PANICCA – PANIC ON CROSS-SECTION AVERAGES*

Simon Reese
Lund University

Joakim Westerlund†
Lund University
and
Financial Econometrics Group
Centre for Research in Economics and Financial Econometrics
Deakin University

March 19, 2015

Abstract

The cross-section average (CA) augmentation approach of Pesaran (2007) and Pesaran et al. (2013), and the principal components-based panel analysis of non-stationarity in idiosyncratic and common components (PANIC) of Bai and Ng (2004, 2010) are among the most popular "second-generation" approaches for cross-section correlated panels. One feature of these approaches is that they have different strengths and weaknesses. The purpose of the current paper is to develop PANICCA, a combined approach that exploits the strengths of both CA and PANIC.

JEL Classification: C12; C13; C33; C36.

Keywords: PANIC; cross-section average augmentation; unit root test; cross-section dependence; common factors.

1 Introduction

Consider the panel data variable $Y_{i,t}$, observable for $t = 1, ..., T$ time periods and $i = 1, ..., N$ cross-section units. It is well known that unattended cross-section dependence can lead to

*A previous version of this paper was presented at the 6-th Italian Congress of Econometrics and Empirical Economics in Salerno, and at a seminar at Lund University. The authors would like to thank conference and seminar participants for many valuable comments and suggestions. Westerlund thanks the Knut and Alice Wallenberg Foundation for financial support through a Wallenberg Academy Fellowship. Both authors thank the Jan Wallander and Tom Hedelius Foundation for financial support under research grant number P2014–0112:1.

†Corresponding author: Department of Economics, Lund University, Box 7082, 220 07 Lund, Sweden. Telephone: +46 46 222 8997. Fax: +46 46 222 4613. E-mail address: joakim.westerlund@nek.lu.se.

deceptive inference when testing the null hypothesis of a unit root in such variables. This is certainly true for panel unit root tests devised to test the hypothesis that $Y_{1,t}, ..., Y_{N,t}$ are jointly unit root non-stationary, but the problem is there also when applying univariate unit root tests to each cross-section unit. This finding has led to the development of factor-based "second-generation" test procedures that are robust to cross-section dependence (see Breitung and Pesaran, 2008; Baltagi, 2013, Chapter 12, for surveys of the literature). Two of the most popular second-generation tests are the cross-section augmented Im–Pesaran–Shin (CIPS) and Sargan–Bhargava (CSB) tests of Pesaran (2007) and Pesaran et al. (2013). In fact, these tests have in a short period of time become two of the industry's workhorses, with a large number of applications and also several theoretical extensions (see, for example, Westerlund, 2015a; Westerlund et al., 2015).

As the name suggests, the idea underlying the cross-section average (CA) augmentation approach, originally put forth by Pesaran (2006) in the context of factor-augmented panel regressions, is to use the cross-section average $\overline{Y}_t$ of $Y_{i,t}$ as a proxy for the common component of the data, which is then included in the regression as additional regressors. But if $Y_{i,t}$ is unit root non-stationary then so is $\overline{Y}_t$, suggesting that, in analogy to the spurious regression phenomenon, the asymptotic distributions of the resulting CIPS and CSB statistics will depend on the Brownian motion generated by $\overline{Y}_t$. They will therefore be highly nonstandard, which in turn makes for complicated implementation. In particular, not only is it necessary to tabulate critical values for each constellation of (N, T), but there is also a need to truncate the test statistics in order to ensure finite moments. As Pesaran et al. (2013) show, however, when properly implemented, the CIPS and CSB tests do seem to enjoy relatively good small-sample performance, which is partly expected given the relatively good performance of the CA components estimator (see, for example, Chudik et al., 2011; Kapetanios and Pesaran, 2005; Westerlund and Urbain, 2015). Another feature of the CIPS and CSB statistics is that they assume that the common and idiosyncratic components of the data have the same order of integration, which of course need not be the case in practice.

An alternative test approach that supports asymptotically normal inference is the panel analysis of non-stationarity in idiosyncratic and common components (PANIC) of Bai and Ng (2004, 2010). This approach, which, in contrast to CA augmentation, does not require the common and idiosyncratic components to be integrated of the same order, is arguably the most popular approach in the literature with even more applications and extensions than CA (see, for example, Bai and Carrion-i-Silvestre, 2009, 2013; Gengenbach et al., 2006; Westerlund, 2014; Westerlund and Hess, 2011; Westerlund and Larsson, 2012). The basic idea in PANIC is to first transform $Y_{i,t}$ by taking first-differences. The method of principal components (PC) is then applied to estimate the first-differenced common and idiosyncratic components, which can be cumulated up to levels. The fact that the components are estimated from a regression in first-differences means that the spurious regression problem is avoided, thereby enabling standard normal inference. As the bulk of the existing Monte Carlo evidence show (see, for example, Gengenbach et al., 2006, 2010; Pesaran et al., 2013; Westerlund and Larsson, 2009; Westerlund and Urbain, 2015), however, the use of PC can render PANIC small-sample distorted, especially when N is "small".

The purpose of the present paper is to propose a test procedure that is both general and simple, yet with good small-sample performance. In view of the above discussion, a natural suggestion towards this end is to use PANIC, but to apply it to the estimated CA components rather than to the estimated PC components. As far as we are aware this is the first attempt to exploit the advantages of both CA and PANIC. The properties of the resulting PANICCA procedure is studied under the condition that the number of panel data variables is at least as large as the number of common factors. Our key findings can be summarized as follows. First, PANICCA inherits the generality of PANIC and enables inference regarding the unit root and cointegration properties of both the common and idiosyncratic components of the data. PANICCA can therefore be seen as a complete panel unit root toolbox. Second, being based on simple CA, PANICCA is very user friendly. In fact, in view of its generality, it is surprisingly simple, requiring nothing but basic averaging and least squares (LS) operations. To facilitate easy implementation, a full suit of GAUSS codes can be downloaded freely from http://sites.google.com/site/perjoakimwesterlund/. Third, PANICCA leads to the same asymptotic theory as PANIC. Appropriate critical values can therefore be taken directly from Bai and Ng (2004, 2010). Fourth, the use of CA rather than PC leads to much improved small-sample performance, especially in the type of small- to medium-N panels often encountered in applied work (see Lanzafame, 2010; Schmidt and Vosen, 2013; Martín, 2013; Joseph et al., 2012, 2013; Örsal and Dilan, 2014; Blomquist and Westerlund, 2014, for a non-exhaustive list).

In our empirical application we consider an old empirical puzzle within financial economics, namely, the failure of the efficient market hypothesis (EMH). Here we demonstrate the usefulness of the generality of PANICCA, as a platform for testing for cointegration both within and between cross-section units. According to EMH, not only should the current forward rate be cointegrated with the future spot rate, but there should also not be any cointegration running across currencies. Interestingly, while separately these cointegrating restrictions have been subject to countless tests (see, for example, Hakkio and Rush, 1989; Baillie and Bollerslev, 1989; Crowder, 1994, for early contributions), as far as we are aware, the current paper is the first to consider a joint test of both restrictions.

The balance of the paper is organized as follows. In Section 2, we lay out the assumptions that we will be working under and explain how these compare to the assumptions of PANIC. Section 3 provides an account of the PANICCA procedure and its asymptotic properties, whose accuracy in small samples is studied by means of Monte Carlo simulation in Section 4. Section 5 contains the empirical application and Section 6 concludes.

2 Model and assumptions

Consider the scalar panel data variable $Y_{i,t}$, observable for $i = 1, ..., N$ cross-section units and $t = 1, ..., T$ time periods. The data generating process (DGP) of this variable is assumed to be given by the following common factor model:

$$Y_{i,t} = \boldsymbol{\alpha}_i'\mathbf{D}_{t,p} + \boldsymbol{\lambda}_i'\mathbf{F}_t + e_{i,t}, \tag{1}$$

where $e_{i,t}$ is a scalar idiosyncratic error, $\mathbf{F}_t$ is an $r \times 1$ vector of common factors with $\boldsymbol{\lambda}_i$ being the associated $(r \times 1)$ vector of loading coefficients, and $\mathbf{D}_{t,p} = (1, ..., t^p)'$ is a $(p + 1) \times 1$ vector of trends for which we consider two specifications; (i) a constant ($p = 0$), and (ii) a constant and trend ($p = 1$). In this paper, $Y_{i,t}$ is considered as the variable of interest. However, we do allow for the presence of an $m \times 1$ vector of additional variables, henceforth denoted $X_{i,t}$, whose data generating process is given by

$$\mathbf{X}_{i,t} = \boldsymbol{\beta}_i'\mathbf{D}_{t,p} + \boldsymbol{\Lambda}_i'\mathbf{F}_t + \mathbf{u}_{i,t}, \tag{2}$$

where $\mathbf{u}_{i,t}$ is a $m \times 1$ vector of idiosyncratic errors. Thus, as in Pesaran et al. (2013), we assume the existence of an additional m variables that are permitted (but not required; see Remark 1) to share the common factors of the variable of interest. This seems very plausible, especially in macroeconomics and finance, where most variables are highly co-moving (see Section 5 for an empirical illustration).

Define $\mathbf{Z}_{i,t} = (Y_{i,t}, \mathbf{X}'_{i,t})'$. In view of (1) and (2), the DGP of this variable is easily seen to be given by

$$\mathbf{Z}_{i,t} = \mathbf{B}'_i \mathbf{D}_{t,p} + \mathbf{C}'_i \mathbf{F}_t + \mathbf{V}_{i,t}, \tag{3}$$

where $\mathbf{B}_i = (\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i)$, $\mathbf{C}_i = (\boldsymbol{\lambda}_i, \boldsymbol{\Lambda}_i)$ and $\mathbf{V}_{i,t} = (e_{i,t}, \mathbf{u}'_{i,t})'$. Note that the dimension of $\mathbf{C}_i$ is $r \times (m+1)$. The conditions under which we will be working are summarized below. Here and throughout this paper $tr(\mathbf{A})$, $rk(\mathbf{A})$ and $||\mathbf{A}|| = \sqrt{tr(\mathbf{A}'\mathbf{A})}$ denote the trace, rank and Frobenius (Euclidean) norm, respectively, of the matrix $\mathbf{A}$, $\overline{\mathbf{A}} = N^{-1}\sum_{i=1}^{N} \mathbf{A}_i$ for any $\mathbf{A}_i$, and $M < \infty$ is a generic positive number.

Assumption 1. $(1 - \rho_i L)e_{i,t} = \phi_i(L)\epsilon_{i,t}$, where $\phi_i(L) = \sum_{n=0}^{\infty} \phi_{i,n} L^n$ with $\sum_{n=0}^{\infty} n|\phi_{i,n}| \leq M$ and $\phi_i(1) > 0$, and $\epsilon_{i,t}$ is independently and identically distributed (iid) across both i and t with $E(\epsilon_{i,t}) = 0$, $E(\epsilon_{i,t}^2) = 1$ and $E(|\epsilon_{i,t}|^8) \leq M$.

Assumption 2. $\Delta \mathbf{u}_{i,t} = \boldsymbol{\Psi}_i(L)\boldsymbol{\varepsilon}_{i,t}$, where $\boldsymbol{\Psi}_i(L) = \sum_{n=0}^{\infty} \boldsymbol{\Psi}_{i,n} L^n$ with $\sum_{n=0}^{\infty} n||\boldsymbol{\Psi}_{i,n}|| \leq M$ and $rk[\boldsymbol{\Psi}_i(1)] = m_1 \in [0, m]$, $\mathrm{var}(\Delta \mathbf{u}_{i,t}) = \sum_{n=0}^{\infty} \boldsymbol{\Psi}_n \boldsymbol{\Psi}'_n$ is positive definite, and $\boldsymbol{\varepsilon}_{i,t}$ is iid across both i and t with $E(\boldsymbol{\varepsilon}_{i,t}) = 0_{m\times 1}$, $E(\boldsymbol{\varepsilon}_{i,t}\boldsymbol{\varepsilon}'_{i,t}) = I_m$ and $E(||\boldsymbol{\varepsilon}_{i,t}||^8) \leq M$.

Assumption 3. $\Delta \mathbf{F}_t = \boldsymbol{\Phi}(L)\boldsymbol{\eta}_t$, where $\boldsymbol{\Phi}(L) = \sum_{n=0}^{\infty} \boldsymbol{\Phi}_n L^n$ with $\sum_{n=0}^{\infty} n||\boldsymbol{\Phi}_n|| \leq M$ and $rk[\boldsymbol{\Phi}(1)] = r_1 \in [0, r]$, $\mathrm{var}(\Delta \mathbf{F}_t) = \sum_{n=0}^{\infty} \boldsymbol{\Phi}_n \boldsymbol{\Sigma}_\eta \boldsymbol{\Phi}'_n$ is positive definite, and $\boldsymbol{\eta}_t$ is iid across t with $E(\boldsymbol{\eta}_t) = 0_{(m+1)\times 1}$, $E(\boldsymbol{\eta}_t\boldsymbol{\eta}'_t) = \boldsymbol{\Sigma}_\eta$ and $E(||\boldsymbol{\eta}_t||^4) \leq M$.

Assumption 4. $\mathbf{C}_i$ is a nonrandom vector satisfying $||\mathbf{C}_i|| \leq M$ and $rk(\overline{\mathbf{C}}) = r \leq m+1$ for any N, including $N \to \infty$.

Assumption 5. $\epsilon_{i,t}$, $\boldsymbol{\varepsilon}_{i,t}$ and $\boldsymbol{\eta}_t$ are mutually independent.

