Note on Panel Data Econometrics

Report by Sebastian Buhai*

August, 2003

1 Introduction

Since longitudinal data sets contain observations over the same units (e.g. individuals, household, firms, industries, regions, countries) repeated over a number of time periods, panel data techniques present major advantages over standard time-series or cross-sectional approaches, by combining their identifying features. The identification of time series parameters was usually based on notions of stationarity, predeterminedness and uncorrelated shocks; identifying cross-sectional parameters traditionally relied on exogenous instrumental variables and random sampling; working with panel data sets allows using all these resources and at the same time determines economists to think more about the nature and applicability of a particular technique to identify a parameter of potential interest.

Several improvements that working with panel data has over working with conventional time-series or cross-sectional data are for instance mentioned in Hsiao (2003). There is for instance a larger number of data points available, increasing therefore the degrees of freedom and reducing the colinearity among explanatory variables. Next, panel data allows constructing and testing more complicated behavioral models than purely cross-sectional or time series data. Yet another plus is that we can better study the dynamics of adjustment, such as for instance the effect of unionism on economic behavior. A particular advantage of the micro-panel data sets is that they eliminate biases resulting from aggre-

*Tinbergen Institute and Erasmus University Rotterdam; contact: Roetersstraat 31, 1018 WB, Amsterdam; phone: 020-551 3561; fax: 020- 551 3555; email: buhai@tinbergen.nl

gation over micro-units; see Blundell (1988) for an interesting survey in this sense or Blundell and Meghir (1990) for a more specific discussion on estimating life-cycle models. In a nutshell, the most acclaimed features of longitudinal sets are probably summarized in that they can often provide the exogenous variation required to identify structural parameters through comparisons across periods covering policy changes and in the possibility of following the same units over time, which facilitates analyzing dynamic responses and modelling unobserved heterogeneity.

The main objective of this paper is to give a general overview of the main trends and current status in panel data econometrics, building on the guidelines given in Manuel Arellano's lectures in the NAKE Workshop from Rotterdam, June 10-13, 2003. Next to adopting a broad-spectrum approach, we will put some more emphasis on a few selected themes. We will insist for instance on the static longitudinal framework with its basic motivations of controlling for unobserved heterogeneity and decomposing error structures, since much of this is recurrent in other sections. An extensive treatment of estimating covariance structures for dynamic models, including the discussion of a concrete example in Abowd and Card (1989), will also be provided. The overall structure of the paper is the following: section 2. will tackle static panel data models; in section 3. the focus will be on time series models and dynamic error components; section 4. will try to give the reader an impression of dynamics with exogeneity versus predeterminedness; section 5. will briefly discuss discrete choice models.

2 Static models

2.1 Unobserved heterogeneity

The rationale econometricians invoke when using panel data in general and static regression models in the narrower sense, seems to be based on two distinct motives: first, controlling for unobserved time-invariant heterogeneity in cross-sectional models and second, disentangling components of variance and studying the dynamics of cross-sectional populations. Loosely speaking, these reasons can be associated with the two main panel data techniques, the fixed effects and the random effects models. In the remainder of this section we shall discuss issues pertaining to both these motivations.

Consider the following cross-sectional regression model:

$$y_{i1} = \beta x_{i1} + \eta_i + v_{i1} \tag{1}$$

Following the introductory discussion above we can approach the estimation of β in different ways, depending on our starting premise. If η_i is observed then β can be immediately identified from a multiple regression of y on x and η. If η_i is not observed however, in order to identify β we either need lack of correlation between x_{i1} and η_i, so $Cov(x_{i1}, \eta_i) = 0$, in which case we will estimate β as

$$\beta = \frac{Cov(x_{i1}, y_{i1})}{Var(x_{i1})} \tag{2}$$

or we need to have an external instrument z_i available so that $Cov(z_i, \eta_i) = 0$, $Cov(z_{i,}v_{i1}) = 0$ but $Cov(z_i, x_{i1}) \neq 0$, so that we can estimate

$$\beta = \frac{Cov(z_i, y_{i1})}{Cov(z_i, x_{i1})} \tag{3}$$

When none of the approaches above is applicable we have a potential problem. Suppose however that we observe y_{i2} and x_{i2} for the same individuals in a second period:

$$y_{i2} = \beta x_{i2} + \eta_i + v_{i2} \tag{4}$$

If we make the strict exogeneity assumption[1] for the idiosyncratic disturbance v_{it} ,

$$E(v_{it}|x_{i1}, x_{i2}, \eta_i) = 0, \text{ for } t = 1, 2 \tag{5}$$

then we can still identify β in a regression in first differences, albeit the individual effect η_i is unobservable:

$$y_{i2} - y_{i1} = \beta(x_{i2} - x_{i1}) + (v_{i2} - v_{i1}) \tag{6}$$

resulting in

$$\beta = \frac{Cov(\Delta x_{i2}, \Delta y_{i2})}{Var(\Delta x_{i2})} \tag{7}$$

Arellano (2003) gives a few concrete examples where determining the coefficient of interest through first-differencing works and where this does not work. For the first case we have the classical Cobb-Douglas production function, present among others also in Chamberlain (1984),

[1] The strict exogeneity assumption implies that explanatory variables in each time period are uncorrelated with the idiosyncratic error in each time period, $E(x_{is}^{'}u_{it}) = 0$, $\forall\ s, t = 1, 2, ..T$ (with $T = 2$ in our case). In particular this assumption is much stronger than the zero contemporaneous correlation $E(x_{it}^{'}u_{it}) = 0$, $t = 1, 2, ..T$. For consistency of panel data estimation the $E(x_{is}^{'}u_{it}) = 0$ condition suffices, allowing for possible correlation between c_i and v_{it} for any t. Nonetheless standard forms of statistical inference rely on expression (5. See Wooldrige(2002) for a clear and in depth treatment of these issues.

where (1) represents a production function for an agricultural product. The following notations are imposed: i - farms, t - time periods (seasons or years), y_{it} - log output, x_{it} - log of a variable input (e.g. labour), η_i - an input remaining constant over time (e.g. soil quality), v_{it} - a stochastic input outside the farmer's control (e.g. rainfall). We suppose that the soil quality is known by the farmer but not by the researcher; hence if the farmers maximize expected profits there will be a cross-sectional correlation between labour and soil quality, or, formally, between the regressor variable x_{it} and the individual effect η_i. Suppose next that data on y_{i2} and x_{i2} for a second period become available. Suppose also that rainfall in the second period is independent of a farm's labour demand in the two periods, thus checking for the exogeneity assumption above having labour uncorrelated with rainfall at all lags and leads. Then even in the absence of data on η_i, the availability of panel data affords identification of the technological parameter β.

