Cointegration Testing in Panels with Common Factors

Christian Gengenbach, Franz C. Palm, Jean-Pierre Urbain*

Department of Quantitative Economics,
Universiteit Maastricht, The Netherlands
March 2006

Abstract

Panel unit root and no-cointegration tests that rely on cross-sectional independence of the panel unit experience severe size distortions when this assumption is violated, as has e.g. been shown by Banerjee, Marcellino and Osbat (2004, 2005) via Monte Carlo simulations. Several studies have recently addressed this issue for panel unit root tests using a common factor structure to model the cross-sectional dependence, but not much work has been done yet for panel no-cointegration tests.

This paper proposes a model for panel no-cointegration using an unobserved common factor structure, following the work by Bai and Ng (2004) for panel unit roots. The model enables us to distinguish two important cases: (i) the case when the non-stationarity in the data is driven by a reduced number of common stochastic trends, and (ii) the case where we have common and idiosyncratic stochastic trends present in the data. We discuss the homogeneity restrictions on the cointegrating vectors resulting from the presence of common factor cointegration. Furthermore, we study the asymptotic behavior of some existing, residual-based panel no-cointegration, as suggested by Kao (1999) and Pedroni (1999, 2004). Under the DGP used, the test statistics are no longer asymptotically normal, and convergence occurs at rate T rather than $\sqrt{N}T$ as for independent panels. We then examine the possibilities to test for various forms of no-cointegration in our model, by extracting the common factors and individual components from the observed data directly and then test for no-cointegration using residual-based panel tests applied to the defactored data.

Keywords: panel cointegration testing, common factors.

*Corresponding author: Jean-Pierre Urbain, Department of Quantitative Economics Universiteit Maastricht, P.O. Box 616, 6200 MD Maastricht, The Netherlands, j.urbain@ke.unimaas.nl. Previous versions of this paper were presented at the 12^{th} International Panel Conference in Copenhagen, at the Conference on Frontiers in Times Series Analysis in Olbia, at a seminar at the Econometric Institute in Rotterdam and at the NSF/NBER Time Series Conference in Heidelberg. The authors would like to thank conference and seminar participants and in particular Jörg Breitung, Josep Carrion-i-Silvestre, Peter Pedroni, Joakim Westerlund and three unknown referees for helpful comments and suggestions on a previous draft. The usual disclaimer applies.

1 Introduction

The effect of persistent cross-sectional dependence on panel unit root tests has been recently analyzed and documented in some detail in the literature. As shown by Monte Carlo simulations (Banerjee et al, 2005) or by asymptotic analysis (Lyhagen, 2000; Pedroni and Urbain, 2001), the standard (Levin-Lin-Chu or IPS) panel unit root tests are severely affected in that either they display dramatic size distortions or even worse can be shown to diverge with the cross-sectional dimension of the panel. To overcome these issues, new panel unit root tests have been proposed that model the possibly persistent cross-sectional dependency using a single common factor model (Pesaran, 2004, Phillips and Sul, 2003) or k-common factor models (Moon and Perron, 2004, Bai and Ng, 2004).

For the case of tests for the null of no-cointegration, not much work has been done yet. Banerjee et al. (2004) conduct an extensive Monte Carlo study where they conclude that while all statistics investigated (residual-based tests or likelihood based trace-type test) are affected, the presence of cross-member cointegration appears much less harmful for single-equation tests than for the panel version of the Johansen test. In many cases, the tests are affected by the presence of cointegration between members is such a way that these tests cannot discriminate between cointegration among members and cointegration within, that is for a single member of the panel. Bai and Kao (2004) and Banerjee and Carrion-i-Silvestre (2005) study tests for panel no-cointegration with cross-sectional dependence. Both studies consider residual-based tests for a single cointegration relationship, where the error term of the cointegrating equation follows a common factor structure as in Bai and Ng (2004). Urbain (2004) on the other hand studies analytically the issue of spurious regression in panels when the units are cointegrated along the cross-sectional dimension, i.e. when there is cross-member cointegration. In contrast to the spurious regression result for independent panels studied by Phillips and Moon (1999), Pedroni (1995) or Kao (1999), in most of the cases considered these estimators are not consistent and actually converge to non-degenerate limiting distributions once the observed non-stationarity is generated by a reduced number of common stochastic trends.

This paper builds on these results and extends the analysis to the analytical study of panel tests for no-cointegration when the cross-sectional dependence in the panel is modelled by a common factor structure following the work of Bai and Ng (2004). Adopting the framework of Bai and Ng (2004) enables to consider essentially two different classes of cases that we believe are both of theoretical and empirical relevance: (i) the case where the observed non-stationarity in the variables originates from a reduced number of cross-sectional common trends only; (ii) the case where we have both cross-sectional common stochastic trends as well as idiosyncratic ones. The spurious regression analysis for the former case is reported in Urbain (2004) and corresponds to the cross-member cointegration case. The second case is actually the one considered by Moon and Perron (2004) or Pesaran (2003) in the context of panel unit root analysis and excludes the existence of cross-unit or cross-member cointegration in the panel since both components are $\mathsf{I}(1)$.

For both classes of DGP's, we discuss the homogeneity restrictions on the cointegrating vectors resulting from the presence of common factor cointegration. These implications of the common factor cointegration are important reasons to propose a sequential approach whereby the data are decomposed into common factors and idiosyncratic components and the (no-)cointegration is tested for these factors and components separately in a sequential approach. Then, we study analytically the behavior of several popular test for panel cointegration including Kao's (1999) and Pedroni's (1999, 2004) residual based panel no-cointegration tests that have been widely used in empirical work in the recent years. For example, when the number of common factors generating the non-stationarity in the panel is kept fixed while the dimension of the panel increases, then the Gaussian limiting results derived for the independent case are not valid anymore. Tests that are based on pooled or LSDV estimation of the underlying panel cointegration static regression may in some cases diverge with $\sqrt{N}$ and hence important size distortions are to be expected already in panels with moderate cross-sectional dimension. Group mean statistics are also affected and not asymptotically Gaussian anymore.

These results complement and help to have a better understanding of some of the Monte Carlo reported by Banerjee et al. (2004). We then examine the possibilities to test for no-cointegration, by extracting the common factors and individual components from the observed data directly and then test for no-cointegration using residual-based panel tests applied to the defactored data.

The remainder of the paper is organized as follows: In Section 2 we present our model for panel no-cointegration with a common factor structure. In Section 3 we examine the asymptotic behavior of some residual-based panel no-cointegration tests when the data is generated by our DGP. Section 4 discusses a possible solution for testing for various forms of no-cointegration when the data contains unobserved common factors. In particular, we examine the possibilities of de-factoring the observed data before testing, using the methodology from Bai and Ng (2004). The finite sample behavior of the proposed approach is analyzed in Section 5. In Section 6 we illustrate the approach by revisiting the issue of the weak purchasing power parity (PPP) hypothesis. Conclusions are drawn in Section 7.

A note on notation: throughout the text, M is used to denote a generic positive number, not depending on T or N. For a matrix $\boldsymbol{A}$, $\boldsymbol{A} > 0$ denotes that $\boldsymbol{A}$ is positive definite. Furthermore, $\|\boldsymbol{A}\| = trace(\boldsymbol{A}'\boldsymbol{A})^{\frac{1}{2}}$. We write the integral $\int_0^1 B(r)\mathbf{d}r$ as $\int B$, and $\int_0^1 B(r)B(r)'\mathbf{d}r$ as $\int BB'$. Furthermore, $\Longrightarrow$ denotes weak convergence, and $\stackrel{p}{\longrightarrow}$ denotes convergence in probability. For any number x, $\lfloor x \rfloor$ denotes the largest integer smaller than x. For any variable $X_{i,t}$, $\tilde{X}_{i,t} = X_{i,t} - \frac{1}{T}\sum_{s=1}^{T} X_{i,s}$. Similarly, for any Brownian motion B, $\tilde{B} = B - \int B$.Throughout the paper we employ sequential limit theory, where we consider $T \to \infty$ followed by $N \to \infty$.

2 The Model

We consider balanced panels with N cross-sectional units and T time-series observations, indexed by $i = 1, \ldots, N$ and $t = 1, \ldots, T$ respectively. For each unit in the panel we observe a $(1+m)$-dimensional vector of variables $Z_{i,t} = (Y_{i,t}, X'_{i,t})'$, where $Y_{i,t}$ is a scalar time series and $X_{i,t}$ is a m-vector time series. We assume that the DGP for $Z_{i,t}$ has a common factor structure as e.g. in Bai and Ng (2004), and we assume the presence of k common factors in the data. Furthermore, we assume the number of common factors to be fixed as $T, N \to \infty$ throughout the paper. Our model is given by

$$Z_{i,t} = D_{i,t} + \Lambda_i F_t + E_{i,t}, \tag{1}$$

$t = 1, \ldots, T$, $i = 1, \ldots, N$. $D_{i,t}$ is an unobserved deterministic component such that either $D_{i,t} = 0$ for all i and t if there are no deterministic components present, $D_{i,t} = d_{0i}$ for all t if the data contains individual specific fixed effects, or $D_i = d_{0i} + d_{1i}t$ if the data contains individual specific deterministic linear time trends, where the coefficients d_{0i} and d_{1i} depend on i only. For the remainder of the paper we assume $D_{i,t} = 0$ unless mentioned otherwise. The common component in $Z_{i,t}$ is given by F_t in (1). F_t is a k-vector of common $\mathsf{I}(1)$ factors given by

$$F_t = F_{t-1} + f_t, \tag{2}$$

where $f_t = \Phi(L)\eta_t$, η_t is a sequence of $(k \times 1)$ $iid(0, I_k)$ random vectors, $\Phi(L) = \sum_{j=0}^{\infty} \Phi_j L^j$. The $(1+m) \times k$ matrix of factor loadings Λ_i is assumed to be of full rank and block-diagonal, with block-diagonality corresponding to the partition of $Z_{i,t}$, such that

$$\Lambda_i = \begin{bmatrix} \lambda'_{1i} & 0 \\ 0 & \lambda'_{2i} \end{bmatrix}.$$

As for the vectors of observations $Z_{i,t}$, we have partitions for the unobserved vector of common factors $F_t = (F_t^{Y'}, F_t^{X'})'$ where F_t^Y and F_t^X have k_Y and k_X elements respectively, and the partition of F_t corresponds to the structure of Λ_i, such that λ_{1i} is $k_Y \times 1$ and λ_{2i} is $k_X \times m$. The block diagonal structure for the factor loadings is necessary to ensure that $Y_{i,t}$ and $X_{i,t}$ are not cointegrated when the non-stationarity in the data is driven by the common factors alone. When the idiosyncratic components are non-stationary as well, this assumption on Λ_i might be relaxed and a more general structure can be considered.

For the idiosyncratic component in (1), $E_{i,t}$, we distinguish two cases, namely stationary and non-stationary idiosyncratic components. For the former case we have

$$E_{i,t} = e_{i,t}, \tag{3}$$

while in the latter case we assume

$$E_{i,t} = E_{i,t-1} + e_{i,t}, \tag{4}$$

where the stationary vector $e_{i,t} = \Gamma_i(L)\varepsilon_{i,t}$ with $\varepsilon_{i,t}$ being a sequence of random $iid(0, \Sigma_i)$ random vectors, $\Gamma_i(L) = \sum_{j=0}^{\infty} \Gamma_{ij} L^j$. Again, we partition $E_{i,t}$ conformable with the data $Z_{i,t}$, such that $E_{i,t} = (E_{i,t}^Y, E_{i,t}^{X'})'$, where $E_{i,t}^Y$ is a scalar time series and $E_{i,t}^X$ has m elements.

For the above given model we specify the following assumptions, where M denotes a generic positive real number:

Assumption 1 *Common factors: (i)* $\eta_t \sim iid(0, I_k)$ *with finite* 4^{th} *moments, (ii) there is an* M *such that* $\sum_{j=0}^{\infty} j \cdot \|\Phi_j\| < M$, *(iii)* $\operatorname{rank}(\Phi(1)) = k$, *(iv)* $\mathsf{E}\|F_0\| \leq M$.

Assumption 2 *Factor loadings: (i) for non-random* λ_{1i} *and* λ_{2i}, $\|\lambda_{1i}\| \leq M$ *and* $\|\lambda_{2i}\| \leq M$; *for random* λ_{1i} *and* λ_{2i}, $\mathsf{E}\|\lambda_{1i}\|^4 \leq M$ *and* $\mathsf{E}\|\lambda_{2i}\|^4 \leq M$, *(ii)* $N^{-1}\sum_{i=1}^{N} \Lambda_i'\Lambda_i \xrightarrow{p} \Sigma_\Lambda > 0$, *(iii) for non-random* λ_{1i} *and* λ_{2i}, $N^{-1}\sum_{i=1}^{N} \lambda_{1i} \neq 0$ *and* $N^{-1}\sum_{i=1}^{N} \lambda_{2i} \neq 0$; *for random* λ_{1i} *and* λ_{2i}, $\mathsf{E}(\lambda_{1i}) \neq 0$ *and* $\mathsf{E}(\lambda_{2i}) \neq 0$.

Assumption 3 *Idiosyncratic components: for each* $i = 1, \ldots, N$, *(i)* $\varepsilon_{i,t} \sim iid(0, \Sigma_i)$ *with finite* 8^{th} *moments, and* $\varepsilon_{i,t}$ *and* $\varepsilon_{j,s}$ *are independent for any* t, s *and* $i \neq j$, *(ii)* $\mathsf{E}\|\varepsilon_{i,0}\| < M$, *(iii)* $\Gamma_i(L)$ *fulfills the random coefficients and summability conditions from Phillips and Moon (1999), Assumptions 1 and 2 on p.1060 and p.1061 respectively, (iv)* $\operatorname{rank}(\Gamma_i(1)) = m + 1$, $\forall i$.

Assumption 4 *The errors,* η_t, $\varepsilon_{i,t}$, *and the factor loadings* Λ_i *form mutually independent groups.*

Under the conditions of Assumption 1, the common factors F_t form a k-dimensional I(1) process and the possibility of cointegration between the common factors is excluded. The full rank assumption on the long-run covariance matrix of F_t could in fact be relaxed, as long as the diagonal blocks corresponding to the long-run covariances of F_t^Y and F_t^X have at least rank 1 each. The long-run covariance matrix of the common factors is given by

$$\Omega = \Phi(1)\Phi(1)' = \Xi + \Theta + \Theta',$$

where $\Xi = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathsf{E}(f_t f_t')$ and $\Theta = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathsf{E}(f_t F_{t-1}')$ (see e.g. Phillips and Durlauf, 1986). Furthermore, an invariance principle holds such that

$$T^{-1/2} F_{\lfloor rT \rfloor} \Longrightarrow B_F(r) \quad as \quad T \to \infty, \tag{5}$$

where B_F is a k-vector Brownian motion with covariance matrix Ω. Assumptions 2(i) and 2(ii) are standard assumptions for factor models and ensure that the factor loadings are identifiable. Assumption 2(iii) is needed for the spurious regression results when the non-stationarity in the data is only driven by the common factors. Assumption 3(iii) specifies that a panel functional central limit theorem holds for $S_{i,t} = \sum_{s=1}^{t} e_{i,t}$, which corresponds to $E_{i,t}$ in case the idiosyncratic components are non-stationary as in (4), or to its cumulative sum if (3) is true. The long-run covariance of $S_{i,t}$ is given by

$$\Psi_i = \Gamma_i(1)\Sigma_i\Gamma_i(1)' = \Upsilon_i + \Delta_i + \Delta_i',$$

where $\Upsilon_i = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathsf{E}(e_{i,t}e_{i,t}')$ and $\Delta_i = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathsf{E}(e_{i,t}S_{i,t-1}')$, and an invariance principle ensures that

$$T^{-1/2} S_{i,\lfloor rT \rfloor} \Longrightarrow B_i(r) \quad as \quad T \to \infty, \tag{6}$$

where B_i is a randomly scaled $(1+m)$-vector Brownian motion with covariance matrix Ψ_i. Assumption 3(iv) ensures that the idiosyncratic terms do not cointegrate in case these are I(1) vectors.

The implications of these assumptions are best understood by considering the Beveridge-Nelson (BN) decomposition for F_t and for $E_{i,t} = \sum_{s=1}^{t} e_{i,s}$:

$$F_t = \Phi(1)\sum_{s=1}^{t} \eta_s + \Phi^*(L)(\eta_t - \eta_0) + F_0, \tag{7}$$

$$E_{i,t} = \Gamma_i(1)\sum_{s=1}^{t} \varepsilon_{i,s} + \Gamma_i^*(L)(\varepsilon_{i,t} - \varepsilon_{i,0}) + E_{i,0}, \tag{8}$$

where $\Phi^*(L) = \sum_{j=0}^{\infty} \Phi_j^* L^j$ with $\Phi_j^* = -\sum_{l=j+1}^{\infty} \Phi_l$, $\Gamma_i^*(L) = \sum_{j=0}^{\infty} \Gamma_{i,j}^* L^j$ with $\Gamma_{i,j}^* = -\sum_{l=j+1}^{\infty} \Gamma_{i,l}$, $\Phi^*(L)(\eta_t - \eta_0)$ and $\Gamma_i^*(L)(\varepsilon_{i,t} - \varepsilon_{i,0})$ are stationary with finite fourth order moments and F_0 and $E_{i,0}$ are $O_p(1)$ by assumption.

If (3) is true the idiosyncratic data components are $\mathsf{I}(0)$, and the $\mathsf{I}(1)$ trends of the common factors contained in $\Lambda_i\Phi(1)\sum_{s=1}^{t}\eta_s$ drive the non-stationarity in the data. Then, we might observe *cross-member cointegration* between some $Y_{i,t}$ and $Y_{j,t}$, and between some $X_{i,t}$ and $X_{j,t}$ for some $i, j,\ i \neq j$, the exact cointegration structure depending on the individual loadings. The assumption on the block-diagonal structure of the factor loadings Λ_i in turn implies that we have separation in a cointegrating system, see Hecq, Palm and Urbain (2002). Note that the assumption of cointegration between $Y_{i,t}$ and $X_{i,t}$ would only be possible if the common factors F_t^Y and F_t^X would cointegrate, which is ruled out by Assumption 1 from which the full rank of the long-run covariance matrix of F_t follows.

When $E_{i,t}$ is given by (4), both common and idiosyncratic data components are non-stationary and drive the nonstationarity in $Z_{i,t}$, $i = 1, \dots, N$. Furthermore, the idiosyncratic components do not cointegrate along the cross-section. Hence, we do not have cointegration "within" units, e.g. between $Y_{i,t}$ or $X_{i,t}$. The BN decomposition of the $Z_{i,t}$ is easily obtained from (1) and (7-8) and shows that the non-stationarity of $Z_{i,t}$ stems from the term $\Lambda_i \Phi(1) \sum_{s=1}^t \eta_s + \Gamma_i(1) \sum_{s=1}^t \varepsilon_{i,s}$.

Remark 1. The purpose of this paper is to investigate tests for no-cointegration so that we need to maintain the assumption that there does exist a full column rank matrix β_i' such that $\beta_i' Z_{it} \sim \mathsf{I}(0)$. The model enables us however to distinguish a variety of different cases. We will concentrate on two different important cases, namely one with cross-member cointegration where we have $\mathsf{I}(1)$ common factors and $\mathsf{I}(0)$ idiosyncratic terms and one where the panel units contain common stochastic trends, but do not cointegrate even along the cross-sectional dimension so that both the common and the idiosyncratic components are $\mathsf{I}(1)$.

