{x}_{it} - \bar{\underset{\sim}{x}}_i)(y_{it} - \bar{y}_i)\right], \tag{3.6}$$

is called *covariance* estimator or *within* estimator because the estimation of $\underset{\sim}{\beta}$ only makes use of within (group) variation of y_{it} and $\underset{\sim}{x}_{it}$ only. The covariance estimator of $\underset{\sim}{\beta}$ turns out to be also the least squares estimator of (3.1) when $\lambda_t = 0$. 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 (3.2) into (3.3) 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,

$$\hat{\underset{\sim}{\beta}} = \left[\sum_{i=1}^{N} X_i' V^{-1} X_i\right]^{-1} \left[\sum_{i=1}^{N} X_i' V^{-1} \underset{\sim}{y}_i\right]. \tag{3.7}$$

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

When α_i is treated as fixed, the covariance estimator is equivalent to applying LS to the transformed model (3.5). If a variable is time-invariant, like gender dummy, $x_{kit} = x_{kis} = \bar{x}_{ki}$, the 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 $\underset{\sim}{x}_i, \psi \neq 0$, the GLS can still estimate the coerfficients of those time-invariant variables.

4. Dynamic Models

When the regressors of a linear model contains lagged dependent variables, say, of the form

$$\underset{\sim}{y}_i = \underset{\sim}{y}_{i,-1}\gamma + X_i\underset{\sim}{\beta} + \underset{\sim}{e}\alpha_i + \underset{\sim}{u}_i = Z_i\underset{\sim}{\theta} + \underset{\sim}{e}\alpha_i + \underset{\sim}{u}_i, \quad i = 1, \ldots, N. \tag{4.1}$$

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 premultiplying (4.1) by the transformation matrix Q $(Q\underset{\sim}{e} = \underset{\sim}{0})$,

$$Q\underset{\sim}{y}_i = QZ_i\underset{\sim}{\theta} + Q\underset{\sim}{u}_i. \tag{4.2}$$

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} is independently, 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 (4.1) into the form

$$(y_{it} - \bar{y}_i) = (y_{i,t-1} - \bar{y}_{i,-1})\gamma + (\underset{\sim}{x}_{it} - \bar{\underset{\sim}{x}}_i)'\underset{\sim}{\beta} + (u_{it} - \bar{u}_i), \quad \begin{matrix} i = 1, \ldots, N, \\ t = 1, \ldots, T, \end{matrix} \tag{4.3}$$

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 $y_{i,t-1}$ and $\bar{u}_i$ is of order (1/T) if $\mid \gamma \mid < 1$. 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 γ=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

$$EW_i\underset{\sim}{u}_i'Q' = \underset{\sim}{0}, \tag{4.4}$$

and

$$\text{rank } (W_iQZ_i) = k, \tag{4.5}$$

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

$$\left[\sum_{i=1}^{N} W_i(Q\underset{\sim}{y}_i - QZ_i\underset{\sim}{\theta})\right]' \left[\sum_{i=1}^{N} W_iQ\underset{\sim}{u}_i\underset{\sim}{u}_i'Q'W_i'\right]^{-1} \left[\sum_{i=1}^{N} W_i(Q\underset{\sim}{y}_i - Q\underset{\sim}{Z}_i\underset{\sim}{\theta})\right], \tag{4.6}$$

with respect to $\underset{\sim}{\theta}$. (e.g. Ahn and Honoré (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

$$D = \begin{bmatrix} -1 & 1 & 0 & \cdot & \cdot \\ 0 & -1 & 1 & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & -1 & 1 \end{bmatrix}, \tag{4.7}$$

then the transformation (4.2) is equivalent to taking the first difference of (4.1) over time to eliminate α_i for $t = 2, \ldots, T$,

$$\Delta y_{it} = \Delta y_{i,t-1}\gamma + \Delta \underset{\sim}{x}'_{it}\underset{\sim}{\beta} + \Delta u_{it}, \begin{array}{l} i = 1, \ldots, N, \\ t = 2, \ldots, T, \end{array} \tag{4.8}$$

