THE CURRENT STATE OF BUSINESS DISCIPLINES
VOL. , PP.

NEW DEVELOPMENTS IN THE ANALYSIS OF PANEL DATA SETS [1]

S.G.HALL
The Management School,
Imperial College, London
E-mail:

G.URGA
Department of Investment, Risk Management and Insurance
City University Business School
Frobisher Crescent, London EC2Y 8HB (UK)
E-mail: g.urga@city.ac.uk

ABSTRACT

The purpose of this paper is to present a selective introduction of the traditional and most recent developments of econometrics of panel data. We consider first the case of stationary variables. Thus we discuss the static models, introducing the crucial distinction between fixed and random effects, and comparing estimation in levels with the within-groups estimator. Then we present the dynamic models, introducing the Anderson and Hsiao (1981) estimators, and the more efficient generalised method of moments (GMM) recently proposed by Arellano and Bond (1991). We also present a useful transformation (orthogonal deviations) and some specification tests. Further, we explore the basic consequences of non-stationarity for panel data and in particular consider the main results in Pesaran and Smith (1995) and we examine a counter example to the Pesaran and Smith case due to Hall and Urga (1999) which provides a new condition for valid aggregation. We finally consider the recent literature on panel data unit roots tests and cointegration, and we provide some guidelines for empirical research.

Keywords: Panel Data, Cointegration, Aggregation, Common Stochastic Trend.

1. Introduction

Aggregate time series econometrics has been revolutionised by the introduction of non-stationary time series analysis and cointegration since the mid-1980s. But this revolution has left panel data estimation techniques largely unaffected. This is surprising since, by definition, a panel data set consists of a standard time series data on a (possibly large) number of economic agents, firms, households, etc. So all the problems which exist in aggregate estimation with non-stationary variables must be equally present for a complete panel. We would suggest that the only reason that little attention has been paid to these issues is that panel data estimation is generally a complex and difficult task simply because of the size of the data set which must be manipulated and researchers have been reluctant to face these complex conceptual issues at the same time as dealing with this data complexity.

This is, of course, not a defensible position. Recently a small number of papers (Robertson and Symons (1992), Pesaran and Smith (1995) and Hall and Urga (1996)) have been appearing which are beginning to face up to these problems and as a result we have begun to learn that non-stationarity is at least as serious a problem for panel data sets as it is for aggregate data. Indeed we are learning that if anything the implications are even more important.

The purpose of this paper is to present a selective introduction of the econometrics of panel data and then to discuss some of the recent papers which have considered the implications of non-Stationarity. The main aim is to provide an accessible guide to the most widely used methods in practice; therefore, we consider only methods for the single equation linear regression model, covering in detail only standard techniques for static models and some recently developed GMM estimators for dynamic models with limited serial correlation[2].

The growing interest in panel data modelling, especially over the '80s, reflects the increasing availability of such data and advances in computing technology (powerful and easy to use). It also reflects the growing interest in estimating models of individual behaviour over time without having to use aggregate time series data.

Panel data provide repeated observations on the cross-section of individuals[3]. Econometric estimates therefore utilise both time series and cross-section variation in the data.

The typical panel data set consists of a large number of cross-section units observed at a few points in time. In general, for a variable y_{it} we have i=1,...,N individuals observed for t=1,..,T time periods. In a micropanel of households, firms or banks, T can be as small as 3 or 4 whilst N may refer to hundreds or thousands of individual units. It is natural to assume in this case that the series follow stationary processes. In the case when N and T are both large, the stationarity hypothesis needs to be tested and then the problem of aggregation over micro units needs to be addressed. Even in the case in which T is large it may often be informative to use panel data estimation, but it is import to test the homogeneity restrictions across the coefficients of the N time series. If they are accepted by the data, this gives rise to more efficient estimates (Zellner, 1969).

It is always the case that panel data analysis allows one to identify effects of economic interest that would not be identified in a single cross-section, an obvious example of this is the dynamics of an economic relationship. However these effects can be seriously biased by the use of some panel data estimators, for example Robertson and Symons (1992) examine the consequences of imposing equality restrictions on parameters in the presence of dynamics and non-stationarity and find the possibility of very serious biases. Pesaran and Smith (1995) and Hall and Urga (1996) follow up and extend this analysis.

The chapter is organised as follows. The case of stationary variables is introduced in the first 4 sections. In section 2 we discuss briefly static models, introducing the distinction between fixed and random effects, and comparing estimation in levels with the within-groups estimator. In section 3 we present dynamic models, starting from the bias of the within-groups estimator in this case, and then introducing the Anderson and Hsiao (1981 and 1982) estimators and the more efficient generalised method of moments estimator (GMM) recently proposed by Holtz-Eakin, Newey and Rosen (1988) and Arellano and Bond (1991). Furthermore, we report the computationally useful orthogonal deviations transformation proposed by Arellano (1988) and Arellano and Bover (1990). In section 4 we present some specification tests. In section 5 we explore the basic consequences of non-Stationarity for Panel data and in particular consider the main results of Pesaran and Smith (1995) (Section 5.1) and in Section 5.2 we examine a counter example to the Pesaran and Smith case due to Hall and Urga (1996) which provides a new condition for valid aggregation and finally examine some new procedures in testing common stochastic trends . In section 6 we consider the recent literature on panel data unit root tests. Finally in section 7 we draw some conclusions and provide some guidelines for empirical research.

2. Static models[4]

Let us consider the simple linear model

$$y_{it} = x'_{i,t}\,\beta + \eta_i + v_{it} \quad i = 1,\dots, N \quad t = 1,\dots, T \tag{1}$$

where x_{it} is a (kx1) vector of time-varying regressors assumed to be strictly exogenous (i.e. uncorrelated with past, present and future realisation of v_{it}), and η_i is fixed or random individual effects. v_{it} is the error term independently and identically distributed over *I* and *t* with zero mean and variance σ^2_ν . We omit time dummies for simplicity although they may capture unobserved aggregate effects.

In the case where observations on y_{it} and x_{it} for I=1,....,N and t=1,....,T are available, an aggregate time series regression would treat η_i as part of the constant and thus unidentified, whilst a cross-section regression will yield a biased estimator of β if η_i is correlated with x_{it} across I.

In the case of the presence of fixed effects (see below), β and η_i can be estimated consistently and efficiently by the following estimators which can be obtained by OLS after the data are transformed by subtracting group means from each observation (see Hsiao, 1985 and 1986)

$$\hat{\beta}_{wg} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)'\right]^{-1}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i)\right] \tag{2}$$

where $\bar{x}_i = \sum_{t=1}^{T}(x_{it}/T)$ and $\bar{y}_i = \sum_{t=1}^{T}(y_{it}/T)$ [5]

Because η_i is treated as a fixed constant, the estimator of β is called least squares dummy variables. This estimator is known also as the "within-groups estimator" (WG),

$$\hat{\eta}_i = \bar{y}_i - \hat{\beta}_{wg}\,\bar{x}_i \quad i = 1, \dots, N$$

or “covariance estimator” (CV), because it can be obtained using an appropriate transformation in the form of the orthogonal projection

$$Q = I_{NT} - P \tag{3}$$

where

$$P = \left[I_N - T^{-1}\, i_T\, i_T' \right] \tag{4}$$

with I as the identity matrix and i a column vector of ones, $\otimes$ denotes the Kronecker product.

If we re-write (1) in the form

$$y = X\beta + \eta + v \qquad (5)$$

where y, X, η and v denote the (TNx1), (TNxk), (TNx1) and (TNx1) matrices respectively, we can pre-multiply (5) by Q yielding

$$Qy = QX\beta + Qv \qquad (6)$$

since $Q\eta = 0$. Given that x_{it} is uncorrelated with v_{is} for each s,t, then in the transformed specification QX and Qv are uncorrelated and thus we can apply OLS giving rise to the estimator [6]

$$\hat{\beta}_{wg} = (X'QX)^{-1} X'Qy \qquad (7)$$

which coincides with (2) above. This can be re-written as

$$\hat{\beta}_{wg} = (\tilde{x}'\tilde{x})^{-1} \tilde{x}'\tilde{y} \qquad (8)$$

where $\tilde{x}_{it} = x_{it} - \bar{x}_i$ and likewise for the other variables. The observations for individual i are transformed to deviations from their mean value over time.

It is worth noting that we can estimate $\hat{\beta}_{wg}$ without calculating $\hat{\eta}$. This procedure is particularly useful when the number of individuals is large. With respect to the statistical properties, $\hat{\beta}_{wg}$ is consistent for $T \to \infty$ and its consistency does not depend on η_i. $\hat{\beta}_{wg}$ is consistent for T fixed and $N \to \infty$ if and only if $E(\tilde{x}_{it}\tilde{v}_{it}) = 0$, which occurs only if $E(x_{it} v_{is}) = 0$ for each t,s=1,..,T.

In the random effects case η_i is assumed to be a random variable such that

$$u_{it} = \eta_i + v_{it} \qquad (9)$$

where we suppose that $E(\eta_i)=E(v_{it})=0$ and $E(x_{it}\,\eta)=E(x_{it}\,v_{is})=0$. Under these conditions the OLS regression of y_{it} on x_{it} (in levels) gives a consistent estimate of β.

