A Principal Components Approach to Cross-Section Dependence in Panels

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

[a]Department of Accounting, Finance and Management, University of Essex

[b]Department of Economics, Birkbeck College, University of London

[c]Faculty of Finance, City University Business School

First version: March 2001. This version: March 2002

Abstract

The use of GLS to deal with cross-section dependence in panels is not feasible where N is large relative to T since the disturbance covariance matrix is rank deficient. Neither is it the appropriate response if the dependence results from omitted global variables or common shocks correlated with the included regressors. These can be proxied by the principal components of the residuals from a baseline regression. It is shown that the OLS estimates from a regression augmented by these principal components are unbiased and consistent using sequential limits for large T, large N. Simulations show that this leads to a substantial reduction in bias even for relatively small T and N panels. An empirical application indicates that the impact of cross section dependence seems to strengthen the case for long run PPP.

Keywords: Factor analysis; global shocks; omittted variable bias

JEL Classification: C32; F31

---

*Corresponding author: 7-15 Gresse St., London W1P 1PA, UK Tel: +44 207 6316413. Fax: +44 207 6316418. E-mail: r.smith@econ.bbk.ac.uk.. Earlier versions of this paper were presented at the 7th SCE International Conference on Computing in Economics and Finance, Yale University, June 2001, and at seminars at the Universities of Maastricht, Leeds and Swansea. We thank participants, Kit Baum, Michael Binder, Bertrand Candelon and Martin Sola for helpful discussions. Smith is grateful for support under ESRC grant L138251003.

1 Introduction

In the panel time-series literature, where both the number of groups, N, and the number of time periods T are both large, it is usual to assume the absence of cross section dependence or uncorrelated disturbances across groups. This seems restrictive for many applications in macroeconomics and finance and neglecting it may be far from innocuous for empirical issues such as purchasing power parity (PPP) or whether real exchange rates display reversion towards a long run value (Bai and Ng 2001a; O'Connell 1998; Moon and Perron 2001; Pedroni 1997).[1] In addition, many theoretical panel results have been derived under the assumption of cross section independence (Baltagi and Kao 2000; 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 the presence of cross section dependence, traditional OLS-based estimators are inefficient and the estimated standard errors are biased producing misleading inference. The traditional remedy, SURE-GLS, is not however feasible when the cross section dimension N is of the same order of magnitude as the time series dimension T because the disturbance covariance matrix is rank deficient. Robertson and Symons (1999) propose an innovative method in this context which imposes a factor structure on the residuals to provide a full-rank estimator of the covariance matrix. However, when the non-zero covariances between the errors of different cross section units are due to common omitted variables, it is not obvious that SURE-GLS is always the correct response. If these common omitted variables — say oil prices or global political events in the case where the units are countries — are correlated with the country-specific regressors, both traditional pooled estimators and GLS estimators will be biased and inconsistent. If there is just one common omitted variable to which all cross-section units react homogeneously and if the slope parameters in the model are identical across units, then a two-way fixed effects (FE) estimator may be appropriate. However these conditions are very restrictive.

This paper, in common with a number of recent contributions in panel

[1] While Banerjee, Marcellino and Osbat (2001) do not examine PPP directly, their discussion of the issue of cross unit cointegrating relations is also germane to this debate. Pedroni (1997) also stresses long run cross section dependence.

methods, employs a factor analysis framework.[2] Suppose the parameters of interest are a set of slope coefficients in a baseline regression (RI), the estimates of which will be biased and inconsistent if there are omitted global variables correlated with each country regressor. We propose extracting the largest factors of the RI residuals as proxies for the global unobservable variables. These factors are then included in an augmented regression (RII) to seek to reduce the bias in the RI coefficient estimators. Established panel estimators such as the pooled OLS (POLS), fixed effects (FE) or the Pesaran and Smith (1995) mean group (MG) procedures can be used in each stage. We show analytically for a simple DGP that the POLS estimator of the RII slope coefficient has substantially smaller bias than that from the baseline regression. In addition, it is shown to be consistent using sequential limit asymptotics, as T converges to infinity with a fixed N and then N converges to infinity.

This suggests that the proposed method is especially suited to many large dimension datasets typically used for macroeconomic and financial analysis. This is confirmed by Monte Carlo simulations for the finite sample performance of the method and the choice of the number of factors included. The bias reduction is confirmed using panel dimensions typical of annual and monthly PPP datasets. Simulations show that information criteria based on Bai and Ng (2002) are quite accurate in our context where the factors are extracted from estimated disturbances rather than observed variables. Finally when the method is applied to a PPP data set there is evidence of cross section dependence and handling it with the proposed method seems to reinforce the support for PPP. Throughout this paper we assume that the omitted variables are I(0) and that the regressors and dependent variable are either I(0) or cointegrate. The interesting case of an I(1) omitted variable is a topic for further research.

The proposed estimation method is outlined in §2 and theoretical results for a simple case are given in §3. The Monte Carlo simulations are presented in §4 and an empirical PPP application follows in §5. A final section concludes.

[2] Factor structures for panels are used by Hall, Lazarova and Urga (1999) to test for the number of common stochastic trends and by Bai and Ng (2001a,b) and Moon and Perron (2001) to test for unit roots and cointegration.

2 Cross section dependence

Consider the following baseline regression (RI) model

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\theta}_i + u_{it}, i = 1, 2, ..., N, t = 1, 2, ..., T, \tag{1}$$

where $\mathbf{x}_{it}$ is a $K-$vector of explanatory variables which would typically include an intercept and lags of y_{it}. Suppose the data are generated by

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta}_i + \mathbf{z}'_t\boldsymbol{\gamma}_i + \varepsilon_{it} \tag{2}$$

where the idiosyncratic error ε_{it} is zero-mean white noise distributed independently over units and $\mathbf{z}_t$ is a $J-$vector of unobserved common shocks or random variables which may be correlated with each country regressor. The auxiliary regression $\mathbf{z}'_t = \mathbf{x}'_{it}\mathbf{D}_i + \boldsymbol{\eta}'_{it}$ where $\mathbf{D}_i$ is a $K \times J$ matrix, decomposes $\mathbf{z}'_t$ into two terms, one which is correlated with the ith country regressors and another which is orthogonal to them. Substituting for $\mathbf{z}'_t$ in (2) gives $y_{it} = \mathbf{x}'_{it}(\boldsymbol{\beta}_i + \mathbf{D}_i\boldsymbol{\gamma}_i) + \boldsymbol{\eta}'_{it}\boldsymbol{\gamma}_i + \varepsilon_{it}$ which implies that the omitted $\mathbf{z}'_t$ creates two problems in (1). First $\hat{u}_{it}$ measures $u_{it} = \boldsymbol{\eta}'_{it}\boldsymbol{\gamma}_i + \varepsilon_{it} = \mathbf{z}'_t\boldsymbol{\gamma}_i - \mathbf{x}'_{it}\mathbf{D}_i\boldsymbol{\gamma}_i + \varepsilon_{it}$ which contains $\boldsymbol{\eta}'_{it}$, the part of $\mathbf{z}'_t$ orthogonal to $\mathbf{x}'_{it}$. This term may be serially correlated and will certainly be correlated across units. Both of these will induce a nondiagonal residual covariance matrix or non-spherical disturbances. This will make the OLS estimator $\hat{\boldsymbol{\theta}}_i$ inefficient and its standard error biased, even if $\mathbf{D}_i$ is the null matrix. Second, u_{it} is correlated with the included explanatory variables unless $\mathbf{D}_i = 0$ and this makes $\hat{\boldsymbol{\theta}}_i$, and also the GLS estimator, biased for $\boldsymbol{\beta}_i$ (since $E(\hat{\boldsymbol{\theta}}_i) = \boldsymbol{\beta}_i + \mathbf{D}_i\boldsymbol{\gamma}_i$) and inconsistent.