While this was an example that worked, let us turn to one which does not. This is the standard case of estimating the structural returns to education. Consider again our regression equation in (1). We label by y_{it} - log wage, x_{it} - years of full time education, η_i - unobserved ability and β - returns to education. The dilemma is that x_{it} typically lacks time series variation. A regression in first differences will not identify β because in (7), $Var(\Delta x_{i2})$ would be virtually 0. So in this case panel data analysis is not very useful; we would manage to solve this problem if we could get some exogenous instrumental variables such as data on siblings for instance.

In the paragraphs above, talking about the unobserved heterogeneity, we have in fact introduced the fixed-effects models within panel data. Let us see more formally what are its basic assumptions. Take a random sample $\{(y_{i1},...,y_{iT},x_{i1},...,x_{iT},\eta_i), i = 1,...N\}$ and consider the model from (1), which rewritten here for multiple t:

$$y_{it} = x'_{it}\beta + \eta_i + v_{it} \tag{8}$$

The assumptions of this model are:

A1.

$$E(v_i|x_i,\eta_i) = 0, \text{ for } t = 1,...,T \tag{9}$$

where $v_i = (v_{i1},...,v_{iT})'$ and $x_i = (x_{i1},...,x_{iT})'$.

A2.

$$Var(v_i|x_i,\eta_i) = \sigma^2 I_T \tag{10}$$

which means that the time-variant errors are conditionally homoskedastic and serially uncorrelated.

The fixed effects models are usually referred to as models where we allow for arbitrary correlation between the unobserved individual effect η_i and the observed explanatory variable x_{it}, contrasting with random effects models where the permanent effect is independent of the regressors (Wooldridge (2002)). Assumption A1 above is the fundamental assumption since it puts forward the strict exogeneity condition. Panel data models without this assumption are more complex and shall be considered later in this paper. The second assumption A2 is just an auxiliary assumption needed for optimality of simple OLS estimators as it will become clearer below.

We shall now introduce the most popular estimator in panel data analysis, namely the within-group estimator (know also under several other names, including covariance estimator, dummy-variable least square estimator[2] or fixed effects estimator). If we consider the equation in (8), estimating first differences works without problems if $T = 2$, when we can apply OLS to the equation

$$\Delta y_{i2} = \Delta x'_{i2}\beta + \Delta v_{i2} \tag{11}$$

When we have $T \geq 3$ however, we could have some problems trying to use first differencing since this technique might induce serial correlation. Subsequent to first-differencing we will have a system of $T-1$ equations which can be written easier in the compact form

$$Dy_i = DX_i\beta + Dv_i \tag{12}$$

with D being the $(T-1) \times T$ matrix operator

$$D = \begin{pmatrix} -1 & 1 & 0 \ldots & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ . & & . & & . \\ . & & & . & . \\ . & & & . & . \\ 0 & 0 & 0 \ldots & -1 & 1 \end{pmatrix} \tag{13}$$

Provided $E(Dv_i|x_i) = 0$, we obtain unbiased and consistent estimates for β by OLS, for large N:

$$\widehat{\beta}_{OLS} = \left(\sum_{i=1}^{N} (DX_i)'DX_i \right)^{-1} \sum_{i=1}^{N} (DX_i)'Dy_i \tag{14}$$

[2]See Verbeek(2000) for a brief discussion of the equivalence between the within-group and the dummy least square estimator.

If assumption A2 is however not respected we will face correlation between the errors in the first differences and the $x's$, so that

$$Var(Dv_i|x_i) = \sigma^2 DD' \tag{15}$$

The optimal estimator in this case is the generalized least squares (GLS) estimator, which takes the form

$$\widehat{\beta}_{WG} = \left(\sum_{i=1}^{N} X_i' D' (DD')^{-1} D X_i \right)^{-1} \sum_{i=1}^{N} X_i{}' D' \left(DD'\right)^{-1} Dy_i \tag{16}$$

This is the within-group estimator and it is equivalent to its alternative form as deviations from time means:

$$\widehat{\beta}_{WG} = \left[\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \overline{x_i})(x_{it} - \overline{x_i})' \right]^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \overline{x_i})(y_{it} - \overline{y_i}) \tag{17}$$

The equivalence between (16) and (17) can be easiest seen if we label by Q the within-group operator

$$Q \equiv I_T - u'/T = D'(DD')^{-1} D \tag{18}$$

Q transforms the original time series into deviations from time means: $\widetilde{y_i} = Qy_i$, with elements are given by $\widetilde{y_{it}} = y_{it} - \overline{y_i}$. We obtain thus equation (17), an OLS in deviations from time means.

There is yet another alternative transformation of the original data that results from first differencing and then applying GLS to the differenced data to remove the serial correlation induced in the first stage. This is the forward orthogonal deviations technique (see Arellano and Bover (1995) for a detailed treatment). Consider the $(T-1) \times T$ matrix

$$A = (DD')^{-1/2} D \tag{19}$$

so that $A'A = Q$ defined above, and $AA' = I_{T-1}$. A $T \times 1$ time series error transformed by A,

$$v_i^* = Av_{it} \tag{20}$$

will consist of elements of the form

$$v_{it}^* = c_t[v_{it} - \frac{1}{T-t}(v_{i,t+1} + ... + v_{iT})] \tag{21}$$

where $c_t^2 = \frac{T-t}{T-t+1}$.

We eliminate individual effects by applying the orthogonal deviations transformation without inducing serial correlation in the transformed error as the case with first-differencing. The within-group estimator can thus be also regarded as OLS in orthogonal deviations.