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 $\underset{\sim}{x}_{is}$, 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

$$W_i = \begin{bmatrix} \underset{\sim}{q}_{i2} & \underset{\sim}{0} & \cdot & \cdot \\ \underset{\sim}{0} & \underset{\sim}{q}_{is} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \underset{\sim}{q}_{iT} \end{bmatrix}, \tag{4.9}$$

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 independently, identically distributed across i, the Arellano-Bover (1991) GMM estimator takes the form

$$\begin{aligned} \underset{\sim}{\hat{\theta}}_{AB,GMM} = & \left\{ \left[\sum_{i=1}^{N} Z'_i D' W'_i \right] \left[\sum_{i=1}^{N} W_i A W'_i \right]^{-1} \left[\sum_{i=1}^{N} W_i D Z_i \right] \right\}^{-1} \\ & \left\{ \left[\sum_{i=1}^{N} Z'_i D W'_i \right] \left[\sum_{i=1}^{N} W_i A W'_i \right]^{-1} \left[\sum_{i=1}^{N} W_i D \underset{\sim}{y}_i \right] \right\} \end{aligned} \tag{4.10}$$

where A is a $(T-1) \times (T-1)$ matrix with 2 on the diagonal elements, -1 on the elements 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 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 all the available sample, hence can yield more efficient estimator than (4.10) (e.g.

Hsiao, Pesaran and Tahmiscioglu (2002), Binder, Hsiao and Pesaran (2004)). Since there is no reason to assume that the data generating process of initial observations, y_{i0}, to be different from the rest of y_{it}, the likelihood approach has to formulate the joint likelihood function of $(y_{i0}, y_{i1}, \ldots, y_{iT})$ (or the conditional likelihood function $(y_{i1}, \ldots, y_{iT} \mid y_{i0})$). However, y_{i0} depends on previous values of $\underset{\sim}{x}_{i,-j}$ and α_i which are unavailable. Bhargava and Sargan (1983) suggest to circumscribe this missing data problem by conditioning y_{i0} on $\underset{\sim}{x}_i$ and α_i if α_i is treated as random while Hsiao, Pesaran and Tahmisciogulu (2002) propose conditioning $(y_{i1} - y_{i0})$ on the first difference of $\underset{\sim}{x}_i$ if α is treated as fixed constants.

5. Random vs Fixed Effects Specification

The advantages of random effects (RE) specifications are:

1. The number of parameters stay constant when sample size increases.
2. It allows the derivation of efficient estimators that make use of both within and between (group) variation.
3. It allows the estimation of the impact of time-invariant variables.

The disadvantages of RE specification is that it typically assumes that the individual- and/or time-specific effects are randomly distributed with a common mean and are independent of $\underset{\sim}{x}_{it}$. 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 fixed effects (FE) specification are that it allows 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:

1'. 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)).

2'. The FE estimator does not allow the estimation of the coefficients that are time-invariant.

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

$$\left(\hat{\underset{\sim}{\theta}}_{FE} - \hat{\underset{\sim}{\theta}}_{RE}\right)' \left[\text{Cov}(\hat{\underset{\sim}{\theta}}_{FE}) - \text{Cov}(\hat{\underset{\sim}{\theta}}_{RE})\right]^{-} \left(\hat{\underset{\sim}{\theta}}_{FE} - \hat{\underset{\sim}{\theta}}_{RE}\right) \tag{5.1}$$

to test RE vs FE specification. The statistic (5.1) is asymptotically chi-square distributed with degrees of freedom equal to the rank of $\left[\text{Cov}(\hat{\underset{\sim}{\theta}}_{GMM}) - \text{Cov}(\hat{\underset{\sim}{\theta}}_{RE})\right]$.

