Birkbeck Working Papers in Economics & Finance

ISSN 1745-8587

Birkbeck UNIVERSITY OF LONDON

School of Economics, Mathematics and Statistics

BWPEF 0403

Unobserved Heterogeneity in Panel Time Series Models

Jerry Coakley
Ana-Maria Fuertes
Ron Smith

October 2004

▪ Birkbeck, University of London ▪ Malet Street ▪ London ▪ WC1E 7HX ▪

Unobserved heterogeneity in panel time series models

Jerry Coakley[a],[*] Ana-Maria Fuertes[b], Ron Smith[c]

[a]Department of Accounting, Finance and Management, University of Essex, Colchester CO4 3SQ, UK.

[b]Faculty of Finance, Cass Business School, 106 Bunhill Row, London EC1Y 8TZ, UK.

[c]Department of Economics, Birkbeck College, Malet Street, London WC1E 7HX, UK.

October 2004

Abstract

Recently, the large T panel literature has emphasized unobserved, time-varying heterogeneity that may stem from omitted common variables or global shocks that affect each individual unit differently. These latent common factors induce cross-section dependence and may lead to inconsistent regression coefficient estimates if they are correlated with the explanatory variables. Moreover, if the process underlying these factors is nonstationary, the individual regressions will be spurious but pooling or averaging across individual estimates still permits consistent estimation of a long-run coefficient. The need to tackle both error cross-section dependence and persistent autocorrelation is motivated by the evidence of their pervasiveness found in three well-known, international finance and macroeconomic examples. A range of estimators is surveyed and their finite-sample properties are examined by means of Monte Carlo experiments. These reveal that a mean group version of the common-correlated-effects estimator stands out as the most robust since it is the preferred choice in rather general (non) stationary settings where regressors and errors share common factors and their factor loadings are possibly dependent. Other approaches which perform reasonably well include the two-way fixed effects, demeaned mean group and between estimators but they are less efficient than the common-correlated-effects estimator.

Keywords: Factor analysis; global shocks; latent variables
JEL Classification: C32; F31

1 Introduction

Panel or longitudinal data which have observations on cross-section units $i = 1, 2, ..., N$, such as individuals, firms or countries, over time periods $t = 1, 2, ..., T$ enable one to model a variety of forms of unobserved heterogeneity in regression models. The standard panel literature, developed

*Corresponding author: Tel. +44-01206-872455; fax: +44-01206-873429. E-mail address: jcoakley@essex.ac.uk (J. Coakley).

for cases where N is large and T is small, emphasizes unit-specific heterogeneity such as unobserved ability in earnings equations. When T is large, one can allow for such unit-specific heterogeneity by estimating a separate time-series equation for each unit. Recent years have witnessed increasing interest in panel data models with unobserved time-varying heterogeneity induced by common shocks that influence all units, perhaps to different degrees. This is particularly important in international finance and macroeconomics where long runs of data are available for many countries, each of which may be subject to global shocks. Such heterogeneity will introduce cross-section dependence or correlation between the errors of different units and will render the conventional estimators inconsistent if the global shocks are correlated with the regressors.

It is also quite plausible that these unobserved factors, such as technology shocks in a production function or financial innovation in a money demand function, may need first differencing to achieve stationarity. Such $I(1)$ shocks cause the variables not to cointegrate and the regression to be spurious, that is, the covariance between the $I(1)$ error and the $I(1)$ regressor does not go to zero even as $T \to \infty$ and so the estimator does not converge to the true parameter value but to a random variable. However, Phillips and Moon (1999, 2000) and Kao (1999) show that panels make it possible to obtain consistent estimators (as $N \to \infty$) of a long-run average parameter even when each of the individual time-series regressions is spurious. The averaging over N attenuates the noise in the individual estimators and thus facilitates a consistent estimator of the mean effect.

In the panel time-series literature where both N and T are large, the usual approach has been either to ignore the possibility of cross-section dependence produced by time-specific heterogeneity or deal with it by including period dummies or fixed effects. But this assumes that the global shocks have identical effects on each unit which seems quite restrictive. When N is of the same order of magnitude or greater than T, the traditional SUR-GLS approach for dealing with cross-section dependence breaks down because the estimated contemporaneous variance-covariance matrix cannot be inverted. If T is only slightly greater than N, estimation is feasible but it will be unreliable.

However, assuming cross-section independence seems restrictive for many applications in macroeconomics and finance and neglecting it may be far from innocuous as has been clear in the purchasing power parity (PPP) debate (see O'Connell, 1998). Phillips and Sul (2003) note that pooling may provide little gain in precision over single-equation estimation if there is substantial cross-section dependence. In addition, many theoretical results have been derived under the assumption of independence (Phillips and Moon, 2000). As Phillips and Moon (1999: p1092) put it "...quite commonly in panel data theory, cross-section independence is assumed in part because of the difficulties of characterizing and modelling cross-section dependence."

In spatial econometrics, quite popular in urban economics and regional science, a natural way to model cross-section dependence is in terms of distance (see Baltagi, 2001). But for most economic problems there is no obvious distance measure. In recent years, characterizing cross-section dependence by means of a factor structure has attracted a lot of attention (Robertson and Symons, 1999; Bai and Ng, 2002; Coakley, Fuertes and Smith, 2002; Phillips and Sul, 2003; Moon and Perron, 2004; Pesaran 2004a). Accordingly, the disturbances are assumed to contain one or more unobserved (latent) factors which may influence each unit differently.

This paper examines the consequences of time-varying heterogeneity that arises from unobserved factors, which are possibly $I(1)$ processes, and the relative effectiveness of various approaches in dealing with this phenomenon. The focus of the analysis is on *estimation* issues rather than inference. Section 2 provides an empirical illustration of the problems. It shows that three standard bivariate economic relations involve substantial cross-section dependence and the residuals resemble $I(1)$ series. Section 3 discusses a range of possible estimators. Since we want to make the paper accessible to a wide audience, we indicate the nature of the issues rather than provide formal proofs or derivations. Section 4 provides Monte Carlo evidence on the finite sample properties of these estimators under various data generation processes and Section 5 concludes.

2 Empirical illustrations

We take three standard applications to assess the extent of the two problems, cross-section dependence and $I(1)$ errors, and to help in designing our Monte Carlo experiments. The applications are PPP, the Fisher relationship and the Feldstein-Horioka (FH) puzzle. Each of them involves a simple bivariate linear relationship that should hold in the long run.

Let s_{it} be the logarithm of the nominal exchange rate and $d_{it} = p_{it} - p_t^*$ the log price differential between country i and the base country (the US) at period t. According to PPP, exchange rates should reflect price fluctuations in the long-run so in the regression

$$s_{it} = \alpha_i + \beta_i d_{it} + e_{it}, \tag{1}$$

the restriction $\beta_i = 1$ should hold. Boyd and Smith (1999) and Coakley, Flood, Fuertes and Taylor (2004) provide further discussion.

Let il_{it} denote the annualized long-term nominal interest rate and π_{it} the log annual inflation rate. Assuming $E_t(\pi_{i,t+1}) = \pi_{it}$, the ex ante real interest rate is $rl_{it} = il_{it} - \pi_{it}$. The Fisher effect suggests that nominal interest rates fully reflect inflation expectations in the long-run. Thus in

$$il_{it} = \alpha_i + \beta_i \pi_{it} + e_{it}, \tag{2}$$

the restriction $\beta_i = 1$ should hold. Coakley, Fuertes and Wood (2004) discuss this in more detail. In both examples, one might expect common (across countries) factors to be present. These would include base country effects, oil price shocks and the long swings in the real dollar rate for PPP and movements in the world real interest rate for the Fisher equation.

Let I_{it} be the share of domestic fixed investment in GDP and S_{it} the share of savings. In a world of free capital mobility, national saving would flow to the countries offering the highest returns and domestic investment would be financed from global capital markets. Thus in

$$I_{it} = \alpha_i + \beta_i S_{it} + e_{it}, \tag{3}$$

$\beta_i = 0$ should hold. The puzzle is that Feldstein and Horioka (1980) found the average β_i for OECD countries to be close to unity, the expected value under no capital mobility. Coakley, Kulasi and Smith (1996, 1998) and Coakley, Fuertes and Spagnolo (2004) provide further discussion.

