# The Performance of Panel Cointegration Methods: Results from a Large Scale Simulation Study

Martin Wagner
Jaroslava Hlouskova

IRIHS INSTITUTIONAL REPOSITORY AT IHS
Vienna

INSTITUT FÜR HÖHERE STUDIEN
INSTITUTE FOR ADVANCED STUDIES
Vienna

# The Performance of Panel Cointegration Methods:
## Results from a Large Scale Simulation Study

**Martin Wagner, Jaroslava Hlouskova**

# The Performance of Panel Cointegration Methods:

Results from a Large Scale Simulation Study

**Martin Wagner, Jaroslava Hlouskova**

**May 2007**

**Contact:**

Martin Wagner
Department of Economics and Finance
Institute for Advanced Studies
Stumpergasse 56
1060 Vienna, Austria
☎: +43/1/599 91-150
email: Martin.Wagner@ihs.ac.at

Jaroslava Hlouskova
Department of Economics and Finance
Institute for Advanced Studies
Stumpergasse 56
1060 Vienna, Austria
☎: +43/1/599 91-142
email: hlouskov@ihs.ac.at

# Abstract

This paper presents results concerning the performance of both single equation and system panel cointegration tests and estimators. The study considers the tests developed in Pedroni (1999, 2004), Westerlund (2005), Larsson, Lyhagen, and Löthgren (2001) and Breitung (2005); and the estimators developed in Phillips and Moon (1999), Pedroni (2000), Kao and Chiang (2000), Mark and Sul (2003), Pedroni (2001) and Breitung (2005). We study the impact of stable autoregressive roots approaching the unit circle, of I(2) components, of short-run cross-sectional correlation and of cross-unit cointegration on the performance of the tests and estimators. The data are simulated from three-dimensional individual specific VAR systems with cointegrating ranks varying from zero to two for fourteen different panel dimensions. The usual specifications of deterministic components are considered.

# Contents

# 1  Introduction

This companion paper to Hlouskova and Wagner (2006), where panel unit root and stationarity tests have been studied, investigates the properties of panel cointegration tests and estimators by means of a large scale simulation study. Our study includes both single equation and system (to be precise vector autoregression, in short VAR) tests and estimators.

The single equation tests (of the null hypothesis of no cointegration) of Pedroni (1999, 2004) and of Westerlund (2005) and the systems tests developed in Larsson, Lyhagen, and Löthgren (2001) and Breitung (2005) are analyzed. We do not consider single equation tests of the null hypothesis of cointegration, as such tests are bound to perform as poorly as their panel stationarity counterparts. For the example of the McCoskey and Kao (1998) test, its panel stationarity test counterpart developed in Hadri (2000) has been found to exhibit devastating performance in Hlouskova and Wagner (2006). Additionally performed simulations also confirm this expectation perfectly.

We have implemented several versions (i.e. with average or individual specific correction factors, normalized, group-mean, see the description in Section 2.1.2) of both the fully modified (FM) and dynamic (D) OLS estimators, as developed in Phillips and Moon (1999) and Pedroni (2000) for FM-OLS and in Kao and Chiang (2000), Mark and Sul (2003) and Pedroni (2001) for D-OLS estimation. As system estimators we include only the two-step panel VAR estimator of the cointegrating space developed in Breitung (2005) and the one-step or group-mean VAR estimator given by the cross-sectional average of appropriately normalized individual specific Johansen estimates of the cointegrating spaces. This latter estimator is included because it is the system counterpart to the group-mean single equation estimators and is also one potential starting value for iterative system estimators. Note here that the two-step estimator of Breitung is not an iterative estimator for the cointegrating space because in its second step only one estimator of the cointegrating space is computed. We abstain from including truly iterative estimators (like Larsson and Lyhagen, 1999; Groen and Kleibergen, 2003) for the following reasons. First, we want to compare 'similar' estimators in terms of (computational) complexity, i.e. we only want to compare simple regression based estimators. Second, the proposed iterative estimators, like the Groen and Kleibergen (2003) estimator, are (in their general version) more demanding in terms of the time dimension of the panel due to the unrestricted set-up of the panel VAR model, see the discussion in Sec-

tion 2.2. Third, more pragmatically, iterative estimators increase the required computer time substantially, which is particularly unpleasant for large scale simulation experiments. For this latter reason we also abstain from performing even one more iterative step in Breitung's two-step estimator.

All described tests and estimators are derived for cross-sectionally independent panels. This for many applications unrealistic assumption is still commonly employed when developing panel cointegration methods, in particular for estimation procedures. Only few and partial results concerning both cointegration testing and estimation are available for cross-sectionally dependent panels to date. Panel cointegration tests that allow for some form of cross-sectional dependence via common factors include Banerjee and Carrion-i-Silvestre (2006) and Westerlund and Edgerton (2006). The results with respect to estimation are even more scarce and include, with a different focus, (approximate) factor models as developed in Bai and Ng (2004), an extension of FM-OLS estimation to panels with short-run cross-sectional correlation developed in Bai and Kao (2005), or Kapetanios, Pesaran, and Yamagata (2006) who consider the properties of so called common correlated effects (CCE) estimators when allowing for nonstationary common factors. Additionally, several authors consider spatial or 'economic distance' formulations to allow for cross-sectional dependence, see e.g. Pesaran, Schuermann, and Weiner (2004).

All in all, however, the panel cointegration literature appears to be relatively nascent and partly ad-hoc with regard to cross-sectional dependence and in particular there does not seem to exist a consensus yet about successful modelling strategies for cross-sectional dependence. Note in this respect that the present paper appears to be the first one to provide a formal definition of cross-unit cointegration, see Definition 1 in Section 3. Therefore we abstain from including methods designed for some form of cross-sectional dependence in our simulation study and focus only on some widely-used tests and estimators designed for cross-sectionally independent panels.

The data generating processes (DGPs) in the simulations are given by individual specific three-dimensional VAR(2) processes with cointegrating ranks ranging from zero to two. Only the cointegrating spaces are restricted to be identical for all cross-section members. We are in particular interested in the following aspects. First, we investigate the performance of the tests and estimators depending upon the time series and cross-section dimensions. The time series dimension assumes the values $T \in \{10, 25, 50, 100\}$ and the cross-section dimension assumes

the values $N \in \{5, 10, 25, 50, 100\}$. We only consider those combinations where $T \geq N$, which results in fourteen different panel dimensions. The restriction is put in place for two reasons: (i) For the panel VAR system methods clearly a 'relatively' large time series dimension is required to mitigate the substantial small sample biases of autoregressive estimation. Taking the time series dimension at least as large as the cross-sectional dimension serves as a simple heuristic lower bound. (ii) Preliminary simulations, available upon request, highlight that a cross-sectional dimension that is too large compared to the time series dimension leads to size divergence, i.e. the actual size tends to one for increasing $N$ and fixed $T$ smaller than $N$. Similar findings have been obtained for panel unit root tests in Hlouskova and Wagner (2006). Second, we investigate the impact of stable autoregressive roots approaching the unit circle on the performance of the tests and estimators. Third, we assess the effects of an I(2) component. Fourth, we study the impact of (three different forms of) short-run cross-sectional correlation on the methods' performance. Fifth, we consider how the methods are affected by the presence of one cross-unit cointegrating relationship that is introduced in addition to the identical within-unit cointegrating relationships. Sixth, we consider the usual variety of specifications of the deterministic components.

Because we compare in part of the analysis single equation tests (where only one test is performed) with system tests (where a test sequence is performed) we use as a commonly applicable performance measure the hit rates, defined as the acceptance frequencies of the correct dimension of the cointegrating space. For the single equation tests we consider in addition the power against stationary alternatives. The performance measure for the estimators is given by the gap (see (40) in Section 3.2) between the true and the estimated cointegrating spaces.

The paper is organized as follows: Section 2 describes the implemented panel cointegration tests and estimators. Section 3 presents the simulation set-up, provides a discussion of different forms of cross-sectional dependence, gives a definition of cross-unit cointegration and discusses the simulation results. Section 4 draws some conclusions. An appendix containing additional figures and tables follows the main text.

## 2   The Panel Cointegration Methods

In this section we describe the implemented single equation and system panel cointegration tests and estimators. We include a relatively detailed discussion for two reasons. First, the

3

detailed description allows the reader to see the differences and similarities across methods and tests in one place. Second, the description is intended to be detailed enough to allow the reader to implement the methods herself. In Section 2.1 the single equation methods are described and in Section 2.2 the system methods are described.

All the methods considered are so called *first generation* methods, as they are all formulated for cross-sectionally independent panels. The cross-sectional independence assumption allows for relatively straightforward asymptotic results using sequential limit theory, employed for all methods described, with first $T \to \infty$ followed by $N \to \infty$. For the derivation of the test statistics the main tools applied to obtain asymptotic normality are for the pooled tests the Delta method and for the group-mean tests (where individual specific statistics are averaged) standard central limit theorems.

## 2.1 Single Equation Methods

The single equation methods are panel extensions of the Engle and Granger (1987) approach to cointegration analysis. The DGP is in its most general form given by

$$y_{it} = \alpha_i + \delta_i t + X_{it}'\beta_i + u_{it} \tag{1}$$

$$X_{it} = X_{it-1} + \varepsilon_{it}, \tag{2}$$

observed for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. Here $y_{it}, u_{it} \in \mathbb{R}$, $X_{it}, \varepsilon_{it} \in \mathbb{R}^l$, $\alpha_i, \delta_i \in \mathbb{R}$ and $\beta_i \in \mathbb{R}^l$.[1] To simplify we use in slight abuse of notation the same notation for random variables (e.g. $u_{it}$) and the corresponding stochastic processes (which should correctly be written as e.g. $\{u_{it}\}_{t \in \mathbb{Z}}$).

Under the assumptions listed below cointegration is equivalent to stationarity of the processes $u_{it}$. The single cointegrating vector is given by $[1, -\beta_i']'$. When cointegration prevails we assume that the stacked processes $v_{it} = [u_{it}, \varepsilon_{it}']' \in \mathbb{R}^{l+1}$ are cross-sectionally independent stationary ARMA processes. The ARMA assumption is stronger than required for the applicability of the functional limit theorems underlying the asymptotic analysis of the described methods. In particular the ARMA assumption guarantees the existence of finite long-run covariance matrices $\Omega_i = \begin{bmatrix} \omega_{ui}^2 & \Omega_{u\varepsilon i} \\ \Omega_{u\varepsilon i}' & \Omega_{\varepsilon i} \end{bmatrix}$. The matrices $\Omega_{\varepsilon i}$ are assumed to have full

---

[1]In the panel literature sometimes also time effects are included. We abstain from including them in both the description and the simulation study, as they are usually extracted in the first step (similarly to the fixed effects) and the analysis is then performed on the adjusted data. Note, however, that in general the presence of time effects may change some of the asymptotic results.

rank, which excludes cointegration amongst the regressors $X_{it}$. Note that in case of ARMA processes the long-run covariance matrices $\Omega_i$ are given by $a_i(1)^{-1}b_i(1)\Sigma_{\xi i}b_i(1)'(a_i(1)^{-1})'$, where $a_i(z)v_{it} = b_i(z)\xi_{it}$ is an ARMA representation of $v_{it}$ with $\det a_i(z) \neq 0$ for $|z| \leq 1$, $\det b_i(z) \neq 0$ for $|z| \leq 1$, $a_i(z)$ and $b_i(z)$ are left co-prime and $\Sigma_{\xi i} > 0$ denotes the covariance matrix of the white noise process $\xi_{it}$. For later reference we furthermore define also the conditional long-run variance $\omega_{u.\varepsilon i}^2 = \omega_{ui}^2 - \Omega_{u\varepsilon i}\Omega_{\varepsilon i}^{-1}\Omega_{u\varepsilon i}'$ and the one-sided long-run variance matrix $\Lambda_i = \sum_{j=0}^{\infty}\mathbb{E}v_{it}v_{it-j}'$, which is partitioned according to the partitioning of $\Omega_i$.

In case of no cointegration, the processes $u_{it}$ are integrated of order one and the assumptions discussed above apply analogously to $v_{it} = [\Delta u_{it}, \varepsilon_{it}']'$. Because it will be clear throughout whether the cointegration or no cointegration case is considered, using the same notation for both cases should not lead to confusion.

We consider the usual three cases for the deterministic variables: Case 1 without deterministic components, case 2 with only fixed effects $\alpha_i$ and case 3 with both intercepts $\alpha_i$ and individual specific linear time trends $\delta_i t$. The methods discussed below all allow the short-run dynamics to differ across the members of the panels. In case of cointegration, however, the usual assumption is that of a homogeneous cointegrating relation, i.e. $\beta_i = \beta$ for $i = 1, \ldots, N$ in (1). A major limitation of the single equation methods is (as in the time series case) the restriction to one cointegrating relationship.

### 2.1.1 Tests for the null hypothesis of no cointegration

Under the null hypothesis of no cointegration (1) is a spurious regression equation. Nevertheless, the cross-sectional dimension allows for meaningful estimation of the so called *long-run average regression coefficient* for increasing cross-sectional dimension $N$, see Phillips and Moon (1999). Their paper establishes many of the required asymptotic results and also includes a detailed discussion concerning joint limits (with $N, T \to \infty$ jointly) versus sequential limits (used in the other papers discussed here) with first $T \to \infty$ followed by $N \to \infty$.

**Pedroni.** Pedroni (1999, 2004) develops in total seven different tests for the null hypothesis of no cointegration. Under the stated assumptions the processes $u_{it}$ can be written as

$$u_{it} = \rho_i u_{it-1} + \eta_{it}, \tag{3}$$

5

where the processes $\eta_{it}$ are stationary ARMA processes. The null hypothesis of the tests is given by $H_0 : \rho_i = 1$ for $i = 1, \ldots, N$. The *pooled* tests are specified against the homogeneous alternative $H_1^1 : -1 < \rho_i = \rho < 1$ for $i = 1, \ldots, N$, i.e. these tests are shown to be consistent against the homogeneous alternative which restricts the first-order serial correlation coefficient $\rho$ of the processes $u_{it}$ to be identical for $i = 1, \ldots, N$. The *group-mean* tests, based on cross-section averages of individual estimators of $\rho_i$, are specified (i.e. consistent) against the heterogeneous alternative $H_1^2 : -1 < \rho_i < 1$ for $i = 1, \ldots, N_1$ and $\rho_i = 1$ for $i = N_1 + 1, \ldots, N$. For consistency of the group-mean tests, a non-vanishing fraction of the individual units has to be stationary under the alternative, i.e. $\lim_{N \to \infty} N_1/N > 0$.

Of course, the tests are computed with estimated $\hat{u}_{it}$ in place of the unobserved errors $u_{it}$. In particular OLS residuals of (1) can be chosen, see Phillips and Ouliaris (1990) in the time series case, and a similar endogeneity correction factor ($\omega_{u.\varepsilon i}^2$) appears, see Phillips and Moon (1999) for a discussion of the properties of the OLS estimator of $\beta_i$. The estimator $\hat{\omega}_{u.\varepsilon i}^2$ is given by the estimator of the long-run variance of the residuals, $\hat{\eta}_{it}$ say, of an OLS regression of $\Delta y_{it}$ on the differenced deterministic components and $\Delta X_{it}$. The estimate $\hat{\omega}_{u.\varepsilon i}^2$ can be obtained by using a kernel estimator, see Andrews (1991) or Newey and West (1987) or alternatively by fitting an ARMA or AR model to $\hat{\eta}_{it}$ (and computing the long-run variance model based).

The correction for serial correlation can be handled either non-parametrically (following Phillips and Perron, 1988) or by using ADF type regressions. Let us start with the non-parametric tests. Denote the residuals of the OLS regressions $\hat{u}_{it} = \rho_i \hat{u}_{it-1} + \mu_{it}$ by $\hat{\mu}_{it}$. Further, denote their estimated variances by $\hat{\sigma}_{\mu i}^2$ and their estimated long-run variances by $\hat{\omega}_{\mu i}^2$. Then, the serial correlation factors are given by $\hat{\lambda}_i = \frac{1}{2}(\hat{\omega}_{\mu i}^2 - \hat{\sigma}_{\mu i}^2)$. For later use we also define $\hat{\omega}_{NT}^2 = \frac{1}{N} \sum_{i=1}^{N} \hat{\omega}_{\mu i}^2 / \hat{\omega}_{u.\varepsilon i}^2$.

With the defined quantities the following pooled test statistics can be computed: the variance ratio statistic $PP_\sigma$, the test based on the autoregressive coefficient $PP_\rho$, and the test based on the $t$-value of the autoregressive coefficient $PP_t$.[2] We write the test statistic below including scaling factors that display the convergence properties of the different components of the test statistics. This allows for a simple understanding of the construction principles, whereas in an actual computation of the test statistics these scaling factors (partly) drop out.

---

[2] $PP$ is used here as acronym for Pedroni pooled test. Below we use $PG$ as acronym for Pedroni group-mean test.