Assumption 6. $E(||\mathbf{F}_0||) \leq M$ and $E(|e_{i,0}|) \leq M$ for all i.

Denote by $\sigma_{\epsilon,i}^2 = \sum_{n=0}^{\infty} \phi_{i,n}^2$, $\omega_{\epsilon,i}^2 = \phi_i(1)^2$ and $\tau_{\epsilon,i} = (\omega_{\epsilon,i}^2 - \sigma_{\epsilon,i}^2)/2$ the contemporaneous, long-run and one-sided long-run variance of $\epsilon_{i,t}$, respectively. Let us further denote by $\overline{\sigma}_\epsilon^2$, $\overline{\omega}_\epsilon^2$ $\overline{\phi}_\epsilon^4$ and $\overline{\lambda}_\epsilon$ the cross-sectional averages of $\sigma_{\epsilon,i}^2$, $\omega_{\epsilon,i}^2$, $\phi_{\epsilon,i}^4$ and $\tau_{\epsilon,i}$, respectively, where $\phi_{\epsilon,i}^4 = \omega_{\epsilon,i}^4$ (to avoid confusion between $(\overline{\omega}_\epsilon^2)^2$ and $N^{-1}\sum_{i=1}^{N} \omega_{\epsilon,i}^4$).

Assumption 7. $\overline{\sigma}_\epsilon^2 \to \sigma_\epsilon^2$, $\overline{\omega}_\epsilon^2 \to \omega_\epsilon^2$ $\overline{\phi}_\epsilon^4 \to \phi_\epsilon^4$ and $\overline{\lambda}_\epsilon \to \lambda_\epsilon$ as $N \to \infty$, where σ_ϵ^2, ω_ϵ^2, $\phi_\epsilon^4 \in (0, M)$ and $|\lambda_\epsilon| \leq M$.

The above conditions are very similar to those employed by Bai and Ng (2004, 2010), and we therefore refer to these previous works for a detailed discussion. The main differences are; (i) the assumed presence of the $m \times 1$ vector $\mathbf{X}_{i,t}$, (ii) the requirement that $rk(\overline{\mathbf{C}}) = r \leq m+1$, (iii) the requirement that $\epsilon_{i,t}$ and $\boldsymbol{\varepsilon}_{i,t}$ are iid across i, and (iv) the assumed nonrandomness of $\mathbf{C}_i$. Assumptions (i)–(iii) ensure that $\mathbf{F}_t$ can be estimated using nothing but the simple cross-section average of $\mathbf{Z}_{i,t}$. The PC equivalent of (ii) is that $rk(N^{-1}\sum_{i=1}^{N}\mathbf{C}_i\mathbf{C}_i') = r \leq k$, where k is the assumed number of common factors, which can be larger or smaller than $m+1$. Hence, the usual problem in PC analysis of finding a suitable upper bound on the true number of factors, r, is in CA tantamount to finding an appropriate number of extra variables. The additional observations required in CA is the "price" paid for the relative simplicity with which the factors are estimated. Of course, in many situations, the model of ultimate interest is a multivariate one, and the unit root testing is just a pre-test step. In situations like this, joint CA estimation of the factors of all the variables of the model is expected to lead to reduced estimation uncertainty when compared to variable-by-variable PC, as in original PANIC (see Westerlund and Urbain, 2015). As pointed out in Westerlund and Urbain (2015), the requirement that $rk(\overline{\mathbf{C}}) = r$ is not testable. It can be relaxed, but then at the cost of additional restrictions on $\mathbf{C}_i$. Indeed, as Westerlund and Urbain (2013) show, if Assumption 4 is violated, then $\boldsymbol{\lambda}_i$ and $\boldsymbol{\Lambda}_i$ have to be random and uncorrelated.

As in Bai and Ng (2010), (iii) is not really necessary and can be relaxed to allow for weak cross-section correlation in the "idiosyncratic" component (see Bai and Ng, 2004). In the terminology of Chudik et al. (2011), $\epsilon_{i,t}$ and $\boldsymbol{\varepsilon}_{i,t}$ may be "semi-weakly" correlated without affecting the results derived in Appendix. The intuition is simple. Suppose for sake of argument that $\phi_i(L) = 1$, $\boldsymbol{\Psi}_i(L) = \mathbf{I}_m$ and $\rho_1 = ... = \rho_N = 1$, implying $\mathbf{v}_{i,t} = \Delta\mathbf{V}_{i,t} = (\epsilon_{i,t}, \boldsymbol{\varepsilon}_{i,t}')'$. A key requirement for the consistency of the estimated factors is that $||\overline{\mathbf{v}}_t|| = O_p(N^{-1/2})$ (see Remark 2 of Section 3), which will be the case if $\epsilon_{i,t}$ and $\boldsymbol{\varepsilon}_{i,t}$ are iid. However, while sufficient, iidness is clearly not a necessary condition. Suppose for example that $\mathbf{v}_{i,t} = \mathbf{A}_i'\Delta\mathbf{F}_t + \boldsymbol{\xi}_{i,t}$, where $\mathbf{A}_i = N^{-\alpha}\mathbf{C}_i$ and $\boldsymbol{\xi}_{i,t}$ is iid across both i and t with mean zero and four finite moments. If $\alpha \in [1/2, 1)$, such that $\mathbf{v}_{i,t}$ is semi-weakly cross-correlated, then $||\overline{\mathbf{v}}_t|| \leq N^{-\alpha}||\overline{\mathbf{C}}|| \cdot ||\Delta\mathbf{F}_t|| + ||\overline{\boldsymbol{\xi}}_t|| = O_p(\max\{N^{-\alpha}, N^{-1/2}\}) = O_p(N^{-1/2})$.

As with (iii), assumption (iv) is only for simplicity, and can be relaxed, provided that $\mathbf{C}_i$ is independent of all the other random elements of the DGP and $E(||\mathbf{C}_i||^4) \leq M$ (see Bai and Ng, 2004, 2010). Alternatively, we may assume that $\mathbf{C}_i$ satisfies Assumption 3 of Pesaran (2004).

Remark 1. The assumption that $Y_{i,t}$ and $\mathbf{X}_{i,t}$ depend on the same set of factors is not a restriction. Suppose, for example, that the factors to $Y_{i,t}$ and $\mathbf{X}_{i,t}$ do not have any elements in common. In order to capture this we introduce the $r \times r$ orthogonal matrix $\mathbf{J} = (\mathbf{J}_1, \mathbf{J}_2)$, which is such that $\mathbf{J}'\mathbf{J} = \mathbf{J}\mathbf{J}' = \mathbf{I}_r$. The component matrices $\mathbf{J}_1$ and $\mathbf{J}_2$, which are of dimension $r \times r_1$ and $r \times (r - r_1)$, respectively, are such that $\mathbf{J}_2'\mathbf{J}_1 = \mathbf{0}_{(r-r_1)\times r_1}$, $\mathbf{J}_1'\boldsymbol{\Lambda}_i = \mathbf{0}_{r_1 \times (m+1)}$ and $\mathbf{J}_2'\boldsymbol{\lambda}_i = \mathbf{0}_{(r-r_1)\times 1}$. The matrix $\mathbf{J}$ allows us to rotate $\mathbf{F}_t$ as $\mathbf{J}'\mathbf{F}_t = (\mathbf{J}_1'\mathbf{F}_t, \mathbf{J}_2'\mathbf{F}_t) = (\mathbf{F}_{1,t}', \mathbf{F}_{2,t}')'$. Thus, defining $\mathbf{J}_1'\boldsymbol{\lambda}_i = \boldsymbol{\lambda}_{1,i}$ and $\mathbf{J}_2'\boldsymbol{\Lambda}_i = \boldsymbol{\Lambda}_{2,i}$, we have $Y_{i,t} = \boldsymbol{\alpha}_i'\mathbf{D}_{t,p} + \boldsymbol{\lambda}_i'\mathbf{F}_t + e_{i,t} = \boldsymbol{\alpha}_i'\mathbf{D}_{t,p} + \boldsymbol{\lambda}_i'\mathbf{J}\mathbf{J}'\mathbf{F}_t + e_{i,t} = \boldsymbol{\alpha}_i'\mathbf{D}_{t,p}\boldsymbol{\lambda}_{1,i}'\mathbf{F}_{1t} + e_i$ and similarly $\mathbf{X}_{i,t} = \boldsymbol{\beta}_i'\mathbf{D}_{t,p} + \boldsymbol{\Lambda}_{2,i}'\mathbf{F}_{2,t} + \mathbf{u}_{i,t}$.

3 PANICCA

The idea behind PANIC is to first transform $\mathbf{Z}_{i,t}$ by taking first differences. Since the transformed variable is stationary by assumption, the uncertainty regarding the order of integration of $\mathbf{Z}_{i,t}$ is gone, which means that the common and idiosyncratic components can be estimated using existing methods for common factor models. While Bai and Ng (2004, 2010) use PC, in the present paper we use CA. The estimated level components are obtained by simply taking partial sums of the estimated first-differenced components. The unit root and cointegration properties of these components can then be tested using the existing battery of tests.

The purpose of the rest of this section is to make the above discussion a little more precise. Let us begin by defining $\mathbf{z}_{i,t} = \Delta\mathbf{Z}_{i,t}$, $\mathbf{d}_{t,p} = \Delta\mathbf{D}_{t,p}$, $\mathbf{f}_t = \Delta\mathbf{F}_t$ and $\mathbf{v}_{i,t} = \Delta\mathbf{V}_{i,t}$. Denote by $\mathbf{G}$ the $p \times (p+1)$ selection matrix of zeroes and ones removing the first element of $\mathbf{d}_{t,p}$, which is zero, that is, $\mathbf{G}\mathbf{d}_{t,p} = \mathbf{D}_{t,p-1}$. Since $rk(\mathbf{G}'\mathbf{G}) = p$, we may further define the $p \times (m+1)$ matrix $\mathbf{b}_i = \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{B}_i$. In this notation, the first-differenced version of (3) may be written as

$$\mathbf{z}_{i,t} = \mathbf{b}_i'\mathbf{D}_{t,p-1} + \mathbf{C}_i'\mathbf{f}_t + \mathbf{v}_{i,t}, \tag{4}$$

or, in matrix form,

$$\mathbf{z}_i = \mathbf{D}_{p-1}\mathbf{b}_i + \mathbf{f}\mathbf{C}_i + \mathbf{v}_i, \tag{5}$$

where $\mathbf{z}_i = (\mathbf{z}_{i,2}, ..., \mathbf{z}_{i,T})'$ and $\mathbf{v}_i = (\mathbf{v}_{i,2}, ..., \mathbf{v}_{i,T})'$ are $(T-1) \times (m+1)$, $\mathbf{f} = (\mathbf{f}_2, ..., \mathbf{f}_T)'$ is $(T-1) \times r$, and $\mathbf{D}_{p-1} = (\mathbf{D}_{2,p-1}, ..., \mathbf{D}_{T,p-1})'$ is $(T-1) \times (p-1)$. Since $\mathbf{f}$ and $\mathbf{C}_i$ are not separately identifiable, the best that we can do is to estimate the space spanned by these matrices. Define $\mathbf{M}_p = \mathbf{I}_{T-1} - \mathbf{D}_{p-1}(\mathbf{D}_{p-1}'\mathbf{D}_{p-1})^{-1}\mathbf{D}_{p-1}'$ for $p = 1$ and $\mathbf{M}_p = \mathbf{I}_{T-1}$ for $p = 0$. Let

$\mathbf{z}_i^p = (\mathbf{z}_{i,2}^p, ..., \mathbf{z}_{i,T}^p)' = \mathbf{M}_p \mathbf{z}_i$ with similar definitions of $\mathbf{f}^p$ and $\mathbf{v}_i^p$. In this notation, (5) can be written alternatively as

$$\mathbf{z}_i^p = \mathbf{f}^p \mathbf{C}_i + \mathbf{v}_i^p. \tag{6}$$

The CA estimator of (the space spanned by) $\mathbf{f}^p$ is given by $\hat{\mathbf{f}}^p = \mathbf{M}_p \overline{\mathbf{z}} = \overline{\mathbf{z}}^p = N^{-1} \sum_{i=1}^N \mathbf{M}_p z_i$, while $\hat{\mathbf{C}}_i = [(\hat{\mathbf{f}}^p)' \hat{\mathbf{f}}^p]^{-1} (\hat{\mathbf{f}}^p)' \mathbf{z}_i^p$ is the LS estimator of $\mathbf{C}_i$ in (6) with $\mathbf{f}^p$ replaced by $\hat{\mathbf{f}}^p$. The estimator of $\mathbf{v}_i^p$ is given by $\hat{\mathbf{v}}_i^p = \mathbf{z}_i^p - \hat{\mathbf{f}}^p \hat{\mathbf{C}}_i$. Note that $\hat{\mathbf{f}}_t^p$ ($\hat{\mathbf{v}}_{i,t}^p$) is an estimator of the first-differenced (and detrended) version of $\mathbf{F}_t$ ($\mathbf{V}_{i,t}$). As an estimator of (the detrended version of) $\mathbf{F}_t$ ($\mathbf{V}_{i,t}$) we use $\hat{\mathbf{F}}_t^p = \sum_{n=2}^t \hat{\mathbf{f}}_n^p$ ($\hat{\mathbf{V}}_{i,t}^p = \sum_{n=2}^t \hat{\mathbf{v}}_{i,n}^p$).