6. Nonlinear Models

The introduction of individual-specific effects, α_i, and/or time-specific effects, λ_t, provide a simple way to capture the unobserved heterogeneity across i and over t. However, the likelihood functions are in terms of observables, $(\underset{\sim}{y}_i, \underset{\sim}{x}_i), i = 1, \ldots, N$. Therefore, we will have to either treat α_i as unknown parameters (fixed effects) and consider the conditional likelihood,

$$f(\underset{\sim}{y}_i \mid \underset{\sim}{x}_i, \underset{\sim}{\beta}, \alpha_i), i = 1, \ldots, N, \tag{6.1}$$

or to treat α_i as random and consider the marginal likelihood

$$f(y_i \mid \underset{\sim}{x}_i; \underset{\sim}{\beta}) = \int f(y_i \mid \underset{\sim}{x}_i, \underset{\sim}{\beta}, \alpha_i) f(\alpha_i \mid \underset{\sim}{x}_i) d\alpha_i, i = 1, \ldots, N, \tag{6.2}$$

where $f(\alpha_i \mid \underset{\sim}{x}_i)$ denotes the conditional density of α_i given $\underset{\sim}{x}_i$.

When the unobserved individual specific effects, α_i, (and or time-specific effects, λ_t) affect the outcome, y_{it}, linearly, 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 (3.3) or first difference transformation (4.8). However, if α_i affects y_{it} nonlinearly, it is not easy to find transformation that can eliminate α_i. For instance, consider the following binary choice model where the observed y_{it} takes the value of either 1 or 0 depending on the latent response function

$$y_{it}^* = \underset{\sim}{\beta}' \underset{\sim}{x}_{it} + \alpha_i + u_{it}, \tag{6.3}$$

and

$$y_{it} = \begin{cases} 1, & \text{if } y_{it}^* > 0, \\ 0, & \text{if } y_{it}^* \leq 0, \end{cases} \tag{6.4}$$

where u_{it} is independently, identically distributed with density function $f(u_{it})$. Let

$$y_{it} = E(y_{it} \mid \underset{\sim}{x}_{it}, \alpha_i) + \epsilon_{it}, \tag{6.5}$$

then

$$\begin{aligned} E(y_{it} \mid \underset{\sim}{x}_{it}, \alpha_i) &= \int_{-(\underset{\sim}{\beta}' \underset{\sim}{x}_{it} + \alpha_i)}^{\infty} f(u) du \\ &= [1 - F(-\underset{\sim}{\beta}' \underset{\sim}{x}_{it} - \alpha_i)]. \end{aligned} \tag{6.6}$$

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},

$$\begin{aligned} y_{it} - y_{i,t-1} = &[1 - F(-\underset{\sim}{\beta}' \underset{\sim}{x}_{it} - \alpha_i)] \\ &- [1 - F(-\underset{\sim}{\beta}' \underset{\sim}{x}_{i,t-1} - \alpha_i)] + (\epsilon_{it} - \epsilon_{i,t-1}). \end{aligned} \tag{6.7}$$

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

$$\Pi_{i=1}^{N} \Pi_{t=1}^{T} [F(-\underset{\sim}{\beta}' \underset{\sim}{x}_{it} - \alpha_i)]^{1-y_{it}} [1 - F(-\underset{\sim}{\beta}' \underset{\sim}{x}_{it} - \alpha_i)]^{y_{it}}. \tag{6.8}$$

If T is large, consistente estimator of $\underset{\sim}{\beta}$ and α_i can abe obtained by maximizing (6.8). If T is finite, there is only limited information about α_i no matter how large N is. 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, consistent estimator of $\underset{\sim}{\beta}$ can be obtained by maximizing the marginal likelihood function,

$$\Pi_{i=1}^{N} \int \Pi_{t=1}^{T} [F(-\underset{\sim}{\beta}' \underset{\sim}{x}_{it} - \alpha_i)]^{1-y_{it}} [1 - F(-\underset{\sim}{\beta}' \underset{\sim}{x}_{it} - \alpha_i)]^{y_{it}} f(\alpha_i \mid \underset{\sim}{x}_i) d\alpha_i. \tag{6.9}$$

However, maximizing (6.9) involves T-dimensional integration. Butler and Moffit (1982), Chamberlain (1984), Heckman (1981), etc., have suggested methods to simplify the computation.