The *essential* parts (i.e. without centering and scaling factors, see below) of the pooled test statistics are given by

$$PP_\sigma^o = N^{1/2} \left( N^{-1} \sum_{i=1}^{N} \hat{\omega}_{u.\varepsilon i}^{-2} \left( T^{-2} \sum_{t=2}^{T} \hat{u}_{it-1}^2 \right) \right)^{-1} \tag{4}$$

$$PP_\rho^o = N^{1/2} \frac{N^{-1} \sum_{i=1}^{N} \hat{\omega}_{u.\varepsilon i}^{-2} \left( T^{-1} \sum_{t=2}^{T} \hat{u}_{it-1} \Delta \hat{u}_{it} - \hat{\lambda}_i \right)}{N^{-1} \sum_{i=1}^{N} \hat{\omega}_{u.\varepsilon i}^{-2} \left( T^{-2} \sum_{t=2}^{T} \hat{u}_{it-1}^2 \right)} \tag{5}$$

$$PP_t^o = N^{1/2} \frac{N^{-1} \sum_{i=1}^{N} \hat{\omega}_{u.\varepsilon i}^{-2} \left( T^{-1} \sum_{t=2}^{T} \hat{u}_{it-1} \Delta \hat{u}_{it} - \hat{\lambda}_i \right)}{\hat{\omega}_{NT} \left( N^{-1} \sum_{i=1}^{N} \hat{\omega}_{u.\varepsilon i}^{-2} \left( T^{-2} \sum_{t=2}^{T} \hat{u}_{it-1}^2 \right) \right)^{1/2}}. \tag{6}$$

The ADF-type test $PP_{df}$ is based on autoregressions to correct for serial correlation. Employing the Frisch-Waugh theorem, two auxiliary regressions are performed

$$\Delta \hat{u}_{it} = \sum_{k=1}^{K_i} \gamma_{1ik} \Delta \hat{u}_{it-k} + \zeta_{1it} \tag{7}$$

$$\hat{u}_{it-1} = \sum_{k=1}^{K_i} \gamma_{2ik} \Delta \hat{u}_{it-k} + \zeta_{2it}, \tag{8}$$

where the lag lengths $K_i$ are determined in our simulations using `AIC` in $\Delta \hat{u}_{it} = \rho_i \hat{u}_{it-1} + \sum_{k=1}^{K_i} \tilde{\gamma}_{ik} \Delta \hat{u}_{it-k} + \tilde{\zeta}_{it}$. Denote the OLS residuals of the above equations (7) and (8) by $\hat{\zeta}_{1it}$ and $\hat{\zeta}_{2it}$, the residuals from the regressions $\hat{\zeta}_{1it} = \rho_i \hat{\zeta}_{2it} + \theta_{it}$ by $\hat{\theta}_{it}$ and their estimated variances (needed later) by $\hat{\sigma}_{\theta_i}^2$. Furthermore define $\hat{\sigma}_{NT}^2 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=K_i+2}^{T} \hat{\theta}_{it}^2$. The essential part of the ADF-type statistic is then given by

$$PP_{df}^o = N^{1/2} \frac{N^{-1} \sum_{i=1}^{N} \hat{\omega}_{u.\varepsilon i}^{-2} \left( T^{-1} \sum_{t=K_i+2}^{T} \hat{\zeta}_{1it} \hat{\zeta}_{2it} \right)}{\hat{\sigma}_{NT} \left( N^{-1} \sum_{i=1}^{N} \hat{\omega}_{u.\varepsilon i}^{-2} \left( T^{-2} \sum_{t=K_i+2}^{T} \hat{\zeta}_{2it}^2 \right) \right)^{1/2}}. \tag{9}$$

By construction, for $N = 1$ these statistics coincide with their time series counterparts. Asymptotic normality using sequential limit theory can easily be established for the above test statistics by applying the so called Delta method (this requires knowledge of the (asymptotic) means and variances of the building blocks, which are obtained for practical purposes by simulation). The mean and variance correction factors, $M_{PP}(r, s, l)$ and $V_{PP}(r, s, l)$ depend upon the test considered ($r \in \{\sigma, \rho, t, df\}$), the case concerning the deterministic variables ($s \in \{1, 2, 3\}$) and upon the number of regressors $l$, i.e.:

$$PP_r = \frac{PP_r^o - N^{1/2} M_{PP}(r, s, l)}{(V_{PP}(r, s, l))^{1/2}} \quad \Rightarrow \quad N(0, 1).$$

7

Pedroni develops three group-mean tests against the heterogeneous alternative. These are: a test based on the first order serial correlation coefficient $PG_\rho$, a test based on its $t$-value $PG_t$ and again an ADF-type test $PG_{df}$. The essential parts of the test statistics are given by

$$
\begin{aligned}
PG_\rho^o &= N^{-1/2} \sum_{i=1}^N PG_{\rho,i}^o &=& N^{-1/2} \sum_{i=1}^N \frac{T^{-1} \sum_{t=2}^T (\hat{u}_{it-1} \Delta \hat{u}_{it} - \hat{\lambda}_i)}{T^{-2} \sum_{t=2}^T \hat{u}_{it-1}^2} \\[2ex]
PG_t^o &= N^{-1/2} \sum_{i=1}^N PG_{t,i}^o &=& N^{-1/2} \sum_{i=1}^N \frac{T^{-1} \sum_{t=2}^T (\hat{u}_{it-1} \Delta \hat{u}_{it} - \hat{\lambda}_i)}{\hat{\omega}_{\mu_i} (T^{-2} \sum_{t=2}^T \hat{u}_{it-1}^2)^{1/2}} \\[2ex]
PG_{df}^o &= N^{-1/2} \sum_{i=1}^N PG_{df,i}^o &=& N^{-1/2} \sum_{i=1}^N \frac{T^{-1} \sum_{t=K_i+2}^T \hat{\zeta}_{1it} \hat{\zeta}_{2it}}{\hat{\sigma}_{\theta_i} (T^{-2} \sum_{t=K_i+2}^T \hat{\zeta}_{2it}^2)^{1/2}}.
\end{aligned}
\tag{10}
$$

Appropriately centered and scaled group-mean test statistics converge to the standard normal distribution in the sequential limit by applying the central limit theorem to the i.i.d. (across N) sequences, i.e.:

$$
PG_r = \frac{PG_r^o - N^{1/2} M_{PG}(r,s,l)}{(V_{PG}(r,s,l))^{1/2}} = N^{-1/2} \sum_{i=1}^N \frac{PG_{r,i}^o - M_{PG}(r,s,l)}{(V_{PG}(r,s,l))^{1/2}} \Rightarrow N(0,1),
\tag{11}
$$

with $r \in \{\rho, t, df\}$, $s \in \{1, 2, 3\}$ and $l$ the number of regressors. The correction factors for all discussed tests are tabulated in Pedroni (1999) for two to seven regressors.

Kao (1999) develops similar tests to those of Pedroni for the special case of panels where not only the vectors $\beta_i$ are assumed to be identical, but also the dynamics of the error processes $v_{it}$ are assumed to be identical for $i = 1, \ldots, N$.

**Westerlund.** Westerlund (2005) develops two simple non-parametric tests that extend the Breitung (2002) approach from the time series to the panel case. One test, $WP$, is pooled and hence specified against the homogeneous alternative and the other one, $WG$, is a group-mean test against the heterogeneous alternative. As for the Pedroni tests, the OLS residuals $\hat{u}_{it}$ from (1) are the starting point. Denote with $\hat{r}_i = \sum_{t=1}^T \hat{u}_{it}^2$, with $\bar{r} = N^{-1} \sum_{i=1}^N \hat{r}_i$ and with $\hat{e}_{it} = \sum_{j=1}^t \hat{u}_{ij}$. The essential parts of the test statistics are then given by

$$
WP^o = N^{1/2} \left( N^{-1} \sum_{i=1}^N \frac{T^2}{\bar{r}} \left( T^{-4} \sum_{t=1}^T \hat{e}_{it}^2 \right) \right)
\tag{12}
$$

$$
WG^o = N^{-1/2} \sum_{i=1}^N \frac{T^2}{\hat{r}_i} \left( T^{-4} \sum_{t=1}^T \hat{e}_{it}^2 \right).
\tag{13}
$$

Applying the Delta method to $WP^o$ and a central limit theorem to $WG^o$ (easily seen again by writing $WG^o = N^{-1/2} \sum_{i=1}^N WG_i^o$) leads to asymptotic standard normality under the null

hypothesis in the sequential limit when applying appropriate mean and variance correction factors, i.e.:

$$WP = \frac{WP^o - N^{1/2}M_{WP}(r,s,l)}{(V_{WP}(r,s,l))^{1/2}} \quad \Rightarrow \quad N(0,1)$$

$$WG = \frac{WG^o - N^{1/2}M_{WG}(r,s,l)}{(V_{WG}(r,s,l))^{1/2}} \quad \Rightarrow \quad N(0,1).$$

### 2.1.2 Estimation of the cointegrating vector

In this subsection we assume that the processes $u_{it}$ are stationary and hence that cointegration prevails for all cross-section members. As mentioned at the beginning of the section, a difference to the time series case is that the cross-section dimension implies that the (pooled) OLS estimator of $\beta$ in (1) (also when the $\beta_i$ are not restricted to be identical for $i = 1, \ldots, N$) converges to a well-defined limit also in the spurious regression and the (heterogeneous) cointegration cases. This limit is given by the so called *average long-run regression coefficient* (for a detailed discussion and the precise assumptions see Theorems 4 and 5 of Phillips and Moon, 1999). As in the time series case, the limiting distribution of the OLS estimator depends upon nuisance parameters. For the rest of this section we focus on the case of *homogeneous* cointegration, i.e. $\beta_i = \beta$ for $i = 1, \ldots, N$, to be estimated by panel methods. Consequently, later on we also consider the system methods for the estimation of a cross-sectionally identical cointegrating space.

As in the time series case regressor endogeneity and serial correlation of the errors, which lead to nuisance parameter dependency of the limiting distribution of the OLS estimator, can be handled in two ways, by either performing *fully modified* OLS estimation (cf. Phillips and Hansen, 1990) or by performing *dynamic* OLS estimation (cf. Saikkonen, 1991). As for the tests both a pooled and a group-mean approach are possible. Furthermore, also the correction factors can be individual specific or cross-sectionally averaged.

In the description of the estimation procedures we focus on the case with fixed effects only (i.e. case 2).[3] Denote with $\bar{y}_i = \frac{1}{N}\sum_{t=1}^{T} y_{it}$ and with $\bar{X}_i = \frac{1}{N}\sum_{t=1}^{T} X_{it}$ and denote the demeaned variables by $\tilde{y}_{it} = y_{it} - \bar{y}_i$ and $\tilde{X}_{it} = X_{it} - \bar{X}_i$.

---

[3]The other two cases are entirely similar. In case 1 the original variables are taken as inputs, and in case 3 the variables are demeaned and detrended at the outset of the procedure. The limiting distributions change accordingly, between case 1 on the one hand and cases 2 and 3 on the other. Also, intercepts in the regressors, i.e. $X_{it} = A_i + X_{it-1} + \varepsilon_{it}$ can be accommodated.

**FM-OLS.** Obtain estimators $\hat{\Omega}_{u\varepsilon i}$, $\hat{\Omega}_{\varepsilon i}$, $\hat{\Lambda}_{u\varepsilon i}$ and $\hat{\Lambda}_{\varepsilon i}$ from the residuals $[\hat{u}_{it}, \varepsilon'_{it}]'$.[4] Defining the endogeneity corrected variable $\tilde{y}^+_{it} = \tilde{y}_{it} - \hat{\Omega}_{u\varepsilon i}\hat{\Omega}^{-1}_{\varepsilon i}\Delta\tilde{X}_{it}$ leads to the pooled FM-OLS estimator:

$$\hat{\beta}_{FM} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}'_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{X}_{it}\tilde{y}^+_{it} - NT(\hat{\Lambda}^+_{u\varepsilon i})'\right)\right), \tag{14}$$

with $\hat{\Lambda}^+_{u\varepsilon i} = \hat{\Lambda}_{u\varepsilon i} - \hat{\Omega}_{u\varepsilon i}\hat{\Omega}^{-1}_{\varepsilon i}\hat{\Lambda}_{\varepsilon i}$.

Phillips and Moon (1999) use in their formulation of the FM-OLS estimator *averaged* correction factors, e.g. $\hat{\Omega}_\varepsilon = \frac{1}{N}\sum_{i=1}^{N}\hat{\Omega}_{\varepsilon i}$ and similarly constructed $\hat{\Omega}_{u\varepsilon}$, $\hat{\Lambda}_\varepsilon$, $\hat{\Lambda}_{u\varepsilon}$ and $\hat{\Lambda}^+_{u\varepsilon}$.[5] The limiting distribution of the FM-OLS estimator (see e.g. Theorem 9 of Phillips and Moon, 1999) is given by:

$$N^{1/2}T(\hat{\beta}_{FM} - \beta) \quad\Rightarrow\quad N(0, 6\omega^2_{u.\varepsilon}\Omega^{-1}_\varepsilon) \tag{15}$$

with $\omega^2_{u.\varepsilon} = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\omega^2_{u.\varepsilon i}$ and $\Omega_\varepsilon = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\Omega_{\varepsilon i}$. For case 1 without deterministic components the factor 6 in the limiting distribution has to be replaced by the factor 2. Note that the limiting covariance matrix is composed of cross-sectional averages.

Standard (up to the factor 2 or 6, depending upon case considered) normally distributed pooled FM-OLS estimators are also easily constructed. These are popular due to their implementation in freely available software. Define $\tilde{y}^0_{it} = \hat{\omega}^{-1}_{u.\varepsilon i}\tilde{y}^+_{it} - \left[(\hat{\omega}^{-1}_{u.\varepsilon i}I_l - \hat{\Omega}^{-1/2}_{\varepsilon i})X_{it}\right]'\hat{\beta}$, $\tilde{X}^0_{it} = \hat{\Omega}^{-1/2}_{\varepsilon i}\tilde{X}_{it}$ and $\hat{\Lambda}^0_{u\varepsilon i} = \hat{\omega}^{-1}_{u.\varepsilon i}\hat{\Lambda}^+_{u\varepsilon i}\hat{\Omega}^{-1/2}_{\varepsilon i}$, where $\hat{\beta}$ denotes the LSDV estimate of $\beta$. Then, the *normalized* FM-OLS estimator is given by

$$\hat{\beta}^0_{FM} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}^0_{it}\tilde{X}^{0\prime}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{X}^0_{it}\tilde{y}^0_{it} - NT(\hat{\Lambda}^0_{u\varepsilon i})'\right)\right) \tag{16}$$

and it holds that $N^{1/2}T(\hat{\beta}^0_{FM} - \beta) \Rightarrow N(0, 6I_l)$, where the factor 6 has to be replaced by the factor 2 in case no deterministic components are included in (1).

*Group-mean* FM-OLS estimation is considered in Pedroni (2000). The group-mean FM-OLS estimator is (in its unnormalized form) given by the cross-sectional average of the individual FM-OLS estimators of $\beta$:

$$\hat{\beta}^G_{FM} = \frac{1}{N}\sum_{i=1}^{N}\left(\left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}'_{it}\right)^{-1}\left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{y}^+_{it} - T(\hat{\Lambda}^+_{u\varepsilon i})'\right)\right). \tag{17}$$

---

[4]By assumption $\varepsilon_{it} = \Delta X_{it}$. Note also that because we assume a homogeneous cointegrating relationship instead of the OLS residuals also the residuals from an LSDV regression (which puts the restriction $\beta_i = \beta$ in place) can be used.

[5]Similar results are also contained in Pedroni (2000) and Kao and Chiang (2000).

**D-OLS.** We now turn to *dynamic* OLS estimation of the cointegrating relationship, as discussed in Kao and Chiang (2000) and Mark and Sul (2003). The idea of D-OLS estimation is to correct for the correlation between $u_{it}$ and $\varepsilon_{it}$ by including leads and lags of $\Delta X_{it}$ as additional regressors in the cointegrating regression. As in the time series case the number of leads and lags (in general) has to increase with the time dimension of the panel at a suitable rate to induce asymptotic uncorrelatedness between the noise processes in the lead and lag augmented cointegrating regression and $\varepsilon_{it}$. Thus, considering again case 2, let the augmented cointegrating regression be given by

$$
\begin{aligned}
\tilde{y}_{it} &= \tilde{X}'_{it}\beta + \sum_{j=-p_i}^{p_i} \Delta \tilde{X}'_{it-j}\gamma_{ij} + u^*_{it} \\
&= \tilde{X}'_{it}\beta + \tilde{Z}'_{it}\gamma_i + u^*_{it},
\end{aligned} \tag{18}
$$

where the last equation defines $\tilde{Z}_{it}$ and $\gamma_i$. The pooled D-OLS estimator for $\beta$ is then obtained from OLS estimation of the above equations (18). Let $\tilde{Q}_{it} = [\tilde{X}'_{it}, 0', \ldots, 0', \tilde{Z}'_{it}, 0', \ldots, 0']' \in \mathbb{R}^{2l(1+\sum_{i=1}^{N} p_i)}$, where the variables $\tilde{Z}'_{it}$ are at the $i$-the position in the second block of the regressors. Using this notation we obtain

$$
\begin{bmatrix} \hat{\beta}_D \\ \hat{\gamma}_1 \\ \vdots \\ \hat{\gamma}_N \end{bmatrix} = \left( \sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{Q}_{it}\tilde{Q}'_{it} \right)^{-1} \left( \sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{Q}_{it}\tilde{y}_{it} \right). \tag{19}
$$

Mark and Sul (2003) derive the asymptotic distribution of $\hat{\beta}_D$ that has a 'sandwich' type limit covariance matrix. Denote with $\bar{V} = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \omega^2_{u.\varepsilon i}\Omega_{\varepsilon i}$. Then it holds that

$$
N^{1/2}T(\hat{\beta}_D - \beta) \Rightarrow N(0, 6\Omega^{-1}_\varepsilon \bar{V}\Omega^{-1}_\varepsilon). \tag{20}
$$

Kao and Chiang (2000) discuss a normalized version of the D-OLS estimator that corresponds to $\hat{\beta}^0_{FM}$. This estimator, $\hat{\beta}^0_D$ say, is obtained when in the above discussion of the D-OLS estimator $\tilde{y}_{it}$ and $\tilde{X}_{it}$ are replaced by $\tilde{y}^0_{it}$ and $\tilde{X}^0_{it}$. These changes lead to an estimator with a limiting covariance matrix proportional to the identity matrix.