2.1 Principal components approach

Equation (1) for group i can be written for the N groups as

$$\mathbf{y}'_t = \mathbf{x}'_t\Theta + \mathbf{u}'_t \tag{3}$$

where $\mathbf{y}_t = [y_{1t}, y_{2t}, ..., y_{Nt}]'$ and $\mathbf{x}_t = [\mathbf{x}_{1t}, \mathbf{x}_{2t}, ..., \mathbf{x}_{Nt}]'$ are an $N-$vector and an $NK-$vector, respectively, and Θ is the block-diagonal $NK \times N$ matrix

$$\Theta = \begin{bmatrix} \boldsymbol{\theta}_1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{\theta}_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \boldsymbol{\theta}_N \end{bmatrix},$$

Similarly, (2) can be written for the N groups as $\mathbf{y}_t' = \mathbf{x}_t'\boldsymbol{\beta} + \mathbf{z}_t'\Gamma + \boldsymbol{\varepsilon}_t'$ where Γ is a $J \times N$ matrix. Stacking the $K \times J$ matrices $\mathbf{D}_i$ for the N groups gives the $NK \times J$ matrix $\mathbf{D} = [\mathbf{D}_1'\mathbf{D}_2' \cdots \mathbf{D}_N']'$ and $\mathbf{z}_t' = \mathbf{x}_t'\mathbf{D} + \boldsymbol{\eta}_{it}'$. We have

$$\mathbf{y}_t' = \mathbf{x}_t'[\boldsymbol{\beta} + \mathbf{D}\Gamma] + (\mathbf{z}_t' - \mathbf{x}_t'\mathbf{D})\Gamma + \boldsymbol{\varepsilon}_t'$$

and it follows that if we estimate (3) by OLS, then $E(\hat{\Theta}) = \Theta = \boldsymbol{\beta} + \mathbf{D}\Gamma$. The residuals measure $\mathbf{u}_t' = (\mathbf{z}_t' - \mathbf{x}_t'\mathbf{D})\Gamma + \boldsymbol{\varepsilon}_t'$ which can be written as $\mathbf{u} = Z\Gamma - XD\Gamma + \boldsymbol{\varepsilon}$ for the T time periods, where Z is $T \times J$, X is $T \times NK$ and $\boldsymbol{\varepsilon}$ and $\mathbf{u}$ are $T \times N$ matrices. Post-multiplying the latter by $\tilde{\mathbf{u}} = \text{diag}\,(u_i'u_i)^{-1/2}$ gives the $T \times N$ matrix of standardized errors $\underline{\mathrm{u}} = Z\tilde{\Gamma} - XD\tilde{\Gamma} + \tilde{\boldsymbol{\varepsilon}}$ where $\tilde{\Gamma} = \Gamma\tilde{\mathbf{u}}$ and $\tilde{\boldsymbol{\varepsilon}} = \boldsymbol{\varepsilon}\tilde{\mathbf{u}}$. Since Z is unobserved, one needs to impose J^2 normalising restrictions on $\tilde{\Gamma}$ to provide estimates of Z. For this purpose assume $\tilde{\Gamma}\tilde{\Gamma}' = I$. With this normalisation one can write $\underline{\mathrm{u}}\tilde{\Gamma}' = Z - XD + \tilde{\boldsymbol{\varepsilon}}\tilde{\Gamma}'$ or, equivalently

$$Z = \underline{\mathrm{u}}\tilde{\Gamma}' + XD - \tilde{\boldsymbol{\varepsilon}}\tilde{\Gamma}' \tag{4}$$

This suggests measuring Z by $W = \underline{\mathrm{u}}A$, the N principal components of $\underline{\mathrm{u}}$ obtained via the spectral decomposition of $R = \underline{\mathrm{u}}'\underline{\mathrm{u}} = A\Lambda A'$, where A is the orthogonal matrix of eigenvectors and Λ is the diagonal eigenvalue matrix. If a few random factors, W_J, account for most of the disturbances covariation then the cross-section dependence can be characterized by means of a factor model $\underline{\mathrm{u}} = W_J A_J' + E$. The $N \times J$ matrix A_J (non-random factor loadings) contains the $J < N$ eigenvectors associated with the largest eigenvalues and E is a $T \times N$ idiosyncratic error matrix. This suggests an augmented regression (RII) for handling the cross-section dependence that biases the estimators of the regression of interest (RI). This is

$$y_{it} = \mathbf{x}_{it}'\mathbf{b}_i + \mathbf{w}_t'c_i + v_{it} \tag{5}$$

where $\mathbf{w}_t$ is a $J-$vector of principal components from the RI errors.

One issue is how to determine J and, relatedly, how well the factors W_J proxy the unobserved variables Z. Another issue regards interpreting the factors because the identifying assumptions $\tilde{\Gamma}\tilde{\Gamma}' = I$ need not be meaningful from an economic viewpoint. However, for a reasonable small J it may be possible to give them an economic or financial interpretation.

Our approach has some commonalities with SURE-GLS. First, it resembles the latter in that (5) includes linear combinations of OLS residuals.[3] However, our procedure is distinctive to the extent that it includes the own residual. This results in endogeneity bias which falls with N as shown in §3.2. For small N a slight modification of (5) can circumvent the endogeneity problem. For each group i, the $T \times (N-1)$ residual matrix $u_{\bar{\imath}} = [u_1, ..., u_{i-1}, u_{i+1}, ..., u_N]$ which excludes group i residuals is used to extract the first J factors, $\mathbf{w}_{\bar{\imath}t}$, and $y_{it} = \mathbf{x}_{it}'\mathbf{b}_i + \mathbf{w}_{\bar{\imath}t}'c_i + v_{it}$ is estimated.[4] Second, since our proxies for the common shocks are calculated as linear combinations of OLS residuals — by construction $\hat{u}_{it}$ is orthogonal to $\mathbf{x}_{it}$ although not necessarily to $\mathbf{x}_{jt}$ for $j \neq i$ — this suggests that the smaller the cross-section correlations among regressors the closer the factors W_J will be to Z and, hence, the more gains are expected from our approach in terms of bias reduction. Conversely, if $\mathbf{x}_{1t} = \mathbf{x}_{2t} = ... = \mathbf{x}_{Nt}$ then the inclusion of W_J will not improve the properties of the estimator $\hat{\mathbf{b}}_i$ in RII (over $\hat{\boldsymbol{\theta}}_i$ in RI) which will be still biased and inconsistent. This is similar to the situation where for identical regressors there are no efficiency gains from SURE-GLS over equation-by-equation OLS.

3 Analytical results

Consider a simple data generating process (DGP) comprising a country specific regression, say a PPP equation, in which a global variable z_t, such as oil prices, is omitted but where it also influences the country specific regressors x_{it}. For instance, oil prices could influence inflation differentials because country-specific inflation differs in its response to oil prices depending on whether the country imports or exports oil. Suppose data are generated by

$$\begin{aligned} x_{it} &= d_{it} + z_t \\ y_{it} &= \beta x_{it} + \gamma z_t + \varepsilon_{it}, \varepsilon_{it} \sim iid(0, \sigma_\varepsilon^2) \end{aligned} \tag{6}$$

where the innovations ε_{it} are uncorrelated across countries. We assume that each regressor has an *idiosyncratic* (or country-specific) and *common* influence, $d_{it} \sim iid(0, \sigma_d^2)$ and $z_t \sim iid(0, \sigma_z^2)$, respectively, which are orthogonal to each other and also to ε_{it}.

[3] Telser (1964) suggested an iterative approach to account for the cross-equation residual correlation which converges to an estimator with the same asymptotic properties as Zellner's SURE-GLS estimator. This consists of including as additional variables in each equation the OLS residuals of all other equations.

[4] An alternative approach to abate the small N endogeneity bias would be to use the IV method, that is, to instrument the factors $\mathbf{w}_t$ in (5).

Suppose data are available on x_{it} and x_{it} and we focus on the regressions

$$\begin{array}{lll} y_{it} = \theta x_{it} + u_{it} & (\mathbf{RI}) \\ y_{it} = b x_{it} + c w_t + v_{it} & (\mathbf{RII}) \end{array} \tag{7}$$

which differ in that the latter is augmented by $w_t = \sum_{i=1}^{N} u_{it}/N$ to proxy the unobserved random factor z_t. It can be shown that, for our baseline DGP, this proxy is equal to the first principal component of $\mathbf{u}$ up to a scaling factor. To study the properties of these regressions, we define the sequential probability limits (plim) for $T \to \infty$ and any fixed N

$$S_{yy}^N = \operatorname*{plim}_{T\to\infty} T^{-1} \textstyle\sum_t y_{it}^2; \quad S_{yx}^N = \operatorname*{plim}_{T\to\infty} T^{-1} \sum_t y_{it} x_{it} \tag{8}$$

and likewise for the other variables in (6) whose (co)variances are:

$$\begin{array}{ll}
E(x_{it}^2) = \sigma_x^2 = \sigma_d^2 + \sigma_z^2 & E(d_{it}^2) = \sigma_d^2 \\
E(y_{it}^2) = \beta^2\sigma_d^2 + (\beta+\gamma)^2\sigma_z^2 + \sigma_\varepsilon^2 & E(\varepsilon_{it}^2) = \sigma_\varepsilon^2 \\
E(x_{it}y_{it}) = \beta\sigma_d^2 + (\beta+\gamma)\sigma_z^2 & E(x_{it}x_{jt}) = \sigma_z^2 \\
E(y_{it}z_t) = (\beta+\gamma)\sigma_z^2 & E(x_{it}d_{it}) = \sigma_d^2 \\
E(y_{it}y_{jt}) = (\beta+\gamma)^2\sigma_z^2 & E(x_{it}z_t) = \sigma_z^2 \\
E(z_t d_{it}) = E(z_t\varepsilon_{it}) = E(d_{it}\varepsilon_{it}) = 0 & E({z_t}^2) = \sigma_z^2
\end{array}$$

