Estimation and inference with non-stationary panel time-series data.

Ron P. Smith
Department of Economics, Birkbeck College, London
email: R.Smith@bbk.ac.uk

September 28, 2001

Abstract

The econometric theory for panel data regressions was largely developed for survey data where N the number of individuals was large and T the number of time periods small. The asymptotic statistical theory was derived by letting $N \to \infty$ for fixed T. In recent years there has been growing interest in cases, such as sets of countries, regions or industries, where there are fairly long time-series for a large number of groups. These large T, large N, panel time-series raise a number of issues. First, since time-series tend to be non-stationary, determining the order of integration and cointegration becomes important. Second, since it is possible to estimate a separate regression for each group, it is natural to think of heterogeneous panels where parameters differ over groups. One can then test for parameter homogeneity rather than having to assume it as one is forced to do in small T panels. Third, one needs to determine the asymptotic properties of standard panel estimators when the data are non-stationary. Fourth, there is an issue of how to do the asymptotic analysis as both N and T can go to infinity. These issues have generated a large amount of recent econometric work and this paper provides an introductory survey for applied workers.

Acknowledgement 1 *I am grateful to the ESRC for support under grant L138251003. Prepared for the RC33 Conference, Cologne October 2000.*

1 Introduction

The econometric theory for panel data (see Hsiao (1986), Baltagi (1995) and Matyas and Sevestre (1996) for surveys) was largely developed for survey data where T, the number of time-series observations, was small (often only four or five observations) but N the number of groups or individuals was large. The asymptotic statistical theory was derived by letting $N \to \infty$, for fixed T, in constrast to time-series analysis which was done letting $T \to \infty$, for fixed N. In recent years, interest has grown in cases, such as sets of countries, industries or regions, where there are fairly long time series for a large number of groups. These large N, large T panel time-series raise a number of issues.

First, since time-series data tend to be non-stationary, determining the order of integration or cointegration of the variables becomes important. The order of integration is the number of times a time-series must be differenced to make it stationary. Many economic time-series appear to be integrated of order one, I(1), needing to be differenced once to make them stationary. They are said to exhibit a unit root. However, it may be the case that equilibrium or arbitrage conditions imply that certain linear combinations of the variables are stationary, I(0). If this is the case the variables are said to be cointegrated. For instance, the logarithms of the exchange rate, e_t, domestic and foreign prices, p_t and p_t^* may all be I(1), but Purchasing Power Parity predicts that the real exchange rate, $r_t = e_t - p_t + p_t^*$, should be I(0). Extending the estimation and testing procedures for integrated and cointegrated series to panels is thus a natural development. In fact, the tendency of individual time series tend to reject Purchasing Power Parity (not reject a unit root in the real exchange rate) has led the emphasis to switch to testing PPP in panels.

Secondly, since it is possible to estimate a separate regression for each group, which is not possible in the small T case, it is natural to think of heterogeneous panels where the parameters can differ over groups. One can then test for equality of the parameters, rather than having to assume it, as one is forced to do in the small T case. When equality of parameters over groups is tested it is very often rejected and the differences in the estimates between groups can be large. Baltagi and Griffin (1997) discuss the dispersion in OECD country estimates for gasoline demand functions, Boyd and Smith (2000) review possible explanations for the large dispersion in models of the monetary transmission mechanism for 57 developing countries.

Third one needs to determine the asymptotic properties of standard panel estimators when the data are non-stationary. These properties are rather different from those of single time-series, in particular spurious regression seems to be less of a problem.

Fourth, there is a question about how to do the asymptotic analysis as both N and T can go to infinity. There are a number of different ways that N and T can go to infinity and the relation between these different ways remains a subject of research: Phillips and Moon (1999, 2000).

This paper provides an introductory survey of some of the issues involved. Section 2 provides a brief introduction to the time-series issues. Section 3 considers panel estimators for I(0), stationary, data. Section 4 considers the properties of panel estimators for I(1) data. Section 5 discusses testing for unit roots and cointegration in panels. Section 6 concludes. Some caveats are in order. Firstly, this is an area that is developing rapidly, so any survey will quickly become dated. Secondly, there are a range of difficult technical issues (e.g. the rather restrictive assumptions necessary to get central limit theorems in this area), which I largely ignore in this paper and the results I summarise are dependent on particular assumptions that I do not fully set out. Thirdly, I have largely ignored Bayesian estimators, though in many ways a Bayesian approach is more natural in this context since it is usual to treat parameters as random. The text-book by Maddala and Kim (1998) provides an excellent discussion of some of the issues, which I summarise very briefly. Banerjee (1999) provides a more technical survey of these issues in his introduction to the November 1999 special issue of the Oxford Bulletin of Economics and Statistics, on Testing for Unit Roots and Cointegration using Panel Data. On the more general issues on cointegration and unit roots see the special issue of the Journal of Economic Surveys, December 1998 on Practical Issues in Cointegration Analysis, reprinted as Oxley and McAleer (1999).

2 Time Series

Suppose we have data on a dependent variable y_{it} and a strictly exogenous independent variable x_{it}, for groups $i = 1, 2, ..., N$ and time periods $t = 1, 2, ..., T$, and we are interested in estimating a dynamic regression relationship between them. The central issue in panel analysis is the extent to which the the processes that generate the data in different groups share common features. To begin, assume that they share no common features, so that the data for each group are appropriately analysed as separate time series.

For many purposes, it is convenient to describe a time-series by its order of integration, the number of times it must be differenced to make it stationary. A variable is said to be (weakly or covariance) stationary if after removal of deterministic elements (e.g. a linear trend or seasonal dummy variables) its mean, variance and covariances with its lagged values are not a function of time. A stationary series is said to be integrated of order zero, I(0), because it does not need to be differenced at all. One example of an I(0) process is a white noise error, ε_{it} where $E(\varepsilon_{it}) = 0$, $E(\varepsilon_{it}^2) = \sigma_i^2$; $E(\varepsilon_{it}\varepsilon_{it-p}) = 0$, $p \neq 0$. Another example is the first order trend stationary process:

$$y_{it} = a_i + \rho_i y_{i,t-1} + c_i t + \varepsilon_{it} \tag{1}$$

where $-1 < \rho_i < 1$. Trend stationary processes were thought to be good descriptions of many economic variables such as the logarithm of real GDP. Substituting for lagged y_t this can be written

$$y_{it} = \frac{a_i}{1-\rho_i} + \sum_{p=0}^{\infty} \rho_i^p \varepsilon_{i,t-p} + \frac{c_i}{1-\rho_i} t, \tag{2}$$

so the effects of shocks die out.