The advantage of 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 consistent estimator of $\underset{\sim}{\beta}$. Moreover, the derivation of marginal likelihood through multiple integration may be computationally infeasible. The advantage of 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. (6.8)) if the errors are assumed 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 to transform 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 (6.3) is logistic, then

$$\text{Prob } (y_{it} = 1 \mid \underset{\sim}{x}_{it}, \alpha_i) = \frac{e^{\underset{\sim}{\beta}' \underset{\sim}{x}_{it} + \alpha_i}}{1 + e^{\underset{\sim}{\beta}' \underset{\sim}{x}_{it} + \alpha_i}}. \tag{6.10}$$

Since, in a logit model, the denominators of Prob$(y_{it} = 1 \mid \underset{\sim}{x}_{it}, \alpha_i)$ and 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 $\exp(\alpha_i s) \cdot \exp\{\sum_{t=1}^{T} (\underset{\sim}{\beta}' \underset{\sim}{x}_{it}) y_{it}\}$, 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

$$\begin{aligned} \text{Prob}(y_{i1} = 1, y_{i2} = 0 \mid y_{i1} + y_{i2} = 1) &= \frac{e^{\underset{\sim}{\beta}' \underset{\sim}{x}_{i1}}}{e^{\underset{\sim}{\beta}' \underset{\sim}{x}_{i1}} + e^{\underset{\sim}{\beta}' \underset{\sim}{x}_{i2}}} \\ &= \frac{1}{1 + e^{\underset{\sim}{\beta}' \Delta \underset{\sim}{x}_{i2}}}, \end{aligned} \tag{6.11}$$

and

$$\text{Prob}(y_{i1} = 0, y_{i2} = 1 \mid y_{i1} + y_{i2} = 1) = \frac{e^{\underset{\sim}{\beta}' \Delta \underset{\sim}{x}_{i2}}}{1 + e^{\underset{\sim}{\beta}' \Delta \underset{\sim}{x}_{i2}}}, \tag{6.12}$$

(Chamberlain (1980), Hsiao (2003)).

Alternatively, Manski (1987) exploits the latent linear structure of (6.3) by noting that for given i,

$$\underset{\sim}{\beta}' \underset{\sim}{x}_{it} \gtreqless \underset{\sim}{\beta}' \underset{\sim}{x}_{i,t-1} \Longleftrightarrow E(y_{it} \mid \underset{\sim}{x}_{it}, \alpha_i) \gtreqless E(y_{i,t-1} \mid \underset{\sim}{x}_{i,t-1}, \alpha_i), \tag{6.13}$$

and suggests maximizing the objective function

$$H_N(b) = \frac{1}{N} \sum_{i=1}^{N} \sum_{t=2}^{T} sgn(\underset{\sim}{b}' \Delta \underset{\sim}{x}_{it}) \Delta y_{it}, \tag{6.14}$$

where $sgn(w) = 1$ if $w > 0, = 0$ if $w = 0$, and -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 (6.13) 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 $\| \underset{\sim}{\beta} \| = 1$. Moreover, the spped of convergence is considerably slower ($N^{1/3}$) and the limiting distribution is quite complicated. Horowitz (19) and Lee ( ) 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, Shen, Wang and Weeks (2004)), 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, often they 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 there does not appear to possess consistent estimator when T is finite.

An alternative to consider consistent estimators is to consider bias reduced estimator. 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 a consistent estimators because the latter often have to throw away a lot of sample, thus tend to have large variances.

Following the idea of Cox and Reid (1987), Arellano (2001) and Carro (2004) propose to derive the modified MLE by maximizing the modified log-likelihood function

$$L^*(\underset{\sim}{\beta}) = \sum_{i=1}^{N} \left[\ell_i^*(\underset{\sim}{\beta}, \hat{\alpha}_i(\underset{\sim}{\beta})) - \frac{1}{2} \log \ell_{i,d_i d_i}^*(\beta_1 \hat{\alpha}_i(\underset{\sim}{\beta}) \right], \tag{6.15}$$

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 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 to the order of $(1/T^2)$, without increasing the asymptotic variance.