The auxiliary regression $z_t = \delta x_{it} + \eta_{it}$ implies that $y_{it} = (\beta+\gamma\delta)x_{it} + \gamma\eta_{it} + \varepsilon_{it}$. Hence, the OLS estimator $\hat{\theta}$ measures $(\beta+\gamma\delta)$ and the residuals $\hat{u}_{it}$ estimate $u_{it} = \gamma\eta_{it} + \varepsilon_{it}$, with variance $\sigma_u^2 \equiv E(u_{it}^2) = \gamma^2(1-\delta)^2\sigma_z^2 + \gamma^2\delta^2\sigma_d^2 + \sigma_\varepsilon^2$ and covariance $\sigma_{ij} \equiv E(u_{it}u_{jt}) = \gamma^2(1-\delta)^2\sigma_z^2$. Then we have $w_t = \sum_{i=1}^{N}(\gamma\eta_{it} + \varepsilon_{it})/N = \gamma(1-\delta)z_t - \gamma\delta\bar{d}_t + \bar{\varepsilon}_t$ where $\bar{d}_t = N^{-1}\sum_i d_{it}$ and it follows that

$$\begin{array}{rcl}
E(w_t^2) & = & \gamma^2(1-\delta)^2\sigma_z^2 + \gamma^2\delta^2 N^{-1}\sigma_d^2 + N^{-1}\sigma_\varepsilon^2 \\
E(y_{it}w_t) & = & (\beta+\gamma)(1-\delta)\gamma\sigma_z^2 - \beta\gamma\delta N^{-1}\sigma_d^2 + N^{-1}\sigma_\varepsilon^2 \\
E(x_{it}w_t) & = & \gamma(1-\delta)\sigma_z^2 - \gamma\delta N^{-1}\sigma_d^2; \quad E(z_t w_t) = \gamma(1-\delta)\sigma_z^2
\end{array}$$

For our baseline DGP we have

$$\begin{array}{ll}
S_{yy}^N = \beta^2\sigma_d^2 + (\beta+\gamma)^2\sigma_z^2 + \sigma_\varepsilon^2 & S_{xx}^N = \sigma_d^2 + \sigma_z^2 \\
S_{ww}^N = \gamma^2(1-\delta)^2\sigma_z^2 + \gamma^2\delta^2 N^{-1}\sigma_d^2 + N^{-1}\sigma_\varepsilon^2 & S_{xz}^N = \sigma_z^2 \\
S_{xw}^N = \gamma[(1-\delta)\sigma_z^2 - N^{-1}\delta\sigma_d^2] & S_{yx}^N = \beta\sigma_d^2 + (\beta+\gamma)\sigma_z^2 \\
S_{yw}^N = \gamma[(\beta+\gamma)(1-\delta)\sigma_z^2 - N^{-1}\beta\delta\sigma_d^2] + N^{-1}\sigma_\varepsilon^2 & S_{zw}^N = \gamma(1-\delta)\sigma_z^2
\end{array}$$

These results are used in the next sections to analyze the question of how well does the OLS estimator $\hat{b}$ (and by comparison $\hat{\theta}$) measure the true β.

3.1 Baseline regression (RI)

RI is misspecified due to an unobserved global effect z_t which is correlated with each country regressor. It follows that the POLS estimator $\hat{\theta} = (X'X)^{-1}X'Y = \theta + (X'X)^{-1}X'u$ is biased for the true parameter since $E(\hat{\theta}) = \theta = \beta + \gamma\delta$. The plim of $\hat{\theta}$ as $T \to \infty$ for any fixed N is

$$\underset{T\to\infty}{\text{plim}}\, \hat{\theta} = \frac{N^{-1}\sum_i S_{yx}^N}{N^{-1}\sum_i S_{xx}^N} = \frac{\beta\sigma_d^2 + (\beta+\gamma)\sigma_z^2}{\sigma_d^2 + \sigma_z^2} = \beta + \gamma\delta$$

and letting $N \to \infty$ subsequently, results in $\text{plim}_{N,T\to\infty}\, \hat{\theta} = \theta = \beta + \gamma\delta$. Hence, $\hat{\theta}$ is inconsistent for β for both large T and N. Its variance is

$$var(\hat{\theta}) \equiv E\left[(\hat{\theta}-\theta)(\hat{\theta}-\theta)'\right] = E\left[(X'X)^{-1}X'uu'X(X'X)^{-1}\right] \tag{9}$$

or $var(\hat{\theta}) = (X'X)^{-1}X'E(uu')X(X'X)^{-1}$ for u_{it} orthogonal to x_{jt}. Assuming spherical disturbances or $E(uu') = \sigma_u^2 I_{NT} = \Sigma \otimes I_T$ with $\Sigma = \sigma_u^2 I_N$, we have $var(\hat{\theta}) = \sigma_u^2(X'X)^{-1}$. However, for (6) the disturbances of RI contain a random omitted variable which makes the latter inappropriate on two accounts.

First, the contemporaneous covariance matrix has the following structure

$$\Sigma = E(u_t u_t') = \begin{bmatrix} \sigma_u^2 & \sigma_{ij} & \cdots & \sigma_{ij} \\ \sigma_{ij} & \sigma_u^2 & & \vdots \\ \vdots & & \ddots & \sigma_{ij} \\ \sigma_{ij} & \cdots & \sigma_{ij} & \sigma_u^2 \end{bmatrix}$$

since the errors are groupwise homoskedastic $E(u_{it}^2) = \sigma_u^2$ and have equal covariances $E(u_{it}u_{jt}) = \sigma_{ij}$. It follows that

$$var(\hat{\theta}) = (\sum_{i=1}^{N} X_i'X_i)^{-1}(\sum_{i=1}^{N}\sum_{j=1}^{N} \sigma_{ij}X_i'X_j)(\sum_{i=1}^{N} X_i'X_i)^{-1} \tag{10}$$

which for our DGP particularizes to

$$\begin{aligned} var(\hat{\theta}) &= \frac{\sigma_u^2}{\sum_t\sum_i(x_{it}-\bar{x})^2} + \sigma_{ij}\frac{\sum_i\sum_{j\neq i}\left[\sum_t(x_{it}-\bar{x}_i)(x_{jt}-\bar{x}_j)\right]}{\left[\sum_t\sum_i(x_{it}-\bar{x})^2\right]^2} \\ &= \frac{\sigma_u^2}{NT\sigma_x^2} + \sigma_{ij}\frac{(N-1)\sigma_{x,ij}}{NT(\sigma_x^2)^2} \end{aligned} \tag{11}$$

where $\sigma_x^2 = \sum_i \sum_t (x_{it} - \bar{x})^2/NT$ and $\sigma_{x,ij} = \sum_i \sum_t (x_{it} - \bar{x}_i)(x_{jt} - \bar{x}_j)/NT$. Since $\hat{\theta}$ is consistent for θ, $var(\hat{\theta})$ can be consistently estimated by substituting $\hat{\sigma}_u^2 = \hat{u}'\hat{u}/(NT-k)$ where k is the number of freely estimated parameters in RII and $\hat{\sigma}_{ij} = \frac{\sum_i \sum_{j\neq i} \hat{u}_i'\hat{u}_j/T}{N(N-1)}$ for σ_u^2 and σ_{ij}, respectively.

Second, although u_{it} is orthogonal to x_{it} by construction, it is correlated with x_{jt} for $j \neq i$.[5] Hence (11) is still incorrect in assuming that elements of the form $E(x_{it}u_{it}x_{jt}u_{jt})$ in $E(X'uu'X)$ are equal to $E(x_{it}x_{jt})E(u_{it}u_{jt})$ or, more generally, that $E(X'uu'X) = X'E(uu')X$ in (9). The appropriate asymptotic covariance matrix is

$$var(\hat{\theta}) = \frac{\sigma_u^2}{NT\sigma_x^2} + \frac{(N-1)\phi}{NT(\sigma_x^2)^2} \tag{12}$$

where $\phi = E(x_{it}u_{it}x_{jt}u_{jt})$. An estimator for $var(\hat{\theta})$ is straightforward to compute using $\hat{\sigma}_u^2$ and $\hat{\phi} = \sum_{i=1}^N \sum_{j\neq i}^N (T^{-1} \sum_{t=1}^T \tilde{x}_{it}\hat{u}_{it}\tilde{x}_{jt}\hat{u}_{jt})/N(N-1)$ where $\tilde{x}_{it} = x_{it} - \bar{x}_i$ and $\tilde{x}_{jt} = x_{jt} - \bar{x}_j$. The first term in (12) is the usual POLS variance formula and the second term may be viewed as a correction for cross-section dependence of residuals, $cov(\hat{u}_{it}, \hat{u}_{jt}) \neq 0$, and dependence between residuals and regressors for different units, $cov(\hat{u}_{it}, x_{jt}) \neq 0$ arising from an omitted global variable or shock z_t. Consequently, POLS not only gives biased and inconsistent (for both $N, T \to \infty$) estimators for RI but also biased standard errors. Formula (12) suggests that the bias of the latter is of order T^{-1} and hence it will be non-negligible for small T even if N is large.