Remark 2: Heterogeneity and cross-sectional dependence. Strictly speaking, with $\mathsf{I}(1)$ common factors as well as $\mathsf{I}(1)$ idiosyncratic components, we actually have two different sets of possible cointegrating vectors that would annihilate the idiosyncratic and the common $\mathsf{I}(1)$ stochastic trends respectively, see also the discussion in Gregoir (2005), Breitung and Pesaran (2005). Combining (1) and (7)-(8), the resulting BN representation of $Z_{i,t}$ shows that it will not be easy to annihilate both. In particular, cointegrating vector(s), say δ, that annihilate the common I(1) components should lie in the left null space of Λ_i such that $\delta \Lambda_i \Phi(1) = 0$ as $\Phi(1)$ is of full rank by Assumption 1, while those for the idiosyncratic components, say γ_i' would have to lie in the left null space of $\Gamma_i(1)$ such that $\gamma_i' \Gamma_i(1) = 0$. If the intersection of these left null spaces is the empty set, then there does not exist a cointegrating relationship that would annihilate both the unit roots from the common stochastic trends and those of the idiosyncratic terms. This would *also* represent a situation where none of the $Z_{i,t}$ vectors are cointegrated. The components taken in isolation could be cointegrated though.

A closer inspection of the possible structure under cointegration shows that there is an important trade-off between the degree of heterogeneity that can allowed for and the existence of cross-sectional dependence modeled by common factors.

Without loss of generality, but to simplify the presentation, we consider the following simple bivariate DGP where we have a single $\mathsf{I}(1)$ common factor in Y and a single $\mathsf{I}(1)$ common factor in X:

$$Y_{i,t} = \lambda_{1,i} F_t^Y + E_{i,t}^Y, \tag{9}$$

$$X_{i,t} = \lambda_{2,i} F_t^X + E_{i,t}^X, \tag{10}$$

from which we see that any linear combination can be written as

$$Y_{i,t} - \beta_i X_{i,t} = \lambda_{1,i}(F_t^Y - \frac{\beta_i \lambda_{2,i}}{\lambda_{1,i}} F_t^X) + E_{i,t}^Y - \beta_i E_{i,t}^X. \tag{11}$$

For the linear combination $(1, -\beta_i)$ to be a cointegrating vector such that $Y_{i,t} - \beta_i X_{i,t} \sim \mathsf{I}(0)$, two conditions need to hold, namely (i) $(F_t^Y - \frac{\beta_i \lambda_{2,i}}{\lambda_{1,i}} F_t^X) \sim \mathsf{I}(0)$ (ii) $(E_{i,t}^Y - \beta_i E_{i,t}^X) \sim \mathsf{I}(0)$.

Given that we here have only two $\mathsf{I}(1)$ common factors, there can be at most a single linear cointegrating combination between these factors and hence $\frac{\beta_i \lambda_{2,i}}{\lambda_{1,i}}$ cannot be individual specific, i.e. should be the same $\forall i$. There are consequently three different cases that are compatible with a constant (over i) ratio:

1. The first case is the situation where we assume *homogeneity* of the factor loadings and of the possible cointegrating vector. In that case $\frac{\beta_i \lambda_{2,i}}{\lambda_{1,i}} = \frac{\beta \lambda_2}{\lambda_1}$ does not depend on i anymore. A similar restriction is considered by Gregoir (2005). Another possibility is the case of homogeneity of the cointegrating vector β_i coupled with the constancy (over i) of the ratio of the factor loadings $\frac{\lambda_{2,i}}{\lambda_{1,i}}$ over the units i which is also excluded by Assumptions 1-4.

2. The second case allows for some degree of heterogeneity. To rule out individual specific ratios, the factor loadings should vary with β_i such that the ratio $\frac{\beta_i \lambda_{2,i}}{\lambda_{1,i}}$ is constant across i. This limits the possible amount of heterogeneity in the panels and is excluded by Assumptions 1-4 where the loadings and Ψ_i are assumed to vary independently from each other.

3. A third case could arise in practice when the variables $Y_{i,t}$ and $X_{i,t}$ have a common source of nonstationarity only, that is they are generated by a single nonstationary common factor F_t (e.g. a common technology shock affecting all companies i). The idiosyncratic component is assumed to be stationary (or could be cointegrated with cointegrating vector β_i). In this case, $Y_{i,t}$ and $X_{i,t}$ are cointegrated with $\beta_i = \frac{\lambda_{1,i}}{\lambda_{2,i}}$. It is ruled out by the assumption of block-diagonality of Λ_i, but it would be a natural alternative hypothesis to the null of no-cointegration. Homogeneity of the cointegrating vector then arises if $\frac{\lambda_{1,i}}{\lambda_{2,i}}$ is constant across entities i.

The conclusion from this brief discussion is that one should keep in mind that if we allow for almost unrestricted (up to the conditions stated in Assumptions 1-4) heterogeneity, then the existence of cointegrating relations that annihilate *both* the common and idiosyncratic $\mathsf{I}(1)$ stochastic trends is a very unlikely stituation. The consequence of this discussion for testing for the null of no-cointegration in this factor set-up will be mentioned in Section 4.

Remark 3. A similar framework is also, independently of the present work, proposed by Dees, di Mauro, Pesaran and Smith (2005) for the study of macroeconomic linkages within the Euro area. The purpose of their work is however different as no attempt to discuss tests for cointegration is made. This work may thus appear as complementary to theirs.

3 The behavior of panel residual based tests

The purpose of this section is to study, given the set-up introduced in the preceding section, the asymptotic behavior of some standard and popular panel tests for no-cointegration. We can view this section as providing some complementary results to the simulation results reported by Banerjee et al. (2004). The statistics we consider are designed to test for the presence of a single cointegration relationship between $Y_{i,t}$ and $X_{i,t}$.[1] Kao (1999) considers

[1] This is a restrictive assumption that we however will make in the sequel by assuming the existence of a single cointegrating vector. Approaches that allow for more than one cointegrating vector, are reviewed in Breitung and Pesaran (2005).

a homogenous cointegrating vector, whereas Pedroni (1999) allows for some heterogeneity. However, both rely on the cross-sectional independence of the panel unit to derive asymptotic normality for their test statistics.

3.1 Kao (1999)

Kao (1999) proposes to estimate the homogeneous cointegrating relationship by pooled regression allowing for individual fixed effects. The regression equation is given by

$$Y_{i,t} = \alpha_i + \beta X_{i,t} + u_{i,t}, \tag{12}$$

where β and $X_{i,t}$ are row and column vectors respectively, and the least squared dummy variable (LSDV) estimator for β is

$$\tilde{\beta} = \big(\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{Y}_{i,t}\tilde{X}'_{i,t}\big)\big(\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{X}_{i,t}\tilde{X}'_{i,t}\big)^{-1},$$

where $\tilde{Y}_{i,t} = Y_{i,t} - \frac{1}{T}\sum_{s=1}^{T} Y_{i,s}$ and $\tilde{X}_{i,t} = X_{i,t} - \frac{1}{T}\sum_{s=1}^{T} X_{i,s}$. The residuals from this first stage regression $\tilde{u}_{i,t} = \tilde{Y}_{i,t} - \tilde{\beta}\tilde{X}_{i,t}$ will still contain a unit root under the null hypothesis of no cointegration. We now estimate a pooled DF regression

$$\Delta\tilde{u}_{i,t} = (\rho - 1)\tilde{u}_{i,t-1} + v_{i,t}, \tag{13}$$

where the pooled ordinary least squares (POLS) estimator of $(\rho - 1)$ is given by

$$(\tilde{\rho} - 1) = \big(\sum_{i=1}^{N}\sum_{t=2}^{T} \Delta\tilde{u}_{i,t}\tilde{u}_{i,t-1}\big)\big(\sum_{i=1}^{N}\sum_{t=2}^{T} \tilde{u}^2_{i,t-1}\big)^{-1}.$$

Kao's (1999) tests are based on $\tilde{\rho}$ and the corresponding t-statistic

$$t_{\tilde{\rho}} = \big(\tilde{\rho} - 1\big)\big(\hat{s}^2_{\tilde{u}}\big(\sum_{i=1}^{N}\sum_{t=2}^{T} \tilde{u}^2_{i,t-1}\big)^{-1}\big)^{-\frac{1}{2}},$$

where $\hat{s}^2_{\tilde{u}} = N^{-1}T^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}(\Delta\tilde{u}_{i,t-1} - (\tilde{\rho} - 1)\tilde{u}_{i,t-1})^2$, corrected for endogeneity and serial correlation. When the panel units are cross-sectionally independent, the test statistics are asymptotically normally distributed as $T \to \infty$ followed by $N \to \infty$. However, for the model given by (1), (2) and (3) or (4), this assumption is clearly violated. Using the results reported in Lemmas 1-3 in Appendix A, we obtain the following limit results, where $\mathrm{vec}(\int \mathbf{d}B_{F\Lambda}B'_{F\Lambda}) = \check{\boldsymbol{\Lambda}}\,\mathrm{vec}(\int \mathbf{d}B_F B'_F)$, $\mathrm{vec}(\Theta_{F\Lambda}) = \check{\boldsymbol{\Lambda}}\,\mathrm{vec}(\Theta)$, $\mathrm{vec}(\int B_{F\Lambda}B'_{F\Lambda}) = \check{\boldsymbol{\Lambda}}\,\mathrm{vec}(\int B_F B'_F)$, $\mathrm{vec}(\int \mathbf{d}B_{F\Lambda}\tilde{B}'_{F\Lambda}) = \check{\boldsymbol{\Lambda}}\,\mathrm{vec}(\int \mathbf{d}B_F\tilde{B}'_F)$ and $\mathrm{vec}(\int \tilde{B}_{F\Lambda}\tilde{B}'_{F\Lambda}) = \check{\boldsymbol{\Lambda}}\,\mathrm{vec}(\int \tilde{B}_F\tilde{B}'_F)$, and $\check{\boldsymbol{\Lambda}} = \mathrm{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}(\Lambda_i \otimes \Lambda_i)$, and B_F and B_i are given in Equations (5) and (6), respectively.

Proposition 1 *Given Assumptions 1, 2, 3 and 4:*

(A) Consider the model given by (1), (2) *and* (3),

(a) $\tilde{\beta} \Longrightarrow (\int \tilde{B}^{Y}_{F\Lambda}\tilde{B}^{X'}_{F\Lambda})(\int \tilde{B}^{X}_{F\Lambda}\tilde{B}^{X'}_{F\Lambda})^{-1} = \tilde{\mathbf{b}}_A$ *as* $T, N \to \infty$ *sequentially,*

(b) $T(\tilde{\rho}-1) \Longrightarrow \frac{(1,-\tilde{\mathbf{b}}_A)(\int \mathbf{d}B_{F\Lambda}\tilde{B}'_{F\Lambda}+\Theta_{F\Lambda}+\gamma_1-\Upsilon)(1,-\tilde{\mathbf{b}}_A)'}{(1,-\tilde{\mathbf{b}}_A)(\int \tilde{B}_{F\Lambda}\tilde{B}'_{F\Lambda})(1,-\tilde{\mathbf{b}}_A)'}$ *as* $T, N \rightarrow \infty$ *sequentially, where* $\gamma_1 = \mathsf{E}(\gamma_{i1})$ *and* $\gamma_{i1} = \mathsf{E}(\tilde{e}_{i,t-1}\tilde{e}'_{i,t})$,

(c) $t_{\tilde{\rho}}$ *diverges at rate* $\sqrt{N}$ *as* $T, N \rightarrow \infty$ *sequentially.*

(B) Consider the model given by (1), (2) *and* (4),

(a) $\tilde{\beta} \Longrightarrow (\int \tilde{B}^{Y}_{F\Lambda}\tilde{B}^{X\prime}_{F\Lambda} + \frac{1}{6}\Psi^{YX})(\int \tilde{B}^{X}_{F\Lambda}\tilde{B}^{X\prime}_{F\Lambda} + \frac{1}{6}\Psi^{XX})^{-1} = \tilde{\mathbf{b}}_B$ *as* $T, N \rightarrow \infty$ *sequentially,*

(b) $T(\tilde{\rho}-1) \Longrightarrow \frac{(1,-\tilde{\mathbf{b}}_B)(\int \mathbf{d}B_{F\Lambda}\tilde{B}'_{F\Lambda}+\Theta_{F\Lambda}-\frac{1}{2}\Psi+\Delta)(1,-\tilde{\mathbf{b}}_B)'}{(1,-\tilde{\mathbf{b}}_B)(\int \tilde{B}_{F\Lambda}\tilde{B}'_{F\Lambda}+\frac{1}{6}\Psi)(1,-\tilde{\mathbf{b}}_B)'}$ *as* $T, N \rightarrow \infty$ *sequentially,*

(c) $t_{\tilde{\rho}}$ *diverges at rate* $\sqrt{N}$ *as* $T, N \rightarrow \infty$ *sequentially.*

Proof: see Appendix B.

The results summarized in Proposition 1 are clearly in contrast to the asymptotic normality Kao (1999) derives for the tests statistics for independent panels, although we have not yet considered corrections for serial correlation and endogenous regressors. Results $A(a)$ and $B(a)$ are similar to those derived by Urbain (2004) for the pooled least squares estimator (PLS). This is in sharp contrast with the $\sqrt{N}$ consistency of the LSDV estimator in the case of a spurious regression estimated from independent panel data, see Phillips and Moon (1999). The statistics proposed by Kao (1999) rely on this consistency, namely on the fact that $\tilde{\beta} \xrightarrow{p} \Psi^{YX}\Psi^{XX-1}$ where Ψ^{YX} is the average long-run covariance between the errors driving $X_{i,t}$ and those driving $Y_{i,t}$ and Ψ^{XX} is the average long covariance matrix of the $X_{i,t}$'s. The presence of common factors destroys this property and consequently also destroys the asymptotic normality of these estimators and of the statistics relying on this result. For the case of stationary idiosyncratic components, our findings are similar to the spurious regression results from time-series analysis. With non-stationary idiosyncratic components we obtain some mixture of time-series and panel spurious regression results in the limiting distributions. It is apparent that the tests are inconsistent when the data has a common factor structure, and size distortions have to be expected which will increase with N. The nuisance parameters in the limiting distributions given in Proposition 1 introduced by the serial correlation in the common factors and idiosyncratic components can be corrected for non-parametrically, i.e. the composite effect of $\Theta_{F\Lambda} + \gamma_1 - \Upsilon$ or $\Theta_{F\Lambda} + \Delta$ can be accounted for. However, it is not possible to identify nuisance parameters associated with the common factors or the idiosyncratic components individually. So, the covariance of $\tilde{B}_{F\Lambda}$ as well as the average long-run covariance matrix of idiosyncratic stochastic trends, Ψ, will in general remain in the limits. The limit of $t_{\tilde{\rho}}$ will be the product of $\sqrt{N}$, the limit of $(\tilde{\rho}-1)$ and the limit of the standard deviation of $(\tilde{\rho}-1)$. Whereas the latter is positive, the driving factor of the limiting distribution of $(\tilde{\rho}-1)$ is $\frac{\int \mathbf{d}B_{F\Lambda}\tilde{B}'_{F\Lambda}}{\int \tilde{B}_{F\Lambda}\tilde{B}'_{F\Lambda}}$ which has a negative expected value. Thus, $t_{\tilde{\rho}}$ can be expected to diverge to $-\infty$.

3.2 Pedroni (1999)

Pedroni (1999) allows for some heterogeneity in the cointegration relationship. He proposes to estimate a first stage regression individually for each panel member to obtain an estimate of β_i from the following cointegrating equation

$$Y_{i,t} = \alpha_i + \beta_i X_{i,t} + u_{i,t}. \tag{14}$$

We now have for each panel unit

$$\tilde{\beta}_i = \Big(\sum_{t=1}^{T} \tilde{Y}_{i,t}\tilde{X}'_{i,t}\Big)\Big(\sum_{t=1}^{T} \tilde{X}_{i,t}\tilde{X}'_{i,t}\Big)^{-1}.$$

Pedroni (1999) proposes two classes of statistics, namely those based on the within-dimension denoted as "panel" statistics, and those based on the between-dimension denoted as "group mean" statistics. For the former group, the residuals from the first stage regression, $\tilde{u}_{i,t} = \tilde{Y}_{i,t} - \tilde{\beta}_i \tilde{X}_{i,t}$, are stacked and a pooled DF regression is estimated as in (13).[2] The group mean statistics are based on averages of individual unit root statistics, derived from

$$\Delta\tilde{u}_{i,t} = (\rho_i - 1)\tilde{u}_{i,t-1} + v_{i,t}, \tag{15}$$

to obtain

$$(\tilde{\rho}_i - 1) = \Big(\sum_{t=2}^{T} \Delta\tilde{u}_{i,t}\tilde{u}_{i,t-1}\Big)\Big(\sum_{t=2}^{T} \tilde{u}^2_{i,t-1}\Big)^{-1}.$$

Consider now the panel-rho statistic denoted by $Z_{\tilde{\rho}_{NT}-1}$ and the group-mean rho statistic $\tilde{Z}_{\tilde{\rho}_{NT}-1}$ given by

$$Z_{\tilde{\rho}_{NT}-1} = \Big(\sum_{i=1}^{N}\sum_{t=2}^{T}(\Delta\tilde{u}_{i,t}\tilde{u}_{i,t-1} - \hat{\boldsymbol{\lambda}}_i)\Big)\Big(\sum_{i=1}^{N}\sum_{t=2}^{T} \tilde{u}^2_{i,t-1}\Big)^{-1}, \tag{16}$$

and

$$\tilde{Z}_{\tilde{\rho}_{NT}-1} = \sum_{i=1}^{N}\bigg(\Big(\sum_{t=2}^{T}(\Delta\tilde{u}_{i,t}\tilde{u}_{i,t-1} - \hat{\boldsymbol{\lambda}}_i)\Big)\Big(\sum_{t=2}^{T} \tilde{u}^2_{i,t-1}\Big)^{-1}\bigg), \tag{17}$$

with $\hat{\boldsymbol{\lambda}}_i = T^{-1}\sum_{s=1}^{J}\omega_{sJ}\sum_{t=s+1}^{T}\tilde{v}_{i,t}\tilde{v}_{i,t-s}$ where $\tilde{v}_{i,t}$ are the residuals of the second stage regression, and J and ω_{sJ} are suitable bandwidth and kernel functions, respectively. For these 2 statistics, we obtain the following limiting results:

Proposition 2 *Given Assumptions 1, 2, 3 and 4:*

(A) Consider the model given by (1), (2) *and* (3),

(a) $\tilde{\beta}_i \Longrightarrow \big(\lambda'_{1i}(\int \tilde{B}^Y_F \tilde{B}^{X'}_F\ \lambda_{2i})\big)\big(\lambda'_{2i}(\int \tilde{B}^X_F \tilde{B}^{X'}_F)\ \lambda_{2i})\big)^{-1} = \tilde{\mathbf{b}}_{iA}$ *as* $T \to \infty$,

(b) $TZ_{\tilde{\rho}_{NT}-1} \Longrightarrow \frac{\sum_{i=1}^{N}\lambda'_{1i}L'_{11}\int \mathbf{d}Q_F\tilde{Q}'_F L_{11}\lambda_{1i}}{\sum_{i=1}^{N}\lambda'_{1i}L'_{11}\int \tilde{Q}_F\tilde{Q}'_F L_{11}\lambda_{i1}}$ *as* $T \to \infty$,

(c) $T\tilde{Z}_{\tilde{\rho}_{NT}-1} \Longrightarrow \sum_{i=1}^{N}\frac{\lambda'_{1i}L'_{11}\int \mathbf{d}Q_F\tilde{Q}'_F L_{11}\lambda_{1i}}{\lambda'_{1i}L'_{11}\int \tilde{Q}_F\tilde{Q}'_F L_{11}\lambda_{i1}}$ *as* $T \to \infty$,

where $\tilde{Q} = \tilde{W}^Y_F - (\int \tilde{W}^Y_F \tilde{W}^X_F{}')(\int \tilde{W}^X_F \tilde{W}^X_F{}')^{-1}\tilde{W}^X_F$, $\tilde{W}_F$ *is a demeaned k-vector standard Brownian motion, and* L_{11} *is upper left element of* L, *the block triangular decomposition of* $\Omega = L'L$.

