



The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

Marc Nerlove and Pietro Balestra

1. FORMULATION AND ESTIMATION OF ECONOMETRIC MODELS FOR PANEL DATA*

The fundamental fact about society as a going concern is that it is made up of individuals who are born and die and give place to others; and the fundamental fact about modern civilization is that it is dependent upon the utilization of three great accumulating funds of inheritance from the past, material goods and appliances, knowledge and skill, and morale. Besides the torch of life itself, the material wealth of the world, a technological system of vast and increasing intricacy and the habituations which fit men for social life must in some manner be carried forward to new individuals born devoid of all these things as older individuals pass out.

Waite Library

Applied Economics - U of M
1994 Buford
St Paul MN 55108-6040 USA

Frank H. Knight (1921)

The moral of Knight's characterization is that history is important and individuals have histories. This should not be forgotten as we embark on this study.

In his famous and influential monograph, *The Probability Approach in Econometrics*, Haavelmo [1944] laid the foundations for the formulation of stochastic econometric models and an approach which has dominated our discipline to this day. He wrote:

... we shall find that two individuals, or the same individual in two different time periods, may be confronted with exactly the same set of specified influencing factors [and, hence, they have the same y^* , ...], and still the two individuals may have different quantities y , neither of which may be equal to y^* . We may try to remove such discrepancies by introducing more "explaining" factors, x . But, usually, we shall soon exhaust the number of factors which could be considered as common to all individuals, and which, at the same time, were not merely of negligible influence upon y . The discrepancies $y - y^*$ for each individual may depend upon a great variety of factors, these factors may be different from one individual to another, and they may vary with time for each individual. (Haavelmo [1944], p. 50).

*Nerlove's contribution was supported by the Maryland Agricultural Experiment Station, Project __, and will appear as MAES Paper No. __.

WP 95-13

Waite Library
Applied Economics - U of M
1994 Buford
St Paul MN 55108-6040 USA

And further that:

... the class of populations we are dealing with does not consist of an infinity of different individuals, it consists of an infinity of possible decisions which might be taken with respect to the value of y.

... we find justification for applying them [stochastic approximations] to economic phenomena also in the fact we usually deal only with — and are interested only in — total or average effects of many individual decisions, which are partly guided by common factors, partly by individual specific factors ...
(Haavelmo [1944], pp. 51 and 56).

Marschak ([1950];[1953]) further amplified Haavelmo's themes in his introductions to Cowles Commission Monographs 10 and 14, observing that:

The numerous causes that determine the error incurred ... are not listed separately; instead their joint effect is represented by the probability distribution of the error, a random variable (1950, p. 18) [, which] ... is called 'disturbance' or 'shock,' and can be regarded as the joint effect of numerous separately insignificant variables that we are unable or unwilling to specify but presume to be independent of observable exogenous variables. (1953, p. 12).

Since the early work of Mundlak [1961] and Balestra and Nerlove [1966], panel or longitudinal data have become increasingly important in econometrics, and methods for the analysis of such data have generated a vast literature much of which was been summarized in the first edition of this volume (Mátyás and Sevestre, 1992). In the last three years there has been an extraordinary further growth, captured here in eleven completely new chapters and seven significantly revised chapters, which appeared in the earlier edition.

In this introduction we examine how the basic principle underlying the formulation of econometric models has been carried forward in the development of econometric models and methods for the analysis of panel

data. We argue that while fixed effects models may be appropriate in cases in which a population is sampled exhaustively (e.g., data from geographic regions over time) or in which it is desired to predict individual behaviour (e.g., the probability that a given individual in a sample will default on a loan), random effects models are more consistent with Haavelmo's view, quoted above, that the "population" we model in econometrics consists *not* of an infinity of individuals, in general; but of an infinity of *decisions*. This is not to say, however, that fixed effects models may not be extremely useful as an analytic device.

Moreover, we shall argue, taking a leaf from Knight [1921], that what differentiates the individuals, who make the decisions with which we are concerned, is largely historical, the "three great accumulating funds of inheritance from the past, material goods and appliances, knowledge and skill, and morale." This view has important implications for the relevance and appropriateness of many of the models and methods for the analysis of panel data which have been developed over the past 30 years. We critically review these developments here and conclude that not only are random effects models most relevant and appropriate but that often our central analytical and modelling concerns are also dynamic. Thus, the most fruitful developments in this enormous literature have been those which deal with the central issues of history and dynamics.

1.1. History and Dynamics: How Should We View the Disturbances?

Before turning to a review of the state of the art in panel econometrics, we illustrate our general view of the central principle involved using a simple illustrative example drawn from a recent paper of Matyas and Rahman [1992].

Let i index individuals and t time periods. Suppose the relationship we are interested in estimating is

$$y_{it} = \sum_{s=0}^{\infty} \beta_s x_{i,t-s} + \varepsilon_{it} . \quad (1-1)$$

The variable x_{it} is assumed to be exogenous and distributed independently of the true disturbances ε_{it} for all finite subsets of the t -index set. We also assume, despite our previous injunction, that

$$\begin{aligned} E\varepsilon_{it} &= 0, \quad \text{all } i \text{ and } t, \\ E\varepsilon_{it'}\varepsilon_{it} &= \sigma_{\varepsilon}^2, \quad i = i' \text{ and } t = t' \\ &= 0, \quad i \neq i' \text{ or } t \neq t'. \end{aligned} \quad (1-2)$$