A variable is said to be integrated of order one, I(1), if it is stationary after differencing once, i.e. $\Delta y_{it} = y_{it} - y_{it-1}$ is stationary. An example of an I(1) process is the random walk with drift:

$$y_{it} = \alpha_i + y_{i,t-1} + \varepsilon_{it}. \tag{3}$$

This is difference stationary, since

$$\Delta y_{it} = \alpha_i + \varepsilon_{it} \tag{4}$$

is stationary. If we back substitute for lagged y_{it} in (3), it can be written:

$$y_{it} = \alpha_i t + \sum_{p=0}^{t} \varepsilon_{i,t-p} + y_{i0}, \tag{5}$$

the sum of a deterministic trend, a stochastic trend (the sum of past errors) and an initial condition. Notice that in I(1) series shocks are permanent, the effects persist for ever, unlike the trend stationary process (1) where the effects of shocks are transitory.

To distinguish between the trend stationary and difference stationary case, write (1) as

$$\Delta y_{it} = a_i + (\rho_i - 1)y_{i,t-1} + (\rho_i - 1)\gamma_i t + \varepsilon_{it}, \tag{6}$$

where $c_i = (\rho_i - 1)\gamma_i$. Then if $\rho_i = 1$ this reduces to (3). One way of testing the null hypothesis of what is called a unit root $H_0 : \rho_i = 1$ is to use the t ratio for the coefficient of the lagged dependent variable in (6). However, under H_0 the regression involves an I(0) variable on the left hand side and an I(1) variable on the right hand side, so the t ratio does not have a standard t distribution. The actual distribution was tabulated by Dickey and Fuller and the 5% critical value is about -3.4. This test is known as the Dickey-Fuller test. The regression can be augmented by lagged changes in Δy_{it} to ensure that the error does not suffer serial correlation and the test based on the t ratio of the lagged level of the dependent variable in this augmented regression is known as the Augmented Dickey Fuller (ADF) test. An alternative is to remove the serial correlation non-parametrically as in the Phillips-Perron test. There are also tests where the null hypothesis is that the series is I(0) and the KPSS test, Kwiatkowski et al. (1992), examines whether the variance of the Random Walk component of the series is greater than zero.

There are a wide range of tests, see Phillips and Xiao (1998) for an introduction, and some seem to be definitely better, e.g. the quasi differencing variant of Elliott et al. (1996). However, ADF tests are widely used and we will focus on them. All these tests have very low power (i.e. a low probability of rejecting the I(1), unit root, null) when in fact the process is stationary with a coefficient close to unity. In addition, inference is sensitive to treatment of possible serial correlation in the disturbance (e.g. the number of lags included in the ADF test); the presence of shifts in mean or trend; and the span of the data, long time-series tend to look I(0), short ones I(1).

One reason the order of integration is interesting is that if y_{it} and x_{it} are unrelated independent random walks, in a regression of y_{it} on x_{it}, say

$$y_{it} = \theta_{0i} + \theta_i x_{it} + v_{it}, \tag{7}$$

the least squares estimate of the regression coefficient, $\widehat{\theta}_i$, and its t ratio will not go to zero as they should but to non-zero random variables and the R^2

of the regression will go to unity and the Durbin-Watson statistic to zero as $T \to \infty$. This problem of spurious regression was pointed out by Granger and Newbold (1974) and Phillips (1986) provides a theoretical treatment. In general, if y_{it} and x_{it} are I(1), any linear combination, say $u_{it} = y_{it} - \theta_i x_{it}$ will also be I(1). Define $\widetilde{y}_{it} = y_{it} - \overline{y}_i$ where $\overline{y}_i = \sum_{i=1}^{T} y_{it}/T$, the mean for each group, and similarly for $\widetilde{x}_{it}$, then

$$\widehat{\theta}_i = \frac{\sum_{t=1}^{T} \widetilde{x}_{it}\widetilde{y}_{it}}{\sum_{t=1}^{T} \widetilde{x}_{it}^2} = \theta_i + \frac{\sum_{t=1}^{T} \widetilde{x}_{it}u_{it}}{\sum_{t=1}^{T} \widetilde{x}_{it}^2}$$

What happens in spurious regression is that the noise in the least squares estimate $\widehat{\theta}_i$ caused by the covariance of two I(1) random variables, x_{it} and u_{it} (which does not go to zero as $T \to \infty$, even when suitably scaled), swamps the signal, the true value of θ_i, which in the case of a spurious regression is zero.

If it happens that there is a linear combination of the I(1) variables, $y_{it} - \theta_i x_{it}$, which is in fact I(0), then the two variables are said to be cointegrated, Engle and Granger (1987). Cointegration of the two variables is the condition required for a regression of y_{it} on x_{it} not to be spurious. The error u_{it} is then I(0) and the noise does not swamp the signal. When there are more than two variables, there is the possibility of more than one cointegrating vector.

For exposition, we will work with a simple Partial Adjustment Model (PAM), although many of the arguments extend to more complicated dynamic models. The partial adjustment model is:

$$y_{it} = \alpha_i + \beta_i x_{it} + \gamma_i y_{it-1} + u_{it}; \tag{8}$$

for each group. Assume u_{it} is distributed $N(0, \sigma_i^2)$ and is independent over time and across groups. If the process is stationary, $-1 < \gamma_i < 1$, there is a long run solution, which specifies the equilibrium value of y_t, of the form:

$$y_{it}^* = \frac{\alpha_i}{1-\gamma_i} + \frac{\beta_i}{1-\gamma_i} x_{it} = \theta_{0i} + \theta_i x_{it}.$$

The adjustment process is:

$$\Delta y_{it} = \lambda_i (y_{it}^* - y_{it-1}) + u_{it}, \tag{9}$$

where λ_i is interpreted as a speed of adjustment, the proportion of the difference between the equilibrium and the previous actual value that is made up in each period. In terms of the original variables, this is:

$$\Delta y_{it} = a_i + \beta_i x_{it} + \lambda_i y_{it-1} + u_{it}, \tag{10}$$

where $\alpha_i = \lambda_i \theta_{oi}$; $\beta_i = \lambda_i \theta_i$; $\lambda_i = \gamma_i - 1$. Define the least squares estimators of the parameters of (10) as $\widehat{\alpha}_i$, $\widehat{\beta}_i$, and $\widehat{\lambda}_i$. The estimates of the long run parameters are $\widetilde{\theta}_i = \widehat{\beta}_i/(1-\widehat{\gamma}_i) = -\widehat{\beta}_i/\widehat{\lambda}_i$ and $\widetilde{\theta}_{0i} = -\widehat{\alpha}_i/\widehat{\lambda}_i$.