Pedroni (2001) considers a group-mean D-OLS estimator. Denote with $\tilde{R}_{it} = [X'_{it}, \tilde{Z}'_{it}]'$ and estimate (separately for $i = 1, \ldots, N$)

$$
\begin{bmatrix} \hat{\beta}_{Di} \\ \hat{\gamma}_i \end{bmatrix} = \left( \sum_{t=1}^{T} \tilde{R}_{it}\tilde{R}'_{it} \right)^{-1} \left( \sum_{t=1}^{T} \tilde{R}_{it}\tilde{y}_{it} \right). \tag{21}
$$

Then the group-mean D-OLS estimator is given by $\hat{\beta}_D^G = \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_{Di}$. Also this estimator can be computed in a normalized version. The limiting distributions coincide with the limiting distributions of the corresponding pooled estimators.

## 2.2 System Methods

The second strand of the panel cointegration literature is based on panel extensions of VAR cointegration analysis (see Johansen, 1995). Compared to the single-equation methods several differences are worth mentioning. First, the systems approach allows to model multiple cointegrating relationships. Second, the cointegrated VAR approach allows to incorporate a richer specification concerning (restricted) deterministic components considered relevant in the applied cointegration literature. Third, specifying a parametric model allows to also consider the dynamic (short-run) characteristics of the data, which are treated as nuisance parameters in the non-parametric single equation approaches. Being based on VAR estimates, the system methods are, as their time series building block, subject to substantial biases for short time series. Thus, for practical applications the time series dimension has to be sufficiently large. This is also required for the reason that specifying a dynamic model requires the estimation of more parameters and hence in general more data. In this respect one, however, has to take into account that also an accurate estimation of the long-run variances used in the nonparametric methods discussed above also requires a sufficiently large time dimension.

Without imposing any homogeneity assumption the panel VAR DGP is given in *error correction form* by

$$\Delta Y_{it} = C_{1i} + C_{2i}t + \alpha_i\beta_i'Y_{it-1} + \sum_{j=1}^{p_i}\Gamma_{ij}\Delta Y_{it-j} + w_{it}, \tag{22}$$

with $Y_{it} \in \mathbb{R}^m$, $C_{1i}, C_{2i} \in \mathbb{R}^m$, $\alpha_i, \beta_i \in \mathbb{R}^{m \times k_i}$ with full rank, $\Gamma_{ij} \in \mathbb{R}^{m \times m}$ and $w_{it}$ cross-sectionally independent $m$-dimensional white noise processes with covariance matrices $\Sigma_i > 0$. To ensure that the processes described by (22) are (up to the deterministic components) I(1) processes, the matrices $\alpha_{i\perp}'\Gamma_i\beta_{i\perp}$ have to be invertible, where $\alpha_{i\perp} \in \mathbb{R}^{m \times (m-k_i)}$, $\beta_{i\perp} \in \mathbb{R}^{m \times (m-k_i)}$ are full rank matrices such that $\alpha_{i,\perp}'\alpha_i = 0$ and $\beta_{i,\perp}'\beta_i = 0$ and $\Gamma_i = I_m - \sum_{j=1}^{p_i}\Gamma_{ij}$.[6] In this case the space spanned by the columns of the matrix $\beta_i$, i.e. $sp\{\beta_i\}$, is the $k_i$-dimensional cointegrating space of unit $i$.[7]

---

[6]One possible choice is given by $\alpha_{i\perp} = I_m - \alpha_i(\alpha_i'\alpha_i)^{-1}\alpha_i'$ and similarly for $\beta_{i\perp}$.

[7]The integer $k_i$ is often referred to as cointegrating rank. Please note that $\beta_i$ as used in this sub-section

In the VAR cointegration literature the following five specifications concerning the deterministic components are usually discussed. Case 1 is without any deterministic components. In case 2 restricted intercepts of the form $C_{1i} = \alpha_i \tau_i$ are contained (in the cointegrating space) and case 3 includes unrestricted intercepts $C_{1i}$ that induce linear time trends in $Y_{it}$. In case 4 unrestricted intercepts and restricted trend coefficients $C_{2i} = \alpha_i \kappa_i$ are included, this allows for linear trends in both the data and the cointegrating relationships. Finally in case 5 unrestricted intercepts and trend coefficients are included. The latter case leads to quadratic time trends in the data and appears to be not too relevant for economic time series. For this reason we do not consider this case in our simulations. A detailed discussion of the specifications of the deterministic variables is given in Johansen (1995, Section 5.7). The statistical analysis, i.e. parameter estimation (via reduced rank regression) as well as testing for the cointegrating rank, is well-developed and known for all the listed cases (see Johansen, 1995) and therefore we do not repeat a discussion of this well-known procedure here.

**Larsson, Lyhagen, and Löthgren.** Larsson, Lyhagen, and Löthgren (2001) consider testing for cointegration in the above framework under the assumption that $\Pi_i = \alpha_i \beta_i' = \alpha \beta' = \Pi$ for $i = 1, \ldots, N$.[8] The null hypothesis of their test is $H_0 : rk(\Pi_i) = k$ for $i = 1, \ldots, N$. The test is consistent against the alternative hypothesis that $H_1 : rk(\Pi_i) = m$ for a non-vanishing fraction of cross-section members. The construction of this test statistic is similar to Im, Pesaran, and Shin (2003) and hence the test statistic is given by a suitably centered and scaled version of the cross-sectional average of the individual trace statistics. Thus, denote with $LR_i^s(k|m)$ the trace statistic for the null hypothesis of a $k$-dimensional cointegrating space for unit $i$, where the superscript $s$ indicates the specification of the deterministic components. Using a central limit theorem in the cross-sectional dimension and the appropriate mean and variance correction factors implies that under the null hypothesis

$$LLL^s(k|m) = N^{-1/2} \sum_{i=1}^{N} \frac{LR_i^s(k|m) - \mathbb{E}(LR_i^s(k|m))}{\sqrt{Var(LR_i^s(k|m))}} \quad \Rightarrow \quad N(0,1) \tag{23}$$

does not coincide with $\beta$ used in the description of the single equation methods, where the single cointegrating vector is given by $[1, -\beta']'$. Also for notational brevity we will not always differentiate between the matrix $\beta$ and the space spanned by its columns. We are confident that this does not lead to any confusion.

[8]As we shall see below, their test is simply based on the cross-sectional average of the Johansen trace statistic, where this restriction is not imposed anywhere in the construction of the test statistic. However, they have only established the asymptotic distribution of their test statistic under this assumption (see Assumption 3' and Theorem 1 of Larsson, Lyhagen, and Löthgren, 2001).

in the sequential limit $T \to \infty$ followed by $N \to \infty$.[9] For $T \to \infty$ the expressions $\mathbb{E}(LR_i^s(k|m))$ and $Var(LR_i^s(k|m))$ converge to the limit of the expected value respectively variance of the trace statistic (corresponding to the case $s$ considered). In principle also finite $T$ correction factors (for different lag lengths and numbers of variables), see again Im, Pesaran, and Shin (2003), can be obtained by simulation. Our simulation study is based on the asymptotic correction factors. Note already here that the simulation results indicate that using finite $T$ correction factors may be beneficial.

**Breitung.** Breitung (2005) proposes a 2-step estimation (and related test) procedure that extends the Ahn and Reinsel (1990) and Engle and Yoo (1991) approach from the time series to the panel case. Breitung considers the homogeneous cointegration case where only the cointegrating spaces are assumed to be identical for all cross-section members. In the first step of his procedure the parameters are estimated individual specifically and in the second step the common cointegrating space $\beta$ is estimated in a pooled fashion.[10]

For simplicity we describe the method here for the VAR(1) model without deterministic components. In the general case lagged differences as well as (restricted) deterministic components are treated in the usual way and are concentrated out in the first step, as described in Johansen (1995). Thus, consider

$$\Delta Y_{it} = \alpha_i \beta' Y_{it-1} + w_{it} \tag{24}$$

Pre-multiplying equations (24) by $T_i = (\alpha_i' \Sigma_i^{-1} \alpha_i)^{-1} \alpha_i' \Sigma_i^{-1}$ leads to

$$(\alpha_i' \Sigma_i^{-1} \alpha_i)^{-1} \alpha_i' \Sigma_i^{-1} \Delta Y_{it} = \beta' Y_{it-1} + (\alpha_i' \Sigma_i^{-1} \alpha_i)^{-1} \alpha_i' \Sigma_i^{-1} w_{it} \tag{25}$$

$$T_i \Delta Y_{it} = \beta' Y_{it-1} + T_i w_{it} \tag{26}$$

$$\Delta Y_{it}^+ = \beta' Y_{it-1} + w_{it}^+, \tag{27}$$

where the last equation defines the variables with superscript $^+$. Note also that $\mathbb{E} w_{it}^+ (w_{it}^+)' = (\alpha_i' \Sigma_i^{-1} \alpha_i)^{-1}$. Now use the normalization $\beta = [I_k, \beta_2']'$ and partition $Y_{it} = [(Y_{it}^1)', (Y_{it}^2)']'$ with $Y_{it}^1 \in \mathbb{R}^k$ and $Y_{it}^2 \in \mathbb{R}^{m-k}$. Using this notation we can rewrite the above equation as

$$\Delta Y_{it}^+ - Y_{it-1}^1 = \beta_2' Y_{it-1}^2 + w_{it}^+. \tag{28}$$

---

[9]The authors actually derive this result for a so called *diagonal* limit with $\frac{N^{1/2}}{T} \to 0$, i.e. for sequences of $(N, T)$ where $N$ grows suitably slower than $T$.

[10]Note that in the first step individual specific estimators of all parameters are obtained and used, including first step estimators of $\hat{\beta}_i$.

14

Breitung suggests to estimate (28) by pooled OLS using the estimate $\hat{T}_i = (\hat{\alpha}_i' \hat{\Sigma}_i^{-1} \hat{\alpha}_i)^{-1} \hat{\alpha}_i' \hat{\Sigma}_i^{-1}$ based on the Johansen estimators. Note that, given that the covariance structure of the errors in (28) is known (and an estimate is available), also pooled feasible GLS estimation of (28) is an option.

Breitung's estimation procedure stops here. However, an iterative estimator is easily conceived, based on the above procedure. With the estimated $\hat{\beta}_2$, one can re-estimate the individual specific parameters in (24) by running separate OLS regressions. With the new estimates of $\alpha_i$ and $\Sigma_i$ (in the VAR(1) example without deterministics) then again equation (28) can be estimated. This process can be continued until convergence occurs. Such an iterative procedure corresponds to the iterative estimator proposed in Larsson and Lyhagen (1999).[11] The only difference is that in the first step Larsson and Lyhagen (1999) propose to take as an initial estimator $\hat{\beta} = \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_i$, where $\hat{\beta}_i$ denotes the Johansen estimator of the cointegrating space for cross-section unit $i$.[12] We refer to this initial estimator later on as *one-step* or *group-mean* VAR estimator, in analogy to the group-mean FM-OLS and D-OLS estimators discussed above.

Breitung shows that the two-step estimator, $\tilde{\beta}_2$ say, is asymptotically normally distributed

$$N^{1/2} T \text{vec}(\tilde{\beta}_2 - \beta_2) \quad \Rightarrow \quad N(0, \Omega_{\mathbf{2}}^{-1} \otimes \Sigma_\alpha), \tag{29}$$

with $\otimes$ denoting the Kronecker product, $\Omega_{\mathbf{2}} = \lim_{N\to\infty} \lim_{T\to\infty} \mathbb{E}\left[ \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} Y_{it-1}^2 (Y_{it-1}^2)' \right]$ and $\Sigma_\alpha = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} (\alpha_i' \Sigma_i^{-1} \alpha_i)^{-1}$.

The test Breitung proposes for the null hypothesis of $rk(\beta) = k$ is based on Saikkonen (1999). The difference to the Larsson, Lyhagen, and Löthgren (2001) test is that Breitung's test incorporates the homogeneity restriction $\beta_i = \beta$ for $i = 1, \dots, N$ in the construction of the

---

[11]Note, however, that Larsson and Lyhagen (1999) consider a much more general specification (using again the VAR(1) example) of the form $\begin{bmatrix} \Delta Y_{1t} \\ \vdots \\ \Delta Y_{Nt} \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \dots & \alpha_{1N} \\ \vdots & \ddots & \vdots \\ \alpha_{N1} & \dots & \alpha_{NN} \end{bmatrix} \begin{bmatrix} \beta' & 0 & \dots \\ 0 & \beta' & \vdots \\ 0 & \dots & \beta' \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ \vdots \\ Y_{Nt-1} \end{bmatrix} +$ $\begin{bmatrix} w_{1t} \\ \vdots \\ w_{Nt} \end{bmatrix}$, with a full covariance matrix of the stacked noise process. It is not clear whether such an – up to the cointegrating space – unrestricted VAR process for the stacked data should really be interpreted as a panel model, as the loose parameterization implies that the time dimension of the panel has to be large compared to the cross-sectional dimension. Also Groen and Kleibergen (2003) consider a relatively general panel VAR model that seems to require panel data sets with a large time dimension, as in their simulations they consider a bivariate example for $N = 1, 3, 5$ and $T = 1000$.

[12]Before averaging the estimators over the cross-section dimension it is important to impose a common normalization, e.g. $\hat{\beta}_i = [I_k, \hat{\beta}_{i,2}']'$.

test statistic. The discussion is again for the VAR(1) case without deterministic components. Denote with $\gamma_i \in \mathbb{R}^{m \times (m-k)}$ matrices with full column rank and consider

$$\Delta Y_{it} = \alpha_i \beta' Y_{it-1} + \gamma_i \beta'_\perp Y_{it-1} + w_{it}. \tag{30}$$

Under the null hypothesis of a $k$-dimensional cointegrating space, $\gamma_i = 0$ for $i = 1, \ldots, N$ and under the alternative (of an $m$-dimensional cointegrating space) $\gamma_i$ is unrestricted (in a non-vanishing fraction of panel members to imply consistency of the test against this alternative) to allow for $\Pi_i = \alpha_i \beta' + \gamma_i \beta'_\perp$ of full rank. Pre-multiply (30) with $\alpha'_{i,\perp}$ to obtain

$$\alpha'_{i,\perp} \Delta Y_{it} = \alpha'_{i,\perp} \gamma_i \beta'_\perp Y_{it-1} + \alpha'_{i,\perp} w_{it} \tag{31}$$

$$\alpha'_{i,\perp} \Delta Y_{it} = \phi_i(\beta'_\perp Y_{it-1}) + \tilde{w}_{it}, \tag{32}$$

where the last equation defines the coefficients and variables. Replacing $\alpha_{i,\perp}$ and $\beta_\perp$ by estimators (as discussed above) allows to estimate equations (32) separately by OLS and to construct test statistics for the hypotheses $H_0 : \phi_i = 0$ for $i = 1, \ldots, N$. Any of the Lagrange multiplier, likelihood ratio or Wald test statistics can be used. Our implementation rests upon the Lagrange multiplier test statistic, which has the advantage that it only requires estimation under the null hypothesis. Denote with $\hat{f}_{it} = \hat{\alpha}'_{i,\perp} \Delta Y_{it}$ and with $\hat{g}_{it} = \hat{\beta}'_\perp Y_{it}$, then the Lagrange multiplier test statistic for unit $i$ is given by

$$LM_i(k|m) = T \mathrm{tr} \left[ \sum_{t=2}^{T} \hat{f}_{it} \hat{g}'_{it-1} \left( \sum_{t=2}^{T} \hat{g}_{it-1} \hat{g}'_{it-1} \right)^{-1} \sum_{t=2}^{T} \hat{g}_{it-1} \hat{f}'_{it} \left( \sum_{t=2}^{T} \hat{f}_{it} \hat{f}'_{it} \right)^{-1} \right], \tag{33}$$

which is sequentially computed for the different values of $k = 0, \ldots, m$. The panel test statistic is then, as usual, given by the corresponding centered and scaled cross-sectional average (putting again the superscript to indicate the dependence upon the deterministic components). Thus, under the null hypothesis

$$B^s(r|m) = N^{-1/2} \sum_{i=1}^{N} \frac{LM_i^s(k|m) - \mathbb{E}(LM_i^s(k|m))}{\sqrt{Var(LM_i^s(k|m))}} \quad \Rightarrow \quad N(0,1). \tag{34}$$

The *asymptotic* mean and variance correction factors coincide with those of Larsson, Lyhagen, and Löthgren (2001) in (23) for all specifications of the deterministic components.