Remark 2. A conceptual difference when compared to Pesaran (2006, 2007), and Pesaran et al. (2013) is that while in these other works the averages are referred to as "factor proxies", in the present study $\hat{\mathbf{f}}^p$ is treated as an estimator for the space spanned by $\mathbf{f}_t^p$. The reason is simple. We begin by noting that $\overline{\mathbf{z}}_t^p = \overline{\mathbf{C}}' \mathbf{f}_t^p + \overline{\mathbf{v}}_t^p$. This implies

$$\hat{\mathbf{f}}_t^p = \overline{\mathbf{z}}_t^p = \overline{\mathbf{C}}' \mathbf{f}_t^p + \overline{\mathbf{v}}_t^p = \overline{\mathbf{C}}' \mathbf{f}_t^p + O_p(N^{-1/2}),$$

where the order of the remainder follows from the fact that $\mathbf{v}_{i,t}^p$ is mean zero and independent across i. Hence, $\hat{\mathbf{f}}_t^p$ is consistent, but not for $\mathbf{f}_t^p$; only for $\overline{\mathbf{C}}' \mathbf{f}_t^p$, which is enough for our purposes. The rotation by $\overline{\mathbf{C}}$ here illustrates the need for the rank condition in Assumption 4. Suppose, for example, that $r = 1$ but $\overline{\mathbf{C}} = \mathbf{0}_{1 \times (m+1)}$. In this case there is a single common factor present. However, since $\overline{\mathbf{C}}' \mathbf{f}_t^p = 0$, $\hat{\mathbf{f}}_t^p$ will be unable to capture it.

3.1 Testing $e_{i,t}$

Note that $\hat{\mathbf{C}}_i$ can be decomposed as $\hat{\mathbf{C}}_i = (\hat{\boldsymbol{\lambda}}_i, \hat{\boldsymbol{\Lambda}}_i)$, where $\hat{\boldsymbol{\lambda}}_i$ is $r \times 1$ and $\hat{\boldsymbol{\Lambda}}_i$ is $r \times m$. We also have $\hat{\mathbf{V}}_{i,t}^p = [\hat{e}_{i,t}^p, (\hat{\mathbf{u}}_{i,t}^p)']'$, where $\hat{e}_{i,t}^p$ is a scalar and $\hat{\mathbf{u}}_{i,t}^p$ is $m \times 1$. Denote by $\hat{\rho}_p$ the least squares slope estimator in a pooled panel regression of $\hat{e}_{i,t}^p$ onto $\hat{e}_{i,t-1}^p$. The test statistics considered herein are all taken from Bai and Ng (2010), and are designed to test the null hypothesis that

$\rho_1 = ... = \rho_N = 1$. The first two test statistics, denoted $P_{a,p}$ and $P_{b,p}$, are defined as follows:

$$
\begin{aligned}
P_{a,0} &= \frac{\sqrt{N}T(\hat{\rho}_0^+ - 1)}{\sqrt{2\hat{\phi}_\epsilon^4/\hat{\omega}_\epsilon^4}}, \\
P_{b,0} &= \frac{\sqrt{N}T(\hat{\rho}_0^+ - 1)}{\sqrt{\hat{\phi}_\epsilon^4/[\hat{\omega}_\epsilon^2 N^{-1}T^{-2}\sum_{i=1}^N(\hat{e}_{i,-1}^0)'\hat{e}_{i,-1}^0]}}, \\
P_{a,1} &= \frac{\sqrt{N}T(\hat{\rho}_1^+ - 1)}{\sqrt{36\hat{\sigma}_\epsilon^4\hat{\phi}_\epsilon^4/5\hat{\omega}_\epsilon^8}}, \\
P_{b,1} &= \frac{\sqrt{N}T(\hat{\rho}_1^+ - 1)}{\sqrt{6\hat{\phi}_\epsilon^4\hat{\sigma}_\epsilon^4/[5\hat{\omega}_\epsilon^6 N^{-1}T^{-2}\sum_{i=1}^N(\hat{e}_{i,-1}^1)'\hat{e}_{i,-1}^1]}},
\end{aligned}
$$

where $\hat{e}_{i,-1}^p = (\hat{e}_{i,2}^p, ..., \hat{e}_{i,T-1}^p)'$ and

$$
\begin{aligned}
\hat{\rho}_0^+ &= \hat{\rho}_0 + \frac{\hat{\tau}_\epsilon}{(NT)^{-1}\sum_{i=1}^N(\hat{e}_{i,-1}^0)'\hat{e}_{i,-1}^0}, \\
\hat{\rho}_1^+ &= \hat{\rho}_1 + \frac{3\hat{\sigma}_\epsilon^2}{T\hat{\omega}_\epsilon^2}.
\end{aligned}
$$

Here $\hat{\sigma}_\epsilon^2$, $\hat{\omega}_\epsilon^2$ $\hat{\phi}_\epsilon^4$ and $\hat{\tau}_\epsilon$ are given by the cross-sectional averages of $\hat{\sigma}_{\epsilon,i}^2$, $\hat{\omega}_{\epsilon,i}^2$ $\hat{\phi}_{\epsilon,i}^4$ and $\hat{\tau}_{\epsilon,i}$, respectively. The first two of these estimated variances are constructed as follows:

$$
\begin{aligned}
\hat{\sigma}_{\epsilon,i}^2 &= \frac{1}{T}\sum_{t=3}^T \hat{\varepsilon}_{i,t}^2, \\
\hat{\omega}_{\epsilon,i}^2 &= \sum_{j=J+1}^{J-1} K(j)\frac{1}{T}\sum_{t=j+3}^T \hat{\varepsilon}_{i,t}\hat{\varepsilon}_{i,t-j},
\end{aligned}
$$

where $\hat{\varepsilon}_{i,t} = \hat{e}_{i,t} - \hat{\rho}_p\hat{e}_{i,t-1}$, $K(j) = 1 - j/(J+1)$ is the Bartlett kernel and J is the associated kernel bandwidth parameter, which is assumed to satisfy Assumption 8 below. The estimators of $\phi_{\epsilon,i}^4$ and $\lambda_{\epsilon,i}$ are given naturally by $\hat{\tau}_{\epsilon,i} = (\hat{\omega}_{\epsilon,i}^2 - \hat{\sigma}_{\epsilon,i}^2)/2$ and $\hat{\phi}_{\epsilon,i}^4 = \hat{\omega}_{\epsilon,i}^4$, respectively.

Assumption 8. $J/\min\{\sqrt{N}, \sqrt{T}\} \to 0$ as J, N, $T \to \infty$.

The second test statistic that we consider, denoted $PMSB_p$, is the panel modified Sargan–Bhargava (PMSB) test statistic of Bai and Ng (2010), as given by

$$
\begin{aligned}
PMSB_0 &= \frac{\sqrt{N}(N^{-1}T^{-2}\sum_{i=1}^N(\hat{e}_{i,-1}^0)'\hat{e}_{i,-1}^0 - \hat{\omega}_\epsilon^2/2)}{\sqrt{\hat{\phi}_\epsilon^4/3}}, \\
PMSB_1 &= \frac{\sqrt{N}(N^{-1}T^{-2}\sum_{i=1}^N(\hat{e}_{i,-1}^1)'\hat{e}_{i,-1}^1 - \hat{\omega}_\epsilon^2/6)}{\sqrt{\hat{\phi}_\epsilon^4/45}}.
\end{aligned}
$$

Theorem 1 reports the asymptotic null distributions of $P_{a,p}$, $P_{b,p}$ and $PMSB_p$.

Theorem 1. *Under Assumptions 1–8 and the null hypothesis that $\rho_1 = ... = \rho_N = 1$, as N, $T \to \infty$ with $N/T \to 0$,*

$$
\begin{aligned}
P_{a,p} &\to_d N(0,1), \\
P_{b,p} &\to_d N(0,1), \\
PMSB_p &\to_d N(0,1),
\end{aligned}
$$

where $\to_d$ signifies convergence in distribution.

According to Theorem 1 all three test statistics converge to $N(0,1)$ under the unit root null. It is not difficult to show, however, that $P_{a,p}$ and $P_{b,p}$ are asymptotically equivalent, which means that for these statistics the convergence is to the same normal variate. Note also that while $P_{a,p}$ and $P_{b,p}$ are (bias-adjusted) t-statistics of the largest autoregressive root, $PMSB_p$ is a ratio of variances. In spite of this difference, however, provided that the alternative formulated as that $|\rho_i| < 1$ for some i, all three statistics are left-tailed. The appropriate 5% critical value is therefore given by -1.645.

Remark 4. The fact that the PANICCA-based statistics are asymptotically $N(0,1)$ stands in sharp contrast to the results reported by Pesaran (2007) and Pesaran et al. (2013), who use $\overline{\mathbf{Z}}_t$ (and $\overline{\mathbf{z}}_t$) as a "proxy" for $\mathbf{F}_t$. This means that if $r_1 > 0$ the asymptotic distributions of their CIPS and CSB test statistics depend on the Brownian motion associated with $\mathbf{F}_t$. As alluded to in Section 1, this difference is due to the fact that here the estimation is done using only $\overline{\mathbf{z}}_t$ (the first-differenced data).

3.2 Testing $\mathbf{F}_t$

The rate of consistency of the CA estimator $\hat{\mathbf{f}}_t^p$ of (the space spanned by) $\mathbf{f}_t^p$ is the same as that of the PC estimator when $\sqrt{N}/T \to c \leq M$ and it is faster when $\sqrt{N}/T \to \infty$. The results reported by Bai and Ng (2004, Theorems 1 and 3) for the tests of the estimated PC factors therefore go through also in case of CA estimation.

The testing is carried out in the following fashion. If $r = m + 1 = 1$, such that $\hat{\mathbf{F}}_t^p$ (and $\mathbf{F}_t^p$) is a scalar, then the testing can be carried out using any existing unit root test. Bai and Ng (2004) only consider the augmented Dickey–Fuller (ADF) test, henceforth denoted ADF_p, and hence so do we. Let us therefore define $\Delta\hat{\mathbf{F}}^p = (\Delta\hat{\mathbf{F}}^p_{3+q}, ..., \Delta\hat{\mathbf{F}}^p_T)'$, $\hat{\mathbf{F}}^p_{-1} = (\hat{\mathbf{F}}^p_{2+q}, ..., \hat{\mathbf{F}}^p_{T-1})'$, $\mathbf{W} = (\mathbf{W}_{2+q}, ..., \mathbf{W}^p_T)'$, where $\mathbf{W}_t = (\Delta\hat{\mathbf{F}}^p_{t-1}, ..., \Delta\hat{\mathbf{F}}^p_{t-q})'$. The ADF statistic is given by

$$ADF_p = \frac{(\hat{\mathbf{F}}^p_{-1})'\mathbf{M}_{p+1}\mathbf{M}_W\mathbf{M}_{p+1}\Delta\hat{\mathbf{F}}^p}{\hat{\sigma}_\eta\sqrt{(\hat{\mathbf{F}}^p_{-1})'\mathbf{M}_{p+1}\mathbf{M}_W\mathbf{M}_{p+1}\hat{\mathbf{F}}^p_{-1}}},$$

where $\hat{\sigma}^2_\eta = T^{-1}(\Delta\hat{\mathbf{F}}^p)'\mathbf{M}_{p+1}\mathbf{M}_W\mathbf{M}_{p+1}\Delta\hat{\mathbf{F}}^p$, $\mathbf{M}_W = \mathbf{I}_{T-q-1} - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'$ and the last q rows of $\mathbf{M}_p$ has been removed to make it conformable with $\mathbf{W}$. Note how the dependence on q has been suppressed in ADF_p.

If $r = m + 1 > 1$, then the following sequential test procedure can be used to determine r_1, the number of unit roots in $\hat{\mathbf{F}}^p_t$:

1. Set $k = r$.

2. Compute $\hat{\mathbf{Y}}^p_k = (\hat{\mathbf{Y}}^p_{2,k}, ..., \hat{\mathbf{Y}}^p_{T,k})' = \mathbf{M}_{p+1}\hat{\mathbf{F}}^p\hat{\boldsymbol{\beta}}_k$, where $\hat{\boldsymbol{\beta}}_k$ is the $(m+1)\times k$ matrix of eigenvectors associated with the k largest eigenvalues of $T^{-1}(\hat{\mathbf{F}}^p)'\mathbf{M}_{p+1}\hat{\mathbf{F}}^p$.

3. The test statistic is given by

$$MQ_p(k) = T[\hat{w}_p(k) - 1],$$

where $\hat{w}_p(k)$ is the smallest eigenvalue of

$$\frac{1}{2}[(\hat{\mathbf{Y}}^p_{k,-1})'\hat{\mathbf{Y}}^p_k + (\hat{\mathbf{Y}}^p_k)'\hat{\mathbf{Y}}^p_{k,-1} - T(\hat{\boldsymbol{\Sigma}}_k + \hat{\boldsymbol{\Sigma}}'_k)][(\hat{\mathbf{Y}}^p_{k,-1})'\hat{\mathbf{Y}}^p_{k,-1}]^{-1},$$

with $\hat{\boldsymbol{\Sigma}}_k = \sum_{j=1}^{J-1} K(j)T^{-1}\sum_{t=j+3}^{T}\hat{\mathbf{U}}_{t-j,k}\hat{\mathbf{U}}'_{t,k}$, $\hat{\mathbf{U}}_{t,k}$ is the residual from a LS fit of $\hat{\mathbf{Y}}_{t,k}$ onto $\hat{\mathbf{Y}}_{t-1,k}$, and where J is assumed to satisfy Assumption 8.