To guarantee some stability in the relationship we are interested in estimating, we must also assume some convergence properties for the sequence of distributed lag weights. Although stronger than necessary, assume they are square-summable:

$$\sum_{s=0}^{\infty} \beta_s^2 < \infty . \quad (1-3)$$

Of course, as Matyas and Rahman note, (1-1) is not estimable with a finite amount of data. Indeed, the time dimension is likely to be very short. Instead, we truncate:

$$\begin{aligned} y_{it} &= \sum_{s=0}^k \beta_s x_{i,t-s} + \sum_{s=k+1}^{\infty} \beta_s x_{i,t-s} + \varepsilon_{it} \\ &\equiv \sum_{s=0}^k \beta_s x_{i,t-s} + \mu_i + \varepsilon_{it} . \end{aligned} \quad (1-4)$$

Equation (1-4) is in the form of a frequently used random effects model, except that now the individual-specific effects are interpreted in terms of the past histories of each individual in the panel prior to the time when observation begins. Moreover, the assumption that x_{it} is stochastic, although exogenous is not innocuous. The implications are:

First, interpreting μ_i as fixed, i.e., nonstochastic, is not appropriate. If one accepts Haavelmo's view that the class of populations which we imagine (1-4) reflects consists of decisions rather than identifiable specific individuals, then, in principle, we should not even condition on μ_i . However, an exception to this rule is if, for the particular sample of individuals we have drawn (now we can specifically identify each), we want to predict future values of y_{it} for that individual.

Second, since the x_{it} are themselves considered to be stochastic, for each individual their values over time will in general be correlated. There may also be correlations among x_{it} 's for different values of i if different individuals have some characteristics in common. But we neglect this possibility here. It follows that μ_i and the values x_{it} observed are correlated. Suppose, for example,

$$x_{it} = \rho_i x_{i,t-1} + v_{it}, \quad (1-5)$$

where $|\rho_i| < 1$ and $Ev_{it} = 0$, $Ev_{it}v_{i't'} = 0$, $i \neq i'$ or $t \neq t'$, and

$Ev_{it}v_{i't'} = \sigma_i^2$, $i = i'$ and $t = t'$, for all i and t . Let $S = \{0, 1, \dots, K\}$ be the set of indices for which, given i , x_{it} is observed (normally k will be chosen much less than K). Since

$$Ex_{it} = 0$$

$$Ex_{it}x_{i,t-\tau} = \frac{\rho_i^\tau}{1 - \rho_i^2} \sigma_i^2, \quad (1-6)$$

it follows that x_{it} , $t \in S$, will be correlated with μ_i , and the correlation will depend on how close to the beginning of the sample period the observation on x_{it} is taken:

$$\begin{aligned}
Ex_{it} \mu_i &= \sum_{s=k+1}^{\infty} \beta_s Ex_{it} x_{i,t-s} \\
&= \frac{\sigma_i^2}{1-\rho_i^2} \sum_{s=k+1}^{\infty} \beta_s \rho_i^{t-s}, \quad (1-7)
\end{aligned}$$

for $\tau \in S$. Clearly, this makes the likelihood of the sample much more difficult to determine and introduces some of the parameters, namely β_s , into the relationship between the individual-specific disturbances in (1-4) and the observed past values of the explanatory exogenous variable. (We would perhaps be willing to regard σ_i^2 and ρ_i as nuisance parameters.)

The important point about this admittedly unrealistic example is that it shows that an entirely new set of questions must be considered. In particular, the error which we make by treating μ_i as independent of the observed values of x_{it} now depends in a complex way on the way in which the distributed lag parameters of interest interact with the nuisance parameters σ_i^2 and ρ_i . Indeed, matters become even more interesting when we note that the unconditional variance of x_{it} is $\sigma_i^2 / (1 - \rho_i^2)$, so that, in general the greater σ_i^2 the greater is the signal to noise ratio in (1-4), on the one hand, but, *ceteris paribus*, the greater is the dependence between x_{it} and μ_i , especially for τ near the beginning of the observation period.

How can we optimally rid ourselves of the nuisance parameters? How badly does a method, which is based on the assumption that μ_i and the observed x_{it} , are uncorrelated, approximate the true ML estimates? What is the appropriate likelihood function in the dynamic case? What constitute appropriate instruments in considering alternative methods to ML? And so forth.

With these general principles in mind, we now turn to a review of the volume and the recent developments in panel data econometrics.

1.2 Methodological Developments

Besides Chapters 2-24 of this volume, which survey methodological developments in great detail, the following are also valuable: the recent survey of Baltagi and Raj [1992], which focuses on estimation of error components models extended to allow for serial correlation, heteroskedasticity, seemingly unrelated regressions and simultaneous equations; the supplement to the *Journal of Econometrics*, 59 [1993] edited by Carraro, Peracchi and Weber, which contains several important methodological surveys as well as a number of applied studies, and the much older survey of Chamberlain [1984]. Chamberlain's paper is more than a survey; it presents a very general framework for the analysis of panel data problems which is summarized and elaborated in a chapter added to this volume in the second edition by Crepon and Mairesse (Chapter 14).

The most common model for the analysis of panel data is the linear model in which explanatory variables are taken to be exogenous, that is independent of the disturbances in the equation or, in the case of the random coefficients model of the distributions of the coefficients. When the coefficients (except for the constant term) in the linear relationship with which we describe the data are assumed to be constant, it is usual to distinguish between *fixed effects* and *error components* models. In the case of the former, the intercepts are assumed to vary across individuals at the same point in time and, possibly, over time for all individuals taken together. In the case of the latter, the variations are assumed to be random and uncorrelated both with the

observed explanatory variables and the latent disturbance in the equation.

A considerable quantity of interesting mathematics has been developed for both types of models. In particular, the Appendix to part I, based on unpublished work of Trognon, presents a series of projection matrices which take deviations between the raw observations and various means, across individuals, across time periods, over all, and of various means from other means. These projections can be used to define different possible estimators in fixed effects models or the spectral decomposition of the disturbance variance-covariance matrix in the case of error components models. A principal result is then the demonstration, first noted by Maddala [1971], that the Generalized Least Squares (GLS) estimators of the slope parameters in the error components case are a weighted combination of estimators in the fixed effects case (the so-called "between" and "within" distinction among possible estimators). See Chapters 2, 3 and 4.