(B) Consider the model given by (1), (2) *and* (4),

[2] Note that although the estimated DF equation is the same for Kao (1999) and Pedroni (1999), the residuals used in the estimation are obtained from individual regressions instead of a pooled one.

$$
\begin{array}{llll}
 & \tilde{\beta}_i & \Longrightarrow & \left(\lambda_{1i}' \int \tilde{B}_F^Y \tilde{B}_F^{X'}\ \lambda_{2i} + \int \tilde{B}_i^Y \tilde{B}_i^{X'} + \lambda_{1i}' \int \tilde{B}_F^Y \tilde{B}_i^{X'} + \int \tilde{B}_i^Y \tilde{B}_F^{X'}\ \lambda_{2i}\right) \\
(a) & & & \left(\lambda_{2i}' \int \tilde{B}_F^X \tilde{B}_F^{X'}\ \lambda_{2i} + \int \tilde{B}_i^X \tilde{B}_i^{X'} + \lambda_{2i}' \int \tilde{B}_F^X \tilde{B}_i^{X'} + \int \tilde{B}_i^X \tilde{B}_F^{X'}\ \lambda_{2i}\right)^{-1} \\
 & & & = \tilde{\mathbf{b}}_{iB}
\end{array}
$$

as $T \to \infty$,

(b) $TZ_{\tilde{\rho}_{NT}-1} \Longrightarrow \frac{\sum_{i=1}^N (1-\tilde{\mathbf{b}}_{iB})\left(\Lambda_i'(\int \mathbf{d}B_F \tilde{B}_F')\Lambda_i' + \int \mathbf{d}B_i \tilde{B}_i' + \Lambda_i \int \mathbf{d}B_F \tilde{B}_i' + \int \mathbf{d}B_i \tilde{B}_F' \Lambda_i'\right)(1-\tilde{\mathbf{b}}_{iB})'}{\sum_{i=1}^N (1-\tilde{\mathbf{b}}_{iB})\left(\Lambda_i \int \tilde{B}_F \tilde{B}_F' \Lambda_i' + \int \tilde{B}_i \tilde{B}_i' + \Lambda_i \int \tilde{B}_F \tilde{B}_i' + \int \tilde{B}_i \tilde{B}_F' \Lambda_i'\right)(1-\tilde{\mathbf{b}}_{iB})'}$ *as* $T \to \infty$,

(c) $T\tilde{Z}_{\tilde{\rho}_{NT}-1} \Longrightarrow \sum_{i=1}^N \frac{(1-\tilde{\mathbf{b}}_{iB})\left(\Lambda_i'(\int \mathbf{d}B_F \tilde{B}_F')\Lambda_i' + \int \mathbf{d}B_i \tilde{B}_i' + \Lambda_i \int \mathbf{d}B_F \tilde{B}_i' + \int \mathbf{d}B_i \tilde{B}_F' \Lambda_i'\right)(1-\tilde{\mathbf{b}}_{iB})'}{(1-\tilde{\mathbf{b}}_{iB})\left(\Lambda_i \int \tilde{B}_F \tilde{B}_F' \Lambda_i' + \int \tilde{B}_i \tilde{B}_i' + \Lambda_i \int \tilde{B}_F \tilde{B}_i' + \int \tilde{B}_i \tilde{B}_F' \Lambda_i'\right)(1-\tilde{\mathbf{b}}_{iB})'}$ *as* $T \to \infty$.

Proof: see Appendix C.

For the panel-rho and group-mean-rho statistics Pedroni (1999, 2004) derives asymptotic normality when they are properly standardized. In particular, $\sqrt{N}TZ_{\tilde{\rho}_{NT}-1} - \sqrt{N}\theta_2\theta_1^{-1}$ and $N^{-\frac{1}{2}}T\tilde{Z}_{\tilde{\rho}_{NT}-1} - \sqrt{N}\tilde{\theta}_1$ are asymptotically normally distributed for independent panels, where θ_1, θ_2 and $\tilde{\theta}_1$ are means of functionals of Brownian motions (for details see Pedroni, 2004). The results from Proposition 2 indicate that under the DGP we consider, $TZ_{\tilde{\rho}_{NT}-1}$ and $N^{-1}T\tilde{Z}_{\tilde{\rho}_{NT}-1}$ converge, so that the two test-statistics diverge at rate $\sqrt{N}$ when standardized as above. Furthermore, due to the presence of the common factors, the individual statistics will not be independent along the cross-section, so that the use of a CLT to derive asymptotic normality of the average statistic will be invalid. The result is similar to that derived by Lyhagen (2000) for the Im-Pesaran-Shin (IPS) statistics. Also, for independent panels the distributions of $Z_{\tilde{\rho}_{NT}-1}$ and $\tilde{Z}_{\tilde{\rho}_{NT}-1}$ will be nuisance parameter free. For the DGP we consider, this is not true in general. Although the composite effect of serial correlation in the common factors and idiosyncratic components can be corrected for non-parametrically, nuisance parameters coming only from the common factors or from the idiosyncratic components cannot be identified. So, the limiting distributions will in general depend on the long-run covariances of the common and/or idiosyncratic stochastic trends. A special case arises when there is a single common factor in $Y_{i,t}$ and the idiosyncratic components are stationary. Then, $\lambda_{1i}L_{11}$ will cancel from the limits given in Proposition 2 A (b) and (c).

4 A two-step procedure to test for (no)cointegration in the presence of common factors

The previous section shows that standard panel tests for the null of no-cointegration suffer from serious theoretical problems when applied to data that have been generated by a common factor structure. In this section, we propose to tackle the problem using a simple approach based on the Bai and Ng (2004) PANIC methodology.[3]

Note that a related, albeit different, idea is exploited in the recent work of Banerjee and Carrion-i-Silvestre (2005). These authors assume a factor structure for the *disturbance* of a panel static regression model:

$$
\begin{aligned}
Y_{i,t} &= \alpha_i + \beta_i X_{i,t} + u_{i,t} \\
u_{i,t} &= \gamma_i' F_t + E_{i,t},
\end{aligned}
$$

[3]Wagner and Müller-Fürstenberger (2004) use similar ideas in an empirical study of the Kuznets curve.

where F_t and $E_{i,t}$ are the common factors and the idiosyncratic components respectively that can be either $\mathsf{I}(1)$ or $\mathsf{I}(0)$. A similar framework is used by Bai and Kao (2004) for the estimation of a cointegrating relationship in the presence of common factors. Under some conditions that bound the possible heterogeneity, this framework leads to panel statistics for the null of no-cointegration that have the same distribution as those of panel unit root tests and hence are not affected by the number of regressors.[4] This framework makes it however difficult to interpret the case of no-cointegration (spurious regression) if the regressors $X_{i,t}$ also have a factor structure since in the absence of cointegration, assuming a factor structure for the residuals essentially boils down to assuming a factor structure for $Y_{i,t}$ only. It is nevertheless worth relating their framework to the discussion presented in Remark 2, Section 2.

Let us again consider the simple bivariate DGP (9)-(10)[5]. In this set-up, we can logically address the issue of no-cointegration at three different levels.

(i) Testing for idiosyncratic component no-cointegration. This would mean to test the null hypothesis that $(E^Y_{i,t} - \beta_i E^X_{i,t}) \sim \mathsf{I}(1)$ against $(E^Y_{i,t} - \beta_i E^X_{i,t}) \sim \mathsf{I}(0)$,

(ii) Testing for common factor no-cointegration. This would boil down to testing the null that $(F^Y_t - \delta F^Y_t) \sim \mathsf{I}(1)$ against $(F^Y_t - \delta F^Y_t) \sim \mathsf{I}(0)$,

(iii) Testing for panel no-cointegration which would be testing the null that $Y_{i,t} - \beta_i X_{i,t} \sim \mathsf{I}(1)$ against $Y_{i,t} - \beta_i X_{i,t} \sim \mathsf{I}(0)$. Rejecting the null of no-cointegration here requires evidence of idiosyncratic component cointegration as well as of common factor cointegration with an *additional* restriction that the cointegrating vector takes the form $(1, -\frac{\beta_i \lambda_{2,i}}{\lambda_{1,i}})$ which should moreover be constant across the individuals i.

Provided the components have been extracted from the data, see below, case (i) is tested using standard panel tests for no-cointegration given in (16) and (17). Case (ii) can be investigated using standard time series no-cointegration tests such as the Johansen rank test. Case (iii) is slightly more problematic since rejecting the null of panel no-cointegration requires not only factor and idiosyncratic cointegration, but also that the cointegrating vector(s) for the factors is of a very specific form. Panel cointegration arises if (1) there is common factor cointegration with cointegrating vector $(1, -\frac{\beta_i \lambda_{2,i}}{\lambda_{1,i}})$, (2) $\frac{\beta_i \lambda_{2,i}}{\lambda_{1,i}}$ is constant for all i, (3) there is idiosyncratic component cointegration with cointegrating vector $(1, -\beta_i)$. The restrictions between the cointegrating coefficients result from the common factor structure and from the condition that the left null spaces of the common factor and idiosyncratic component cointegration must have a non empty intersection.

There is however an useful indirect way of addressing this question. Consider (11) and write $(F^Y_t - \frac{\beta_i \lambda_{2,i}}{\lambda_{1,i}} F^X_t) \equiv G_t$ and $(E^Y_{i,t} - \beta_i E^X_{i,t}) \equiv E^*_{i,t}$ such that (11) becomes:

$$Y_{i,t} - \beta_i X_{i,t} = \lambda_{1,i} G_t + E^*_{i,t} \tag{18}$$

which is nothing but the parametrization considered in Banerjee and Carrion-I-Silvester (2005). Under this parametrization, $(1, -\frac{\beta_i \lambda_{2,i}}{\lambda_{1,i}})$ will be a cointegrating vector for the common factors if and only if $G_t \sim \mathsf{I}(0)$. One may consequently investigate the hypothesis of panel

[4] A similar set-up is retained by Westerlund (2005) who proposes Durbin-Hausman tests for cointegration in panels.

[5] The discussion extents to a more general set-up.

cointegration using the approach proposed by these authors. If G_t turns out to be $\mathsf{I}(1)$ this does not rule out common factor cointegration. But common factor cointegration can only be investigated when the factors are extracted from the data.

Now we shall outline a sequential testing procedure based on the framework presented in Section 2 that does not restrict the heterogeneity. The approach starts with a decomposition of the data into common factors and idiosyncratic components. It investigates the cointegration properties of the extracted factors and components.

The basic idea of the procedure we propose is to exploit results derived in Bai and Ng (2004) for their PANIC methodology. In particular, under the DGP (1), (2) and (3) or (4), and using the partitioning of $Z_{i,t} = (Y_{i,t}, X'_{i,t})'$, both $Y_{i,t}$ and $X_{i,t}$ are characterized by a factor structure. Since the problems encountered with the usual panel cointegration tests stem from the unit root in the factors, we propose a simple procedure that essentially consists in two different steps.

Step 1. Conduct a preliminary PANIC analysis of each variable $X_{i,t}$ and $Y_{i,t}$ individually to extract the common factors, using for example the principal components approach advocated by Bai and Ng (2004). Test for unit roots in both the factors and the idiosyncratic components using the approach proposed by Bai and Ng (2004) or using a related panel unit root tests valid in the presence of a factor structure such as the one proposed by Breitung and Das (2005).

Step 2.

a. If $\mathsf{I}(1)$ common factors and $\mathsf{I}(0)$ idiosyncratic components are detected, then we face the situation of *cross-member cointegration* and consequently the nonstationarity in the panel is entirely due to a reduced number of common stochastic trends. Cointegration between $Y_{i,t}$ and $X_{i,t}$ can only occur if the common factors for $Y_{i,t}$ cointegrate with those of $X_{i,t}$. The null of no-cointegration between these estimated factors can be tested using a Johansen type of likelihood ratio test for example.

b. If $\mathsf{I}(1)$ common factors and $\mathsf{I}(1)$ idiosyncratic components are detected, we carry out step a on the estimated common factors and we will work with *defactored* series. In contrast to the work of to Banerjee and Carrion-i-Silvestre (2005) however, instead of defactoring the residuals from a static regression we will defactor separately $Y_{i,t}$ and $X_{i,t}$. The defactored $Y_{i,t}$ (e.g. the estimated idiosyncratic components) is simply obtained as $\widehat{E}^Y_{i,t} = \sum_{s=1}^t \widehat{e}^Y_{i,s} = \sum_{s=1}^t (\Delta Y_{i,s} - \widehat{\lambda}'_{1,i}\widehat{f}_s)$ where $\widehat{f}_s$ is a consistent factor estimate of f_t in (2) and $\widehat{\lambda}'_{1,i}$ a consistent estimate of the loading.[6]

Testing for no-cointegration between the defactored data (the estimated idiosyncratic components) can be conducted using standard panel tests for no-cointegration such as those of Pedroni (1999, 2004) given in (16) and (17) while testing for no-cointegration between the common factors (the $\mathsf{I}(1)$ common trends) can again be performed using Johansen's likelihood ratio type tests as in step a.

The rejection of no-cointegration between $Y_{i,t}$ and $X_{i,t}$ only occurs if *both* statistics reject. However, this is a necessary condition. If the three restrictions mentioned above hold as well for the cointegrating vectors, then panel cointegration will hold. If the outcome of step 2.b is that both the common factors and the idiosyncratic components cointegrate one might want to jointly or sequential test the restrictions

[6] In the case of a single factor, the moment estimator defactoring approach of Phillips and Sul (2003) which does not require large N could also be used.

on the cointegrating vectors. The required tests are currently not available with the exception of a homogeneity test on the idiosyncratic component cointegrating vectors proposed by Pedroni (2004b). While in an empirical analysis one would at least compare point estimates of the parameters involved to get further insight into the structure of the model, formal testing of panel no-cointegration could be done using the Banerjee and Carrion-i-Silvestre (2005) test mentioned above.

One should make a few remarks at this stage.

Remark 4. The procedure outlined in steps 1-2 is a sequential test of panel no-cointegration. It is defined as a multiple comparison procedure with the panel no-cointegration hypothesis rejected if both the separate hypotheses of common factor no-cointegration and idiosyncratic component no-cointegration are rejected and the restrictions between the cointegrating vector parameters are not rejected. An approximate test of the joint hypothesis could use the Bonferroni procedure (see e.g. Savin, 1980). In a Monte Carlo simulation, we studied the properties of the joint hypothesis test of factor and idiosyncratic component (no-)cointegration. The results are available upon request. We find that the test is undersized due to the idiosyncratic component (no-)cointegration. Its power properties are shown to be fine in simulations.

Remark 5. The theoretical justification and motivation for this sequential procedure is analogous to the one behind the PANIC approach for panel unit root analysis. In particular, since the DGP implies that all series have a Bai and Ng (2004) representation, we will proceed by analogy with the results derived in Bai and Ng (2004)[7], which exploits the fact that, provided the number of common factors is known or consistently selected using one of the asymptotically consistent model selection procedures discussed in Bai and Ng (2004), then it holds that $T^{-1/2}\sum_{s=2}^{t}\hat{e}_{i,s}^{Y} = T^{-1/2}\sum_{s=2}^{t}e_{i,s}^{Y} + O_p(C_{NT}^{-1})$ where $\hat{e}_{i,t}^{Y}$ is the estimated idiosyncratic component (e.g. the defactored variables), $\hat{e}_{i,t}^{Y} = \Delta Y_{i,s} - \widehat{\lambda}_{1,i}'\widehat{f}_s$, $\widehat{f}_s$ a consistent factor estimate of f_t, $\widehat{\lambda}_{1,i}'$ a consistent estimate of the loading and $C_{NT}^{-1} = \min(N^{1/2}, T^{1/2})$. It holds consequently that $\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor}\hat{e}_{i,t}^{Y} \Longrightarrow B_i^{Y}(r)$, $\forall i$ where $B_i^{Y}(r)$ are the first element of the $(1+m)$-vector Brownian motion $B_i(r)$. $B_i^{Y}(r)$ and $B_j^{Y}(r)$ are furthermore uncorrelated Brownian motions for $i \neq j$. The same holds for $X_{i,t}$.

Consequently, standard panel no-cointegration tests derived under the maintained assumption of independent panel unit, such as those proposed by Pedroni (2004) for example, can be used on the defactored observations.

Remark 6. This approach requires both large N and T which is one of the important limitation. Notice also that this approach will have finite sample properties that can, *at best*, be close to those observed for the tests when applied to a panel data set with independent units. This will be analyzed in the next subsection by Monte Carlo simulations.

Remark 7. If the rank of the long-run covariance matrix of the factors turns out to be smaller than k, that is if the factors cointegrate, then a further step is needed to assess overall lack of cointegration between $Y_{i,t}$ and $X_{i,t}$. No cointegration then requires separability in cointegration as discussed and analyzed in details in Hecq et al. (2002).

[7] A similar result is given in Kapetanios (2004, Theorem 2)

5 Some Monte Carlo Evidence

The approach we propose in the preceding section has several characteristics that call for a Monte Carlo analysis of some of its finite sample properties. In particular, the theoretical foundation requires both large N and T which is not always met in typical applications of panel cointegration techniques. We will focus on the empirical size properties of the proposed approach, namely testing for no-cointegration using "defactored" data, as it was shown that tests designed for cross-sectionally independent data may suffer from dramatic size distortions when applied to panel with cross-member cointegration for example as pointed out by Banerjee et al. (2004). The DGP is a simple bivariate process (i.e. $m = 1$) with $k = 2$ common factors that obeys the representation (1)-(4).

$$
\begin{aligned}
Z_{i,t} &= \Lambda_i F_t + E_{i,t}, \qquad E_{i,t} = \begin{cases} e_{i,t} \\ E_{i,t-1} + e_{i,t} \end{cases}, \\
e_{i,t} &= \varepsilon_{i,t} + \Gamma_i \varepsilon_{i,t-1} \\
F_t &= F_{t-1} + f_t, \\
f_t &= \eta_t + \Phi_1 \eta_{t-1},
\end{aligned}
$$

where $\varepsilon_{i,t} \sim \text{i.i.d.}\, N(0, \Sigma_i), \eta_t \sim \text{i.i.d.}\, N(0, I_2)$. The loading matrix has a diagonal structure

$$
\Lambda_i = \begin{bmatrix} \lambda_{1i} & 0 \\ 0 & \lambda_{2i} \end{bmatrix},
$$

with $\lambda_{1i}, \lambda_{2i} \sim U[-1, 3]$ where U denotes uniform distributions. The remaining parameters are also drawn from independent uniform distributions to allow for some degree of heterogeneity: $\Phi_{11,22} \sim U[0.5, 0.7], \Phi_{12,21} \sim U[0, 0.5], \Sigma_{i,11,22} \sim U[1, 1.4], \Gamma_{11,22} \sim U[0.5, 0.7], \Phi_{12,21} \sim U[0, 0.5]$ and $\Sigma_{i,12,21} \sim U[0, 0.2]$. The sample size has been set to $T \in \{50, 100, 250\}$ and the number of units in the panel is set to $N \in \{25, 50, 100\}$. We consider the rejection frequencies based on 1000 replications[8] for the following statistics:

1. Kao's pooled normalized coefficient (the ρ test) test based on the raw data,
2. Kao's pooled ADF test based on the raw data,
3. Pedroni's panel-t statistics based on raw data,
4. Pedroni's panel-ρ statistics based on raw data,
5. Pedroni's group mean t statistics based on raw data,
6. Pedroni's group mean ρ statistics based on raw data,
7. Pedroni's panel ρ statistics based on defactored data,
8. Pedroni's group mean ρ statistics based on defactored data ,
9. Johansen trace test for the estimated common factors, using the information criterion of Aznar and Salvador (2002) to select the lag length of the VECM.

[8] All experiments are carried out using GAUSS 6.0.

For the last two statistics based on the defactored data, we estimate the number of common factors k using the IC_1 criterion from Bai and Ng (2002) with $k_{max} = 4$. For the ADF type statistics the lag length was selected using the AIC. For the non-parametric correction for serial correlation, we used a quadratic spectral kernel with a bandwidth of $3.21T^{\frac{1}{3}}$ as proposed in Andrews (1991).

The two polar cases that we consider in the simulations are the cases discussed earlier namely:

- the case of cross-member cointegration in which the common factors are $\mathsf{I}(1)$ and the idiosyncratic component are $\mathsf{I}(0)$,
- the case where both common factors and idiosyncratic components are $\mathsf{I}(1)$.