So far we have been assuming that the auxiliary assumption A2 (equation (10) above) was satisfied. The within-group estimator is optimal under this assumption. The question is now how to estimate the model when homoskedasticity and serial correlation are violated and when there would be thus inconsistent standard errors subsequent to estimation. There are two relevant cases to be considered. One is when we have panel data sets of fixed T and large N dimensions while the second situation is the treatment of the case for large T and fixed N dimensions. For the second case there are not clear directions of performing the analysis, some variations being Newey-West estimators (see Arellano(2003) for a discussion). For the first case, if

$$Var(v_i^*|x_i) = \Omega(x_i) \tag{22}$$

where $\Omega(x_i)$ is a symmetric matrix of order T, then the optimal estimator can be written as follows

$$\widehat{\beta}_{UGLS} = \left(\sum_{i=1}^{N} X^{*'} \Omega^{-1}(x_i) X_i^* \right)^{-1} \sum_{i=1}^{N} X^{*'} \Omega^{-1}(x_i) y_i^* \tag{23}$$

Nevertheless this estimator is unfeasible (UGLS) because the matrix Ω is unknown. We would need to estimate this matrix using nonparametric methods or similar techniques. The only case where Ω could be estimated in a straightforward manner is one where the conditional variance of v_i^* in (22) is a constant but non-scalar matrix of the form

$$Var(v_i^*|x_i) = \Omega \tag{24}$$

In this case the GLS estimator is feasible (FGLS) and has the following form

$$\widehat{\beta}_{FGLS} = \left(\sum_{i=1}^{N} X^{*'} \widehat{\Omega}^{-1} X_i^* \right)^{-1} \sum_{i=1}^{N} X^{*'} \widehat{\Omega}^{-1} y_i^* \tag{25}$$

where

$$\widehat{\Omega} = \frac{1}{N} \sum_{i=1}^{N} \widehat{v}_i^* \widehat{v}_i^{*'} \tag{26}$$

Since however most of the time we cannot ensure that we have (24), we have to think about intermediate possibilities. To this end we could consider a larger set of moments, as follows

$$E(x_{it} v_{is}^*) = 0, \, \forall s, t \tag{27}$$

so as to use all past and future values to produce a GMM estimator. If the auxiliary assumption A2 in (10) is not satisfied, then this estimator will be more efficient that the within-groups estimator, but less efficient than the benchmark unfeasible estimator β_{UGLS} in (23). Such an estimator is obtained for example by using the π- matrix approach, also referred to as the minimum-distance approach, of Chamberlain (1982, 1984). We will however not cover this topic here, indicating Hsiao (2003) for a thorough review of this case.

2.2 Error components

We have been talking in the previous section about controlling for unobserved heterogeneity as one of the main reasons for working with panel data. It is time we moved to the second motivation, namely separating out permanent from transitory components of variation. Let us consider a very basic variance-components model

$$y_{it} = \mu + \eta_i + v_{it} \tag{28}$$

where μ - intercept, $\eta_i \sim iid(0, \sigma_\eta^2)$ and $v_{it} \sim iid(0, \sigma^2)$. η_i and v_{it} are independent of each other and

$$Var(y_{it}) = \sigma_\eta^2 + \sigma^2 \tag{29}$$

Having (29) above, we see that $\sigma_\eta^2/(\sigma_\eta^2 + \sigma^2)$ of the total variance is due to the permanent time-invariant component.

This model allows us to make a distinction between aggregate and individual transition probabilities. The individual transition probabilities given η_i are independent of the state of origin:

$$\Pr(y_{it} \in [a, b] | y_{i,t-1} \in [c, d], \eta_i) = \Pr(y_{it} \in [a, b] | \eta_i) \tag{30}$$

Using (30), the aggregate probability will be

$$\int \Pr(y_{it} \in [a, b] | \eta_i) dF(\eta_i | y_{i,t-1} \in [c, d]) \tag{31}$$

We see thus that the decomposition in (28) allows us to distinguish between individual probability statements for units with given η_i and aggregate probabilities for groups of observationally equivalent units. In order to estimate this variance-components models we first try to estimate conditionally on η_i, in other words implementing an estimation of the realizations of permanent effects η_i that occur in the sample and the variance σ^2 (Arellano (2003)). The straightforward unbiased estimates would then be

$$\widehat{\eta_i} = \overline{y_i} - \overline{y}, \ i = 1, ..., N \tag{32}$$

and respectively

$$\widehat{\sigma}^2 = \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \overline{y_i})^2 \tag{33}$$

The only problem is that we do not have yet an estimate of σ_η^2 and typically this is also of interest. The seemingly immediate way to get this estimate would be the following. Having

$$Var(\overline{y_i}) \equiv \overline{\sigma}^2 = \sigma_\eta^2 + \frac{\sigma^2}{T} \tag{34}$$

an unbiased estimator of our statistics of interest can be obtained as

$$\widehat{\sigma}_\eta^2 = \widehat{\overline{\sigma}}^2 - \frac{\widehat{\sigma}^2}{T} \tag{35}$$

The dilemma is that the estimator above can be negative by construction. There will be thus cases where we won't be able to estimate $\widehat{\sigma}_\eta^2$. We also have to note that for large N and short T dimensions one can obtain precise estimates of σ_η^2 and σ^2 but not of η_i, while for small N and large T sets we would be able to obtain good estimates of η_i and σ^2 but not of σ_η^2.

The regression version of the model discussed previously in (28) is

$$y_{it} = \mu(x_{it}, f_i) + \eta_i + v_{it} \tag{36}$$

where

$$\mu(x_{it}, f_i) = x_{it}'\beta + f_i'\nu \tag{37}$$

We allow for both time-invariant as well as time-varying conditioning variables (f_i, respectively x_{it}). A very important assumption is that both error terms are uncorrelated with any of the regressors. This is in contrast with the unobserved heterogeneity discussion in the previous subsection where the very rationale was to allow the unobserved individual effect η_i to be potentially correlated with the regressor. So here we will have

$$E(\eta_i | x_{i1}, ..., x_{iT}, f_i) = 0 \tag{38}$$

and the variance

$$Var(\eta_i | x_{i1}, ..., x_{iT}, f_i) = \sigma_\eta^2 \tag{39}$$

Optimal estimation of the model is achieved using a specific GLS known under the name Balestra-Nerlove estimator (Balestra and Nerlove (1966)). A short review of this technique as well as discussion on tests for correlated unobserved heterogeneity can be found in Arellano (2003).