The analysis for the PPP and Fisher equations is based on quarterly data for 18 countries (Australia, Austria, Belgium, Canada, Denmark, France, Germany, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, South Africa, Sweden, Switzerland, UK and US) over the 1973Q1-1998Q4 period. The panel dimensions for the PPP analysis are $N = 17$ (US is excluded) and $T = 104$ while those for the Fisher regression are $N = 18$ (US is included) and $T = 100$ (four observations are lost in calculating the annual inflation series $\pi_{it} = p_{it} - p_{i,t-4}$). Nominal exchange rates and prices are scaled (1995=100) to remove the effect of units of measurement on the intercepts. Long-term interest rates are average yields to maturity on bellwether government bonds with residual maturities between 9 and 10 years. All the price indexes are CPI series except for Australia where the PPI is used due to data unavailability. The FH regression is based on quarterly national saving, domestic investment and GDP observations for 12 OECD economies (Australia, Canada, Finland, France, Italy, Japan, Netherlands, Norway, Spain, Switzerland, UK and US) over 1980Q1-2000Q4.

Table 1 gives various summary statistics for the variables and two sets of residuals coming from individual OLS and from two-way fixed effects (2FE). Both levels and first differences are considered. The 2FE estimator imposes slope coefficient (and error variance) homogeneity but allows for country effects α_i and time effects α_t. The latter may pick up any common factor.

[Table 1 around here]

On the one hand, Table 1 reports the average (absolute) correlation as an indication of the degree of cross-section dependence — Pesaran (2004b) proposes a test for cross-section dependence based on the average correlation of the residuals and compares it with the Breusch-Pagan (1980) test based on the average of the squared correlations. On the other hand, Table 1 reports the proportion of the variance accounted for by the first two principal components (PCs), as an indication of how well a factor structure works, and the average ADF t-statistic of Im, Pesaran and Shin (2003) [IPS] as an indication of the possibility of a unit root. The PCs are the linear combinations of the standardized time series that account for the maximal amount of the total variation. The eigenvectors of the relevant correlation matrix are the weights and the ordered eigenvalues over the cumulated eigenvalues give the variance proportions. The first PC often has roughly equal weights and so it is close to the cross-section mean of the data for each time period.

The average absolute correlations between OLS residuals are 0.67 for PPP, 0.55 for Fisher and 0.26 for FH. Using the 2FE estimator reduces the average absolute correlation in the PPP and (somewhat in) Fisher but not the FH case. There is little difference between the average absolute correlation and the average correlation (except for FH) since the residuals are mainly positively correlated. This is not always the case for the variables. In particular, the log price differential has an average absolute correlation of 0.84 but an average correlation of only 0.06 because large positive and negative correlations cancel out. The first PC accounts for 72% of the OLS residual variance in the PPP case, 61% in Fisher and 29% in FH and similarly for the 2FE residuals. In the PPP and Fisher cases, the first two factors explain about 80% of the total residual variation.

The IPS test is designed for variables (not residuals) and it assumes cross-section independence. Therefore, the average ADF statistics should be treated as descriptive rather than as formal tests. The fact that these statistics are rather small (around -2) suggests that a unit root is likely to be present in the disturbances for many of the countries. There is slightly more evidence for a unit root in the residuals from 2FE than in those from individual OLS, which is the reverse of what one would expect if there was an $I(1)$ factor that the time fixed effects have removed. The first-differenced series yield much larger (absolute) average ADF statistics, as expected, and lower cross-section correlations. However, the residual dependence is still quite marked in the PPP and Fisher cases. This analysis illustrates that both cross-section dependence and potentially $I(1)$ errors are a pervasive feature of the levels regressions (1)-(3).

3 Alternative panel estimators

3.1 The model

Suppose that the data generating process is a linear heterogeneous panel model

$$y_{it} = \alpha_i + \beta_i x_{it} + u_{it}, \quad i = 1, ..., N; \quad t = 1, ..., T, \tag{4}$$

where the parameters are distributed randomly over units, $\alpha_i = \alpha + \eta_{\alpha i}$ and $\beta_i = \beta + \eta_{\beta i}$ with $\eta_{\alpha i} \sim iid(0, \sigma_\alpha^2)$ and $\eta_{\beta i} \sim iid(0, \sigma_\beta^2)$, and independently of x_{it} and u_{it}. Such random coefficient models (RCM) are discussed by Hsiao (2003) and Hsiao and Pesaran (2004). The variables and disturbances may be $I(1)$ or $I(0)$. The cross-section and time dependence structure is given by

$$\begin{aligned} u_{it} &= \rho_{ui} u_{it-1} + \gamma_i f_t + \varepsilon_{u,it}, \quad \varepsilon_{u,it} \sim iid(0, \sigma_{ui}^2), & (5) \\ x_{it} &= \rho_{xi} x_{it-1} + \phi_i f_t + \psi_i \chi_t + \varepsilon_{x,it}, \quad \varepsilon_{x,it} \sim iid(0, \sigma_{xi}^2), & (6) \end{aligned}$$

where iid denotes independence across t and i. Both f_t and χ_t are latent common factors such that f_t may influence both errors (loading γ_i) and regressors (loading ϕ_i) whereas χ_t is regressor specific. If $\gamma_i \neq 0$ and $\phi_i \neq 0$, the error and regressor in (4) are correlated. We assume that $\varepsilon_{u,it}$ and $\varepsilon_{x,it}$ are independently distributed. The factors may be $I(0)$ or $I(1)$ processes such as

$$\begin{aligned} f_t &= \rho_f f_{t-1} + \varepsilon_{ft}, \quad \varepsilon_{ft} \sim iid(0, \sigma_f^2), & (7) \\ \chi_t &= \rho_\chi \chi_{t-1} + \varepsilon_{\chi t}, \quad \varepsilon_{\chi t} \sim iid(0, \sigma_\chi^2), & (8) \end{aligned}$$

where ε_{ft} and $\varepsilon_{\chi t}$ are independently distributed.

We do not consider lagged dependent variables as regressors because this raises a variety of quite different issues central to a distinct literature on panel unit root testing surveyed by Trapani (2004). The parameter of interest is the mean effect β. The estimators we consider differ in how they deal with: a) unobserved heterogeneity, b) error cross-section dependence and c) dependence between x_{it} and u_{it} induced by latent common factors. These issues are discussed below.

3.2 Pooled OLS (POLS)

This approach simply pools the data and ignores parameter heterogeneity. It estimates

$$y_{it} = \alpha + \beta x_{it} + e_{it},\ e_{it} \sim iid(0, \sigma^2),\ i = 1, ..., N,\ t = 1, ..., T, \tag{9}$$

by OLS. The POLS residuals measure $e_{it} = u_{it} + \eta_{\alpha i} + \eta_{\beta i} x_{it}$ where u_{it} is the true disturbance. Even if homogeneity is wrongly imposed, there is no correlation between e_{it} and x_{it} because $\eta_{\alpha i}$ and $\eta_{\beta i}$ are independent of x_{it}, so the $\hat{\beta}^{POLS}$ estimator is unbiased and consistent provided that x_{it} and u_{it} are not influenced by the same factor. But if $\gamma_i \neq 0$ and $\phi_i \neq 0$ then POLS is inconsistent.

For non-stationary variables that cointegrate homogeneously ($\eta_{\beta i} = 0$), the POLS estimator is $T\sqrt{N}$-consistent. For the $I(1)$ error case (no cointegration), Phillips and Moon (1999) show that POLS is $\sqrt{N}$-consistent for a *long run average*, namely, the ratio of the expected (over $i = 1, ..., N$) long-run covariance between y_{it} and x_{it} to the expected long-run variance of x_{it}. In their particular random coefficients setting, the latter is different from the *average long run* defined as the expected (over $i = 1, ..., N$) value of the ratio of the long-run covariance over the variance.

3.3 Individual fixed effects (FE)

The FE approach introduces dummies to allow the intercepts to differ by unit and estimates

$$y_{it} = \alpha_i + \beta x_{it} + e_{it},\ e_{it} \sim iid(0, \sigma^2), \tag{10}$$

by OLS. This amounts to regressing $(y_{it} - \overline{y}_i)$ on $(x_{it} - \overline{x}_i)$ where $\overline{y}_i = T^{-1} \sum_{t=1}^{T} y_{it}$ and $\overline{x}_i = T^{-1} \sum_{t=1}^{T} x_{it}$ are the unit means. The issues discussed above for the POLS estimator regarding non-stationary variables and $I(0)$ or $I(1)$ errors apply equally to this estimator as does inconsistency when regressors and disturbances are influenced by the same latent factor.