4. If the null hypothesis that $r_1 = k$ is rejected using $MQ_p(k)$, set $k = k - 1$ and return to step 2. Otherwise, set $\hat{r}_1 = k$ and stop.

Remark 5. MQ_p is the MQ_c test statistic of Bai and Ng (2004) applied to the estimated CA factors. Bai and Ng (2004) also consider another statistic, denoted MQ_f. However, since MQ_c is more general, in this paper we only consider the CA version of this statistic.

Assumption 9. $q^3/\min\{N, T\} \to 0$ as q, N, $T \to \infty$.

In Theorem 2 we report the asymptotic null distributions of ADF_p and $MQ_p(k)$. In so doing it is useful to introduce the following detrended Brownian motion:

$$\mathbf{W}^p_r(s) = \mathbf{W}_r(s) - \int_0^1 \mathbf{W}_r(v)\mathbf{d}_p(v)'dv\left(\int_0^1 \mathbf{d}_p(v)\mathbf{d}_p(v)'dv\right)^{-1}\mathbf{d}_p(s),$$

where $\mathbf{d}_p(s) = (1, ..., s^p)'$ is the limiting trend function and $\mathbf{W}_r(s)$ is an $r \times 1$ vector standard Brownian motion.

Theorem 2. *Under Assumptions 1–9 the following results hold as N, $T \to \infty$:*

(i) *Suppose that $r = m + 1 = 1$. Under the null hypothesis that $\mathbf{F}_t$ has a unit root,*

$$ADF_p \to_w \frac{\int_0^1 \mathbf{W}_1^p(s) d\mathbf{W}_1(s)}{\sqrt{\int_0^1 [\mathbf{W}_1^p(s)]^2 ds}},$$

where $\to_w$ signifies weak convergence.

(ii) *Suppose that $r = m + 1 > 1$. Under the null hypothesis that $\mathbf{F}_t$ has k unit roots,*

$$MQ_p(k) \to_w T[w_p(k) - 1],$$

where $w_p(k)$ is the smallest eigenvalue of

$$\frac{1}{2}[\mathbf{W}_r^p(1)\mathbf{W}_r^p(1)' - \mathbf{I}_k] \left(\int_0^1 \mathbf{W}_r^p(s)\mathbf{W}_r^p(s)' ds \right)^{-1}.$$

The asymptotic distribution of ADF_p is identically the ADF test distribution, for which critical values are readily available (see, for example, MacKinnon, 1996). Appropriate 1%, 5% and 10% critical values for $MQ_p(k)$ ($k = 1, ..., 6$) can be found in Table 1 of Bai and Ng (2004).

Remark 6. The results reported so far make use of Assumption 4, which only requires that the true number of factors, r, is less than or equal to $m + 1$. If one would like to pinpoint r, one possibility is to employ an information criterion. This approach has been shown to work in the context of PC estimation (see Bai and Ng, 2002), and, as pointed out by Pesaran et al. (2013, Section 4.1), it is expected to work well also for CA. The information criterion considered in this paper, which can be seen as a multivariate analog of the PC-specific IC_{p3} criterion of Bai and Ng (2002), takes the form

$$IC(s) = \ln[\det(\hat{\mathbf{\Sigma}}_s)] + s \cdot N^{-1} \ln(N), \tag{7}$$

where $\hat{\mathbf{\Sigma}}_s = (NT)^{-1} \sum_{i=1}^N (\mathbf{v}_i^p)' \mathbf{v}_i^p$ is the sum of squared residuals in (6) based on using $s \leq m + 1$ cross-section averages. The penalty, $s \cdot N^{-1} \ln(N)$, is the same as in IC_{p3} with $T = 0$. The reason for this difference is that while the rate of consistency of the PC estimator depends on both N

and T, as we explain in Remark 2 of Section 3, the rate of consistency of $\hat{\mathbf{f}}^p$ only depends on N (see Bai and Ng, 2002, page 219, for a discussion). The estimator $\hat{r}$ of r is given naturally by

$$\hat{r} = \arg \min_{s=0,\ldots,m+1} IC(s).$$

The consistency of $\hat{r}$ is a direct consequence of Corollary 2 of Bai and Ng (2002), which is not PC-specific but applies to any estimator of $\mathbf{f}^p$.

4 Monte Carlo simulations

4.1 Testing $e_{i,t}$

The relative performance of PANICCA when compared to original PANIC is assessed through a small-scale Monte Carlo simulation study. The DGP used for this purpose is given by a simplified version of (1)–(3) that sets $r = 3$, $m = 2$, $\boldsymbol{\alpha}_i \sim U(0,1)$, $\boldsymbol{\beta}_i \sim U(0,1)$, $\boldsymbol{\lambda}_i' = (1, l_i, l_i)$ and

$$\boldsymbol{\Lambda}_i' = \begin{bmatrix} l_i & 1 & l_i \\ l_i & l_i & 1 \end{bmatrix},$$

where $l_i = 1.5 \cdot 1(i > N/2) - 0.5$ and $1(A)$ is the indicator function for the event A. This parametrization of $\boldsymbol{\lambda}_i$ and $\boldsymbol{\Lambda}_i$ ensures that $\overline{\mathbf{C}}$ has ones on the main diagonal and 0.25 elsewhere, which means that Assumption 4 is met. Also, $e_{i,t} = \rho e_{i,t-1} + \epsilon_{i,t}$, $\mathbf{u}_{i,t} = \rho \mathbf{u}_{i,t-1} + \boldsymbol{\varepsilon}_{i,t}$ and $\mathbf{F}_t = \delta \mathbf{F}_{t-1} + \boldsymbol{\eta}_t$, where $(\epsilon_{i,t}, \boldsymbol{\varepsilon}_{i,t}', \boldsymbol{\eta}_t')' \sim N(\mathbf{0}_{6\times 1}, \mathbf{I}_6)$. We begin by considering the 5% size and size-corrected power of $P_{a,p}$, $P_{a,p}$ and $PMSB_{a,p}$. In the size experiments, $\rho = \delta = 1$, while in the power experiments, $\rho = 0.95$ and $\delta = 0.5$. All results are based on making 5,000 draws of panels where N and T are chosen so as to illustrate the main difference between PANIC and PANICCA, which occurs naturally when the sample size is relatively small. We chose $N, T \in \{10, 20, 35, 50\}$, which is consistent with the bulk of empirical work based on PANIC (see, for example, Lanzafame, 2010; Schmidt and Vosen, 2013; Martín, 2013; Joseph et al., 2012, 2013; Örsal and Dilan, 2014; Blomquist and Westerlund, 2014).

We begin by considering the size results for the intercept-only case when $p = 0$, which are reported in Table 1. The information content of this table may be summarized as follows.

- While there are some noticeable distortions, these are mainly among the smaller values of N and T. In fact, size accuracy is quite good already with $T = 50$ and $N = 35$, and it increases with increasing values of T and to a lesser extent with increasing values N.

That the effect of increasing T is relatively more pronounced is in agreement with the condition that $N/T \to 0$.

- The distortions are generally somewhat smaller for PANICCA than for PANIC. This corroborates the findings of Westerlund and Urbain (2015) in the factor-augmented regression case, suggesting that CA tend to be more accurate than PC.
- Looking across the three types of tests, the best size accuracy is generally obtained by using the PSMB tests.
- The PANICCA-based tests are uniformly more powerful than their PANIC-based counterparts. The difference in power is large enough not to be ignored and can in fact be quite substantial.
- In agreement with their relatively high rejection frequencies under the null, the best power is generally obtain by using the P_a- and P_b-type tests. This finding is consistent with the results of Westerlund (2015b), showing how the local asymptotic power of the PANIC versions of these tests is higher than that of the PANIC-based PSMB test.
- The fact that the difference in size and power is decreasing in N and T is consistent with the asymptotic equivalence of PANICCA and PANIC.

The results reported in Table 2 for the case with an intercept and trend ($p = 1$) are very similar to those reported in Table 1, and we therefore just briefly describe them. The first thing to note is that the size distortions are actually reduced as the linear trend is added, which is somewhat unexpected, because usually the distortions are increasing in p. Another difference worth noting is the power, which is much lower in Table 2 than in Table 1. In fact, the power in the linear trend case only rarely raises above the nominal 5% level. That the power is reduced by the inclusion of the linear trend is a reflection of the well-known "incidental trends problem" (see Westerlund, 2015b), and is therefore expected.

4.2 Testing $\mathbf{F}_t$

In this subsection, we investigate the performance of the sequential procedure to determine r_1, the number of unit root factors. The DGP is the same as before. The only difference is the common factors, which are now generated according to $\mathbf{F}_t = \text{diag}(\delta_0 \cdot \mathbf{1}'_{(r-r_1)\times 1}, \mathbf{1}'_{r_1\times 1})\mathbf{F}_{t-1} + \boldsymbol{\eta}_t$, where $\boldsymbol{\eta}_t$ is as before, $\mathbf{1}_{r\times 1} = (1, ..., 1)'$ is an $r \times 1$ vector of ones, and $|\delta_0| < 1$ is the autoregressive coefficient of the stationary factors (see Bai and Ng, 2004, for a similar parametrization). In interest of space, we focus on the results for the case when $N = 20$ and $T = 50$.

The most striking observation that can be made from Table 3 is that the proposed CA-based estimator $\hat{r}_1$ of r_1 is much more robust to variations in ρ than the corresponding (PANIC) estimator based on PC. In fact, in a majority of cases the PC bias was twice as large as the corresponding CA bias. According to the results reported by Bai and Ng (2004), the PC-based estimator of r_1 is more robust than both the trace test-based procedure of Johansen (1995), and the information criterion of Aznar and Salvador (2002). Being more accurate than PC, the CA-based estimator is expected to outperform also these other estimation approaches.

5 An application to the EMH

A financial market is said to be efficient if prices fully reflect all available information and no profit opportunities are left unexploited. The agents form their expectations rationally and rapidly arbitrage away any deviations of the expected returns consistent with supernormal profits. Therefore, if currency markets are efficient, the spot (forward) exchange rate should embody all relevant information, and it should not be possible to forecast one spot (forward) rate as a function of another. In what follows we refer to this proposition of the EMH as the efficient cross-market hypothesis (ECMH). Also, provided that agents are risk neutral and that the risk premium is stationary, the current forward rate should be an unbiased predictor of the future spot rate. This is the forward rate unbiasedness hypothesis (FRUH).

The validity of the above propositions the EMH has been, and still is, one of the most heavily researched areas in the financial literature. However, a lot of controversy still exists about the method that must be applied to test for its existence. In particular, the use of cointegration techniques has become very popular, and is by now the workhorse of the industry (see Zivot, 2000, for a survey of the cointegration-based literature). Indeed, since the seminal work of Hakkio and Rush (1989), it is well recognized that the FRUH requires that the future spot and current forward rates are cointegrated and one-to-one. Also, if the ECMH holds, then spot and forward rates cannot be cointegrated across markets.

Interestingly, while each of these propositions of the EMH occupies a huge literature (see, for example, Hakkio and Rush, 1989; Baillie and Bollerslev, 1989; Crowder, 1994, for early contributions), we know of no previous study that has tried to formalize the connection between the two. In particular, since both spot and forward rates from across a variety of markets exhibit unit root-like behavior, a natural question concerns the source of the non-stationarity. To formalize matters slightly, let us denote by $s_{i,t}$ ($f_{i,t}$) the log spot (forward) rate of market i at time t. In terms of the model of Section 2, $Y_{i,t} = s_{i,t+1}$ and $\mathbf{X}_{i,t} = f_{i,t}$. Consider $s_{i,t}$. According to the ECMH, this variables must not be cointegrated across markets. In order to appreciate the implications of this, it is useful to note that

$$s_{i,t+1} - \theta s_{j,t+1} = (\boldsymbol{\alpha}_i - \theta \boldsymbol{\alpha}_j)' \mathbf{D}_{t,p} + (\boldsymbol{\lambda}_i - \theta \boldsymbol{\lambda}_j)' \mathbf{F}_t + e_{i,t} - \theta e_{j,t}.$$

Obviously, being idiosyncratic, $e_{i,t}$ and $e_{j,t}$ cannot be cointegrated for $i \neq j$. Hence, for the ECMH to hold it must be that $(\boldsymbol{\lambda}_i - \theta \boldsymbol{\lambda}_j)' \mathbf{F}_t$ and/or $e_{1,t}, ..., e_{N,t}$ are unit root non-stationary, such that $s_{i,t+1} - \theta s_{j,t+1}$ is unit root non-stationary too. Similarly, for $f_{i,t} - \theta f_{j,t}$ to be non-stationary, we require that $(\boldsymbol{\Lambda}_i - \theta \boldsymbol{\Lambda}_j)' \mathbf{F}_t$ and/or $\mathbf{u}_{1,t}, ..., \mathbf{u}_{N,t}$ are unit root non-stationary. Of course, only one of the conditions have to be met for the EMH not to fail, and in the current paper we therefore test whether $e_{1,t}, ..., e_{N,t}$ and $\mathbf{u}_{1,t}, ..., \mathbf{u}_{N,t}$ are unit root non-stationary. But we also have

$$s_{i,t+1} - f_{i,t} = (\boldsymbol{\alpha}_i - \boldsymbol{\beta}_i)' \mathbf{D}_{t,p} + (\boldsymbol{\lambda}_i - \boldsymbol{\Lambda}_i)' \mathbf{F}_t + e_{i,t} - \mathbf{u}_{i,t},$$

which means that for $s_{i,t+1}$ and $f_{i,t}$ to be cointegrated and one-to-one, as dictated by the FRUH, the following additional conditions must be satisfied: $\boldsymbol{\alpha}_i = \boldsymbol{\beta}_i$, $(\boldsymbol{\lambda}_i - \boldsymbol{\Lambda}_i)' \mathbf{F}_t$ is stationary, and $e_{i,t}$ and $\mathbf{u}_{i,t}$ must be either stationary, or cointegrated and one-to-one.