Seemingly unrelated regression (SUR) models and simultaneous equations models (Chapter 9) as well as extension to various forms of nonspherical disturbances follow more-or-less effortlessly through creative use of "stacking."

An important distinction is made between fully *asymptotic* theory in which the limiting properties of estimators are analysed when both the number of time periods and the number of individuals goes to infinity and *semi-asymptotic* theory in which the number of individuals (or the number of time observations) is assumed to increase without bound, that is, asymptotics in only one of two dimensions. Clearly, in the case of random effects models, the moments of the distribution of the effect whose dimension is not increased in the calculation cannot be semi-asymptotically consistently estimated.

As long as the model is not dynamic, that is, does not contain a distributed lag, lagged values of the dependent variable, or the equivalent stock or state variable, the GLS estimators of these coefficients have the usual good small sample and asymptotic properties. The problem, then, is that the elements of the disturbance variance-covariance matrix are unknown. Since consistency of the variance components estimates depends on the asymptotics assumed, the usual justification for a two-stage procedure (feasible GLS or FGLS) based on first-stage consistent estimates of the variances and covariances of the panel model disturbances does not clearly apply. Indeed, in some senses the FGLS may not even be consistent. Moreover, as we have argued, most relationships of interest are likely to be dynamic and the past histories of individuals are almost always important determinants of current behavior. In a very important chapter (7 in this edition), Sevestre and Trognon establish the inconsistency of the coefficient of the lagged dependent variable *for all feasible non-ML estimates and for GLS*.

The GLS estimates are obtained by transforming the observations to weighted sums of between and within, using appropriate weights based on the characteristic roots of

$$\begin{aligned}\sigma^2\Omega &= \sigma^2\{ \rho(I_N \otimes J_T) + (1 - \rho)I_{NT}\} \\ &= \sigma^2\{ \rho(I_N \otimes J_T) + (1 - \rho)(I_N \otimes I_T)\} \\ &= \sigma^2\{ I_N \otimes (\rho J_T + (1 - \rho)I_T)\},\end{aligned}$$

$$\xi = (1-\rho) + T\rho \text{ and } \eta = (1-\rho),$$

where $\sigma^2 = \sigma_\mu^2 + \sigma_\epsilon^2$, transforms the variance covariance matrix of the disturbances $u_{it} = \mu_i + \epsilon_{it}$ in (1-8) below to $\sigma^2 I_{NT}$. Applying this

transformation to (1-8) and replacing $\frac{\eta}{\xi} = \frac{1}{\theta^2} = \lambda$, the normal equations to

be solved for the GLS estimates become:

$$\begin{pmatrix} W_{yx} + \lambda B_{yx} \\ W_{yy-1} + \lambda B_{yy-1} \end{pmatrix} = \begin{bmatrix} W_{xx} + \lambda B_{xx} & W_{y-1x} + \lambda B_{y-1x} \\ W_{xy-1} + \lambda B_{xy-1} & W_{y-1y-1} + \lambda B_{y-1y-1} \end{bmatrix} \begin{pmatrix} \beta \\ \gamma \end{pmatrix}.$$

In this case, the calculated RSS/NT estimates not σ^2 but $\eta^{-1}\sigma^2$. As Maddala [1971] points out, the GLS estimates with $\lambda = 1/\theta^2$ can be considered members of a more general class of estimators obtained through different choices of λ . Let $\hat{\gamma}(\lambda)$ be the estimator of γ obtained by solving the above equations for an arbitrary value of λ . Sevestre and Trognon show that for the case in which $\beta = 0$, the purely autoregressive case, the following inequality holds:

$$\begin{matrix} p\lim \hat{\gamma}(0) < \gamma < p\lim \hat{\gamma}(\theta^2) < p\lim \hat{\gamma}(1) < p\lim \hat{\gamma}(\infty) \\ \text{within} & \text{GLS} & \text{OLS} & \text{between} \end{matrix}.$$

Remarkably, therefore, the GLS estimate is inconsistent in this case. The problem is that the lagged dependent variable is correlated even with the transformed disturbance.

Since $p\lim \hat{\gamma}(\lambda)$ is a continuous function of λ , there exists a value λ^* in the interval $[0, \theta^2]$ for which $p\lim \hat{\gamma}(\lambda) = \gamma$. In an earlier paper, Sevestre and Trognon [1983] have derived this value. They also show that when $\beta \neq 0$, the estimate $\hat{\gamma}(\lambda)$ behaves almost the same as in the purely autoregressive case. Since the λ^* estimate is consistent when there are no exogenous variables, it remains so when there are. The trick is to obtain a consistent estimate of λ^* which can be accomplished by finding an appropriate instrumental variable for y_{-1} . Even in this case the results depend heavily on the distribution of the estimate of λ^* .

In the dynamic error-components model, not only are the OLS pooled regression estimates, the fixed-effect or within estimates, and the between estimates inconsistent, but so are the GLS estimates using the true value of ρ . However, the method of instrumental variables may be used to obtain a feasible member of the λ -class of estimates which is consistent.

Unfortunately, this estimate may have a very large variance.

The method of choice in most cases is Maximum Likelihood (ML), provided, of course, that associated computational difficulties can be resolved. But even when the matrix of observed regressors is assumed to be nonstochastic, the properties of ML estimators may no longer be fully optimal asymptotically. See for example Chapter 4, Table 4-2. Although consistent ML estimates of the coefficients of observed exogenous and of the nonspecific residual variance can be obtained either in the asymptotic or the semi-asymptotic sense, consistent ML estimates of the individual specific residual variance cannot be obtained except in the semi-asymptotic sense. In the dynamic case, however, maximum likelihood based on the likelihood function conditional on the initial observation, or more generally the state, can yield inconsistent estimates (Trognon [1978]).