However, the specification (9) generates autocorrelation and heteroskedasticity in u_{it} even if it is absent in v_{it}. Note that with $E(v_{it}\,v_{is})=0$ we have $E(u_{it}\,u_{i,t-s})=\mathrm{Var}(\eta_i)$ for all $s \neq t$. This suggests that we seek a more efficient estimator than OLS. The appropriate GLS estimator of β, under these assumptions turns out to be a weighted average of the above within-groups estimator and the so-called "between-groups" estimator

$$\hat{\beta}_{bg} = (X'PX)^{-1} X'Py \qquad (10)$$

that is

$$\hat{\beta}_{GLS} = D\,\hat{\beta}_{bg} + (I_k - D)\,\hat{\beta}_{wg} \qquad (11)$$

where $D = (V_{bg} + V_{wg})^{-1} V_{wg}$, V_{bg} and V_{wg} are the covariance matrices of $\hat{\beta}_{bg}$ and $\hat{\beta}_{wg}$ respectively, and where the estimator $\hat{\beta}_{bg}$ can be obtained by applying OLS to the transformed model (see Maddala, 1971)

$$PY = PX\,b + Ph + Pu. \qquad (12)$$

This GLS estimator is often known in the literature as the Balestra and Nerlove (1966) estimator,

$\hat{b}_{BN}$ (with $\hat{b}_{BN} = \hat{b}_{GLS}$)[7].

Hausman and Taylor (1981) show that if $h_i \sim iid\ (0, s_h^2)$ and $v_{it} \sim iid\ (0\ ,\ s_v^2)$ a GLS transformation of (1) is given by

$$y_{it} - (1-J)\,\bar{y}_i = b\left[x_{it} - (1-J)\,\bar{x}_i\right] + Jh_i + \left[v_{it} - (1-J)\,\bar{v}_i\right] \qquad (13)$$

where $J = \left[\, s_v^2 \ /\ [\ s_v^2 + T\ s_h^2\,]\right]^{1/2}$, $\bar{y}_i$ and $\bar{x}_i$ are once again the time means of the variables. Estimates of s_v^2 and s_h^2 can be obtained as follows (see Hausman and Taylor, 1981)

$$\hat{s}_v^2 = \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\tilde{y}_{it} - \hat{b}_{wg}\,\tilde{x}_{it}\right]^2 \qquad (14)$$

and

$$\hat{s}_h^2 = N^{-1} \sum_{i=1}^{N} \left[\bar{y}_i - \hat{b}_{wg}\,\bar{x}_i\right]^2 - T^{-1}\,\hat{s}_v^2 \qquad (15)$$

The feasible GLS estimator can then be obtained by applying OLS to (13), having (14) and (15), to construct an estimate of J.

It is worth clarifying the following point. When η_i is treated as a fixed constant, the model is referred to as a fixed effects model, whilst when η_i is treated as a random variable, it is called a random effects model. More generally, we can unify these two formulations, and we may assume from the outset that the effects are always random. What is crucial, however, is to investigate if η_i is correlated or not with the observed variables x_{it}. If η_i is correlated with x_{it}, the fixed effects model is viewed as one where investigators make inferences conditional on the effects that are in the sample. Whilst if η_i is not correlated with x_{it} the random effects model is viewed as one where investigators make unconditional or marginal inferences with respect to the population of all effects. "It is up to the investigator to decide whether he wants to make inference with respect to the population characteristics or only with respect to the effects that are in the sample" (Hsiao, 1985, p. 131).

Therefore, in applied work the first step is to compare the estimates in levels and deviations. Significant differences between the two indicate that correlated individual effects are omitted from the regression in levels. It is worth noting that this is equivalent to testing whether the effects are correlated or not with x_{it}.

To perform this experiment, we can use the traditional Hausman test[8] (1978) based on the comparison between the within-groups estimator and the Balestra and

$$h = \left[\hat{b}_{GLS} - \hat{b}_{wg}\right]' \left[var\,(\hat{b}_{wg}) - var\,(\hat{b}_{GLS})\right]^{-1} \left[\hat{b}_{GLS} - \hat{b}_{wg}\right] \qquad (16)$$

Nerlove estimator under the null that the effects are not correlated with the regressors, the $\text{plim}(\hat{\beta}_{GLS} - \hat{\beta}_{wg}) = 0$ and h is distributed as $\chi^2_{(k)}$ with k degrees of freedom[9].

This procedure is valid in the case in which $\hat{\beta}_{GLS}$ is efficient relative to $\hat{\beta}$ under the null. In the presence of heteroskedasticity and autocorrelation, this condition is not satisfied.

Alternatively, Arellano and Bover (1989) and Arellano (1992) propose to form a system by combining the equations in levels and in deviations. For instance[10], for T=2 we have

$$\begin{bmatrix} \tilde{y}_{i1} \\ \tilde{y}_{i2} \\ \bar{y}_{i2} \end{bmatrix} = \begin{bmatrix} \tilde{x}_{i1} \\ \tilde{x}_{i2} \\ \bar{x}_{i2} \end{bmatrix} \beta + \begin{bmatrix} 0 \\ 0 \\ \bar{x}_{i2} \end{bmatrix} \alpha + \begin{bmatrix} \tilde{v}_{i1} \\ \tilde{v}_{i2} \\ \bar{v}_{i3} \end{bmatrix} \quad i = 1, \ldots, N.$$

β and α can be estimated using OLS and robust estimates of their variances may be obtained. A Wald statistic or an equivalent t-test of the null $H_0 : \alpha = 0$, (or an $\chi^2_{(k)}$ or F-test depending on whether α is a scalar or a vector), tests the null hypothesis of no correlation.

This test is robust to heteroskedasticy and serial correlation. For further details about the various versions of this Hausman type test are in Arellano (1993) and Arellano and Bover (1995). Finally, Baltagi (1996) presents this test plus other specific forms of misspecification tests for the error component models.

3. Dynamic models

Dynamic specifications are of particular interest in modelling panel data, part of the richness of a panel data set is precisely due to the fact that we can analyse the process of dynamic adjustment which is impossible in a cross section data set. The simplest model is a first-order autoregression of the form

$$y_{it} = \alpha\, y_{i,t-1} + \eta_i + v_{it} \qquad (17)$$

i=1,...,N, t=1,....,T and y_{i0} observed. Note that $y_{i,t-1}$ and η_i are necessarily correlated.

What are the methods of eliminating the permanent effects η_i? First, we can use the within-groups estimator, $\hat{\alpha}_{wg}$, that is OLS on the transformed equation

$$(y_{it} - \bar{y}_i) = \alpha\,(y_{i,t-1} - \bar{y}_{i(-1)}) + (v_{it} - \bar{v}_i) \qquad (18)$$

where $\bar{y}_i = T^{-1} \sum_{t=1}^{T} y_{it}$, $\bar{y}_{i(-1)} = T^{-1} \sum_{t=0}^{(T-1)} y_{it}$

and $\bar{v}_i = T^{-1} \sum_{t=1}^{T} v_{it}$.

Because $\left(y_{i,t-1} - \bar{y}_{i,(-1)}\right)$ and $(v_{it} - \bar{v}_i)$ are correlated through their mean components, then $\hat{\alpha}_{wg}$ will be seriously biased even for large N when T is small. Nickell (1981) provides the asymptotic bias

$$\operatorname*{plim}_{N\to\infty} (\hat{\alpha}_{wg} - \alpha) = - \left[\frac{(1+\alpha)\, h(\alpha\ T)}{T-1}\right] \left[1 - \frac{2\alpha h(\alpha, T)}{(T-1)(1-\alpha)}\right]^{-1} \quad (19)$$

where $h(\alpha, T) = 1 - T^{-1}\dfrac{(1-\alpha^T)}{1-\alpha}$.

Under the hypotheses that η_i is a fixed constant, the model is stationary and v_{it} is white noise. Nickell shows that the bias goes to zero when $T\to\infty$, whilst for positive values of α the bias is negative and does not tend to zero for $\alpha\to 0$, nor when var $(\eta_i)\to 0$.

Alternatively, the unobservable effects η_i can be eliminated by taking the first differences, e.g.

$$(y_{it} - y_{i,t-1}) = \alpha\ (y_{i,t-1} - y_{i,t-2}) + (v_{i,t} - v_{i,t-1}). \quad (20)$$

The OLS estimator of (20), $\hat{\alpha}_{\Delta}$ is also biased, but the bias does not vanish as $T\to\infty$ (Arellano and Bover, 1989)

$$\operatorname*{plim}_{N\to\infty} (\hat{\alpha}_{\Delta} - \alpha) = -\frac{1+\alpha}{2}. \quad (21)$$

This coincides with the bias of the WG estimator when T=2; however, GLS in the first difference equation will yield $\hat{\alpha}_{wg}$. Finally, the asymptotic bias of OLS in the levels, $\hat{\alpha}_1$, take the form

$$\operatorname*{plim}_{N\to\infty} (\hat{\alpha}_1 - \alpha) = \frac{\lambda}{\lambda\,(1-\alpha)^{-1} + (1+\alpha)^{-1}} \quad (22)$$

where $\lambda = \sigma_\eta^2 / \sigma_v^2$

with $\sigma_\eta^2 = \lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} \eta_i^2$ and $v_{it} \sim iid(0, \sigma_v^2)$.

Note that the biases of $\hat{\alpha}_{wg}$ and $\hat{\alpha}_{\Delta}$ do not depend on the η_i's and are negative. On the contrary, the bias of $\hat{\alpha}_1$ depends on σ_η^2 and is positive for a>0.

