

Hanck, Christoph; Demetrescu, Matei; Tarcolea, Adina

**Conference Paper**

## IV-Based Cointegration Testing in Dependent Panels with Time-Varying Variance

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2012: Neue Wege und Herausforderungen für den Arbeitsmarkt des 21. Jahrhunderts - Session: Time Series Econometrics II, No. E09-V1

**Provided in Cooperation with:**

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Hanck, Christoph; Demetrescu, Matei; Tarcolea, Adina (2012) : IV-Based Cointegration Testing in Dependent Panels with Time-Varying Variance, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2012: Neue Wege und Herausforderungen für den Arbeitsmarkt des 21. Jahrhunderts - Session: Time Series Econometrics II, No. E09-V1, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft

This Version is available at:

<https://hdl.handle.net/10419/62072>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# IV-Based Cointegration Testing in Dependent Panels with Time-Varying Variance\*

Matei Demetrescu,<sup>†</sup> Christoph Hanck<sup>‡</sup> and Adina I. Tarcolea<sup>§</sup>

This version: November 10, 2011

## Abstract

While the limiting null distributions of cointegration tests are invariant to a certain amount of conditional heteroskedasticity as long as global homoskedasticity conditions are fulfilled, they are certainly affected when the innovations exhibit time-varying volatility. Worse yet, distortions from single units accumulate in panels, where one must anyway pay special attention to dependence among cross-sectional units, be it time-dependent or not. To obtain a panel cointegration test robust to both global heteroskedasticity and cross-unit dependence, we start by adapting the nonlinear instruments method proposed for the Dickey-Fuller test by Chang (2002, *J Econometrics* 110, 261–292) to an error-correction testing framework. We show that IV-based testing of the null of no error-correction in individual equations results in asymptotic standard normality of the test statistic as long as the  $t$ -type statistics are computed with White heteroskedasticity-consistent standard errors. Remarkably, the result holds even in the presence of endogenous regressors, irrespective of the number of integrated covariates, and for any variance profile. Furthermore, a test for the null of no cointegration—in effect, a joint test against no error correction in any equation of each unit—retains the nice properties of the univariate tests. In panels with fixed cross-sectional dimension, both types of test statistics from individual units are shown to be asymptotically independent even in the presence of correlation or cointegration across units, leading to a panel test statistic robust to cross-unit dependence and unconditional heteroskedasticity. The tests perform well in panels of usual dimensions with innovations exhibiting variance breaks and a factor structure.

**Key words:** Nonstationary volatility, dependent units, cross-member cointegration, cointegrating relation, endogeneity, instrumental variable, integrable transformation

**JEL classification:** C12; C22; C23

---

\*A previous version of the paper was circulated under the title “Panel Cointegration Testing Using Nonlinear Instruments.” The authors would especially like to thank Yoosoon Chang, as well as Anindya Banerjee, Jörg Breitung, Uwe Hassler, Vladimir Kuzin, Helmut Lütkepohl, Richard Spady, Jean-Pierre Urbain and Jürgen Wolters for very helpful comments and suggestions. Adina I. Tarcolea was supported in her research by the Deutsche Forschungsgemeinschaft (DFG) within the project “Integration and Cointegration Analysis with Panel Data”.

<sup>†</sup>**Corresponding author:** Hausdorff Center for Mathematics and Department of Economics, University of Bonn, Adenauerallee 24-42, D-53113 Bonn, Germany, tel. +49 228 733925, email: [Matei.Demetrescu@uni-bonn.de](mailto:Matei.Demetrescu@uni-bonn.de).

<sup>‡</sup>Rijksuniversiteit Groningen, Department of Economics, Econometrics and Finance.

<sup>§</sup>Statistics and Econometric Methods, Goethe University Frankfurt, and Banque PSA Finance S.A.

# 1 Motivation

Two particular aspects distinguish cointegration in (macroeconomic) panels from “usual” time series cointegration. On the one hand, cross-unit correlation is not as easily taken into account as the correlation between the innovations of a unit: even for a relatively small number of units, the number of e.g. covariances involved can be quite large, so estimating all of them in panels is usually infeasible. And ignoring cross-unit correlation can have disastrous consequences, as first pointed out by O’Connell (1998) for panel unit root tests. On the other hand, the null and alternative hypotheses have a particular structure, since one is interested in cointegration within the units, irrespective of cointegration across the units. Cross-unit cointegration is indeed an issue, see also Banerjee, Marcellino and Osbat (2004); but, roughly speaking, one is interested in “marginal” cointegration of panel units, irrespective of “joint” long-run dynamics. Along these lines, Gengenbach, Palm and Urbain (2006) study the asymptotics of the first-generation panel cointegration tests due to Kao (1999) and Pedroni (1999, 2004) under several forms of cross-sectional dependence and highlight the extent of the resulting distortions.