Various interesting extensions of both the fixed effects and error components linear models have recently been made:

- (a) To random coefficients (Chapter 5);
- (b) To linear models with random regressors (Chapter 6);
- (c) To data with measurement errors (Chapter 10);
- (d) To dynamic models (Chapters 7 and 8);
- (e) To simultaneous equations models (Chapter 9).

In addition there are chapters on specification problems (12), the Bayesian approach to pooling (13), and an exposition of the Chamberlain

approach (14), which permits a unified treatment of both fixed effects and random effects models in an instrumental variables context.

Maximum-Likelihood Estimation of the Dynamic Model

As suggested above, from the conceptual standpoint the dynamic model is the most interesting and, as indicated, the properties of both dynamic fixed effects and error components models are dramatically different than we might expect on the basis of experience with both standard linear econometrics (OLS, MS, etc.) or with panel models with strictly exogenous regressors. In the sense that the likelihood function summarizes all that is relevant of the empirical data at hand, it is useful to see what the implications of a simple random-effects model are. Following Chapter 6, suppose the model is

$$y_{it} = \alpha y_{i,t-1} + x'_{it}\beta + \mu_i + \varepsilon_{it},$$

$$i = 1, \dots, N, \quad t = 1, \dots, T. \quad (1-8)$$

If only semi-asymptotics on N is considered (T finite), we need not assume $|\alpha| < 1$. On the other hand, the process generating the initial observations is very important. As suggested above, this means that the individuals' past history with respect to both the observed variables x and the latent variables ε become crucial.

We can rewrite (1-8) as

$$y_{it} = \alpha^t y_{i0} + \sum_{j=0}^{t-1} \alpha^j x'_{i,t-j} \beta + \frac{1 - \alpha^t}{1 - \alpha} \mu_i + v_{it}, \quad (1-9)$$

where

$$v_{it} = \sum_{j=0}^{t-1} \alpha^j \varepsilon_{i,t-j}.$$

Thus, each observation on the dependent variable y_{it} can be written as the sum of four terms:

The first, $\alpha^t y_{i0}$, depends on the initial values which, as long as T is finite, do influence the behaviour of any estimators. Moreover, there is no good reason (as in Balestra-Nerlove [1966]) to assume that these are fixed (i.e., to condition upon their values) and independent of individual specific effects. Indeed, unless there is something special about the initial date of observation (see Chapters 19, 26, and 31), there is no justification for treating the initial observation differently from subsequent observations or from the past, but unobserved, previous values.

The second term in (1-9) depends on the current and past values of the exogenous variables x'_{it} . The form that this dependence takes depends not only on the dynamics of the model, but also on the way in which individuals' past histories differ (Knight [1921]).

The third term depends on remaining individual specific effects which are assumed to be orthogonal to the individual's past history.

Finally, the last term is a moving average in past values of the remaining disturbances, which may also be written:

$$v_{it} = \alpha v_{i,t-1} + \varepsilon_{it}, \quad t \geq 1,$$

$$v_{i0} = 0, \quad t = 0.$$

Conditioning on the initial observations implies that they can be treated as fixed constants independently of μ and v_{it} . They need not be independent of any of the lagged values of the explanatory x 's which are included. But if any truncation within-sample occurs, the truncation remainder will be part of the individual specific disturbance, as shown above, and thus the initial values of the endogenous variable are independent of the disturbance and cannot be treated as fixed.

This point can be made in another way (following Chapter 7):
Write the cross section of initial observations as a function of past x 's, μ_i , and ε_{i0} ,

$$y_{i0} = f(x'_{i,0}, x'_{i,-1}, \dots, \mu_i, \varepsilon_{i0}) . \quad (1-10)$$

The problem is now related to whether or not we choose to regard μ_i as fixed or random. If μ_i is fixed and thus independent, cross-sectionally, of ε_{i0} , and if $x'_{i,t-j}$, $j = 0, 1, \dots$, are cross-sectionally exogenous, then the y_{i0} can be conditioned on. They are still, however, random variables. But, if the μ_i are random variables, the y_{i0} are not exogenous. This shows that in a dynamic context fixed effects versus error components assumptions make a big difference. Our preceding argument suggests that the error components assumption is the more appropriate.

In this case, recent literature suggests a variety of different assumptions about the initial observation leading to different optimal estimation procedures and implying different properties for suboptimal estimates. One line takes the generating process of the initial observations to be different from that of subsequent observations. Anderson and Hsiao [1982], for example, suggest a general form

$$y_{i0} = k_0 + k_1\mu_i + k_2\varepsilon_{i0} . \quad (1-11)$$

If $k_1 = k_2 = 0$, the initial observations are fixed and identical. If $k_0 = k_1 = 0$ and $k_2 \neq 0$, the y_{i0} are random variables independent of the disturbances in (1-8). If $k_0 = 0$, $k_1 = 1/(1 - \alpha)$ and $k_2 = 1/(1 - \alpha^2)^{1/2}$ the individual autoregressive processes which generate the y 's are stationary, etc.

But, although convenient, it is not very reasonable to suppose the initial observation to be generated by a mechanism much different than

that which generates subsequent observations. Bhargava and Sargan [1983] suggest

$$y_{i0} = k_0 + x_{i0}''\gamma + k_1\mu_1 + k_2\varepsilon_{i0}, \quad (1-12)$$

where the x_{i0}'' are exogenous variables, possibly different from x'_{i0} but quite possibly correlated with subsequent observed x'_{i0} 's and where γ may or may not equal β . This formulation obviously encompasses the stricter assumption that the same mechanism generates y_{i0} and subsequent y_{it} 's and allows the exogenous variables themselves to be generated by other independent dynamical systems.