In addition, we consider cases where

- only the common factors are cointegrated but the idiosyncatic components are not cointegrated,
- not cointegrated common factors combined with cointegrated idiosyncratic components,
- cointegration in both the common factors and the idiosyncratic components.

Tables 1 to 5 report the rejection frequencies the 5 cases but where we have in excluded any serial correlation in e_t and in f_t and assume k is known. All statistics considered in these two tables are then also constructed without correction for serial correlation. As expected from the results of the previous section, the usual panel statistics perform badly in cases of cross-member cointegration where the non-stationary is essentially due to the common factors (Table 1). In these cases, as was already observed in Banerjee et al. (2004), standard tests assuming uncorrelated units are very severely oversized even for cases where N and T are both large. When both the common factors and the idiosyncratic components are I(1) (Table 2), we observe that the distributions for Kao's DF_ρ and DF_t statistics diverge to the right. Therefore, these statistics have rejection frequencies of 0 in that case. A similar behaviour is observed when only the common factors cointegrate (Table 3). When there is only cointegration in the idiosyncratic components, Kao's DF_ρ is oversized, while the DF_t has rejection frequencies close to the nominal 5% size. We observe strong size distortions for the Pedroni tests when either the common factors or the idiosyncratic components are cointegrated. When there is no cointegration in the common factors, the size distortions of the group mean statistics are smaller than those for the panel statistics. When cointegration is present in both F_t and $E_{i,t}$ with a common cointegrating vector and homogenous factor loadings such that there is cointegration between $Y_{i,t}$ and $X_{i,t}$ (Table 5), the test statistics of Kao and Pedroni have rejection frequencies of 1.

The last three rows of the tables are the rejection frequencies of the panel tests using the defactored data (column 2-7) and of the system test for the estimated factors (columns 8-10). These are denoted by *Idiosyncratic Panel-ρ*, *Idiosyncratic Group-ρ* and *Aznar/Johansen*. As could be expected, the tests on the defactored data always reject the null in the presence of cross-member cointegration (Table 1) since these components are stationary. When the idiosyncratic components are not cointegrated, we observe that the Idiosyncratic Panel-ρ has rejection frequencies close to the nominal 5% size, while the Idiosyncratic Group-ρ test is slightly undersized. When the idiosyncratic components cointegrate, both tests have rejection frequencies of 1. The Aznar/Johansen test applied to the extracted common factors has approximately the correct size when the common factors do not cointegrate, and a power ranging from 70% to 90% when cointegrated common factors are combined with not cointegrated idiosyncratic components. When both data components are cointegrated, the power of the Aznar/Johansen test is larger than 90% for all combinations of N and T.

Insert Tables 1 to 5 about here

Tables 6 to 10 present simulation results for cases the 5 cases with MA(1) dynamics in the error terms and $k = 2$ common factors, i.e. one common factor in $Y_{i,t}$ and one in $X_{i,t}$. Furthermore, the number of common factors is estimated using the IC_1 criterion of Bai and Ng (2002) with $k_{max} = 4$. We observe that the criterion always picks the correct number of common factors. Both Kao test statistics show strong size distortions when either the common factors or the idiosyncratic components (or both) cointegrate. The Pedroni tests exhibit very strong size distortions in the cross-member cointegration case (Table 6). When non-stationary idiosyncratic components are combined with non-cointegrated or cointegrated common factors (Tables 7 and 8) size distortions are reduced, and the tests are even undersized for some combinations of N and T. When both the common factors and the idiosyncratic components cointegrate (Table 10), the Pedroni tests have rejection frequencies of up to 1. However, as we do not impose homogeneity of the factor loadings, panel cointegration is not present (see the discussion in Section 4).

The tests applied to the estimated idiosyncratic components show rejection frequencies of (close to) 1 when those are stationary or cointegrated. When the idiosyncratic components are not cointegrated, the Idiosyncratic Panel-ρ and Idiosyncratic Group-ρ tests are undersized. The Aznar/Johansen test applied to the estimated common factors is slightly oversized when the common factors do not cointegrate with rejection frequencies between 8% and 15%. When there is cointegration among the common factors, the test has a power between 61% and 92%.

Insert Tables 6 to 10 about here

Tables 11 to 15 present simulation results where we have introduced a second factor in $X_{i,t}$, such that $k = 3$ now. Again estimating the number of common factors using the IC_1 criterion of Bai and Ng (2002), we note that the second common factor of $X_{i,t}$ is not picked up[9]. Nevertheless, simulation results for the Kao and Pedroni tests applied to the raw data and the Aznar/Johansen test applied to the extracted common factors do not change qualitatively compared to the results obtained for $k = 2$. However, the Idiosyncratic Panel-ρ and Idiosyncratic Group-ρ applied to the estimated common components exhibit a reduced power when the common components are cointegrated (Tables 14 and 15), in particular when T=50.

Insert Tables 11 to 15 about here

6 Empirical Illustration

Among the economic hypotheses that are selected to illustrate the feasibility of new techniques in nonstationary econometrics, the purchasing power parity (PPP) hypothesis is without any doubts among the most popular ones. This applications was also chosen in Pedroni (2004) to

[9]Similarly, the PC_1 or BIC_3 criteria from Bai and Ng (2002) only select a sinlge common factor for $X_{i,t}$

illustrate the feasibility of his tests. In his illustration, he also highlights the importance of taking cross-sectional dependence into account even if this was outside the scope of his initial study. Since PPP is a field where persistent cross-sectional dependence is present in the data due to the very nature of the economic problems studied (see Lyhagen, 2000; Banerjee et al, 2004), we will also consider PPP to illustrate the approach discussed earlier in this paper. Following Pedroni (2004), we will focus on a version of the hypothesis known as weak long-run PPP which posits that although nominal exchange rates and aggregate price ratios may move together over long periods, there are reasons to think that in practice the movements may not be perfectly proportional. Motivations for this deviation from the perfect PPP hypothesis are numerous and will not be discussed here (transportation costs, measurement errors, productivity differential, ...). All these factors however speak in favor of allowing for substantial heterogeneity, since under the alternative hypothesis of cointegration there are no reasons to expect the cointegrating vector to be the same for all countries.

The empirical model adopted can be written as

$$s_{i,t} = \alpha_i + \beta_i p_{i,t} + \epsilon_{i,t}, \tag{19}$$

where $s_{i,t}$ is the log nominal bilateral US dollar exchange rate at time t for country i and $p_{i,t}$ is the log price level differential between country i and the United States at time t. A rejection of the null hypothesis of no cointegration in this equation is taken as empirical evidence in favor of the weak PPP hypothesis.

We employ quarterly data (in contrast to Pedroni who uses both monthly and annual IFS data) on nominal exchange rates and consumer price index (CPI) deflators for the postBretton Woods/pre-Euro period from the 1st quarter of 1974 to 3rd quarter of 1998. The cross-section consists of 18 countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Ireland, Italy, Japan, The Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the UK. We then proceed in the usual way by testing for no-cointegration for each series separately in a time series framework. Then we carry out the standard tests for no-cointegration in a panel setting under cross-sectional independence. Finally, we check for a common factor structure and apply the procedure proposed above under cross-sectional dependence. The empirical evidence supports the presence of common factors. It does not lead to rejection of the null of no-cointegration.

For the Kao-ADF statistics the lag length was selected using the AIC. For the non-parametric correction used in the construction of the Pedroni statistics and the Kao-ρ^* statistics, we used a quadratic spectral kernel with a bandwidth of $3.21T^{\frac{1}{3}}$.

INSERT TABLES 16-17 ABOUT HERE

The results are reported in Table 8. Although not reported to save space,[10], we note that the number of rejections of the null of no-cointegration when using the individual time series statistics is too low to favor weak PPP. A similar observation is made in Pedroni (2004). When we consider the standard panel tests applied to the raw data, we see that the panel statistics of Kao clearly reject the null. The same holds for the "within" panel tests of Pedroni, while the group-mean statistics do not reject, globally only lending some weak support to the PPP hypothesis. The aforementioned tests do however not take into account any cross-sectional dependence.

[10]The results are available upon request from the authors.

Following the procedure proposed in Section 3 we first consider a Bai and Ng (2004) type PANIC analysis for the series for $s_{i,t}$ and $p_{i,t}$ separately. The number of common factors is selected using Bai and Ng's (2002) IC1 criterion. For both the exchange rates and the prices a single common factor is selected. The pooled p-value Fisher type unit root tests on the extracted idiosyncratic components do not reject the null hypothesis (the statistics take the value -2.25 and -1.46 respectively). Similarly, the ADF tests for the extracted common factors do not reject the unit root hypothesis since these take the value -2.79 and -2.63 respectively for $\hat{F}_Y$ and $\hat{F}_X$ which is larger than the 5% critical of -3.416. Given these outcomes, the second step of the procedure consists in testing for no-cointegration between the idiosyncratic components as well as testing for the no-cointegration between the estimated factors. The results are reported in Table 9. As it is clear from the entries in the Table, none of the reported statistics reject the null of no-cointegration.

7 Conclusions

In this paper we have considered the problem of testing for (no-)cointegration in panel data set characterized by strong permanent cross-sectional dependencies that take the form of an approximate factor representation inspired by the work of Bai and Ng (2004). We focus on two polar cases that we believe are of empirical relevance namely the case of cross-member cointegration and the case where the panel units have both common and individual specific stochastic trends that are not cointegrated.

For both classes of DGP's, we discuss the homogeneity restrictions for the cointegrating vectors resulting from the presence of common factor cointegration. We study analytically the behavior of several popular test for panel cointegration including Kao (1999) and Pedroni's (1999, 2004) residual-based panel no-cointegration tests that have been widely used in empirical work over the recent years. The results complement and help to understand some of the Monte Carlo reported by Banerjee et al. (2004). For example, when the number of common factors generating the non-stationarity in the panel is kept fixed while the dimension of the panel increases, then the Gaussian limiting results derived for the independent case are not valid anymore. Tests that are based on pooled or LSDV estimation of the underlying panel cointegration static regression may in some cases diverge with $\sqrt{N}$ and hence important size distortions are to be expected already in panels with moderate cross-sectional dimension. Group mean statistics are also affected and are not asymptotically Gaussian anymore.

These observations provide sufficient reason to propose a two-step procedure to address the issue of cointegration testing in panels with common factors. Specifically, we propose to first conduct a PANIC analysis of each series, to defactor the data if $\mathsf{I}(1)$ common factors are found and then to conduct "standard" (panel) cointegration analysis on defactored data and the estimated common factors. This is similar in spirit to the recent work of Banerjee and Carrion-i-Silvestre (2005) and complementary to their approach. One of its advantages is that it covers many sub-cases of interest and allows to have a clear picture of the common/global and iniosyncratic/individual specific components in the panel and about the homogeneity requirements resulting from the occurrence of common factor cointegration. The procedure is simple to apply and makes use of existing tools. Some simulation results show that the procedure we propose seems to have reasonable size properties.

While being attractive due, among other things, to its ease of application and nice properties, some limitations inherent in this approach should be mentioned. A first limitation of the proposed procedure, as well as of that proposed by Banerjee and Carrion-i-Silvestre (2005), is that the theoretical validity relies on both large N and large T which may be unrealistic for applications with "moderate" N and large T. The performance of the proposed procedures, in particular the power properties, in such situation still needs to be further studied even if the size properties reported in Monte Carlo section are promising. If considering large N analysis is clearly inappropriate for the problem under study, then an alternative would be to follow the work of Demetrescu and Tarcolea (2005) who propose a non-linear IV testing approach or to consider the use of bootstrapping techniques that seem to work well from an empirical point of view (see Fachin, 2005). Future work should study the relative merits of these different approaches both theoretically and empirically.

A second limitation worth mentioning lies in the fact that the approach is a residual-based testing procedure and hence suffers from the usual critiques against residual-based tests such as the maintained assumptions of a single cointegrating relationship (if it exists) as well as the imposition of the common factor restriction. Nothing however precludes conceptually to extend the idea developed in this paper to other cointegration techniques that could not suffer from these drawbacks.

References

[1] Andrews, D.W.K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, *Econometrica* 59, No.3, 817-858.

[2] Andrews, D.W.K. and Monahan, C. (1992). An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator, *Econometrica* 60, No.4, 953-966.

[3] Aznar, A. and Salvador, M. (2002). Selecting the Rank of the Cointegration Space and the Form of the Intercept Using an Information Criterion, *Econometric Theory* 18, 926-947.

[4] Bai, J. and Kao, C. (2004). On the Estimation and Inference of a Panel Cointegration Model with Cross-Sectional Dependence, mimeo, Presented at ESEM2004.

[5] Bai, J. and Ng, S. (2004). A PANIC Attack on Unit Roots and Cointegration, *Econometrica* 72, 1127-1177.

[6] Bai, J. and Ng, S. (2002), Determining the number of Factors in Approximate Factor Models, *Econometrica* 70, No.1, 191-221.

[7] Banerjee, A. and Carrion-i-Silvestre, J.L. (2005). Breaking Panel Data Cointegration, mimeo.

[8] Banerjee, A., Marcellino, M. and Osbat, C. (2004). Some Cautions on the Use of Panel Methods for Integrated Series of Macro-Economic Data, *Econometrics Journal* 7, 2, 322-340.

[9] Banerjee, A., Marcellino, M. and Osbat, C. (2005). Testing for PPP: Should We Use Panel Methods?, *Empirical Economics* 30(1), 77-91.

[10] Breitung, J. and Das, S. (2005). Testing for Unit Roots in Panels With a Factor Structure, University of Bonn, unpublished.

[11] Breitung, J. and Pesaran, H.M. (2005). Unit Roots and Cointegration in Panels, forthcoming in Matyas, L. and Sevestre, P. (eds.) *The Econometrics of Panel Data*, Kluwer Academic Publishers, (forthcoming).

[12] Davidson, J. and de Jong, R. (2000). The Functional Central Limit Theorem and Weak Convergence to Stochastic Integrals I: Weakly Dependent Processes, *Econometric Theory* 16, 621-642.

[13] Dees, S., F. di Mauro, M. H. Pesaran and Smith, L.V. (2005). Exploring the International Linkages of the Euro Area: A Global VAR Analysis , CESifoWorking Paper No. 1425. Presented at CESifo Area Conference on Macro, Money and International Finance, February 2005.

[14] Demetrescu, M. and Tarcolea, A.-I. (2005). Panel Cointegration Testing using Nonlinear Instruments, Working Paper, Goethe-University Frankfurt.

[15] Fachin, S. (2005), Long-Run Trends in Internal Migrations in Italy: a Study in Panel Cointegration with Dependent Units, mimeo, University of Rome 'La Sapienza'.

[16] Gengenbach, C., Palm, F.C. and Urbain, J.-P. (2004). Panel Unit Root Tests in the Presence of Cross-Sectional Dependencies: Comparison and Implications for Modelling, METEOR Research Memorandum 04039 Universiteit Maastricht.

[17] Gregoir, S. (2005). Representation and Statistical Analysis of Weakly Linearly Exchangeable Dynamic Panels, Working Paper, CREST.

[18] Hecq, A., Palm, F.C. and Urbain, J.-P. (2002). Separation, Weak Exogeneity and P-T Decomposition in Cointegrated VAR Systems with Common Features, *Econometric Reviews* 21, 273-307.

[19] Im, K.S., Pesaran, M.H. and Shin, Y. (2003). Testing for Unit Roots in Heterogenous Panels, *Journal of Econometrics* 115, 53-74.

[20] Kao, C. (1999). Spurious Regression and Residual-Based Tests for Cointegration in Panel Data, *Journal of Econometrics* 90, 144

[21] Kapetanios, G. (2004). Dynamic Factor Extraction of Cross-Sectional Dependence in Panel Unit Root Tests, Working Paper 509, Queen Mary College, University of London.

[22] Levin, A., Lin, C.F. and Chu C.S.J. (2002). Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties, *Journal of Econometrics* 108, 1-24.

[23] Lyhagen, J. (2000), Why not Use Standard Panel Unit Root Test for Testing PPP, mimeo, Stockholm School of Economics.

[24] Moon, H.R. and Perron, B. (2004). Testing for a Unit Root in Panels with Dynamic Factors, *Journal of Econometrics* 122, 81-126.

[25] Pedroni, P. (1999). Critical Values for Cointegration Tests in Heterogeneous Panels with Multiple Regressors, *Oxford Bulletin of Economics and Statistics,* 61, 653-670.

[26] Pedroni, P. (2004). Panel Cointegration, Asymptotic and Finite Sample Properties of Pooled Time Series tests with an Application to the PPP hypothesis, *Econometric Theory* 20, 597-625.

[27] Pedroni, P. (2004). Social Capital, Barriers to Production and Capital Shares; Implications for the Importance of Parameter Heterogeneity from a Nonstationary Panel Approach, mimeo, Williams College.

[28] Pedroni, P. and Urbain, J.-P. (2001). Cross Member Cointegration in Non-Stationary Panels, mimeo, Universtiteit Maastricht.

[29] Pesaran, M.H. (2003). A Simple Panel Unit Root Test in the Presence of Cross Section Dependence, mimeo, Cambridge University.

[30] Phillips, P.C.B. and Durlauf, S. (1986). Multiple Time Series Regression with Integrated Processes, *Review of Economic Studies* 53, 473-495.

[31] Phillips, P.C.B. and Moon, H.R. (1999). Linear Regression Limit Theory for Nonstationary Panel Data, *Econometrica* 67, 1057-1111.

[32] Phillips, P.C.B. and Sul, D. (2003). Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence, *Econometrics Journal* 6, 217-259.

[33] Savin, N.E. (1980). The Bonferroni and Scheffe' Multiple Comparison Procedures, *Review of Economic Studies* 47, 255-273.

[34] Urbain, J.-P. (2004). Spurious Regression in Nonstationary Panels with Cross-Member Cointegration, mimeo, Presented at ESEM2004.

[35] Wagner, M. and Müller-Fürstenberger, G. (2004). The Carbon Kuznets Curve: A Cloudy Picture Emitted by Bad Econometrics?, Discussion paper 4.18, University of Bern.

[36] Westerlund, J. (2005). New Simple Tests for Panel Cointegration, Working Paper 2005:8, Lund University, Department of Economics.

A Proofs of Lemmas 1 to 3

Given Assumptions 1 to 4, we can summarize some convergence results. In the following lemmas, M is used to denote a generic positive number, not depending on T or N. For a matrix $\boldsymbol{A}$, $\boldsymbol{A} > 0$ denotes that $\boldsymbol{A}$ is positive definite. Furthermore, $\|\boldsymbol{A}\| = trace(\boldsymbol{A}'\boldsymbol{A})^{\frac{1}{2}}$. We write the integral $\int_0^1 B(r)\mathbf{d}r$ as $\int B$, and $\int_0^1 B(r)B(r)'\mathbf{d}r$ as $\int BB'$. Furthermore, $\Longrightarrow$ denotes weak convergence, and $\xrightarrow{p}$ denotes convergence in probability. For any number x, $\lfloor x \rfloor$ denotes the largest integer smaller than x. For any variable $X_{i,t}$, $\tilde{X}_{i,t} = X_{i,t} - \frac{1}{T}\sum_{s=1}^{T} X_{i,s}$. Similarly, for any Brownian motion B, $\tilde{B} = B - \int B$.Throughout the paper we employ sequential limit theory, where we consider $T \to \infty$ followed by $N \to \infty$. Furthermore, for non-random factor loadings, $\bar{\Lambda} = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \Lambda_i$, while for random factor loadings $\bar{\Lambda} = \mathsf{E}(\Lambda_i)$, $\Psi = \mathsf{E}(\Psi_i)$ and $\Delta = \mathsf{E}(\Delta_i)$.

Lemma 1 presents convergence results for the common data component $\Lambda_i F_t$. The limiting distributions are functionals of Brownian motions weighted by the factor loadings, even as $N \to \infty$. These results are intuitive, as we assume a fixed number of common factors. Lemma 2 summarizes the convergence for the idiosyncratic components, where we recover the panel spurious regression results for Phillips and Moon (1999). In Lemma 3, the limits for the cross-products of the common and individual specific components are given. It is evident that these cross-products will only affect limiting distributions for finite N, but as $N \to \infty$ these effects will vanish due to the independence of the shock driving F_t and $E_{i,t}$.