In a more general K-regressor setup, $Y = X\Theta + u$, the appropriate asymptotic covariance matrix of the POLS estimator $\hat{\Theta}$ is

$$var(\hat{\Theta}) = (X'X)^{-1}E(X'uu'X)(X'X)^{-1} = (X'X)^{-1}\Pi(X'X)^{-1} \tag{13}$$

where Π is a $K \times K$ matrix with elements

$$\pi_{pp} = NT\sigma_{x_p}^2\sigma_u^2 + N(N-1)T\phi_{pp},\ \pi_{pq} = NT\sigma_{x_{pq}}\sigma_u^2 + N(N-1)T\phi_{pq}$$

for $p, q = 1, ..., K$, where $\sigma_{x_p}^2 = \sum_{i=1}^N \sum_{t=1}^T (x_{p,it} - \bar{x}_p)^2/NT$ and $\sigma_{x_{pq}} = \sum_{i=1}^N \sum_{t=1}^T (x_{p,it} - \bar{x}_p)(x_{q,it} - \bar{x}_q)/NT$. A consistent estimator $\widehat{var}(\hat{\Theta})$ can

[5] For our simple DGP $u_{it} = \gamma(z_t - \delta x_{it}) + \varepsilon_{it}$ with $\delta = \sigma_z^2/(\sigma_z^2 + \sigma_d^2)$. It follows that $cov(u_{jt}, x_{jt}) = -\gamma\delta\sigma_d^2 + \gamma(1-\delta)\sigma_z^2 = 0$. However, $cov(u_{jt}, x_{it}) = \gamma(1-\delta)\sigma_z^2 \neq 0$.

be obtained substituting

$$\hat{\phi}_{pp}=\frac{\sum_i\sum_{j\neq i}(T^{-1}\sum_t\tilde{x}_{p,it}\hat{u}_{it}\tilde{x}_{p,jt}\hat{u}_{jt})}{N^2},\ \hat{\phi}_{pq}=\frac{\sum_i\sum_{j\neq i}(T^{-1}\sum_t\tilde{x}_{p,it}\hat{u}_{it}\tilde{x}_{q,jt}\hat{u}_{jt})}{N^2}$$

for ϕ_{pp} and ϕ_{pq}, respectively.

3.2 Augmented regression (RII)

We start by asking how well w_t proxies the unobserved global variable z_t. The plim of their squared sample correlation as $T\to\infty$ for any fixed N is

$$\underset{T\to\infty}{\text{plim}}\,\hat{\rho}^2_{zw}=\frac{(S^N_{zw})^2}{S^N_{zz}S^N_{ww}}=\frac{\sigma^2_z}{\sigma^2_z+N^{-1}\delta^2\tau^2\sigma^2_d+N^{-1}\gamma^{-2}\tau^2\sigma^2_\varepsilon}$$

where $\tau=(1-\delta)^{-1}$. Letting $N\to\infty$ also, we have $\text{plim}_{T,N\to\infty}\,\hat{\rho}^2_{zw}=1$ and it follows that for large N and T the first factor for RI residuals is a consistent estimator of z_t. The plim of $\hat{b}$ as $T\to\infty$ for any fixed N is

$$\begin{aligned}\underset{T\to\infty}{\text{plim}}\,\hat{b} &= \frac{(N^{-1}\sum_i S^N_{ww})(N^{-1}\sum_i S^N_{xy})-(N^{-1}\sum_i S^N_{xw})(N^{-1}\sum_i S^N_{yw})}{(N^{-1}\sum_i S^N_{xx})(N^{-1}\sum_i S^N_{ww})-(N^{-1}\sum_i S^N_{xw})^2} \qquad (14)\\ &= \frac{\beta\gamma^2(1-\delta)^2\sigma^2_z\sigma^2_d+N^{-1}f_1+N^{-2}g_1}{\gamma^2(1-\delta)^2\sigma^2_z\sigma^2_d+N^{-1}f_2+N^{-2}g_2}\end{aligned}$$

where

$$\begin{aligned}f_1 &= \gamma^2\delta\beta\sigma^4_d+(2\gamma^2\beta\delta-\gamma^2\beta\delta^2+\gamma^3\delta)\sigma^2_d\sigma^2_z+(\beta+\gamma\delta)\sigma^2_\varepsilon\sigma^2_z+\beta\sigma^2_\varepsilon\sigma^2_d,\\ f_2 &= \gamma^2\delta^2\sigma^4_d+(2\gamma^2\delta-\gamma^2\delta^2)\sigma^2_d\sigma^2_z+\sigma^2_\varepsilon\sigma^2_z+\sigma^2_\varepsilon\sigma^2_d),\\ g_2 &= -\gamma^2\delta^2\sigma^4_d \text{ and } g_1=\gamma\delta\sigma^2_\varepsilon\sigma^2_d-\gamma^2\delta^2\beta\sigma^4_d\end{aligned}$$

Making $N\to\infty$ also it follows that $\hat{b}$ is a consistent estimator for β since

$$\underset{N,T\to\infty}{\text{plim}}\,\hat{b}=\frac{\underset{N\to\infty}{\text{plim}}\,(\beta\gamma^2(1-\delta)^2\sigma^2_z\sigma^2_d+N^{-1}f_1+N^{-2}g_1)}{\underset{N\to\infty}{\text{plim}}\,(\gamma^2(1-\delta)^2\sigma^2_z\sigma^2_d+N^{-1}f_2+N^{-2}g_2)}=\beta$$

Some remarks are in order. First, the inconsistency of $\hat{b}$ for fixed N reflects the endogeneity bias which arises because the principal components are averages of RI disturbances, $w_t = \sum_i a_i u_{it}$ where $u_{it} = \gamma\eta_{it} + \varepsilon_{it}$. Hence, there is correlation between w_t and RII disturbances which also contain ε_{it}. However, the weight of ε_{it} in the former (a_i) falls as N increases and so does the endogeneity bias. Second, since $corr(x_{it}, x_{jt}) = \sigma_z^2/(\sigma_z^2 + \sigma_d^2)$, if we let $\sigma_d^2 \to 0$ then the regressors are identical across equations. In this case $\text{plim}_{T,N\to\infty}\,\hat{b} = \beta + \gamma\delta$ and there are no gains from RII over RI. Third, although the foregoing theoretical results (large T, large N consistency of $\hat{b}$) are derived for a homogeneous panel with identical β and γ across units, we conjecture that similar results apply to heterogenous panels. This is investigated via Monte Carlo simulations in §4.

The variance of the POLS estimator of $a = [b\ c]'$ in RII is

$$var(\hat{a}) \equiv E\left[(\hat{a} - a)(\hat{a} - a)'\right] = (X'X)^{-1}E\left(X'vv'X\right)(X'X)^{-1} \tag{15}$$

where X is an $NT \times 2$ matrix. As shown above $\hat{\rho}_{zw}$ converges in probability to 1 as $T \to \infty$ and $N \to \infty$. However, to the extent that w_t is not a perfect measure of the unobserved z_t in finite samples, there will be cross-equation dependence in the disturbances, $E(v_{it}, v_{jt}) \neq 0$, and between disturbances and regressors $E(v_{it}, x_{jt}) \neq 0$ for $i \neq j$. Hence $E(X'vv'X) \neq X'E(vv')X$. This suggests estimating $var(\hat{a})$ using an analogous formula to (13).

3.3 Serial dependence and heteroskedasticity

The assumption that d_{it} and z_t are iid processes is relaxed by letting

$$\begin{aligned} d_{it} &= \rho_d d_{i,t-1} + \varepsilon_{d,it},\ \ \varepsilon_{d,it} \sim iid(0, \sigma_d^2) \\ z_t &= \rho_z z_{t-1} + \varepsilon_{z,t},\ \ \varepsilon_{z,t} \sim iid(0, \sigma_z^2) \end{aligned}$$

with $-1 < \rho_d, \rho_z < 1$. This introduces the serial correlation typical of economic variables in the disturbances of RI (and RII). In particular, the matrix $E(X'vv'X)$ in (15) will have also nonzero elements $E(x_{it}v_{it}x_{js}v_{js})$ for $t \neq s$. By letting $\varepsilon_{d,it} \sim iid(0, \sigma_{d,i}^2)$ the heterogeneous variance of the idiosyncratic influence in x_{it} introduces groupwise heteroskedasticity in RI and RII errors. In particular, for RI we have

$$E(u_{it}^2) = \gamma^2(1-\delta)^2\sigma_z^2 + \gamma^2\delta^2\sigma_{d,i}^2$$

By rewriting (6) as $y_{it} = \beta x_{it} + \gamma_i w_t + \gamma_i(z_t - w_t) + \varepsilon_{it}$ it follows that $v_{it} = \gamma_i(z_t - w_t) + \varepsilon_{it}$ in RII. Hence to the extent that $w_t \neq z_t$, letting $\gamma_i \neq \gamma_j$ will also introduce heteroskedasticity in RII residuals.