Assuming fixed effects in dynamic framework and estimating them as if they were constants (or eliminating them by taking deviations from individual means) together with the autoregressive coefficient α leads to inconsistent estimates of the latter. This was noted in Nerlove [1971], Nickell [1981] and proved by Sevestre and Trognon [1985]. Although $y_{i,t-1}$ and ε_{it} are uncorrelated, their respective individual means are correlated with each other, with ε_{it} and with $y_{i,t-1}$.

Instrumental variable methods have been proposed to get around this problem (e.g., Balestra and Nerlove [1966]), but as shown in Nerlove [1971], they can result in very erratic estimates if the instruments themselves have relatively low explanatory value.

Conditioning on the initial values of the endogenous variable also leads to troublesome problems. As noted in Nerlove [1971], the estimates of α appear to be inconsistent even when an error components model is assumed and σ_{jt}^2 and σ_e^2 are estimated together with other parameters of the model. This was proved in Trognon [1978]. Bhargava and Sargan [1983] show that this does not happen when the likelihood function is unconditional, i.e., when it takes into account the density

function of the first observation, e.g., as determined by (1-12) and assumptions about the k 's, γ , and the densities of μ_1 and ε_{10} . Our opinion on this matter is that it is most plausible and appropriate to assume that the mechanism which generates the initial observation is highly similar, if not identical, to that which generates subsequent observations. If observations on past values of the exogenous variables are not generally available, it would be preferable to model their independent determination rather than to assume their joint effect, $X_{10}'\gamma$, to be fixed constants. At least, such an approach would be more consistent with Haavelmo's views as quoted above.

When the solution to the likelihood equations (scores = 0) is not on a boundary and when the likelihood function is locally concave at such a solution, the solution with the largest value is consistent, asymptotically efficient, and root-N asymptotically normally distributed with variance-covariance matrix equal the inverse information matrix. Provided the marginal distribution of the initial values y_{10} , $i = 1, \dots, N$, can be correctly specified, the unconditional density of y_{11}, \dots, y_{10} , conditional only on the values of observed exogenous variables gives rise to a likelihood function which has an interior maximum with probability one. If the marginal density of the initial values is misspecified, ML estimates are no longer consistent.

It is not, in fact, difficult to obtain the unconditional likelihood function once the marginal distribution of the initial values is specified. The problem is a correct specification of this distribution.

Suppose that the dynamic relationship to be estimated is stationary so that $|\gamma| < 1$. Consider equation (1-10) for y_{10} and the infinite past:

$$y_{i0} = \sum_{j=1}^{\infty} \gamma^j \beta x_{i-j} + \frac{1}{1-\gamma} \mu_i + v_{i0}, \text{ where } v_{it} = \gamma v_{it-1} + \varepsilon_{it} \quad (1-13)$$

(Recall that all variables are expressed as deviations of from their overall means.)

If $\beta = 0$, so that the relationship to be estimated is a pure autoregression, the vector of initial values $y_0 = (y_{10}, \dots, y_{N0})'$ has a joint normal distribution with means 0 and variance-covariance matrix

$$\left(\frac{\sigma_{\mu}^2}{(1-\gamma)^2} + \frac{\sigma_{\varepsilon}^2}{1-\gamma^2} \right) I_N. \text{ The unconditional likelihood is therefore}$$

$$\begin{aligned} & \log L(\gamma, \sigma_{\mu}^2, \sigma_{\varepsilon}^2 | y_{11}, \dots, y_{NT}; x_{11}, \dots, x_{NT}; y_{10}, \dots, y_{N0}) \\ &= -\frac{NT}{2} \log 2\pi - \frac{NT}{2} \log \sigma^2 - \frac{N}{2} \log \xi - \frac{N(T-1)}{2} \log \eta \\ & \quad - \frac{1}{2\sigma^2} \sum_{i=1}^N \sum_{t=1}^T (y_{it} - \gamma y_{it-1})^2 - \frac{N}{2} \log \left(\frac{\sigma_{\mu}^2}{(1-\gamma)^2} + \frac{\sigma_{\varepsilon}^2}{1-\gamma^2} \right). \quad (1-14) \end{aligned}$$

To maximize, express σ_{μ}^2 , σ_{ε}^2 , ξ and η in terms of ρ . For given ρ in the interval $[0,1)$, concentrate the likelihood function with respect to σ^2 and γ . This is a little more complicated than the usual minimization of the SS in the penultimate term because γ enters the final term as well. Then do a gradient search on ρ .

When $\beta \neq 0$, things are more complicated still. Various alternative specifications considered in the literature are reported and analyzed in Sevestre and Trognon, Chapter 7.¹ Considerable simplification, however, can

¹ One interesting possibility discussed by them is to choose y_{i0} a linear function of some observed individual-specific time-invariant exogenous variables and a disturbance which is decomposed as the sum of the individual-specific disturbances μ_i and a remainder. The first-order

Other Methodological Issues

Problems associated with measurement errors are more important than they might seem at first, because of the increasing importance of so-called "pseudo panel" data (Chapter 11) and the application of measurement error models to the analysis of such data as if they were true panel data. Measurement errors in panel data are treated extensively in Chapter 10. Griliches [1986] persuasively argues the need to understand and model the processes generating errors in economic data in the estimation of economic relations. Griliches and Hausman [1986] provide a pioneering application to panel data.