# 3    The Simulation Study

In this section we present a representative selection of results obtained from our large scale simulation study. Due to space constraints we only report a small subset of results and focus

on some of the main observations that emerge. The full set of results is available from the authors upon request.

The computations have been performed in GAUSS, except for the kernel density estimates (see the discussion below) that have been computed using MATLAB. The number of replications is 5,000 for each DGP and panel size. The time dimension $T$ assumes the values in the set $\{10, 25, 50, 100\}$ and the cross-section dimension $N$ assumes values in the set $\{5, 10, 25, 50, 100\}$, with the additional constraint that $T \geq N$. This leads to in total fourteen different panel sizes. The constraint $T \geq N$ is also imposed for the following reason: A cross-sectional dimension that is too large compared to the time series dimension leads to size divergence, i.e. the actual size tends to 1 for increasing $N$ and fixed $T$ smaller than $N$. This finding is analogous to similar findings for panel unit root tests in Hlouskova and Wagner (2006).

**Baseline.** We start the description of the DGPs with the baseline case of cross-sectionally independent processes. We consider three-dimensional VAR(2) processes, i.e. $p_i = p = 1$ in the error correction representation (22), with cointegrating ranks $k = 0, 1, 2$. For $k = 1$ the common cointegrating space is given by $\beta = \begin{bmatrix} 1 & -1 & -1 \end{bmatrix}'$ and for $k = 2$ we use $\beta = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}'$. The data series $Y_{it}$ are generated by static linear transformations of processes $\tilde{Y}_{it}$ as

$$M_i Y_{it} = \tilde{Y}_{it} = \tilde{C}_{1i} + \tilde{C}_{2i} t + \begin{bmatrix} \tilde{a}_{11}^i & 0 & 0 \\ 0 & \tilde{a}_{12}^i & 0 \\ 0 & 0 & \tilde{a}_{13}^i \end{bmatrix} \tilde{Y}_{it-1} + \begin{bmatrix} \tilde{a}_{21}^i & 0 & 0 \\ 0 & \tilde{a}_{22}^i & 0 \\ 0 & 0 & \tilde{a}_{23}^i \end{bmatrix} \tilde{Y}_{it-2} + \tilde{w}_{it}, \quad (35)$$

with $M_i = [\beta, R_i]'$ and where $R_i \in \mathbb{R}^{3 \times (3-k)}$ are individual specific (parts of) positive definite matrices that are generated as the product of matrices with uniformly distributed entries with their transpose. The *diagonal* processes $\tilde{Y}_{it}$ have their coordinates de-coupled and thus the stochastic properties (like stationarity or integratedness) of these processes are separated across coordinates. The first $k$ coordinates of $\tilde{Y}_{it}$ are generated as stationary VAR(2) processes and the remaining coordinates as I(1) VAR(2) processes. Note that this also implies that in case of cointegration the cointegrated linear combinations of the processes $Y_{it}$, i.e. $\beta' Y_{it}$, are stationary (de-coupled) AR(2) process(es). Note that in general the cointegrated linear combinations, when starting from cointegrated VAR processes, are stationary ARMA processes (see Zellner and Palm, 1974) and the AR structure is an implication of our con-

17

struction via diagonal processes. This implies that the processes $u_{it}$ in (1) when considering the estimation problem, for $k = 1$, with single equation methods are AR processes.

| Cointegrating rank $k$ | Coordinate $j$ | $q_{1j}^i$ | $q_{2j}^i$ |
|---|---|---|---|
| 0 | 1 | 1 | U[1.8,3.0] |
|   | 2 | 1 | U[1.8,3.0] |
|   | 3 | 1 | U[1.8,3.0] |
| 1 | 1 | {1.1,1.3,1.5} | U[1.5,2.5] |
|   | 2 | 1 | U[1.8,3.0] |
|   | 3 | 1 | U[1.8,3.0] |
| 2 | 1 | U[1.5,2.5] | U[1.5,2.5] |
|   | 2 | U[1.5,2.5] | U[1.5,2.5] |
|   | 3 | 1 | U[1.8,3.0] |

Table 1: Specification of autoregressive roots for the diagonal processes $\tilde{Y}_{it}$.

Given that the coefficients in the diagonal autoregressive matrices are functions of the roots (of polynomials of degree 2) the integration properties are most easily controlled by appropriately choosing the roots, $q_{1j}^i, q_{2j}^i$ say, for $j = 1, 2, 3$, see Table 1.[13] For $k = 0$ all three coordinates of $\tilde{Y}_{it}$ are I(1) processes where the stable roots are generated (cross-section specifically) uniformly in the interval $[1.8, 3.0]$. For $k = 1$ we generate the smaller stable root $q_{11}^i = q_{11} \in \{1.1, 1.3, 1.5\}$ to assess the the sensitivity of the cointegration tests and estimators when stable roots tends to 1. For the single equation cointegration tests this allows us to study the power and for the system methods we use this to study the sensitivity of the hit rates of the correct cointegrating rank. The larger stable root of the first coordinate is generated unit specifically $U[1.5, 2.5]$. The stable roots of the I(1) coordinates are generated $U[1.8, 3.0]$. For the case $k = 2$ all four stable roots of the two stationary coordinates of $\tilde{Y}_{it}$ are generated as $U[1.5, 2.5]$ and the stable root of the I(1) coordinate is generated as $U[1.8, 3.0]$. All roots are drawn separately for each unit.

The companion paper Hlouskova and Wagner (2006) studies in detail the impact of moving average roots (approaching 1) on the performance of panel unit root tests. Because the single equation tests are panel unit root tests on the residuals of the panel spurious regression we do not repeat this analysis here and only note that the results along this dimension can only be worse than in the panel unit root case, because the unobserved errors are replaced in the cointegration analysis with e.g. the OLS residuals, which introduces additional finite

---

[13]To be precise we have $\tilde{a}_{1j}^i = \frac{1}{q_{1j}^i} + \frac{1}{q_{2j}^i}$ and $\tilde{a}_{2j}^i = -\frac{1}{q_{1j}^i q_{2j}^i}$.

|          | Case 1 | Case 2                                    | Case 3        | Case 4                                                |
|----------|--------|-------------------------------------------|---------------|-------------------------------------------------------|
| $C_{1i}$ | $--$   | $\alpha_i C_1, C_1 \in \mathbb{R}^k$ <br> $U[0.01, 0.05]$ | U[0.01,0.05]  | U[0.01,0.05]                                          |
| $C_{2i}$ | $--$   | $--$                                      | $--$          | $\alpha_i C_2, C_2 \in \mathbb{R}^k$ <br> $U[0.002, 0.0125]$ |

Table 2: Specification of deterministic components.

sample biases. Also for VAR systems cointegration methods the impact of moving average roots is quite well documented in the time series literature (see e.g. Bauer and Wagner, 2006). Therefore we do not focus in this study again on the impact of moving average roots and refer to reader to the cited papers to get an impression of the effects that have to be expected.

The coefficients corresponding to the deterministic components in (35) are chosen as follows. Define $\tilde{C}_{1i} = M_i C_{1i}$ and $\tilde{C}_{2i} = M_i C_{2i}$. Then the coefficients $C_{1i}$ and $C_{2i}$ corresponding to the processes $Y_{it}$ are described in Table 2. The magnitude of the individual specific intercept and trend coefficients is chosen to generate data that 'resemble' macro-economic time series in their trend behavior. Note that in cases 2 and 4 with restrictions upon the deterministic components, where $C_1$ respectively $C_2$ assume identical values for $i = 1, \ldots, N$ the individual specific adjustment matrices $\alpha_i$ nevertheless lead to individual specific means and trends despite the cross-sectionally common underlying means and trend components in the cointegrating relationships.[14] We use this set-up to achieve in all four cases concerning the deterministic components identical cointegrating spaces (including the deterministic components in the cointegrating space) for all cross-section units.

The white noise processes $\tilde{w}_{it} = M_i w_{it}$ are generated normally distributed with covariance matrices $\tilde{\Sigma}_i = M_i \Sigma M_i'$, with $\Sigma = \frac{1}{25} \begin{bmatrix} 0.47 & 0.20 & 0.18 \\ 0.20 & 0.32 & 0.27 \\ 0.18 & 0.27 & 0.30 \end{bmatrix}$, which is up to the factor 1/25 as in Saikkonen and Luukkonen (1997). In the baseline simulations the noise processes are assumed to be cross-sectionally independent.

On top of the above baseline specifications we also try to quantify the impact of the presence of one I(2) component in each cross-sectional unit, short-run cross-sectional correlation, and cross-unit cointegration.

---

[14]The matrices $\alpha_i$ are given by $\alpha_i = -M_i^{-1} \tilde{a}_i(1)[I_k \ 0_{k \times (3-k)}]'$, with $\tilde{a}_i(z)$ denoting the autoregressive polynomial for $\tilde{Y}_{it}$.

**I(2) component.** For the cases $k = 0$ and $k = 1$ we also assess the robustness with respect to the presence of one I(2) trend in $Y_{it}$ for $i = 1, \ldots, N$. For $k = 0$ this is achieved by setting $q_{21}^i = 1$ and for $k = 1$ we set $q_{22}^i = 1$ for $i = 1, \ldots, N$, in which case we take $q_{11}^i$ individually specifically distributed as $U[1.5, 2.5]$. For both cases of $k$ we investigate the robustness of the tests and for $k = 1$ we investigate in addition the robustness of the estimators.

**Short-run cross-sectional correlation.** As the description of the methods in the previous section has made clear, the cross-sectional independence assumption allows to use sequential limit theory and simple central limit theorems to establish the results. Depending upon the extent of violation of this assumption asymptotic normality may still prevail or fail. Consider the following simple example to illustrate the issue: Let $\tau_i, i \in \mathbb{N}$ be a sequence of $N(0, 1)$ distributions with for simplicity stationary covariance function $cov(\tau_i, \tau_j) = \rho_{ij} = \rho_{|i-j|}$. Then, if the covariance function is absolutely summable we obtain $\lim_{N \to \infty} N^{-1/2} \sum_{i=1}^{N} \tau_i = N(0, \sum_{i=-\infty}^{\infty} \rho_i)$. However, if the correlation (or more generally the dependence) between the elements of the sequence is too large, asymptotic normality when scaled with $N^{-1/2}$ may fail. The extreme case is of course a sequence with $\tau_i = \tau$ for all $i \in \mathbb{N}$, with $N^{-1/2} \sum_{i=1}^{N} \tau_i = N^{1/2} \tau$ diverging and $N^{-1} \sum_{i=1}^{N} \tau_i = \tau$.

From the above example we may *heuristically* conclude that 'moderate' short-run cross-sectional dependence leads to asymptotic normality of test statistics, but generally with the variance of the limiting distribution influenced by the form and extent of cross-sectional dependence. 'Large' cross-sectional dependencies will in general imply a more dramatic change of the first generation methods' asymptotic behavior. Short-run cross-sectional correlation can in principle be mitigated (if $T$ is large enough compared to $N$) by using some GLS type correction to the methods employed and to re-establish standard normal asymptotic distributions thereby. Theoretically this is more intricate than it is often presented, as it requires to study joint limits in both $T$ and $N$.

In our simulations we assess the sensitivity of the methods with respect to cross-sectional correlation in the innovations with three different cases: constant correlation, Toeplitz correlation and a factor structure in the errors. The first two cases are the multivariate analogues of cross-sectional correlation studied for panel unit root tests in Hlouskova and Wagner (2006). The constant correlation covariance matrix $\mathbf{\Sigma^{CC}} \in \mathbb{R}^{Nm \times Nm}$ and the Toeplitz covariance

matrix $\boldsymbol{\Sigma}^{\mathbf{TP}} \in \mathbb{R}^{Nm \times Nm}$ are given by

$$\boldsymbol{\Sigma}^{CC} = \begin{bmatrix} 1 & \kappa & \dots & \kappa \\ \kappa & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \kappa \\ \kappa & \dots & \kappa & 1 \end{bmatrix} \otimes \Sigma, \qquad \boldsymbol{\Sigma}^{TP} = \begin{bmatrix} 1 & \kappa & \dots & \kappa^{N-1} \\ \kappa & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \kappa \\ \kappa^{N-1} & \dots & \kappa & 1 \end{bmatrix} \otimes \Sigma, \qquad (36)$$

where we choose $\kappa \in \{0.3, 0.6, 0.9\}$. Because the innovation covariance matrix $\Sigma$ is identical for all cross-section units, the constant correlation case implies that the correlations between any pair of series from different units is $\kappa$ times as large as the correlation between the same pair of series from the same cross-sectional unit. The cross-sectional correlation function is therefore not summable. The Toeplitz case corresponds to a spatial autoregression of order 1 (interpreting the cross-section dimension spatially), with the correlations decreasing geometrically with 'distance' and hence in this case the cross-sectional correlation function is summable.

The third formulation of cross-sectional correlation considered is a factor structure with two common stationary factors, i.e. $w_{it} = \lambda_i F_t + w_{it}^*$, with $w_{it}^*$ cross-sectionally independent white noise processes with covariance matrix $\Sigma$, $\lambda_i \in \mathbb{R}^{3 \times 2}$, where the entries are uniformly distributed $U[-0.4, 0.4]$ and $F_t \sim N(0, I_2)$. For simulated, given factor loadings $\lambda_i$, the stacked vector $w_t = [(w_{1t})', \dots, (w_{Nt})']'$ is normally distributed with covariance matrix

$$\boldsymbol{\Sigma}^{\mathbf{F}} = \begin{bmatrix} \Sigma + \lambda_1 \lambda_1' & \lambda_1 \lambda_2' & \dots & \lambda_1 \lambda_N' \\ \lambda_2 \lambda_1' & \Sigma + \lambda_2 \lambda_2' & \dots & \lambda_2 \lambda_N' \\ \vdots & \ddots & \ddots & \vdots \\ \lambda_N \lambda_1' & \dots & \lambda_N \lambda_{N-1}' & \Sigma + \lambda_N \lambda_N' \end{bmatrix}. \qquad (37)$$

Given that $\det \boldsymbol{\Sigma}^{\mathbf{CC}}$, $\det \boldsymbol{\Sigma}^{\mathbf{TP}}$ and $\det \boldsymbol{\Sigma}^{\mathbf{F}} > 0$ (verified in each simulation for the latter which depends upon random $\lambda_i$) no *additional* cointegrating relationships over and above the cointegrating relationships comprising only variables from one cross-section member (i.e. of the form $\beta' Y_{it}$) are introduced in the joint I(1) process $Y_t = [Y_{1t}', \dots, Y_{Nt}']'$.[15] Thus, with this set-up we generate simulation data that exhibit short-run cross-sectional correlation to assess

---

[15]Note here that it is not generally true that a process composed of stacking I(1) (or stationary) processes is again an I(1) (or stationary) process. The probably simplest example for this observation is given by $\varepsilon_t$ being an (i.i.d normally distributed) white noise process and the stacked process $[\varepsilon_t', \varepsilon_{qt}']'$, where the second block-component is given by the (also white noise) sub-sampled process $\varepsilon_{qt}$ with e.g. $q \in \mathbb{N}, q \geq 2$. It is straightforward to verify by computing the autocovariance function that the stacked process is not stationary. The example also shows that it is not generally true that the sum of stationary processes is again a stationary process. We do not run into the problem of non I(1)-ness of the stacked panel process $Y_t$ in our study, because we construct $Y_t$ based on either cross-sectionally independent and hence also jointly stationary processes $w_{it}$ or a jointly stationary process $w_t = [w_{1t}', \dots, w_{Nt}']'$ in case of cross-sectional correlation.

the relative sensitivity of the different tests and estimators.

**Cross-unit cointegration.** Under any set of assumptions that ensures that the stacked process $Y_t$ is also jointly an I(1) process, the issue of cross-unit cointegration can be meaningfully discussed. Let us start with a definition of the *cross-unit cointegrating space* and the *cross-unit cointegrating rank*. Denote with $B \in \mathbb{R}^{Nm \times K}$ (a basis of) the $K$-dimensional cointegrating space of the stacked process $Y_t$ and with $\beta_i \in \mathbb{R}^{m \times k_i}$ bases of the cointegrating spaces of the processes $Y_{it}$, considering here for generality of the definition cross-section specific cointegrating spaces $sp\{\beta_i\}$. Furthermore define

$$\vec{\beta} = \begin{bmatrix} \beta_1 & 0 & \ldots & 0 \\ 0 & \beta_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \beta_N \end{bmatrix} \in \mathbb{R}^{Nm \times \sum k_i}. \tag{38}$$

Note that it always holds that $sp\{B\} \supseteq sp\{\vec{\beta}\}$.