We conjecture that the unbiasedness and consistency of $\hat{b}$, the (POLS) estimator of RII slope coefficient, carry over to more complex setups. However, an appropriate asymptotic $var(\hat{b})$ is needed. The latter has to account not only for contemporaneous dependences — such as $cov(\hat{v}_{it}, \hat{v}_{jt})$ and $cov(\hat{v}_{it}, x_{jt})$ which arise in finite samples to the extent that w_t does not perfectly measure z_t — but also serial dependence and groupwise heteroskedasticity.

4 Monte Carlo Analysis

Monte Carlo simulations are employed to assess the finite sample properties of the proposed estimation approach. The following DGP is used

$$\begin{aligned} d_{it} &= \rho_{di} d_{i,t-1} + \varepsilon_{di,t}; \varepsilon_{di,t} \sim NID(0, \sigma^2_{d,i}) \\ z_t &= \rho_z z_{t-1} + \varepsilon_{zt}; \varepsilon_{zt} \sim NID(0, \sigma^2_z) \\ x_{it} &= d_{it} + \lambda_i z_t \\ y_{it} &= \beta_i x_{it} + \gamma_i z_t + \varepsilon_{it}; \varepsilon_{it} \sim NID(0, \sigma^2_{\varepsilon,i}) \end{aligned}$$

In the first set of simulations, reported in Table 1, it is assumed $\rho_{di} = \rho_z = 0, \sigma^2_z = \sigma^2_{d,i} = \sigma^2_{\varepsilon,i} = 1$ and $\lambda_i = \beta_i = \gamma_i = 1$. In a second set, reported in Table 2, we assume $\sigma^2_{d,i} \sim U[0.5, 1.5]$ and $\gamma_i \sim U[0.5, 1.5]$ to introduce groupwise heteroskedasticity and $\rho_{di} = \rho_z = 0.9$ to introduce serial correlation. The panel dimensions $N, T = \{30, 300\}$ and $N, T = \{30, 25\}$ which are typical of monthly and annual PPP studies, respectively, are used. $R = 500$ replications are employed throughout. We estimate RI and RII — where w_t is the first factor extracted from the equation-by-equation OLS residuals of RI — using the POLS, FE and MG estimators.[6]

The comparison focuses on the bias of the slope estimator and on the difference between the true standard errors (s.e.) and various estimates of them. The former is measured by the sample standard deviation (SSD) of the slope estimates over replications while the latter is measured by the sample mean (SM) of the estimated s.e. For comparative purposes, Table 1 reports three s.e. estimates for the pooled estimators. These are SE_1 based on the conventional formula $s^2(X'X)^{-1}$, SE_2 using (10) and SE_3 based on (13).[7]

[6] All computations were performed in GAUSS version 3.2.32.

[7] For the MG estimator $SE(\hat{\theta}^{MG}) = \sigma(\hat{\theta}_i)/\sqrt{N}$ where $\sigma(\hat{\theta}_i)$ is the standard deviation of the individual OLS estimates.

The mean slope estimates for RI reported in Table 1(A-B) are in line with the theoretical bias at 0.5 for the baseline DGP. Those for RII reveal that the proposed approach succeeds in reducing bias. As expected for RI, SE_1 underestimates the true s.e. because of neglected non-zero $cov(\hat{u}_{it}, \hat{u}_{jt})$ and $cov(\hat{u}_{it}, x_{jt})$ caused by z_t.[8] The SE_2 underestimate the true s.e. also because of failing to account for $cov(\hat{u}_{it}, x_{jt}) \neq 0$. By contrast, the SM($SE_3$) matches the SSD($\hat{\theta}$) quite well. However, this is not the case for RII where there is still some underestimation in SE_3 particularly for the annual panel. The reasons underlying this bias warrant further investigation.

If our baseline DGP is modified to have a common global influence orthogonal to the regressors by letting $x_{it} = d_{it} + \lambda_i z_{2t}$ and $y_{it} = \beta_i x_{it} + \gamma_i z_{2t} + \varepsilon_{it}$ where $cov(z_{2t}, z_{1t}) = 0$ *ceteris paribus,* it follows that $cov(\hat{u}_{it}, x_{jt}) = cov(\hat{v}_{it}, x_{jt}) = 0$. Unsurprisingly, simulations show that SM(SE_2) and SM(SE_3) are reasonably close for both RI and RII. Again these estimators match the true s.e. for RI (and SE_1 underestimates it) and they are biased downwards for RII. For a simpler DGP where $\lambda_i = 0$ *ceteris paribus*, $SE_1 \simeq SE_2 \simeq SE_3$ are correct in RI. The latter can be explained by the fact that, although $cov(\hat{u}_{it}, \hat{u}_{jt}) \neq 0$ the second term in (11) and (12) vanishes because $cov(x_{it}, x_{jt}) = 0$. This result (correct SE_1) also emerges in RII.[9] These experiments suggest that there is an additional effect in RII (over RI) not captured by any of the covariance matrices considered and which becomes apparent when $cov(x_{it}, x_{jt}) \neq 0$.

When the assumptions of groupwise homoskedastic and non autocorrelated errors are relaxed, the proposed approach continues to reduce bias substantially as Table 2(A-B) shows. Unsurprisingly, for this DGP none of the available covariance matrices provides accurate estimates of the true standard errors, not even for RI. Formula (13) fails to account for the autocorrelation pattern while the Newey-West covariance estimator for panels does not account for the cross-equation correlations $cov(\hat{u}_{it}, \hat{u}_{jt})$ and $cov(\hat{u}_{it}, x_{jt})$.[10] Finally the baseline DGP is modified to introduce slope coefficient hetero-

[8] The biased SE_1 for the MG estimator may stem from non-zero covariances between the coefficient estimates for each group driven by a common bias term $\delta\gamma$.

[9] In the case where the unobservable global variables are uncorrelated with the regressors our approach offers efficiency gains — borne out by SSD($\hat{b}$) $\ll$SSD($\hat{\theta}$) — like SURE-GLS. However, our approach has the additional advantage that no restriction is imposed on the relation between N and T.

[10] The s.e. from (13) and Newey-West (L=2) are .0561(.0277) and .0408(.0266) for FE and POLS, respectively, for RI and .0285(.0215) and .0197(.0148) for RII (annual panel).

geneity and differences in the response of cross-section units to the common influence by letting $\beta_i \sim U[0.5, 1.5]$, $\lambda_i \sim U[0.5, 1.5]$ and $\gamma_i \sim U[0.5, 1.5]$. The mean slope estimate over replications for RI(RII) is 1.492(1.081), 1.493(1.082) and 1.517(1.118) for the POLS, FE and MG estimators, respectively. These results support our conjecture that the proposed approach is suitable for heterogeneous panels also.

There remain two unresolved questions regarding the standard errors. One is to ascertain why SE_3 are incorrect in RII while they work quite well in RI for our baseline DGP. The other is to derive the theoretical covariance matrix for the estimators of RII in a DGP with autocorrelation. One possible solution is to implement a bootstrap technique. However, the consistency (or otherwise) of the bootstrap s.e. in our setup may well depend on the answer to the first question.