We move now to discuss an ever recurrent theme in the panel data econometrics literature, measurement error in variables. Start with the following cross-sectional regression model

$$y_i = \alpha + x_i^{\dagger}\beta + v_i \tag{40}$$

Suppose $x_i^{\dagger}$ is observed with an additive noise effect ε_i under the form x_i:

$$x_i = x_i^{\dagger} + \varepsilon_i \tag{41}$$

Further assume independence between all unobservables $x_i^{\dagger}, \varepsilon_i$ and v_i. Then β is given by

$$\beta = \frac{Cov(y_i, x_i^{\dagger})}{Var(x_i^{\dagger})} \tag{42}$$

but $x_i^{\dagger}$ is unobservable. We try to get a better expression for β from

$$\frac{Cov(y_i, x_i)}{Var(x_i)} = \frac{Cov(y_i, x_i^{\dagger})}{Var(x_i) + Var(\varepsilon_i)} = \frac{\beta}{1+\lambda} \tag{43a}$$

with $\lambda = Var(\varepsilon_i)/Var(x_i^{\dagger})$. If we assume that we have the means of assessing the measurement error, so that we either know or we can estimate λ and $\sigma_{\varepsilon}^2 = Var(\varepsilon_i)$ then from (43a) β can be determined :

$$\beta = (1+\lambda)\frac{Cov(y_i, x_i)}{Var(x_i)} = \frac{Cov(y_i, x_i)}{Var(x_i) - \sigma_{\varepsilon}^2} \tag{44}$$

More generally however we need to find other ways to estimate β since we cannot always rely on approximating the size of the measurement error. One solution would be to have a second noisy measure of $x_i^{\dagger}$ and to use this as an instrumental variable. Suppose

$$z_i = x_i^{\dagger} + \xi_i \tag{45}$$

If ξ_i is independent of ε_i and other unobservables, z_i can be successfully used as an IV so that β is obtained as[3]

$$\beta = \frac{Cov(z_i, y_i)}{Cov(z_i, x_i)} \tag{46}$$

[3] In fact we even have an overidentifying restriction in this problem since we can also write

$$\frac{Cov(x_i, y_i)}{Cov(x_i, z_i)} = \beta$$

We have noticed that for linear regression problems the treatment of measurement errors is possible; the problem becomes really difficult when we have a nonlinear regression framework where the measurement error is no longer additively separable from the true value of the regressor. One good reference for further reading in this sector is Hausman et al (1995). A detailed introduction to the literature on measurement error in panel data is Baltagi (2001).

We have not considered yet a model which combines unobserved heterogeneity and measurement error. Write the following linear regression in a cross-section:

$$y_i = x_i^{\dagger}\beta + \eta_i + v_i \tag{47}$$

having again a measurement error in $x_i^{\dagger}$ as in (41). All unobservables are independent of each other, with the exception now that and η_i are not independent. There will be in this case two bias components: the first is due to the measurement error and it depends on σ_{ε}^2 and the second is due to the unobserved heterogeneity and it depends on the $Cov(\eta_i, x_i^{\dagger})$ term[4]. Trying to bypass this problem by using first differences in a panel data series may exacerbate the error rather then eliminating it (see Arellano (2003) and the further references he gives). In principle the problem arises when there is more time series dependence in $x_i^{\dagger}$ than in ε_{it}. One important observations that follows is that finding significantly different results between regressions in first-differences and regressions in orthogonal deviations indicate a high chance that we have have measurement error in the model.

In general the availability of panel data helps to solve the problem of measurement error bias by providing internal instruments but only as long as we can restrict the serial dependence of the measurement error. Take a model with unobserved heterogeneity and a white noise measurement error, having $T > 2$. The moment conditions for the errors in first-differences are

$$E\left[\begin{pmatrix} x_{i1} \\ \cdots \\ x_{i,t-2} \\ x_{i,t+1} \\ \cdots \\ x_{iT} \end{pmatrix} (\Delta y_{it} - \Delta x_{it}\beta)\right] = 0, \text{ for } t = 2, .., T \tag{48}$$

This sort of moments and GMM estimators based on them have been proposed before in the literature (see for instance Griliches and Hausman

[4]Sometimes these two biases tend to offset each other, for instance if $\beta > 0$ and $Cov(\eta_i, x_i^{\dagger}) > 0$, but a full offsetting only happens if $Cov(\eta_i, x_i^{\dagger}) = \sigma_{\varepsilon}^2\beta$ (Arellano(2003)), something very unlikely to happen however.

(1986)). One thing to pay attention to is that if the latent variable $x_i^{\dagger}$ is also white noise than the moment conditions above are still satisfied but for any β, so the rank condition fails and we cannot identify the true value of the coefficient. The basic identification assumption is therefore the time-series persistence in the x's relative to the ε's. These estimators are very useful in situations where first differencing aggravates measurement error bias, as mentioned above.

3 Time series models for panel data

3.1 Covariance structures for error components

The time series models for panel data are motivated by a general interest in the time series properties of longitudinal data sets. Albeit we are interested in separating transitory from permanent components of variation as we saw before, or we want to test a theoretical model with specific predictions and map this to the data (e.g. Hall and Mishkin (1982)), or we like to use a predictive distribution in optimization problems under uncertainty (e.g. Chamberlain (2000)), we find all these issues tackled within the chapter of time series analysis in panel data sets.

One particular matter often arising in the panel data context is the heterogeneity in the individuals and the quest of distinguishing it from inherent individual dynamics in the data set. This problem is relevant especially in short panels. Consider the following heterogeneous error components model

$$y_{it} = \eta_i + v_{it} \tag{49}$$

where v_{it} is white noise. As we saw in previous sections, for $T = 2$ we will have

$$Cov(y_{i1}, y_{i2}) = \sigma_\eta^2 \tag{50}$$

since

$$Var(y_{i1}) = Var(y_{i2}) = \sigma_\eta^2 + \sigma^2 \tag{51}$$

Take now a homogeneous AR(1) model

$$y_{it} = \eta + v_{it} \tag{52}$$

where

$$v_{it} = \alpha v_{i,t-1} + \varepsilon_{it} \tag{53}$$

with η a constant, $\varepsilon_{it} \sim iid(0, \sigma_\varepsilon^2)$, $v_{it} \sim iid(0, \sigma^2)$ and $\sigma^2 = \sigma_\varepsilon^2/(1-\alpha)$. For $T = 2$ we will get