If the variables are I(0), as $T \to \infty$ least squares provides $\sqrt{T}$ consistent estimates for α_i, β_i, γ_i (λ_i) and θ_i. These estimators will be biased for small T, because of the presence of the lagged dependent variable; in particular $\widehat{\gamma}_i$ will be biased downwards, towards zero. In the I(1) case, where y_{it} and x_{it} cointegrate, then $\widetilde{\theta}_i$ is a superconsistent (converging to its true value at rate T rather than $\sqrt{T}$) estimator of θ_i, and the short run parameters are $\sqrt{T}$ consistent. Cointegration implies that β_i and λ_i are not equal to zero and this provides a test for cointegration, in the case where there is a single cointegrating vector; see Pesaran, Shin and R. J. Smith (1999).

We could also run the static regression (7), with least squares estimators $\widehat{\theta}_{0i}$ and $\widehat{\theta}_i$. In the I(0) case, (7) suffers omitted variable bias and $\widehat{\theta}_{0i}$ and $\widehat{\theta}_i$ are not consistent for θ_{0i} and θ_i. However in the I(1) case, $\widehat{\theta}_i$ is T consistent and this is the regression that Engle and Granger (1987) suggest using to estimate a single cointegrating vector. Engle and Granger also suggest testing for cointegration using a ADF type test on the estimated residuals from (7), though the critical values for the ADF on estimated residuals are different from those for an ADF on a variable. However, the coefficients of this regression are likely to be badly biased in small samples and other methods have tended to be adopted. In the case of a single cointegrating vector direct estimation of the dynamic equation like (10) seems to have good properties and to be robust to whether the series are I(0) or I(1). The dynamic regression can also be used to test whether a long-run relationship exists, cointegration in the case of I(1) variables.

When there is more than one cointegrating vector, the situation is rather different. Suppose that we have an $mx1$ vector of variables for each country Y_{it} generated by a first order Vector Autoregression (VAR) with trend

$$Y_{it} = \mu_{0i} + \mu_{1i} t + A_{1i} Y_{i,t-1} + \varepsilon_{it} \tag{11}$$

where A_{1i} is an $m \times m$ matrix. Notice that this is just a vector version of (1). This can be written in Vector Error Correction Model (VECM) form corresponding to (6):

$$\Delta Y_{it} = \mu_{0i} + \mu_{1i}t + \Pi_i Y_{i,t-1} + \varepsilon_{it} \tag{12}$$

where $\Pi_i = (A_{1i} - I)$. If the elements of Y_i are all $I(0)$, then Π_i is a full rank $m \times m$ matrix. If the elements of Y_i are all $I(1)$, and not cointegrated then $\Pi_i = 0$, a pure first difference model is appropriate. If there are r linear combinations of Y_i that are $I(0)$, say $Z_{it} = \beta' Y_{it}$,where Z_{it} is $r \times 1$ and β is $m \times r$ then Π_i has rank r and can be written $\Pi_i = \alpha\beta'$. Johansen (1988) provides a widely used maximum likelihood procedure that integrates testing for the number of cointegrating vectors, determining r, and (subject to indentifying conditions) estimating the cointegrating vectors, β, and the feedback coefficients α. As with the ADF test, lagged changes in ΔY_{it} can be added to the right hand side of (12). There are also a number of different ways that the coefficients of the deterministic variables, μ_0 and μ_1 can be treated, including restricting the trend as in (6). Pesaran and Smith (1998) discuss the issues.

3 Pooling with I(0) variables

Panel estimators use the fact that if there are similarities or links between the processes generating the data in the different groups, combining the data can improve the efficiency of the estimation of the parameters. In addition, panel data allows one to answer questions that cannot be answered with time-series data alone (e.g. the effect of variables that are relatively constant over time within countries, such as institutional structures) or with cross-section data alone (e.g. the patterns of adjustment to change). In the large T small N case, the links are often assumed to be through the disturbances, with $E(u_{it}u_{jt}) = \sigma_{ij} \neq 0$. Then the covariance matrix can be estimated and the Zellner (1962) Seemingly Unrelated Regression Equation (SURE) estimator is used. If $N > T$ the estimated covariance matrix is rank deficient and SURE estimation is not feasible unless some structure is imposed on the covariance matrix. Robertson and Symons (1999) suggest using a factor analytic structure and this seeems to result in substantial gains in efficiency. In other cases one may be able to specify a spatial structure for the covariances between groups. We will not pursue the covariance links.

A second approach is to regard the parameters as random, drawn from some probability distribution with a finite number of parameters. In the

PAM case we will assume a mixed fixed-random structure treating the α_i as fixed parameters and the slopes as random. Notice this contrasts with the Random Effects model which treats the slopes as fixed and homogenous and the α_i as random. Define $W_{it} = (x_{it}, y_{it})'$ and $\delta_i = (\beta_i, \lambda_i)'$ then stacking the observations for each group (8) can be written:

$$y_i = \alpha_i + W_i\delta_i + u_i. \tag{13}$$

Then it is assumed that $\delta_i = \delta + \eta_i$, $E(\eta_i) = 0$, $E(\eta_i\eta_j') = \quad , i = j$, $E(\eta_i\eta_j') = 0$ otherwise and that the η_i are independent of W_i. The crucial assumption here is the independence of the randomly varying parameters from the regressors. See Pesaran, Haque and Sharma (1999) for a case where this assumption fails producing misleading inferences in pooled regressions. There are a large number of estimators for δ, the expected value of the random coefficients. The simplest is to take the least squares estimates of (13) for each group, say $\widehat{\delta}_i$ and compute the average $\overline{\delta} = \sum_i \widehat{\delta}_i/N$, estimating by $\widehat{\ } = \sum_i(\widehat{\delta}_i - \overline{\delta})(\widehat{\delta}_i - \overline{\delta})'/N$. Pesaran and Smith (1995) call this the Mean Group Estimator. Swamy (1970) suggests a weighted average of the $\widehat{\delta}_i$, which can be interpreted either as a Generalised Least Squares estimator or an empirical Bayes estimator. Hsiao Pesaran and Tahmiscioglu (1999) review a variety of Bayes and empirical Bayes estimators of this sort and show that the Mean Group estimator is asymptotically normal for large N and T as long as $\sqrt{N}/T \to 0$ as both N and $T \to \infty$, but is unlikely to perform well when N or T are small. In particular, the Mean Group Estimator is very sensitive to outliers which are a common feature of the group specific estimates in many applications.