**Definition 1** *Under the assumption that the stacked process $Y_t = [Y'_{1t}, \ldots, Y'_{Nt}]'$ is an I(1) process the cross-unit cointegrating space is defined as the span of $B^{CU} = (I_{Nm} - \vec{\beta}(\vec{\beta}'\vec{\beta})^{-1}\vec{\beta}')B$, i.e. by the projection of the cointegrating space $B$ of $Y_t$ on the ortho-complement of $\vec{\beta}$ defined in (38). The cross-unit cointegrating rank is defined as the dimension of $sp\{B^{CU}\}$.*

This is to the best of our knowledge the first formal definition of the cross-unit cointegrating space. It formalizes the notion that relationships like in the simplest example $[\beta'_1, \beta'_2, 0, \ldots, 0]'$ that involve variables from different cross-sections but lead to a stationary transformed process $\beta'_1 Y_{1t} + \beta'_2 Y_{2t}$ via combinations of already stationary transformations of variables from different cross-section units ($\beta'_i Y_{it}$, $i = 1, 2$), should not be considered as *genuine* cross-unit cointegrating relationships. The above definition of the space $B^{CU}$ as the projection of $B$ on the ortho-complement of $\vec{\beta}$ delivers all cointegrating relationships that are not given by linear combinations of individual specific cointegrating relationships, i.e. $B = \vec{\beta} \oplus B^{CU}$, with $\oplus$ denoting the direct sum. The second important component of the definition is the restriction to situations where the joint process is also an I(1) process.

In our simulation study we introduce cross-unit cointegration of dimension one in the generation of the stacked process $\tilde{Y}_t$ (and hence also in $Y_t$) via $\mathbf{M}Y_t = \tilde{Y}_t = \left[\tilde{Y}'_{1t}, \ldots, \tilde{Y}'_{Nt}\right]'$,

with

$$
\mathbf{M} = \begin{bmatrix}
\beta' & 0 & \ldots & 0 \\
\psi' & \psi' & \ldots & \psi' \\
\dot{R}_1' & 0 & \ldots & 0 \\
0 & \begin{bmatrix} \beta' \\ R_2' \end{bmatrix} & \ddots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
0 & \ldots & 0 & \begin{bmatrix} \beta' \\ R_N' \end{bmatrix}
\end{bmatrix}.
\tag{39}
$$

We take $\psi = [1, 1, 0]'$, which is not contained in the cointegrating spaces $\beta$ as defined above. In this case both stable roots in the coordinate of the diagonal process $\tilde{Y}_{1t}$ corresponding to this cointegrating relationship are generated as $U[1.5, 2.5]$. The matrix $\dot{R}_1 \in \mathbb{R}^{3 \times (3-k-1)}$ is up to the different dimensions generated as the matrices $R_i \in \mathbb{R}^{3 \times (3-k)}$ for $i = 2, \ldots, N$, as described in the baseline case. We refer to this case below as the cross-unit cointegration case, which by construction (because $\mathbf{M}^{-1}$ from (39) is not block-diagonal) also implies cross-sectional correlation between the processes $Y_{it}$ despite the cross-sectionally independent innovations $w_{it}$. Note also that the set-up in (39) implies that the marginal DGPs for $Y_{it}$ are VARMA processes and not VAR processes. Thus, VAR estimation for $Y_{it}$ is only an approximation in the cross-unit cointegration case. Asymptotic validity of the Johansen approach in this context is established in Saikkonen (1992) when the VAR lag lengths increase at a suitable rate with the sample size.

Introducing cross-unit cointegration in addition to the individual cointegrating relationships is a benevolent form of misspecification, as in this case the estimation problem for the individual and identical cointegrating relationships is still well defined and thus the corresponding impacts on the performance constitute a 'lower bound' for the effects of cross-unit cointegration. In situations where there are no (or not only) cross-sectionally identical unit specific cointegrating relationships the asymptotic behavior of the discussed estimators is potentially much more fundamentally adversely affected.

We only report results for the deterministic specifications 2 to 4, because case 1 has limited empirical relevance for economic time series. The discussion of the methods in the previous section has made clear that in the single equation methods only the deterministic components in the (single) cointegrating relationship are estimated. Consequently, the 'correspondence' of the deterministic specifications between the VAR and single equation methods is as follows.

23

If the data are generated according to VAR case 2 we include only fixed effects in the single equation methods, in cases 3 and 4 we include both fixed effects and linear time trends.

## 3.1 The Performance of the Tests

In this subsection we report the results concerning the performance of the tests. We use the word size to denote the type I error rate at the actual DGP. This is not the size as defined by the maximal type I error rate over all feasible DGPs under the null hypothesis, see Horowitz and Savin (2000) for an excellent discussion of this issue. Also based on insights in that paper we do not base the analysis of the power of the tests on so-called *size-corrected* critical values, because size correction based on arbitrary points in the set of feasible DGPs under the null hypothesis in general leads to empirically irrelevant critical values. This occurs unless the test statistic is pivotal, which is not the case in finite sample for any of the tests discussed in this paper.

When applying the system tests one performs a sequence of tests (with each test performed at a certain nominal critical level, 5% in our case) until the first rejection of the null hypothesis occurs. The percentage of correct decisions that results from these test sequences is called *hit rate*. By definition, for the single equation tests the hit rate is given by one minus the actual size. In cases where we compare single equation tests and system test sequences we use hit rates as a commonly applicable performance measure.

The relative performance of the tests is practically unchanged across the different deterministic specifications, therefore we abstain from a separate discussion of the different deterministic cases. It has to be noted, however, that the absolute performance of the two system tests varies across the different deterministic specifications. On average the best performance is obtained for case 4, followed by case 2 and often the worst performance occurs for case 3.

In Figure 1 we start with displaying the impact of the three considered forms of short-run cross-sectional correlation on the size of the single equation tests. For the cases of constant and Toeplitz correlation the findings are remarkably similar to those obtained in Hlouskova and Wagner (2006) for panel unit root and stationarity tests. As in that paper, strong effects only occur if the coefficient $\kappa$ is set to the largest considered value of 0.9, see the lower row of Figure 1. For smaller values of $\kappa$ the effects are quite modest and for $\kappa = 0.6$ comparable for those obtained when using a factor structure for the errors to model cross-sectional correlation.
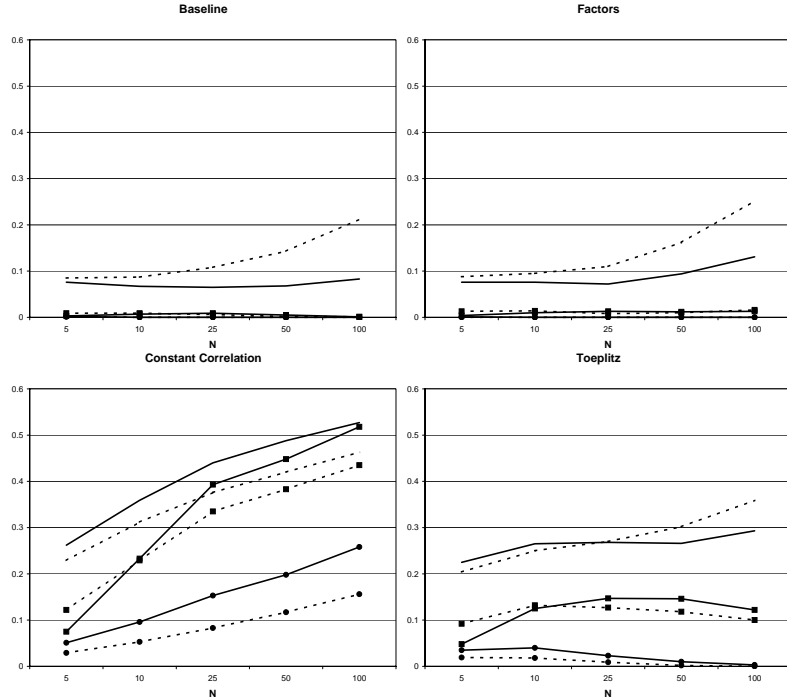
Figure 1: Size of the single equation tests for case 3 for $T = 100$. For the constant correlation and Toeplitz covariance matrices $\kappa = 0.9$.

The solid line corresponds to $PP_{df}$, the dashed line to $PG_{df}$, the solid line with bullets to $PP_t$, the dashed line with bullets to $PG_t$, the solid line with squares to $WP$ and the dashed line with squares to $WG$.

Similar findings prevail also for the system tests. Thus, for the remainder of the section we throughout refer to the factor model case when talking about cross-sectional correlation.[16]

Amongst the single equation tests the generally best performing tests are the two ADF type tests of Pedroni, $PP_{df}$ and $PG_{df}$, and from Westerlund's tests it is the $WG$ test. The good performance of the ADF type tests is potentially partly due to the fact that the true but unobserved processes $u_{it}$ in equation (1) on which the tests are performed are due to our diagonal VAR set-up AR processes and the observable quantities $\hat{u}_{it}$ are ARMA processes, which appear to be well approximated by AR processes. The second general remark with respect to Pedroni's tests is that the test type, i.e. whether based on $\rho$, on its $t$-value or using the ADF approach, has a larger impact on the performance than whether the test is computed in a pooled or group-mean fashion.

---

[16]Note also that under certain assumptions Bai and Kao (2005) derive consistency of FM-OLS estimation for the case of short-run cross-sectional correlation introduced via stationary factors in the errors, with the limiting distribution depending upon the factors and their loadings.

| N | T | Baseline | | | | | I(2) Component | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $PP_{df}$ | $PG_{df}$ | WG | LLL | B | $PP_{df}$ | $PG_{df}$ | WG | LLL | B |
| 5 | 10 | 0.63 | 1.00 | 1.00 | 0.00 | 1.00 | 0.63 | 1.00 | 1.00 | 0.00 | 1.00 |
| | 25 | 0.85 | 0.78 | 1.00 | 0.04 | 0.86 | 0.87 | 0.79 | 1.00 | 0.01 | 0.56 |
| | 50 | 0.91 | 0.88 | 1.00 | 0.41 | 0.84 | 0.88 | 0.84 | 0.96 | 0.06 | 0.37 |
| | 100 | 0.92 | 0.92 | 0.99 | 0.71 | 0.88 | 0.90 | 0.89 | 0.89 | 0.19 | 0.36 |
| 10 | 10 | 0.71 | 1.00 | 1.00 | 0.00 | 1.00 | 0.71 | 1.00 | 1.00 | 0.00 | 1.00 |
| | 25 | 0.87 | 0.75 | 1.00 | 0.00 | 0.76 | 0.93 | 0.80 | 1.00 | 0.00 | 0.26 |
| | 50 | 0.92 | 0.86 | 1.00 | 0.18 | 0.75 | 0.94 | 0.87 | 0.98 | 0.00 | 0.12 |
| | 100 | 0.93 | 0.91 | 0.99 | 0.58 | 0.84 | 0.92 | 0.89 | 0.86 | 0.03 | 0.13 |
| 25 | 25 | 0.80 | 0.55 | 1.00 | 0.00 | 0.47 | 0.94 | 0.68 | 1.00 | 0.00 | 0.01 |
| | 50 | 0.91 | 0.81 | 1.00 | 0.01 | 0.56 | 0.94 | 0.81 | 0.98 | 0.00 | 0.00 |
| | 100 | 0.94 | 0.89 | 0.99 | 0.32 | 0.74 | 0.91 | 0.83 | 0.58 | 0.00 | 0.00 |
| 50 | 50 | 0.91 | 0.73 | 1.00 | 0.00 | 0.31 | 0.93 | 0.72 | 0.95 | 0.00 | 0.00 |
| | 100 | 0.93 | 0.86 | 1.00 | 0.10 | 0.56 | 0.88 | 0.77 | 0.31 | 0.00 | 0.00 |
| 100 | 100 | 0.92 | 0.79 | 1.00 | 0.01 | 0.31 | 0.89 | 0.70 | 0.11 | 0.00 | 0.00 |
| N | T | Cross-sectional correlation | | | | | Cross-unit cointegration | | | | |
| | | $PP_{df}$ | $PG_{df}$ | WG | LLL | B | $PP_{df}$ | $PG_{df}$ | WG | LLL | B |
| 5 | 10 | 0.62 | 1.00 | 1.00 | 0.00 | 1.00 | 0.61 | 1.00 | 1.00 | 0.00 | 1.00 |
| | 25 | 0.87 | 0.79 | 1.00 | 0.04 | 0.81 | 0.85 | 0.78 | 1.00 | 0.04 | 0.83 |
| | 50 | 0.91 | 0.88 | 0.99 | 0.35 | 0.76 | 0.91 | 0.87 | 0.99 | 0.38 | 0.81 |
| | 100 | 0.92 | 0.91 | 0.99 | 0.61 | 0.78 | 0.92 | 0.91 | 0.99 | 0.69 | 0.86 |
| 10 | 10 | 0.69 | 1.00 | 1.00 | 0.00 | 1.00 | 0.72 | 1.00 | 1.00 | 0.00 | 1.00 |
| | 25 | 0.82 | 0.69 | 1.00 | 0.00 | 0.65 | 0.80 | 0.66 | 1.00 | 0.00 | 0.74 |
| | 50 | 0.91 | 0.85 | 1.00 | 0.13 | 0.62 | 0.92 | 0.85 | 0.99 | 0.17 | 0.75 |
| | 100 | 0.92 | 0.91 | 0.99 | 0.41 | 0.67 | 0.93 | 0.91 | 0.99 | 0.58 | 0.84 |
| 25 | 25 | 0.78 | 0.54 | 1.00 | 0.00 | 0.32 | 0.78 | 0.52 | 1.00 | 0.00 | 0.44 |
| | 50 | 0.90 | 0.79 | 1.00 | 0.01 | 0.34 | 0.91 | 0.77 | 1.00 | 0.01 | 0.55 |
| | 100 | 0.93 | 0.89 | 0.99 | 0.15 | 0.46 | 0.92 | 0.86 | 0.99 | 0.32 | 0.75 |
| 50 | 50 | 0.88 | 0.71 | 1.00 | 0.00 | 0.14 | 0.90 | 0.70 | 1.00 | 0.00 | 0.29 |
| | 100 | 0.91 | 0.84 | 0.99 | 0.04 | 0.27 | 0.92 | 0.83 | 1.00 | 0.10 | 0.57 |
| 100 | 100 | 0.87 | 0.75 | 0.98 | 0.01 | 0.12 | 0.90 | 0.73 | 1.00 | 0.01 | 0.32 |

Table 3: Hit rates for $k = 0$ and case 3 for Pedroni's ADF type tests ($PP_{df}$, $PG_{df}$), for Westerlund's group-mean test ($WG$), the Larsson et al. test (LLL) and Breitung's test (B). The upper left panel displays the results for the baseline case, the upper right panel displays the results when an I(2) component is present, the lower left panel displays the results for the case with cross-sectional correlation ($\mathbf{\Sigma^{F}}$) and the lower right panel displays the results for the case with cross-unit cointegration.

Some illustrative results concerning the hit rates for $k = 0$ are displayed in Table 3 for case 3. This table shows in its four blocks ordered clockwise the results for the baseline case, the case of an I(2) component present in the data, short-run cross-sectional correlation and cross-unit cointegration. One of the strongest findings is that Westerlund's tests are undersized, an observation that also holds for some of Pedroni's tests that are not displayed in the table (in particular $PP_t$, $PG_t$ and $PG_\rho$). The ADF type tests are generally a bit oversized, with the exception of the $PG_{df}$ test for $T = 10$. The $PP_{df}$ test has hit rates between 0.9 and 0.95 for many of the experiments with the exceptions occurring mainly, as expected, for small $T$. The $PG_{df}$ test is generally a bit more oversized than its pooled counterpart. The $PP_{df}$ test is least affected by the non-baseline experiments considered. Westerlund's test $WG$ remains undersized in the presence of cross-sectional correlation and cross-unit cointegration but has its size going up and correspondingly its hit rates going down in the presence of an I(2) component to 0.11 for $T = N = 100$.

The system tests are strongly affected by the presence of an I(2) component, with the hit rates going down to zero in many instances. In the I(2) experiment Breitung's test leads mostly to the conclusion of a one-dimensional cointegrating space for $T \geq 25$ and for the $LLL$ test a three-dimensional cointegrating space is the most likely outcome. However, the LLL test performs very poorly already in the baseline case and consequently also in all additional robustness experiments. Generally also Breitung's test performs rather poorly for $N \geq 25$ for all values of $T$, albeit not as bad as the $LLL$ test. Compared to the performance in the baseline case, cross-sectional correlation and cross-unit cointegration do not contribute strongly to a further deterioration of the system tests' overall performance. The system tests have a general tendency, except for $T$ very small, to lead to an over-estimation of the dimension of the cointegrating space (see also Tables 4 and 5). Consequently, $k = 0$ is the situation where this tendency has the largest impact on the performance.