Anderson and Hsiao (1981) proposed a consistent estimator of α which is an instrumental variables estimator for the first differenced equations. For T=3, and supposing that hereafter y_{i0} is not observed, we can have

$$(y_{i3} - y_{i2}) = \alpha\,(y_{i2} - y_{i1}) + (v_{i3} - v_{i2}) \quad (23)$$

where y_{i1} is the instrument for $(y_{i2} - y_{i1})$. y_{i1}, in fact, is correlated with $(y_{i3} - y_{i2})$ but is not correlated with $(v_{i3} - v_{i2})$ provided v_{it} is serially uncorrelated. The same applies for T>3, giving the estimator

$$\hat{a}_{AH} = \frac{\sum_{i=1}^{N} \sum_{t=3}^{T} y_{i,t-2}\,(y_{it} - y_{i,t-1})}{\sum_{i=1}^{N} \sum_{t=3}^{T} y_{i,t-2}\,(y_{i,t_1} - y_{i,t-2})} . \tag{24}$$

An alternative estimator proposed by Anderson and Hsiao is to use $\Delta\ y_{i,t-2}$, rather than $y_{i,t-2}$, as instrument, and consequently the summation over time periods goes from 4 to T in (24).

$\hat{a}_{AH}$ is consistent for T fixed and N $\rightarrow\infty$, for T $\rightarrow\infty$ and N fixed, and when both T and N tend to ∞.

Special to the case when T is small

Arellano and Bond (1991) (AB hereafter) provide a more efficient estimator when T is small and N is large. From the following table it is clear that y_{i1} is the only

Table A.

Equation	**Valid Instruments**
$\Delta y_{i3} = \alpha \Delta y_{i2} + \Delta v_{i3}$	y_{i1}
$\Delta y_{i4} = \alpha \Delta y_{i3} + \Delta v_{i4}$	y_{i1}, y_{i2}
$\Delta y_{i,T} = \alpha \Delta y_{i(T-1)} + \Delta v_{i,T}$	$y_{i1}, y_{i2} \ldots, y_{i(T-2)}$

valid instrument available in the equation for the first differences for period 3. For the second, we have two instruments, and so on, until the last equation when we have (T-2) instruments.

If we define

$$\bar{v}_i = (\Delta v_{i3}, \ldots, \Delta v_{iT})\ as\ a\ [(T-2)\ x\ 1] \quad \text{vector}^{11} \tag{25}$$

and

$$z_i = \begin{bmatrix} y_{i1} & 0 & 0 \\ 0 & y_{i1}, y_{i2} & 0 \\ . & . & . \\ 0 & 0 & y_{i1}, \ldots, y_{iT-2} \end{bmatrix}$$

a ((T - 2) x m) matrix with $m = \left[\frac{(T-1)(T-2)}{2}\right]$, then the moment restrictions can be expressed by $E\ (Z_i' - \bar{v}_i\ ) = 0$.

If the errors v_{it} are independently distributed with constant variance s^2, we have $E\ (\bar{v}_i\ \bar{v}_i') = s^2\ H$, where H is a symmetric matrix (T-2) x (T-2) which has twos in the main diagonal, minus ones in the first sub-diagonal and zeros otherwise. Then AB show that the variance matrix of $z_i' \bar{v}_i$ is

$$E\ (z_i' \bar{v}_i\ \bar{v}_i'\ z_i) = s^2\ E\ (z_i'\ Hz_i) \tag{26}$$

and the estimator of α is the following

$$\hat{a}_{AB1} = \frac{\Sigma_i\ \bar{y}_{i(-1)}'\ z_i\ (\Sigma_i\ z_i'\ H\ z_i)^{-1}\ \Sigma_i\ z_i'\ \bar{y}_i}{\Sigma_i\ \bar{y}_i'\ z_i\ (\Sigma_i\ z_i'\ H\ z_i)^{-1}\ \Sigma_i\ z_i'\ \bar{y}_{i(-1)}} \tag{27}$$

where $\bar{y}_i = (\Delta\ y_{i3}, ..., \Delta\ y_{iT})'$ and $\bar{y}_{i(-1)} = (\Delta y_{i2}, \ldots, \Delta y_{i,T-1})'$.

This estimator has been proposed by AB for dynamic models from panel data[12]. It is called the generalised method of moments (GMM), and it minimises the discrepancy between the sample moments $\Sigma_i\ z_i'\ \bar{v}_i\ /\ N$ and their values in probability (Hansen, 1982).

$\hat{a}_{AB1}$ represents the one-step estimator which is consistent provided v_{it} is serially uncorrelated. For the case where the v_{is} are heteroskedastic, however, we can obtain a more efficient two-step estimator $\hat{a}_{AB2}$, if, instead of $\Sigma_i\ z_i'\ Hz_i$, in (27) we use a more general estimate of the variance of $z_i'\ \bar{v}_i$, that is $\Sigma_i\ z_i' \hat{v}_i \hat{v}_i'\ z_i$ where $\hat{v}_i$ are the residuals from the preliminary consistent estimator $\hat{a}_{AB1}$.

Finally, Arellano and Bover (1990) present a class of valid transformations alternative to the first differences in the context of models with predetermined variables. Among these is the computationally convenient orthogonal deviations transformation (see below for more detail). In addition, the authors show the invariance of the optimal estimators to the choice of transformation.

Ahn and Schmidt (1995) show that by exploiting some additional moments restrictions with non-linear GMM estimator leads to substantial improvements in the efficiency of estimation, in particular when the model contains exogenous variables in addition to the lagged dependent variable, as introduced in the next section.[13]

Models with exogenous regressors

Let us now consider a model which includes independent explanatory variables

$$y_{it} = a\ \ y_{i,t-1} + b\ \ x_{it} + h_i + v_{it}\ . \tag{28}$$

The previous method of assigning different instruments to different equations always applies. Building on this, we must now make some hypothesis concerning the correlation between x_{it} and v_{it} .

With T>3, for instance, we can consider an equation in first differences which eliminate the permanent effects

$$y_{i3} - y_{i2} = \boldsymbol{a} (y_{i2} - y_{i1}) + \boldsymbol{b} (x_{i3} - x_{i2}) + (v_{i3} - v_{i2}) \qquad (29)$$

Again, in this case, an OLS or a GLS (e.g. WG) regression does not consistently estimate α and β because Δ_{yi2} and Δ_{vi3} are negatively correlated (through the correlation of y_{i2} and v_{i2}). However, the model can be consistently estimated by IV methods using (x_{i1}, x_{i2} and x_{i3}) as instruments for (Δ_{yi2} Δx_{yi3}) if x_{it} is strictly exogenous, i.e. uncorrelated with all transitory errors.

If this is the case, it makes no difference to the validity of the instruments (and therefore to the consistency of α and β) whether the v_{it} are serially correlated or not. If, moreover, x_{it} is uncorrelated with the η_i 's, we can use the same instruments for the equations in levels [14].

More generally, and following AB, let us suppose that the x_{it} is correlated with η_i, and no external instruments are available. If v_{it} is not serially correlated, and x_{it} is predetermined (e.g. $E(x_{it}v_{is}) = 0$ if and only if s>t), the GMM estimator of α and β is similar to the autoregressive specification. The valid instruments for period t for the equation in first differences or orthogonal deviations are given by

$$z_{it} = [y_{i1} , \ldots\ldots\ldots, y_{i,t-2} , x_{i1} \ldots\ldots\ldots x_{i,t-1}) ;$$

that is the optimal z_i matrix is

$$z_i = \begin{vmatrix} y_{i1} \; x_{i1} \; x_{i2} \; 0 \; 0 \; 0 \ldots\ldots\ldots 0 \ldots\ldots 0 \;\; 0 \ldots\ldots\ldots 0 \\ 0 \;\; 0 \;\; 0 \; y_{i1} \; y_{i2} \; x_{i1} \; x_{i2} \; x_{i3} \;\; 0 \ldots\ldots 0 \;\; 0 \ldots\ldots\ldots 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ 0 \;\; 0 \;\; 0 \ldots\ldots\ldots\ldots\ldots\ldots \; y_{i1} \ldots\ldots y_{i,T-2} \; x_{i1} \ldots\ldots\ldots x_{i,T-1} \end{vmatrix}$$

If, instead, x_{it} is correlated with η_i and is strictly exogenous ($E(x_{it} v_{is})=0$ for all t and s) then all x's are valid instruments for all equations in first differences. In this case the valid instruments are

$$z_i = diag \;\; [y_{i1} \; , \ldots, \; y_{is} \; , x_{i1} \; , \ldots, \; x_{iT}] \quad \text{for } s = 1 \; , \ldots, T - 2$$

where diag[] represents a block diagonal matrix, of the type presented above.

Moreover, AB discuss other cases such as that in which x_{it} can be partitioned into (x_{1it} x_{2it}) where x_{1it} only is uncorrelated with η_i. This implies additional moment restrictions that can be exploited by stacking equations in levels and equations in first differences (Arellano and Bover, 1995). Moreover, other instruments have been suggested by Keane and Runkle (1992) and Schmidt, Ahn and Wyhoski (1992); they receive an excellent treatment in Baltagi (1995).

AB, finally, report some Monte Carlo experiments which show "negligible finite sample biases in the GMM estimators and substantially smaller variances than those associated with simpler IV estimators of the kind introduced by Anderson and Hsiao (1981)". They "also find that the distributions of the serial-correlation tests (see below) are well-approximated by their asymptotic counterpart". Kiviet (1995) finally presents interesting findings regarding bias, inconsistency and efficiency of alternative estimators utilised in dynamic models of panel data.

Special to the case when T is large

In the case in which the data set presents large T, the application of the "covariance" or "within-groups" estimator is straightforward (see Urga, 1991, for instance).