A third approach is to assume some degree of homogeneity among the parameters. The most common panel estimator for dynamic models takes the model (8) and imposes the restriction that the slope coefficients and variances are identical across groups, allowing the intercepts to differ. This gives

$$y_{it} = \alpha_i + \beta x_{it} + \gamma y_{it-1} + w_{it}. \tag{14}$$

This estimator goes under a variety of names, including: Least Squares Dummy Variables (because it can be implemented by applying least squares to the pooled data, including a dummy variable for each group); the Within estimator (to distinguish it from the Between estimator which uses a regression of the means for each group); the analysis of Covariance estimator; and the Fixed Effect estimator (to distinguish it from the Random Effect estimator). We will refer to it as the Fixed Effect estimator. There is also a two way fixed effect estimator, which allows for time effects:

$$y_{it} = \alpha_i + \alpha_t + \beta x_{it} + \gamma y_{it-1} + w_{it}. \tag{15}$$

This estimator is usually implemented by using deviations from group and period means for all the variables.

Let us consider the properties of the one way Fixed Effect estimators in (14), say $\widehat{\beta}_{FE}$ and $\widehat{\gamma}_{FE}$. If $\beta_i = \beta$ and $\gamma_i = \gamma$ the fixed effect estimators are consistent as $T \rightarrow \infty$, for fixed N; though they are biased in small samples, because of the lagged dependent variable bias. However, they are inconsistent as $N \rightarrow \infty$ for fixed T. The latter is the result of the fact that the lagged dependent variable bias arising from the initial conditions, is not removed by increasing N. There is a large literature dealing with dynamic models in the small T case. The problems arise from the treatment of the initial conditions and the incidental parameters (the number of intercepts α_i increases with N). There is a further difficulty that the traditional GMM estimator breaks down if the variables are I(1). Binder, Hsiao and Pesaran (1999) suggest a maximum likelihood estimator that is consistent whether the variables are I(0) or I(1).

If $\beta_i \neq \beta$, or $\gamma_i \neq \gamma$, i.e. the true model is (8) the fixed effect estimator is an inconsistent estimator of β and γ, the expected values of β_i and γ_i. This inconsistency in heterogenous dynamic panel models was first noted by Robertson and Symonds (1992) and is analysed in detail in Pesaran and Smith (1995) and Pesaran, Smith and Im (1996). It arises because under heterogeneity if the true model is (8) the error in (14) takes the form:

$$w_{it} = (\beta_i - \beta)x_{it} + (\gamma_i - \gamma)y_{it-1} + u_{it}.$$

This composite disturbance will be serially correlated, if x_{it} is serially correlated, as it usually is, and will not be independent of the lagged dependent variable. This heterogeneity bias, which depends on the serial correlation in the x and the variance of the random parameters, can be quite severe. Suppose x_{it} is generated by an AR1 process:

$$x_{it} = \mu_i(1 - \rho) + \rho x_{it-1} + \varepsilon_{it}$$

where μ_i is the unconditional mean. Then as x_{it} tends towards being an I(1) variable, i.e. as $\rho \rightarrow 1$, the Probability Limits of the fixed effects estimator (taken by first letting $T \to \infty$. then letting $N \to \infty$) are given by:

$$\underset{\rho \to 1}{P\lim} \, (\widehat{\beta}_{FE}) = 0; \quad \underset{\rho \to 1}{P\lim} \, (\widehat{\gamma}_{FE}) = 1$$

irrespective of the true values of β and γ, the expected values of β_i and γ_i. Notice that these results are not valid for $\rho = 1$, the I(1) case considered in the next section. The asymptotic bias in the estimator of the long-run coefficient $\widehat{\theta}_{FE} = \widehat{\beta}_{FE}/(1 - \widehat{\gamma}_{FE})$ is not as severe, because the biases in the numerator and denominator tend to cancel out. Notice that if ρ is positive, the usual case, the bias in $\widehat{\gamma}_{FE}$ is upwards, the opposite of the lagged dependent variable bias, so the two biases may offset each other at certain sample sizes. The bias cannot be dealt with by traditional instrumental variable estimators, but the random coefficient estimators discussed at the beginning of this section provide consistent estimators. This analysis has been extended to stationary Vector Autoregressions by Rebucci (2000).

The heterogeneity issue is important since when the hypothesis that the slope coefficients are in fact identical is tested, it is almost always rejected. Pesaran, Shin and R.P. Smith (1999) argue that it is more likely that long run effects are homogenous and suggest a pooled mean group estimator that constrains the long-run coefficients to be the same while allowing the short-run coefficients and variances differ over groups. In the applied examples they present, consumption functions and energy demand functions, the hypothesis that the long-run coefficients are identical is also rejected at conventional significance levels. However, in large samples conventional significance levels may not be appropriate. With sufficiently large samples and fixed significance level every hypothesis will be rejected. An alternative to testing is to use a model selection criterion to decide whether or not to pool. One possibility which does correct for the effect of sample size is to choose the model with the largest value of the Schwarz Bayesian Criteria. This is $SBC = MLL_i - 0.5k_i \ln(NT)$, where MLL_i is the maximised log-likelihood for model i, k_i the number of parameters in model i. Whether N and T should be treated symmetrically in panel applications, as the SBC does, is an open question. For large samples the SBC penalises over-parameterization more heavily than tests at conventional significance levels and is thus more likely to choose the pooled model. The decision whether to pool or not will depend both on the degree of heterogeneity and on the purpose of the exercise. Baltagi and Griffin (1997) use a large number of different estimators to estimate demand for gasoline in OECD countries. They find that there is a very large degree of heterogeneity in the individual country estimates but that for the purpose of medium term forecasting simple pooled estimators do well. Attanasio et al. (2000) use a variety of estimators to measure the dynamic interaction between savings, investment and growth.

4 Pooling with I(1) Variables

This section is largely based on Phillips and Moon (1999, 2000) and Kao (1999) and research in this area is proceeding rapidly. When an estimator is said to be consistent ($T \to \infty$, $N \to \infty$), this usually means first $T \to \infty$, then $N \to \infty$. However, some results can be obtained having them go to infinity simultaneously with $N/T \to 0$. These results are based on a set of assumptions which include that there are no correlations between groups and that the coefficients of the moving average representations of the data are distributed randomly. Define $Z_{it} = (y_{it}, x_{it})'$ and assume that the data are generated by:

$$Z_{it} = Z_{it-1} + U_{it}$$

where

$$U_{it} = \sum_{s=1}^{\infty} C_{is} V_{i,t-s}$$

and V_{it} are white noise independent across i and t, C_{is} are a set of random matrices independently distributed across i, and C_{is} and V_{it} are independent. Let $C_i(1) = \sum_{s=0}^{\infty} C_{is}$. The long-run covariance matrix of Z is given by

$$C_i(1)C_i(1)' = {}_i = \begin{bmatrix} {}^{i}_{yy} & {}^{i}_{yx} \\ {}^{i}_{xy} & {}^{i}_{xx} \end{bmatrix},$$

where the elements of the matrix would be scalars in the bivariate case. The long-run average covariance matrix of Z is given by

$$E({}_i) = \quad = \begin{bmatrix} {}_{yy} & {}_{yx} \\ {}_{xy} & {}_{xx} \end{bmatrix}.$$