Figure 2 (see also Figure 5 in the appendix) displays the power of the single equation tests as a function of the stable root $q_{11}^i$, see the description in Table 1. Pedroni's tests $PP_{df}$ and $PG_{df}$ have the highest power throughout, which is substantially larger than the power of the other tests in many cases. For $q_{11}^i = 1.1$ and case 3 (see Figure 5 in the appendix) the two favorite Pedroni tests $PP_{df}$ and $PG_{df}$ are the only ones that have substantial power for $T = 100$ and $N$ large. For the smaller values of $T \leq 25$ not displayed, power is essentially zero for all values of $q_{11}^i$, with the exception of the $PG_{df}$ test whose power is positive in many
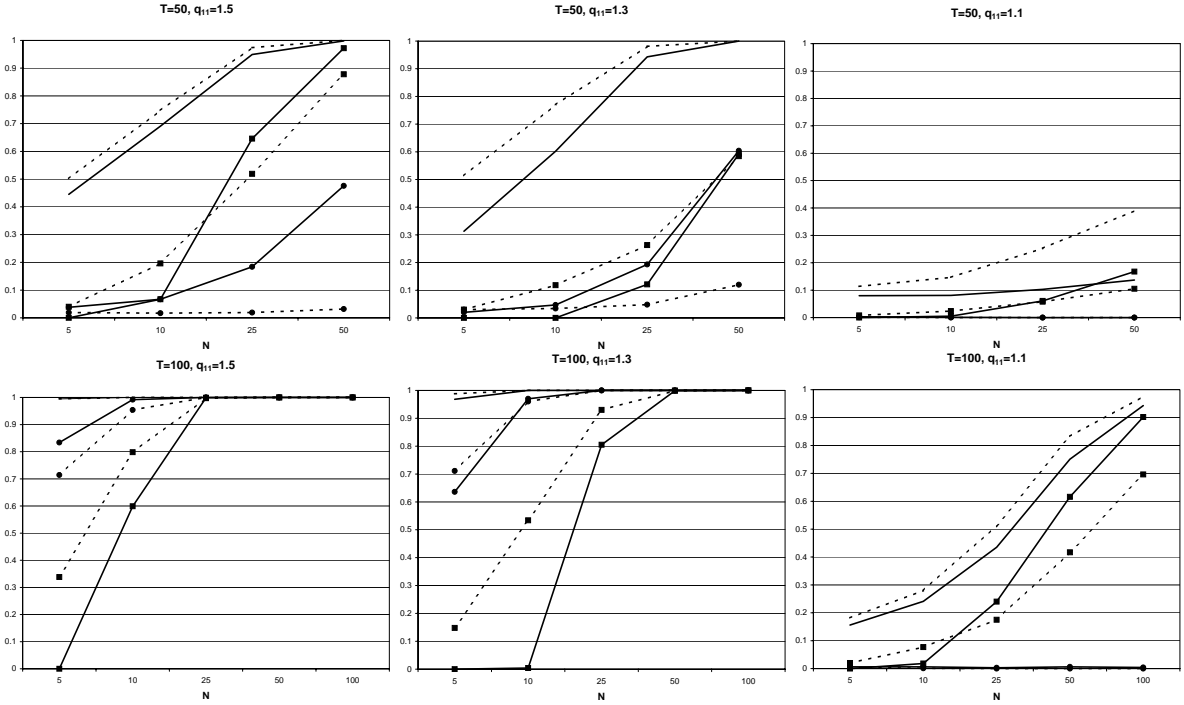
Figure 2: Power of single equation cointegration tests for case 2. The first column corresponds to $q_{11}^i = 1.5$, the second to $q_{11}^i = 1.3$ and the third to $q_{11}^i = 1.1$. The first row displays the results for $T = 50$ and the second displays the results for $T = 100$.

The solid line corresponds to $PP_{df}$, the dashed line to $PG_{df}$, the solid line with bullets to $PP_t$, the dashed line with bullets to $PG_t$, the solid line with squares to $WP$ and the dashed line with squares to $WG$.

circumstances from $T = 25$ onwards.

We now turn to a more detailed discussion of the two system tests. Table 4 displays the hit rates for case 2 and $k = 2$ and Table 5 in the appendix displays the results for case 3 and $k = 1$. Comparing the two tables already displays that there are sizeable differences in the performance between the different deterministic specifications. The baseline, the cross-sectional correlation and the cross-unit cointegration case are shown and Table 5 in the appendix displays in addition the results obtained in the presence of an I(2) component. As for $k = 0$ discussed above, the two tests exhibit quite different performance. Both tests often lead to an over-estimation of the dimension of the cointegrating space, a tendency that becomes more pronounced with increasing $N$ and even dominant for the Larsson, Lyhagen, and Löthgren (2001) test. For $T = 10$ the $LLL$ test sequence leads to the conclusion of a three-dimensional cointegrating space throughout all experiments, whereas Breitung's test leads to the conclusion of no cointegration. On the other hand, for $T = 100$ and $N = 5$ the

| | | | LLL | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Baseline | | | | Cross-sectional correlation | | | | Cross-unit cointegration | | | |
| N | T | 0 | 1 | 2 | 3 | 0 | 1 | 2 | 3 | 0 | 1 | 2 | 3 |
| 5 | 10 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| | 25 | 0.00 | 0.26 | 0.60 | 0.14 | 0.00 | 0.18 | 0.65 | 0.18 | 0.00 | 0.18 | 0.64 | 0.18 |
| | 50 | 0.00 | 0.10 | 0.76 | 0.14 | 0.00 | 0.05 | 0.78 | 0.17 | 0.00 | 0.05 | 0.74 | 0.21 |
| | 100 | 0.00 | 0.00 | 0.80 | 0.20 | 0.00 | 0.00 | 0.78 | 0.22 | 0.00 | 0.00 | 0.74 | 0.26 |
| 10 | 10 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| | 25 | 0.00 | 0.04 | 0.70 | 0.26 | 0.00 | 0.02 | 0.67 | 0.31 | 0.00 | 0.02 | 0.65 | 0.33 |
| | 50 | 0.00 | 0.00 | 0.75 | 0.25 | 0.00 | 0.00 | 0.71 | 0.29 | 0.00 | 0.00 | 0.70 | 0.30 |
| | 100 | 0.00 | 0.00 | 0.68 | 0.32 | 0.00 | 0.00 | 0.66 | 0.34 | 0.00 | 0.00 | 0.63 | 0.37 |
| 25 | 25 | 0.00 | 0.00 | 0.33 | 0.67 | 0.00 | 0.00 | 0.37 | 0.63 | 0.00 | 0.00 | 0.26 | 0.75 |
| | 50 | 0.00 | 0.00 | 0.43 | 0.57 | 0.00 | 0.00 | 0.46 | 0.54 | 0.00 | 0.00 | 0.37 | 0.63 |
| | 100 | 0.00 | 0.00 | 0.35 | 0.65 | 0.00 | 0.00 | 0.39 | 0.61 | 0.00 | 0.00 | 0.32 | 0.68 |
| 50 | 50 | 0.00 | 0.00 | 0.13 | 0.87 | 0.00 | 0.00 | 0.22 | 0.78 | 0.00 | 0.00 | 0.11 | 0.89 |
| | 100 | 0.00 | 0.00 | 0.09 | 0.92 | 0.00 | 0.00 | 0.18 | 0.82 | 0.00 | 0.00 | 0.08 | 0.92 |
| 100 | 100 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.06 | 0.94 | 0.00 | 0.00 | 0.00 | 1.00 |

| | | | B | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Baseline | | | | Cross-sectional correlation | | | | Cross-unit cointegration | | | |
| N | T | 0 | 1 | 2 | 3 | 0 | 1 | 2 | 3 | 0 | 1 | 2 | 3 |
| 5 | 10 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| | 25 | 0.51 | 0.39 | 0.09 | 0.01 | 0.36 | 0.40 | 0.18 | 0.07 | 0.37 | 0.59 | 0.04 | 0.00 |
| | 50 | 0.00 | 0.35 | 0.60 | 0.05 | 0.00 | 0.28 | 0.45 | 0.27 | 0.00 | 0.71 | 0.26 | 0.03 |
| | 100 | 0.00 | 0.01 | 0.89 | 0.10 | 0.00 | 0.07 | 0.61 | 0.33 | 0.00 | 0.32 | 0.59 | 0.09 |
| 10 | 10 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| | 25 | 0.14 | 0.72 | 0.13 | 0.01 | 0.09 | 0.46 | 0.33 | 0.13 | 0.09 | 0.84 | 0.07 | 0.00 |
| | 50 | 0.00 | 0.22 | 0.72 | 0.07 | 0.00 | 0.08 | 0.53 | 0.39 | 0.00 | 0.47 | 0.50 | 0.03 |
| | 100 | 0.00 | 0.00 | 0.88 | 0.12 | 0.00 | 0.00 | 0.54 | 0.46 | 0.00 | 0.02 | 0.87 | 0.11 |
| 25 | 25 | 0.00 | 0.77 | 0.21 | 0.02 | 0.00 | 0.14 | 0.48 | 0.38 | 0.00 | 0.83 | 0.16 | 0.01 |
| | 50 | 0.00 | 0.03 | 0.86 | 0.12 | 0.00 | 0.00 | 0.30 | 0.70 | 0.00 | 0.07 | 0.86 | 0.07 |
| | 100 | 0.00 | 0.00 | 0.84 | 0.16 | 0.00 | 0.00 | 0.27 | 0.73 | 0.00 | 0.00 | 0.89 | 0.11 |
| 50 | 50 | 0.00 | 0.00 | 0.83 | 0.17 | 0.00 | 0.00 | 0.19 | 0.81 | 0.00 | 0.00 | 0.87 | 0.13 |
| | 100 | 0.00 | 0.00 | 0.77 | 0.23 | 0.00 | 0.00 | 0.17 | 0.83 | 0.00 | 0.00 | 0.85 | 0.15 |
| 100 | 100 | 0.00 | 0.00 | 0.72 | 0.28 | 0.00 | 0.00 | 0.04 | 0.96 | 0.00 | 0.00 | 0.74 | 0.26 |

Table 4: Hit rates for $k = 2$ and case 2 for the Larsson et al. test (LLL) in the upper panel and Breitung's test (B) in the lower panel. The first column (from the left) displays the results for the baseline case, the second column displays the results for the case with cross-sectional correlation ($\mathbf{\Sigma^F}$) and the third column displays the results for the case with cross-unit cointegration.

behavior of both tests is, as expected, in the vicinity of the Johansen test when performed on the same DGP with $N = 1$ (which leads to a hit rate of about 0.8). The performance of both system tests is generally deteriorating (at least slightly) with increasing $N$. Note for completeness that the system tests are less sensitive than the single equation tests with respect to stable roots approaching the unit circle.

Cross-sectional correlation has a stronger detrimental impact on the performance of the system tests than the presence of one additional cross-unit cointegrating vector, whose effect tends to become less pronounced for increasing $N$. The effect of cross-sectional correlation tends to become stronger for increasing $N$ and leads for both test sequences generally to the conclusion that the dimension of the cointegrating space is equal to 3. These findings hold for all considered values of $k = 0, 1, 2$ and also across the deterministic specifications. The presence of an I(2) component has a rather strong negative impact on the system test performance, as discussed for $k = 0$ above and as can be seen in Table 5 for $k = 1$ in the appendix.

The over-estimation of the cointegrating space (increasing in $N$) is most likely due to the use of asymptotic correction factors for both tests. Therefore, using either finite sample critical values in the spirit of Im, Pesaran, and Shin (2003) or resorting to bootstrap inference as laid out for the time series case in Park (2002) and in Swensen (2006) for VAR systems may lead to improved performance. For the single equation tests of Pedroni the use of asymptotic correction factors seems to be much less influential, which partly is also due to the fact that in that case only one test decision is made and not a test sequence.

The finding that both cross-sectional correlation and the additional cross-unit cointegrating relationship do not exert a devastatingly negative impact on the performance of the tests compared to the baseline shows that obtaining detailed understanding of their properties in case of DGPs that exhibit these features is very important.

## 3.2 The Performance of the Estimators

We now turn to an assessment of the approximation quality of the estimated cointegrating spaces to the true cointegrating spaces. The measure of quality employed is the *gap* between the true and the estimated cointegrating space. The gap is defined as follows: Let $M$ and $N$

denote two linear subspaces of $\mathbb{R}^s$, then the gap $d(M, N)$ is given by

$$d(M, N) = \max \left( \sup_{x \in M, \|x\|=1} \| (I - Q)x \|, \sup_{x \in N, \|x\|=1} \| (I - P)x \| \right), \qquad (40)$$

where $Q$ denotes the orthogonal projection onto $N$, $P$ the orthogonal projection onto $M$ and $\| x \|$ denotes the Euclidean norm on $\mathbb{R}^s$. The gap is between zero and one and it is e.g. equal to one for spaces of different dimensions. As an illustration, in $\mathbb{R}^2$ the gap between two vectors with a five degree angle between them is about 0.087 (and the logarithm – because we display the logarithms of the gaps in the figures – is about -3.44).

We present results concerning four different estimators: two single equation estimators (only for $k = 1$) and two system estimators. The two reported single equation estimators are the FM-OLS estimator in the formulation of Phillips and Moon (1999), see equation (14) and the definition of the averaged correction factors below, and the D-OLS estimator as described in (18) and (19). The window lengths for FM-OLS estimation are chosen according to Newey and West (1987) and the lead and lag length selection for D-OLS is performed individual specifically using `AIC` in (18). The two systems estimators are the one-step estimator (as in the discussion above equation (29)) and Breitung's two-step estimator. We include the one-step VAR estimator to gauge the effect of using a panel estimator for $\beta_i = \beta$ for $i = 1, \ldots, N$ as opposed to simply taking the average over the cross-sectional units (which corresponds to the group-mean versions of the FM-OLS and D-OLS estimators discussed above). The VAR lag lengths are chosen individual specifically according to `AIC` in (22).

We have also implemented the other described variants of both the FM-OLS and the D-OLS estimators, normalized to have (up to the scaling factor) standard normal asymptotic distribution and the group-mean estimators. These estimators all perform worse than the versions for which the results are displayed below. For most of the experiments it turns out that the effect of the construction principle of the estimators, i.e. normalized or group-mean, is larger than the effect of using fully modified or dynamic OLS estimation. Especially the two group-mean estimators behave often similarly and perform very poorly for small $T$. The group-mean estimators are also the most sensitive ones with respect to stable roots approaching the unit circle, i.e. $q_{11}^i$ tending to 1, a feature that is shared by the group-mean VAR estimator (see below). Also the normalized estimators are more sensitive in this respect than the versions discussed in this section.

We report the results in the form of density estimates of the logarithms of the gaps between

the estimated and the true cointegrating spaces. The density estimates are based on normal kernels and bandwidths chosen according to Silverman's rule of thumb. Logarithms are taken to increase the variability because consistency of the estimators implies that for increasing sample sizes the gaps tend to zero. In all figures displaying density plots the dashed line corresponds to the FM-OLS estimator, the dotted line to the D-OLS estimator (both only for $k = 1$), the grey solid line to the one-step or group-mean VAR estimator and the black solid line to the two-step VAR estimator of Breitung.

We start our discussion with the case of cointegrating spaces of dimension one. Figure 3 displays the performance of the four estimators for the baseline case of cross-sectionally independent processes (with $q_{11}^i = 1.5$). This figure already shows some of the main findings concerning the properties of the estimators (also for experiments and panel sizes not displayed). First, the D-OLS estimator performs best across a large variety of experiments, with its relative performance even improving with increasing sample size in many cases. Second, going from one-step (i.e. mere averaging of the cross-section specific Johansen estimates) to Breitung's two-step estimator generally leads to a large improvement. The cross-sectional average of Johansen estimates (the group-mean VAR estimator) performs – as expected – especially poor for small values of $T$. The performance of this estimator is improving with a larger cross-sectional dimension. As mentioned, the feature of extremely bad performance for small $T$ is shared by the group-mean FM-OLS (17) and D-OLS (21) estimators. Third, for a variety of our experiments FM-OLS and Breitung's two-step estimator show comparable performance. However, there is also a considerable amount of experiments in which Breitung's estimator outperforms FM-OLS. Fourth, for small values of $T$ (up to about 25) and $N$ (up to 10) the system estimators often perform worse than the single equation estimators. Fifth, as illustrated by Figure 6 in the appendix, the two single equation estimators are less affected by the stable root $q_{11}^i$ tending to one than the system estimators. Amongst all estimators the one-step estimator is affected most strongly by stable roots approaching the unit circle, which essentially reflects the sensitivity of the Johansen estimator with respect to stable roots approaching the unit circle in the time series case.
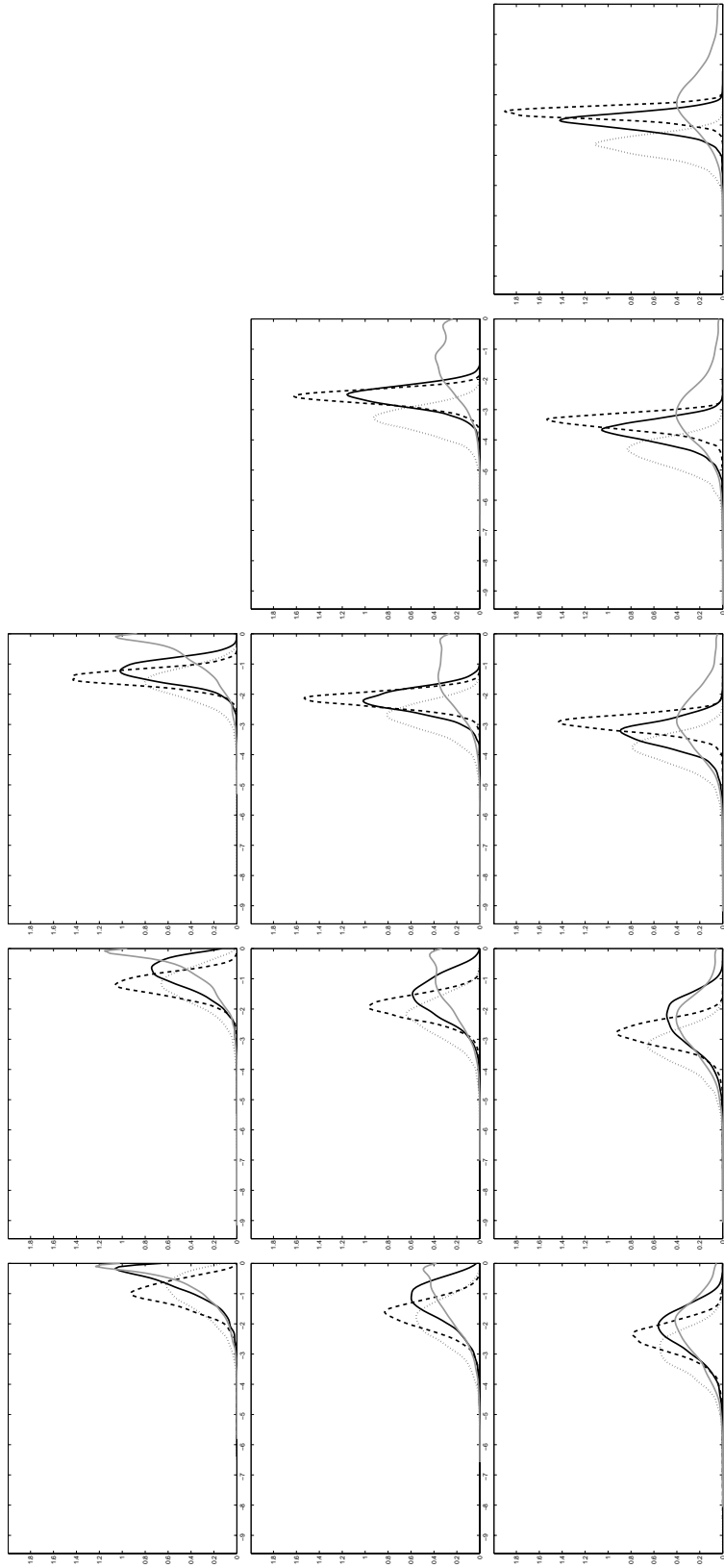
Figure 3: Density plots of the logarithms of the gaps between estimated and true cointegrating spaces for $k = 1$ and cross-sectionally independent processes $Y_{it}$. The graphs display the results for $q_{11}^i = 1.5$ and case 3 of the deterministic components. The rows correspond to $T = 25, 50, 100$ and the columns to $N = 5, 10, 25, 50, 100$.