3.4 Two-way fixed effects (2FE)

This approach allows the intercepts to differ, both by unit and time period, and estimates

$$y_{it} = \alpha_i + \alpha_t + \beta x_{it} + e_{it},\ e_{it} \sim iid(0, \sigma^2), \tag{11}$$

by OLS or equivalently, a regression of $(y_{it} - \overline{y}_i - \overline{y}_t + \overline{y})$ on $(x_{it} - \overline{x}_i - \overline{x}_t + \overline{x})$ where $\overline{y}_t = N^{-1}\sum_{i=1}^{N} y_{it}$ are the time means and $\overline{y} = (NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N} y_{it}$ is the overall mean and similarly for x. If the true country slopes and variances are homogeneous ($\beta_i = \beta$; $\sigma_i^2 = \sigma^2$) and there is a single unobserved factor f_t that has an identical influence on each unit, then this is captured by the time effects ($\alpha_t = \gamma f_t$) and the estimator $\hat{\beta}^{2FE}$ is unbiased and efficient.

If $\gamma_i \neq 0$ and $\phi_i \neq 0$ so that regressors and errors are correlated, the 2FE estimator remains unbiased as long as γ_i and ϕ_i are independent since 2FE amounts to FE for the demeaned data $\tilde{y}_{it} = y_{it} - \overline{y}_t$ and $\tilde{x}_{it} = x_{it} - \overline{x}_t$. For equations (5) and (6), assuming $\rho_{xi} = \rho_{ui} = 0$ for simplicity, we have $\tilde{x}_{it} = (\phi_i - \overline{\phi})f_t + (\psi_i - \overline{\psi})\chi_t + (\varepsilon_{x,it} - \overline{\varepsilon}_{xt})$ and $\tilde{u}_{it} = (\gamma_i - \overline{\gamma})f_t + (\varepsilon_{u,it} - \overline{\varepsilon}_{ut})$. The covariance between $\tilde{x}_{it}$ and $\tilde{u}_{it}$, equal to $E\{(\phi_i - \overline{\phi})(\gamma_i - \overline{\gamma})f_t^2\}$, is zero if ϕ_i and γ_i are independent.

3.5 Fixed effects with principal components (FE-PC)

Coakley, Fuertes and Smith (2002) suggest estimating individual OLS regressions of y_{it} on x_{it} to extract the residual PCs as proxies for the latent factors. The second stage consists of estimating

$$y_{it} = \alpha_i + \beta x_{it} + \mathbf{c}'\mathbf{z}_t + e_{it},\ e_{it} \sim iid(0, \sigma^2) \tag{12}$$

where $\mathbf{c} = (c_1, ..., c_J)'$ and $\mathbf{z}_t = (z_{1t}, ..., z_{Jt})'$ are the $J < N$ largest PCs of the first-stage standardized residuals. Factor-model information criteria, such as those derived by Bai and Ng (2002), can be used to choose J. The estimator of β in (12), called FE-PC, is consistent if regressors and errors in (4) are uncorrelated and more generally ($\gamma_i \neq 0$ and $\phi_i \neq 0$), provided that f_t can be perfectly measured by the cross-section mean of the regressor ($\psi_i = 0$) as noted by Pesaran (2004a).

3.6 Mean group (MG)

None of the above estimators allows for heterogeneity in the slopes. Pesaran and Smith (1995) propose a MG approach which does so by estimating individually (OLS) the equations

$$y_{it} = \alpha_i + \beta_i x_{it} + e_{it},\ e_{it} \sim iid(0, \sigma_i^2), \tag{13}$$

and define the estimator $\hat{\beta}^{MG} = N^{-1}\sum_{i=1}^{N}\widehat{\beta}_i$ with variance $V(\hat{\beta}^{MG}) = \frac{1}{N(N-1)}\sum_{i=1}^{N}(\widehat{\beta}_i - \overline{\beta})^2$. Hsiao and Pesaran (2004) review this and other RCM estimators.

If the variables are $I(1)$ and cointegrated, then $\widehat{\beta}_i$ is superconsistent (rate T) for the long-run coefficient β_i. However, the estimates $\widehat{\beta}_i$ will be spurious if e_{it} is $I(1)$. But again, as with POLS and FE, averaging over the units will attenuate the noise allowing a consistent estimator of β for large N. The response surface estimates in Coakley, Fuertes and Smith (2001) suggest that the dispersion of $\hat{\beta}^{MG}$ falls at rate $\sqrt{N}$ in the $I(1)$ error case, just like that of POLS and FE.

3.7 SUR mean group (SUR-MG)

In the SUR approach introduced by Zellner (1962), the individual OLS residuals are used to construct a covariance matrix estimate which, in turn, facilitates the FGLS estimate $\hat{\boldsymbol{\beta}} \equiv (\hat{\beta}_1, ..., \hat{\beta}_N)'$. The SUR-MG estimator is defined as the average of $\hat{\beta}_i, i = 1, ..., N$. When regressors and errors are uncorrelated ($\phi_i = 0$), the SUR-MG estimator is unbiased and more efficient than MG because it accounts for the non-zero cross section covariances. However, it does not fully use the information that the latter arise from a factor structure, so there may be more efficient estimators.

If the same latent factor affects regressors and errors ($\gamma_i \neq 0$ and $\phi_i \neq 0$), then SUR is no longer consistent. The bias of SUR-MG will generally differ from that of MG. One might expect the former to be smaller because SUR gives less weight than individual OLS to observations with large variances, those where the factors are important. Moreover, the SUR approach is infeasible for $N > T$ because the estimated covariance matrix cannot be inverted. Robertson and Symons (1999) suggest exploiting the factor structure to tackle this problem. But their estimator is quite complicated and will not be consistent if the unobserved factors are correlated with the regressors.

3.8 Demeaned mean group (DMG)

Another approach is to demean the data for the OLS estimation of the individual regressions

$$\tilde{y}_{it} = \alpha_i + \beta_i \tilde{x}_{it} + e_{it}, \; e_{it} \sim iid(0, \sigma_i^2), \tag{14}$$

where $\tilde{y}_{it} = y_{it} - \overline{y}_t$ and $\overline{y}_t = N^{-1} \sum_i y_{it}$. The DMG estimator is defined as the average of the $\widehat{\beta}_i$. For the RCM with one factor (and $\rho_{ui} = 0$ for simplicity) we have $y_{it} = \alpha_i + \beta_i x_{it} + \gamma_i f_t + \varepsilon_{u,it}$ with time means $\overline{y}_t = \alpha + \beta \overline{x}_t + \overline{\gamma} f_t + \overline{\varepsilon}_{ut} + N^{-1} \sum_{i=1}^N \eta_{\beta i} x_{it} + \overline{\eta}_\alpha$. Noting that $\beta_i x_{it} - \beta \overline{x}_t = \beta_i \tilde{x}_{it} + \eta_{\beta i} \overline{x}_t$, it follows that the true relation between the demeaned variables $\tilde{y}_{it}$ and $\tilde{x}_{it}$ is

$$\tilde{y}_{it} = \eta_{\alpha i} + \beta_i \tilde{x}_{it} + \tilde{\varepsilon}_{u,it} + v_{it}, \tag{15}$$

where $\tilde{\varepsilon}_{u,it} = \varepsilon_{u,it} - \overline{\varepsilon}_{ut}$, $v_{it} = (\gamma_i - \overline{\gamma}) f_t + \eta_{\beta i} \overline{x}_t - N^{-1} \sum_{i=1}^N \eta_{\beta i} x_{it}$ and $\overline{\eta}_\alpha \simeq 0$. Hence, the residuals from (14) measure $e_{it} = \tilde{\varepsilon}_{u,it} + v_{it}$. If the latent factors have identical effects on each unit ($\bar{\gamma} = \gamma_i = \gamma$), demeaning removes the cross-section dependence because $(\gamma_i - \bar{\gamma}) f_t = 0$ but it adds new error terms due to the slope heterogeneity ($\eta_{\beta i} \neq 0$). If in addition, there is a regressor-specific factor ($\psi_i \neq 0$), demeaning removes it and so the regressor variance in (14) falls which may adversely affect the estimation efficiency. Since the DMG and 2FE approaches only differ in that the latter imposes slope and error variance homogeneity, they raise similar issues. As with 2FE, if disturbances and regressor are correlated ($\gamma_i \neq 0$ and $\phi_i \neq 0$), the DMG estimator remains unbiased as long as γ_i and ϕ_i are mutually independent.