Monte Carlo experiments conducted by Carro (2004) have shown that when $T = 8$, the bias of modified MLE 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 will not affect the asymptotic properties of the modified MLE.

7. Modeling Cross-Sectional Dependence

Most panel studies assume that apart from the possible presence of individual in-

variant but period varying time specific effects, λ_t, the effects of omitted variables are independently distributed across cross-sectional units. However, often economic theory predicts that agents take actions that lead 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, Le Gallo and Jayet (2005)). 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, $w_{ij} = 1$. 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 to the dependent variable, to the explanatory variables, or to 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

$$y = \rho(W \otimes I_T)\underset{\sim}{y} + X\underset{\sim}{\beta} + \underset{\sim}{u} \tag{7.1}$$

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,

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{v}, \tag{7.2}$$

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

$$\underset{\sim}{v} = \theta(W \otimes I_T)\underset{\sim}{v} + \underset{\sim}{u}, \tag{7.3}$$

or a *spatial moving average* form,

$$\underset{\sim}{v} = \gamma(W \otimes I_T)\underset{\sim}{u} + \underset{\sim}{u}. \tag{7.4}$$

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". Another approach to model cross-sectional dependence is to assume that the error of a model, say model (7.3) follows a linear factor model,

$$v_{it} = \sum_{j=1}^{r} b_{ij} f_{jt} + u_{it}, \tag{7.5}$$

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 (7.5) 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 load matrix $B = (b_{ij})$ can be empirically identified if both N and T are large. However, when T is large, one can estimate the covariance between i and j, σ_{ij}, by $\frac{1}{T}\sum_{t=1}^{T} \hat{v}_{it}\hat{v}_{jt}$ directly, then apply the generalized least squares method, where $\hat{v}_{it}$ is some preliminary estimate of v_{it}.

8. 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 are 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 the modeling of joint dependence or simultaneous equations models, time-varying parameter models (e.g. Hsiao (1992, 2003), tests of unit root or cointegration (e.g., Levin, Lin and Chu (2003), Pesaran, Shin and Smith (2004), Hsiao and Pesaran (2004)), the asymptotics for panels with large N and T (e.g. Phillips and Moon (1999)), unbalanced panel, measurement errors (Griliches and Hausman (1986), Wansbeek and Konig (1989)), etc. that were not discussed, but could be found in Baltagi (2001) or Hsiao (2003).

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 the 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, when using panel data, how can we increase the efficiency of parameter estimates? Fourth, are the assumptions underlying the statistical inference procedures and the data-generating process compatible.

References

Ahn, S.C. and P. Schmidt (1995), " Efficient Estimation of Models for Dynamic Panel Data", *Journal of Econometrics*, 68, 5-27.

Anderson, T.W. and C. Hsiao (1981), "Estimation of Dynamic Models with Error Components", *Journal of the American Statistical Association*, 767, 598-606.

_____ (1982), "Formulation and Estimation of Dynamic Models Using Panel Data", *Journal of Econometrics*, 18, 47-82.

Anselin, L., J. Le Gallo and H. Jayet (2005), "Spatial Panel Econometrics", *mimeo*.

Arellano, M. and S.R. 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 Variable Estimation of Error-Components Models", *Journal of Econometrics*, 68, 29-51.

Bai, J. and S. Ng (2002), "Determining the Number of Factors in Approximate Factor Models", *Econometrica*, 70. 91-121.

Balestra, P. and H. 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.

Baltagi, B.H. (2001), "Econometric Analysis of Panel Data", Second edition, New York: Wiley.

Ben-Porath, Y. (1973), "Labor Force Participation Rates and the Supply of Labor", *Journal of Political Economy*, 81, 697-704.