For many types of problems true panel data are not available, but rather several cross sections at different points in time are. For example, surveys of consumer expenditures based on a sample of individual households are made every few years in the UK or the US. Surveys to determine unemployment and labour force participation are made monthly on the basis of a rotating sample. Pseudo panel methods for treating such data are described in Chapter 14. These methods go

simply maximize $\log L(\beta, \gamma, \sigma_\eta^2, \sigma_\epsilon^2, \sigma_\phi^2, \gamma_{11}, \dots, \gamma_{NT}, x_{11}, \dots, x_{NT}, \gamma_{10}, \dots, \gamma_{N0})$ with respect to $\beta, \gamma, \sigma_\eta^2, \sigma_\epsilon^2$, and σ_ϕ^2 . While omitting estimation of λ and σ_λ^2 leads to some loss of efficiency, the ML estimates obtained in this way remain consistent as long as the random variables $\phi'' = \sum_{j=0}^f \gamma_j' \beta x_{j''-j}$ have well-defined variances and covariances, which they will if the x_{it} are generated by stationary processes. Besides, since the x_{it} are assumed to be exogenous, we really have no basis on which to model their determination and are likely to misspecify this part of the model. In this sense we ought to prefer this kind of "almost full-information" maximum likelihood.

back to Deaton [1985] who proposed dividing the sample into "cohorts" sharing common demographic, socio-economic, or historical characteristics, then treating the "cohort" averages as observations on "representative" individuals in a panel. Because each "cohort" observation is based on a sample of the true population cohort, the averages, treated as observations, contain sampling errors. Thus, Deaton proposed that the observations be considered as measurements of the "true" values with errors.

What should we make of this approach from the standpoint of the fundamental issues of history and dynamics? It goes without saying that we want to make use of whatever data is available in an appropriate way. The question is what do the cohort averages mean and how should relationships among them be interpreted? Deaton's cohorts and his proposed treatment of cohort averages is similar to the notion of a representative economic agent, introduced by Alfred Marshall in the last century, and in widespread theoretical use today. Kirman [1992] has recently given a detailed critique of the concept and many of his points apply in the present context. Essentially, relationships among averages, or for representative individuals, are often not interpretable directly in terms of individual behaviour since the relationships among the aggregates is often a result of the aggregation. Another way of saying the same thing is that the aggregate relationships are reduced from which the underlying structural relations (i.e., at the individual level) will not generally be identifiable. This is particularly the case when differences among individuals are historical to a significant degree and when the relationships of interest are dynamic. To the extent that the cohort-defining variables succeed in classifying individuals together who share common histories and exhibit common forms of (dynamic) behaviour, the use of pseudo panel data as if they were true panel data subject to sampling error will be successful.

But to the extent that unobserved heterogeneity in either respect remains, the relationships obtained from pseudo panel data may not permit identification of the underlying structure of interest.

Chapters 15-24 in Part II deal with latent variables and other forms of nonlinear models in a panel data context. Two points are worth making in this respect: first, it is frequently more difficult to see how elements of individual heterogeneity should be introduced, in contrast to the simple way in which such heterogeneity is introduced in equations (whether linear or not) in terms of disturbances. In Chapter 15, it is pointed out that much of the randomness cannot in general be interpreted in terms of omitted explanatory variables. Second, also pointed out, even in the case in which all the explanatory variables are truly exogenous, failure to take account of heterogeneity may result in bias, not merely inefficiency, in nonlinear cases, whereas no bias results in the linear case.

The solution in principle is to formulate a model in terms of the probability of individual observations and then to "integrate out" the heterogeneity factors if these can be parametrically specified. In practice, of course, this is rarely possible analytically and may even be extremely difficult computationally. Methods of simulated moments (see McFadden [1989]) are of considerable utility in this connection. An important application of latent variables models (which are largely highly nonlinear) is to selection bias and incompleteness in panel data (Chapter 13). In the case of selection bias, a rule other than simple random sampling determines how sampling from the underlying population takes place. Ignoring the nature of the selectivity mechanism may seriously distort the relationship obtained with respect to the true underlying structure. Heckman (e.g., [1990]), and references cited therein) has pioneered in this analysis. The greatest problem in

Applications of panel data are very diverse, depending, of course, on the availability of such data in specific substantive contexts. This volume contains chapters on labor demand (Chapter 25), labor supply (Chapter 28), individual labor market transitions (Chapter 29), consumption dynamics (Chapter 27), investment (Chapter 26 and 31), and

1.3 Applications of Panel Data Econometrics

In Part II there are also a number of key methodological chapters new to the second edition. These include chapters on Generalized Method of Moments (22) and on Simulation Estimation (23), and a long chapter on duration models (19), which supplements a chapter on point processes already included in the first edition.

Panel data in this connection is attrition (sometimes resolved through partial rotation which has its own problems). The probability of nonresponse increases when the same individual is repeatedly sampled. In Chapter 18, it is shown that the crucial question is whether the observed values in the sample can be considered as the result of a simple random drawing or whether, on the contrary, they are "selected" by some other rule, random or not. In the case of simple random selection, standard estimation and inference are appropriate and we say the selection rule is ignorable. On the other hand, if selection is nonrandom with respect to factors reflecting heterogeneity, that is correlated with them, standard techniques yield biased estimates and inferences. In this case the selection rule must be explicitly modelled to correct for selection biases. The authors of Chapter 18 show how this can be done for both fixed and random effects models. Because consistent estimation in the case of a non-ignorable selection rule is much more complicated than in the ignorable case, several tests are proposed to check whether the selection rule is ignorable.

dividend policy (Chapter 30), and on the estimation of production functions and measuring efficiency (Chapter 32). In addition to surveying important substantive areas of research these chapters are particularly useful in illustrating our message.

Obviously, panel data (or pseudo panel data) are essential if we want to estimate dynamic relationships at an individual or disaggregated level. As soon as the focus is on dynamics, historically generated heterogeneity becomes a central issue.