We investigate the number of factors problem via simulations. We examine the rule of thumb (known as Kaiser criterion) that only those factors whose associated eigenvalues $\lambda_i > 1$ are retained. The intuition for the latter is that — since the factors are extracted from standardized residuals — unless a factor extracts as much variation as one original variable it is dropped. For our baseline DGP the number of factors thus chosen ranges between 9 and 12 (with mean 10.61 and standard deviation .027) for the annual data panel and between 7 and 11 (mean 9.07 and standard deviation .028) for the monthly panel over 500 replications. Hence the Kaiser criterion is too conservative and this raises issues of interpretation in an economic context.[11]

Next we explore the performance of the recent Bai and Ng (2002) approach for selecting the number of factors in approximate factor models. They formulate this non-standard problem as a model selection problem and propose minimizing, inter alios, the following two information criteria

$$IC_{p_1}(\tau) = \ln V(\tau, \hat{W}_\tau) + \tau \left(\frac{N+T}{NT}\right) \ln \left(\frac{NT}{N+T}\right) \tag{16}$$

$$IC_{p_2}(\tau) = \ln V(\tau, \hat{W}_\tau) + \tau \left(\frac{N+T}{NT}\right) \ln C_{NT}^2 \tag{17}$$

where $C_{NT}^2 = \min\{N, T\}$ and $V(\tau, \hat{W}_\tau) = \hat{e}'\hat{e}/NT$ is the average residual variance of a factor model where τ factors are assumed for each cross-section

[11] Nevertheless, including too many factors does not seem to have an adverse effect on bias reduction. This is borne out by the sample mean of $\hat{b}$ at some 1.069 (POLS) and 1.072 (FE) for the annual panel and 1.096 (POLS, FE) for the monthly panel.

unit. They demonstrate the consistency of these and other criteria and their simulations show that they are fairly robust provided $\min\{N,T\} > 40$. One important difference between our setup and that in Bai and Ng is that our factors are extracted from residuals rather than observed variables such as returns. Accordingly, we considered two cases for $V(\tau, \hat{W}_\tau)$. One is based on the equation-by-equation OLS residuals of the $\tau-$factor model $u = W_\tau A'_\tau + E$. A second version which is more in the spirit of our approach is based on the equation-by-equation OLS residuals of RII. Setting as $\tau_{\max}$ the number of factors dictated by Kaiser criterion and using the two DGPs described in the foregoing analysis (where the true number of factors is J=1) the outcome of these criteria is averaged over 500 replications. In all cases the criteria systematically choose one factor for both the small and large panels.

Next our DGP is generalized to $J \geq 1$ omitted global variables *ceteris paribus* by letting

$$x_{it} = d_{it} + \sum_{j=1}^{J} \lambda_{ij} z_{jt};\ y_{it} = \beta_i x_{it} + \sum_{j=1}^{J} \gamma_{ij} z_{jt} + \varepsilon_{it}$$

where z_{jt} for $j = 1, ..., J$ are orthogonal to each other and λ_{ij} and γ_{ij} are $N(1,1)$ variates. Setting $J = 2$ the average of $\hat{\tau}$ over replications for the annual panel dimensions is 2.052 (with standard deviation of 0.4536) for the baseline DGP and 4.146 (2.375) for the DGP with autocorrelation and heteroskedasticity. For the monthly panel these figures are 2 and 2.889 (1.881), respectively.[12] Notwithstanding that $\min\{N,T\} < 40$ for both the annual and monthly panels and despite extracting the factors from residuals, these criteria appear to perform reasonably well in this context.

5 Application to PPP equations

The issue of cross section dependence in empirical tests of PPP was recently hightlighted by O'Connell (1998). He showed that assuming one (stationary) common shock reversed the positive verdict on PPP (from tests ignoring such dependence) for three out of four interational panels. We employ monthly

[12] Using $x_{it} = \sqrt{J} d_{it} + \sum_{j=1}^{J} \lambda_{ij} z_{jt}$ instead so that the idiosyncratic and common component of x_{it} have the same variance, the results for the annual and monthly panel are 2.296 (1.013) and 2.0 for the baseline DGP.

observations for the nominal exchange rate and price index variables, 1972:1-1998:12, giving $T = 324$ observations. Four panels are constructed which represent different combinations of price indexes and numeraire currencies. The CPI-DM and CPI-US$ panels use consumer price index data and nominal exchange exchange rates against the German mark and US dollar, respectively. The WPI-DM and WPI-US$ panels are based on wholesale price indexes.[13]

The following two PPP equations are considered:

i) Augmented Dickey-Fuller (ADF) type regression:

$$\Delta q_{it} = \alpha_{0i} + \alpha_{1i} t + \rho_i q_{i,t-1} + \sum_{j=1}^{k} \gamma_j \Delta q_{i,t-j} + u_{it} \tag{18}$$

where $q_{it} = e_{it} - d_{it}$ represents the real exchange rate, $d_{it} = p_{it} - p_t^*$ is the price differential between domestic country i and the foreign country, and all variables are in logarithms.

ii) Autoregressive distributed lag (ARDL) regression:

$$\Delta e_{it} = \alpha_i + \beta_i \Delta d_t + \delta_{1i} q_{i,t-1} + \gamma_i d_{i,t-1} + \sum_{j=1}^{k} \gamma_{ji}^1 \Delta e_{i,t-j} + \sum_{j=1}^{k} \gamma_{ji}^2 \Delta d_{i,t-j} + u_{it} \tag{19}$$

which is a reparameterization of an analogous equation with $e_{i,t-1}$ as regressor instead of $q_{i,t-1}$, where the coefficients on the former and on $d_{i,t-1}$ are δ_{1i} and δ_{2i}, respectively, with $\gamma_i = \delta_{1i} + \delta_{2i}$.

The average correlations between the cross-section OLS (standardized) residuals are shown in Table 3 together with the condition number and p-values of a diagonality test. All these indicate a considerable degree of cross-section dependence. This is especially so for the dollar panels where the depdendence may well reflect the large swings in the dollar during the 1980s. The number of retained factors ($\lambda_i > 1$) is also reported in Table 3.

Panel regressions are estimated with and without factors. Table 4 reports the estimated coefficients and t-ratios on the factors. The t-ratio on the single

[13] The CPI panels comprise $N = 17$ countries, Austria, Canada, Denmark, Finland, Germany, Greece, Japan, Netherlands, Sweden, Spain, Switzerland, United Kingdom, United States, Belgium, France, Italy, Norway, Portugal. The WPI panels exclude the latter five countries and instead include Ireland to make $N = 13$.

factor included in the US$ panel regressions is quite high. In the DM panels where $\hat{\tau} = 3$ and 2, the t-ratio on the first factor is dramatically larger than that for the other two. Since — as suggested by the Monte Carlo simulations for the monthly panel dimensions — the conventional standard errors for these panel estimators underestimate the true ones by almost 50%, these results can be taken as prima facie evidence that just one factor is significant in the DM panels also.[14] This is clearly corroborated by the Bai and Ng (2002) criteria (16) and (17). For instance, using RII residuals in $V(\tau, \hat{W}_\tau)$, for the ADF model IC_{p1} gives $3.655(\tau = 3)$, $3.611(\tau = 2)$ and $3.550(\tau = 1)$ for the CPI-DM panel and $3.608(\tau = 2)$ and $3.552(\tau = 1)$ for the WPI-DM panel. Inspection of the factors graphs does not suggest any obvious interpretation.

Tables 5(A-B) summarise the estimation results.[15] While the restrictions $-2 < \rho_i < 0$ (ADF model) and $\gamma_i = 0$ (ARDL) both imply long-run PPP, there is a subtle difference between them. The latter restriction only implies that changes in relative prices are reflected one-for-one in nominal exchange rates, but the former is more restrictive. It additionally requires real exchange rates to be I(0) variables.[16] The regression results indicate that the inclusion of factors has a small but consistent impact on the coefficient estimates.[17] In 11 out of 12 cases, the inclusion of factors makes ρ more negative strengthening the evidence in favour of PPP. In 10 out of the 12 cases, the included factors move γ closer to zero, again making the evidence more favourable toward PPP. The effects seem stronger for the dollar than the DM panels.

How do our PPP results compare with those from panel unit root or cointegration approaches accounting for cross section dependence? On one hand, our findings contrast with the results of both Moon and Perron (2001) and O'Connell (1998) which strongly reject the PPP hypothesis. On the other, they are in line with those of Pedroni (1997) whose panel cointegration tests support weak PPP and with those of Bai and Ng (2001a) who find some evidence for PPP using a common-idiosyncratic decomposition.

[14] Repeating the simulations with the panel dimensions of our PPP data, $N = 13$ and $T = 324$, gives a qualitatively similar ratio SM($SE_{\hat{b}}$)/SSD($\hat{b}$) of 0.56 (FE) and 0.57 (POLS).

[15] Note that the nominal exchange rate is normalized on the base year (1995) for the POLS regression to prevent biases arising from pooling observations in different metrics.

[16] On the basis of panel regressions of spot rates on price differentials a consistent estimator may suggest a long-run slope coefficient equal to one while real exchange rates are non-stationary. This issue is discussed in Coakley, Fuertes and Smith (2001).

[17] This is probably related to the high correlation between regressors in PPP equations due to a common numeraire which, as with SURE-GLS, erodes the gains from our approach.