But taking cross-unit dependence into account is not straightforward due to the non-standard asymptotic null distributions of the involved tests. Moreover, the distributions are not invariant to unconditional heteroskedasticity in the time dimension: Cavaliere, Rahbek and Taylor (2010) show that the usual trace statistic for the cointegration rank is only unaffected by time-varying variance (nonstationary volatility) when a so-called global homoskedasticity condition is fulfilled.<sup>1</sup> Hanck (2009) points out that the distortions resulting from such non-stationary volatility increase for panel unit root tests with the number of cross-sectional units; we find similar behavior for panel cointegration tests. And, with macroeconomic data, global homoskedasticity assumptions are not easily justified.

Thus, with integrated panel data, applied research should rely on inferential procedures that are robust to both cross-sectional dependence and time-varying volatility. In dealing with cross-sectional dependence, approximate factor models—used with the PANIC methodology of Bai and Ng (2004)—are well-suited to deal with cross-unit dependence. E.g. Banerjee and Carrion-i-Silvestre (2006) allow too for a factor structure of the errors and consider models with structural breaks, while Gengenbach, Palm and Urbain (2006) work with factor models allowing for cross-unit cointegration; see also Westerlund and Larsson (2009) and Wang et al. (2010). Combining the significance of tests from individual units with a correction for cross-unit dependence (e.g. as in Hartung, 1999 and Demetrescu, Hassler and Tarcolea, 2006, or in Hanck, 2011) is also able to deal (to some extent) with the issue of cross-unit dependence. But neither factor models nor significance-based panel methods are automatically robust to global heteroskedasticity. Alternatively, one could

---

<sup>1</sup>See Cavaliere and Taylor (2009) and their references for the intimately related case of unit root tests.

panel-bootstrap the respective test statistics to account for cross-unit correlation; this was first proposed by Maddala and Wu (1999) for the so-called Fisher test for panel unit roots, and adopted for panel cointegration testing by Fachin (2005), Westerlund (2007) or Westerlund and Edgerton (2007). This does come at a cost, as bootstrap hypothesis tests are not always trivial to implement.<sup>2</sup> But a particular bootstrapping scheme, the wild bootstrap, was shown by Cavaliere, Rahbek and Taylor (2010) to be able to deal with global heteroskedasticity. We are not aware though of a bootstrap scheme for panel cointegration tests capable of accounting for both cross-sectional dependence and time-varying variance or covariance.

We therefore discuss a different approach to panel cointegration testing in this paper, one capable of handling both dependence across units and unconditional heteroskedasticity. Following the elegant treatment of cross-sectional dependence proposed by Chang (2002) for panel unit root testing, we use nonlinear instrumental variables [NIV] to test the null of no cointegration in the single equation framework advocated by Kremers, Ericsson and Dolado (1992) and Banerjee, Dolado and Mestre (1998). Single-equation testing in the panel framework is also discussed by Westerlund (2007), who considers an independent-units setup for asymptotic results, and then uses the bootstrap as a way to deal with cross-unit dependence.

The testing methodology discussed here has several remarkable asymptotic properties. For individual units, we show the NIV test for no error correction to have an asymptotic standard normal distribution, even when regressors are not weakly exogenous, irrespective of the number of integrated covariates, and notwithstanding unconditional heteroskedasticity. We show that the main ingredient for our result is the use of White standard errors, whereas usual standard errors, as originally proposed by Chang (2002) in a homoskedastic framework, lead to distorted tests. This is in sharp contrast to the OLS case, where White standard errors do not help with heteroskedasticity (cf. Demetrescu, 2010). The discussion is made possible by the recent results of Wang and Phillips (2009) who discuss convergence to local time of arbitrary continuous-time Gaussian processes, including the time-transformed Wiener processes involved in the asymptotics of the globally heteroskedastic case.

Moreover, a test statistic for the null of no cointegration can be built by combining evidence against no error-correction from all equations of the system in a very simple manner. In cross-dependent panels, test statistics from individual units (be they for no error correction or or no cointegration) are shown to be asymptotically independent in the presence of cross-unit correlation or cross-unit cointegration. This leads to a no-cointegration panel

---

<sup>2</sup>Rigorous asymptotic justification of sieve panel bootstrap schemes is provided for the case of panel unit root testing by Chang (2004). See Palm, Smeekes and Urbain (2011) for a block-bootstrap panel unit root test.

test statistic which is not affected by either cross-unit dependence or nonstationary volatility asymptotically. Furthermore, standard limiting distributions result without so-called  $N$  asymptotics, and unbalanced panels are allowed for. Chang and Nguyen (2011) discuss a nonlinear IV procedure for residual-based panel cointegration testing, having properties similar to those of our tests, and Miller (2010) adapts the NIV procedure to a trace-type test following Johansen (1995). Neither of the two tests is robust to global heteroskedasticity.

After describing our model in Section 2, we provide in Section 3 the analysis of the NIV cointegration test under time-varying volatility and its extension to panels exhibiting cross-unit dependence. The small-sample properties of the NIV test are discussed in Section 4. Section 5 concludes, and the proofs have been gathered in the Appendix.