$$Var(y_{i1}) = Var(y_{i2}) = \sigma^2 \tag{54}$$

and subsequently

$$Cov(y_{i1}, y_{i2}) = \alpha\sigma^2 \tag{55}$$

The point is that there is no way to distinguish empirically between the models in (49) and (52) when $T = 2$ and $\alpha \geq 0$. It is easy to see this if we consider the observed autocorrelation ρ_1. In the case of the heterogeneous model this is given by

$$\rho_1 = \frac{\sigma_\eta^2/\sigma^2}{1 + \sigma_\eta^2/\sigma^2} \tag{56}$$

while in the AR(1) model

$$\rho_1 = \alpha \tag{57}$$

So if for instance it happens that $\sigma_\eta^2/\sigma^2 = 4$ and $\alpha = 0.8$, then ρ_1 will have the same value in both models (numerical example in Arellano (2003)). Following this argument, one can further check that a heterogeneous model of the type (49) with $T = 3$ this time will be indistinguishable from a homogeneous $AR(2)$ model. This discussion can be also carried out using moving average processes instead of autoregressive ones. These few examples suggest is that in order to carry out a nonparametric test of heterogeneity we need to have large N and T and we need to ensure the absence of structural breaks.

One important feature in analyzing panel data is trying to distinguish time effects; it might be for instance often useful to remove business cycles or seasonal effects, to control for trends or demographic characteristics and so on. Consider for discussion a model with individual-specific trends used in the income modelling literature (e.g. Lillard and Weiss (1979), Hause (1980)):

$$y_{it} = \eta_{0i} + \eta_{1i}t + v_{it} \tag{58}$$

This can be written in a vector notation as

$$y_i = S\eta_i + v_i \tag{59}$$

with $y_i = (y_{i1}, ..., y_{iT})'$, $\eta_i = (\eta_{0i}, \eta_{1i})'$, $v_i = (v_{i1}, ..., v_{iT})'$ and S the $T \times 2$ matrix

$$S = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \cdots & \cdots \\ 1 & T \end{pmatrix} \tag{60}$$

If $Var(\eta_i) = \Omega_\eta$ and $v_{it} \sim iid(0, \sigma^2)$ and independent of η_i, the covariance matrix of y_i is

$$\Omega = S\Omega_\eta S' + \sigma^2 I_T \tag{61}$$

In order to identify the parameters of interest, Ω_η and σ^2, we need $T \geq 3$, since with $T = 2$ we will have that the variances of η_{0i}, η_{1i} and v_{it} are well identified only when η_{0i} and η_{1i} are assumed to be uncorrelated. For the case $T = 3$, the identification proceeds as follows. We find the covariance matrix of the variables y_{i1}, Δy_{i2}, and $\Delta y_{i3} - \Delta y_{i2}$, with

$$\begin{aligned} y_{i1} &= \eta_{0i} + \eta_{1i} + v_{i1} \\ \Delta y_{i2} &= \eta_{1i} + v_{i2} - v_{i1} \\ \Delta y_{i3} - \Delta y_{i2} &= v_{i3} - 2v_{i2} + v_{i1} \end{aligned} \tag{62}$$

Then the variance-covariance matrix is given by

$$Var \begin{pmatrix} y_{i1} \\ \Delta y_{i2} \\ \Delta y_{i3} - \Delta y_{i2} \end{pmatrix} = \begin{pmatrix} \sigma_{00} + \sigma_{11} + 2\sigma_{01} + \sigma^2 & \sigma_{11} + \sigma_{01} - \sigma^2 & \sigma^2 \\ & \sigma_{11} + 2\sigma 2 & -3\sigma^2 \\ & & 6\sigma^2 \end{pmatrix} \tag{63}$$

From (63) we can identify both Ω_η and σ^2.

We consider next the issues of stationarity and nonstationarity of moving average processes and the covariance structure estimation. Let us look first at some definitions. Covariance stationarity requires that

$$Cov(y_{it}, y_{i,t-j}) = \gamma_j, \forall t, j \tag{64}$$

In other words the covariances at any lag do not depend on t. A stationary moving average process of order q ($MA(q)$) with individual effects will impose next to (64) above a further restriction

$$\lambda_{q+1} = ... = \lambda_{T-1}, \; q < j \tag{65}$$

If we do not have individual effects the constant in (65) is 0. It is almost intuitive that a nonstationary moving average of order q without individual effects, provided $q < T - 1$, will satisfy

$$Cov(y_{it}, y_{i,t-j}) = 0, \; j > q \tag{66}$$

Estimating and testing the covariance structures is usually done using method of moments or maximum likelihood estimators. We will not insist here on the theoretical part, suggesting for instance the paper of Arellano (1990) for further reading and references, but pass on to the empirical stream and describe a related research therein. An application where one uses the concepts introduced above in a multivariate context and models dynamics through moving average processes appears in Abowd and Card (1989). We will discuss this paper below. Another relevant application in this area is done by Hall and Mishkin (1982), where predictions of the permanent income model are tested; a detailed review of the econometrics in this latter paper is presented in Arellano (2003).

Abowd and Card (1989) put forward an empirical analysis of individual earnings and hours data. Their paper summarizes the main features of the covariance structure of earnings and hours changes and compares it with a structure implied by a simple version of the life-cycle labor supply model. The authors use three different longitudinal surveys, two samples from the Panel Study of Income Dynamics (PSID), a sample of older men from the National Longitudinal Survey of Men 49-59 (NLS) and a sample from the control group of the Seattle and Denver Income Maintenance Experiment (SIME/DIME). For each data set they find evidence that supports the restrictions implied by a nonstationary bivariate MA(2) process without individual effects[5]. In order to control for differences in experience within and between the samples, the covariances between the changes in logs of annual earnings and annual hours were computed using the residuals from multivariate regressions of changes in earnings and hours on time dummies and potential experience. A comparison of these results with previous literature dealing with covariance structure of earnings such as Lillard and Weiss (1979), Hause (1980) or MaCurdy (1982), reveals some possible general results. For instance negative serial correlation between consecutive changes in log earnings seems to be a pervasive phenomenon, a bivariate MA(2) moving average process seems to be adequate for describing the data (with an exception in Lillard and Weiss who find significant large higher-order autocovariances of earnings) and nonstationarity appears to be the rule.