The good performance of D-OLS in our simulations is potentially due to the fact that for our DGPs the true unobserved errors of the single equation cointegrating regression (1) are AR processes and the regressors in (1) are ARMA processes (see again Zellner and Palm, 1974). Therefore, the improving relative performance of D-OLS over FM-OLS for increasing $T$ shows that the serial correlation and endogeneity correction by augmenting the cointegrating regressions with leads and lags improves faster than the corrections based on non-parametric spectral estimates.
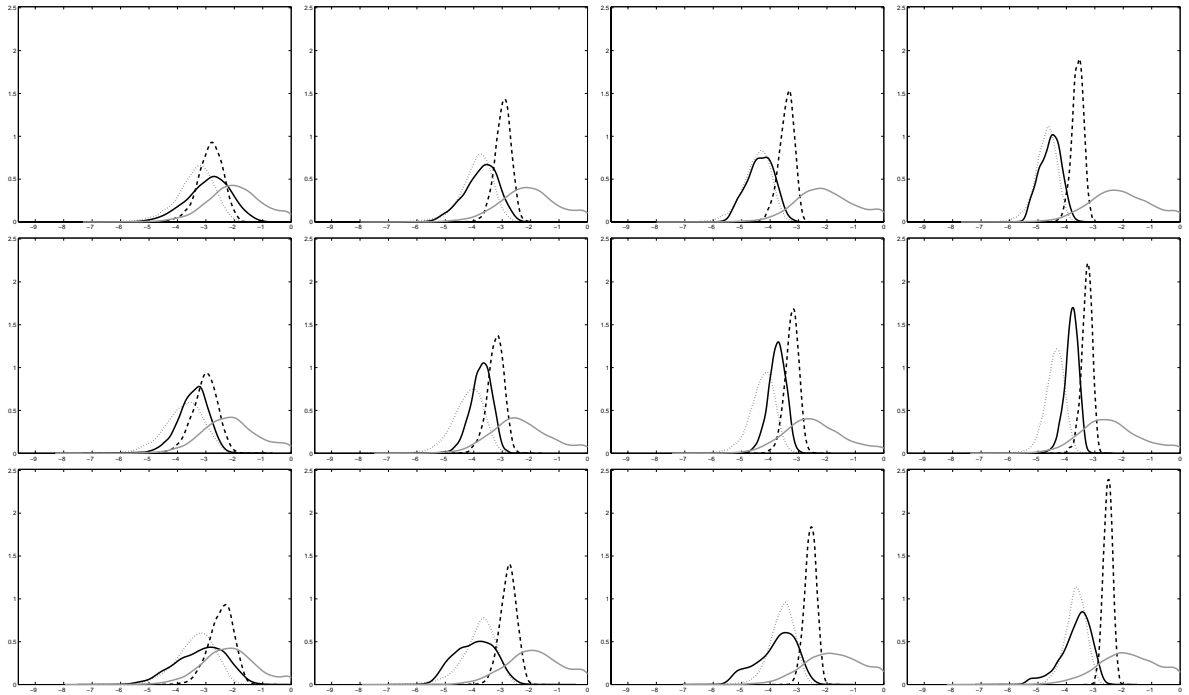


Figure 4: Density plots of the logarithms of the gaps between estimated and true cointegrating spaces for $k = 1$, $q_{11}^i = 1.5$, case 4 and $T = 100$. The columns correspond to $N = 10, 25, 50, 100$. The first row displays the results for cross-sectionally independent processes, the second row displays the results for cross-sectionally correlated ($\mathbf{\Sigma^F}$) processes and the third row displays the results for the case with cross-unit cointegration.

It is quite well known from many simulation studies that VAR cointegration analysis leads to inaccurate estimation of the cointegrating spaces despite super-consistency even for time series of lengths usually available in macroeconomic applications, see Bauer and Wagner (2006) and Wagner (2004). What is interesting to note in this respect is that D-OLS outperforms or has at least comparable performance as Breitung's estimator even for the

largest panels in which also VAR estimation (given that the data are generated according to VAR(2) processes) should not suffer from degrees of freedom problems too strongly. FM-OLS estimation on the other hand seems to suffer even for the larger panels from the 'imprecision' in the estimated correction factors. This better performance of parametric methods is not necessarily confined to a situation where the true DGP follows an ARMA process, as even in case of a non-rational transfer function it may well be the case that the spectrum is better approximated by a rational function (i.e. an AR or ARMA model) in small (or finite) samples than by a non-parametric estimate (see Chapter 7.4 in Hannan and Deistler, 1988).

In Figure 4 we illustrate the impact of cross-sectional correlation (via the factor model for the error processes) and cross-unit cointegration on the performance of the estimators. The performance of all estimators deteriorates. The one-step VAR estimator is affected most, with large deterioration for $T$ up to 50. For the other estimators the impact of short-run cross-sectional correlation is not too large, which holds true also for other experiments. The detrimental impact of cross-unit cointegration on the estimators is slightly larger, but the impact is generally quite limited. However, as discussed, the effects may well be much stronger for other specifications of cross-unit cointegration, with more cross-unit cointegrating relationships that are not so clearly separated from the individual specific cointegrating relationships. Further understanding of the impacts of cross-unit cointegration is a key open issue for future research in this area.

For our specifications of cross-sectional correlation Breitung's two-step estimator as well as the D-OLS estimator theoretically suffer *at least* from efficiency losses, because feasible GLS (implementable for $T$ large enough compared to $N$) is an estimation technique to incorporate the cross-sectional correlation structure, under appropriate assumptions under which consistency can be established. The fact that FM-OLS estimation performs worse, must be driven by the finite sample properties of the non-parametric estimation of the correction factors, as the limiting distributions are identical e.g. for the normalized versions of the estimators.

Again the effect of cross-unit cointegration is relatively small, which is as discussed related to the set-up chosen to study the effects of cross-unit cointegration in a situation, where we simply add one additional cointegrating relationship and keep (cross-sectionally identical) within-unit cointegrating relationships. With some few exceptions, as illustrated in Figure 4 for case 4, where Breitung's two-step procedure performs a bit better, D-OLS remains the best performing method overall.

The results for two dimensional cointegrating spaces are very similar to the ones obtained for $k = 1$, see Figure 7 in the appendix for some illustrative results. The group-mean or one-step VAR estimator performs very poor for small values of $T$. Using the two-step procedure again leads to performance gains. Similarly to the time series case, the precision of the estimation of the cointegrating space decreases with increasing dimension of the cointegrating space. As for $k = 1$ the impact of short-run cross-sectional correlation and of one cross-unit cointegrating relationship is very small. The somewhat surprising result is that the cointegrating spaces are estimated with slightly higher precision in these two cases than in the baseline case.

The simulation results obtained for the studied cases of cross-sectional correlation and cross-unit cointegration imply that it may indeed be a fruitful task to study the asymptotic behavior (i.e. consistency and asymptotic distribution) of the considered simple estimators also for cross-sectionally dependent panels to gain an understanding about situations under which consistency prevails. Positive results in this respect would greatly improve the applicability and usefulness of the discussed estimators or appropriate extensions (see e.g. Bai and Kao, 2005).

## 4   Conclusions

Under the premise that the results from simulation studies have to be interpreted carefully and that one should be cautious with generalizations, some relatively clear observations emerge from our results.

Amongst the single equation tests for the null hypothesis of no cointegration the two tests of Pedroni applying the ADF principle perform best, whereas all other tests are partly severely undersized and have very low power in many circumstances (and virtually none for $T \leq 25$). Pedroni's $PP_{df}$ and $PG_{df}$ tests are also the ones least affected by the presence of an I(2) component, short-run cross-sectional correlation or cross-unit cointegration of the form considered. Both of Westerlund's tests are severely undersized. These findings highlight that the use of finite sample correction factors (in the spirit of Im, Pesaran, and Shin, 2003) or bootstrap inference may have strong beneficial effects on the performance of the tests. Taking additionally into account that for the case of no cointegration the mentioned tests of Pedroni outperform the system tests (which exhibit a tendency to conclude a too high-dimensional cointegrating space), we conclude that in a situation where the null hypothesis

of no cointegration is of particular relevance or importance, these two tests of Pedroni are the first choice.

The system tests perform – as expected – very poor for small values of $T$, but also suffer partly from problems when $N$ is *too* large. Even more than for the single equation tests, because a whole test sequence and hence several comparisons with critical values have to be performed, the use of finite sample correction factors or bootstrap inference may be fruitful. Especially the Larsson, Lyhagen, and Löthgren (2001) test often leads to the conclusion of a three-dimensional (i.e. maximal dimensional) cointegrating space, but also the test sequence of Breitung (2005) has a tendency to result in a too high-dimensional cointegrating space. Both system tests are sensitive with respect to the presence of an I(2) component (a feature less prominent for the single equation tests) and are not very sensitive with respect to stable autoregressive roots approaching the unit circle (which affects the single equation tests more strongly). The studied forms of cross-sectional correlation and cross-unit cointegration do not lead to a sizeable deterioration of the tests' performance compared to the baseline case. These findings imply that obtaining a detailed understanding of the methods' performance for cross-sectionally dependent panels is an important question for further research.

In the case of one-dimensional cointegrating spaces, the D-OLS estimator in the version of equation (19) outperforms all other estimators, both single equation and system estimators, even for large samples. The D-OLS estimator is also the least sensitive estimator with respect to the discussed additional experiments (stable root approaching the unit circle, I(2) component, cross-sectional correlation, cross-unit cointegration). The version of the FM-OLS estimator as given by (14) and cross-sectionally averaged correction factors appears to suffer from imprecise estimation of the required correction factors, which however could be related to the fact that all the individual series in our simulations are ARMA processes (given that they jointly form a VAR process). For small values of $T \leq 25$ and small $N \leq 10$ the system estimators are in many cases outperformed by the single equation estimators. For larger samples the FM-OLS estimator performs in many occasions comparable to the Breitung (2005) estimator. The (normalized) D-OLS and FM-OLS estimators based on individual specific correction factors and even more the group-mean versions of D-OLS and FM-OLS perform worse than their counterparts based on averaged correction factors. Especially the group-mean versions are very sensitive with respect to short time series dimension or stable roots approaching the unit circle. The latter is a feature shared by the one-step or group-mean

37

VAR estimator. Throughout the experiments applying Breitung's two-step estimator leads to large improvements over the one-step estimator, with the advantages being generally more pronounced in case of two-dimensional cointegrating spaces.

## Acknowledgements

## References

Ahn, S.K. and G.C. Reinsel (1990). Estimation for Partially Nonstationary Multivariate Autoregressive Models. *Journal of the American Statistical Association* **85**, 813–823.

Andrews, D.W.K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica* **59**, 817–858.

Bai, J. and S. Ng (2004). A PANIC Attack on Unit Roots and Cointegration. *Econometrica* **72**, 1127–1178.

Bai, J. and C. Kao (2005). On the Estimation and Inference of a Panel Cointegration Model with Cross-Sectional Dependence. Mimeo.

Banerjee, A. and J.L. Carrion-i-Silvestre (2006). Cointegration in Panel Data with Breaks and Cross-Section Dependence. ECB Working Paper No. 591.

Bauer, D. and M. Wagner (2006). The Performance of Subspace Algorithm Cointegration Analysis: A Simulation Study. Mimeo.

Breitung, J. (2002). Nonparametric Tests for Unit Root and Cointegration. *Journal of Econometrics* **108**, 343–363.

Breitung, J. (2005). A Parametric Approach to the Estimation of Cointegration Vectors in Panel Data. *Econometric Reviews* **24**, 151–173.

Engle, R.F. and C.W.J. Granger (1987). Cointegration and Error Correction: Representation, Estimation and Testing. *Econometrica* **55**, 251–276.

Engle, R.F. and B.S. Yoo (1991). Cointegrated Economic Time Series: An Overview with New Results. In Fomby, T.B. and G.F. Rhodes (eds.) *Advances in Econometrics: Co-Integration, Spurious Regression, and Unit Roots*. JAI Press, Greenwich (CT), 237–266.

Groen, J.J.J. and F. Kleibergen (2003). Likelihood-Based Cointegration Analysis in Panels of Vector Error Correction Models. *Journal of Business and Economic Statistics* **21**, 295–318.

Hadri, K. (2000). Testing for Stationarity in Heterogeneous Panel Data. *Econometrics Journal* **3**, 148–161.

Hannan, E.J. and M. Deistler (1988). *The Statistical Theory of Linear Systems*. John Wiley and Sons, New York.

Hlouskova, J. and M. Wagner (2006). The Performance of Panel Unit Root and Stationarity Tests: Results from a Large Scale Simulation Study. *Econometric Reviews* **25**, 85–116.

Horowitz, J.L. and E.N. Savin (2000). Empirically Relevant Critical Values for Hypothesis Tests: A Bootstrap Approach. *Journal of Econometrics* **95**, 375–389.

Im, K.S., M.H. Pesaran, and Y. Shin (2003). Testing for Unit Roots in Heterogeneous Panels. *Journal of Econometrics* **115**, 53–74.

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

Kao, C. (1999). Spurious Regression and Residual Based Tests for Cointegration in Panel Data. *Journal of Econometrics* **90**, 1–44.

Kao, C. and M.-H. Chiang (2000). On the Estimation and Inference of a Cointegrated Regression in Panel Data. In Baltagi, B.H. (Ed.) *Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, Elsevier, Amsterdam.

Kapetanios, G., M.H. Pesaran, and T. Yamagata (2006). Panels with Nonstationary Multi-factor Error Structure. Mimeo.

Larsson, R. and J. Lyhagen (1999). Likelihood-Based Inference in Multivariate Panel Cointegration Models. Working Paper Series in Economics and Finance 331, Stochholm School of Economics.

Larsson, R., J. Lyhagen, and M. Löthgren (2001). Likelihood-Based Cointegration Tests in Heterogeneous Panels. *Econometrics Journal* **4**, 109–142.

Mark, N.C. and D. Sul (2003). Cointegration Vector Estimation by Panel Dynamic OLS and Long-Run Money Demand. *Oxford Bulletin of Economics and Statistics* **65**, 655–680.

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

Newey, W. and K. West (1987). A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator. *Econometrica* **50**, 703–708.

Park, J.Y. (2002). An Invariance Principle for Sieve Bootstrap in Time Series. *Econometric Theory* **18**, 469–490.

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

Pedroni, P. (2000). Fully Modified OLS for Heterogeneous Cointegrated Panels. In Baltagi, B.H. (Ed.) *Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, Elsevier, Amsterdam.

Pedroni, P. (2001). Purchasing Power Parity Tests in Cointegrated Panels. *Review of Economics and Statistics* **83**, 1371–1375.

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.

Pesaran, H.M., T. Schuermann, and S.M. Weiner (2004). Modelling Regional Interdependencies Using a Global Error-Correcting Macroeconometric Model. *Journal of Business and Economic Statistics* **22**, 129–162.

Phillips, P.C.B. and B.E. Hansen (1990). Statistical Inference in Instrumental Variables Regression with I(1) Processes. *Review of Economic Studies* **57**, 99–125.

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

Phillips, P.C.B. and S. Ouliaris (1990). Asymptotic Properties of Residual Based Tests for Cointegration. *Econometrica* **58**, 165–193.