As the above discussion makes clear, the ECMH and FRUH arise naturally as restrictions on the general factor model considered here. All-in-all, we have the following four restrictions:

R1. $e_{1,t}, ..., e_{N,t}$ and $\mathbf{u}_{1,t}, ..., \mathbf{u}_{N,t}$ are unit root non-stationary;

R2. $\boldsymbol{\alpha}_i = \boldsymbol{\beta}_i$;

R3. $(\boldsymbol{\lambda}_i - \boldsymbol{\Lambda}_i)' \mathbf{F}_t$ is stationary;

R4. $e_{i,t}$ and $\mathbf{u}_{i,t}$ are either stationary, or cointegrated and one-to-one.

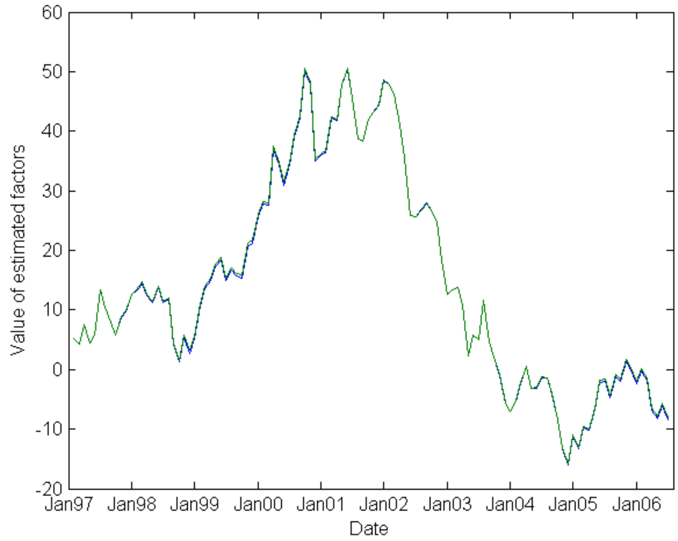
While R1 is a test of the ECMH, R2–R4 test the FRUH. In what remains we test each of restrictions. The test machinery developed in the present paper is ideally suited for this task, as it does not place any restrictions on the source of the (potential) non-stationarity of the data.

Also, unlike original PANIC, in which the common component of each variable would be estimated separately, in PANICCA the common component of both variables is estimated jointly, leading to increased efficiency (see Westerlund and Urbain, 2015).[1] The data set that we use is the same as in Westerlund (2007), and consists of monthly spot and forward exchange rates relative to the United States dollar. The sample covers 15 OECD countries between February 1997 and July 2006. Hence, $N = 15$ and $T = 115$. The choice of data set is motivated in part by comparability, in part by the preference of Westerlund (2007) to treat the factors as stationary, a restriction that is never tested.

We begin by testing R1, that is, we test if the estimated idiosyncratic components of both the spot and forward rates can be characterized as unit root non-stationary. The tests are implemented as described in Sections 3 and 4. Also, since both variables do not appear to be trending, we focus on the constant-only specification. The results reported in Table 4 are mixed. In particular, while according to $P_{a,0}$ the unit root null should be rejected, according to $PMSB_0$ it should not. The evidence based on $P_{b,0}$ is more ambiguous, favoring a rejection at the 10% level but not at the 5% level. Of course, given the broad formulation of the alternative hypothesis (as that there is at least one country for which the idiosyncratic component is stationary), ideally the results should be overwhelmingly against the null in case of a rejection. However, this is not what we observe. In view of this, and the tendency of $P_{a,0}$ and $P_{b,0}$ to overreject in small-N panels (see Section 4), in what follows we treat the idiosyncratic components of both $s_{i,t}$ and $f_{i,t}$ as unit root non-stationary, a conclusion that is supported by some (unreported) unit-by-unit ADF test results. The implication of this result is that $s_{i,t}$ and $f_{i,t}$ cannot be cointegrated across countries, which is consistent with the ECMH.

[1] As a referee of this journal point out, one could also consider a multivariate extension of original PANIC.

Figure 1: Estimated common factors.



Having tested R1 we now continue to R2–R4, which represent the FRUH part of the EMH. We begin by examining the the common component. Application of the $MQ_0(k)$-based sequential test procedure to the common component of $\mathbf{Z}_{i,t} = (s_{i,t+1}, f_{i,t})'$ yields $\hat{r}_1 = 1$. Hence, there is a unique cointegrating relationship between the two factors, $F_{1,t}$ and $F_{2,t}$ say, in $\mathbf{F}_t$. In order to appreciate the implication of this for R3, note how $(\boldsymbol{\lambda}_i - \boldsymbol{\Lambda}_i)'\mathbf{F}_t = (\lambda_{1,i} - \Lambda_{1,i})[F_{1,t} - \gamma_i F_{2,t}]$, where $\gamma_i = -(\lambda_{2,i} - \Lambda_{2,i})/(\lambda_{1,i} - \Lambda_{1,i})$, and $\lambda_{n,i}$ ($\Lambda_{n,i}$) is the n-th row of $\boldsymbol{\lambda}_i$ ($\boldsymbol{\Lambda}_i$). Cointegration between $F_{1,t}$ and $F_{2,t}$ therefore implies that $(\boldsymbol{\lambda}_i - \boldsymbol{\Lambda}_i)'\mathbf{F}_t$ is stationary (see also Gengenbach et al., 2006). The finding that $F_{1,t}$ and $F_{2,t}$ are cointegrated is supported by Figure 1, which plots the estimated CA factors. Both factors exhibit clear and strikingly similar unit root-like behavior. In fact, the lines representing the factors almost coincide. R3 is therefore supported by the data. The non-stationarity of the factors implies that previous results based on assuming either that the factors are stationary or indeed absent altogether (see, for example, Westerlund, 2007) should be reconsidered.

A test of R4 involves testing for cointegration between $\hat{e}^0_{i,t}$ and $\hat{\mathbf{u}}^0_{i,t}$, which is can be carried out in a very straightforward fashion. Note in particular that since $\hat{e}^0_{i,t}$ and $\hat{\mathbf{u}}^0_{i,t}$ are (asymptotically) cross-section independent, we may apply any first-generation test statistic designed for such cross-section independent panels. We choose the panel-t and group-t statistics of Pedroni (2004), which are two of the most popular (and scrutinized) test statistics in the literature. The main difference between the two is that while the panel-t statistic is based on within pooling, the group-t statistic is based on between pooling. The results reported in Table 4 suggests that no cointegration null is strongly rejected even at the 1% level, which we take as evidence in favor of cointegration.

The next step in the test of R4 involves testing if the cointegrating slope on $f_{i,t}$ is indeed unity, as postulated by theory. The estimated cointegrating slopes of both the common and idiosyncratic components are reported in Table 5. Again, given the consistency of the component estimates, the estimation of the cointegrating relationship can be carried out as if the components are in fact observed. We therefore follow the usual practice and apply fully modified LS (FMLS) and dynamic LS (DLS) techniques (see, for example, Pedroni, 2001). These are robust to endogeneity, but in the panel case not to cross-section dependence, which is also not necessary since the idiosyncratic components are cross-section independent by assumption. Analogous to the cointegration testing, we consider both a group estimator and a panel estimator, which both allow for country-specific fixed effects. The first thing to note is that the slope estimates are very close to one. In case of the common factors, the evidence against the null hypothesis of a unit slope is insignificant, as is the evidence for the idiosyncratic component based on the group mean estimator. However, according to the panel estimator, the slope is significantly different from one. But since the point estimate is very close to one, our overall interpretation of the results is still in support of the unit slope hypothesis. The bulk of the evidence is therefore in favor of R4.

Since the deterministic component is eliminated prior to estimating the components of the data, unlike R1, R3 and R4, in PANICCA there is no natural test of R2. Westerlund and Blomquist (2013) develop a (PANIC-based) test for the presence of a linear trend in (1), which is based on testing if the average of the first-differenced data is zero. We test if $(\bar{s}_{i,+1} - \overline{f}_i)$ is zero on average, which can be done using a simple t-test. The logic behind this test stems from the fact that under R3 and R4, $(\bar{s}_{+1} - \overline{f}) = N^{-1}\sum_{i=1}^{N}(\bar{s}_{i,+1} - \overline{f}_i)$ is a consistent estimator of $(\overline{\alpha} - \overline{\beta})$, which is zero under R2. Applying this test to the data, we find that $(\bar{s}_{+1} - \overline{f}) \approx 0.0003$ and the associated t-statistic is 0.25, leading to a clear non-rejection of the zero intercept null. Hence, this test does not provide any evidence against R2.

The results reported in this section suggest that the evidence against the EMH is weak, at best. In fact, most of the restrictions of the hypothesis seem to be satisfied in our sample. This is noteworthy because, despite the wide acceptance of the EMH in theory, most cointegration-based studies tend to reject the EMH (see Zivot, 2000, for a review of the literature). One explanation of this difference in the results is the generality of the DGP considered here, which does not impose any assumptions on the nature of the non-stationarity of the data. In fact, PANICCA seems to provide a natural platform for testing the unit root and cointegration implications of the EMH.

6 Conclusions

The CA approach of Pesaran (2006) is one of the most convenient approaches around for dealing with the effects of cross-section dependence. However, the way that this approach is implemented when testing for unit roots has resulted in test statistics with nonstandard asymptotic distributions and, as a result, complicated implementation. The current paper can be seen as a reaction to this. The purpose is to develop CA-based tests that support asymptotically normal inference. As a starting point we take the PANIC approach of Bai and Ng (2004, 2010), which is one of the most general panel unit root test approaches around. Original PANIC uses PC to estimate the common and idiosyncratic components of the data. CA is more convenient and has been shown to perform relatively well in small samples. These considerations lead naturally to PANICCA, PANIC based on CA rather than PC.

References

Aznar, A., and M. Salvador (2002). Selecting the Rank of the Cointegration Space and the Form of the Intercept Using an Information Criteria. *Econometric Theory* **18**, 926–947.

Bai, J., and S. Ng (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica* **70**, 191–221.

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

Bai, J., and S. Ng (2010). Panel Unit Root Tests with Cross-Section Dependence: A Further Investigation. *Econometric Theory* **26**, 1088–1114.

Bai, J., and J. L. Carrion-i-Silvestre (2009). Structural Changes, Common Stochastic Trends, and Unit Roots in Panel Data. *Review of Economic Studies* **76**, 471–501.

Bai, J., and J. L. Carrion-i-Silvestre (2013). Testing Panel Cointegration with Unobservable Dynamic Common Factors That are Correlated with the Regressors. *Econometrics Journal* **16**, 222–249.

Baillie, R. T., and T. Bollerslev (1989). Common Stochastic Trends in a System of Exchange Rates. *Journal of Finance* **44**, 167–181.

Baltagi, B. (2013). *Econometric Analysis of Panel Data, Fifth Edition*. John Wiley and Sons, New York.

Blomquist, J., and J. Westerlund (2014). A Non-Stationary Panel Data Investigation of the Unemployment–Crime Relationship. *Social Science Research* **44**, 114–125.

Breitung J., and M. H. Pesaran (2008). Unit Roots and Cointegration in Panels. In Matyas, L., and P. Sevestre (Eds.), *The Econometrics of Panel Data*, Klüwer Academic Publishers, 279–322.

Chudik, A., M. H. Pesaran and E. Tosetti (2011). Weak and Strong Cross Section Dependence and Estimation of Large Panels. *Econometrics Journal* **14**, C45–C90.

Crowder, W. J. (1994). Foreign Exchange Market Efficiency and Common Stochastic Trends. *Journal of International Money and Finance* **13**, 551–564.

Gengenbach, C., F. C. Palm and J.-P. Urbain (2006). Cointegration Testing in Panels with Common Factors. *Oxford Bulletin of Economics and Statistics* **68**, 683–719.

Gengenbach, C., F. C. Palm and J.-P. Urbain (2010). Panel Unit Root Tests in the Presence of Cross-Sectional Dependencies: Comparison and Implications for Modelling. *Econometric Reviews* **29**, 111–145.

Hakkio, C. S., and M. Rush (1989). Market Efficiency and Cointegration: An Application to the Sterling and Deutschmark Exchange Markets. *Journal of International Money and Finance* **8**, 75–88.

Johansen, S. (1995). *Likelihood Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford.

Joseph, B. P., F. Giorgio and F. Norbert (2012). Interest Rate Co-Movements, Global Factors and the Long end of the Term Spread. *Journal of Banking and Finance* **36**, 183–192.

Joseph, B. P., F. Giorgio and F. Norbert (2013). Primary Commodity Prices: Co-Movements, Common Factors and Fundamentals. *Journal of Development Economics* **101**, 16–26.

Kapetanios, G., and M. H. Pesaran (2005). Alternative Approaches to Estimation and Inference in Large Multifactor Panels: Small Sample Results with an Application to Modelling of Asset Returns. CESifo Working Paper Series 1416.

Lanzafame, M. (2010). The Nature of Regional Unemployment in Italy. *Empirical Economics* **39**, 877–895.

MacKinnon, J. G. (1996). Numerical Distribution Functions for Unit Root and Cointegration Tests. *Journal of Applied Econometrics* **11**, 601–618.