The authors subsequently examine three statistical models that could be considered the generators for the structure of earnings and hours changes above. It is found that a relatively simple components-of-variance model explains the data from all three surveys. This constructed model has three sources of earnings and hours variation:

$$\Delta y_{it} = \binom{\mu}{1} \Delta z_{it} + \Delta u_{it} + \varepsilon_{it} \tag{67}$$

with Δy_{it} a vector of growth rates in earnings and hours of work, Δz_{it} a shared component of earnings and hours variation (a scalar $MA(2)$ component) and Δu_{it} and ε_{it} bivariate white noise processes accounting for the measurement error and respectively the permanent shocks.

[5] In other words, Abowd and Card (1989) obtain a representation where $cov[\Delta \log g_{it}, \Delta \log g_{it-j}]$, $cov[\Delta \log h_{it}, \Delta \log h_{it-j}]$ and $cov[\Delta \log g_{it}, \Delta \log h_{it-j}]$ are constant for all t and are zero for $|j| > 2$, where g represents earnings and h represents worked hours.

As regards interpreting the predicted outcome, according to the life-cycle model the variation in the individual productivity affects earnings more than hours. Abowd and Card's (1989) empirical findings show however that on the contrary, the earnings and hours covary proportionally, which casts doubt on the labor supply interpretation of earnings and hours variation and puts forward the view that most changes in earnings and hours occur at fixed hourly wage rates.

3.2 Autoregressive models with individual effects

The form of regression covariances within autoregressive structures are more complex than those in moving average processes. If the MA processes limit persistence to a given number of periods and imply linear moment restrictions in the covariance matrix of the data, as we have seen in the previous section, with autoregressive processes the situation is different. They imply nonlinear covariance restrictions but provide at the same time instrumental-variable conditions that are linear in the autoregressive coefficients. Moreover the autoregressive model discussion usually centers around themes such as stationarity of initial conditions, homoskedasticity and unit roots, which were typical issues in time-series analysis. Consider the following model

$$y_{it} = \lambda y_{i,t-1} + \alpha_i + \varepsilon_{it},\ t = 0, 1..., T \tag{68}$$

Let us assume that $|\lambda| < 1$. The within-group estimator of the regression equation above is given then by

$$\widehat{\lambda}_{WG} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \bar{y}_i)(y_{i,t-1} - \bar{y}_{i,-1})}{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t-1} - \bar{y}_{i,-1})^2} \tag{69}$$

with $\overline{y}_i = (1/T)\sum_{t=1}^{T} y_{it}$ and $\bar{y}_{i,-1} = (1/T)\sum_{t=1}^{T} y_{i,t-1}$.

It has been shown, using expressions (68) and (69) above (Nickell (1981)) that the within-group estimator is inconsistent for fixed T and it is biased and inconsistent for large N and fixed T. This inconsistency has not been caused by any assumption about the individual effects α, since they are eliminated anyway when estimating the within-group regression. The problem is that the within transformed lagged error is correlated with the within transformed error. The bias disappears with $T \to \infty$, but it is very important for small values of T. Fortunately there are ways to go around this inconsistency. Take first the first differences in (68):

$$y_{it} - y_{it-1} = \lambda(y_{i,t-1} - y_{i,t-1}) + \varepsilon_{it} - \varepsilon_{i,t-1},\ t = 2, ..., T \tag{70}$$

Anderson and Hsiao (1981) proposed an instrumental variable estimator for λ, using $y_{i,t-2}$ as an instrument for (70):

$$\widehat{\lambda}_{IV} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} y_{i,t-2}(y_{it} - y_{i,t-1})}{\sum_{i=1}^{N} \sum_{t=1}^{T} y_{i,t-2}(y_{i,t-1} - y_{i,t-2})} \tag{71}$$

This estimator is consistent given that our assumption of no autocorrelation of the idiosyncratic error term ε_{it} is met and provided that

$$p \lim \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=2}^{T} (\varepsilon_{it} - \varepsilon_{i,t-1}) y_{i,t-2} = 0 \tag{72}$$

Anderson and Hsiao also proposed an alternative instrument, $y_{i,t-2} - y_{i,t-3}$, having a similar estimator expression and consistency condition as in (71) and respectively (72), but having $y_{i,t-2} - y_{i,t-3}$ instead of the lag value $y_{i,t-2}$. Of course this second estimator also requires an additional lag.

A method of moments approach can unify these estimators and eliminate the disadvantages of the small sample sizes. It was thus suggested that the list of instruments is extended by exploiting additional moment conditions and letting their number vary with t. GMM estimators using all available lags at each period as instruments for equations in first-differences were proposed by Holtz-Eakin et al (1988) and Arellano and Bond (1991). This GMM estimator is represented as follows

$$\widehat{\lambda}_{GMM} = \left[\Delta \left(y'_{-1} Z\right) V_N^{-1} \left(Z' \Delta y_{-1}\right)\right]^{-1} \left(\Delta y'_{-1} Z\right) V_N^{-1} \left(Z' \Delta y_{-1}\right) \tag{73}$$

where

$$Z_i = \begin{pmatrix} y_{i0} & 0 & 0 & \ldots & 0 & \ldots & 0 \\ 0 & y_{i0} & y_{i1} & & 0 & & 0 \\ \ldots & & & & \ldots & & \ldots \\ 0 & 0 & 0 & \ldots & y_{i0} & \ldots & y_{i,T-2} \end{pmatrix} \tag{74}$$

According to standard GMM theory, an optimal choice of the inverse weighting matrix V_N is one that gives the most efficient estimator, in other words the one that gives the smallest asymptotic covariance matrix for $\widehat{\lambda}_{GMM}$. Since the optimal weighting matrix is asymptotically proportional to the inverse of the covariance matrix of the sample moments (Verbeek (2000)), we have that the optimal matrix ought to satisfy

$$p \lim_{N \to \infty} V_N = E(Z'_i \Delta\varepsilon_i \Delta\varepsilon'_i Z_i)^{-1} \tag{75}$$

In the standard case where we do not impose any restrictions on the covariance matrix of ε_{it} we can obtain the optimal weighting matrix using a first-step consistent estimator of λ and replacing the expectation in (75) by a sample average:

$$\hat{V}_N^{opt} = (\frac{1}{N}\sum_{i=1}^{N} Z_i'\Delta\widehat{\varepsilon}_i\Delta\widehat{\varepsilon}_i' Z_i)^{-1} \tag{76}$$

where $\Delta\widehat{\varepsilon}_i$ is a residual vector obtained for instance by using $V_N = I$ as a first-step consistent estimator. We have estimated here the optimal weighting matrix without imposing restrictions of homoskedascity; in particular for the estimation we do not even need the non-autocorrelation assumption but this is needed for the validity of the moment conditions.