3.9 Mean group with principal components (MG-PC)

The homogeneity restriction in the FE-PC approach can be relaxed by individually estimating

$$y_{it} = \alpha_i + \beta_i x_{it} + \mathbf{c}_i' \mathbf{z}_t + e_{it},\ e_{it} \sim iid(0, \sigma_i^2),$$

by OLS. The MG-PC estimator is defined as the average of the individual $\hat{\beta}_i$ estimates. This has similar properties to FE-PC, namely, it is consistent when a common factor drives errors and regressors provided that $\psi_i = 0$. Telser (1964) noted that SUR-GLS could be implemented by augmenting each equation with the OLS residuals from the remaining $N-1$ equations. This is not feasible for $N > T$ but including the $J < N$ largest residual PCs provides a parsimonious approximation to it. Hence, if regressors and errors are uncorrelated ($\phi_i = 0$), the consistent MG-PC estimator can be seen as a feasible alternative to SUR-GLS in large N and T panels.

3.10 Common correlated effects mean group (CMG)

Pesaran (2004a) suggests including the cross-section averages of the observed variables as proxies for the latent factors, that is, the mean effect β is estimated through the augmented regression

$$y_{it} = \alpha_i + \beta_i x_{it} + c_{1i}\overline{y}_t + c_{2i}\overline{x}_t + e_{it}, e_{it} \sim iid(0, \sigma_i^2), \tag{16}$$

where, although $\bar{y}_t$ and e_{it} are not independent, their correlation goes to zero as $N \to \infty$. For the RCM with $\eta_{\beta i} = 0$, $\rho_{xi} = 0$ and $\rho_{ui} = 0$ without loss of generality, we have $\overline{y}_t = \overline{\alpha} + \beta\overline{x}_t + \overline{\gamma} f_t + \overline{\varepsilon}_{u,t}$ which suggests that $\overline{y}_t - \beta\overline{x}_t$ can capture the effect of f_t for large N as long as $\overline{\gamma} \neq 0$. Pesaran shows that this estimator is consistent for β in a RCM with general cross-section dependence such as that implied by (5)-(6) with $\gamma_i \neq 0$, $\phi_i \neq 0$ and $\psi_i \neq 0$. The consistency proof holds for any linear combination of the variables, i.e. $\bar{y}_t = \sum_i w_i y_{it}$ and $\overline{x}_t = \sum_i w_i x_{it}$ subject to the assumptions

$$(a)\ \ w_i = O\left(\frac{1}{N}\right),\quad (b)\ \sum_{i=1}^{N} \mid w_i \mid < K,\quad (c)\ \sum_{i=1}^{N} w_i \gamma_i \neq 0,$$

where K is a finite constant. These clearly hold for the arithmetic mean since $w_i = 1/N$, $\sum_{i=1}^{N} |w_i| = 1$ and $N^{-1}\sum_{i=1}^{N} \gamma_i = \overline{\gamma}$. Here we focus on a cross-sectionally augmented MG estimator (referred to as CMG) defined as the average of the individual OLS estimates from regression (16). Pesaran discusses the latter and a one-way fixed effects variant also.

3.11 Between or cross-section (CS)

Pesaran and Smith (1995) noted that the OLS estimator of the between or cross-section regression

$$\overline{y}_i = \alpha + \beta \overline{x}_i + e_i,\ e_i \sim iid(0, \sigma^2),\ i = 1, ..., N, \tag{17}$$

remains consistent for the mean effect β in the presence of $I(1)$ errors. This requires the RCM assumptions and strict exogeneity. Furthermore, if the data are generated by (4) with error cross-section dependence due to a latent factor that influences the regressors also, this between estimator is unbiased provided that the regressor and error loadings (ϕ_i and γ_i) are mutually independent.

4 Small sample properties

4.1 Monte Carlo design

The purpose of this section is to compare the small sample properties of the ten estimators discussed above in settings with error cross-section dependence. The errors may be either $I(0)$ or $I(1)$ processes. Each experiment involves 5,000 replications of $(N, T + T_0)$ observations where the first $T_0 = 50$ observations are discarded for each time series to minimize the (zero) initialization effects. We employ $(N, T) = \{(30, 100), (20, 30)\}$ which roughly typify macroeconomic panels of quarterly and annual frequency, respectively. In both cases $T > N$ so that SUR estimation is feasible.

The data generating process (DGP) for the experiments is

$$y_{it} = \alpha_i + \beta_i x_{it} + u_{it},\quad i = 1, ..., N,\quad t = 1, ..., T, \tag{18}$$

$$u_{it} = \rho_{ui} u_{it-1} + \boldsymbol{\gamma}_i' \mathbf{f}_t + \varepsilon_{u,it},\quad \varepsilon_{u,it} \sim iidN(0, \sigma_{ui}^2), \tag{19}$$

$$x_{it} = \rho_{xi} x_{it-1} + \boldsymbol{\phi}_i' \mathbf{f}_t + \psi_i \chi_t + \varepsilon_{x,it},\quad \varepsilon_{x,it} \sim iidN(0, \sigma_{xi}^2), \tag{20}$$

where $\mathbf{f}_t = (f_{1t}, f_{2t})'$, $\boldsymbol{\gamma}_i = (\gamma_{1i}, \gamma_{2i})'$ and $\boldsymbol{\phi}_i = (\phi_{1i}, \phi_{2i})'$. The factors are generated as

$$f_{mt} = \rho_m f_{m,t-1} + \varepsilon_{f_{mt}}, \quad \varepsilon_{f_{mt}} \sim iidN(0, \sigma^2_{f_m}), \; m = 1, 2, \tag{21}$$

$$\chi_t = \rho_\chi \chi_{t-1} + \varepsilon_{\chi t}, \quad \varepsilon_{\chi t} \sim iidN(0, \sigma^2_\chi). \tag{22}$$

Thus the Monte Carlo design resembles the setup in Section 2.1 but is more general in that it allows for one or two common factors, f_{mt}, driving both errors and regressors.

Heterogeneous intercepts are generated as $\alpha_i \sim iidU[-0.5, 0.5]$ such that $\alpha \equiv E(\alpha_i) = 0$. We focus on the homogenous slope case, $\beta_i = \beta$, and set $\beta_i = 1$. Experiments using $\beta_i \sim iidU[0.5, 1.5]$ such that $\beta \equiv E(\beta_i) = 1$ gave quite similar results on parameter bias but had effects on estimated standard errors and efficiency ranking as noted below. Throughout, the factors have different effects (loadings) on each unit, $\psi_i \sim iidU[0.5, 1.5]$, $\gamma_{mi} \sim iidU[0.5, 1.5]$ and $\phi_{mi} \sim iidU[0.5, 1.5]$, $m = 1, 2$. We use $\sigma^2_{ui} = \sigma^2_{f_1} = \sigma^2_{f_2} = \sigma^2_\chi = 1$ but the regressor variance differs randomly across units, $\sigma_{xi} \sim iidU[0.5, 1.5]$, so that the FE and MG estimators are not identical. Two specifications are considered when the same factor influences errors and regressors: *a*) $\boldsymbol{\phi}_i$ is drawn independently from $\boldsymbol{\gamma}_i$ for each i; *b*) independence is introduced simply by using $\boldsymbol{\phi}_i = \boldsymbol{\gamma}_i$ for all i.

Each experiment is summarized by the sample mean (SM) and standard deviation (SSD) of $\hat{\beta}$ and the sample mean of the estimated standard error of $\widehat{\beta}$ (denoted by SE) over replications. These can be used to gauge the bias and variance of $\widehat{\beta}$, and the reliability of the conventional standard errors. In some settings, e.g. $I(1)$ disturbances, we already know that the conventional standard errors of particular estimators will be misleading. These issues are discussed below.

By varying the parameter specifications in (19)-(22) we have several settings which differ according to: *i*) the stationarity properties of variables, factors and errors *ii*) the number of common factors, *iii*) whether errors and regressors are independent, *iv*) whether the factor loadings in errors and regressors are independent, *v*) whether there is a regressor-specific factor. The analysis focuses on the two extreme cases of stationarity and unit root autocorrelation. In the stationary settings below we focus on $\rho_{ui} = \rho_{xi} = \rho_m = \rho_\chi = 0$ but results (available upon request) using autocorrelation of 0.3 are very similar. For the non-stationary settings, the variables are $I(1)$ throughout and the latent factors can be $I(0)$ or $I(1)$. The average absolute cross-section correlation (u_{it}) is in the 0.5-0.8 range for all experiments (except the baseline). The reported results for FE-PC and MG-PC are based on $J = 1$ but those for $J = 2$ are very similar. The settings are as follows:

i) Stationary settings

Case a (baseline): $\boldsymbol{\gamma}_i = \boldsymbol{\phi}_i = \mathbf{0}, \psi_i = 0$. No cross-section dependence.