A.1 Lemma 1: Common Component

Lemma 1 *Given Assumptions 1, 2 and 4,*

(a)
$$\begin{array}{rcll} \frac{1}{T}\sum_{t=1}^{T}\Lambda_i f_t F_{t-1}'\Lambda_i' & \Longrightarrow & \Lambda_i(\int \mathbf{d}B_F B_F' + \Theta)\Lambda_i' & \text{as } T\to\infty, \text{ and} \\ \frac{1}{N}\sum_{i=1}^{N}\Lambda_i(\int \mathbf{d}B_F B_F' + \Theta)\Lambda_i' & \xrightarrow{p} & \int \mathbf{d}B_{F\Lambda}B_{F\Lambda}' + \Theta_{F\Lambda} & \text{as } N\to\infty, \end{array}$$

(b)
$$\begin{array}{rcll} \frac{1}{T^2}\sum_{t=1}^{T}\Lambda_i F_t F_t'\Lambda_i' & \Longrightarrow & \Lambda_i(\int B_F B_F')\Lambda_i' & \text{as } T\to\infty, \text{ and} \\ \frac{1}{N}\sum_{i=1}^{N}\Lambda_i(\int B_F B_F')\Lambda_i' & \xrightarrow{p} & \int B_{F\Lambda}B_{F\Lambda}' & \text{as } N\to\infty, \end{array}$$

(c)
$$\begin{array}{rcll} \frac{1}{T}\sum_{t=1}^{T}\Lambda_i f_t \tilde{F}_{t-1}'\Lambda_i' & \Longrightarrow & \Lambda_i(\int \mathbf{d}B_F \tilde{B}_F' + \Theta)\Lambda_i' & \text{as } T\to\infty, \text{ and} \\ \frac{1}{N}\sum_{i=1}^{N}\Lambda_i(\int \mathbf{d}B_F \tilde{B}_F' + \Theta)\Lambda_i' & \xrightarrow{p} & \int \mathbf{d}B_{F\Lambda}\tilde{B}_{F\Lambda}' + \Theta_{F\Lambda} & \text{as } N\to\infty, \end{array}$$

(d)
$$\begin{array}{rcll} \frac{1}{T^2}\sum_{t=1}^{T}\Lambda_i \tilde{F}_t \tilde{F}_t'\Lambda_i' & \Longrightarrow & \Lambda_i(\int \tilde{B}_F \tilde{B}_F')\Lambda_i' & \text{as } T\to\infty, \text{ and} \\ \frac{1}{N}\sum_{i=1}^{N}\Lambda_i(\int \tilde{B}_F \tilde{B}_F')\Lambda_i' & \xrightarrow{p} & \int \tilde{B}_{F\Lambda}\tilde{B}_{F\Lambda}' & \text{as } N\to\infty, \end{array}$$

where $\mathrm{vec}(\int \mathbf{d}B_{F\Lambda}B_{F\Lambda}') = \check{\boldsymbol{\Lambda}}\,\mathrm{vec}(\int \mathbf{d}B_F B_F')$, $\mathrm{vec}(\Theta_{F\Lambda}) = \check{\boldsymbol{\Lambda}}\,\mathrm{vec}(\Theta)$, $\mathrm{vec}(\int B_{F\Lambda}B_{F\Lambda}') = \check{\boldsymbol{\Lambda}}\,\mathrm{vec}(\int B_F B_F')$, $\mathrm{vec}(\int \mathbf{d}B_{F\Lambda}\tilde{B}_{F\Lambda}') = \check{\boldsymbol{\Lambda}}\,\mathrm{vec}(\int \mathbf{d}B_F \tilde{B}_F')$ *and* $\mathrm{vec}(\int \tilde{B}_{F\Lambda}\tilde{B}_{F\Lambda}') = \check{\boldsymbol{\Lambda}}\,\mathrm{vec}(\int \tilde{B}_F \tilde{B}_F')$, *and* $\check{\boldsymbol{\Lambda}} = \mathrm{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}(\Lambda_i \otimes \Lambda_i)$.

A.2 Proof of Lemma 1

For the common factors given in (2) we find the following Beveridge-Nelson (BN) decomposition:

$$F_t = \Phi(1)\sum_{s=1}^{t}\eta_s + \Phi^*(L)(\eta_t - \eta_0) + F_0, \tag{20}$$

where $\Phi^*(L) = \sum_{j=0}^{\infty}\Phi_j^* L^j$ with $\Phi_j^* = -\sum_{l=j+1}^{\infty}\Phi_l$. Now, $\frac{1}{\sqrt{T}}\Phi(1)\sum_{s=1}^{\lfloor rT \rfloor}\eta_s \Longrightarrow \Phi(1)W_F(r) \equiv B_F(r)$ by the FCLT, where W_F is standard Brownian Motion. Furthermore, $\Phi^*(L)(\eta_t - \eta_0)$ is stationary with finite fourth order moments such that $\frac{1}{\sqrt{T}}\Phi^*(L)(\eta_t - \eta_0) \xrightarrow{p} 0$, and F_0 is $O_p(1)$ by assumption.

(a) We have $\frac{1}{T}\sum_{t=1}^{T}\Lambda_i f_t F_{t-1}'\Lambda_i' = \Lambda_i(\frac{1}{T}\sum_{t=1}^{T} f_t F_{t-1}')\Lambda_i'$. Now, $\frac{1}{T}\sum_{t=1}^{T} f_t F_{t-1}' \Longrightarrow \int \mathbf{d}B_F B_F' + \Theta$ as $T \to \infty$ as shown in e.g. Davidson and de Jong (2000), and the result of Lemma 1 (a) follows immediately. Furthermore, $\mathrm{vec}(\Lambda_i(\int \mathbf{d}B_F B_F' + \Theta)\Lambda_i') = (\Lambda_i \otimes \Lambda_i)\,\mathrm{vec}(\int \mathbf{d}B_F B_F' + \Theta)$. As $\mathsf{E}\|(\Lambda_i \otimes \Lambda_i)\|^2 = \mathsf{E}\|\Lambda_i\|^4 \leq M$ by Assumption 2 (i), we can apply a LLN to $\frac{1}{N}\sum_{i=1}^{N}(\Lambda_i \otimes \Lambda_i)$. Denote $\mathrm{plim}(\Lambda_i \otimes \Lambda_i) = \check{\boldsymbol{\Lambda}}$ to obtain the second result of Lemma 1 (a).

(b) The proof of Lemma 1 (b) is similar to that of (a), except that $\frac{1}{T^2}\sum F_t F_t' \Longrightarrow \int B_F B_F'$ as shown in e.g. Phillips and Durlauf (1986) in the first step.

(c) $\frac{1}{T}\sum_{t=1}^T f_t \tilde{F}_{t-1}' = \frac{1}{T}\sum_{t=1}^T f_t F_{t-1}' - \big(\sum_{t=1}^T \frac{f_t}{\sqrt{T}}\big)\big(\frac{1}{T^{\frac{3}{2}}}\sum_{s=1}^T F_s'\big)$. Now, $\frac{1}{T}\sum_{t=1}^T f_t F_{t-1}' \Longrightarrow \int \mathbf{d}B_F B_F' + \Theta$ while $\big(\sum_{t=1}^T \frac{f_t}{\sqrt{T}}\big)\big(\frac{1}{T^{\frac{3}{2}}}\sum_{s=1}^T F_s'\big) \Longrightarrow \int \mathbf{d}B_F(\int B_F)'$ as $T \to \infty$, so that $\frac{1}{T}\sum_{t=1}^T f_t \tilde{F}_{t-1}' \Longrightarrow \int \mathbf{d}B_F \tilde{B}_F' + \Theta$. The remainder of the proof follows the same arguments as above.

(d) Now, $\sum_{t=1}^T \tilde{F}_t \tilde{F}_t' \Longrightarrow \int \tilde{B}_F \tilde{B}_F'$ as shown in Phillips and Moon (1999), and the limit as $N \to \infty$ follows as above.

A.3 Lemma 2: Idiosyncratic Components

Lemma 2 *Given Assumption 3,*

(a)
$$\begin{array}{lcll} \frac{1}{T}\sum_{t=1}^T e_{i,t} S_{i,t-1}' & \Longrightarrow & \big(\int \mathbf{d}B_i B_i' + \Delta_i\big) & \textit{as } T \to \infty, \textit{ and} \\ \frac{1}{N}\sum_{i=1}^N \big(\int \mathbf{d}B_i B_i' + \Delta_i\big) & \stackrel{p}{\longrightarrow} & \Delta & \textit{as } N \to \infty, \end{array}$$

(b)
$$\begin{array}{lcll} \frac{1}{T^2}\sum_{t=1}^T S_{i,t} S_{i,t}' & \Longrightarrow & \int B_i B_i' & \textit{as } T \to \infty, \textit{ and} \\ \frac{1}{N}\sum_{i=1}^N \int B_i B_i' & \stackrel{p}{\longrightarrow} & \frac{1}{2}\Psi & \textit{as } N \to \infty, \end{array}$$

(c)
$$\begin{array}{lcll} \frac{1}{T}\sum_{t=1}^T e_{i,t} \tilde{S}_{i,t-1}' & \Longrightarrow & \big(\int \mathbf{d}B_i \tilde{B}_i' + \Delta_i\big) & \textit{as } T \to \infty, \textit{ and} \\ \frac{1}{N}\sum_{i=1}^N \big(\int \mathbf{d}B_i \tilde{B}_i' + \Delta_i\big) & \stackrel{p}{\longrightarrow} & -\frac{1}{2}\Psi + \Delta & \textit{as } N \to \infty, \end{array}$$

(d)
$$\begin{array}{lcll} \frac{1}{T^2}\sum_{t=1}^T \tilde{S}_{i,t} \tilde{S}_{i,t}' & \Longrightarrow & \int \tilde{B}_i \tilde{B}_i' & \textit{as } T \to \infty, \textit{ and} \\ \frac{1}{N}\sum_{i=1}^N \int \tilde{B}_i \tilde{B}_i' & \stackrel{p}{\longrightarrow} & \frac{1}{6}\Psi & \textit{as } N \to \infty. \end{array}$$

A.4 Proof of Lemma 2

For the partial sum process $S_{i,t} = \sum_{s=1}^t e_{i,s}$ we obtain a BN decomposition

$$S_{i,t} = \Gamma_i(1)\sum_{s=1}^t \varepsilon_{i,s} + \Gamma_i^*(L)(\varepsilon_{i,t} - \varepsilon_{i,0}) + E_{i,0}, \tag{21}$$

where $\Gamma_i^*(L) = \sum_{j=0}^\infty \Gamma_{i,j}^* L^j$ with $\Gamma_{i,j}^* = -\sum_{l=j+1}^\infty \Gamma_{i,l}$. Now, $\frac{1}{\sqrt{T}}S_{i,\lfloor rT\rfloor} \Longrightarrow \Gamma_i(1)\Sigma_i^{\frac{1}{2}} W_i(r) \equiv B_i(r)$ as $T \to \infty$ for all i, where W_i is standard Brownian motion and $\Sigma_i^{\frac{1}{2}}$ is the Cholesky decomposition of Σ_i such that $\Sigma_i^{\frac{1}{2}}\Sigma_i^{\frac{1}{2}\prime} = \Sigma_i$ as shown in Phillips and Moon (1999). Furthermore, B_i and B_j are *i.i.d* over the i-dimension.

(a) We have $\frac{1}{T}\sum_{t=1}^T e_{i,t} S_{i,t-1}' \Longrightarrow \int \mathbf{d}B_i B_i' + \Delta_i$ as $T \to \infty$ as shown in Davidson and de Jong (2000). Now, $\int \mathbf{d}B_i B_i'$ are *i.i.d* across the i-dimension with $\mathsf{E}(\int \mathbf{d}B_i B_i') = 0$ and $E\|\operatorname{vec}(\int \mathbf{d}B_i B_i')\|^2 < M$. So, we can apply a LLN to find $\frac{1}{N}\sum_{i=1}^N \int \mathbf{d}B_i B_i' \stackrel{p}{\longrightarrow} 0$. Furthermore, a LLN also applies such that $\frac{1}{N}\sum_{i=1}^N \Delta_i \stackrel{p}{\longrightarrow} \Delta \equiv \mathsf{E}(\Delta_i)$, which proves the first result.

(b) This result is proven in Phillips and Moon (1999).

(c) $\frac{1}{T}\sum_{t=1}^T e_{i,t} \tilde{S}_{i,t-1}' = \frac{1}{T}\sum_{t=1}^T e_{i,t} S_{i,t-1}' - \frac{1}{T}\sum_{t=1}^T e_{i,t} \bar{S}_i'$, where $\bar{S}_i' = \frac{1}{T}\sum_{t=1}^T S_{i,t}$. Now, $\frac{1}{T}\sum_{t=1}^T e_{i,t} S_{i,t-1}' \Longrightarrow \int \mathbf{d}B_i B_i' + \Delta_i$, while $\frac{1}{T}\sum_{t=1}^T e_{i,t}\bar{S}_i' = \frac{1}{\sqrt{T}}S_T \frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^T S_{i,t}' \Longrightarrow B_i(1)\int B_i$ as $T \to \infty$. So, $\frac{1}{T}\sum_{t=1}^T e_{i,t}\tilde{S}_{i,t-1}' \Longrightarrow \int \mathbf{d}B_i \tilde{B}_i' + \Delta_i$ as $T \to \infty$. Furthermore, $\mathsf{E}(\int \mathbf{d}B_i \tilde{B}_i') = -\frac{1}{2}\Psi$ and hence, using similar arguments as in (a) $\frac{1}{N}\sum_{i=1}^N \int \mathbf{d}B_i \tilde{B}_i' + \Delta_i \stackrel{p}{\longrightarrow} -\frac{1}{2}\Psi + \Delta$ as $N \to \infty$.

(d) See Phillips and Moon (1999).

A.5 Lemma 3

Lemma 3 *Given Assumptions 1, 2, 3 and 4*

$$
\begin{array}{llll}
\text{(a)} & \frac{1}{T}\sum_{t=1}^{T}\Lambda_i F_{t-1}e_{i,t}' & \Longrightarrow \Lambda_i \int B_F \mathbf{d}B_i' & \text{as } T\to\infty, \text{ and} \\
 & \frac{1}{N}\sum_{i=1}^{N}\Lambda_i \int B_F \mathbf{d}B_i' & \stackrel{p}{\longrightarrow} 0 & \text{as } N\to\infty, \\
\text{(b)} & \frac{1}{T}\sum_{t=1}^{T}\Lambda_i f_t S_{i,t-1}' & \Longrightarrow \Lambda_i \int \mathbf{d}B_F B_i' & \text{as } T\to\infty, \text{ and} \\
 & \frac{1}{N}\sum_{i=1}^{N}\Lambda_i \int \mathbf{d}B_F B_i' & \stackrel{p}{\longrightarrow} 0 & \text{as } N\to\infty, \\
\text{(c)} & \frac{1}{T^2}\sum_{t=1}^{T}\Lambda_i F_t S_{i,t}' & \Longrightarrow \Lambda_i \int B_F B_i' & \text{as } T\to\infty, \text{ and} \\
 & \frac{1}{N}\sum_{i=1}^{N}\Lambda_i \int B_F B_i' & \stackrel{p}{\longrightarrow} 0 & \text{as } N\to\infty, \\
\text{(d)} & \frac{1}{T}\sum_{t=1}^{T}\Lambda_i \tilde{F}_{t-1}e_{i,t}' & \Longrightarrow \Lambda_i \int \tilde{B}_F \mathbf{d}B_i' & \text{as } T\to\infty, \text{ and} \\
 & \frac{1}{N}\sum_{i=1}^{N}\Lambda_i \int \tilde{B}_F \mathbf{d}B_i' & \stackrel{p}{\longrightarrow} 0 & \text{as } N\to\infty, \\
\text{(e)} & \frac{1}{T}\sum_{t=1}^{T}\Lambda_i \tilde{F}_{t-1}\tilde{e}_{i,t}' & \Longrightarrow \Lambda_i \int \tilde{B}_F \mathbf{d}B_i' & \text{as } T\to\infty, \text{ and} \\
 & \frac{1}{N}\sum_{i=1}^{N}\Lambda_i \int \tilde{B}_F \mathbf{d}B_i' & \stackrel{p}{\longrightarrow} 0 & \text{as } N\to\infty, \\
\text{(f)} & \frac{1}{T}\sum_{t=1}^{T}\Lambda_i f_t \tilde{S}_{i,t-1}' & \Longrightarrow \Lambda_i \int \mathbf{d}B_F \tilde{B}_i' & \text{as } T\to\infty, \text{ and} \\
 & \frac{1}{N}\sum_{i=1}^{N}\Lambda_i \int \mathbf{d}B_F \tilde{B}_i' & \stackrel{p}{\longrightarrow} 0 & \text{as } N\to\infty, \\
\text{(g)} & \frac{1}{T^2}\sum_{t=1}^{T}\Lambda_i \tilde{F}_t \tilde{S}_{i,t}' & \Longrightarrow \Lambda_i \int \tilde{B}_F \tilde{B}_i' & \text{as } T\to\infty, \text{ and} \\
 & \frac{1}{N}\sum_{i=1}^{N}\Lambda_i \int \tilde{B}_F \tilde{B}_i' & \stackrel{p}{\longrightarrow} 0 & \text{as } N\to\infty,
\end{array}
$$

A.6 Proof of Lemma 3

For each i, the stacked error vector $w_{i,t} = (f_t', e_{i,t}')'$ and the corresponding partial sum process $W_{i,t} = \sum_{s=1}^{t} w_{i,s} = (F_t', S_{i,t}')'$ fulfill the conditions for a FCLT, such that $\frac{1}{\sqrt{T}}W_{i,\lfloor rT\rfloor} \Longrightarrow B_{w_i}(r) = (B_F(r)', B_i(r)')'$. Due to the independence of f_t and $e_{i,t}$, the covariance matrix of B_{w_i} will have zero off-diagonal blocks. Now, for every panel unit i we obtain time series spurious regression results as $T\to\infty$. Furthermore, the functionals of B_F and B_i we obtain in the first step have zero mean and finite variance, and are uncorrelated across the i-dimension of the panel. So, we can apply a LLN to the average to find the limits as $N\to\infty$. We present the proof for (a), (b)-(g) are obtained using a similar line of argumentation.

(a) The limit as $T\to\infty$ follows from applying a spurious regression result as above and noting that $\mathsf{E}(F_{t-1}e_{i,t}) = 0$ for all i and t. Now, taking expectations we find $\mathsf{E}(\Lambda_i\int B_F\mathbf{d}B_i) = 0$, while $\mathsf{E}\|\operatorname{vec}(\Lambda_i\int B_F\mathbf{d}B_i)\|^2 < M$. for all i. Applying a LLN, we find $\frac{1}{N}\sum_{i=1}^{N}\Lambda_i\int B_F\mathbf{d}B_i \stackrel{p}{\longrightarrow} 0$.