Models of factor demand (labor and capital-investment) reveal the crucial role of expectations. In this connection it is interesting to note the special impact of heterogeneity on expectations. Panel data provide a unique opportunity to study expectation formation and to test various hypotheses about expectation formation. (See for example, Nerlove [1983], Nerlove and Schuermann [1995].) Often, however, panel data do not contain direct observations on expectations but, as is typically the case with time series data, only on other variables affected by expectations. In this case, we formulate a model of expectation formation and infer indirectly the parameters of both the behavioural and the expectational model. To see how heterogeneity plays a critical role, it is useful to consider two simple examples: adaptive expectations and rational expectations.

Suppose that the model we wish to estimate is

$$y_t = \alpha x_t + u_t \quad t = 1, \dots, N; \quad t = 1, \dots, T, \quad (1-17)$$

where expectations are adaptive:

$$x_t^* = \beta x_{t-1}^* + (1-\beta)x_{t-1} + v_t. \quad (1-18)$$

Even if the disturbances in the behavioural equation (1-17) are i.i.d. random variables, it is unlikely that past history and past experience

Rational expectations imply

$$\Omega_{i,t-1} = \{y_{i,t-1}, \dots, x_{i,t-1}, \dots, z_{i,t-1}, \dots\}. \quad (1-23)$$

Then, for the i -th individual,

$$z_i = \sum_{n=1}^N y_n. \quad (1-22)$$

suppose

which individual decisions interact to produce aggregates. For example, on aggregates of individuals, and may include knowledge of the way in only contains that individual's own past history, but also observations at the time when his expectations are formed. In principle, $\Omega_{i,t-1}$ not where $\Omega_{i,t-1}$ is the set of information available to the i -th individual

$$x_n^* = E(x_n | \Omega_{i,t-1}), \quad (1-21)$$

expectations. In this case (1-18) is replaced by

Still more interesting things happen in the case of rational

values of these will no longer serve as instruments.

specific disturbances, μ_i , are correlated with past x_{it}^* 's, the lagged

disturbance is serially correlated. Moreover, if the individual

the correlation between $y_{i,t-1}$ and μ_i , but the third term of the

Not only do the usual difficulties, discussed above, arise because of

$$y_n'' = \beta y_{i,t-1} + \alpha(1 - \beta)x_{i,t-1} - \beta\mu_i' + \beta\varepsilon_{i,t-1}'' + u_n'' - \beta u_{i,t-1}''. \quad (1-20)$$

transformation of (1-17)-(1-18) then yields

past x_{it}^* 's and also, presumably, with past u_{it}^* 's and x_{it}^* 's. The usual

where the individual specific effects are likely to be correlated with

$$v_n'' = \mu_i + \varepsilon_n'', \quad (1-19)$$

no fully taken into account by $x_{i,t-1}^*$. Thus, write

will play no latent role in the determination of current expectations,

$$E(y_{it} | \Omega_{i,t-1})$$

can be solved for the N values

Equations (1-31) are N equations for each t , which, in principle,

$$i = 1, \dots, N.$$

$$(1-30) \quad E(y_{it} | \Omega_{i,t-1}) = \alpha \gamma \sum_{i=1}^N \{ \alpha \gamma E(y_{it} | \Omega_{i,t-1}) \} + \sum_{i=1}^N E(\theta_i | \Omega_{i,t-1}),$$

Hence, if $u_i = \theta_i + v_i$,

$$= \alpha \gamma \sum_{i=1}^N E(y_{it} | \Omega_{i,t-1}) + u_i.$$

$$(1-29) \quad = \alpha \gamma E \left(\sum_{i=1}^N y_{it} | \Omega_{i,t-1} \right) + u_i$$

$$y_{it} = \alpha \gamma E(z_i | \Omega_{i,t-1}) + u_i$$

the nature of the difficulties involved. Then

zero. Suppose it is. Such a simplification does not essentially affect

The last term on the right hand side of (1-28) will not generally be

$$(1-28) \quad E(x_{it} | \Omega_{i,t-1}) = \gamma E(z_i | \Omega_{i,t-1}) + E(\delta_i | \Omega_{i,t-1}).$$

then

$$(1-27) \quad x_{it} = \gamma z_i + \delta_i,$$

So, for example, if

$$(1-26) \quad E(x_{it} | \Omega_{i,t-1}) = E(f_i(z_i) | \Omega_{i,t-1}).$$

which may also be stochastic, then

$$(1-25) \quad x_{it} = f_i(z_i),$$

to that individual, of z_i :

Now if the value of x_{it} faced by each individual is a function, peculiar

$$(1-24) \quad = \alpha E(x_{it} | y_{i,t-1}, \dots, x_{i,t-1}, \dots) + u_i.$$

$$y_{it} = \alpha E(x_{it} | \Omega_{i,t-1}) + u_i$$

points:

In this introductory chapter we have tried to bring out the following

1.4 Conclusions

is included in a final chapter.

Computational issues and a review of currently available software

applications discussed in Part III have this problem.

information set $\Omega_{i,t-1}$ when expectations are formed, all of the

Unless future values of the exogenous variables are in the

expectations, one is in deep trouble in the case of panel data.

identifiable. The bottom line is that, if one believes in rational

parameters of g_i , a_1 , and a_2 and α are not generally separately

expectation conditional on $\Omega_{i,t-1}$. Finally, it can be seen that the

This effect is correlated with the element in $\Omega_{i,t-1}$ since it is an

time-varying, individual-nonspecific, effect in addition to θ_i and v_{it} .