Martín, S. (2013). A Market-Based Approach to Sector Risk Determinants and Transmission in the Euro Area. *Journal of Banking and Finance 37*, 4534–4555.

Pedroni, P. (2001). Purchasing Power Parity Tests in Cointegrated Panels. *Review of Economics and Statistics* **83**, 727–731.

Pesaran, M. H. (2006). Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure. *Econometrica* **74**, 967–1012.

Pesaran, H. M. (2007). A Simple Panel Unit Root Test in Presence of Cross-Section Dependence. *Journal of Applied Econometrics* **22**, 265–312.

Pesaran, H. M., L. V. Smith and T. Yamagata (2013). Panel Unit Root Test in the Presence of a Multifactor Error Structure. *Journal of Econometrics* **175**, 94–115.

Schmidt, T., and S. Vosen (2013). Demographic Change and the Labour Share of Income. *Journal of Population Economics* **26**, 357–378.

Westerlund, J. (2007). Unbiased Estimation of Cointegrated Panel Regressions with Cross Section Dependence. *Journal of Financial Econometrics* **5**, 491–522.

Westerlund, J. (2014). Heteroskedasticity Robust Panel Unit Root Tests. *Journal of Business & Economic Statistics* **32**, 112–135.

Westerlund, J., and W. Hess (2011). A New Poolability Test for Cointegrated Panels. *Journal of Applied Econometrics* **26**, 56–88.

Westerlund, J., and R. Larsson (2009). A Note on the Pooling of Individual PANIC Unit Root Tests. *Econometric Theory* **25**, 1851–1868.

Westerlund, J., and R. Larsson (2012). Testing for Unit Roots in a Panel Random Coefficient Model. *Journal of Econometrics* **167**, 254–273.

Westerlund, J., and J. Blomquist (2013). PANIC in the Presence of Uncertainty about the Deterministic Trend. *Oxford Bulletin of Economics and Statistics* **75**, 123–135.

Westerlund, J., and J.-P. Urbain (2013). On the Estimation and Inference in Factor-Augmented Panel Regressions with Correlated Loadings. *Economics Letters* **119**, 247–250.

Westerlund, J. (2015a). The Effect of Recursive Detrending on Panel Unit Root Tests. *Journal of Econometrics* **185**, 453–467.

Westerlund, J. (2015b). The Power of PANIC. *Journal of Econometrics* **185**, 495–509.

Westerlund, J, and J.-P. Urbain (2015). Cross-Sectional Averages versus Principal Components. *Journal of Econometrics* **185**, 372–377.

Zivot, E. (2000). Cointegration and Forward and Spot Exchange Rate Regressions. *Journal of International Money and Finance* **19**, 785–812.

Örsal, K., and D. Dilan (2014). Do the Global Stochastic Trends Drive the Real House Prices in OECD Countries? *Economics Letters* **123**, 9–13.

Appendix: Proofs

Define $e_{i,t}^{p} = \sum_{n=2}^{t}(\Delta e_{i,n})^{p}$. Let $\mathbf{s}_t = \hat{f}_t^{p} - \overline{\mathbf{C}}'\mathbf{f}_t^{p}$, $\mathbf{S}_t = \sum_{n=2}^{t}\mathbf{s}_n$ and $\mathbf{d}_i = \hat{\boldsymbol{\lambda}}_i - \overline{\mathbf{C}}^{-}\boldsymbol{\lambda}_i$.

Lemma A.1. *Under Assumptions 1–7, uniformly in $i = 1, ..., N$ and $t = 2, ..., T$,*

$$||\mathbf{s}_t|| = O_p(N^{-1/2}), \tag{i}$$

$$||\mathbf{S}_t|| = O_p(\sqrt{T}N^{-1/2}), \tag{ii}$$

$$||\mathbf{d}_i|| = O_p(T^{-1/2}) + O_p(N^{-1}). \tag{iii}$$

Proof of Lemma A.1.

From $\hat{\mathbf{f}}_t^{p} = \overline{\mathbf{z}}_t^{p} = \overline{\mathbf{C}}'\mathbf{f}_t^{p} + \overline{\mathbf{v}}_t^{p}$, we obtain $\mathbf{s}_t = \hat{\mathbf{f}}_t^{p} - \overline{\mathbf{C}}'\mathbf{f}_t^{p} = \overline{\mathbf{v}}_t^{p} = O_p(N^{-1/2})$, as required for (i). The result in (ii) is a direct consequence of this;

$$||T^{-1/2}\mathbf{S}_t|| = N^{-1/2}\left|\left|\frac{1}{\sqrt{T}}\sum_{n=2}^{t}\sqrt{N}\overline{\mathbf{v}}_t^{p}\right|\right| = O_p(N^{-1/2}). \tag{A1}$$

For $\mathbf{d}_i$, note that $\hat{\mathbf{C}}_i = [T^{-1}(\hat{\mathbf{f}}^{p})'\hat{\mathbf{f}}^{p}]^{-1}T^{-1}(\hat{\mathbf{f}}^{p})'\mathbf{z}_i^{p}$, where, via $\mathbf{f}^{p} = (\hat{\mathbf{f}}^{p} - \overline{\mathbf{v}}^{p})\overline{\mathbf{C}}^{-}$,

$$\begin{aligned}
T^{-1}(\mathbf{z}_i^{p})'\hat{\mathbf{f}}^{p} &= T^{-1}\mathbf{C}_i'(\mathbf{f}^{p})'\hat{\mathbf{f}}^{p} + T^{-1}(\mathbf{v}_i^{p})'\hat{\mathbf{f}}^{p} \\
&= T^{-1}\mathbf{C}_i'(\overline{\mathbf{C}}^{-})'(\hat{\mathbf{f}}^{p})'\hat{\mathbf{f}}^{p} - T^{-1}\mathbf{C}_i'(\overline{\mathbf{C}}^{-})'(\overline{\mathbf{v}}^{p})'(\mathbf{f}^{p}\overline{\mathbf{C}} + \overline{\mathbf{v}}^{p}) + T^{-1}(\mathbf{v}_i^{p})'(\mathbf{f}^{p}\overline{\mathbf{C}} + \overline{\mathbf{v}}^{p}) \\
&= \mathbf{C}_i'(\overline{\mathbf{C}}^{-})'T^{-1}(\hat{\mathbf{f}}^{p})'\hat{\mathbf{f}}^{p} - \mathbf{C}_i'(\overline{\mathbf{C}}^{-})'T^{-1}(\overline{\mathbf{v}}^{p})'\mathbf{f}^{p}\overline{\mathbf{C}} - \mathbf{C}_i'(\overline{\mathbf{C}}^{-})'T^{-1}(\overline{\mathbf{v}}^{p})'\overline{\mathbf{v}}^{p} \\
&\quad + T^{-1}(\mathbf{v}_i^{p})'\mathbf{f}^{p}\overline{\mathbf{C}} + T^{-1}(\mathbf{v}_i^{p})'\overline{\mathbf{v}}^{p}.
\end{aligned}$$

Here $||T^{-1}(\overline{\mathbf{v}}^{p})'\mathbf{f}^{p}|| = O_p((NT)^{-1/2})$ and $||T^{-1}(\overline{\mathbf{v}}^{p})'\overline{\mathbf{v}}^{p}|| = O_p(N^{-1})$, which are dominated by $T^{-1}(\mathbf{v}_i^{p})'\mathbf{f}^{p}$ and $T^{-1}(\mathbf{v}_i^{p})'\overline{\mathbf{v}}^{p}$. The first of these is $O_p(T^{-1/2})$. For the second, we use

$$\begin{aligned}
NT^{-1}(\mathbf{v}_i^{p})'\overline{\mathbf{v}}^{p} &= \frac{1}{T}\sum_{t=2}^{T}\sum_{j=1}^{N}\mathbf{v}_{i,t}^{p}\mathbf{v}_{j,t}^{p} = \frac{1}{T}\sum_{t=2}^{T}(\mathbf{v}_{i,t}^{p})^2 + \frac{\sqrt{N}}{\sqrt{T}}\frac{1}{\sqrt{NT}}\sum_{t=2}^{T}\sum_{j\neq i}^{N}\mathbf{v}_{i,t}^{p}\mathbf{v}_{j,t}^{p} \\
&= O_p(1) + O_p(\sqrt{N}T^{-1/2}),
\end{aligned}$$

implying that $||T^{-1}(\mathbf{v}_i^{p})'\overline{\mathbf{v}}^{p}|| = O_p(N^{-1}) + O_p((NT)^{-1/2})$. It follows that

$$\begin{aligned}
&||T^{-1}(\mathbf{z}_i^{p})'\hat{\mathbf{f}}^{p} - \mathbf{C}_i'(\overline{\mathbf{C}}^{-})'T^{-1}(\hat{\mathbf{f}}^{p})'\hat{\mathbf{f}}^{p}|| \\
&\leq ||\mathbf{C}_i||||\overline{\mathbf{C}}^{-}||||T^{-1}(\overline{\mathbf{v}}^{p})'\mathbf{f}^{p}||||\overline{\mathbf{C}}|| + ||\mathbf{C}_i||||\overline{\mathbf{C}}^{-}||||T^{-1}(\overline{\mathbf{v}}^{p})'\overline{\mathbf{v}}^{p}|| + ||T^{-1}(\mathbf{v}_i^{p})'\mathbf{f}^{p}||||\overline{\mathbf{C}}|| \\
&+ ||T^{-1}(\mathbf{v}_i^{p})'\overline{\mathbf{v}}^{p}|| \\
&= O_p(T^{-1/2}) + O_p(N^{-1}).
\end{aligned}$$

Substitution of this result into $\hat{\mathbf{C}}_i$ yields, with $||T^{-1}(\hat{\mathbf{f}}^p)'\hat{\mathbf{f}}^p|| = O_p(1)$,

$$||\hat{\mathbf{C}}_i - \overline{\mathbf{C}}^{-}\mathbf{C}_i|| = ||[T^{-1}(\hat{\mathbf{f}}^p)'\hat{\mathbf{f}}^p]^{-1}T^{-1}(\hat{\mathbf{f}}^p)'\mathbf{z}_i^p - \overline{\mathbf{C}}^{-}\mathbf{C}_i|| = O_p(T^{-1/2}) + O_p(N^{-1}). \tag{A2}$$

Hence,

$$||\mathbf{d}_i|| = ||\hat{\boldsymbol{\lambda}}_i - \overline{\mathbf{C}}^{-}\boldsymbol{\lambda}_i|| = O_p(T^{-1/2}) + O_p(N^{-1}), \tag{A3}$$

as was to be shown. ■

Lemma A.2. *Under the conditions of Lemma A.1,*

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}(\hat{e}_{i,t}^p)^2 = \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}(e_{i,t}^p)^2 + O_p(N^{-1}) + O_p(T^{-1}).$$

Proof of Lemma A.2.

This proof is analogous to Proof of Lemma 1 in Bai and Ng (2010). Let us denote by $\mathbf{A}^{-}$ the Moore–Penrose inverse of the matrix A. Note in particular that if A has full row rank, then $\mathbf{A}^{-} = \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}$, whereas if $\mathbf{A}$ has column row rank, then $\mathbf{A}^{-} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$. Thus, since $\overline{\mathbf{C}}$ has full row rank, we have $\overline{\mathbf{C}}^{-} = \overline{\mathbf{C}}'(\overline{\mathbf{C}\mathbf{C}}')^{-1}$, such that $\overline{\mathbf{C}\mathbf{C}}^{-} = \mathbf{I}_r$. Making use of this result,

$$y_{i,t}^p = \boldsymbol{\lambda}_i'\mathbf{f}_t^p + (\Delta e_{i,t})^p = \boldsymbol{\lambda}_i'(\overline{\mathbf{C}}^{-})'\overline{\mathbf{C}}'\mathbf{f}_t^p + (\Delta e_{i,t})^p. \tag{A4}$$

Moreover,

$$y_{i,t}^p = \hat{\boldsymbol{\lambda}}_i'\hat{\mathbf{f}}_t^p + (\widehat{\Delta e}_{i,t})^p. \tag{A5}$$

Subtracting (A4) from (A5), we obtain

$$\begin{aligned}
(\widehat{\Delta e}_{i,t})^p &= (\Delta e_{i,t})^p + \boldsymbol{\lambda}_i'(\overline{\mathbf{C}}^{-})'\overline{\mathbf{C}}'\mathbf{f}_t^p - \hat{\mathbf{C}}_i'\hat{\mathbf{f}}_t^p \\
&= (\Delta e_{i,t})^p - \boldsymbol{\lambda}_i'(\overline{\mathbf{C}}^{-})'(\hat{\mathbf{f}}_t^p - \overline{\mathbf{C}}'\mathbf{f}_t^p) - (\hat{\boldsymbol{\lambda}}_i - \overline{\mathbf{C}}^{-}\boldsymbol{\lambda}_i)'\hat{\mathbf{f}}_t^p \\
&= (\Delta e_{i,t})^p - \boldsymbol{\lambda}_i'(\overline{\mathbf{C}}^{-})'\mathbf{s}_t - \mathbf{d}_i'\hat{\mathbf{f}}_t^p.
\end{aligned} \tag{A6}$$