Following Hsiao (1985, 1986) and on the basis of Nickell's (1981) results, let us consider the following lagged-dependent variable model with exogenous regressors (omitting the time effects for simplicity):

$$y_{it} = \boldsymbol{a} \, y_{i,t-1} + \boldsymbol{b}' \, x_{i,t} + \boldsymbol{h}_i + v_{it} \qquad (30)$$

If η_i is treated as a fixed constant, the covariance estimator for α and β is biased if T is fixed. The bias occurs because of the particular way we have eliminated the unknown individual effects η_i from each observation. This creates the correlation of order 1/T between $(y_{i,t-1} - \bar{y}_{i.,-1})$ and the residuals in the transformed model

$$(y_{i,t} - \bar{y}_{i.}) = a \, (y_{i,t-1} - \bar{y}_{i,-1}) + b' (x_{i,t} - \bar{x}_{i.}) + (v_{i,t} - \bar{v}_{i.}) \qquad (31)$$

where $\bar{y}_{i.} = 1/T \sum_{t=1}^{T} y_{i,t}$ and likewise for the other dotted variables.

When T is very large the right-hand side variables become asymptotically uncorrelated and the CV or WG estimator is consistent. For small T, instead, the bias for α is negative and the bias for β is positive if $\alpha > 0$ (Anderson-Hsiao, 1981, Nickell, 1981, and Nerlove,1971). When η_i is treated as random, the error terms $u_{i,t} = \eta_i + v_{it}$ are serially correlated and correlated with $y_{i,t-1}$. Hence the OLS estimator is biased. Thus, the within-groups transformation has this significant disadvantage. If the explanatory variables are correlated with the past shocks, but not with the present and the future ones, one normally has available lags of these variables as instruments. However, this is not possible after one performs the within-groups transformation, because these lags will now be correlated with the transformed error term. The orthogonal deviations (OD) transformation, proposed by Arellano (1988)[15], gives an equivalent within-groups estimator, whereby we remove the individual effects by subtracting the mean of all future values of the variable, i.e.

$$X_{i,t} = \left[x_{i,t} - (x_{i,t+1} + \ldots\ldots + x_{i,T}) \, / \, T - t \right] \left[\frac{T - t}{T - t - 1} \right]^{1/2} \qquad (32)$$

for t = 1, ..., T - 1.

It is clear that the purpose of using the orthogonal deviations transformation instead of the within-groups transformation is that the alternative transformation of the model eliminates the individual effects while preserving the orthogonality among the transformed errors[16]. That is if the original random errors are independently and identically distributed (iid), then the transformed errors are also i.i.d. From this we can obtain optimal generalised method of moments (GMM) or standard instrumental variables (IV) estimators of the kind proposed by Anderson and Hsiao for models with predetermined variables[17].

4. Specification tests[18]

The consistency of the estimators discussed so far relies on the assumption of the lack of serial correlation of v_{it}. AB provide tests based on the standardised residual autocovariances which are asymptotically N(O,1) variables under the null of no autocorrelation. More specifically, AB report tests for the lack of the first-order (m_1) and second-order (m_2) serial correlation in the residuals. If v_{it} are serially uncorrelated, then the first differences transformation induces first-order serial correlation, but not second-order.

On the other hand, the orthogonal transformation preserves the orthogonality among the transformed errors if the original error are i.i.d., but the serial correlation in the transformed errors can be induced either by serial correlation in the original errors or by heteroskedasticity across time[19].

To discriminate between these two hypotheses, we can set up a test for serial correlation using constructed first-difference residuals from the orthogonal deviations estimates. If the residuals have been transformed to first-differences, first-order serial correlation is to be expected[20] but not second-order (AB, 1991, p.281)[21].

The **generalised m_2 test**[22] statistic for second-order serial correlation has the following form. Let us consider first the GMM estimator

$$\hat{d} = (\bar{X}' Z A_N Z' \bar{X})^{-1} \bar{X} Z A_N Z' \bar{y} \qquad (33)$$

where $\bar{X}$, $\bar{y}$ denote the transformation used in calculating the estimator (first differences or orthogonal deviations), whilst $A_N = (\frac{1}{N}\sum_{i=1}^{N} z_i' \hat{\bar{v}}_i \hat{\bar{v}}_i' z_i)^{-1}$. From this we form the first difference residuals

$$\hat{v} = \Delta y - \hat{d}' \Delta X \qquad (34)$$

Thus the generalised m_2 test can be expressed as

$$m_2 = \frac{\hat{v}_{-2}' \hat{v}_*}{\hat{v}^{1/2}} \sim N(0,1) \quad \text{under the assumption that } E(v_{it} v_{i,t-2}) = 0 \qquad (35)$$

[23]

where

$$\hat{v} = \sum_{i=1}^{N} \hat{v}_{i(-2)}' \hat{v}_{i*} \hat{v}_{i*}' \hat{v}_{i(-2)} - 2(\hat{v}_{-2}' \Delta X_*)(\bar{X} Z A_N Z' \bar{X})^{-1} \bar{X}' Z A_N$$

Here $\hat{v}_{-2}$ is the vector of first differenced residuals lagged twice, $\hat{v}_*$ is a vector of trimmed $\hat{v}$ to match $\hat{v}_{-2}$ and similarly for ΔX_* ; A_N has already been defined and $\hat{\bar{v}}_i$ is a vector (T- 2)x1 of the two-step GMM residuals, and avar ( ) represents the asymptotic covariance matrix.

$$\left[\sum_{i=1}^{N} Z_i' \hat{\hat{v}}_i \hat{v}_{i*}' v_{i(-2)}\right] + \left(\hat{v}_{-2}' \Delta X_*\right) av\hat{a}r\ (\hat{d}) \left(\Delta X_*' \hat{v}_{-2}\right). \tag{36}$$

Finally it is useful to report the Sargan test of over-identifying restrictions[24] (Sargan, 1958 and 1988; Hansen,1982): if A_N is optimal then under the null of validity of the instruments in z the test statistic is

$$s = (\sum_{i=1}^{N} \hat{v}_i^{*'} z_i) A_N (\sum_{i=1}^{N} z_i' \hat{v}_i^*) \sim c_r^2 \tag{37}$$

Where r represents the difference between the number of columns in z and the number of columns in X[25].

5. Consequences of Non-Stationarity for Panel Data

So far we have considered the conventional case of a typical panel which is implicitly assumed to consist of stationary variables. In this section we extend the analysis to consider the consequences of non-stationarity. The following section can, in practice only be relevent when we have a sufficiently large number of time series observations to begin to exploit the non-stationarity. This will not be true of all panels of data as in some cases we may be dealling with T as small as 4 or 5.

This section also deals with the conditions for valid aggregation of a set of micro economic relationships to provide a valid macro relationship. A series of papers by Pesaran and associates (Pesaran Pierse and Kumar(1989)), Lee, Pesaran and Pierse(1990)) for instance concludes that econometrics should proceed at as micro a level as possible, but not using "panel data" estimators. They argue that if micro relationships are dealing with non-stationary data then, even if these relationships cointegrated, the properties of a derived aggregate model or pooled models will be even worse than we previously thought. Robertson and Symons, (1992) and, more recently, Pesaran and Smith (1995) have shown that, with a data set of this type, inference often proceeds by imposing equality restrictions on parameters across individuals or through time (i.e. panel data estimators). As we have already shown, this may produce serious problems in a stationary world (especially in the presence of lagged dependent variable). They go on to show that in the presence of non stationarity but cointegrated micro relationships aggregation can completely invalidate the macro relationship (and/or the panel estimator).

It should be clarified here that the principal interest in this section has to be achieving a good estimate of the average or aggregate effect of an economic relationship. If we are interested in understanding the individual components of a panel then individual time series regressions is really the only way to proceed. However an important question remains unresolved as to the best way to achieve this average estimate. Pooling is often assumed to give the best estimate of the average effect across the panel, but we are beginning to understand that this may often not be the case.

Pesaran and Smith (1995) in particular state that the common practice of aggregating and pooling by assuming homogeneity in dynamic models is "far from being innocuous"; instead they suggest estimating the individual micro equations and then taking the means of the estimated micro-parameters and relative standard errors. We think that the results from these authors are very important in practice, thus we will briefly summarise them as follows.

5.1 The Pesaran and Smith (1995) case

Pesaran and Smith (1995) address the problem of estimating the average long run relationship between a set of variables when the micro relationships are made up of I(1) variables which cointegrate but with different cointegrating vectors. They conclude that the micro single equation approach gives consistent estimates of the long-run parameters, whilst the conventional view (Zellner(1969), Malinvaud(1956)) that the pooled and aggregate time-series estimators will also provide consistent estimators of the mean effects, is no longer valid.

In order to demonstrate this we can make use of a very simple example. Let us suppose that x_{it} are I(1) and there is a single cointegrating relationship between y_{it} and x_{it} for each group, with the parameters varying randomly across groups, i.e. suppose that

$$y_{it} = \boldsymbol{b}_i \; x_{it} + \in_{it} \quad i = 1, \ldots\ldots, N \quad t = 1, \ldots.., T \tag{38}$$

where $\in_{it}$ is a stationary process which is integrated of degree 0, I(0), which implies that each of the relationships cointegrate, and that

$$\boldsymbol{b}_i \neq \boldsymbol{b}_j \quad i \neq j \tag{39}$$

Now suppose that we aggregate, then the variables will be

$$\overline{y} = \sum_{i=1}^{n} y_i, \quad \overline{x} = \sum_{i=1}^{n} x_i \tag{40}$$

Pesaran and Smith (1995) argue that the aggregated relationship does not cointegrate as exact aggregation of (38) over n gives

$$y_1 + \ldots + y_n = \boldsymbol{b}_1 \; x_1 + \ldots + \boldsymbol{b}_n \; x_n + \boldsymbol{e}_1 + \ldots + \boldsymbol{e}_N \tag{41}$$

which gives

$$\overline{y} = \boldsymbol{b}_1 \; \overline{x} - \sum_{i=2}^{n} (\boldsymbol{b}_1 - \boldsymbol{b}_i) \; x_i + \sum_{i=1}^{n} \boldsymbol{e}_i \tag{42}$$

given that x is I(1) we would only therefore expect to find cointegration between $\overline{y}$ and $\overline{x}$ when $\boldsymbol{b}_1 = \boldsymbol{b}_j$ all j. So if we perform the standard aggregate regression with dynamic terms we will be dealing with non stationary aggregates which do not cointegrate and we would expect the parameter value of the aggregate long run relationship to tend to zero even though all the micro relationships do in fact cointegrate. This is the result pointed out by Robertson and Symons(1992).