6 Conclusions

In the context of large N and T panels it is important to confront the issue of cross-section dependence. Such dependence may arise from omitted unobservable global variables or common shocks which may be correlated with each group regressor. In this paper we propose a two-stage estimation method to deal with these. First one extracts principal components from the residuals of the model of interest. These factors can be used to augment the original regression equations to proxy possible omitted variables. A comparison of parameter estimates across the regression models with and without factors may provide some insights into the nature of these group dependences, namely, whether they represent exogenous world shocks or omitted global variables correlated with the regressors. In contrast to two-way FE, this approach can be used in models with heterogeneous slope coefficients across countries and when there are multiple omitted variables to which each country reacts differently. Moreover, in contrast to SURE-GLS it can be applied to large N panels. Using sequential limit theory it is shown that, for a simple DGP with no autocorrelation or groupwise heteroskedasticity, the POLS slope coefficient estimator of the augmented regression is consistent. For panel dimensions typical of the PPP literature and using POLS, FE and MG estimators, Monte Carlo simulations also confirm the bias reduction for a DGP with serial dependence and heteroskedasticity.

The proposed approach is illustrated by means of a PPP application for a group of 17 OECD countries 1973:1-1998:12. We find that between-group dependence is clearly significant in PPP equations and that it is much stronger in the US dollar than in the German mark panels. In all panels and PPP equations a minimum of one of the unobserved factors is clearly significant. Moreover, the impact of the factors on the augmented regression coefficients seems to strengthen support for long run PPP.

Finally the paper suggests various issues which warrant further research. One is to derive the appropriate asymptotic covariance matrix of the above panel estimators for the augmented regression. Absent the latter, another is to investigate whether a bootstrap technique might provide consistent standard errors in this setup. The case where the omitted variables are I(1) also needs consideration. It is hoped that the present contribution may motivate a more formal treatment of the ideas and approach elaborated in this paper.

References

[1] Bai, J. and S. Ng (2001a) A New Look at Panel Testing of Stationarity and the PPP Hypothesis, Mimeo, Dept. of Economics, Boston College.

[2] Bai, J. and S. Ng (2001b) A PANIC Attack on Unit Roots and Cointegration, Mimeo, Dept. of Economics, Boston College.

[3] Bai, J. and S. Ng (2002) Determining the Number of Factors in Approximate Factor Models, Econometrica, 70, 191-221.

[4] Baltagi, B.H. and Kao, C.W. (2000) Nonstationary Panels, Cointegration in Panels and Dynamic Models: A Survey, Advances in Econometrics 15, 7-51.

[5] Banerjee, A., M. Marcellino and C. Osbat (2001) Testing for PPP: Should We Use Panel Methods?, Mimeo, European University Institute.

[6] Coakley, J., A. M. Fuertes and Smith, R. (2001) Small Sample Properties of Panel Time-Series Estimators with I(1) Errors, Birkbeck College Discussion Paper 03/2001.

[7] Hall, S., Lazarova, S. and G. Urga (1999) A Principal Components Analysis of Common Stochastic Trends in Heterogeneous Panel Data, Oxford Bulletin of Economics and Statistics, 749-64.

[8] Moon, H. R. and B. Perron (2001) Testing for a Unit Root in Panels with Dynamic Factors, Mimeo, Department of Economics, University of Southern California.

[9] O'Connell, P.J.G. (1998) The Overvaluation of Purchasing Power Parity. Journal of International Economics 44: 1-19.

[10] Pedroni, P. (1997) Cross Sectional Dependence in Cointegration Tests of Purchasing Power Parity in Panels, Mimeo, Indiana University.

[11] Pesaran, M.H. and R. Smith (1995) Estimating Long-Run Relationships from Dynamic Heterogeneous Panels, Journal of Econometrics 68, 79-113.

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

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

[14] Robertson, D. and J. Symons. (1999) Factor Residuals in Panel Regressions: a Suggestion for Estimating Panels Allowing for Cross-Section Dependence, Mimeo, University of Cambridge.

[15] Telser, L.G. (1964) Iterative Estimation of a Set of Linear Regression Equations, Journal of the American Statistical Association, 59, 845-862.

Table 1A. Monte Carlo simulation results
$N = 30, T = 300, \gamma_i = 1, \rho_{di} = 0, \rho_z = 0, \sigma^2_{di} = 1$

	Mean	SSD	Min.	Max.	Skew.	Kurt-3
A) Regression I						
MG estimator						
$\hat{\theta}$	**1.4983**	.0217	1.4243	1.5584	.0544	.0215
$SE_1(\hat{\theta})$	.0083	.0011	.0056	.0116	.1819	-.2505
FE estimator						
$\hat{\theta}$	**1.4975**	.0216	1.4244	1.5579	.0593	.0299
$SE_1(\hat{\theta})$	.0092	.0002	.0087	.0096	-.1818	.0738
$SE_2(\hat{\theta})$	.0169	.0003	.0161	.0179	.2264	.2835
$SE_3(\hat{\theta})$	.0222	.0014	.0189	.0291	.7026	1.3561
POLS estimator						
$\hat{\theta}$	**1.4967**	.0216	1.4232	1.5573	.0589	.0382
$SE_1(\hat{\theta})$	.0092	.0002	.0087	.0096	-.1829	.0829
$SE_2(\hat{b})$	.0169	.0003	.0161	.0179	.2299	.3072
$SE_3(\hat{\theta})$	.0221	.0015	.0180	.0277	.5705	.6412

	Mean	SSD	Min.	Max.	Skew.	Kurt-3
B) Regression II						
MG						
$\hat{b}$	**1.0987**	.0165	1.0503	1.1464	.2764	-.1859
$SE_1(\hat{b})$	.0100	.0013	.0063	.0137	.1158	-.0758
FE						
$\hat{b}$	**1.0980**	.0163	1.0517	1.1459	.2769	-.1551
$SE_1(\hat{b})$	.0097	.0001	.0094	.0101	.0156	-.0846
$SE_2(\hat{b})$	.0098	.0001	.0095	.0103	.2678	.1741
$SE_3(\hat{b})$	.0109	.0006	.0092	.0133	.2375	.3445
POLS						
$\hat{b}$	**1.0977**	.0163	1.0506	1.1453	.2678	-.1418
$SE_1(\hat{b})$	.0097	.0001	.0094	.0100	.0219	-.1046
$SE_2(\hat{b})$	.0098	.0001	.0095	.0103	.2678	.1635
$SE_3(\hat{b})$	.0109	.0006	.0093	.0131	.2703	.2165

Note: SE_1 is based on the usual covariance matrix formulae for the MG, FE and POLS estimators. SE_2 is based on (10) and SE_3 is based on (13).

Table 1B. Monte Carlo simulation results
$N = 30, T = 25, \gamma_i = 1, \rho_{di} = 0, \rho_z = 0, \sigma^2_{di} = 1$

A) Regression I						
	Mean	SSD	Min.	Max.	Skew.	Kurt-3
MG						
$\hat{\theta}$	**1.5070**	.0766	1.2645	1.7224	-.2967	.0069
$SE_1(\hat{\theta})$	.0305	.0048	.0171	.0479	.2111	.0868
FE						
$\hat{\theta}$	**1.4966**	.0754	1.2460	1.7053	-.3233	.0326
$SE_1(\hat{\theta})$	.0322	.0020	.0259	.0378	-.0169	-.1542
$SE_2(\hat{\theta})$	.0567	.0035	.0465	.0659	.0043	-.0825
$SE_3(\hat{\theta})$	.0716	.0137	.0410	.1318	.8815	1.2679
POLS						
$\hat{\theta}$	**1.4869**	.0751	1.2406	1.6989	-.3207	.0052
$SE_1(\hat{\theta})$	.0317	.0019	.0256	.0375	-.0124	-.1353
$SE_2(\hat{\theta})$	.0567	.0034	.0461	.0656	.0082	-.0888
$SE_3(\hat{\theta})$	.0717	.0132	.0366	.1296	.4597	.2201

B) Regression II						
	Mean	SSD	Min.	Max.	Skew.	Kurt-3
MG						
$\hat{b}$	**1.1384**	.0734	.9472	1.4178	.3728	.2719
$SE_1(\hat{b})$	.0374	*.0052*	.0201	.0539	.1556	-.1378
FE						
$\hat{b}$	**1.1279**	.0709	.9565	1.3694	.3425	.2523
$SE_1(\hat{b})$	.0341	.0015	.0278	.0378	-.0736	.1013
$SE_2(\hat{b})$	.0355	.0024	.0297	.0488	.6921	1.3834
$SE_3(\hat{b})$	.0406	.0076	.0246	.0801	.8646	1.8745
POLS						
$\hat{b}$	**1.1243**	.0689	.9539	1.3793	.3591	.2305
$SE_1(\hat{b})$	.0335	*.0016*	.0279	.0386	-.1084	.1745
$SE_2(\hat{b})$	.0355	.0024	.0299	.0492	.7604	1.8043
$SE_3(\hat{b})$	.0392	.0076	.0220	.0642	.5005	.2672