Case b: $\gamma_{2i} = 0, \boldsymbol{\phi}_i = \mathbf{0}$. Factor f_{1t} drives the errors. Factor χ_t drives the regressors.

Case c: $\gamma_{2i} = \phi_{2i} = \psi_i = 0$. Factor f_{1t} drives both the errors and the regressors.

Case $\widetilde{\mathbf{c}}$: Like **c** but with $\gamma_{1i} = \phi_{1i}$ for all i to introduce factor-loading dependence.

Case d: $\gamma_{2i} = \phi_{2i} = 0$. Factor f_{1t} drives errors and regressors. Factor χ_t drives the regressors.

Case e: $\psi_i = 0$. Two factors, $\mathbf{f}_t = (f_{1t}, f_{2t})'$, drive both the errors and the regressors.

Case $\widetilde{\mathbf{e}}$: Like **e** but with $\boldsymbol{\gamma}_i = \boldsymbol{\phi}_i$ for all i. Factor-loading dependence.

Case f: Two factors, $\mathbf{f}_t$, drive errors and regressors. Factor χ_t drives the regressors.

ii) Non-stationary settings

Case A (baseline): $\rho_{ui} = 0, \boldsymbol{\gamma}_i = \boldsymbol{\phi}_i = \mathbf{0}, \psi_i = 0$. Cointegration. No cross-section dependence.

Case B: $\rho_{ui} = 1, \boldsymbol{\gamma}_i = \boldsymbol{\phi}_i = \mathbf{0}, \psi_i = 0$. No cointegration. No cross-section dependence.

Case C: $\rho_{ui} = 1, \gamma_{2i} = 0, \boldsymbol{\phi}_i = \mathbf{0}$. No cointegration. An $I(0)$ factor f_{1t} drives the errors. An $I(0)$ factor χ_t drives the regressors.

Case D: $\rho_{ui} = 1, \gamma_{2i} = \phi_{2i} = \psi_i = 0$. No cointegration. An $I(0)$ factor f_{1t} drives the errors and the regressors.

Case $\widetilde{\mathbf{D}}$: Like **D** but with $\gamma_{1i} = \phi_{1i}$ for all i. Factor-loading dependence.

Case E: $\rho_{ui} = 0, \gamma_{2i} = \phi_{2i} = \psi_i = 0$. Cointegration. An $I(0)$ factor f_{1t} drives errors and regressors.

Case F: $\rho_{ui} = 1, \gamma_{2i} = \phi_{2i} = 0$. No cointegration. An $I(0)$ factor f_{1t} drives errors and regressors. An $I(0)$ factor χ_t drives the regressors.

Case G: $\rho_1 = \rho_\chi = 1, \rho_{xi} = \rho_{ui} = 0, \gamma_{2i} = \phi_{2i} = 0$. No cointegration. An $I(1)$ factor f_{1t} drives both the errors and the regressors. An $I(1)$ factor χ_t drives the regressors.

4.2 Monte Carlo results

i) Stationary settings

The results for the experiments on stationary data are summarized in Tables 2(I) and 2(II) for the $N = 30, T = 100$ and $N = 20, T = 30$ panels, respectively.

[Table 2 around here]

The baseline results (**Case a**) for the two panel dimensions are similar except that the estimates from the larger sample have smaller standard errors. Since $\beta_i = \beta$ was used in generating the data, the estimators that allow for heterogeneous β_i suffer quite large losses of efficiency. For instance, in the small panel the reported SSD of the MG and FE estimators is 0.0536 and 0.0396, respectively. In contrast, the MG estimator is more efficient when $\beta_i \sim iidU[0.5, 1.5]$ in (18) instead. For instance, the SSD of MG and FE is 0.080 and 0.087, respectively. Similarly, since there is no cross-section dependence in the disturbances, estimators that allow for it lose efficiency.

When two independent common factors drive errors and regressors, respectively (**Case b**), there is cross-section dependence but since the errors are independent from the regressors, all the estimators remain unbiased. They differ in their efficiency as one would expect. Among the estimators that impose homogeneous slopes, 2FE is the most efficient whereas CS is the least efficient because by averaging x_{it} over time, the regressor variance falls. Among the MG variants, the MG-PC is more efficient than SUR-MG because it explicitly accounts for the factor structure.

When the same factor influences the errors and the regressors but there is independence between their loadings (**Case c**), the conventional POLS, FE and MG estimators are substantially biased. Their mean is about 1.5 rather than 1. The SUR-MG bias at about 0.4 is smaller as one would expect. The FE-PC and MG-PC approaches reduce the bias to about 0.15. In sharp contrast, the 2FE, DMG, CMG and CS estimators are unbiased. But when the factor loadings of errors and regressors are dependent (**Case $\widetilde{\mathbf{c}}$**), the 2FE, DMG and CS estimators show biases of the order of 0.08 whereas CMG remains unbiased. Additionally including a regressor-specific factor (**Case d**) gives similar results, except that all the biases, given by $Cov(x_{it}, e_{it})/V(x_{it})$ where e_{it} is the regression error, tend to be smaller because $V(x_{it})$ has now increased.

When $\mathbf{f}_t$ influences regressors and errors and $\boldsymbol{\phi}_i$ and $\boldsymbol{\gamma}_i$ are drawn independently (**Case e**), then 2FE, DMG, CMG and CS are all unbiased. Hence, adding a second factor does not change the results as long as there is independence between the factor loadings of errors and regressors. Absent the latter in a two-factor setting (**Case $\widetilde{\mathbf{e}}$**), all the estimators are now biased but the smallest bias is clearly that of CMG. Hence, the factor loading dependence in a multiple-factor setting appears to cause difficulties for the CMG. But this could potentially be dealt with by using as augmentation terms in (16) other weighted averages of the observed variables. Adding a regressor-specific factor (**Case f**) gives very similar results but again all the biases are now smaller.

In sum, these results suggest that the CMG estimator is quite robust. It is unbiased unless there are multiple common factors in disturbances and regressors together with factor-loading dependence. However, even in the latter case its bias is relatively small and it is clearly the most efficient estimator. The 2FE, DMG and CS estimators do quite well provided that there is factor-loading independence, but it would be difficult to judge a priori whether this is a plausible assumption in empirical applications. The CS estimator is also unbiased under factor-loading independence but it has very large variance.

ii) Non-stationary settings

The results for the non-stationary panels are summarized in Tables 3(I) and 3(II).

[Table 3 around here]

We first examine the baseline cointegration case with cross-sectional independence in regressors and errors (**Case A**). All the estimators are unbiased and, compared to the stationary counterpart case, their dispersion is substantially reduced because of the larger variance of the $I(1)$ regressor. For instance, in the small panel the SSD of FE (the efficient estimator) falls from 0.0396 to 0.0178. The improvement in the CS estimator is even more noticeable from 0.4596 to 0.0109.

The $I(1)$ disturbances setting (**Case B**) implies a substantial dynamic misspecification and the appropriate model is one in first differences. However, as established by Phillips and Moon (1999), averaging across spurious regressions produces unbiased estimates. Unsurprisingly, the $I(1)$ errors lead to much larger sampling variation. For instance, in the small panel the SSD of FE has increased from 0.0178 to 0.1387. Conventional standard errors are very misleading except those for the MG estimator (and variants) because they are based on the distribution of the individual $\widehat{\beta}_i$, and for the CS estimator because it averages out the time variation in the data.

When an $I(0)$ factor f_{1t} is introduced in the errors and another $I(0)$ factor χ_t in the regressors (**Case C**), all the estimators remain unbiased despite the lack of cointegration. This is because errors and regressors are uncorrelated. The theory in Phillips and Moon (1999) builds on the assumption of cross-section independence so this design (and others that follow) is of particular interest. The cross-section dependence induced by f_{1t} substantially increases the SSD of the FE and MG estimators but not so for the estimators that control for it. An exception to the latter is

SUR-MG (and the parsimonious approximation to it, MG-PC) whose SSD increases from 0.1482 to 0.2019. The most efficient estimator is CMG closely followed by 2FE.