The implications of this are profound for the choice of panel data estimation technique. The aggregate regression is obviously invalid as it does not represent a cointegrating relationship.

The pooled estimator is also invalid as the assumption of a common parameter across all the individuals produces non stationary individual relationships which implies serial correlation in the residuals. If lagged variables are included to deal with this serial correlation then the coefficient on the lagged dependent variable tends to one and the long run relationship disappears. Just as in the case of aggregate time series models under non-cointegration.

The cross section regression(aggregating over time) may perform fairly well as a model of the long run relationship as, by suppressing the dynamics it becomes broadly equivalent to the time series notion of a static regression and the non-cointegration of the individual groups actually tends to average out. However there is no way to detect whether there are actually underlying cointegrating relationships in this framework.

Pesaran and Smith therefore always recommend performing the full set of individual time series regressions as a starting point to a panel data exercise. They then test each member for cointegration and see if the cointegrating parameters may be similar. Then perform the pooled and cross section regressions and compare the implied long run relationships. If all approaches give compatible results then the results are likely to be robust. Otherwise an understanding of the sources of bias may help to determine the reliable estimator in specific cases.

5.2 A Special Case of Valid Aggregation (Hall and Urga, 1996)

In this section we present the argument of Hall and Urga (1996) that, while the basic point made above is quite correct, there is a special case which does allow valid estimation even when the parameters of the micro relationships are heterogeneous. And, moreover we believe that this special case is valid for many real world situations. The basic argument put forward here is that if the exogenous variables in the micro relationships are driven by a common stochastic trend then the simple aggregate relationship can be shown to cointegrate. Moreover in many real world examples we might expect the nonstationary component to be common across a set of micro data, for example wages in different sectors may all be non-stationary but relative wages across sectors might well be stationary. For completeness we will consider a full multivariate case of p regressors x_{jt} ,j=1...p for each of the individual components of the panel(I). This argument may be seen formally quite simply, suppose that the exogenous variables are all driven by the following common trend model

$$\begin{aligned} x_{ijt} &= \boldsymbol{a}_{ij}\; x_{jt} + \boldsymbol{m}_{ijt} \\ \Delta\; x_{jt} &= \boldsymbol{x}_{jt} \end{aligned} \tag{43}$$

where m_{ijt} are $\boldsymbol{x}_{jt}$ stationary ARMA error processes, that is they are integrated of degree 0, I(0). Then x_{jt} becomes the common stochastic trend which drives all the individual x_{ij}'s.

We can then express the aggregate relationship as

$$\begin{aligned} \overline{y}_t &= \sum_{j=1}^{p}\sum_{i=1}^{n} \boldsymbol{b}_{ij}\; x_{ijt} + \sum_{i=1}^{n} \boldsymbol{e}_{it} \\ &= \sum_{j=1}^{p}\sum_{i=1}^{n} \boldsymbol{b}_{ij}\; \boldsymbol{a}_{ij}\; x_{jt} + \sum_{i=1}^{n} \boldsymbol{e}_{it} + \sum_{j=1}^{p}\sum_{i=1}^{n} \boldsymbol{b}_{ij}\; \boldsymbol{m}_{ijt} \end{aligned} \tag{44}$$

and in terms of the aggregates this becomes

$$\overline{y_t} = \sum_{j=1}^{p}\left(\sum_{i=1}^{n} b_{ij}\, a_{ij} \,/\, \sum_{i=1}^{n} a_{ij}\right)\bar{x}_{j,t} + \sum_{i=1}^{n} e_{it} + \sum_{j=1}^{p}\sum_{i=1}^{n} b_{ij}\, m_{ijt} + \sum_{j=1}^{p}\sum_{i=1}^{n} m_{ijt}\left(\sum_{i=1}^{n} b_{ij}\, a_{ij} \,/\, \sum_{i=1}^{n} a_{ij}\right) \quad (45)$$

so that the aggregate equation cointegrates (the error term consists of a linear combination of weighted stationary ARMA components, which is of course stationary) and moreover the aggregate coefficients are a weighted average of the coefficients in the micro relationships.

The key to this special case is, of course, the validity of the common factor linking the x variables. We can speculate that in many cases the non stationary part of a group of related micro series might well be common, panels of wage data, consumption, prices, etc...might well have this property. Indeed it would be surprising if relative wages or prices between sectors were non-stationary and so we might expect that the common factor representation would often be a good one in terms of the main non-stationary component in most data series. We would also suggest that if this is a common property of many data sets then it is a formal explanation of why aggregate econometric estimation work as well as it does, despite the standard conditions for aggregation which are highly implausible.

Formally this suggests that an important stage in analysing a panel of data should be an investigation of the existence of common stochastic trends amongst the individual components of the panel. This can be done in the autoregressive representation by testing for the presence of (n-1) cointegrating vectors amongst a set of n series (thus implying one common stochastic trend) or it can be accomplished in the moving average representation by testing for the presence of a single common factor amongst the series following Geweke (1977).

5.3 Hall , Lazarova and Urga (1999)

In this paper the author propose a new approach to test for the number of common stochastic trends driving the nonstationary series in a panel data set based around principal component techniques. The procedure enables one to carry out the testing even if we have a mixture of I(0) and I(1) series in the sample.

With a set of Monte Carlo experiments they assess the empirical relevance of the testing procedure. The test is shown to have reasonable size and power when the sample size T is larger than the number of series N. The test performs best when there are relatively few stochastic trends underlying the data. The size of the test improves with increasing numbers of stationary series present in the sample while the power deteriorates.

The principal components approach allows one to carry out the test even when N is equal or greater than T. However, from the first simple experiments reported in the paper, the power of the test deteriorates. Further, the estimation of the number of common stochastic trends is done in order to validate the aggregate relationship.

6. Unit Roots and Cointegration in Panel Data

The main studies in testing for unit roots in panel data are Breitung and Meyer (1994), Quah (1994), Levin and Lin (1992, 1993), Im,Pesaran and Shin (1997), Maddala and Wu (1998) and Hall, Lazarova and Urga (1999). Breitung and Meyer (1994) derived the asymptotic normality of the Dickey and Fuller test statistic for panel data with a large cross-section dimension and a small time-series dimension. Quah (1994) studied a unit root test for panel data that have simultaneous extensive cross-section and time-series variation. He showed that the asymptotic distribution for the proposed test is a mixture of the standard normal and Dickey-Fuller asymptotics. Levin and Lin (1993) derived the asymptotic distributions for unit roots on panel data and showed that the power of these tests increases dramatically as the cross-section dimension increases. Im, Pesaran and Shin (1997) criticised the Levin and Lin panel unit root statistics and proposed alternative procedures. Maddala and Wu (1998) provides a comparison of the Levin-Lin and Im-Pesaran-Shin tests. They also suggests a new test based on the Fisher test. However, to date, little is known about cointegration tests and estimation with regression models in panel data. Exceptions are Kao (1997), Kao and Chiang (1997), McCoskey and Kao (1998a,1998b), Pedroni (1996, 1997), Philipps and Moon (1997) and Hall, Lazarova and Urga (1999). In the first half of Kao (1997), the author studies a spurious regression in panel data. Asymptotic properties of the ordinary least squares (OLS) estimator and other conventional statistics were examined. Kao (1997) showed that the OLS estimator is consistent for its true value, but the t-statistic diverges so that inferences about the regression coefficient are wrong with a probability that goes to one. Furthermore, Kao (1997) examined the Dickey-Fuller and the augmented Dickey-Fuller tests to test the null hypothesis of no cointegration in panel data. Kao and Chiang (1997) studied the asymptotic results for a least-squares dummy variable (LSDV) estimator, a fully modified estimator and a dynamic least-squares estimator in a cointegrated regression in panel data. McCoskey and Kao (1998a) proposed further tests for the null hypothesis of cointegration in panel data. McCoskey and Kao (1998) surveyed various tests for cointegration in panel data providing Monte Carlo comparisons. Pedroni (1997) derived asymptotic distributions for residual-based tests of cointegration for both homogenous and heterogeneous panels. Pedroni (1996) proposed a fully modified estimator for heterogeneous panels. Phillips and Moon (1997) developed a sequential limit theory for non-stationary panel data. As mentioned in the previous section, Hall, Lazarova and Urga (1999) propose a new approach to test for the number of common stochastic trends driving the nonstationary series in a panel data set. In the remainder of this section we briefly summarise the IPS and MW tests concerning unit roots testing procedure in panel data. We finally conclude with a brief summary of the three main recent works on cointegration in panel data.

6.1 Panel Data Unit Root Tests: Im-Pesaran and Shin (1997) vs Maddala and Wu(1998)

We begin by describing the two main testing procedures for unit roots in heterogeneous panel data. The first papers which proposed unit roots test in panel are the ones by Quah (1992,1994-Q henceforth) and Levin and Lin (1993-LL henceforth). Both of them do not consider the case of heterogeneous panels even though the LL test allows for some individual specific effects as well as for some heterogeneity across groups. The LL test also requires that N/T tend to zero. Because of all those limitations, the two tests are not considered further in this survey.