It follows that the appropriate equation now contains a specific

$$y_{it} = \alpha a_1 \sum_{i=1}^N g_i(\Omega_{i,t-1}) + \alpha a_2 N \lambda_{i,t-1} + u_{it}. \quad (1-33)$$

So (1-17) becomes

$$x_{it} = a_1 \sum_{i=1}^N g_i(\Omega_{i,t-1}) + a_2 N \lambda_{i,t-1}. \quad (1-32)$$

Then we can replace the left hand side of (1-28) by

$$E(y_{it} | \Omega_{i,t-1}) = a_1 g_i(\Omega_{i,t-1}) + a_2 \lambda_{i,t-1}. \quad (1-31)$$

In general

$$\lambda_{i,t-1} = \sum_{i=1}^N E(\theta_i | \Omega_{i,t-1}).$$

expectations

in terms of the contents of $\Omega_{i,t-1}$ for all N individuals and the sum of

- (a) One of the main reasons for being interested in panel data is the unique possibility of uncovering disaggregated dynamic relationships using such data sets.
- (b) In a dynamic context, one of the primary reasons for heterogeneity among individuals is the different history which each has.
- (c) If the relevant "population" is, following Haavelmo, the space of possible decisions, different past histories take the form of individual specific random variables which are generally correlated with all of the variables taken as explanatory, not just the lagged values of the endogenous variable. The former therefore cannot be conditioned upon in the usual way.
- (d) Finally, although the adaptive expectations model does not introduce any new complications, rational expectations introduce a time specific, individual non-specific, component in the error component formulation, as well as a fundamental failure of identifiability.
- Panel data econometrics is one of the most exciting fields of inquiry in econometrics today. Many interesting and important problems remain to be solved, general as well as specific to particular applications. This volume represents the definitive work with which to begin.
- This paper is a draft of the introductory Chapter to the second edition of L. Mátyás and P. Sevestre, eds., *The Econometrics of Panel Data*. Boston: Kluwer Academic Publishers, forthcoming 1995.

References

- Anderson, T.W. and C. Hsiao [1982]: Formulation and Estimation of Dynamic Models Using Panel Data, *Journal of Econometrics*, 18, 578-606.
- Balestra, P., and M. Nerlove [1966]: Pooling Cross-Section and Time-Series Data in the Estimation of a Dynamic Economic Model: The Demand for Natural Gas, *Econometrica*, 34, 585-612.
- Baltagi, B.H., and B. Raj [1992]: A survey of Recent Theoretical Developments in the Econometrics of Panel Data, *Empirical Economics*, 17, 85-109.
- Bhargava, A., and D. Sargan [1983]: "Estimating Dynamic Random-Effects Models from Panel Data Covering Short Time Periods," *Econometrica*, 51, 1635-1659.
- Carraro, C., F. Peracchi and G. Weber, eds. [1993]: *The Econometrics of Panels and Pseudo Panels*, Annals 1993-94, Supplement to the *Journal of Econometrics*, 59.
- Chamberlain, G., [1984]: "Panel Data," pp. 1247-1318 in Z. Griliches and M. Intriligator, eds. *Handbook of Econometrics*, Vol. 2. New York: Elsevier.
- Deaton, A., [1985]: "Panel Data from a Time Series of Cross-Sections," *Journal of Econometrics*, 30, 109-126.
- Griliches, Z., [1986]: "Economic Data Issues." Pp. 1465-1514 in Z. Griliches and M. Intriligator, eds. *Handbook of Econometrics*, Vol. 3. New York: Elsevier.
- Griliches, Z., and J.A. Hausman [1986]: Errors in Variables in Panel Data, *Journal of Econometrics*, 31, 93-118.
- Haavelmo, T., [1944]: The Probability Approach in Econometrics: Supplement to *Econometrica*, 12.

- Heckman, J.J., [1990]: Varieties of Selection Bias, *American Economic Review, Papers and Proceedings*, 80, 313-318.
- Kirman, A., [1992]: Whom or what does the Representative Individual Represent? *Journal of Economic Perspectives*, 6, 117-136.
- Knight, F.H., [1921]: *Risk, Uncertainty and Profit*. Boston: Houghton Mifflin Co.
- Maddala, G. S., [1971]: "The Use of Variance Components Models in Pooling Cross-Section and Time Series Data," *Econometrica*, 39:341-358.
- Marschak, J., [1950]: Statistical Inference in Economics: An Introduction. Pp. 1-50 in *Statistical Inference in Dynamic Economic Models*, ed. T.C. Koopmans. New York: John Wiley.
- Marschak, J., [1953]: *Economic Measurements for Policy and Prediction*. Pp. 1-26 in *Studies in Econometric Method*. ed. W.C. Hood and T.P. Koopmans. New York: John Wiley.
- Matyas, L., and S. Rahman [1992]: An Alternative Approach for the Estimation of Distributed Lag Models in Panel Data, Working Paper No. 4/92, Department of Econometrics, Monash University.
- McFadden, D.L., [1989]: A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration, *Econometrica*, 57, 995-1026.
- Mundlak, Y., [1961]: Empirical Production Functions Free of Management Bias, *Journal of Farm Economics*, 43, 44-56.
- Nerlove, M., [1971]: Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections, *Econometrica*, 39, 359-382.
- Nerlove, M., [1983]: Expectations, Plans and Realizations in Theory and Practice, *Econometrica*, 51, 1251-1280.

Nerlove, M., and T. Schuermann [1995]: "Expectations: Are They

Rational, Adaptive, or Naïve? An Essay in Simulation-Based

Inference." Pp. 354-381 in G.S. Maddala, P.C.B. Phillips, and T.

N. Srinivasan eds., *Advances in Econometrics and Quantitative*

Economics. Oxford: Basil Blackwell.

Nickell, S., [1981]: Biases in Dynamic Models with Fixed Effects,

Econometrica, 49, 1417-1426.

Sevestre, P., and A. Trognon [1985]: A Note on Autoregressive Error-

Components Models, *Journal of Econometrics*, 29, 231-245.

Trognon, A., [1978]: Miscellaneous Asymptotic Properties of OLS and ML

Estimators in Dynamic Error Components Models, *Annales de l'INSEE*,

30-31, 631-658.

d:\wordfile\papers95\panelrev\chaplrev.doc REVISION 6/26/95 M. Nerlove