Consider the fixed effects pooled regression

$$y_{it} = \theta_{0i} + \theta x_{it} + v_{it} \tag{16}$$

where y_{it} and x_{it} are I(1), with least squares estimator $\widehat{\theta}$. Define $\widetilde{y}_{it} = y_{it} - \overline{y}_i$ where $\overline{y}_i = \sum_{i=1}^{T} y_{it}/T$, the mean for each group, and similarly for $\widetilde{x}_{it}$, and define

$$\widehat{\theta} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \widetilde{x}_{it} \widetilde{y}_{it}}{\sum_{i=1}^{N} \sum_{t=1}^{T} \widetilde{x}_{it}^2} = \theta + \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \widetilde{x}_{it} v_{it}}{\sum_{i=1}^{N} \sum_{t=1}^{T} \widetilde{x}_{it}^2}.$$

Consider a number of cases. First, suppose v_{it} is I(1) and y_{it} and x_{it} are independent random walks, so that the true value of θ is zero. In a single time-series this is the spurious regression case and $\widehat{\theta}_i$ will have a non-zero probability limit as $T \to \infty$, because the noise, $\sum_{t=1}^{T} \widetilde{x}_{it} v_{it}$, swamps the signal. However, in the pooled estimator we are also averaging over i, and this attenuates the noise and allows us to obtain a consistent estimate of the parameter. Thus in the case of a spurious regression, we get a consistent ($T \to \infty$, $N \to \infty$) estimate of the true value of the parameter, zero, Kao (1999). By using pooled data we can avoid the problem of spurious regression. The conventional standard errors of $\widehat{\theta}$ will be wrong, and the standard t statistics will diverge from zero, but correct standard errors can be calculated. Pesaran and Smith (1995) noted that the problem of spurious regression does not arise in a cross-section regression of the form

$$\overline{y}_i = \theta_0 + \theta \overline{x}_i + v_i$$

even if x_{it} and y_{it} contain a unit root, under the strong assumptions of random parameters and strictly exogenous x_{it}. The pooled estimator will be more efficient than the cross-section estimator.

Secondly, suppose y_{it} and x_{it} are not cointegrated, but they are related. This could be either through a homogenous θ or a heterogenous θ_i :

$$y_{it} = \theta_{0i} + \theta_i x_{it} + u_{it},$$

where u_{it} is $I(1)$. Again the averaging over i will attenuate the noise in the relationship and $\widehat{\theta}$, the pooled estimator, will consistently estimate that relationship. It is important to be clear what is being estimated here. $\widehat{\theta}$ provides a consistent estimate of the ratio of the average across groups of the long-run covariance between x and y to the average across groups of the long-run variance of the x,: $\theta^A = {}_{yx}({}_{xx})^{-1}$. Phillips and Moon (1999) call θ^A the 'long-run average regression coefficient'. Notice that in general, this will be different from the 'average long-run regression coefficient' $E(\theta_i)$ estimated by the average across groups of $\widehat{\theta}_i = \ ^{i}_{yx}(\ ^{i}_{xx})^{-1}$, since the expected value of a ratio is not equal to the ratio of the expected values.

The two estimators use the cross-section within-group variation differently. There are some cases where the 'long-run average' and 'average long-run' coefficients coincide, e.g. if the variance of x is the same for each group. The relation between the two estimators is discussed further below. Coakley, Fuertes and Smith (2000) present Monte Carlo evidence of the performance of various panel estimators when the error term is I(1), i.e. there is no cointegration, and show that the variance of the estimators falls with $\sqrt{N}$.

Thirdly suppose the model is again

$$y_{it} = \theta_{0i} + \theta_i x_{it} + u_{it},$$

but now u_{it} is I(0), there is cointegration within each group. The pooled estimator $\widehat{\theta}$ again provides a consistent estimator of the 'long-run average regression coefficient' θ^A. Notice that the error in the pooled regression, (16) is $v_{it} = u_{it} + (\theta_i - \theta)x_{it}$ which will be I(1) if $\theta_i \neq \theta$ and x_{it} is I(1), but as in the spurious regression case averaging the variance and covariance across groups attenuates the noise. Again, in general, the 'long-run average' regression coefficient is different from the 'average long run coefficient', the expected value of the θ_i the estimated cointegrating coefficient. In contrast to the previous case, θ_i is consistently estimated as $T \to \infty$. To see the difference, note that the 'long-run average' regression coefficient can be written as a weighted average of the $\widehat{\theta}_i$:

$$\begin{aligned}
\widehat{\theta} &= \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{y}_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}^2} \\
&= \sum_{i=1}^{N}\frac{\sum_{t=1}^{T}\widetilde{x}_{it}^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}^2}\left(\frac{\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{y}_{it}}{\sum_{t=1}^{T}\widetilde{x}_{it}^2}\right) \\
&= \sum_{i=1}^{N} W_i\widehat{\theta}_i \\
&= \sum_{i=1}^{N} W_i\theta_i + \sum_{i=1}^{N} W_i\left\{\frac{\sum_{t=1}^{T}\widetilde{x}_{it}u_{it}}{\sum_{t=1}^{T}\widetilde{x}_{it}^2}\right\}
\end{aligned}$$

In the case where there is a cointegrating relationship, the component in braces, $\{\}$, in the second term goes to zero as as $T \to \infty$. If $\theta = E(\theta_i)$, then since $\sum_{i=1}^{N} W_i = 1$

$$E(\widehat{\theta}) = \theta^A = \theta + E(\sum_{i=1}^{N} W_i(\theta_i - \theta)).$$

Thus the expected value of $\widehat{\theta}$ will only be the same as the expected value of θ_i where the weights, the share of a groups variance in the total variance of x_{it}, $(\ {}^{i}_{xx} / \sum_i \ {}^{i}_{xx})$ are uncorrelated with the θ_i. Informally, one can think of this as independence of the θ_i from the x_{it}, though the conditions Phillips and Moon use to establish that the relevant limits exist involve different assumptions about which parameters are random.