B Proof of Proposition 1

B.1 Proposition 1 (a): Convergence of $\tilde{\beta}$

The LSDV estimator of β is given by $\tilde{\beta} = (\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Y}_{i,t}\tilde{X}_{i,t}')(\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{i,t}\tilde{X}_{i,t}')^{-1}$. Consider the numerator

$$
\begin{aligned}
\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Y}_{i,t}\tilde{X}_{i,t}' &= \sum_{i=1}^{N}\sum_{t=1}^{T}(\lambda_{1i}'\tilde{F}_t^Y + \tilde{E}_{i,t}^Y)(\lambda_{2i}'\tilde{F}_t^X + \tilde{E}_{i,t}^X)' \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T}(\lambda_{1i}'\tilde{F}_t^Y\tilde{F}_t^{X\prime}\lambda_{21} + \tilde{E}_{i,t}^Y\tilde{E}_{i,t}^{X\prime} + \lambda_{1i}'\tilde{F}_t^Y\tilde{E}_{i,t}^{X\prime} + \tilde{E}_{i,t}^Y\tilde{F}_t^{X\prime}\lambda_{21}). \qquad (22)
\end{aligned}
$$

If the idiosyncratic term is given by (3), we have $\sum_{i=1}^{N}(O_p(T^2)+O_p(T)+O_p(T)+O_p(T))$ in (22). So, as $T\to\infty$, $\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}\tilde{Y}_{i,t}\tilde{X}_{i,t}' \Longrightarrow \sum_{i=1}^{N}\lambda_{1i}'\int\tilde{B}_F^Y\tilde{B}_F^{X\prime}\lambda_{2i}$ from the first result of Lemma 1 (d). Now, using the second result we obtain $\frac{1}{N}\sum_{i=1}^{N}\lambda_{1i}'\int\tilde{B}_F^Y\tilde{B}_F^{X\prime}\lambda_{2i} \stackrel{p}{\longrightarrow} \int\tilde{B}_{F\Lambda}^Y\tilde{B}_{F\Lambda}^{X\prime}$ as $N\to\infty$, where $\int\tilde{B}_{F\Lambda}^Y\tilde{B}_{F\Lambda}^{X\prime}$ is the $1\times m$ upper right block of $\int\tilde{B}_{F\Lambda}\tilde{B}_{F\Lambda}'$ defined in Lemma 1.

If the idiosyncratic terms are also I(1), such that the DGP includes (4), all terms in (22) are $O_p(T^2)$ when summed over T. Using Lemmas 1 (d), 2 (d) and 3 (g) we find as $T \to \infty$,

$$\sum_{i=1}^{N} \frac{1}{T^2} \sum_{t=1}^{T} \tilde{Y}_{i,t} \tilde{X}'_{i,t} \Longrightarrow \sum_{i=1}^{N} (\lambda'_{1i} \int \tilde{B}_F^Y \tilde{B}_F^{X\prime} \lambda_{2i} + \int \tilde{B}_i^Y \tilde{B}_i^{X\prime} + \lambda'_{1i} \int \tilde{B}_F^Y \tilde{B}_i^{X\prime} + \int \tilde{B}_i^Y \tilde{B}_F^{X\prime} \lambda_{21}).$$

The terms given above are $O_p(N) + O_p(N) + o_p(N) + o_p(N)$, and we obtain

$$\frac{1}{N} \sum_{i=1}^{N} (\lambda'_{1i} \int \tilde{B}_F^Y \tilde{B}_F^{X\prime} \lambda_{2i} + \int \tilde{B}_i^Y \tilde{B}_i^{X\prime} + \lambda'_{1i} \int \tilde{B}_F^Y \tilde{B}_i^{X\prime} + \int \tilde{B}_i^Y \tilde{B}_F^{X\prime} \lambda_{21}) \xrightarrow{p} \int \tilde{B}_{F\Lambda}^Y \tilde{B}_{F\Lambda}^{X}{}' + \frac{1}{6} \Psi^{YX}$$

as $N \to \infty$, where Ψ^{YX} is the upper right $1 \times m$ block of Ψ.

Now the denominator of $\tilde{\beta}$ is given by

$$\begin{aligned}
\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{X}_{i,t} \tilde{X}'_{i,t} &= \sum_{i=1}^{N} \sum_{t=1}^{T} (\lambda'_{1i} \tilde{F}_t^X + \tilde{E}_{i,t}^X)(\lambda'_{2i} \tilde{F}_t^X + \tilde{E}_{i,t}^X)' \\
&= \sum_{i=1}^{N} \sum_{t=1}^{T} (\lambda'_{2i} \tilde{F}_t^X \tilde{F}_t^{X\prime} \lambda_{21} + \tilde{E}_{i,t}^X \tilde{E}_{i,t}^{X\prime} + \lambda'_{2i} \tilde{F}_t^X \tilde{E}_{i,t}^{X\prime} + \tilde{E}_{i,t}^X \tilde{F}_t^{X\prime} \lambda_{21}). \qquad (23)
\end{aligned}$$

Similar to the results for the numerator, the terms in (23) are $\sum_{i=1}^{N} \big(O_p(T^2) + O_p(T) + O_p(T) + O_p(T)\big)$, if the DGP contains (3). Hence, $\sum_{i=1}^{N} \frac{1}{T^2} \sum_{t=1}^{T} \tilde{X}_{i,t} \tilde{X}'_{i,t} \Longrightarrow \sum_{i=1}^{N} \lambda'_{2i} \int \tilde{B}_F^X \tilde{B}_F^{X\prime} \lambda_{2i}$ as $T \to \infty$. Furthermore, the remaining term is $O_p(N)$, and we obtain $\frac{1}{N} \sum_{i=1}^{N} \frac{1}{T^2} \sum_{t=1}^{T} \tilde{X}_{i,t} \tilde{X}'_{i,t} \xrightarrow{\mathcal{L},p} \int \tilde{B}_{F\Lambda}^X \tilde{B}_{F\Lambda}^{X}{}'$ as $T \to \infty$ followed by $N \to \infty$, where $\int \tilde{B}_{F\Lambda}^X \tilde{B}_{F\Lambda}^{X}{}'$ is the lower right $m \times m$ block of $\int \tilde{B}_{F\Lambda} \tilde{B}'_{F\Lambda}$.

If the true DGP contains (4), all terms in the summation over T in (23) are $O_p(T^2)$ and we have

$$\sum_{i=1}^{N} \frac{1}{T^2} \sum_{t=1}^{T} \tilde{X}_{i,t} \tilde{X}'_{i,t} \Longrightarrow \sum_{i=1}^{N} (\lambda'_{2i} \int \tilde{B}_F^X \tilde{B}_F^{X\prime} \lambda_{2i} + \int \tilde{B}_i^X \tilde{B}_i^{X\prime} + \lambda'_{2i} \int \tilde{B}_F^X \tilde{B}_i^{X\prime} + \int \tilde{B}_i^X \tilde{B}_F^{X\prime} \lambda_{21}),$$

as $T \to \infty$. As above, the cross-products between common and idiosyncratic components will vanish in the cross-sectional average as $N \to \infty$, and we find

$$\frac{1}{N} \sum_{i=1}^{N} (\lambda'_{2i} \int \tilde{B}_F^X \tilde{B}_F^{X\prime} \lambda_{2i} + \int \tilde{B}_i^X \tilde{B}_i^{X\prime} + \lambda'_{2i} \int \tilde{B}_F^X \tilde{B}_i^{X\prime} + \int \tilde{B}_i^X \tilde{B}_F^{X\prime} \lambda_{21}) \xrightarrow{p} \int \tilde{B}_{F\Lambda}^X \tilde{B}_{F\Lambda}^{X}{}' + \frac{1}{6} \Psi^{XX}$$

as $N \to \infty$, where Ψ^{XX} is the lower right $m \times m$ block of Ψ.

Combining the results given above yields Proposition 1 A(a) and B(a).

B.2 Proposition 1 (b): Convergence of $\tilde{\rho}$

The residuals from the first stage PLS regression are given by $\tilde{u}_{i,t} = (1 \; - \tilde{\beta}) Z_{i,t} = Y_{i,t} - \tilde{\beta} X_{i,t}$. For the pooled regression given in (13) we have

$$(\tilde{\rho} - 1) = \Big(\sum_{i=1}^{N} \sum_{t=2}^{T} (1 \; - \tilde{\beta}) \Delta Z_{i,t} \tilde{Z}'_{i,t-1} (1 \; - \tilde{\beta})' \Big) \Big(\sum_{i=1}^{N} \sum_{t=2}^{T} (1 \; - \tilde{\beta}) \tilde{Z}_{i,t-1} \tilde{Z}'_{i,t-1} (1 \; - \tilde{\beta})' \Big)^{-1}. \qquad (24)$$

For the numerator consider

$$\begin{aligned}
\sum_{i=1}^{N} \sum_{t=2}^{T} \Delta Z_{i,t} \tilde{Z}'_{i,t-1} &= \sum_{i=1}^{N} \sum_{t=2}^{T} (\Lambda_i f_t + \Delta E_{i,t})(\Lambda_i \tilde{F}_{t-1} + \tilde{E}_{i,t-1})' \\
&= \sum_{i=1}^{N} \sum_{t=2}^{T} (\Lambda_i f_t \tilde{F}'_{t-1} \Lambda'_i + \Delta E_{i,t} \tilde{E}'_{i,t-1} + \Lambda_i f_t \tilde{E}'_{i,t-1} + \Delta E_{i,t} \tilde{F}'_{t-1} \Lambda'_i). \qquad (25)
\end{aligned}$$

From Lemma 1 (c), $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=2}^{T}\Lambda_i f_t \tilde{F}_{t-1}\Lambda_i' \implies \int \mathbf{d}B_{F\Lambda}\tilde{B}_{F\Lambda}' + \Theta_{F\Lambda}$ as $T \to \infty$ followed by $N \to \infty$. If the idiosyncratic terms are I(0), i.e. the true DGP is given by (3),

$$\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta E_{i,t}\tilde{E}_{i,t-1}' = \sum_{i=1}^{N}\sum_{t=2}^{T}\big((e_{i,t}-e_{i,t-1})e_{i,t-1}' - (e_{i,t}-e_{i,t-1})\bar{e}_i'\big),$$

where $\bar{e}_i = \frac{1}{T}\sum_{t=1}^{T} e_{i,t}$. Now, $\frac{1}{T}\sum_{t=2}^{T} e_{i,t}e_{i,t-1}' \xrightarrow{p} \gamma_{i1}$ as $T \to \infty$, with $\gamma_{i1} = \lim_{T\to\infty}\frac{1}{T}\sum_{i=1}^{T}\mathsf{E}(e_{i,t}e_{i,t-1}')$, and $\frac{1}{N}\sum_{i=1}^{N}\gamma_{i1} \xrightarrow{p} \gamma_1$ as $N \to \infty$, with $\gamma_1 \equiv \mathsf{E}(\gamma_{i1})$. Also, $\frac{1}{T}\sum_{t=2}^{T} e_{i,t-1}e_{i,t-1}' \xrightarrow{p} \Upsilon_i$ as $T \to \infty$ and $\frac{1}{N}\sum_{i=1}^{N}\Upsilon_i \xrightarrow{p} \Upsilon$ as $N \to \infty$. Furthermore, $\frac{1}{T}\sum_{t=2}^{T} e_{i,t}\bar{e}_i' \xrightarrow{p} 0$ and $\frac{1}{T}\sum_{t=2}^{T} e_{i,t-1}\bar{e}_i' \xrightarrow{p} 0$ as $T \to \infty$. Hence, $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=2}^{T}\Delta E_{i,t}\tilde{E}_{i,t-1}' \xrightarrow{p} \gamma_1 - \Upsilon$ as $T \to \infty$ followed by $N \to \infty$.

For the third term in (25) we have,

$$\begin{aligned}
\frac{1}{T}\sum_{t=2}^{T}\Lambda_i f_t \tilde{E}_{i,t-1}' &= \frac{1}{T}\sum_{t=2}^{T}\Lambda_i f_t (e_{i,t-1}-\bar{e}_i)' \\
&= \frac{1}{T}\sum_{t=2}^{T}\Lambda_i f_t e_{i,t-1}' - \frac{1}{T}\sum_{t=2}^{T}\Lambda_i f_t \bar{e}_i' \\
&\xrightarrow{p} 0 - 0,
\end{aligned}$$

as $T \to \infty$.

Finally,

$$\begin{aligned}
\frac{1}{T}\sum_{t=2}^{T}\Delta E_{i,t}\tilde{F}_{t-1}'\Lambda_i' &= \frac{1}{T}\sum_{t=2}^{T}(e_{i,t}-e_{i,t-1})\tilde{F}_{t-1}'\Lambda_i' \\
&= \frac{1}{T}e_{i,T}\tilde{F}_{T-1}'\Lambda' - \frac{1}{T}e_{i,1}\tilde{F}_1'\Lambda' - \frac{1}{T}\sum_{t=2}^{T} e_{i,t-1}f_{t-1}'\Lambda_i' \\
&\xrightarrow{p} 0 - 0 - 0,
\end{aligned}$$

as $T \to \infty$. Hence,

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=2}^{T}\Delta Z_{i,t}\tilde{Z}_{i,t-1}' \implies \int \mathbf{d}B_{F\Lambda}\tilde{B}_{F\Lambda}' + \Theta_{F\Lambda} + \gamma_1 - \Upsilon,$$

as $T \to \infty$ followed by $N \to \infty$.

If the idiosyncratic components are $\mathsf{I}(1)$ and their true DGP includes (4), such that $\Delta E_{i,t} = e_{i,t}$ and $\tilde{E}_{i,t-1} = \tilde{S}_{i,t-1}$, using Lemmas 1 (c), 2 (c) and 3 (d) and (f), we obtain

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=2}^{T}\Delta Z_{i,t}\tilde{Z}_{i,t-1}' \implies \int \mathbf{d}B_{F\Lambda}\tilde{B}_{F\Lambda} + \Theta_{F\Lambda} - \frac{1}{2}\Psi + \Delta,$$

as $T \to \infty$ followed by $N \to \infty$.

For the denominator in (24) consider

$$\begin{aligned}
\sum_{i=1}^{N}\sum_{t=2}^{T}\tilde{Z}_{i,t-1}\tilde{Z}_{i,t-1}' &= \sum_{i=1}^{N}\sum_{t=2}^{T}(\Lambda_i\tilde{F}_{t-1}+\tilde{E}_{i,t-1})(\Lambda_i\tilde{F}_{t-1}+\tilde{E}_{i,t-1})' \\
&= \sum_{i=1}^{N}\sum_{t=2}^{T}\big(\Lambda_i\tilde{F}_{t-1}\tilde{F}_{t-1}'\Lambda_i' + \tilde{E}_{i,t-1}\tilde{E}_{i,t-1}' \\
&\qquad + \Lambda_i\tilde{F}_{t-1}\tilde{E}_{i,t-1}' + \tilde{E}_{i,t-1}\tilde{F}_{t-1}'\Lambda_i'\big). \qquad (26)
\end{aligned}$$

If the idiosyncratic components are given by (3), we have $\sum_{i=1}^{N}\big(O_p(T^2)+O_p(T)+O_p(T)+O_p(T)\big)$, in (26). So, $\frac{1}{T^2}\sum_{t=2}^{T}\tilde{Z}_{i,t-1}\tilde{Z}'_{i,t-1} \Longrightarrow \Lambda_i\int\tilde{B}_F\tilde{B}'_F\Lambda'_i$ as $T\to\infty$, and $\frac{1}{N}\sum_{i=1}^{N}\Lambda_i\int\tilde{B}_F\tilde{B}'_F\Lambda'_i \xrightarrow{p} \int\tilde{B}_{F\Lambda}\tilde{B}'_{F\Lambda}$ as $N\to\infty$.

For I(1) idiosyncratic components given by (4), we find using Lemmas 1 (d), 2 (d) and 3(g)

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=2}^{T}\tilde{Z}_{i,t-1}\tilde{Z}'_{i,t-1}\Longrightarrow\int\tilde{B}_{F\Lambda}\tilde{B}'_{F\Lambda}+\frac{1}{6}\Psi$$

as $T\to\infty$ followed by $N\to\infty$. Combining the above given results with those of A (a) or B(a) yields Proposition 1 A (b) and B(b).

B.3 Proposition 1 (c): Divergence of $t_{\tilde{\rho}}$

The t-statistic for $\tilde{\rho}=1$ is given by

$$t_{\tilde{\rho}}=(\tilde{\rho}-1)s^{-1}(\sum_{i=1}^{N}\sum_{t=2}^{T}\tilde{u}_{i,t-1}^2)^{\frac{1}{2}},$$

where

$$\begin{aligned} s^2 &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=2}^{T}(\Delta\tilde{u}_{i,t}-(\tilde{\rho}-1)\tilde{u}_{i,t-1})^2 \\ &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=2}^{T}\big(\Delta\tilde{u}_{i,t}^2+0_p(1)\big). \end{aligned}$$

As

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=2}^{T}\Delta\tilde{u}_{i,t}^2=\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=2}^{T}(1\ -\tilde{\beta})\Delta Z_{i,t}\Delta\tilde{Z}'_{i,t}(1\ -\tilde{\beta})',$$

which is $O_p(1)$ whether the idiosyncratic components are I(0) or I(1), s^2 is $O_p(1)$. Furthermore, $T(\tilde{\rho}-1)$ and $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=2}^{T}\tilde{u}_{i,t-1}\tilde{u}'_{i,t-1}$ are $O_p(1)$ as well whether $E_{i,t}$ is given by (3) or (4), as shown above. Hence,

$$\begin{aligned} t_{\tilde{\rho}} &= \sqrt{N}T(\tilde{\rho}-1)s^{-1}(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=2}^{T}\tilde{u}_{i,t-1}\tilde{u}'_{i,t-1})^{\frac{1}{2}} \\ &= \sqrt{N}O_p(1), \end{aligned}$$

which diverges at rate $\sqrt{N}$ as $T\to\infty$ followed by $N\to\infty$.

C Proof of Proposition 2

C.1 Proposition 2 (a): Convergence of $\tilde{\beta}_i$

For each panel unit i, the estimator of β_i is given by $\tilde{\beta}_i=(\sum_{t=1}^{T}\tilde{Y}_{i,t}\tilde{X}'_{i,t})(\sum_{t=1}^{T}\tilde{X}_{i,t}\tilde{X}'_{i,t})^{-1}$. Consider the numerator

$$\begin{aligned} \sum_{t=1}^{T}\tilde{Y}_{i,t}\tilde{X}'_{i,t} &= \sum_{t=1}^{T}(\lambda'_{1i}\tilde{F}_t^Y+\tilde{E}_{i,t}^Y)(\lambda'_{2i}\tilde{F}_t^X+\tilde{E}_{i,t}^X)' \\ &= \sum_{t=1}^{T}(\lambda'_{1i}\tilde{F}_t^Y\tilde{F}_t^{X\prime}\lambda_{21}+\tilde{E}_{i,t}^Y\tilde{E}_{i,t}^{X\prime}+\lambda'_{1i}\tilde{F}_t^Y\tilde{E}_{i,t}^{X\prime}+\tilde{E}_{i,t}^Y\tilde{F}_t^{X\prime}\lambda_{21}). \qquad (27) \end{aligned}$$

If the idiosyncratic term is given by (3), we have $O_p(T^2)+O_p(T)+O_p(T)+O_p(T)$ in (27). So, as $T \to \infty$, $\frac{1}{T^2}\sum_{t=1}^{T} \tilde{Y}_{i,t}\tilde{X}_{i,t}' \Longrightarrow \lambda_{1i}' \int \tilde{B}_F^Y \tilde{B}_F^{X\prime}\lambda_{2i}$ from the first result of Lemma 1 (d).

If the idiosyncratic terms are also I(1), such that the DGP includes (4), all terms in (27) are $O_p(T^2)$ when summed over T. Using Lemmas 1 (d), 2 (d) and 3 (g) we find as $T \to \infty$,

$$\frac{1}{T^2}\sum_{t=1}^{T} \tilde{Y}_{i,t}\tilde{X}_{i,t}' \Longrightarrow (\lambda_{1i}' \int \tilde{B}_F^Y \tilde{B}_F^{X\prime}\lambda_{2i} + \int \tilde{B}_i^Y \tilde{B}_i^{X\prime} + \lambda_{1i}' \int \tilde{B}_F^Y \tilde{B}_i^{X\prime} + \int \tilde{B}_i^Y \tilde{B}_F^{X\prime}\lambda_{21}).$$

Now the denominator of $\tilde{\beta}_i$ is given by

$$\begin{aligned} \sum_{t=1}^{T} \tilde{X}_{i,t}\tilde{X}_{i,t}' &= \sum_{t=1}^{T} (\lambda_{1i}'\tilde{F}_t^X + \tilde{E}_{i,t}^X)(\lambda_{2i}'\tilde{F}_t^X + \tilde{E}_{i,t}^X)' \\ &= \sum_{t=1}^{T} (\lambda_{2i}'\tilde{F}_t^X \tilde{F}_t^{X\prime}\lambda_{21} + \tilde{E}_{i,t}^X \tilde{E}_{i,t}^{X\prime} + \lambda_{2i}'\tilde{F}_t^X \tilde{E}_{i,t}^{X\prime} + \tilde{E}_{i,t}^X \tilde{F}_t^{X\prime}\lambda_{21}). \end{aligned} \tag{28}$$

Similar to the results for the numerator, the terms in (28) are $O_p(T^2) + O_p(T) + O_p(T) + O_p(T)$, if the DGP contains (3). Hence, $\frac{1}{T^2}\sum_{t=1}^{T} \tilde{X}_{i,t}\tilde{X}_{i,t}' \Longrightarrow \lambda_{2i}' \int \tilde{B}_F^X \tilde{B}_F^{X\prime}\lambda_{2i}$ as $T \to \infty$.