In a non-cointegration setting with an $I(0)$ common factor that drives errors and regressors (**Case D**), all estimators but 2FE, DMG, CMG and CS are biased, as in the stationary counterpart case, because errors and regressors are correlated. The FE-PC and MG-PC estimators do rather badly because the $I(1)$ dynamics of the residuals makes it difficult to extract the $I(0)$ factor. However, this problem could be tackled by deploying a modified version of these estimators where the residual PC extraction is based on differencing (and recumulating) as proposed by Bai and Ng (2004) to estimate factors consistently whether they are $I(0)$, $I(1)$ or a mixture. Interestingly, POLS with a bias of only 0.058 does well relative to FE with a bias of 0.390, having similar variances. This is because the bias inversely depends on the regressor variance and, by taking deviations from unit means in the FE approach, the regressor variance falls and the bias increases. If dependence between the factor loadings of errors and regressors is introduced (**Case $\widetilde{\mathbf{D}}$**), only the CMG remains unbiased as in the stationary case.

We then consider the cointegration setting where an $I(0)$ factor affects errors and regressors (**Case E**). Here the biases are rather small (particularly for the large $T = 100$ panel) because the correlation between the $I(1)$ regressor and the $I(0)$ disturbance goes to zero with T. With no cointegration and an $I(0)$ factor driving errors and regressors and a regressor-specific $I(0)$ factor (**Case F**) the results are similar to Case D where the latter is absent. But the biases are now somewhat smaller because the regressor variance has increased. Finally, we simulate the factors f_{1t} and χ_t as $I(1)$ processes (**Case G**). Again the 2FE, DMG, CMG and CS estimators remain unbiased with CMG the most efficient. But the biases of the remaining estimators are now larger, particularly for POLS, FE, MG and SUR-MG, due to the $I(1)$ dynamics of the factors.

We carried out experiments for other $I(1)$ factor settings and the results also suggest that the CMG estimator is the most robust. For instance, if another $I(1)$ factor f_{2t} is included, the only issue is that the biases above increase further — in the large panel the mean for FE, MG and SUR-MG is 1.77, 1.92 and 1.84, respectively. With factor-loading dependence ($\gamma_{1i} = \phi_{1i}$) in Case G, there are now biases in the 2FE, DMG and CS estimators but not in CMG — in the large panel the means are 1.24 (2FE), 1.34 (DMG), 1.39 (Between) and 1.00 (CMG).

5 Concluding remarks

We have considered the estimation of the mean slope coefficient in a linear heterogeneous panel regression where the disturbances are correlated across units due to unobserved factors, such as global shocks, that may also influence the regressors. The disturbances can be $I(1)$ as well as $I(0)$ processes. The analysis is motivated by the need to tackle both error cross-section dependence and persistent autocorrelation found in three empirical macroeconomic examples. We discuss the impact of these phenomena on ten alternative estimation approaches. Their small sample properties are compared through Monte Carlo experiments. It turns out that in panels one can obtain unbiased estimates of average long-run parameters even in the context of $I(1)$ disturbances.

Overall, the novel CMG estimator stands out as the most robust in the sense that it is the preferred choice in rather general (non) stationary settings where regressors and errors share common factors and their factor loadings are possibly dependent. It is based on the common-correlated-effects approach of Pesaran (2004a) which simply augments the regression of interest with the time means of the variables to approximate the factor structure that induces the cross-section dependence. Other approaches which perform reasonably well include 2FE, DMG and CS but they are relatively less efficient than CMG. These estimators show essentially zero bias in most of the experiments except when there is factor-loading dependence. Under several of the factor structures considered, the remaining estimators are inconsistent although POLS and FE-PC exhibit less bias than FE, MG, SUR-MG and MG-PC.

The theoretical literature on cross-section dependence is growing rapidly but many issues await further research. As yet there is a relatively small empirical literature that deals with cross section dependence and so it is unclear which of the available estimators is most appropriate. The answer depends on what the true data generating process is. Application of these methods to our three empirical examples is a matter for further research which will have to consider a number of other specification issues. In particular, our assumption that the parameters are randomly distributed may not be appropriate for those examples. The analysis in this paper assumed a static relationship between (non-)stationary variables. The dynamic case, including $I(0)$ or $I(1)$ unobserved common factors, raises a number of different issues and warrants consideration in a separate paper.

Acknowledgements

Earlier versions of this paper were presented at the 2002 Panel Data Conference, Academy of Sciences, Berlin, the 2001 SCE Conference on Computing in Economics and Finance, Yale University, and at seminars at the Econometric Institute, Erasmus University Rotterdam, the Tinbergen Institute, Amsterdam, the University of Maastricht, Leeds University Business School, and the University of Swansea. We thank Hashem Pesaran, George Kapetanios and Michael Binder for helpful discussions and the Editor Stephen Pollock and four referees for insightful comments.

References

[1] Bai, J., Ng, S., 2004. A PANIC attack on unit roots and cointegration. Econometrica 72 1127-1178.

[2] Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191-221.

[3] Baltagi, B.H., 2001. Econometric Analysis of Panel Data. Second Edition, John Wiley, New York.

[4] Boyd, D., Smith, R.P., 1999. Testing for purchasing power parity: Econometric issues and an application to developing countries, Manchester School 67, 287-303.

[5] Breusch, T.S., Pagan, A.R., 1980. The Lagrange multiplier test and its application to model specifications in econometrics. Review of Economic Studies 47, 239-53.

[6] Coakley, J., Flood, R., Fuertes, A.M., Taylor, M.P., 2004. Purchasing power parity and the theory of general relativity: The first tests. Journal of International Money and Finance, Forthcoming.

[7] Coakley, J., Fuertes, A.M., Smith, R.P., 2001. The effect of I(1) errors in Large T, large N panels. Birkbeck College Discussion Paper in Economics 03/2001 (http://papers.ssrn.com).

[8] Coakley, J., Fuertes, A.M., Smith, R.P., 2002. A principal components approach to cross-section dependence in panels. Mimeo, Birkbeck College, University of London.

[9] Coakley, J., Fuertes, A.M., Spagnolo, F., 2004. Is the Feldstein-Horioka puzzle history? Manchester School 72, 569-590.

[10] Coakley, J., Fuertes, A.M., Wood, A., 2004. Reinterpreting the real exchange rate - yield differential nexus. International Journal of Finance and Economics 9, 201-218

[11] Coakley, J., Kulasi, F., Smith. R., 1996. Current account solvency and the Feldstein-Horioka puzzle, Economic Journal 106, 620-627.

[12] Coakley, J., Kulasi, F., Smith. R., 1998. The Feldstein-Horioka puzzle and capital mobility: A review. International Journal of Finance and Economics 3, 169-188.

[13] Feldstein, M., Horioka, C., 1980. Domestic saving and international capital flows. Economic Journal 90, 314-329.

[14] Hsiao, C., 2003. Analysis of Panel Data. Cambridge University Press, Cambridge.

[15] Hsiao, C. and Pesaran, M.H., 2004. Random coefficient panel data models, Mimeo, University of Cambridge.

[16] Im, K.S. Pesaran M.H. and Shin Y., 2003. Testing for unit roots in heterogeneous panels. Journal of Econometrics 115, 53-74.

[17] Kao, C., 1999. Spurious regression and residual-based tests for cointegration in panel data. Journal of Econometrics 90, 1-44.

[18] Moon, H. R., Perron, B., 2004. Testing for a unit root in panels with dynamic factors. Journal of Econometrics 122, 81-126.

[19] Pagan, A., 1984. Econometric issues in the analysis of regressions with generated regressors. International Economic Review 25, 221-247.

[20] Pesaran, M.H., Smith, R., 1995. Estimating long-run relationships from dynamic heterogeneous panels. Journal of Econometrics 68, 79-113.

[21] Pesaran, M.H., 2004a. Estimation and inference in large heterogeneous panels with a multifactor error structure. Mimeo Cambridge University.

[22] Pesaran, M.H., 2004b. General diagnostic tests for cross-section dependence in panels. Mimeo Cambridge University.

[23] Phillips, P.C.B., Moon, H.R., 1999. Linear regression theory for nonstationary panel data. Econometrica 67, 1057-1111.

[24] Phillips, P.C.B., Moon, H.R., 2000. Nonstationary panel data analysis: An overview of some recent developments. Econometric Reviews 19, 263-286.

[25] Phillips, P.C.B., Sul, D., 2003. Dynamic panel estimation and homogeneity testing under cross section dependence. Econometrics Journal 6, 217-259.

[26] Robertson, D., Symons, J., 1999. Factor residuals in panel regressions: A suggestion for estimating panels allowing for cross-section dependence. Mimeo, University of Cambridge.