Fourthly, suppose that there is a common θ, then the least squares estimator $\widehat{\theta}$ from (16) will consistently estimate it for large T and N. Phillips and Moon (1999) propose a Fully Modified (FM) pooled estimator of the long-run parameter for this case. This case is the same as that considered in Pesaran, Shin and R.P. Smith (1999) for cointegrating I(1) variables with common long-run parameters. The difference is that Pesaran, Shin and Smith have parameteric short-run dynamics, e.g. like (8) rather than relying on non-parametric methods to remove the effects of the dynamics (and endogeneity) in v_{it} on the long-run coefficients as Phillips and Moon do. In many cases the parameters of interest are those of the short run dynamic processes, e.g. speeds of adjustment or rates of convergence, and FM estimators do not provide estimates of those short-run parameters.

The sensitivity of estimates of dynamic parameters to estimation method is indicated by measures of the speed of convergence to the steady state growth path considered in Lee, Pesaran and Smith (1997) using data for 102 countries 1965-1989. One method is to regress growth rates over a period on the logarithm of initial income. This gives estimates of the speed of convergence of about 4% per annum. A second method is to pool the data and use a fixed effects estimator. This gives an estimate about 10%. A third method is to estimate (6) separately for each country. This gives an estimate of about 30% with a standard deviation of just under 20%. However, panel unit root tests cannot reject the hypothesis that the speed of convergence is in fact zero. The heterogeneity in growth rates is also very marked.

5 Testing for Unit Roots and Cointegration in Panels

Since the work of Quah (1994) and Breitung (1994) there has been a vast amount of work on testing for unit roots and cointegration in panels.When testing for unit roots or cointegration, there are two fundamental questions. First, should one test the null of a unit root (no cointegration) against a stationary alternative (cointegration) as for instance the ADF (Johansen) tests do or should one test a null of stationarity (cointegration) against an alternative of a unit root (no cointegration), as KPSS type tests do? Second should one use a parameteric correction for serial correlation as the ADF tests (Johansen estimators) do by including lagged changes in the dependent variable or a non-parameteric correction as the Phillips-Perron type tests (fully-modified estimators) do? Different answers to these questions give rise to different tests and the time-series literature has not produced any clear conclusion as to what the best answer to these questions is. The panel unit root and cointegration tests add two further questions: how much heterogeneity does one allows between groups? and how does one combines the statistics for different groups? It should be emphasised that there are difficult technical issues in deriving the properties of these tests. The papers setting out the standard tests, which have been used in a very large number of applications, are Levin and Lin (1993) and Im, Pesaran and Shin (1997), the orginal version of which dates from 1995. Neither of these papers have been published, because of technical difficulties with the proofs. Harris and Tzavalis (1999a) is a published paper that does derive results, but only on the assumption of no serial correlation. They consider the case of heteroskedastic and serially correlated, MA1, errors in Harris and Tzavalis (1999b), for T fixed and $N \to \infty$. Mosts of the tests discussed below assume that T is large. Binder, Hsiao and Pesaran (1999) examine the small T case. They provide tests for unit roots and cointegrating rank for a dynamic fixed effects model (i.e. with slope homogeneity) using a maximum likelihood estimator which is consistent for large N whether the underlying series are trend stationary or I(1).

5.1 Unit Roots

Consider the Dickey Fuller regression (6) above:

$$\Delta y_{it} = a_i + (\rho_i - 1)y_{i,t-1} + (\rho_i - 1)\gamma_i t + \varepsilon_{it},$$

or

$$\Delta y_{it} = a_i + b_i y_{i,t-1} + c_i t + \varepsilon_{it} \tag{17}$$

where lagged values of Δy_{it} may also be included to ensure that the disturbances are not serially correlated. The unit root hypothesis, $\rho_i = 1$ implies that $b_i = 0$, for all i. Given T sufficiently large, this can be tested on each group using the t ratio for b_i and the non-standard critical values. But as noted above such tests lack power. The hope of panel unit-root tests is that power can be increased by increasing the sample through use of the cross-section dimension.

Levin and Lin (1993) consider a two way fixed effect version of this model:

$$\Delta y_{it} = a_i + \alpha_t + b y_{i,t-1} + c_i t + \varepsilon_{it} \tag{18}$$

and devise a test for the null $H_0 : b = 0$, against the alternative $b < 0$. They also allow for augmenting by lagged changes. The assumption, under the alternative, of homogeneity, that $b_i = b$ for all groups, is clearly restrictive and subject to the possible heterogeneity bias of the fixed effect estimator discussed in the previous section. Im Pesaran and Shin (1997), IPS, allow the b_i to differ; use the estimates of (17) directly (allowing for augmenting); calculate the average ADF statistics (t ratios for b_i) and provide simulated test statistics for the mean and variance of the average, which allows testing of the hypothesis $H_0 : b_i = 0$, for all i. They subtract the means of the data for each year to allow for time effects. Maddala and Wu (1999) agree that a heterogenous alternative is better but argue that averaging the ADF statistics is not the most efficient way to use the information. They propose a test statistic, based on a suggestion of R.A. Fisher, which is $-2\sum_{i=1}^{N}(\ln P_i)$ where P_i is the P value for the ith test. Under the null of no cointegration, this is distributed $\chi^2(2N)$. Their simulations suggest that in a variety of situations the Fisher test is more powerful than the IPS test which is more powerful than the Levin and Lin test.

Before one can apply the individual ADF tests one has to determine the degree of augmentation, the number of lags, and determine whether a trend should be included. This raises questions about the criterion that should be used and the possibility of pre-test biases. In a panel setting, one must determine whether to choose augmentation and trend on a group specific basis or use a common specification for each group. Similar issues arise in choosing window size in non-parameteric tests. None of these tests allow for covariances between groups, though the use of year effects or data measured as deviations from year means allows for a simple covariance structure. An alternative is to use the Robertson and Symons (1999) factor analytic structure discussed above. Maddala and Wu (1999) show that correlations between groups can induce serious size distortions for the tests and suggest using boostrapped tests. They also suggest that the Elliott et al. (1996) quasi difference variant of the ADF regression may be better than the ADF. Andersson and Lyhagen (1999) also investigate a long-memory panel data test.

While there is scope for improvements of the panel unit root tests, there are serious questions about the interpretation of such tests. The null hypothesis is that $b_i = 0$, for all i. This can be rejected (in sufficiently large samples) if any one of the N coefficients b_i is non-zero. Rejection of the null certainly does not indicate that all the series are stationary, though the way the alternative is written in the Levin-Lin variant may give that impression. If the hypothesis of interest is that all the series are stationary, then the appropriate test would be a panel variant of the KPSS test. But again rejection could reflect the fact that a single series was non-stationary, which may not be interesting. As always with pooling, some judgement on the commonality or degree of similarity between groups is needed. There is a further problem that is apparent in (18). Suppose $b < 0$, and the test correctly rejects the hypothesis $b = 0$. However, suppose α_t is $I(1)$, then the test will have rejected a unit root, whereas every series does in fact have a unit root arising from the common stochastic trend in the mean of the series, $\overline{y}_t$. A similar problem arises with the IPS test, where deviations from time-period means are used.