Table 2A. Monte Carlo simulation results

$N = 30, T = 300, \gamma_i \sim U[0.5, 1.5], \rho_{di} = \rho_z = 0.9, \sigma^2_{di} \sim U[0.5, 1.5]$

A) Regression I						
	Mean	SSD	Min.	Max.	Skew.	Kurt-3
MG						
$\hat{\theta}$	**1.4110**	.0637	1.2563	1.6184	.1564	-.1977
$SE_1(\hat{\theta})$	.0234	.0034	.0137	.0365	.0983	.0359
FE						
$\hat{\theta}$	**1.4053**	.0629	1.2459	1.5887	.1382	-.2571
$SE_1(\hat{\theta})$	.0065	.0002	.0060	.0072	.1899	-.0861
POLS						
$\hat{\theta}$	**1.3899**	.0619	1.2368	1.5744	.1564	-.1977
$SE_1(\hat{\theta})$	.0064	.0002	.0059	.0071	.0770	-.1795

B) Regression II						
	Mean	SSD	Min.	Max.	Skew.	Kurt-3
MG						
$\hat{b}$	**1.0397**	.0154	1.0065	1.1210	.9317	1.7659
$SE_1(\hat{b})$	.0060	.0013	.0031	.0121	.7812	1.1088
FE						
$\hat{b}$	**1.0373**	.0158	1.0077	1.1058	.7595	.8561
$SE_1(\hat{b})$	.0044	.0002	.0039	.0053	.6836	.6280
POLS						
$\hat{b}$	**1.0351**	.0150	1.0054	1.0959	.7117	.6410
$SE_1(\hat{b})$	.0043	.0002	.0039	.0051	.6538	.4204

Table 2B. Monte Carlo simulation results
$N = 30, T = 25, \gamma_i \sim U[0.5, 1.5], \rho_{di} = \rho_z = 0.9, \sigma^2_{di} \sim U[0.5, 1.5]$

A) Regression I						
	Mean	SSD	Min.	Max.	Skew.	Kurt-3
MG						
$\hat{\theta}$	**1.4443**	.1499	1.1098	1.9150	.3441	-.2593
$SE_1(\hat{\theta})$	.0456	.0090	.0248	.0750	.4310	-.0399
FE						
$\hat{\theta}$	**1.4104**	.1405	1.0860	1.8757	.2773	-.2972
$SE_1(\hat{\theta})$	.0253	.0018	.0192	.0319	.0469	.5201
POLS						
$\hat{\theta}$	**1.2865**	.1238	1.0495	1.7040	.6528	.2045
$SE_1(\hat{\theta})$	.0224	.0020	.0169	.0324	.4613	1.2229

B) Regression II						
	Mean	SSD	Min.	Max.	Skew.	Kurt-3
MG						
$\hat{b}$	**1.1053**	.1215	.9606	1.9167	2.9186	12.1884
$SE_1(\hat{b})$	.0312	.0071	.0173	.0669	1.2112	2.5237
FE						
$\hat{b}$	**1.0793**	.1022	.9661	1.8435	3.1859	14.5451
$SE_1(\hat{b})$	.0230	.0018	.0186	.0291	.3763	-.1251
POLS						
$\hat{b}$	**1.0514**	.0804	.8759	1.6655	3.5172	18.5723
$SE_1(\hat{b})$	.0187	.0025	.01389	.0344	1.2047	3.1074

Table 3. Between-group correlations and retained factors

	ADF				ARDL			
Panel	$\bar{r}$	γ	λ_{LR}	τ	$\bar{r}$	γ	λ_{LR}	τ
CPI-DM	.265	8.641	1956	3	.274	8.288	2023	3
			[.000]				[.000]	
CPI-US$	.645	24.85	7274	1	.651	19.75	7059	1
			[.000]				[.000]	
WPI-DM	.272	7.836	1550	2	.275	7.419	1544	2
			[.000]				[.000]	
WPI-US$	.619	19.60	4859	1	.613	17.70	4869	1
			[.000]				[.000]	

Note: $\bar{r}$ is the average cross-section correlation, $\gamma = (\frac{\lambda_{\max}}{\lambda_{\min}})^{1/2}$ is the condition number, λ_{LR} is the p-value of Breusch-Pagan diagonality test on the residual correlation matrix and τ is the number of factors retained by Kaiser criterion. All of these are computed from the equation-by-equation OLS residuals.

Table 4. Estimated coefficients and t-ratios on included factors

	ADF model						
	CPI-DM			CPI-US$	WPI-DM		WPI-US$
MG	.0927	.0129	-.0174	.1375	.1131	.0124	.1570
	[6.79]	[.68]	[-.75]	[15.31]	[5.59]	[.42]	[12.02]
FE	.0924	.0126	-.0166	.1372	.1127	.0119	.1568
	[47.56]	[3.57]	[-3.83]	[107.69]	[45.67]	[2.39]	[91.38]
POLS	.0924	.0122	-.0156	.1370	.1125	.0116	.1567
	[47.44]	[3.46]	[-3.58]	[105.21]	[45.43]	[2.32]	[90.95]
	ARDL model						
	CPI-DM			CPI-US$	WPI-DM		WPI-US$
MG	.0881	.0109	.0227	.1360	.1087	.0029	.1530
	[6.36]	[.63]	[1.15]	[15.32]	[5.36]	[.10]	[11.71]
FE	.0882	.0091	.0223	.1356	.1092	.0024	.1533
	[48.39]	[2.64	[5.52]	[112.45]	[45.91]	[.52]	[92.29]
POLS	.0883	.0088	.0218	.1356	.1092	.0026	.1533
	[48.46]	[2.55]	[5.39]	[110.72]	[45.85]	[.56]	[92.29]

Table 5(A). Long run PPP coefficient estimates (CPI)

	MG	FE	POLS
i)DM			
ADF			
$\hat{\rho}$ $[SE_1]$	-.0377[.0037]	-.0237[.0029]	-.0127[.0018]
ADF($\hat{W}_\tau$)			
$\hat{\rho}$ $[SE_1]$	-.0398[.0036]	-.0247[.0024]	-.0131[.0015]
ARDL			
$\hat{\gamma}$ $[SE_1]$	-.01967[.0069]	-.0044[.00079]	-.0045[.00071]
ARDL($\hat{W}_\tau$)			
$\hat{\gamma}$ $[SE_1]$	-.01989[.0087]	-.0041[.00066]	-.0041[.00059]
ii) US$			
ADF			
$\hat{\rho}$ $[SE_1]$	-.0256[.0017]	-.0225[.0028]	-5.96e-5[.00021]
ADF($\hat{W}_\tau$)			
$\hat{\rho}$ $[SE_1]$	-.0274[.0019]	-.0245[.0016]	-4.86e-5[.00012]
ARDL			
$\hat{\gamma}$ $[SE_1]$	-.0064[.0056]	-.0025[.0014]	-.0034[.0012]
ARDL($\hat{W}_\tau$)			
$\hat{\gamma}$ $[SE_1]$	-.0024[.0030]	-.0024[.00077]	-.0040[.00067]

Note: Tables 5A-5B report the estimation results for two dynamic PPP equations, ADF and ARDL. The number of augmentation lags is conservatively set at k=6 in all equations to eliminate serial correlation. ADF($\hat{W}_\tau$) and ARDL($\hat{W}_\tau$) denote the models with τ factors as additional regressors. The conventional s.e. for the MG, POLS and FE estimators are in brackets. These are likely to be biased downwards in all regressions.

Table 5(B). Long run PPP coefficient estimates (WPI)

	MG	FE	POLS
i) DM			
ADF			
$\hat{\rho}$ $[SE_1]$	-.0524[.0050]	-.0317[.0041]	-.0178[.0029]
ADF$(\hat{W}_\tau)$			
$\hat{\rho}$ $[SE_1]$	-.0559[.0053]	-.0346[.0033]	-.0192[.0023]
ARDL			
$\hat{\gamma}$ $[SE_1]$	-.0069[.0047]	-.0025[.0011]	-.0033[.00092]
ARDL$(\hat{W}_\tau)$			
$\hat{\gamma}$ $[SE_1]$	-.0053[.0054]	-.0023[.00089]	-.0028[.00075]
ii) US$			
ADF			
$\hat{\rho}$ $[SE_1]$	-0.0311[.0019]	-0.0279[.0038]	-0.0209[.0032]
ADF$(\hat{W}_\tau)$			
$\hat{\rho}$ $[SE_1]$	-0.0346[.0018]	-0.0314[.0022]	-0.0234[.0019]
ARDL			
$\hat{\gamma}$ $[SE_1]$	-0.0099[.0034]	-0.0037[.0017]	-0.0022[.0013]
ARDL$(\hat{W}_\tau)$			
$\hat{\gamma}$ $[SE_1]$	-0.0030[.0024]	-0.0021[.00097]	-0.0020[.00072]