There is a whole literature on assumptions about the initial conditions, homoskedasticity and mean stationarity with subtopics such as estimation under stationarity, estimation under unrestricted initial conditions, estimation under heteroskedasticity, time effects in AR models, initial condition bias in short panels. Next to these, quite a number of relatively recent papers discuss problems of unit roots, cointegration and spurious regression within panel data. These latter problems are particularly relevant in very long time-series dimension of the longitudinal data sets (if $T \to \infty$). These are all very important themes and much can be written on them, however we will not cover them in this general report. The interested reader is advised to consult Arellano (2003) and the further references therein for an excellent treatment of the subject or Arellano and Honore (2001) for a shorter but concise overview.

4 Dynamic regression models

4.1 Models with exogenous regressors and unrestricted serial correlation

Consider models of the type

$$y_{it} = \alpha y_{i,t-1} + x_{it}'\beta + \eta_i + v_{it} \tag{77}$$

where we assume that

$$E(v_{it}|x_{i1}, ..., x_{iT}, \eta_i) = 0 \tag{78}$$

Although this would seem exactly the autoregressive model analyzed in the previous section plus some exogenous variables, this is not the case. While the autoregressive model was a time-series perspective of modelling the persistence in $y's$, hence we assumed the $v's$ serially uncorrelated, now we have an entirely different view. Lagged y is correlated by construction with η and lagged v, but it might even be correlated with contemporaneous v if v is serially correlated, which is not precluded by (78). Thus lagged y is in fact an endogenous explanatory variable in (77) with respect to both η and v.

From assumption (78) we can work out, as in Arellano (2003),

$$E[x_{is}(\Delta y_{it} - \alpha \Delta y_{i,t-1} - \Delta x'_{it}\beta)] = 0, \forall s, t \tag{79}$$

We have therefore internal moment conditions that will ensure identification of the model despite the unrestricted serial correlation and the endogeneity of y.

Applications of this model arise in testing life-cycle models of consumption of labor supply with habits. In such models the parameter of interest is the coefficient α of the lagged regressor, which measures the extent of habits. This parameter could not be identified in the absence of exogenous IV's since the effect of the genuine habits would be indistinguishable from the serial correlation in the unobservable. One cited example in this stream of literature is Becker et al (1994). They consider a model for cigarette consumption using US panel data:

$$c_{it} = \theta c_{i,t-1} + \beta\theta c_{i,t+1} + \gamma p_{it} + \eta_i + \delta_t + v_{it} \tag{80}$$

where c_{it} - annual per capita cigarette consumption by state, in packs and p_{it} - average cigarette price per pack; β - discount factor. Becker et al are interested in investigating whether smoking is addictive by considering the response of cigarette consumption to a change in prices.

The idiosyncratic errors capture unobserved life-cycle utility shifters, likely to be serially correlated. Thus even in the absence of addiction ($\theta = 0$), we would still expect c_{it} to be autocorrelated, and in particular to find a non-zero effect of $c_{i,t-1}$ in (80). The identification strategy relies thus on identifying θ, β and γ from the assumption that prices are strictly exogenous relative to the unobserved utility shift variables.

4.2 Models with predetermined variables

The models in this section have idiosyncratic errors that are uncorrelated with current and lagged values of the conditioning variables, but might be correlated with their future values, being "predetermined" in this sense. The conditioning variables can be explanatory variables in the equation or lagged values of these variables, but also external instruments.

The moment conditions satisfied by the errors in the models with predetermined variables are sequential moment conditions of the type

$$E(v_{it} | z_{i1}, ..., z_{it}) = 0 \tag{81}$$

Take for illustration the partial adjustment model in (77) in the preceding subsection and formulate its sequential moment condition requirement:

$$E(v_{it}|y_{i,t-1},x_{it},\eta_i) = 0 \tag{82}$$

It is interesting to seek the differences between this assumption above and the corresponding assumption in the strictly exogenous model in the previous subsection, namely expression (78). Arellano and Honore (2001) tell us that there are two main differences. In the first place (82) implies lack of autocorrelation in v_{it} whereas the previous model allowed for unrestricted serial correlation. Secondly, the assumption in the strictly exogenous variables model from (78) implies that forecasts of x_{it} given $x_{i,t-1}$, $y_{i,t-1}$ and additive effects are not affected by $y_{i,t-1}$.

In terms of identification, let us take the example of $T = 3$ in Arellano and Honore (2001) where we had the following equation in first differences

$$y_3 - y_2 = \alpha(y_2 - y_1) + \beta_0(x_3 - x_2) + \beta_1(x_2 - x_1) + (v_3 - v_2) \tag{83}$$

Using the exogenous variable model assumption, the parameters α, β_0 and β_1 are potentially identifiable just from the moment conditions

$$E(x_{is}\Delta v_{i3}) = 0,\ s = 1, 2, 3 \tag{84}$$

Conversely, in the predetermined variables model, the moment conditions that potentially identify α, β_0 and β_1 are

$$\begin{cases} E(y_{i1}\Delta v_{i3}) = 0 \\ E(x_{i1}\Delta v_{i3}) = 0 \\ E(x_{i2}\Delta v_{i3}) = 0 \end{cases} \tag{85}$$

The problem is that the two models have only two moment conditions in common which are not however sufficient to identify the three parameters (Arellano (2003)).

5 Discrete choice models

The discrete or limited dependent variables combined with panel data often complicates considerably the estimation. This happens because in longitudinal data sets often the different observations on the same unit are not independent which leads to correlation between different error terms and subsequently complicates the likelihood functions of such models. One good overview of the literature in this specific fields is Maddala (1987), while for an up-to-date presentation of the current status of research in the area the reader can consult Arellano (2001).

Consider for illustration a binary choice model:

$$y_{it}^* = x_{it}'\beta + \alpha_i + \varepsilon_{it} \tag{86}$$

where we observe $y_{it} = 1$ if $y_{it}^* > 0$ and $y_{it} = 0$ otherwise. We can treat α as a fixed unknown parameter which means that the log likelihood function is given by

$$\log L = \sum_{i,t} y_{it} \log F(\alpha_i + x_{it}'\beta) + \sum_{i,t} (1 - y_{it}) \log[1 - F(\alpha_i + x_{it}'\beta)] \tag{87}$$

Maximizing this function with respect to β and α_i results in consistent estimators only provided that $T \to \infty$. This is known as the incidental parameters problem.