One of the reasons investigators were concerned about the presence of unit roots was to avoid the problem of spurious regressions. But in panels, where the spurious regression problem is reduced by averaging this motivation is less pressing. A similar point applies to testing for cointegration to which we now turn.

5.2 Cointegration

Where it is hypothesised that there is a single cointegrating vector, so Engle-Granger type procedures are appropriate, one can estimate by least squares either the heterogeneous individual levels equation (7) or the homogeneous pooled levels equation (16). Then panel unit root tests, with either heterogeneous or homogeneous autoregressive parameters, can be applied to the residuals from these regressions; with appropriate adjustments to the critical values. This is essentially the procedure suggested by Pedroni (1995, 1999). These test the null hypothesis of no cointegration. KPSS like variants, where the null is cointegration, rather than lack of cointegration, are given in McCoskey and Kao (1998) and Hadri (1999). Rather than using unit root tests on the residuals one could use Johansen tests for each group. Larsson, Lyhagen and Lothgren (1998) and Larsson and Lyhagen (1999) suggest using the average of the Johansen trace statistics. Rather than use IPS like averages one could also use Maddala and Wu like Fisher combinations over groups of the tests.

The interpretation of panel cointegration tests raises even more difficulties than the interpretation of panel unit root tests. For a vector of data Y_{it}, with r_i I(0) cointegrating relationships $Z_{it} = \beta_i' Y_{it}$, there are a variety of possible hypotheses. It could be that there are the same number of cointegrating vectors, $r_i = r$, in each group or that there are at least $r_{\min}$ cointegrating vectors or that there are identical cointegrating vectors, β in each group. As with panel unit root tests it is not clear how rejection should be interpreted. The pre-testing problems in cointegrating models are much more severe than in the unit root tests, where the issues are just determination of lag lengths and treatment of the deterministic elements. The large number of choices involved in cointegration analysis, even in the case of a single time-series, is discussed in Pesaran and Smith (1998). In addition, there are issues of identification. If there are m variables in the system and $0 < r < m$ cointegrating vectors, one needs r restrictions on each cointegrating vector to identify the estimates. If $r = 1$ this is just a normalisation restriction, specifying which variable is the dependent variable. The Engle Granger estimates are not invariant to normalisation (though the Johansen estimates are). In panels, there is the possibility that different normalisations would be appropriate for different groups. The normalisation assumption is not innocuous, it presumes that the variable has a non-zero coefficient in that particular cointegrating vector. With multiple cointegrating vectors one needs more just-identifying restrictions, and these are not testable. Again there is the issue as to whether the just identifying restrictions should be homogeneous across groups.

With cointegration, there is a further issue associated with dependence between groups, discussed by Banerjee et al. (2000). Normally, in time series one considers cointegration within a group, i.e. investigates whether there is a linear combination of the data for a particular group $Z_{it} = \beta_i' Y_{it}$ which is $I(0)$. But in many economic examples it is equally or more plausible that there are linear combinations across groups, e.g. linear combinations of Y_{it} and Y_{jt} which are $I(0)$. Their results suggest that between group cointegration can distort the results of within group cointegration tests. If the number of groups, N, is small one can, as with the SURE case, work with the full system, but in many cases this is not feasible.

6 Conclusion

It should be emphasised, that this area is developing very rapidly, with very many interesting and often surprising results emerging. However, even given this there appear to be some general points that applied workers might bear in mind. First, one should be very careful about using standard pooled estimators such as fixed effects to estimate dynamic models (i.e. including lagged dependent variables) from panel data. The dynamic parameters are subject to large potential biases when the parameters differ across groups and the regressors are serially correlated. However, for some purposes, such as forecasting or estimating long-run parameters, the pooled estimators may perform well. Second, pooled (or cross-section) regressions can be measuring very different parameters from the averages of the corresponding parameters in time-series regressions. In many cases this difference can be expressed as a consequence of a dependence between the time-series parameters and the regressors. The interpretation of this difference will depend on the theory related to the substantive application. It is not primarily a matter of statistical technique. Third, the mechanical application of panel unit-root or cointegration tests is to be avoided. To apply these tests requires that the hypotheses involved are interesting in the context of the substantive application, which is again a question of theory rather than statistics.

References

Andersson, J. and J. Lyhagen (1999) A long-memory panel unit root test: PPP revisited, Stockholm School of Economics, Working paper 303.

Attanasio, O.P., L. Picci and A. Scorcu (2000) Saving Growth and Investment: A Macroeconomic Analysis using a Panel of Countries, Review of Economics and Statistics, LXXXII, May, 182-211.

Baltagi, B.H. (1995) *Econometric Analysis of Panel Data*, New York: Wiley.

Baltagi, B.H. and J.M. Griffin (1997) Pooled Estimators versus their Heterogeneous Counterparts in the context of dynamic demand for gasoline, *Journal of Econometrics*, 77, 303-327.

Banerjee, A (1999) Panel Data, Unit Roots and Cointegration: An Overview, *Oxford Bulletin of Economics and Statistics*, Special Issue on Testing for Unit Roots and Cointegration using Panel Data, Theory and Applications, 61, November, 607-629.

Banerjee, A., M. Marcellino, C. Osbat (2000) Some Cautions on the Use of Panel Methods for Integrated Series of Macro-Economic Data, mimeo, European University Institute.

Binder, M., C. Hsiao and M.H. Pesaran (1999) Likelihood Based Inference for Panel Vector Autoregressions: Testing for Unit Roots and Cointegration in Short Panels, mimeo, Cambridge, available on www.econ.cam.ac.uk /faculty/pesaran.

Boyd, D and R.P. Smith (2000) Some Econometric Issues in Measuring the Monetary Transmission Mechanism, with an application to Developing Countries, mimeo Birkbeck College.

Breitung, J.M. (1994) Testing for unit roots using panel data: are wages on different bargaining levels cointegrated, *Applied Economics*, 26, 353-361.

Coakley J, A-M Fuertes and R.P. Smith (2000) Panel PPP estimators with I(1) real exchange rates, mimeo Birkbeck College.

Elliott, G., T.J. Rothenberg, and J.H. Stock (1996) Efficient Tests for an Autoregressive Unit Root, *Econometrica,* 64, 813-816.

Engle, R.F. and C.W.J. Granger (1987) Cointegration and Error Correction: representations, estimation and testing, *Econometrica,* 55, 252-276.