Insertion into the definition of $\hat{e}_{i,t}^p$ now yields

$$\hat{e}_{i,t}^p = \sum_{n=2}^{t}\widehat{\Delta e}_{i,n}^p = e_{i,t}^p - \boldsymbol{\lambda}_i'(\overline{\mathbf{C}}^{-})'\mathbf{S}_t - \mathbf{d}_i'\hat{\mathbf{F}}_t^p = e_{i,t}^p + a_{i,t},$$

where $a_{i,t} = -\boldsymbol{\lambda}_i'(\overline{\mathbf{C}}^-)'\mathbf{S}_t - \mathbf{d}_i'\hat{\mathbf{F}}_t$. Consequently,

$$\begin{aligned}\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}(\hat{e}_{i,t}^p)^2 &= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}(e_{i,t}^p)^2 + \frac{2}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{i,t}^p a_{i,t} + \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}a_{i,t}^2 \\ &= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}(e_{i,t}^p)^2 + I + II,\end{aligned}$$

with implicit definitions of I and II. By Lemma A.1,

$$\begin{aligned}II &= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}a_{i,t}^2 = \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}[-\boldsymbol{\lambda}_i'(\overline{\mathbf{C}}^-)'\mathbf{S}_t - \mathbf{d}_i'\hat{\mathbf{F}}_t^p]^2 \\ &\leq \frac{2}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}[(\boldsymbol{\lambda}_i'(\overline{\mathbf{C}}^-)'\mathbf{S}_t)^2 + (\mathbf{d}_i'\hat{\mathbf{F}}_t)^2] \\ &\leq 2||\overline{\mathbf{C}}^-||^2\frac{1}{N}\sum_{i=1}^{N}||\boldsymbol{\lambda}_i||^2\frac{1}{T}\sum_{t=2}^{T}||T^{-1/2}\mathbf{S}_t||^2 + \frac{2}{N}\sum_{i=1}^{N}||\mathbf{d}_i||^2\frac{1}{T^2}\sum_{t=2}^{T}||\hat{\mathbf{F}}_t^p||^2 \\ &= O_p(N^{-1}) + O_p(T^{-1}). \end{aligned} \tag{A7}$$

Consider I;

$$I = \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{i,t}^p a_{i,t} = -\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{i,t}^p\boldsymbol{\lambda}_i'(\overline{\mathbf{C}}^-)'\mathbf{S}_t - \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{i,t}^p\mathbf{d}_i'\hat{\mathbf{F}}_t^p.$$

By the Cauchy–Schwarz inequality,

$$\begin{aligned}&\left|\left|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{i,t}^p\boldsymbol{\lambda}_i'(\overline{\mathbf{C}}^-)'\mathbf{S}_t\right|\right| \\ &\leq \frac{1}{\sqrt{N}T}\sum_{t=2}^{T}\left|\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\boldsymbol{\lambda}_i'e_{i,t}^p\right|\right|||\overline{\mathbf{C}}^-||||T^{-1/2}\mathbf{S}_t|| \\ &\leq N^{-1/2}||\overline{\mathbf{C}}^-||\left(\frac{1}{T}\sum_{t=2}^{T}\left|\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\boldsymbol{\lambda}_i'e_{i,t}^p\right|\right|^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=2}^{T}||T^{-1/2}\mathbf{S}_t||^2\right)^{1/2} \\ &= O_p(N^{-1}),\end{aligned}$$

and

$$\begin{aligned}\left|\left|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{i,t}^p\mathbf{d}_i'\hat{\mathbf{F}}_t^p\right|\right| &\leq \frac{1}{NT^2}\sum_{t=2}^{T}\left|\left|\sum_{i=1}^{N}\mathbf{d}_i'e_{i,t}^p\right|\right|||\hat{\mathbf{F}}_t^p|| \\ &\leq T^{-1/2}\left(\frac{1}{T}\sum_{t=2}^{T}\left|\left|\frac{1}{N}\sum_{i=1}^{N}\mathbf{d}_i'e_{i,t}^p\right|\right|^2\right)^{1/2}\left(\frac{1}{T^2}\sum_{t=2}^{T}||\hat{\mathbf{F}}_t^p||^2\right)^{1/2} \\ &= O_p(T^{-1}) + O_p(T^{-1/2}N^{-1}),\end{aligned}$$

where the last result holds, because

$$\left|\left|\frac{1}{N}\sum_{i=1}^{N}\mathbf{d}_i'e_{i,t}^{p}\right|\right| \leq \left(\frac{1}{N}\sum_{i=1}^{N}||\mathbf{d}_i||^2\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}||e_{i,t}^{p}||^2\right)^{1/2} = O_p(T^{-1/2}) + O_p(N^{-1}).$$

It follows that

$$I = O_p(T^{-1}) + O_p(N^{-1}), \tag{A8}$$

which in turn implies

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}(\hat{e}_{i,t}^{p})^2 = \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}(e_{i,t}^{p})^2 + O_p(N^{-1}) + O_p(T^{-1}), \tag{A9}$$

as was to be shown. ■

Lemma A.3. *Under the conditions of Lemma A.1,*

$$\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}[(\hat{e}_{i,1}^{p})^2 - (e_{i,1}^{p})^2] = O_p(\sqrt{N}T^{-1}), \tag{i}$$

$$\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}[(\hat{e}_{i,T}^{p})^2 - (e_{i,T}^{p})^2] = O_p(\sqrt{N}T^{-1}) + O_p(N^{-1/2}), \tag{ii}$$

$$\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T}[(\Delta\hat{e}_{i,t}^{p})^2 - (\Delta e_{i,t}^{p})^2] = O_p(\sqrt{N}T^{-1}) + O_p(T^{-1/2}) + O_p(N^{-1/2}). \tag{iii}$$

Proof of Lemma A.3.

Part (i) is obvious. Consider (ii). From $\hat{e}_{i,t}^{p} = e_{i,t}^{p} + a_{i,t}$,

$$\begin{aligned}\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}[(\hat{e}_{i,T}^{p})^2 - (e_{i,T}^{p})^2] &= \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}(2e_{i,T}^{p}a_{i,T} - a_{i,T}^2)\\ &= \frac{2}{\sqrt{N}T}\sum_{i=1}^{N}e_{i,T}^{p}a_{i,T} - \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}a_{i,T}^2.\end{aligned} \tag{A10}$$

Making use of the fact that $a_{i,t} = -\boldsymbol{\lambda}_i'(\overline{\mathbf{C}}^{-})'\mathbf{S}_t - \mathbf{d}_i'\hat{\mathbf{F}}_t$, we can show that

$$\begin{aligned}\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}a_{i,T}^2 &= \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}[-\boldsymbol{\lambda}_i'(\overline{\mathbf{C}}^{-})'\mathbf{V}_T - \mathbf{d}_i'\hat{\mathbf{F}}_T]^2\\ &\leq 2\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}||\boldsymbol{\lambda}_i||^2||\overline{\mathbf{C}}^{-}||^2||T^{-1/2}\mathbf{V}_T||^2 + \frac{1}{N}\sum_{i=1}^{N}||\mathbf{d}_i||^2||T^{-1/2}\hat{\mathbf{F}}_T||^2\right)\\ &= \sqrt{N}[O_p(T^{-1}) + O_p(N^{-1})] = O_p(\sqrt{N}T^{-1}) + O_p(N^{-1/2}),\end{aligned}$$

and

$$
\begin{aligned}
&\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}e_{i,T}^{p}a_{i,T}\right| \\
&\leq \left|\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}e_{i,T}^{p}\boldsymbol{\lambda}_i'\right|\right| ||\overline{\mathbf{C}}^{-}|| ||T^{-1/2}\mathbf{V}_T|| + \sqrt{N}T^{-1/2}\left|\left|\frac{1}{N}\sum_{i=1}^{N}e_{i,T}^{p}\mathbf{d}_i'\right|\right| ||T^{-1/2}\hat{\mathbf{F}}_T|| \\
&= O_p(N^{-1/2}) + \sqrt{N}T^{-1/2}[O_p(T^{-1/2}) + O_p(N^{-1})] = O_p(N^{-1/2}) + O_p(\sqrt{N}T^{-1}).
\end{aligned}
$$

The result in (ii) is implied by this.

For (iii), note that $\Delta\hat{e}_{i,t}^{p} = \Delta e_{i,t}^{p} + \Delta a_{i,t}$, where $\Delta a_{i,t} = -\boldsymbol{\lambda}_i'(\overline{\mathbf{C}}^{-})'\mathbf{s}_t - \mathbf{d}_i'\hat{\mathbf{f}}_t$. Therefore, $(\Delta\hat{e}_{i,t}^{p})^2 = (\Delta e_{i,t}^{p})^2 + 2\Delta e_{i,t}^{p}\Delta a_{i,t} + (\Delta a_{i,t})^2$. By using this result, we obtain

$$
\begin{aligned}
\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}[(\Delta\hat{e}_{i,t}^{p})^2 - (\Delta e_{i,t}^{p})^2] &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}[2\Delta e_{i,t}^{p}\Delta a_{i,t} + (\Delta a_{i,t})^2] \\
&= \frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta e_{i,t}^{p}\Delta a_{i,t} + \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}(\Delta a_{i,t})^2 \\
&= I + II, \qquad \text{(A11)}
\end{aligned}
$$

with implicit definitions of I and II. The order of II can be obtained in the following fashion:

$$
\begin{aligned}
II &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}[-\boldsymbol{\lambda}_i'(\overline{\mathbf{C}}^{-})'\mathbf{s}_t - \mathbf{d}_i'\hat{\mathbf{f}}_t]^2 \leq \frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}(||\boldsymbol{\lambda}_i'(\overline{\mathbf{C}}^{-})'\mathbf{s}_t||^2 + ||\mathbf{d}_i'\hat{\mathbf{f}}_t||^2) \\
&\leq 2\sqrt{N}\left(||\overline{\mathbf{C}}^{-}||^2\frac{1}{N}\sum_{i=1}^{N}||\boldsymbol{\lambda}_i||^2\frac{1}{T}\sum_{t=2}^{T}||\mathbf{s}_t||^2 + \frac{1}{N}\sum_{i=1}^{N}||\mathbf{d}_i||^2\frac{1}{T}\sum_{t=2}^{T}||\hat{\mathbf{f}}_t||^2\right) \\
&= \sqrt{N}[O_p(N^{-1}) + O_p(T^{-1})] = O_p(N^{-1/2}) + O_p(\sqrt{N}T^{-1})
\end{aligned}
$$

For I,

$$
I = -\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta e_{i,t}^{p}\boldsymbol{\lambda}_i'(\overline{\mathbf{C}}^{-})'\mathbf{s}_t - \frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta e_{i,t}^{p}\mathbf{d}_i'\hat{\mathbf{f}}_t.
$$

By using the fact that $\mathbf{s}_t = \overline{\mathbf{v}}_t^{p}$, we obtain

$$
\begin{aligned}
&\left|\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\Delta e_{i,t}^{p}\boldsymbol{\lambda}_i'(\overline{\mathbf{C}}^{-})'\mathbf{s}_t\right|\right| \\
&\leq \frac{1}{T}\sum_{t=1}^{T}\left|\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Delta e_{i,t}^{p}\boldsymbol{\lambda}_i\right|\right| \left|\left|\overline{\mathbf{C}}^{-}\right|\right| ||\overline{\mathbf{v}}_t^{p}|| \\
&\leq N^{-1/2}\left|\left|\overline{\mathbf{C}}^{-}\right|\right|\left(\frac{1}{T}\sum_{t=1}^{T}\left|\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Delta e_{i,t}^{p}\boldsymbol{\lambda}_i\right|\right|^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}\left|\left|\sqrt{N}\overline{\mathbf{v}}_t^{p}\right|\right|^2\right)^{1/2} \\
&= O_p(N^{-1/2}),
\end{aligned}
$$

and

$$
\begin{aligned}
&\left|\left|\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta e_{i,t}^{p}\mathbf{d}_i'\hat{\mathbf{f}}_t^{p}\right|\right| \\
&\leq \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}||\mathbf{d}_i||^2\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}\left|\left|\frac{1}{T}\sum_{t=2}^{T}\hat{\mathbf{f}}_t^{p}\Delta e_{i,t}^{p}\right|\right|^2\right)^{1/2} \\
&= \sqrt{N}[O_p(T^{-1/2})+O_p(N^{-1})][O_p(T^{-1/2})+O_p(N^{-1/2})] \\
&= O_p(\sqrt{N}T^{-1})+O_p(T^{-1/2})+O_p(N^{-1}),
\end{aligned}
$$

where the last result follows form

$$
\begin{aligned}
\left|\left|\frac{1}{T}\sum_{t=2}^{T}\hat{\mathbf{f}}_t^{p}\Delta e_{i,t}^{p}\right|\right| &\leq \left|\left|\frac{1}{\sqrt{T}}\sum_{t=2}^{T}T^{-1/2}\overline{\mathbf{C}}'\mathbf{f}_t^{p}\Delta e_{i,t}^{p}\right|\right| + \left|\left|\frac{1}{T}\sum_{t=2}^{T}\mathbf{s}_t\Delta e_{i,t}^{p}\right|\right| \\
&\leq T^{-1/2}||\overline{\mathbf{C}}||\left|\left|\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\mathbf{f}_t^{p}\Delta e_{i,t}^{p}\right|\right| + \left(\frac{1}{T}\sum_{t=2}^{T}||\mathbf{s}_t||^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=2}^{T}(\Delta e_{i,t}^{p})^2\right)^{1/2} \\
&= O_p(T^{-1/2})+O_p(N^{-1/2}).
\end{aligned}
$$

Hence,

$$
I = O_p(\sqrt{N}T^{-1})+O_p(T^{-1/2})+O_p(N^{-1/2}),
$$

and so we obtain

$$
\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T}[(\Delta\hat{e}_{i,t}^{p})^2-(\Delta e_{i,t}^{p})^2] = O_p(\sqrt{N}T^{-1})+O_p(T^{-1/2})+O_p(N^{-1/2}). \tag{A12}
$$

This establishes (iii) and hence the proof of the lemma is complete. ■

Lemma A.4. *Under the condition of Lemma A.1,*

$$
\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T}\hat{e}_{i,t-1}^{p}\Delta\hat{e}_{i,t}^{p} = \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{i,t-1}^{p}(\Delta e_{i,t})^{p}+O_p(\sqrt{N}T^{-1})+O_p(T^{-1/2})+O_p(N^{-1/2}).
$$

Proof of Lemma A.4.