If the true DGP contains (4), all terms in (28) are $O_p(T^2)$ and we have

$$\frac{1}{T^2}\sum_{t=1}^{T} \tilde{X}_{i,t}\tilde{X}_{i,t}' \Longrightarrow (\lambda_{2i}' \int \tilde{B}_F^X \tilde{B}_F^{X\prime}\lambda_{2i} + \int \tilde{B}_i^X \tilde{B}_i^{X\prime} + \lambda_{2i}' \int \tilde{B}_F^X \tilde{B}_i^{X\prime} + \int \tilde{B}_i^X \tilde{B}_F^{X\prime}\lambda_{21}),$$

as $T \to \infty$.

Combining the results given above yields Proposition 2 A(a) and B(a).

C.2 Proposition 2 (b): Convergence of $Z_{\tilde{\rho}_{NT}-1}$ and $\tilde{Z}_{\tilde{\rho}_{NT}-1}$

The residuals from the individual first stage regression are given by $\tilde{u}_{i,t} = (1,\ -\tilde{\beta}_i)Z_{i,t} = Y_{i,t} - \tilde{\beta}_i X_{i,t}$. Consider first

$$\sum_{t=2}^{T} \Delta\tilde{u}_{i,t}\tilde{u}_{i,t-1} = \sum_{t=2}^{T} (1,\ -\tilde{\beta}_i)\Delta Z_{i,t}\tilde{Z}_{i,t-1}'(1,\ -\tilde{\beta}_i)'. \tag{29}$$

Now,

$$\begin{aligned} \sum_{t=2}^{T} \Delta Z_{i,t}\tilde{Z}_{i,t-1}' &= \sum_{t=2}^{T} (\Lambda_i f_t + \Delta E_{i,t})(\Lambda_i \tilde{F}_{t-1} + \tilde{E}_{i,t-1})' \\ &= \sum_{t=2}^{T} (\Lambda_i f_t \tilde{F}_{t-1}'\Lambda_i' + \Delta E_{i,t}\tilde{E}_{i,t-1}' + \Lambda_i f_t \tilde{E}_{i,t-1}' + \Delta E_{i,t}\tilde{F}_{t-1}'\Lambda_i'). \end{aligned} \tag{30}$$

From Lemma 1 (c), $\frac{1}{T}\sum_{t=2}^{T} \Lambda_i f_t \tilde{F}_{t-1}\Lambda_i' \Longrightarrow \int \Lambda_i(\mathbf{d}B_F \tilde{B}_F + \Theta)\Lambda_i'$ as $T \to \infty$. If the idiosyncratic terms are I(0), i.e. the true DGP is given by (3),

$$\sum_{t=2}^{T} \Delta E_{i,t}\tilde{E}_{i,t-1}' = \sum_{t=2}^{T} \big((e_{i,t} - e_{i,t-1})e_{i,t-1}' - (e_{i,t} - e_{i,t-1})\bar{e}_i\big),$$

where $\bar{e}_i = \frac{1}{T}\sum_{t=1}^{T} e_{i,t}$. Now, $\frac{1}{T}\sum_{t=2}^{T} e_{i,t}e_{i,t-1}' \xrightarrow{p} \gamma_{i1}$ as $T \to \infty$, with $\gamma_{i1} = \lim_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T} \mathsf{E}(e_{i,t}e_{i,t-1})$. Also, $\frac{1}{T}\sum_{t=2}^{T} e_{i,t-1}e_{i,t-1}' \xrightarrow{p} \Upsilon_i$ as $T \to \infty$. Furthermore, $\frac{1}{T}\sum_{t=2}^{T} e_{i,t}\bar{e}_i' \xrightarrow{p} 0$ and $\frac{1}{T}\sum_{t=2}^{T} e_{i,t-1}\bar{e}_i' \xrightarrow{p} 0$ as $T \to \infty$. Hence, $\frac{1}{T}\sum_{t=2}^{T} \Delta E_{i,t}\tilde{E}_{i,t-1}' \xrightarrow{p} \gamma_{i1} - \Upsilon_i$ as $T \to \infty$.

For the third term in (30) we have,

$$\begin{aligned}\frac{1}{T}\sum_{t=2}^{T}\Lambda_i f_t\tilde{E}'_{i,t-1} &= \frac{1}{T}\sum_{t=2}^{T}\Lambda_i f_t(e_{i,t-1}-\bar{e}_i)' \\ &= \frac{1}{T}\sum_{t=2}^{T}\Lambda_i f_t e'_{i,t-1} - \frac{1}{T}\sum_{t=2}^{T}\Lambda_i f_t\bar{e}'_i \\ &\xrightarrow{p} 0-0,\end{aligned}$$

as $T\to\infty$. Finally,

$$\begin{aligned}\frac{1}{T}\sum_{t=2}^{T}\Delta E_{i,t}\tilde{F}'_{t-1}\Lambda'_i &= \frac{1}{T}\sum_{t=2}^{T}(e_{i,t}-e_{i,t-1})\tilde{F}'_{t-1}\Lambda'_i \\ &= \frac{1}{T}e_{i,T}\tilde{F}'_{T-1}\Lambda' - \frac{1}{T}e_{i,1}\tilde{F}'_1\Lambda' - \frac{1}{T}\sum_{t=2}^{T}e_{i,t-1}f'_{t-1}\Lambda'_i \\ &\xrightarrow{p} 0-0-0,\end{aligned}$$

as $T\to\infty$. Hence,

$$\frac{1}{T}\sum_{t=2}^{T}\Delta Z_{i,t}\tilde{Z}'_{i,t-1} \Longrightarrow \Lambda_i(\int \mathbf{d}B_F\tilde{B}_F+\Theta)\Lambda'_i+\gamma_{i1}-\Upsilon_i,$$

as $T\to\infty$.

If the idiosyncratic components are I(1) and their true DGP includes (4), such that $\Delta E_{i,t}=e_{i,t}$ and $\tilde{E}_{i,t-1}=\tilde{S}_{i,t-1}$, using Lemmas 1 (c), 2 (c) and 3 (d) and (f), we obtain

$$\frac{1}{T}\sum_{t=2}^{T}\Delta Z_{i,t}\tilde{Z}'_{i,t-1} \Longrightarrow \big(\Lambda'_i(\int \mathbf{d}B_F\tilde{B}'_F+\Theta)\Lambda'_i+\int \mathbf{d}B_i\tilde{B}'_i+\Delta_i+\Lambda_i\int \mathbf{d}B_F\tilde{B}'_i+\int \mathbf{d}B_i\tilde{B}'_F\Lambda'_i\big),$$

as $T\to\infty$.

Furthermore, note that the residuals $\tilde{v}_{i,t}=\Delta\tilde{u}_{i,t}+o_p(1)$ regardless of whether they were obtained from the poooled regression (13) or the individual regression (15). Now,

$$\begin{aligned}\hat{\boldsymbol{\lambda}}_i &= T^{-1}\sum_{s=1}^{J}\omega_{sJ}\sum_{t=s+1}^{T}\tilde{v}_{i,t}\tilde{v}_{i,t-s} \\ &= T^{-1}\sum_{s=1}^{J}\omega_{sJ}\sum_{t=s+1}^{T}\Delta\tilde{u}_{i,t}\Delta\tilde{u}_{i,t-s}+o_p(1) \\ &= T^{-1}\sum_{s=1}^{J}\omega_{sJ}\sum_{t=s+1}^{T}(1,\ -\tilde{\beta}_i)\Delta\tilde{Z}_{i,t}\Delta\tilde{Z}'_{i,t-s}(1,\ -\tilde{\beta}_i)'+o_p(1).\end{aligned}$$

Expanding $\Delta\tilde{Z}_{i,t}\Delta\tilde{Z}'_{i,t-s}$ in terms of the common factors and unobserved components we obtain the following four terms and convergence results for suitable choices of bandwidth J and kernel function ω_{sJ}. First,

$$T^{-1}\sum_{s=1}^{J}\omega_{sJ}\sum_{t=s+1}^{T}\Lambda_i\tilde{f}_{i,t}\tilde{f}'_{i,t-s}\Lambda'_i \xrightarrow{p} \Lambda_i\Omega\Lambda'_i. \tag{31}$$

Next,

$$T^{-1}\sum_{s=1}^{J}\omega_{sJ}\sum_{t=s+1}^{T}\Lambda_i\tilde{f}_{i,t}\Delta\tilde{E}'_{i,t-s} \xrightarrow{p} 0, \tag{32}$$

and

$$T^{-1}\sum_{s=1}^{J}\omega_{sJ}\sum_{t=s+1}^{T}\Delta\tilde{E}_{i,t}\tilde{f}'_{i,t-s}\Lambda_i \xrightarrow{p} 0, \tag{33}$$

due to the independence of common factors and idiosyncratic components. Finally,

$$T^{-1}\sum_{s=1}^{J}\omega_{sJ}\sum_{t=s+1}^{T}\Delta\tilde{E}_{i,t}\Delta\tilde{E}'_{i,t-s} \xrightarrow{p} \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathsf{E}(e_{i,t}\tilde{E}_{i,t}), \tag{34}$$

which is $\gamma_{1i} - \Upsilon_i$ if the idiosyncratic components are stationary, and Δ_i if they are I(1).

Now consider

$$\sum_{t=2}^{T}\Delta\tilde{u}_{i,t}\tilde{u}_{i,t-1} = \sum_{t=2}^{T}(1,\ -\tilde{\beta}_i)\Delta Z_{i,t}\tilde{Z}'_{i,t-1}(1,\ -\tilde{\beta}_i)'. \tag{35}$$

We have

$$\begin{aligned}\sum_{t=2}^{T}\tilde{Z}_{i,t-1}\tilde{Z}'_{i,t-1} &= \sum_{t=2}^{T}(\Lambda_i\tilde{F}_{t-1}+\tilde{E}_{i,t-1})(\Lambda_i\tilde{F}_{t-1}+\tilde{E}_{i,t-1})' \\ &= \sum_{t=2}^{T}\big(\Lambda_i\tilde{F}_{t-1}\tilde{F}'_{t-1}\Lambda'_i + \tilde{E}_{i,t-1}\tilde{E}'_{i,t-1} \\ &\qquad + \Lambda_i\tilde{F}_{t-1}\tilde{E}'_{i,t-1} + \tilde{E}_{i,t-1}\tilde{F}'_{t-1}\Lambda'_i\big). \end{aligned}\tag{36}$$

If the idiosyncratic components are given by (3), when summed over T the first term in (36) is $O_p(T^2)$, while the remaining three are $O_p(T)$. So, $\frac{1}{T^2}\sum_{t=2}^{T}\tilde{Z}_{i,t-1}\tilde{Z}'_{i,t-1} \Longrightarrow \Lambda_i\int\tilde{B}_F\tilde{B}'_F\Lambda'_i$ as $T\to\infty$.

For $\mathsf{I}(1)$ idiosyncratic components given by (4), we find using Lemmas 1 (d), 2 (d) and 3(g)

$$\frac{1}{T^2}\sum_{t=2}^{T}\tilde{Z}_{i,t-1}\tilde{Z}'_{i,t-1} \Longrightarrow \big(\Lambda_i\int\tilde{B}_F\tilde{B}'_F\Lambda'_i + \int\tilde{B}_i\tilde{B}'_i + \Lambda_i\int\tilde{B}_F\tilde{B}'_i + \int\tilde{B}_i\tilde{B}'_F\Lambda'_i\big)$$

as $T\to\infty$.

Next, we partition the long-run covariance matrix of the common non-stationary factors Ω conformable to the partition of $B_F = (B_F^{Y\prime}, B_F^{X\prime})'$, such that

$$\Omega = \begin{bmatrix}\Omega_{11} & \Omega'_{21}\\ \Omega_{21} & \Omega_{22}\end{bmatrix}.$$

We use the block-triangular decomposition $\Omega = L'L$, with

$$L = \begin{bmatrix}L_{11} & 0\\ L_{21} & L_{22}\end{bmatrix},$$

where $L_{11} = (\Omega_{11} - \Omega'_{21}\Omega_{22}^{-1}\Omega_{21})^{\frac{1}{2}}$, $L_{21} = \Omega_{22}^{-\frac{1}{2}}\Omega_{21}$, and $L_{22} = \Omega_{22}^{\frac{1}{2}}$. Note that $\Omega_{22} > 0$ by Assumption 1.

Now, $\tilde{B}_F = L'\tilde{W}_F$, where $\tilde{W}_F$ is a demeaned k-vector standard Brownian motion. Furthermore, denote $\boldsymbol{\eta}'_i = (1, -\tilde{\boldsymbol{b}}_{iA})$, and $\boldsymbol{\kappa}' = (\boldsymbol{I}_{k_Y},\ -(\int\tilde{W}_F^Y\tilde{W}_F^{X\prime})(\int\tilde{W}_F^X\tilde{W}_F^{X\prime})^{-1}))$. Then, $L\Lambda'_i\boldsymbol{\eta}_i = \boldsymbol{\kappa}L_{11}\lambda_{1i}$, and $\boldsymbol{\eta}'_i\tilde{B}_F = \lambda'_{1i}L'_{11}\tilde{Q}_F$, with $\tilde{Q}_F = \tilde{W}_F^Y - (\int\tilde{W}_F^Y\tilde{W}_F^{X\prime})(\int\tilde{W}_F^X\tilde{W}_F^{X\prime})^{-1}\tilde{W}_F^X$. Finally,

$$\boldsymbol{\eta}'_i\int\mathbf{d}B_F\tilde{B}'_F\boldsymbol{\eta}_i = \lambda'_{1i}L'_{11}\int\mathbf{d}Q_F\tilde{Q}'_F L_{11}\lambda_{1i},$$

and

$$\boldsymbol{\eta}'_i\int\tilde{B}_F\tilde{B}'_F\boldsymbol{\eta}_i = \lambda'_{1i}L'_{11}\int\tilde{Q}_F\tilde{Q}'_F L_{11}\lambda_{1i}.$$

Combining the above given results with those of A (a) or B (a) yields the convergence results for $Z_{\tilde{\rho}_{NT}-1}$ and $\tilde{Z}_{\tilde{\rho}_{NT}-1}$.

D Tables

Table 1: $k = 2$ common factors; cross-member cointegration without serial correlation.

N	25	50	100	25	50	100	25	50	100
Raw data									
T	Kao DF_ρ			Kao DF_t			Pedroni $Panel-\rho$		
50	0.80	0.83	0.85	0.27	0.27	0.32	0.90	0.91	0.92
100	0.81	0.85	0.87	0.26	0.30	0.33	0.90	0.93	0.93
250	0.83	0.89	0.88	0.28	0.34	0.37	0.92	0.94	0.95
Raw data									
T	Pedroni $Panel-t$			Pedroni $Group-\rho$			Pedroni $Group-t$		
50	0.93	0.93	0.93	0.77	0.78	0.83	0.86	0.85	0.88
100	0.92	0.93	0.93	0.80	0.84	0.84	0.85	0.87	0.86
250	0.94	0.94	0.95	0.84	0.85	0.88	0.87	0.88	0.91
Estimated components									
T	Idiosyncratic $Panel-\rho$			Idiosyncratic $Group-\rho$			Aznar/Johansen		
50	1.00	1.00	1.00	1.00	1.00	1.00	0.08	0.08	0.06
100	1.00	1.00	1.00	1.00	1.00	1.00	0.04	0.04	0.05
250	1.00	1.00	1.00	1.00	1.00	1.00	0.05	0.05	0.05

$F_t = \sum_{s=1}^{t} \eta_s$ where $\eta_t \sim iid\ N(0, I_2)$, $E_{i,t} = \varepsilon_{i,t}$ where $\varepsilon_{i,t} \sim iid\ N(0, I_2)$ and $\Lambda_i = I_2$ for all i. Rejection frequencies are based on 5% asymptotic critical values.

Table 2: $k = 2$ common factors; I(1) common factors and I(1) idiosyncratic components without serial correlation.

N	25	50	100	25	50	100	25	50	100
Raw data									
T	Kao DF_ρ			Kao DF_t			Pedroni $Panel-\rho$		
50	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.23	0.33
100	0.00	0.00	0.00	0.00	0.00	0.00	0.19	0.31	0.36
250	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.29	0.38
Raw data									
T	Pedroni $Panel-t$			Pedroni $Group-\rho$			Pedroni $Group-t$		
50	0.46	0.51	0.57	0.10	0.12	0.18	0.34	0.42	0.49
100	0.44	0.49	0.58	0.13	0.21	0.24	0.30	0.40	0.45
250	0.44	0.48	0.54	0.19	0.21	0.33	0.30	0.33	0.43
Estimated components									
T	Idiosyncratic $Panel-\rho$			Idiosyncratic $Group-\rho$			Aznar/Johansen		
50	0.04	0.04	0.12	0.01	0.02	0.01	0.08	0.08	0.06
100	0.06	0.05	0.05	0.02	0.02	0.01	0.04	0.04	0.05
250	0.07	0.06	0.04	0.06	0.03	0.03	0.06	0.04	0.05

$F_t = \sum_{s=1}^{t} \eta_s$ where $\eta_t \sim iid\ N(0, I_2)$, $E_{i,t} = \sum_{s=1}^{t} \varepsilon_{i,s}$ where $\varepsilon_{i,t} \sim iid\ N(0, I_2)$ and $\Lambda_i = I_2$ for all i. Rejection frequencies are based on 5% asymptotic critical values.

Table 3: $k = 2$ common factors; cointegration in F_t but not in $E_{i,t}$ without serial correlation.

N	25	50	100	25	50	100	25	50	100
					Raw data				
T	Kao DF_ρ			Kao DF_t			Pedroni $Panel - \rho$		
50	0.00	0.00	0.00	0.00	0.00	0.00	0.45	0.59	0.73
100	0.00	0.00	0.00	0.00	0.00	0.00	0.60	0.79	0.85
250	0.00	0.00	0.00	0.00	0.00	0.00	0.70	0.85	0.93
					Raw data				
T	Pedroni $Panel - t$			Pedroni $Group - \rho$			Pedroni $Group - t$		
50	0.74	0.82	0.92	0.38	0.51	0.65	0.70	0.82	0.92
100	0.79	0.91	0.93	0.59	0.80	0.88	0.72	0.89	0.93
250	0.83	0.92	0.96	0.78	0.89	0.96	0.77	0.90	0.94
					Estimated components				
T	Idiosyncratic $Panel - \rho$			Idiosyncratic $Group - \rho$			Aznar/Johansen		
50	0.04	0.04	0.01	0.01	0.02	0.01	0.73	0.83	0.87
100	0.06	0.05	0.04	0.02	0.02	0.01	0.82	0.87	0.92
250	0.08	0.06	0.04	0.06	0.03	0.03	0.68	0.83	0.90

$F_t^Y = \sum_{s=1}^{t} \eta_s^Y$, $F_t^X = F_t^Y + \eta_t^X$, where $\eta_t = (\eta_t^Y, \eta_t^X)' \sim iid\ N(0, I_2)$,
$E_{i,t} = \sum_{s=1}^{T} \varepsilon_{i,s}$ where $\varepsilon_{i,t} \sim iid\ N(0, I_2)$ and $\Lambda_i = I_2$ for all i.
Rejection frequencies are based on 5% asymptotic critical values.