Bhargava, A. and J.D. Sargan (1983), "Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods", *Econometrica*, 51, 1635-1659.

Binder, M., C. Hsiao and M.H. Pesaran (2004), "Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration", *Econometric Theory* (forthcoming).

Butler, J.S. and R. Moffitt (1982), "A Computationally Efficient Quadrature Procedure for the One Factor Multinomial Probit Model", *Econometrica*, 50, 761-764.

Carro, J.M. (2004), "Estimating Dynamic Panel Data Discrete Choice Models with Fixed Effects", *mimeo*.

Chamberlain, G. (1980), "Analysis of Covariance with Qualitatiave Data", *Review of Economic Studies*, 47, 225-238.

_____ (1993), "Feedback in Panel Data Models", *mimeo*, Department of Economics, Harvard University.

Granger, C.W.J. (1990), "Aggregation of Time-Series Variables: A Survey", in *Disag-*

gregation in Econometric Modeling, ed. by T. Barker and M.H. Pesaran, London: Routledge.

Griliches, Z. and J.A. Hausman (1986), "Errors-in-Variables in Panel Data", *Journal of Econometrics*, 31, 93-118.

Hausman, J.A. (1978), "Specification Tests in Econometrica", *Econometrica*, 46. 1251-71.

Heckman, J.J. (1981a), "Statistical Models for Discrete Panel Data", in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C.F. Manski and D. McFadden, Cambridge, Mass., MIT Press, 114-178.

Honoré, Bo and E. Kyriazidou (2000),. "Panel Data Discrete Choice Models with Lagged Dependent Variables", *Econometrica*, 68, 839-874.

Honoré, Bo (1992), "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects", *Econometrica*, 60, 533-567.

Hsiao, C., (1992), "Rnadom Coefficient Models" in *The Econometrics of Panel Data*, ed. by L. Matyas and P. Sevestres, Kluwer, 1st edition, 223-241, 2nd ed. (1996), 410-428.

_____ (2003), *Analysis of Panel Data*, 2nd edition, Cambridge: Cambridge University Press (Econometric Society monograph no. 34).

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., Y. Shen and H. Fujiki (2004), "Aggregate vs Disaggregate Data Analysis — A Paradox in the Estimation of Money Demand Function of Japan Under the Low Interest Rate Policy", *Journal of Applied Econometrics* (forthcoming).

Kyriazidou, E. (1997), "Estimation of a Panel Data Sample Selection Model", *Econometrica*, 65, 1335-1364.

Levin, A., C. Lin, and J. Chu (2002), "Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties", *Journal of Econometrics*.

MaCurdy, T.E. (1981), "An Empirical Model of Labor Supply in a Life Cycle Setting", *Journal of Political Economy*, 89, 1059-85.

Manski, C.F. (1987), "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data", *Econometrica*, 55, 357-362.

Moon, H.R. and B. Perron (2004), "Testing for a Unit Roots in Panels with Dynamic Factors", *Journal of Econometrics*, 122, 81-126.

Neyman, J. and E.L. Scott (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. and Z. Griliches (1984), "Estimating Distributed Lags in Short Panels with

an Application to the Specification of Depreciation Patterns and Capital Stock Constructs", *Review of Economic Studies*, 51, 243-262.

Pesaran, M.H. (2003), "A Simple Panel Unit Root Test in the Presence of Cross-Section Dependence", *mimeo*.

Phillips, P.C.B. (1986), "Understanding Spurious Regressions in Econometrics", *Journal of Econometrics*, 33, 311-340.

Phillips, P.C.B. and H.R. Moon (1999), "Linear Regression Limit Theory for Nonstationary Panel Data", *Econometrica*, 67, 1057-1111.

Rao, C.R., (1973), *Linear Statistical Inference and Its Applications*, 2nd ed., New York: Wiley.

Wansbeek, T.J. and R.H. Koning (1989), "Measurement Error and Panel Data", *Statistica Neerlandica*, 45, 85-92.

Zilak, 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.