Fortunately for linear models β can be identified using the within-group estimator as long as $T \geq 2$. The consistent estimator for β is given by the solution $\widehat{\beta}$ to the equation (Arellano (2001)):

$$\frac{1}{TN} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \overline{x_i})[(y_{it} - \overline{y_i}) - (x_{it} - \overline{x_i})'\widehat{\beta}] = 0 \tag{88}$$

However this does not solve our problem for nonlinear models and thus we need to think of a way to maximize the likelihood. One idea is to find a sufficient statistic for α_i so that the distribution of the data given this statistic does not depend on α_i, and to use the likelihood conditional on it so as to make inferences about β. This has been implemented first in Andersen (1970) and it is known under the name of conditional maximum likelihood estimator.

For the linear model with normal errors the sufficient statistic for α is the mean $\overline{y_i}$. It is easy to show that for these linear setting, maximizing the conditional maximum likelihood by using this statistic is equivalent with using the within-group estimator for β. However the result cannot be generalized to nonlinear models since it has been shown for instance that no statistic of the type required here exists for the probit model, reason why a fixed effects probit model cannot consistently be estimated for fixed T (see Verbeek (2000) for a review).

In the random effects setting we assume lack of correlation in the idiosyncratic error term

$$u_{it} \equiv a_i + \varepsilon_{it} \tag{89}$$

which implies that one can write the joint probability as

$$f(y_{i1},...,y_{iT}|x_{i1},...,x_{iT},\beta)=\int_{-\infty}^{\infty}\left[\prod_{t}f(y_{it}|x_{it},\alpha_i,\beta)\right]f(\alpha_i)d\alpha_i \quad (90)$$

which requires in principle numerical integration over one dimension. However this is extremely difficult and hence the most common practice is to start with an assumption of multivariate normal distribution of u_{i1},...,u_{iT} (see Maddala (1987) for a more in depth discussion). This brings us to the random effects probit model. In practice using standard probit maximum likelihood is consistent, but inefficient. The results can be used however in an iterative maximum likelihood procedure based on the expression in (90).

There is a fast growing literature on discrete choice models with panel data. It is not our purpose to cover here for instance maximum score estimation (see Manski (1987)) or maximizing alternative types of likelihood functions As also mentioned above, a good survey of recent research in the field is contained in Arellano (2001).

References

Abowd, J.M. and D. Card (1989), "On the Covariance Structure of Earnings and Hours Changes", *Econometrica*, 57, 411-445

Andersen, E., B. (1970), "Asymptotic Properties of Conditional Maximum Likelihood Estimation", *Journal of the Royal Statistical Society*, Series B, 32, 283-301

Anderson, T. and C. Hsiao (1981), "Estimation of Dynamic Models with Error Components", *Journal of the American Statistical Association*, 76, 598-606

Arellano, M. (1990), "Testing for Autocorrelation in Dynamic Random Effects Models", *Review of Economic Studies*, 57, 127-134

Arellano, M. (2001), "Discrete Choices with Panel Data", CEMFI, Working Paper 0101

Arellano, M. (2003), *Panel Data Econometrics*, Oxford University Press: Advanced Texts in Econometrics

Arellano, M. and S.R. Bond (1991), "Some Tests of Specifications 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

Arellano, M. and B. Honore (2001), ”Panel Data Models: Some Recent Developments”, in J.J. Heckman and E. Leamer (eds.): *Handbook of Econometrics*, vol 5, ch. 53, North-Holland

Balestra, P. and M. Nerlove (1966), " Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas", *Econometrica*, 34, 585-612

Baltagi, B. H. (2001), *Econometric Analysis of Panel Data*, Wiley, 2nd edition

Becker, G., M. Grossman and K. Murphy (1994), " An Empirical Analysis of Cigarette Addiction", *American Economic Review*, 84, 396-418

Blundell, R. (1988), "Consumer behavior: Theory and Empirical Evidence- A Survey", *The Economic Journal* 98, 16-65

Blundell, R. and C. Meghir (1990), "Panel Data and Life-Cycle Models", in J. Hartog, G. Ridder and J. Theeuwes (eds.): *Panel Data and Labor Market Studies*, North-Holland, Amsterdam, 231-252

Chamberlain, G. (1982), "Multivariate Regression Models for Panel Data", *Journal of Econometrics*, 18, 5-46

Chamberlain, G. (1984), "Panel Data", in Z. Griliches and M.D. Intriligator (eds.): *Handbook of Econometrics*, vol. 2, ch. 22, North-Holland

Chamberlain, G. (2000), "Econometrics and Decision Theory", *Journal of Econometrics*, 95, 255-283

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

Hall, R. and F. Mishkin (1982), "The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households", *Econometrica*, 50, 461-481

Hause, J.C. (1980), "The Fine Structure of Earnings and On-The-Job Training Hypothesis", *Econometrica*, 48, No. 4, 1013-29

Hausman, J.A., W.K. Newey, and J.L. Powell (1995), "Nonlinear Errors in Variables Estimation of Some Engel Curves", *Journal of Econometrics*, 65, 205-233

Holtz-Eakin, D., W. Newey and H. Rosen (1988), "Estimating Vector Autoregressions with Panel Data", *Econometrica*, 56, 1371-1395

Hsiao, C. (2003), *Analysis of Panel Data*, Cambridge University Press, 2nd Edition

Lillard, L. and Y. Weiss (1979), "Components of Variation in Panel Earnings Data: American Scientists 1960-1970", Econometrica, 47, 437-454

MaCurdy, T. E. (1982), "The Use of Time-Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis", *Journal of Econometrics*, 18, 83-114

Maddala, G.S. (1987), "Limited Dependent Variable Models Using Panel Data", *The Journal of Human Resources*, 22, 307-338

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

Nickell, S. (1981), "Biases in Dynamic Models with Fixed Effects", *Econometrica*, 49, 1417-1426

Verbeek, M. (2000), *A guide to Modern Econometrics,* Wiley

Wooldridge, J. M. (2002), *Econometric Analysis of Cross-Section and Panel Data*, MIT