Table 4: $k = 2$ common factors; no cointegration in F_t but cointegration in $E_{i,t}$ without serial correlation.

N	25	50	100	25	50	100	25	50	100
				Raw data					
T	Kao DF_ρ			Kao DF_t			Pedroni $Panel-\rho$		
50	0.16	0.18	0.17	0.02	0.04	0.03	0.58	0.60	0.64
100	0.20	0.24	0.25	0.02	0.04	0.05	0.60	0.66	0.69
250	0.25	0.30	0.32	0.03	0.04	0.07	0.61	0.70	0.75
				Raw data					
T	Pedroni $Panel-t$			Pedroni $Group-\rho$			Pedroni $Group-t$		
50	0.72	0.73	0.74	0.39	0.39	0.46	0.55	0.57	0.61
100	0.70	0.73	0.77	0.43	0.50	0.53	0.54	0.60	0.63
250	0.70	0.78	0.78	0.47	0.52	0.57	0.52	0.58	0.64
				Estimated components					
T	Idiosyncratic $Panel-\rho$			Idiosyncratic $Group-\rho$			Aznar/Johansen		
50	1.00	1.00	1.00	1.00	1.00	1.00	0.07	0.08	0.07
100	1.00	1.00	1.00	1.00	1.00	1.00	0.05	0.03	0.05
250	1.00	1.00	1.00	1.00	1.00	1.00	0.05	0.05	0.05

$F_t = \sum_{s=1}^{t} \eta_s$, where $\eta_t \sim iid\ N(0, I_2)$, $E_{i,t}^Y = \sum_{s=1}^{T} \varepsilon_{i,s}^Y$, $E_{i,t}^X = E_{i,t}^Y + \varepsilon_{i,s}^X$,
where $\varepsilon_{i,t} = (\varepsilon_{i,t}^Y, \varepsilon_{i,t}^X)' \sim iid\ N(0, I_2)$ and $\Lambda_i = I_2$ for all i.
Rejection frequencies are based on 5% asymptotic critical values.

Table 5: $k = 2$ common factors; cointegration in F_t and $E_{i,t}$ with common cointegrating vector without serial correlation.

N	25	50	100	25	50	100	25	50	100
				Raw data					
T	Kao DF_ρ			Kao DF_t			Pedroni $Panel-\rho$		
50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
250	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
				Raw data					
T	Pedroni $Panel-t$			Pedroni $Group-\rho$			Pedroni $Group-t$		
50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
250	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
				Estimated components					
T	Idiosyncratic $Panel-\rho$			Idiosyncratic $Group-\rho$			Aznar/Johansen		
50	1.00	1.00	1.00	1.00	1.00	1.00	0.90	0.90	0.91
100	1.00	1.00	1.00	1.00	1.00	1.00	0.97	0.96	0.97
250	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.99	0.99

$F_t^Y = \sum_{s=1}^t \eta_s^Y$, $F_t^X = F_t^Y + \eta_t^X$, where $\eta_t \sim iidN(0, I_2)$, $E_{i,t}^Y = \sum_{s=1}^T \varepsilon_{i,s}^Y$, $E_{i,t}^X = E_{i,t}^Y + \varepsilon_{i,s}^X$, where $\varepsilon_{i,t} = (\varepsilon_{i,t}^Y, \varepsilon_{i,t}^X)' \sim iidN(0, I_2)$ and $\Lambda_i = I_2$ for all i.
Rejection frequencies are based on 5% asymptotic critical values.

Table 6: $k = 2$ common factors; cross-member cointegration with MA(1) errors

N	25	50	100	25	50	100	25	50	100
				Raw data					
T	Kao - ρ^*			Kao - ADF			Pedroni - $Panel - \rho$		
50	0.27	0.32	0.35	0.48	0.50	0.54	0.68	0.90	0.88
100	0.39	0.47	0.49	0.54	0.62	0.62	0.84	0.96	0.95
250	0.52	0.54	0.55	0.64	0.67	0.69	0.93	1.00	0.96
				Raw data					
T	Pedroni - $Panel - t$			Pedroni - $Group - \rho$			Pedroni - $Group - t$		
50	0.76	0.92	0.91	0.33	0.67	0.62	0.52	0.79	0.77
100	0.83	0.96	0.94	0.67	0.94	0.89	0.67	0.89	0.85
250	0.92	0.99	0.95	0.88	1.00	0.94	0.78	0.95	0.88
				Estimated components					
T	Idiosyncratic- $Panel - t$			Idiosyncratic - $Group - \rho$			Aznar/Johansen		
50	1.00	1.00	1.00	1.00	1.00	1.00	0.12	0.12	0.11
100	1.00	1.00	1.00	1.00	1.00	1.00	0.09	0.11	0.09
250	1.00	1.00	1.00	1.00	1.00	1.00	0.08	0.10	0.08

Rejection frequencies are based on 5% asymptotic critical values.

Table 7: $k = 2$ common factors; I(1) common factors and I(1) idiosyncratic components with MA(1) errors

N	25	50	100	25	50	100	25	50	100
				Raw data					
T	Kao - ρ^*			Kao - ADF			Pedroni - $Panel - \rho$		
50	0.17	0.17	0.23	0.59	0.65	0.69	0.00	0.00	0.00
100	0.23	0.28	0.36	0.63	0.74	0.75	0.02	0.02	0.03
250	0.34	0.39	0.45	0.74	0.81	0.80	0.10	0.08	0.14
				Raw data					
T	Pedroni - $Panel - t$			Pedroni - $Group - \rho$			Pedroni - $Group - t$		
50	0.03	0.02	0.04	0.00	0.00	0.00	0.03	0.02	0.04
100	0.06	0.04	0.08	0.00	0.00	0.01	0.04	0.03	0.06
250	0.13	0.10	0.18	0.02	0.01	0.04	0.07	0.05	0.10
				Estimated components					
T	Idiosyncratic- $Panel - t$			Idiosyncratic - $Group - \rho$			Aznar/Johansen		
50	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.14	0.12
100	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.12	0.09
250	0.02	0.01	0.00	0.00	0.00	0.00	0.11	0.10	0.09

Rejection frequencies are based on 5% asymptotic critical values.

Table 8: $k = 2$ common factors; cointegration in F_t but not in $E_{i,t}$ with MA errors.

N	25	50	100	25	50	100	25	50	100
				Raw data					
T	Kao-ρ^*			Kao-ADF			Pedroni $Panel - \rho$		
50	0.18	0.17	0.24	0.55	0.60	0.65	0.01	0.00	0.01
100	0.25	0.29	0.39	0.59	0.69	0.73	0.07	0.03	0.11
250	0.37	0.40	0.50	0.70	0.77	0.78	0.18	0.10	0.32
				Raw data					
T	Pedroni $Panel - t$			Pedroni $Group - \rho$			Pedroni $Group - t$		
50	0.05	0.03	0.06	0.00	0.00	0.00	0.03	0.02	0.05
100	0.09	0.06	0.19	0.05	0.01	0.10	0.07	0.03	0.15
250	0.17	0.11	0.33	0.25	0.13	0.57	0.19	0.10	0.42
				Estimated components					
T	Idiosyncratic $Panel - \rho$			Idiosyncratic $Group - \rho$			Aznar/Johansen		
50	0.00	0.00	0.00	0.00	0.00	0.00	0.61	0.61	0.74
100	0.00	0.00	0.00	0.00	0.00	0.00	0.70	0.78	0.85
250	0.02	0.00	0.00	0.00	0.00	0.00	0.68	0.85	0.87

$F_t^Y = \sum_{s=1}^t f_s^Y$, $F_t^X = F_t^Y + f_t^X$, $E_{i,t} = \sum_{s=1}^T e_{i,s}$,
where f_t and $e_{i,t}$ are MA processes generated as described in Section 5.
Rejection frequencies are based on 5% asymptotic critical values.

Table 9: $k = 2$ common factors; no cointegration in F_t but cointegration in $E_{i,t}$ with MA errors.

N	25	50	100	25	50	100	25	50	100
				Raw data					
T	Kao-ρ^*			Kao-ADF			Pedroni $Panel-\rho$		
50	0.26	0.28	0.33	0.57	0.60	0.64	0.05	0.13	0.10
100	0.36	0.43	0.46	0.62	0.72	0.72	0.16	0.39	0.32
250	0.46	0.52	0.55	0.71	0.77	0.78	0.31	0.53	0.48
				Raw data					
T	Pedroni $Panel-t$			Pedroni $Group-\rho$			Pedroni $Group-t$		
50	0.14	0.23	0.23	0.01	0.03	0.03	0.10	0.17	0.18
100	0.20	0.42	0.38	0.06	0.32	0.18	0.14	0.35	0.29
250	0.29	0.52	0.50	0.21	0.68	0.40	0.24	0.55	0.44
				Estimated components					
T	Idiosyncratic $Panel-\rho$			Idiosyncratic $Group-\rho$			Aznar/Johansen		
50	0.92	0.99	1.00	0.94	1.00	1.00	0.12	0.12	0.11
100	1.00	1.00	1.00	1.00	1.00	1.00	0.10	0.11	0.10
250	1.00	1.00	1.00	1.00	1.00	1.00	0.09	0.10	0.08

$F_t = \sum_{s=1}^{t} f_s$, $E_{i,t}^Y = \sum_{s=1}^{T} e_{i,s}^Y$, $E_{i,t}^X = E_{i,t}^Y + e_{i,s}^X$,
where f_t and $e_{i,t}$ are MA processes generated as described in Section 5.
Rejection frequencies are based on 5% asymptotic critical values.

Table 10: $k = 2$ common factors; cointegration in F_t and $E_{i,t}$ and MA errors.

N	25	50	100	25	50	100	25	50	100
				Raw data					
T	Kao-ρ^*			Kao-ADF			Pedroni $Panel - \rho$		
50	0.41	0.43	0.47	0.57	0.58	0.61	0.58	0.63	0.69
100	0.50	0.56	0.59	0.60	0.67	0.69	0.84	0.91	0.97
250	0.60	0.63	0.63	0.69	0.72	0.74	0.96	0.99	1.00
				Raw data					
T	Pedroni $Panel - t$			Pedroni $Group - \rho$			Pedroni $Group - t$		
50	0.69	0.77	0.84	0.74	0.81	0.91	0.86	0.94	0.98
100	0.87	0.94	0.98	1.00	1.00	1.00	1.00	1.00	1.00
250	0.95	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00
				Estimated components					
T	Idiosyncratic $Panel - \rho$			Idiosyncratic $Group - \rho$			Aznar/Johansen		
50	0.98	1.00	1.00	0.99	1.00	1.00	0.67	0.65	0.75
100	1.00	1.00	1.00	1.00	1.00	1.00	0.79	0.83	0.89
250	1.00	1.00	1.00	1.00	1.00	1.00	0.78	0.87	0.92

$F_t^Y = \sum_{s=1}^{t} f_s^Y$, $F_t^X = F_t^Y + f_t^X$, $E_{i,t}^Y = \sum_{s=1}^{T} e_{i,s}^Y$, $E_{i,t}^X = E_{i,t}^Y + e_{i,s}^X$,
where f_t and $e_{i,t}$ are MA processes generated as described in Section 5.
Rejection frequencies are based on 5% asymptotic critical values.

Table 11: $k = 3$ common factors; cross-member cointegration with MA(1) errors

N	25	50	100	25	50	100	25	50	100
Raw data									
T	Kao - ρ^*			Kao - ADF			Pedroni - $Panel - \rho$		
25	0.27	0.31	0.35	0.48	0.51	0.54	0.62	0.88	0.86
50	0.40	0.44	0.46	0.55	0.62	0.65	0.82	0.98	0.96
100	0.49	0.52	0.56	0.61	0.65	0.69	0.89	0.99	0.99
Raw data									
T	Pedroni - $Panel - t$			Pedroni - $Group - \rho$			Pedroni - $Group - t$		
25	0.69	0.92	0.90	0.27	0.67	0.60	0.47	0.81	0.88
50	0.82	0.98	0.97	0.65	0.97	0.95	0.64	0.92	0.88
100	0.89	0.99	0.99	0.84	1.00	1.00	0.71	0.96	0.95
Estimated components									
T	Idiosyncratic- $Panel - t$			Idiosyncratic - $Group - \rho$			Aznar/Johansen		
25	1.00	1.00	1.00	1.00	1.00	1.00	0.12	0.15	0.12
50	1.00	1.00	1.00	1.00	1.00	1.00	0.10	0.12	0.13
100	1.00	1.00	1.00	1.00	1.00	1.00	0.08	0.09	0.08

Rejection frequencies are based on 5% asymptotic critical values.

Table 12: $k = 3$ common factors; I(1) common factors and I(1) idiosyncratic components with MA(1) errors

N	25	50	100	25	50	100	25	50	100
				Raw data					
T	Kao - ρ^*			Kao - ADF			Pedroni - $Panel - \rho$		
50	0.15	0.18	0.22	0.56	0.64	0.72	0.01	0.00	0.00
100	0.24	0.27	0.32	0.67	0.74	0.79	0.04	0.03	0.03
250	0.36	0.39	0.44	0.70	0.77	0.81	0.13	0.12	0.13
				Raw data					
T	Pedroni - $Panel - t$			Pedroni - $Group - \rho$			Pedroni - $Group - t$		
50	0.04	0.03	0.03	0.00	0.00	0.00	0.04	0.03	0.04
100	0.08	0.07	0.08	0.01	0.00	0.00	0.04	0.03	0.04
250	0.14	0.15	0.17	0.00	0.00	0.00	0.04	0.03	0.04
				Estimated components					
T	Idiosyncratic- $Panel - t$			Idiosyncratic - $Group - \rho$			Aznar/Johansen		
50	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.14	0.13
100	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.13	0.14
250	0.02	0.01	0.00	0.00	0.00	0.00	0.9	0.10	0.08

Rejection frequencies are based on 5% asymptotic critical values.

Table 13: $k = 3$ common factors; cointegration in F_t but not in $E_{i,t}$ with MA errors.

N	25	50	100	25	50	100	25	50	100
Raw data									
T	Kao-ρ^*			Kao-ADF			Pedroni $Panel-\rho$		
50	0.07	0.08	0.11	0.20	0.26	0.38	0.33	0.26	0.19
100	0.13	0.19	0.22	0.27	0.37	0.53	0.71	0.69	0.60
250	0.27	0.30	0.34	0.39	0.44	0.60	0.80	0.73	0.73
Raw data									
T	Pedroni $Panel-t$			Pedroni $Group-\rho$			Pedroni $Group-t$		
50	0.49	0.44	0.40	0.10	0.06	0.04	0.36	0.28	0.24
100	0.70	0.71	0.65	0.69	0.60	0.40	0.65	0.31	0.48
250	0.77	0.76	0.67	0.91	0.89	0.84	0.80	0.76	0.71
Estimated components									
T	Idiosyncratic $Panel-\rho$			Idiosyncratic $Group-\rho$			Aznar/Johansen		
50	0.00	0.00	0.00	0.00	0.00	0.00	0.60	0.66	0.67
100	0.00	0.00	0.00	0.00	0.00	0.00	0.79	0.79	0.84
250	0.02	0.01	0.01	0.00	0.00	0.00	0.78	0.80	0.90

$F_t^X = \sum_{s=1}^t f_s^X$, $F_t^X = (F_{1t}^X + F_{2t}^X) + f_t^Y$, $E_{i,t} = \sum_{s=1}^T e_{i,s}$,
where f_t is a (3×1) and $e_{i,t}$ is a (2×1) MA process generated as described in Section 5, and $F_t = (F_t^Y, F_t^{X\prime})'$ with $F_t^X = (F_{1t}^X, F_{2t}^X)'$.
Rejection frequencies are based on 5% asymptotic critical values.

Table 14: $k = 3$ common factors; no cointegration in F_t but cointegration in $E_{i,t}$ with MA errors.

N	25	50	100	25	50	100	25	50	100
Raw data									
T	Kao-ρ^*			Kao-ADF			Pedroni $Panel-\rho$		
50	0.21	0.23	0.26	0.54	0.63	0.70	0.03	0.03	0.02
100	0.30	0.34	0.40	0.65	0.72	0.77	0.12	0.13	0.09
250	0.42	0.46	0.49	0.70	0.77	0.80	0.27	0.29	0.27
Raw data									
T	Pedroni $Panel-t$			Pedroni $Group-\rho$			Pedroni $Group-t$		
50	0.09	0.09	0.08	0.01	0.00	0.00	0.07	0.08	0.07
100	0.19	0.19	0.16	0.04	0.04	0.03	0.13	0.12	0.10
250	0.26	0.30	0.30	0.18	0.30	0.35	0.21	0.28	0.26
Estimated components									
T	Idiosyncratic $Panel-\rho$			Idiosyncratic $Group-\rho$			Aznar/Johansen		
50	0.37	0.33	0.37	0.17	0.15	0.24	0.11	0.15	0.13
100	0.75	0.69	0.80	0.85	0.92	1.00	0.10	0.12	0.13
250	0.83	0.84	0.95	0.97	1.00	1.00	0.08	0.09	0.08

$F_t = \sum_{s=1}^{t} f_s$, $E_{i,t}^Y = \sum_{s=1}^{T} e_{i,s}^Y$, $E_{i,t}^X = E_{i,t}^Y + e_{i,s}^X$,
where f_t and $e_{i,t}$ are MA processes generated as described in Section 5.
Rejection frequencies are based on 5% asymptotic critical values.

Table 15: $k = 3$ common factors; cointegration in F_t and $E_{i,t}$ and MA errors.

N	25	50	100	25	50	100	25	50	100
Raw data									
T	Kao-ρ^*			Kao-ADF			Pedroni $Panel - \rho$		
50	0.13	0.17	0.19	0.24	0.29	0.41	0.85	0.83	0.78
100	0.26	0.32	0.34	0.32	0.40	0.52	0.99	0.99	1.00
250	0.40	0.41	0.46	0.43	0.46	0.58	1.00	1.00	1.00
Raw data									
T	Pedroni $Panel - t$			Pedroni $Group - \rho$			Pedroni $Group - t$		
50	0.92	0.92	0.91	0.71	0.67	0.56	0.90	0.89	0.87
100	0.99	0.99	1.00	1.00	1.00	1.00	0.99	0.99	0.99
250	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00
Estimated components									
T	Idiosyncratic $Panel - \rho$			Idiosyncratic $Group - \rho$			Aznar/Johansen		
50	0.40	0.39	0.77	0.23	0.24	0.36	0.65	0.68	0.67
100	0.75	0.70	0.83	0.88	0.95	1.00	0.84	0.83	0.85
250	0.84	0.85	0.95	0.98	1.00	1.00	0.86	0.85	0.90

$F_t^X = \sum_{s=1}^t f_s^X$, $F_t^X = (F_{1t}^X + F_{2t}^X) + f_t^Y$, $E_{i,t}^Y = \sum_{s=1}^T e_{i,s}^Y$, $E_{i,t}^X = E_{i,t}^Y + e_{i,s}^X$,
where f_t is a (3×1) and $e_{i,t}$ is a (2×1) MA process generated as described in Section 5, and $F_t = (F_t^Y, F_t^{X\prime})'$ with $F_t^X = (F_{1t}^X, F_{2t}^X)'$.
Rejection frequencies are based on 5% asymptotic critical values.

Table 16: Panel no-cointegration tests for observed data

	Kao - ρ^*	Kao - ADF	Pedroni - $Panel - \rho$
Statistic	-4.36	-3.58	-2.16
p-value	0.00	0.00	0.02
	Pedroni - $Panel - t$	Pedroni - $Group - \rho$	Pedroni - $Group - t$
Statistic	-1.57	-0.18	-0.50
p-value	0.06	0.43	0.31

Table 17: Panel no-cointegration tests for defactored data

	Statistic	p-value
$Panel - \rho$	-1.58	0.57
$Panel - t$	-1.31	0.10
$Group - \rho$	0.14	0.56
$Group - t$	-0.05	0.48
	Statistic	Critical value
Johansen	8.19	12.53