[27] Telser L.G., 1964. Iterative estimation of a set of linear regression equations. Journal of the American Statistical Association 59, 845-862.

[28] Trapani, L., 2004. Testing for unit roots in heterogeneous panels under cross-sectional dependence. Cass Business School, CEA Working Paper 06-04.

[29] Zellner, A., 1962. An efficient method of estimating seemingly unrelated regressions and tests of aggregation bias. Journal of the American Statistical Association 57, 348-368.

TABLE 1
SUMMARY STATISTICS

	Levels regression				First-difference regression			
	y_{it}	x_{it}	$\hat{e}_{it}^{OLS}$	$\hat{e}_{it}^{2FE}$	Δy_{it}	Δx_{it}	$\hat{e}_{it}^{OLS}$	$\hat{e}_{it}^{2FE}$
a) Purchasing power parity								
ave($\lvert\rho_{ij}\rvert$)	0.5845	0.8421	0.6717	0.5994	0.5547	0.2762	0.5546	0.5550
ave(ρ_{ij})	0.3422	0.0588	0.6717	0.5764	0.5547	0.2656	0.5546	0.5550
$\%V_1$	0.5602	0.8554	0.7244	0.6486	0.6411	0.3355	0.6401	0.6419
$\%V_2$	0.3768	0.0910	0.1031	0.1795	0.0969	0.1355	0.0946	0.0964
$\bar{t}_{ADF}$	-1.618(1)	-2.195(4)	-2.047(1)	-1.923(1)	-7.586(1)	-2.644(4)	-7.649(1)	-7.678(1)
b) Fisher relationship								
ave($\lvert\rho_{ij}\rvert$)	0.6152	0.6682	0.5542	0.5272	0.3068	0.1724	0.2827	0.3008
ave(ρ_{ij})	0.5558	0.6680	0.5440	0.4891	0.3053	0.1616	0.2796	0.2991
$\%V_1$	0.6558	0.7070	0.6144	0.5758	0.3774	0.2342	0.3484	0.3717
$\%V_2$	0.1776	0.0930	0.1154	0.1775	0.1003	0.1031	0.1062	0.1016
$\bar{t}_{ADF}$	-1.072(1)	-2.539(4T)	-2.043(1)	-1.912(1)	-6.759(1)	-4.999(4)	-6.680(1)	-6.823(1)
c) Feldstein-Horioka puzzle								
ave($\lvert\rho_{ij}\rvert$)	0.3752	0.3479	0.2617	0.3579	0.0998	0.1166	0.1004	0.1047
ave(ρ_{ij})	0.2522	0.2357	0.1709	0.2558	0.0368	0.0913	0.0177	0.0185
$\%V_1$	0.3971	0.4088	0.2935	0.4009	0.1400	0.1841	0.1595	0.1633
$\%V_2$	0.2317	0.2507	0.2325	0.2282	0.1343	0.1186	0.1172	0.1167
$\bar{t}_{ADF}$	-1.809(1)	-1.627(1)	-2.093(1)	-1.811(1)	-6.089(1)	-5.661(1)	-6.003(1)	-6.026(1)

ave($\lvert\rho_{ij}\rvert$) is the average absolute cross-section correlation and ave(ρ_{ij}) is the average correlation. $\%V_j$ is the proportion of variability explained by the jth principal component (j=1,2) extracted from the relevant correlation matrix. $\hat{e}^{OLS}$ and $\hat{e}^{2FE}$ are the individual OLS and 2-way FE residuals, respectively. $\bar{t}_{ADF}$ is the mean ADF statistic for the unit root null; the lag order is in parentheses and T denotes a time trend.

Table 2(i)
Stationary Panel: $N = 30, T = 100$

	Case a		Case b		Case c		Case $\tilde{c}$		Case d		Case e		Case $\tilde{e}$		Case f	
	SM	SSD	SM	SSD	SM	SSD	SM	SSD	SM	SSD	SM	SSD	SM	SSD	SM	SSD
Estimator		*SE*		*SE*		*SE*		*SE*		*SE*		*SE*		*SE*		*SE*
POLS	1.000	.0191	1.001	.0269	1.454	.0515	1.495	.0512	1.420	.0475	1.610	.0470	1.661	.0434	1.572	.0458
		.0183		*.0220*		*.0162*		.0159		*.0156*		*.0144*		.0137		*.0142*
FE	.9999	.0186	1.002	.0239	1.457	.0515	1.497	.0512	1.422	.0474	1.613	.0470	1.663	.0432	1.574	.0458
		.0178		*.0213*		*.0159*		.0155		*.0153*		*.0142*		.0134		*.0140*
2FE	.9998	.0186	1.001	.0170	.9990	.0227	1.072	.0245	1.001	.0204	1.000	.0254	1.134	.0285	1.000	.0228
		.0181		*.0172*		*.0180*		.0180		*.0167*		*.0181*		.0179		*.0167*
FE-PC	.9999	.0186	1.000	.0170	1.082	.0343	1.157	.0391	1.072	.0296	1.197	.0642	1.304	.0568	1.165	.0519
		.0177		*.0165*		*.0165*		.0163		*.0154*		*.0159*		.0154		*.0150*
MG	.9994	.0215	1.001	.0294	1.499	.0557	1.499	.0483	1.483	.0533	1.648	.0499	1.666	.0416	1.624	.0491
		.0211		*.0293*		*.0423*		.0378		*.0431*		*.0380*		.0301		*.0400*
SUR-MG	.9989	.0237	1.001	.0237	1.274	.0449	1.284	.0442	1.266	.0429	1.388	.0486	1.437	.0510	1.374	.0481
		.0229		*.0235*		*.0367*		.0839		*.0368*		*.0383*		.0389		*.0396*
DMG	.9996	.0213	1.001	.0201	.9982	.0277	1.086	.0282	.9999	.0274	1.000	.0311	1.156	.0324	1.000	.0294
		.0208		*.0207*		*.0267*		.0259		*.0260*		*.0298*		.0289		*.0282*
MG-PC	.9989	.0217	1.000	.0206	1.115	.0403	1.140	.0416	1.108	.0370	1.242	.0713	1.312	.0600	1.215	.0620
		.0218		*.0210*		*.0247*		.0271		*.0238*		*.0308*		.0317		*.0295*
CMG	.9996	.0215	1.001	.0207	.9990	.0214	1.002	.0218	1.001	.0216	1.000	.0274	1.084	.0298	1.000	.0256
		.0212		*.0207*		*.0210*		.0211		*.0204*		*.0261*		.0274		*.0251*
Between	1.050	.5580	.9768	.5143	.9983	.5582	1.071	.5604	.9843	.4923	1.005	.5390	1.131	.5434	.9999	.5049
		.5612		*.5191*		*.5453*		.5451		*.5135*		*.5290*		.5286		*.4948*

SM and SSD are the sample mean and standard deviation of $\hat{\beta}$ over 5,000 replications. *SE* is the sample mean of the estimated standard error.