Granger, C.W.J. and P. Newbold (1974) Spurious Regression in Econometrics, *Journal of Econometrics*, 2, 111-120.

Hadri, K (1999) Testing the null hypothesis of stationarity against the alternative of a unit root in panel data with serially correlated errors, Liverpool Research Papers in Economics, 9905.

Harris, R.D.F. and E. Tzavalis (1999a) Inference for Unit Roots in Dynamic Panels Where the Time Dimension is fixed, *Journal of Econometrics*, 91, 201-226.

Harris, R.D.F. and E. Tzavalis (1999b) Inference for unit roots in dynamic panels with heteroscedastic and serially correlated errors, mimeo University of Exeter.

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

Hsiao, C., M.H. Pesaran and A. K. Tahmiscioglu (1999) Bayes Estimation of short-run Coefficients in Dynamic Panel Data Models, in *Analysis of Panels and Limited Dependent Variable Models: A Volume in Honour of G.S. Maddala* edited by C. Hsiao, K. Lahiri, L-F Lee and M.H. Pesaran, Cambridge: Cambridge University Press.

Im, K.S., M.H. Pesaran and Y. Shin (1997) Testing for unit roots in heterogenous panels, mimeo Cambridge, available on www.econ.cam.ac.uk /faculty/pesaran.

Johansen, S. (1988) Statistical Analysis of Cointegrating Vectors, *Journal of Economic Dynamics and Control*, 12, 231-254.

Kao, C (1999) Spurious Regression and Residuals based tests for cointegration in panel data, *Journal of Econometrics*, 90, 1-44.

Kwiatkowski, D., P.C. Phillips, P. Schmidt and Y. Shin (1992) Testing the null hypothesis of stationarity against the alternative of a unit root, *Journal of Econometrics*, 54, 159-178.

Larsson, R. and J. Lyhagen (1999) Likelihood-based inference in multivariate panel cointegration models, Stockholm School of Economics, Working paper 331.

Larsson, R., J. Lyhagen and M. Lothgren (1998) Likelihood-based cointegration tests in heterogenous panels, Stockholm School of Economics, Working paper 250.

Lee, K., M.H. Pesaran and R.P. Smith (1997) Growth and Convergence in a Multi-country Empirical Stochastic Solow Model, *Journal of Applied Econometrics, 12, 4, 357-392.*

Levin, A. and C. Lin (1993) Unit Root Tests in Panel Data: New Results, U.C. San Diego Working paper.

Maddala, G.S. and I.-M. Kim (1998) *Unit Roots, Cointegration and Structural Change*, Cambridge: Cambridge University Press.

Maddala, G.S. and S. Wu (1999) A Comparative Study of Unit Root Tests with panel data and a new simple test, *Oxford Bulletin of Economics and Statistics*, Special Issue on Testing for Unit Roots and Cointegration using Panel Data, Theory and Applications, 61, November, 631-652.

Matyas, L. and P. Sevestre (Eds.) (1996) *The Econometrics of Panel Data*, 2nd edition, Dordrecht: Kluwer Academic Publishers.

McCoskey, S. and C. Kao (1998) A Residual Based Test of the Null of Cointegration in Panel Data, *Econometric Reviews*, 17, 57-84.

Oxley, L. and M. McAleer (1999) *Practical Issues in Cointegration Analysis*, Oxford: Basil Blackwell.

Pedroni, P. (1995) Panel Cointegration; Asymptotic and Finite Sample Properties of Pooled Time-series Tests with applications to the PPP hypothesis, Indiana Working papers in Economics No 95-013.

Pedroni, P (1999) Critical values for cointegration tests in heterogenous panels with multiple regressors, *Oxford Bulletin of Economics and Statistics*, Special Issue on Testing for Unit Roots and Cointegration using Panel Data, Theory and Applications, 61, November, 653-670.

Pesaran, M.H., N. U. Haque and S. Sharma (1999) Neglected heterogeneity and dynamics in Cross-country Savings Regressions, forthcoming in J. Krishnakumar and E. Rouchetti (Eds) *Panel Data Econometrics - Future Directions: Papers in Honour of Prof. Balestra*, Contributions to Economic Analysis, Elsevier Science, available on www.econ.cam.ac.uk/ faculty/pesaran.

Pesaran, M.H. and R.P. Smith (1995) Estimating Long-run relationships from Dynamic Heterogenous Panels, *Journal of Econometrics, 68, 79-113.*

Pesaran, M.H. and R.P. Smith (1998) Structural Analysis of Cointegrating VARs, *Journal of Economic Surveys*, 12, 471-506, reprinted in Oxley and McAleer (1999).

Pesaran, M.H. Y. Shin and R.J. Smith (1999) Bounds Testing approaches to the Analysis of Long-run Relationships, University of Cambridge DAE Working Paper 9907 available on www.econ.cam.ac.uk /faculty/pesaran.

Pesaran, M.H., Y. Shin and R.P. Smith (1999) Pooled Mean Group Estimation of Dynamic Heterogenous Panels, *Journal of the American Statistical Association*, 94, 621-634.

Pesaran, M.H. R.P. Smith and K.S. Im (1996) Dynamic Linear Models for Heterogenous Panels p145-195 of Matyas and Sevestre (1996).

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

Phillips, P.C.B. and Z. Xiao (1998) A Primer on Unit Root Testing, *Journal of Economic Surveys*, 12, 423-470, reprinted in Oxley & McAleer (1999).

Phillips, P.C.B., and H.R. Moon (1999) Linear Regression Limit Theory for Nonstationary Panel Data, *Econometrica,* 67,5, 1057-1112.

Phillips, P.C.B., and H.R. Moon (2000) Nonstationary Panel Data Analysis: an overview of some recent developments, *Econometric Reviews* 19, 3, 263-286.

Quah, D (1994) Exploiting Cross-section Variation for Unit Root Inference in Dynamic Data, Economics Letters, 44, 9-19.

Rebucci, A (2000) Estimating Vector Autoregressions with Long Heterogeneous panels, mimeo, IMF.

Robertson, D. and J. Symons (1992) Some Strange Properties of Panel Data Estimators, *Journal of Applied Econometrics*, 7, 175-89.

Robertson, D. and J. Symons (1999) Factor Residuals in Panel Regressions: a suggestion for estimating panels allowing for cross-sectional dependence, mimeo University of Cambridge.

Swamy, P.A.V.B. (1970) Efficient Inference in a Random Coefficient Regression Model, *Econometrica* 38, 311-323.

Zellner, A. (1962) An Efficient Method of Estimating Seemingly Unrelated Regressions and Test for Aggregation Bias, *Journal of the American Statistical Association*, 57, 348-368.