We therefore only give detailed descriptions of the two more suitable tests for a unit root. The first test is the one proposed by Im-Pesaran and Shin (1997, IPS henceforth) and the second is the one introduced by Maddala and Wu (1998, MW henceforth).

In terms of the Q and LL tests, the procedure test for a unit root by setting up the following model.

$$y_{it} = \boldsymbol{r}_i y_{it-1} + e_{it}$$

The null of a unit root is then given by imposing the hypothesis that $\boldsymbol{r}_1 = \boldsymbol{r}_2 = ... = \boldsymbol{r}_N = \boldsymbol{r}$ and then test for $\boldsymbol{r} = 1$ against the alternative that $\boldsymbol{r} < 1$. IPS and MW relax the assumption that $\boldsymbol{r}_1 = \boldsymbol{r}_2 = ... = \boldsymbol{r}_N = \boldsymbol{r}$. In fact the null hypothesis become $\boldsymbol{r}_i = \boldsymbol{r} = 1$ for all i and $H_1 : \boldsymbol{r}_i < 1$ for at least one i. It is clear that the hypothesis that all components have a unit root against the alternative that all components do not have one is too restrictive. MW present some experiments in which the power of their test is evaluated under the alternative hypothesis that some i $H_1 : \boldsymbol{r}_i = 1$ while for other i is $H_1 : \boldsymbol{r}_i < 1$. Having said this, the main disadvantage of the MW test remains in the way in which the test is constructed. The test is computationally intensive. The p-values of the test statistics have to be derived by Monte Carlo simulations, so a large volume of values must be simulated and stored for every combination of number of lags and deterministic regressors.

Im-Pesaran and Shin (1997)

Let us consider now the IPS procedure. This test explicitly accommodates heterogeneity across groups and for different patterns of serial correlation. The test is based on averaging individual unit root test statistics for panels. In particular they propose a test based on the average of the individual t-test (t-bar test) that in a second version of the paper is replaced by the average of the Lagrange Multiplier (LM) statistics computed for each group/firm in the panel (LM-bar test).

Let

$$\Delta y_{i,t} = \boldsymbol{a}_i + \boldsymbol{r}_i y_{i,t-1} + e_{i,t} \quad i = 1,,2,....., N \text{ and } t = 1,2,......., N \tag{46}$$

The null hypothesis may be defined as
$H_0 : \boldsymbol{r}_i = 0$ for all i against the alternative

$$H_A : \boldsymbol{r}_i < 0,\ i = 1,2,...N_A,\ \ \boldsymbol{r}_i = 0,\ i = N_A + 1, N_A + 2,........, N\ .$$

The ADF regressions

$$\Delta y_{i,t} = \boldsymbol{r}_i y_{i,t-1} + \sum_{j=1}^{p_i} \boldsymbol{q}_{i,j} \Delta y_{i,t-j} + \boldsymbol{a}_i + \boldsymbol{h}_{i,t}\ ,\ t = 1,2,...., T \tag{47}$$

are estimated and the LM-statistics testing $\boldsymbol{r}_i = 0$ is computed, i.e.

$$(\overline{LM})_{N,T} = N^{-1} \sum_{i=1}^{N} LM_{i,T}(p_i, \boldsymbol{J}_i) \tag{48}$$

with $\boldsymbol{J}_i = (\boldsymbol{J}_{,1}, \boldsymbol{J}_{i,2},...., \boldsymbol{J}_{ip_i})'$ and $LM_{i,T}(p_i, \boldsymbol{J}_i)$ is the individual statistic for testing $\boldsymbol{r}_i = 0$.

The author show that the test only requires that N/T tends to k (a finite positive constant) and not to 0 as in the LL case. Finally, the authors in addition to the $\overline{LM}$ test also report $\Gamma()$ $\overline{LM}$ and $\Psi(\overline{LM})$. $\overline{LM}$ is the statistics which tests for $r_i=1$, while the $\Gamma(\overline{LM})$ and $\Psi(\overline{LM})$ represent standardised version of the $\overline{LM}$ statistics which are asymptotically distributed as N(0,1). It is worth noticing that the $\Psi(\overline{LM})$ specifically allows for the different lag structure of the single ADF test.

The are two main setbacks associated to this test. The first concerns the way the evidence of several independent unit root test are combined. The second is that we are evaluating unit roots at the panel data level in which the alternative hypothesis is quite general (All have a unit root against some of them do not). The IPS paper does not investigate the behaviour of the power of the test under different combinations of stationary/non-stationary series present in the panel. Although we do have this information from the experiments undertaken by MW, where it is shown that the IPS test is less powerful than the MW test.

Maddala and Wu(1998)

The test proposed by Maddala and Wu seems to overcome some of those limitations. As in IPS, the test is based on the significance of different independent tests. MW use the Fisher test, which is a non-parametric procedure and is based on the *p*-values p_i with the statistics calculated as

$$F = -2\Sigma \log_e p_i \sim c^2 \qquad (49)$$

with 2N degrees of freedom and N is the number of separated samples. The Fisher test is an exact test and not an asymptotic test as the IPS. MW report some simulations which reveal the power of the test as the number of stationary series in the sample increases.

6.2 Panel Data Cointegration

In this section we provide a brief survey on recent developments on various tests proposed for cointegration in panels data (McCoskey and Kao (1998) and Banerjee (1999) provide a further detailed presentation of the various tests).

We will mainly outline and compare three recent studies which present panel data test for cointegration: Kao (1997), Pedroni (1997) and McCoskey and Kao (1998). It is important to note that the first two papers present tests where the null hypothesis of no cointegration is considered and use residual based tests constructed (derived) from Engle and Granger type static regression. While the last paper introduce a test of the null of cointegration and the test is also residual based with its analogue in the time series literature proposed by Harris and Inder (1994) and Shin (1994).

6.2.1 Testing for cointegration in panels with the null hypothesis of no cointegration

The first set of residual based cointegration tests for panels are based on the null hypothesis of no cointegration just as in the time series counterpart (see McCoskey and Kao(1998) for more details). There are two possible regressions whose residuals are analyzed for stationarity. We have (a) the case of varying intercepts and common slopes and (b) of varying intercepts and varying slopes. The first set of tests assumes a form of homogeneity in the relationship of the variable allowing for heterogeneity only in the intercepts. We are not reporting this case which is fully described by McCoskey and Kao (1998). Here we concentrate on case (b) which is more relevant to the case of heterogeneous panels.

Kao (1997)

Kao proposed testing for cointegration using an ADF test for varying slopes and varying intercepts. Suppose we have the following model

$$y_{i,t} = \alpha_i + x_{i,t}^{'}\beta_i + e_{i,t} \quad \text{with } i = 1,2,....,N\text{ , } t = 1,2,.....,T \tag{50}$$

The peculiarity o f this model is that we test for cointegration for each cross-section under the assumption that the cross sections are assumed independent of each other while we allow for heteroskedasticity across the cross-sections.

The ADF test is constructed as

$$\Delta\hat{e}_{i,t} = \rho_i\hat{e}_{i,t-1} + \sum_{j=1}^{p}\vartheta_{ij}\Delta\hat{e}_{i,t-j} + v_{i,tp} \tag{51}$$

The null hypothesis $H_0 : \rho_i = 0$ of no cointegration for each i is evaluated via the test statistic

$$t_{i,ADF} = \frac{(\hat{e}_{-1}^{'}Q_{xp}\hat{e}_{-1})^{1/2}\,\hat{\rho}_i}{s_v} \tag{52}$$

where $\hat{e}_{-1}$ is the vector of observations of $\hat{e}_{t-1}$, $Q_{Xp} = I - X_p(X_p^{'}X_p)^{-1}X_p^{'}$ where X_p is the matrix of observations of the p regressors ($(\Delta\hat{e}_{t-1},......,\Delta\hat{e}_{t-p})$; $s_v^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{v}_{tp}^2$.

To test for cointegration for the whole panel the t-statistic is

$$\bar{t}_{ADF} = \frac{1}{N}\sum_{i=1}^{N} t_{i,ADF} \tag{53}$$

which, using the result from Phillips and Ouliaris(1990) who show that the ADF converges to a functional of Brownian motion, it is possible to prove that the $\bar{t}_{ADF}$ is asymptotically distributed as

$$\sqrt{N}(\bar{t}_{ADF} - \mu_{ADF}) \Rightarrow N(0, \sigma_{ADF}^2) \tag{54}$$

Pedroni (1997)

Pedroni proposes a different test based on the average, across the cross-sections, of the Phillips Z_t statistics, valid for the model for varying intercepts and slopes (see Phillips and Ouliaris (1990) for details on how to calculate the Phillips Z_t statistics).

His first test considers averages of test for cointegration for each cross-section and Pedroni proposes the following variant of the Z_t statistics:

$$Z_{\hat{a}-1} = \sum_{i=1}^{N} \frac{\sum_{t=1}^{T} (\hat{e}_{i,t-1} \Delta \hat{e}_{i,t} - \hat{\boldsymbol{l}}_i)}{(\sum_{t=1}^{T} \hat{e}_{i,t-1}^{2})} \quad (55)$$

where $\hat{e}_{i,t}$ represents the residual from the cointegrating regression with $\boldsymbol{l}_i$ correction coefficient from variances and covariances for the single regression (see the original paper for details).