This proof follows from the same steps used in the proof of Lemma 2 in Bai and Ng (2010). We begin by noting that $(\hat{e}_{i,t}^{p})^2 = (\hat{e}_{i,t-1}^{p}+\Delta\hat{e}_{i,t}^{p})^2 = (\hat{e}_{i,t-1}^{p})^2+2\hat{e}_{i,t-1}^{p}\Delta\hat{e}_{i,t}^{p}+(\Delta\hat{e}_{i,t}^{p})^2$, which implies

$$
\frac{1}{T}\sum_{t=2}^{T}\hat{e}_{i,t-1}^{p}\Delta\hat{e}_{i,t}^{p} = \frac{1}{2T}(\hat{e}_{i,T}^{p})^2-\frac{1}{2T}(\hat{e}_{i,1}^{p})^2-\frac{1}{2T}\sum_{t=2}^{T}(\Delta\hat{e}_{i,t}^{p})^2.
$$

A similar result applies to $T^{-1}\sum_{t=2}^{T} e_{i,t-1}^{p}\Delta e_{i,t}^{p}$. Hence,

$$
\begin{aligned}
&\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T}(\hat{e}_{i,t-1}^{p}\Delta\hat{e}_{i,t}^{p} - e_{i,t-1}^{p}\Delta e_{i,t}^{p}) \\
&\leq \frac{1}{2\sqrt{N}T}\sum_{i=1}^{N}[(\hat{e}_{i,T}^{p})^2 - (e_{i,T}^{p})^2] + \frac{1}{2\sqrt{N}T}\sum_{i=1}^{N}[(\hat{e}_{i,1}^{p})^2 - (e_{i,1}^{p})^2] \\
&+ \frac{1}{2\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T}[(\Delta\hat{e}_{i,t}^{p})^2 - (\Delta e_{i,t}^{p})^2] \\
&= O_p(\sqrt{N}T^{-1}) + O_p(T^{-1/2}) + O_p(N^{-1/2}),
\end{aligned} \tag{A13}
$$

as required. ■

Proof of Theorem 1.

When $p = 0$ ($p = 1$) Lemmas A.2 and A.4 correspond to Lemmas 1 and 2 (Lemma 4) in Bai and Ng (2010). The proof of Theorem 1 therefore follows from using the same steps as in the proofs of Theorems 1 and 2 in this other paper. ■

Proof of Theorem 2.

As mentioned in Section 3.2, the rate of consistency of $\hat{\mathbf{F}}_t^p$ is faster than in the case of PC estimation. In view of this, the proof of Theorem 2 when $p = 0$ ($p = 1$) follows directly from the proof of Theorem 1 (3) in Bai and Ng (2004). ■

Table 1: 5% size and size corrected power for intercept-only case ($p = 0$).

N	T	$P_{a,0}^{PC}$	$P_{a,0}$	$P_{b,0}^{PC}$	$P_{b,0}$	$PMSB_0^{PC}$	$PMSB_0$
				Size			
10	10	0.2252	0.2088	0.1304	0.1162	0.0094	0.0094
10	20	0.1756	0.1478	0.098	0.0734	0.0138	0.0064
10	35	0.1834	0.1574	0.1098	0.088	0.0226	0.0138
10	50	0.1776	0.1504	0.1074	0.085	0.0248	0.0136
20	10	0.213	0.2104	0.134	0.131	0.0696	0.0732
20	20	0.145	0.1452	0.0888	0.0842	0.0392	0.0372
20	35	0.123	0.1174	0.0782	0.0718	0.0284	0.0308
20	50	0.1234	0.1116	0.0746	0.0664	0.0316	0.027
35	10	0.237	0.2314	0.1726	0.1676	0.1382	0.1468
35	20	0.1298	0.1392	0.0844	0.0892	0.06	0.0654
35	35	0.1086	0.1014	0.072	0.0664	0.0408	0.0422
35	50	0.1044	0.104	0.0706	0.0698	0.0392	0.0378
50	10	0.2526	0.2564	0.193	0.1988	0.193	0.1976
50	20	0.1362	0.1422	0.095	0.0994	0.0874	0.0838
50	35	0.0992	0.1008	0.065	0.0666	0.0526	0.0528
50	50	0.0976	0.0994	0.0692	0.0698	0.051	0.051
				Size-corrected power			
10	10	0.0548	0.0708	0.0542	0.0744	0.057	0.0642
10	20	0.1064	0.1572	0.1094	0.158	0.0994	0.138
10	35	0.2296	0.3246	0.2384	0.3298	0.2	0.2738
10	50	0.395	0.5482	0.401	0.5496	0.3434	0.4588
20	10	0.04	0.0926	0.0402	0.095	0.0436	0.0836
20	20	0.1864	0.264	0.1874	0.2684	0.1544	0.215
20	35	0.5046	0.6094	0.5142	0.6116	0.416	0.5014
20	50	0.7996	0.8804	0.797	0.8782	0.6894	0.7794
35	10	0.0282	0.0972	0.0278	0.0976	0.0304	0.0886
35	20	0.2832	0.455	0.2872	0.4536	0.2278	0.3622
35	35	0.778	0.8638	0.7798	0.8632	0.668	0.7574
35	50	0.9778	0.9906	0.9772	0.9888	0.9334	0.9604
50	10	0.0206	0.1172	0.0214	0.1164	0.0256	0.1014
50	20	0.3658	0.5666	0.371	0.5688	0.2906	0.4624
50	35	0.924	0.9642	0.9236	0.962	0.8378	0.9094
50	50	0.9984	0.9998	0.9984	0.9996	0.988	0.9924

Notes: A "PC" superscript signifies that the test is based on original PANIC.

Table 2: 5% size and size corrected power for trend case ($p = 1$).

N	T	$P_{a,0}^{PC}$	$P_{a,0}$	$P_{b,0}^{PC}$	$P_{b,0}$	$PMSB_0^{PC}$	$PMSB_0$
				Size			
10	10	0.1868	0.1908	0.1692	0.166	0	0
10	20	0.1716	0.151	0.1438	0.1214	0.0082	0.0048
10	35	0.1638	0.139	0.1338	0.1118	0.019	0.0088
10	50	0.1646	0.1422	0.1394	0.1146	0.0236	0.0168
20	10	0.2264	0.2274	0.2168	0.2198	0.017	0.0198
20	20	0.1498	0.1398	0.1342	0.1278	0.0202	0.0186
20	35	0.1308	0.1192	0.1152	0.1048	0.0232	0.0224
20	50	0.1226	0.1158	0.1082	0.0998	0.029	0.026
35	10	0.2668	0.2694	0.2656	0.271	0.065	0.0674
35	20	0.1696	0.1632	0.1654	0.159	0.0356	0.0396
35	35	0.1218	0.1198	0.1184	0.1162	0.0342	0.0352
35	50	0.1096	0.1186	0.1054	0.113	0.0362	0.0356
50	10	0.2978	0.301	0.3034	0.306	0.1072	0.1078
50	20	0.19	0.1896	0.192	0.1908	0.0582	0.0524
50	35	0.13	0.1296	0.13	0.1302	0.0398	0.0444
50	50	0.1068	0.1094	0.1066	0.1094	0.0392	0.0388
				Size-corrected power			
10	10	0.0448	0.0494	0.0468	0.0494	0.0442	0.0468
10	20	0.0554	0.0558	0.0554	0.0568	0.0546	0.058
10	35	0.0744	0.0902	0.0748	0.0902	0.0736	0.0886
10	50	0.1158	0.1302	0.1156	0.1302	0.1162	0.1316
20	10	0.0464	0.0474	0.0472	0.0474	0.0456	0.0454
20	20	0.0578	0.065	0.058	0.0662	0.057	0.0664
20	35	0.0986	0.1112	0.0982	0.1122	0.0982	0.1108
20	50	0.1766	0.2076	0.1756	0.2074	0.1746	0.207
35	10	0.0466	0.0506	0.046	0.0512	0.047	0.0544
35	20	0.0556	0.0756	0.0554	0.0756	0.0566	0.0736
35	35	0.1236	0.1506	0.1228	0.1494	0.1228	0.1474
35	50	0.248	0.2976	0.2484	0.297	0.2478	0.2942
50	10	0.0434	0.0418	0.0438	0.042	0.045	0.043
50	20	0.0622	0.0798	0.063	0.0802	0.0614	0.0762
50	35	0.1526	0.197	0.1536	0.198	0.1508	0.194
50	50	0.331	0.39	0.3314	0.3892	0.3308	0.3882

Notes: See Table 1 for an explanation.

Table 3: Correct selection frequency of the estimated number of unit root factors.

				$p = 0$		$p = 1$	
r	r_1	ρ	δ_0	$\hat{r}_1$	$\hat{r}_1^{PC}$	$\hat{r}_1$	$\hat{r}_1^{PC}$
3	3	0	-	0.9576	0	0.9798	0.0002
3	3	0.5	-	0.977	0.2502	0.991	0.6368
3	3	0.8	-	0.9764	0.9266	0.9896	0.9794
3	3	0.9	-	0.974	0.9788	0.9892	0.9898
3	3	1	-	0.9762	0.9892	0.9916	0.9948
3	2	0	0	0.77	0	0.5922	0.0098
3	2	0.5	0	0.77	0.0188	0.603	0.2532
3	2	0.8	0	0.7792	0.753	0.6074	0.8276
3	2	0.9	0	0.779	0.9176	0.5946	0.8532
3	2	1	0	0.7284	0.9018	0.5724	0.8162
3	2	0	0.5	0.452	0.1274	0.1676	0.481
3	2	0.5	0.5	0.3634	0.238	0.1234	0.3492
3	2	0.8	0.5	0.361	0.3066	0.1084	0.12
3	2	0.9	0.5	0.3738	0.315	0.1168	0.1066
3	2	1	0.5	0.2326	0.2162	0.085	0.0734
3	1	0	0	0.6988	0.6384	0.5042	0.5
3	1	0.5	0	0.707	0.6586	0.5026	0.401
3	1	0.8	0	0.712	0.244	0.5174	0.1986
3	1	0.9	0	0.7128	0.2	0.5156	0.2056
3	1	1	0	0.7014	0.204	0.5124	0.1738
3	1	0	0.5	0.5522	0.7056	0.169	0.447
3	1	0.5	0.5	0.4528	0.5442	0.1182	0.165
3	1	0.8	0.5	0.4502	0.0824	0.1158	0.0092
3	1	0.9	0.5	0.44	0.035	0.1194	0.0068
3	1	1	0.5	0.2334	0.0158	0.0766	0.002
3	0	0	0	1	1	0.989	1
3	0	0.5	0	0.9998	1	0.9888	0.984
3	0	0.8	0	1	0.7762	0.9856	0.3128
3	0	0.9	0	1	0.2952	0.988	0.0994
3	0	1	0	0.9744	0.0382	0.9778	0.024
3	0	0	0.5	0.8622	0.9726	0.4022	0.656
3	0	0.5	0.5	0.777	0.8654	0.3172	0.2818
3	0	0.8	0.5	0.7356	0.3024	0.2864	0.0224
3	0	0.9	0.5	0.732	0.1006	0.29	0.0048
3	0	1	0.5	0.4054	0.0042	0.184	0.0008

Notes: r, r_1, ρ and δ_0 refer to the true number of factors, the number of unit root factors, the autoregressive coefficient of the idiosyncratic component, and the autoregressive coefficient of stationary factors, respectively. The "PC" superscript signifies that the estimated number of unit root factors is based on original PANIC.

Table 4: Unit root and cointegration test results.

Common component		
k	$MQ_0(k)$	Reject?
2	-60.9015	yes
1	-2.2407	no
0	-	-
Idiosyncratic component		
Unit root tests		
Test	$F_{i,t}$	$S_{i,t}$
$P_{a,0}$	-3.1485	-3.0226
	(0.0016)	(0.0025)
$P_{b,0}$	-1.7501	-1.7107
	(0.0801)	(0.0871)
$PMSB_0$	-0.8461	-0.8321
	(0.3975)	(0.4054)
Pedroni cointegration tests		
Test	Value	p-value
Panel-t	-3.2713	0.0005
Group-t	-3.3109	0.0005

Table 5: Estimation results of the cointegrating slope.

Common component				
	MFLS	p-value	DLS	p-value
Slope	1.0013	0.4853	0.9989	0.5126
Idiosyncratic component				
Group mean estimation				
	MFLS	p-value	DLS	p-value
Slope	0.9363	0.0246	0.9354	0.5135
Panel estimation				
	MFLS	p-value	DLS	p-value
Slope	1.0221	0.0000	1.0235	0.0000

Notes: "FMLS" and "DLS" refer to the fully modified LS and dynamic LS estimator, respectively. While the group mean estimator is based on between pooling, the panel estimator is based on within pooling. The reported p-values test if the slope is equal to zero.