Table 2(ii)
Stationary Panel: $N = 20, T = 30$

	Case a		Case b		Case c		Case $\tilde{c}$		Case d		Case e		Case $\tilde{e}$		Case f	
	SM	SSD	SM	SSD	SM	SSD	SM	SSD	SM	SSD	SM	SSD	SM	SSD	SM	SSD
Estimator		*SE*		*SE*		*SE*		*SE*		*SE*		*SE*		*SE*		*SE*
POLS	.9976	.0407	1.004	.0514	1.450	.0780	1.486	.0847	1.389	.0765	1.601	.0778	1.652	.0730	1.546	.0757
		.0410		*.0507*		*.0366*		.0356		*.0346*		*.0326*		.0309		*.0318*
FE	.9974	.0396	1.003	.0526	1.458	.0781	1.493	.0849	1.395	.0829	1.608	.0776	1.659	.0725	1.553	.0756
		.0405		*.0509*		*.0361*		.0351		*.0342*		*.0322*		.0304		*.0315*
2FE	.9972	.0400	1.002	.0382	.9984	.0434	1.072	.0478	.9990	.0367	1.000	.0472	1.134	.0510	1.001	.0407
		.0412		*.0380*		*.0414*		.0412		*.0364*		*.0414*		.0410		*.0371*
FE-PC	.9975	.0399	1.002	.0381	1.135	.0743	1.213	.0854	1.099	.0557	1.300	.1329	1.400	.1165	1.230	.0997
		.0400		*.0359*		*.0365*		.0357		*.0328*		*.0340*		.0326		*.0318*
MG	.9987	.0536	1.002	.0683	1.503	.0843	1.501	.0800	1.470	.0965	1.649	.0817	1.667	.0689	1.615	.0774
		.0497		*.0676*		*.0593*		.0546		*.0640*		*.0533*		.0439		*.0559*
SUR-MG	.9982	.0581	1.001	.0621	1.390	.0760	1.397	.0768	1.360	.0836	1.518	.0813	1.557	.0745	1.490	.0763
		.0546		*.0619*		*.0569*		.0560		*.0587*		*.0527*		.0489		*.0548*
DMG	.9982	.0513	1.003	.0492	1.000	.0520	1.085	.0550	.9992	.0444	.9994	.0548	1.154	.0575	1.001	.0502
		.0481		*.0464*		*.0508*		.0500		*.0476*		*.0528*		.0515		*.0490*
MG-PC	.9983	.0545	1.002	.0512	1.183	.0871	1.217	.0934	1.149	.0776	1.361	.1443	1.425	.1204	1.301	.1158
		.0523		*.0488*		*.0510*		.0531		*.0494*		*.0526*		.0514		*.0511*
CMG	.9979	.0554	1.001	.0521	.9998	.0513	1.002	.0518	.9964	.0432	.9985	.0540	1.083	.0571	1.002	.0512
		.0510		*.0484*		*.0496*		.0494		*.0482*		*.0522*		.0529		*.0499*
Between	1.017	.4596	1.019	.3894	.9782	.4379	1.074	.4407	.9915	.2814	1.009	.4283	1.116	.4284	1.004	.3748
		.4302		*.3905*		*.4163*		.4194		*.3883*		*.4102*		.4110		*.3771*

See footnote in Table 2(I).

Table 3(i)
Non-Stationary Panel: $N = 30, T = 100$

	Case A		Case B		Case C		Case D		Case $\tilde{\text{D}}$		Case E		Case F		Case G	
	SM	SSD	SM	SSD	SM	SSD	SM	SSD	SM	SSD	SM	SSD	SM	SSD	SM	SSD
Estimator		*SE*		*SE*		*SE*		*SE*		*SE*		*SE*		*SE*		*SE*
POLS	1.000	.0053	1.002	.1545	.9926	.1578	1.111	.1605	1.182	.1757	1.004	.0052	1.109	.1525	1.394	.2109
		.0019		*.0179*		*.0182*		*.0177*		*.0173*		*.0024*		*.0166*		*.0117*
FE	1.000	.0044	1.004	.1046	.9946	.1545	1.397	.1878	1.440	.1903	1.016	.0093	1.362	.1787	1.682	.1534
		.0043		*.0176*		*.0229*		*.0163*		*.0157*		*.0046*		*.0156*		*.0104*
2FE	1.000	.0045	1.005	.1057	1.002	.1144	.9982	.1145	1.071	.1232	1.000	.0045	1.006	.1075	.9945	.0648
		.0044		*.0180*		*.0173*		*.0182*		*.0181*		*.0045*		*.0168*		*.0122*
FE-PC	1.000	.0044	1.004	.1045	.9984	.1169	1.310	.2252	1.373	.2190	1.008	.0059	1.264	.2067	1.400	.2288
		.0043		*.0177*		*.0173*		*.0161*		*.0156*		*.0033*		*.0152*		*.0107*
MG	.9999	.0072	.9992	.1308	.9926	.1779	1.493	.2050	1.503	.1869	1.026	.0125	1.470	.1903	1.900	.1105
		.0067		*.1294*		*.1740*		*.1064*		*.1005*		*.0067*		*.1049*		*.0768*
SUR-MG	1.000	.0081	.9998	.1173	.9940	.1557	1.464	.1938	1.475	.1772	1.008	.0065	1.441	.1787	1.781	.1515
		.0074		*.1159*		*.1528*		*.0989*		*.0939*		*.0050*		*.0975*		*.0682*
DMG	.9999	.0071	.9994	.1336	1.003	.1288	.9922	.1294	1.088	.1367	1.000	.0065	1.008	.1245	.9921	.0830
		.0067		*.1279*		*.1270*		*.1263*		*.1257*		*.0066*		*.1214*		*.0764*
MG-PC	.9999	.0074	.9968	.1077	.9925	.1281	1.468	.2765	1.507	.2403	1.010	.0073	1.424	.2589	1.664	.2275
		.0069		*.1076*		*.1270*		*.1064*		*.0977*		*.0047*		*.1032*		*.0582*
CMG	.9999	.0084	1.002	.1045	.9973	.1012	.9897	.1072	1.002	.1058	.9999	.0075	1.004	.0996	1.000	.0201
		.0079		*.1022*		*.0991*		*.1013*		*.1016*		*.0077*		*.0981*		*.0198*
Between	.9999	.0065	1.002	.1836	.9925	.1869	.9866	.1909	1.067	.2031	1.000	.0063	1.005	.1839	.9958	.1351
		.0062		*.1863*		*.1768*		*.1840*		*.1820*		*.0060*		*.1729*		*.1122*

SM and SSD are the sample mean and standard deviation of $\hat{\beta}$ over 5,000 replications. *SE* is the sample mean of the estimated standard error.

TABLE 3(II)
NON-STATIONARY PANEL: $N = 20, T = 30$

	Case A		Case B		Case C		Case D		Case D̃		Case E		Case F		Case G	
	SM	SSD	SM	SSD	SM	SSD	SM	SSD	SM	SSD	SM	SM	SM	SSD	SM	SSD
Estimator		*SE*		*SE*		*SE*		*SE*		*SE*		*SE*		*SE*		*SE*
POLS	.9998	.0097	1.001	.2041	1.005	.2055	1.058	.1980	1.131	.2216	1.007	.0103	1.045	.2008	1.243	.2022
		.0054		*.0395*		*.0389*		*.0403*		*.0394*		*.0070*		*.0365*		*.0242*
FE	1.001	.0178	1.002	.1387	.9969	.1841	1.390	.1946	1.436	.2082	1.051	.0346	1.357	.1894	1.568	.1790
		.0181		*.0404*		*.0502*		*.0369*		*.0357*		*.0189*		*.0345*		*.0271*
2FE	1.001	.0182	1.003	.1422	.9949	.1355	.9872	.1361	1.070	.1487	.9998	.0184	.9881	.1231	.9982	.0569
		.0186		*.0414*		*.0384*		*.0414*		*.0413*		*.0185*		*.0375*		*.0306*
FE-PC	1.001	.0178	1.003	.1393	.9924	.1408	1.297	.2260	1.366	.2363	1.024	.0233	1.262	.2191	1.314	.2175
		.0181		*.0403*		*.0392*		*.0361*		*.0351*		*.0137*		*.0333*		*.0262*
MG	1.002	.0279	.9999	.1637	.9987	.2288	1.487	.2190	1.497	.2079	1.081	.0462	1.468	.2077	1.726	.1631
		.0277		*.1604*		*.2113*		*.1315*		*.1235*		*.0263*		*.1294*		*.0806*
SUR-MG	1.002	.0300	.9999	.1482	.9990	.2019	1.459	.2067	1.471	.1980	1.051	.0350	1.440	.1964	1.619	.1693
		.0310		*.1454*		*.1873*		*.1231*		*.1160*		*.0244*		*.1206*		*.0734*
DMG	1.003	.0264	1.001	.1614	.9930	.1559	.9897	.1654	1.084	.1647	.9994	.0275	.9887	.1482	1.002	.0674
		.0272		*.1559*		*.1502*		*.1545*		*.1536*		*.0265*		*.1439*		*.0648*
MG-PC	1.002	.0286	1.008	.1427	.9900	.1711	1.460	.2728	1.493	.2592	1.032	.0298	1.423	.2682	1.473	.2508
		.0289		*.1375*		*.1573*		*.1320*		*.1213*		*.0194*		*.1284*		*.0648*
CMG	1.003	.0330	.9998	.1318	.9938	.1333	.9948	.1412	1.001	.1304	.9998	.0335	.9961	.1230	1.004	.0457
		.0329		*.1264*		*.1199*		*.1243*		*.1262*		*.0322*		*.1163*		*.0465*
Between	.9999	.0109	1.002	.2356	1.012	.2341	.9919	.2254	1.068	.2517	.9997	.0103	.9858	.2304	1.004	.1419
		.0102		*.2278*		*.2189*		*.2341*		*.2286*		*.0099*		*.2105*		*.1137*

See footnote in Table 3(I).