Pedroni shows that

$$\sqrt{N}\,(T\hat{Z}_{\hat{a}-1} + 9.05) \Rightarrow N(0{,}35.98). \quad (56)$$

This second test groups the statistics such that the averaging is done in pieces so that the limiting distributions are based on limits of piecewise numerator and denominator terms (see the original paper plus McKoskey and Kao (1998) and Banerjee (1999) for more details about the test)

6.2.2 Testing for cointegration in panels with the null hypothesis of cointegration

This test, fully discussed in McCoskey and Kao (1998), is a residual based test adapted from the cointegration test of the null of cointegration in time series proposed by Harris and Inder (1994) and Shin (1994). It is important to note that "for models s which consider the case of cointegration vector to change across the cross-sectional observation, the asymptotics depend merely on the time series results as each cross-section is estimated independently. For models with common slopes, the estimation is done jointly and therefore the asymptotic theory is based on the joint estimation of a cointegrated relationship in panel data".

The model with varying slopes and intercepts may be written as follows:

$$y_{i,t} = \boldsymbol{a}_i + x_{i,t}^{'} \boldsymbol{b}_i + e_{i,t} \quad \text{with } i = 1,2,....,N\ ,\ t = 1,2,.....,T \quad (57)$$

where $x_{i,t} = x_{i,t-1} + \boldsymbol{e}_{i,t}$, $e_{i,t} = \boldsymbol{g}_{i,t} + u_{i,t}$ and $\boldsymbol{g}_{i,t} = \boldsymbol{g}_{i,t-1} + \boldsymbol{q} u_{i,t}$.

There is cointegration if $\boldsymbol{q} = 0$ and the test proposed by McCoskey and Kao is

$$\overline{LM} = \frac{\frac{1}{N} \sum_{i=1}^{N} \frac{1}{T^2} \sum_{t=1}^{T} S_{i,t}^{+2}}{s^{+2}} \quad (58)$$

where $S_{i,t}^{+2} = \sum_{j=1}^{t} \hat{e}_{i,t}^{+2}$ is the partial sum of the residuals and

$$s^{+2} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{e}_{i,t}^{+2} \quad (59)$$

The test is asymptotically distributed:

$$\sqrt{N(\overline{LM} - \boldsymbol{m}_n)} \Rightarrow N(0, \boldsymbol{s}_n^2) \quad (60)$$

where $m_n = .1162$ and $s_n^2 = .0109$ fully defined in McKoskey and Kao (1998). Finally, the authors show that the limiting distribution of the $\overline{LM}$ test is free of nuisance parameters and robust to heteroskedasticity.

7. Conclusions

This chapter began by outlining the relatively simple traditional models used in panel data estimation. It then illustrated the problems, which arise as dynamics are introduced into a model which allows for the possibility of heterogeneous panel members. This complication undermines a great deal of the standard panel data results. It then discussed some of the techniques which have been proposed to deal with these problems. We then went on to explore some recent work which has been developing on the importance of non-stationarity in panel data estimation. This elaboration again proves to have profound implications and it is not established that it is important to investigate both the existence of non-stationarity in a panel and the possibility of cointegration. Finally we outlined a range of tests which have been developed recently to allow us to achieve this final objective.

In conclusion, we offer the following guidelines. When we are dealing with a panel of stationary data then for an accurate dynamic panel data analysis first, in according with the characteristics of the data set, one has to compare the estimates in levels and deviations or first differences. Statistically significant differences between the two indicate that correlated individual effects are omitted from the regression in levels. To detect this both the traditional Hausman and the Arellano-Bover tests can be utilised.

The efficient estimator when T is small and N is large is a "generalised method of moments", taking the first differences or orthogonal deviations to eliminate the fixed effects (Arellano and Bover, 1995). When T is large, the "within-groups" estimator provides consistent estimates. Alternatively, the equivalent orthogonal deviations transformation allows one to obtain optimal generalised method of moments estimators for models with only predetermined variables.

If the panel consists of non-stationary data then the correct treatment of the nonstationarity becomes an over riding consideration. Recent work suggests that aggregate relationships may perform very poorly if the micro relationships are cointegrated but with different cointegrating vectors. This is not the case however when the exogenous variables in the micro relationships are all driven by a single stochastic trend. We argue that this is an empirically relevant special case by outlining a testing procedure for this condition and showing that a well known panel data set conforms to this condition. In that case aggregate estimation seems to perform well, as expected. Once we have detected non-stationarity in the panel it is then important to asses the degree to which this is generated by a small number of common stochastic trends. If this is the case the aggregate estimation may work well. If it is not then aggregating over time and using a cross section regression is preferable. In either case a useful way to proceed is to estimate a range of estimators and to compare the results in the light of our improved understanding of the performance of these estimators under non-stationarity.

Endnotes

1. G.Urga wishes to acknowledge that this work was part of an ESRC research project (Grant N. R022251032) entitled "An Analysis of the Importance of Common Stochastic Trends and the Methods of Selecting Them" joint with Stepana Lazarova and Stephen Hall.
2. Amongst the topics omitted are the maximum likelihood (Bhargava and Sargan 1983) and Chamberlain's (1982, 1984) π-matrix procedures to these models (see Sevestre and Trognon 1990, and especially, Arellano and Bover 1989, for an overview), as well as all simultaneous equation systems and non-linear models (see Hsiao 1986). See Matyas and Sevestre (1996) for further details.
3. In many cases one does not have panel data but samples of surveys of independent cross-sections, called "cohorts". "A "cohort" is defined as a group with fixed membership, individuals of which can be identified as they show up in the surveys" (Deaton 1985, p. 109).
4. This section applies only to the large T case, or to static models with strictly exogenous regressors for small T. Static models with predetermined regressors require methods like those of Section 3 when T is small.
5. $\hat{b}$ is consistent as N $\to\infty$ or T $\to\infty$ or both, whilst $\hat{h}_i$ is only consistent as T $\to\infty$
6. Noting that Q' Q = Q2 = Q since Q is idempotent
7. See Hausman and Taylor (1981), Hsiao (1985, 1986) and Arellano and Bover (1989) for a full discussion
8. See Hsiao (1986) and Holly (1982) for a full discussion on tests for mis-specification, and on the relationship between Hausman's specification test and conventional test procedures.
9. Note that an equivalent but computationally simpler statistic than (16) can be based on the within-groups and between groups estimators (see Hausman and Taylor 1981).
10. Arellano and Bover (1989, p. 14). This test requires xit to be strictly exogenous.
11. Note that the "bar" notation no longer indicates the mean as it did earlier in the paper.
12. Holtz-Eakin, Newey and Rosen (1988) use a similar estimator for vector autogressions
13. See Baltagi (1995), chapter 8, for a very comprehensive presentation of the estimator.
14. Bhargava and Sargan (1983), and Chamberlain (1982) developed this type of model using the maximum likelihood estimator (MLE) and the minimum distance estimator (MDE), respectively.
15. See also Arellano and Bover (1989).
16. It is worth noting that in a large T setting it is a matter of indifference whether one uses the WG or the OD transformation to obtain the "withing-groups" estimator, since they will be identical. The advantage of OD over WG arises, when doing IV, in the small T setting, since OD (like first differencing) preserves orthogonality between the transformed errors and lagged values of predetermined regressors, whilst WG does not. It is this orthogonality that is important, not whether the transformed errors are serially uncorrelated. If the original errors are serially uncorrelated but heteroscedastic over time, the transformed errors after OD will be serially correlated, but remain uncorrelated with lagged levels of predetermined regressors.
17. AB show that the framework so far presented can be easily extended to the case of unbalanced panel data, i.e. when the number of time periods may vary from unit to unit as well as the historical points to which the observations correspond (see also Hsiao

1986, ch. 8).

18. At the end of Section 2 we reported the Hausman and Arellano and Bover tests to detect the correlation between the unobservable individual effects and the right-hand side variables for static models. Arellano (1993) extends the Hausman test to the case of dynamic models, by proposing a Wald type test robust in the presence of heteroscedasticity and autocorrelation.
19. These causes have very different implications. For example, serial correlation in the original errors affects instruments validity.
20. E(vit vi,t-1) need not be zero.
21. Note that this is not equivalent to taking first differences of the OD residuals.
22. We thank Manuel Arellano for providing us in 1989 this still unpublished result.
23. Note that if $\overline{X}$, $\overline{y}$ also denote first differences, then (35) reduces to the ordinary m2 test given in AB, equation (8).
24. AB also report the Sargan difference tests to discriminate between nested hypotheses which concern serial correlation in sequential way.
25. An increasing number of econometric packages can be utilised to undertake analysis of panel data in economics. LIMDEP and TSP, for instance, produce standard estimators (pooled, within groups and between groups, and random effects) for panel data (balanced and unbalanced). The minimum distance estimator in LSQ and GMM can be utilised to estimate linear, non linear, and dynamic models. Arellano and Bond (1988), finally, have developed the DPD program written in Gauss to compute optimal estimators and specification tests reported in this paper.

Extended bibliography

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

Ameniya, T. And T.E. Macurdy (1986): "Instrumental - Variable Estimation of an Error-Components Model", Econometrica, 54 pp. 869-881.

Anderson, T.W. and C. Hsiao (1981): "Estimation of Dynamic Models Using Panel Data", *Journal of American Statistical Association*, 76 pp. 598-606.

Anderson, T.W. and C. Hsaio (1982): "Formulation and Estimation of Dynamic Models Using Panel Data", *Journal of Econometrics*, 18, pp. 47-82.

Arellano, M. (1988): "An Alternative Transformation for Fixed Effects Models with Predetermined Variables", Institute of Economics and Statistics and Applied Economics Discussion Paper no. 57.

Arellano, M. (1992): "On the Testing of Correlated Effects with Panel Data", *Journal of Econometrics*, 59, 87-97.

Arellano, M. and Bond S. (1988): "Dynamic Panel Data Estimation Using DPD-A Guide for Users", Institute for Fiscal Studies W.P.n. 88/15.

Arellano, M. and Bond S. (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", *Review of Economic Studies*, 58, 277-297.