Phillips, P.C.B. and P. Perron (1988). Testing for a Unit Root in Time Series Regression. *Biometrika* **75**, 335–346.

Saikkonen, P. (1991). Asymptotically Efficient Estimation of Cointegrating Regressions. *Econometric Theory* **7**, 1–21.

Saikkonen, P. (1992). Estimation and Testing of Cointegrated Systems by an Autoregressive Approximation. *Econometric Theory* **8**, 1–27.

Saikkonen, P. (1999). Testing the Normalization and Overidentification of Cointegrating Vectors in Vector Autoregressive Processes. *Econometric Reviews* **18**, 235–257.

Saikkonen, P. and R. Luukkonen (1997). Testing Cointegration in Infinite Order Vector Autoregressive Processes. *Journal of Econometrics* **81**, 93–126.

Swensen, A.R. (2006). Bootstrap Algorithms for Testing and Determining the Cointegration Rank in VAR Models. *Econometrica* **74**, 1699–1714.

Wagner, M. (2004). A Comparison of Johansen's, Bierens' and the Subspace Algorithm Method for Cointegration Analysis. *Oxford Bulletin of Economics and Statistics* **66**, 399–424.

Westerlund, J. (2005). New Simple Tests for Panel Cointegration. *Econometric Reviews* **24**, 297–316.

Westerlund, J. and D. Edgerton (2006). Simple Tests for Cointegration in Dependent Panels with Structural Breaks. Mimeo.

Zellner, A. and F.C. Palm (1974). Time Series Analysis and Simultaneous Equation Econometric Models. *Journal of Econometrics* **2**, 17–54.
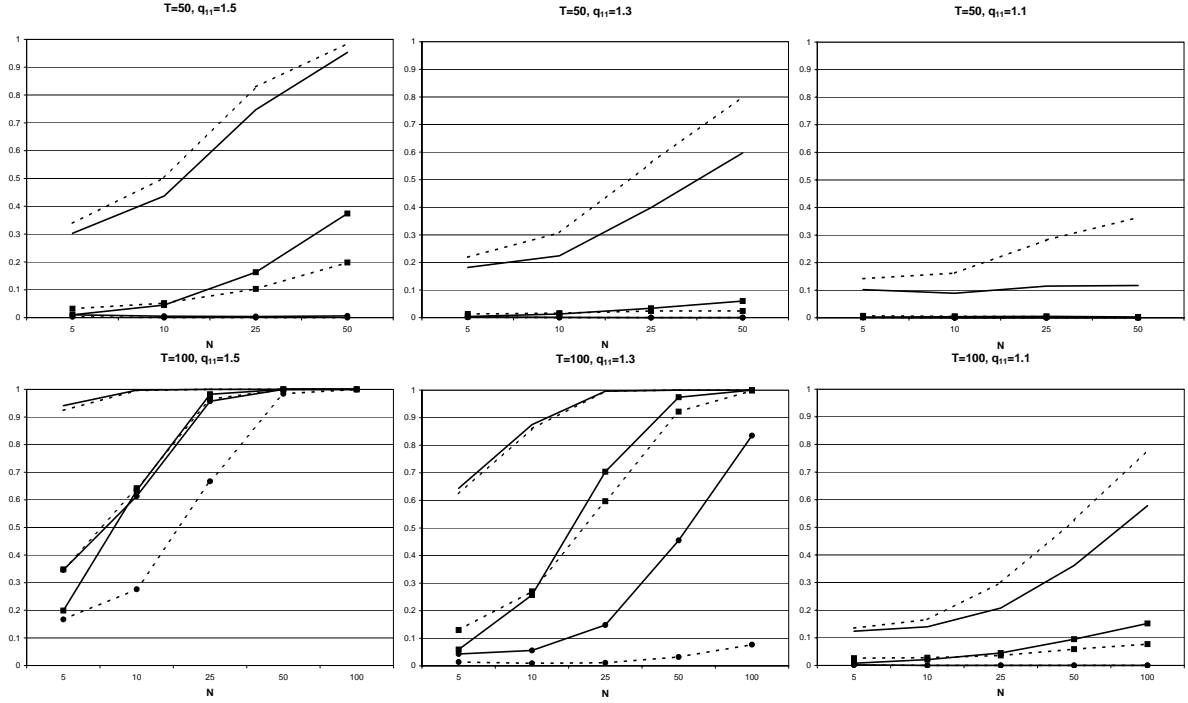
# Appendix: Additional Figures and Tables



Figure 5: Power of single equation cointegration tests for case 3. The first column corresponds to $q_{11}^i = 1.5$, the second to $q_{11}^i = 1.3$ and the third to $q_{11}^i = 1.1$. The first row displays the results for $T = 50$ and the second displays the results for $T = 100$.

The solid line corresponds to $PP_{df}$, the dashed line to $PG_{df}$, the solid line with bullets to $PP_t$, the dashed line with bullets to $PG_t$, the solid line with squares to $WP$ and the dashed line with squares to $WG$.

**LLL**

| N | T | Baseline 0 | 1 | 2 | 3 | I(2) Component 0 | 1 | 2 | 3 | Cross-sectional correlation 0 | 1 | 2 | 3 | Cross-unit cointegration 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 10 | 0.00 | 0.00 | 0.02 | 0.98 | 0.00 | 0.00 | 0.02 | 0.98 | 0.00 | 0.00 | 0.02 | 0.98 | 0.00 | 0.00 | 0.02 | 0.98 |
|   | 25 | 0.02 | 0.47 | 0.16 | 0.35 | 0.00 | 0.12 | 0.25 | 0.63 | 0.01 | 0.38 | 0.15 | 0.45 | 0.02 | 0.48 | 0.17 | 0.33 |
|   | 50 | 0.03 | 0.71 | 0.12 | 0.15 | 0.00 | 0.23 | 0.30 | 0.47 | 0.01 | 0.56 | 0.11 | 0.32 | 0.03 | 0.70 | 0.12 | 0.15 |
|   | 100 | 0.00 | 0.80 | 0.12 | 0.07 | 0.00 | 0.32 | 0.33 | 0.35 | 0.00 | 0.61 | 0.12 | 0.27 | 0.00 | 0.80 | 0.12 | 0.08 |
| 10 | 10 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 |
|   | 25 | 0.00 | 0.23 | 0.19 | 0.58 | 0.00 | 0.01 | 0.13 | 0.87 | 0.00 | 0.15 | 0.13 | 0.72 | 0.00 | 0.25 | 0.21 | 0.54 |
|   | 50 | 0.00 | 0.59 | 0.18 | 0.24 | 0.00 | 0.04 | 0.22 | 0.74 | 0.00 | 0.36 | 0.10 | 0.54 | 0.00 | 0.61 | 0.17 | 0.22 |
|   | 100 | 0.00 | 0.74 | 0.18 | 0.09 | 0.00 | 0.07 | 0.34 | 0.59 | 0.00 | 0.43 | 0.12 | 0.44 | 0.00 | 0.75 | 0.16 | 0.09 |
| 25 | 25 | 0.00 | 0.01 | 0.08 | 0.91 | 0.00 | 0.00 | 0.01 | 0.99 | 0.00 | 0.01 | 0.03 | 0.97 | 0.00 | 0.04 | 0.10 | 0.86 |
|   | 50 | 0.00 | 0.24 | 0.26 | 0.50 | 0.00 | 0.00 | 0.04 | 0.96 | 0.00 | 0.06 | 0.06 | 0.87 | 0.00 | 0.44 | 0.21 | 0.35 |
|   | 100 | 0.00 | 0.54 | 0.32 | 0.14 | 0.00 | 0.00 | 0.14 | 0.86 | 0.00 | 0.15 | 0.08 | 0.77 | 0.00 | 0.64 | 0.26 | 0.10 |
| 50 | 50 | 0.00 | 0.03 | 0.17 | 0.80 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.01 | 0.02 | 0.98 | 0.00 | 0.18 | 0.19 | 0.63 |
|   | 100 | 0.00 | 0.25 | 0.46 | 0.29 | 0.00 | 0.00 | 0.03 | 0.97 | 0.00 | 0.04 | 0.04 | 0.93 | 0.00 | 0.44 | 0.38 | 0.18 |
| 100 | 100 | 0.00 | 0.06 | 0.44 | 0.50 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.01 | 0.02 | 0.98 | 0.00 | 0.17 | 0.49 | 0.34 |

**B**

| N | T | Baseline 0 | 1 | 2 | 3 | I(2) Component 0 | 1 | 2 | 3 | Cross-sectional correlation 0 | 1 | 2 | 3 | Cross-unit cointegration 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 10 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
|   | 25 | 0.76 | 0.18 | 0.03 | 0.03 | 0.24 | 0.65 | 0.11 | 0.00 | 0.67 | 0.23 | 0.05 | 0.06 | 0.75 | 0.17 | 0.05 | 0.03 |
|   | 50 | 0.27 | 0.59 | 0.09 | 0.04 | 0.00 | 0.77 | 0.23 | 0.00 | 0.14 | 0.62 | 0.11 | 0.13 | 0.29 | 0.56 | 0.11 | 0.04 |
|   | 100 | 0.00 | 0.88 | 0.09 | 0.03 | 0.00 | 0.75 | 0.25 | 0.00 | 0.00 | 0.72 | 0.12 | 0.16 | 0.00 | 0.87 | 0.11 | 0.02 |
| 10 | 10 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
|   | 25 | 0.50 | 0.42 | 0.05 | 0.04 | 0.02 | 0.87 | 0.11 | 0.00 | 0.38 | 0.44 | 0.08 | 0.10 | 0.55 | 0.34 | 0.07 | 0.04 |
|   | 50 | 0.03 | 0.82 | 0.10 | 0.06 | 0.00 | 0.68 | 0.32 | 0.00 | 0.01 | 0.62 | 0.15 | 0.23 | 0.07 | 0.73 | 0.16 | 0.04 |
|   | 100 | 0.00 | 0.86 | 0.12 | 0.03 | 0.00 | 0.59 | 0.41 | 0.00 | 0.00 | 0.59 | 0.16 | 0.26 | 0.00 | 0.85 | 0.14 | 0.01 |
| 25 | 25 | 0.06 | 0.77 | 0.13 | 0.05 | 0.00 | 0.86 | 0.15 | 0.00 | 0.04 | 0.55 | 0.14 | 0.28 | 0.20 | 0.66 | 0.10 | 0.04 |
|   | 50 | 0.00 | 0.70 | 0.26 | 0.04 | 0.00 | 0.53 | 0.47 | 0.00 | 0.00 | 0.31 | 0.18 | 0.51 | 0.00 | 0.73 | 0.25 | 0.02 |
|   | 100 | 0.00 | 0.77 | 0.22 | 0.01 | 0.00 | 0.42 | 0.58 | 0.00 | 0.00 | 0.31 | 0.16 | 0.52 | 0.00 | 0.80 | 0.20 | 0.00 |
| 50 | 50 | 0.00 | 0.51 | 0.46 | 0.03 | 0.00 | 0.15 | 0.85 | 0.00 | 0.00 | 0.10 | 0.07 | 0.83 | 0.00 | 0.58 | 0.40 | 0.02 |
|   | 100 | 0.00 | 0.61 | 0.39 | 0.00 | 0.00 | 0.12 | 0.88 | 0.00 | 0.00 | 0.12 | 0.06 | 0.82 | 0.00 | 0.68 | 0.32 | 0.00 |
| 100 | 100 | 0.00 | 0.34 | 0.65 | 0.00 | 0.00 | 0.09 | 0.91 | 0.00 | 0.00 | 0.03 | 0.02 | 0.95 | 0.00 | 0.49 | 0.52 | 0.00 |

Table 5: Hit rates for $k = 1$ and case 3 for the Larsson et al. test (LLL) in the upper panel and Breitung's test (B) in the lower panel. The first column (from the left) displays the results for the baseline case, the second column displays the results when an I(2) component is present, the third column displays the results for the case with cross-sectional correlation and the forth column displays the results for the case with cross-unit cointegration.
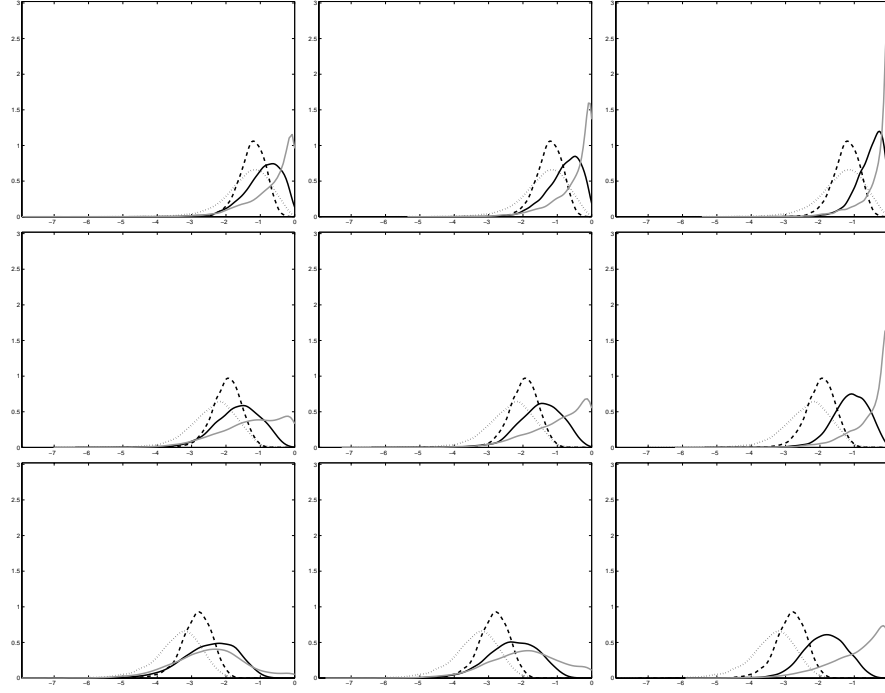
Figure 6: Density plots of the logarithms of the gaps between estimated and true cointegrating spaces for $k = 1$, $q^i_{12} = 1.5$, case 3 and $N = 10$. The time series dimension is $T = 25, 50, 100$ from top to bottom. The columns correspond to $q^i_{11} = 1.5, 1.3, 1.1$ from left to right.
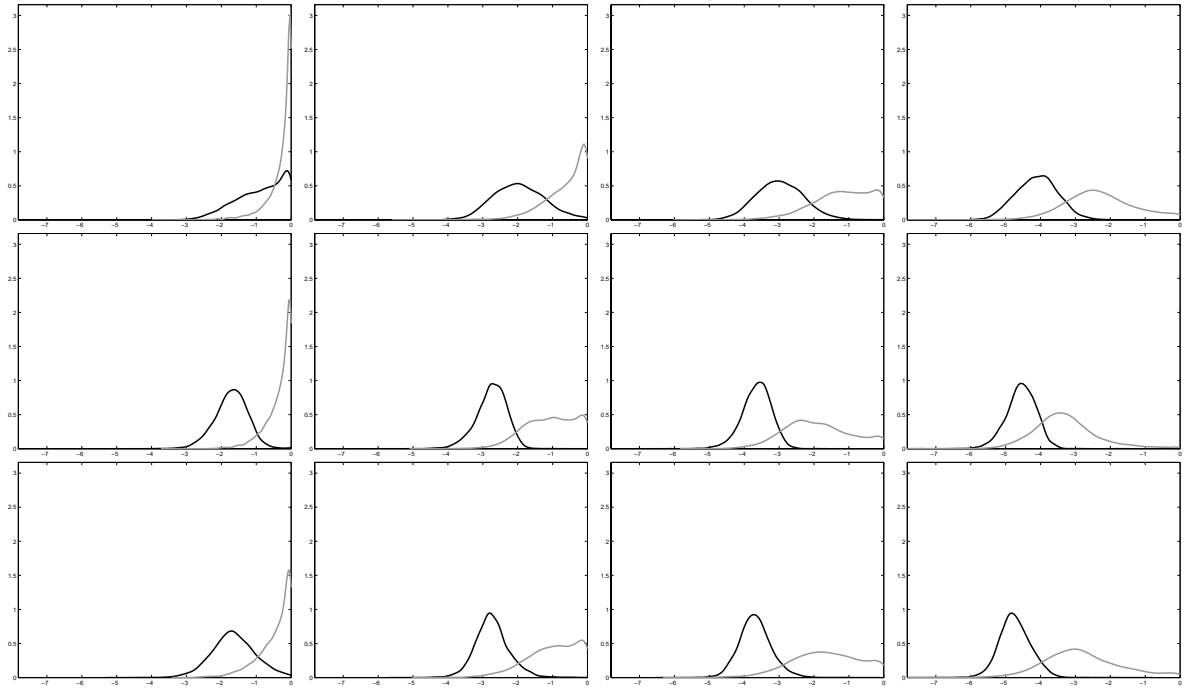
Figure 7: Density plots of the logarithms of the gaps between estimated and true cointegrating spaces for $k = 2$, case 4 and $N = 10$. The time series dimension is $T = 10, 25, 50, 100$ from left to right. The first row displays the results for cross-sectionally independent processes, the second row displays the results for cross-sectionally correlated ($\mathbf{\Sigma^F}$) processes and the third row displays the results for the case with cross-unit cointegration.