Arellano, M. and Bover O. (1989): "La Econometria de Datos de Panel", mimeo London School of Economics. The paper is published in *Investigaciones Economicas*, Vol. XIV n. 1 pp. 3-47.

Arellano, M. and Bover O. (1995): "Another Look at the Instrumental-Variable Estimation of Error-Componenets Models", Journal of Econometrics, 68, 29-51.

Baker, T. and Pesaran M.H. (1990): *Disaggregation in Econometric Modelling*. Routledge, London.

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

Baltagi, B.H. (1995), *Econometric Analysis of Panel Data*, John Wiley and Sons.

Baltagi, BH (1996), "Specification Issues", Chapter 12 in Matyas, L. and Sevestre, P. (eds) *The Econometrics of Panel Data*, Handbook of Theory and Applications, Kluwer Academic Publisher (second edition).

Banerjee, A. (1999), "Panel Data Unit Roots and Cointegration: An Overview", in *Oxford Bulletin of Economics and Statistics*, forthcoming (November 1999).

Bhargava, A. (1991): "Identification and Panel Data Models with Endogenous Regressors", *Review of Economic Studies*, 58, 129-140.

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

Blanchard, P. (1996), "Software Review", in Matyas and Sevestre 91996), op. cit.

Breitung,J. and Mayer,W. (1994), "Testing for Unit Roots in Panel Data: Are Wages on Different Bargaining Levels Cointegrated?", *Applied Economics*, 26, 353-361.

Breusch, T.S., Mizon, G.E. and Schmidt, P. (1989): "Efficient Estimation Using Panel Data", *Econometrica*, 57, 695-700.

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

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

Chamberlain, G. (1984): "Panel Data", in Z.Griliches and M.D. Intreiligator, (eds), *Handbook of Econometrics*, vol. II. Elsevier Science.

Deaton, A. (1985): "Panel Data from Time Series of Cross-Sections", *Journal of Econometrics*, 30, 109-126.

Geroski,P., Lazarova,S. and G.Urga (1999), Modelling the dispersion in productivity growth across firms over time, mimeo, Centre for Economic Forecasting, London Business School.

Ghose, D., 1995, Linear aggregation in cointegrated systems, *Journal of Economic Dynamic and Control*, 19, 1011-1032.

Gonzalo, J., 1993, Cointegration and aggregation, *Ricerche Economiche*, 47, 281-291.

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

Hall, S., S.Lazarova and Urga, G. (1999), "A Principal Components Analysis of Common Stochastic Trends in Heterogeneous Panel Data: Some Monte Carlo Evidence", forthcoming in *Oxford Bulletin of Economics and Stastistics* (November 1999).

Hall, S. and Urga, G. (1998), "Stochastic Common Trends and Long-Run Relationships in Heterogeneous Panels", D.P. 27-95 (revised), Centre for Economic Forecasting, London Business School.

Hansen, L.P. (1982): "Large Sample Properties of Generalised Method of Moments Estimators", *Econometrica*, 50, 1029-1054.

Hansen, L.P. and Singleton K.J. (1982): “Generalised Instrumental Variables Estimation of Nonlinear Rational Expectation Methods”, *Econometrica*, 50, 1269-1286.

Harris, D. and B. Inder (1994), “A Test of the Null Hypothesis of Cointegration”, in *Non-Stationary Time Series Analysis and Cointegration*, ed. C. Hargreaves, Oxford University Press, New York.

Hausman, J.A. (1978): Specification Tests in Econometrics”, *Econometrica*, 46, 1251-1271.

Hausman, J.A., and Taylor, W.E. (1981): “Panel Data and Unobservable Individual Effects”, *Econometrica*, 49 1377-1398.

Hausman, J.A. and Wise D.A. (1979): “Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment”, *Econometrica*, 47, 455-473.

Heckman, J.J. (1979): “Sample Selection Bias as a Specification Error”, *Econometrica* 47, 153-161.

Holly, A. (1982): “A Remark of Hausman’s Specification Test”, *Econometrica*, 749-759.

Holtz-Eakin, D. (1988): “Testing for Individual Effects in Autoregressive Models”, *Journal of Econometrics*, 39, 297-307.

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

Hsiao, C. (1985): “Benefits and Limitations of Panel Data”, *Econometric Review*, 4 (1). 121-174.

Hsiao, C. (1986): *Analysis of Panel Data*, Cambridge: Cambridge University Press.

Hsiao,C., Lahiri,K., L.F. Lee, and M.H.Pesaran (1999), *Analysis of Panels and Limited dependent Variables Models*, Cambridge University Press.

Im,K.S., Pesaran,M.H. and Shin,Y. (1997), “Testing for Unit Roots in Heterogeneous Panles”, Department of Applied Economics, Cambridge University.

Kao, C. and M.H. Chiang, 1998, On the estimation and inference of a cointegrated regression in panel data, manuscript, Center for Policy Research, Syracuse University, January.

Keane, M.P. and Runkle, D.E. (1992), “On the Estimation of Panel Data Models with Serial Correlation when Instruements are not Strictly Exogenous” plus “Comments”, *Journal of Business and Economic Statistics*, 10, 1-29.

Kiefer, N.M. (1980): “Estimation of Fixed Effect Models for Time Series of Cross-Sections with Arbitrary Intertemporal Covariance”, *Journal of Econometrics*, 14, 195-202.

Kiviet, J.F. (1995), “On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Heterogeneous Panels”, *Journal of Econometrics*, Vol. 68, 1, pp. 53-78.

Levin,A. and Lin, C.F. (1992), “Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties”, Department of Economics, University of California at San Diego, D.P. No.92-93 (revised 1993).

Levin,A. and Lin, C.F. (1993), “Unit Root Tests in Panel Data: New Results”, Department of Economics, University of California at San Diego, D.P.No. 92-93,

Maddala, G.S. and Wu,S. (1999), “A Comparative Study of Unit Root Tests with Panel data and a New Simple Test”, forthcoming in *Oxford Bulletin of Economics and Stastistics* (November 1999).

Matyas, L. and Sevestre, P. (eds) (1996) *The Econometrics of Panel Data*, Handbook of Theory and Applications, Kluwer Academic Publisher (Second edition).

McCoskey, S. and C. Kao, 1998a, A residual based test of the null of cointegration in panel data, *Econometric Reviews*, 17(1), 57-84.

McCoskey, S. and C.Kao, 1998b, A Monte Carlo comparison of tests for cointegration in panel data, manuscript, Center for Policy Research, Syracuse University, Syracuse, February.

Mudlak, Y. (1978): “On the Pooling of Time Series and Cross Section Data”, *Econometrica*, 46, 69-85.

Nerlove, M. (1971): “Further Evidence on the Estimation of Dynamic Relations from a Time Series of Cross Section”, *Econometrica*, 39, pp. 359-382.

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

Pedroni, P., 1997, Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis - new results, manuscript, Indiana University, April.

Pedroni, P., 1998, Approximate critical values for cointegration tests in heterogeneous panels with multiple regressors, manuscript, Indiana University, February.

Pesaran,M.H. and Smith,R.(1995), "Estimation of Long-run Relationships from Dynamic Heterogeneous Panels", *Journal of Econometrics*, 68, 79-113.

Phillips, P.C.B. and Ouliaris,S. (1990), “Asymptotic Properties of Residual Based Tests for Cointegration”, *Econometrica*, 58,165-193.

Quah, D. (1994), “Exploiting Cross-Section Variations for Unit Root Inference in Dyanamic Data”, *Economics Letters*, 44, 9-19.

Robertson, D. and J. Symons (1991): “Some Strange Properties of Panel Data Estimators”, *Journal of Applied Econometrics*, 7, 175-189.

Robinson, P.M. (1987): “Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form”, *Econometrica*, 55, 875-891.

Sargan, J.D. (1958): “The Estimation of Economic Relationships Using Instrumental Variables”, *Econometrica*, 26, 393-415.

Sargan, J.D. (1980): “Some Tests of Dynamic Specification for a Single Equation”, *Econometrica*, 48, 879-897.

Sargan, J.D. (1988a): “Testing for Misspecification after Estimating Using Instrumental Variables”, in E. Maasoumi, ed., *Contributions to Econometrics: John Denis Sargan*, vol. 1. Cambridge: Cambridge University Press.

Sargan, J.D. (1988b): *Lectures on Advanced Econometric Theory*. Blackwell, Oxford.

Sevestre P. and A. Trognon (1990): “Estimation Methods for Dynamic Error Components Models: Small and Large Sample Properties”, Paper prepared for the Conference on “The Econometrics of Panels and Pseudo Panels”, Venice, October 25-27, 1990.

Schmidt, P., Ahn, S.C. and Wyhoski, D. (1992), Comment, *Journal of Business and Economic Statistics*, 10, pp. 10-14.

Shin, Y. (1994), “A Residual-Based Test of the Null of Cointegration Against the Alternative of No Cointehration”, *Econometric Theory*, 19, 91-115.

Urga, G. (1991): *Employment and Investment Functions for Italy - Using Individual Firm Data*, D. Phil. thesis, University of Oxford.

Urga, G. (1992): “The Econometrics of Panel Data: a Selective Introduction”, DP. 151, Institute of Economics and Statistics, Oxford, and also in *Ricerche Economiche*.

White, H. (1982): “Instrumental Variables Regression with Independent Observations”, *Econometrica*, 50, 483-499.

Zellner,A. (1962), “On the Aggregation Problem: A New Approach to a Troublesome Problem” in K.A. Fox et al (eds), Economic Models, Estimation and Risk Programming: Essays in Honour of Gerhard Tintner, Springer-Verlag, Berlin.