## 2 Assumptions and test procedure

Denote by  $\mathbf{w}_{it} = (w_{1,it}, \dots, w_{K,it})'$ ,  $i = 1, \dots, N$ ,  $t = 1, \dots, T_i$ , the observations the possibly unbalanced panel consists of. We start with the simpler model allowing only for cross-unit correlation and proceed to the more general case later. Thus, we assume the data for each unit to be generated by an integrated vector autoregressive process [VAR] of order  $p + 1$  with  $K$  components,  $K \geq 2$ :

**Assumption 1.** *Let the model for each unit  $i$  be as follows:*

$$\Delta \mathbf{w}_{it} = \Pi_i \mathbf{w}_{it-1} + \sum_{j=1}^p A_{ij} \Delta \mathbf{w}_{it-j} + \boldsymbol{\epsilon}_{it}, \quad t = 1, \dots, T,$$

where the  $p + 1$  starting values are set to zero.

We shall model the shocks as being unconditionally heteroskedastic. When discussing the dynamics of such systems, the question arises of what cointegration actually stands for. The issue is that linear combinations of  $\mathbf{w}_{it}$  cannot be examined for strict or weak stationarity, since they would only have stationary volatility in particular cases. This is easily avoided, though, when understanding cointegration as mean reversion; see Cavaliere, Rahbek and Taylor (2010). Cointegration is then characterized as usual by the parameter matrices  $\Pi_i$  and  $A_{ij}$ ,  $j = 1, \dots, p$ .

The process  $\mathbf{w}_{it}$  is possibly cointegrated; let the rank of  $\Pi_i$  be denoted  $r_i$ . Under no cointegration, it holds that  $\Pi_i = \mathbf{0}$ , or  $r_i = 0$ . If  $\mathbf{w}_{it}$  is cointegrated, one has  $0 < r_i < K$ . Also, the known factorization of  $\Pi_i$ ,  $\Pi_i = \boldsymbol{\alpha}_i \boldsymbol{\beta}'_i$ , as the product of two  $K \times r_i$  matrices of adjustment speed coefficients and of parameters of the long-run relations, holds for  $r_i > 0$ .

Under the alternative hypothesis of cointegration, we assume  $r_i = 1$  for the units violating the null: the assumption is needed to motivate the test statistic, see Banerjee,

Dolado and Mestre (1998), but the test has power against higher cointegrating rank as well, so the restriction is not substantial. The following technical assumption guarantees that the process  $\mathbf{w}_{it}$  either follows a stable vector autoregressive process in differences (no cointegration), or, when  $\Pi_i \neq 0$ , I(2) processes are avoided.

**Assumption 2.** *Let the roots of the characteristic polynomial associated to  $\mathbf{w}_{it}$  defined in Assumption 1 be either 1 or have absolute values larger than 1. Further, if  $\Pi_i \neq \mathbf{0}$ , let  $\det\left(\boldsymbol{\alpha}'_{i\perp}\left(I - \sum_{j=1}^p A_{ij}\right)\boldsymbol{\beta}_{i\perp}\right) \neq 0$ , where  $\boldsymbol{\alpha}_{i\perp}$  and  $\boldsymbol{\beta}_{i\perp}$  are the orthogonal complements of  $\boldsymbol{\alpha}_i$  and  $\boldsymbol{\beta}_i$  w.r.t.  $\mathbb{R}^K$ .*

Let us examine a single equation; without loss of generality assume it is the first of the VAR system. Let  $\mathbf{w}_{e,it} = (w_{2,it}, \dots, w_{K,it})'$ . In the single equation framework, the error correction representation can be then written as follows:

$$\Delta w_{1,it} = \alpha_i (w_{1,it-1} + \boldsymbol{\theta}'_i \mathbf{w}_{e,it-1}) + \delta_i(L) \Delta w_{1,it-1} + \boldsymbol{\gamma}'_i(L) \Delta \mathbf{w}_{e,it-1} + \varepsilon_{it}, \quad (1)$$

$$\Delta \mathbf{w}_{e,it} = \boldsymbol{\alpha}_{ei} (w_{1,it-1} + \boldsymbol{\theta}'_i \mathbf{w}_{e,it-1}) + \boldsymbol{\delta}_{ei}(L) \Delta w_{1,it-1} + \boldsymbol{\Gamma}_i(L) \Delta \mathbf{w}_{e,it-1} + \boldsymbol{\nu}_{it}, \quad (2)$$

for  $t = 1, 2, \dots, T_i$ , where  $\alpha_i \in \mathbb{R}$ ,  $\boldsymbol{\alpha}_{ei} \in \mathbb{R}^{K-1}$ ,  $\boldsymbol{\theta} \in \mathbb{R}^{K-1}$ , the respective lag polynomials, and the innovations  $\varepsilon_{it}$  and  $\boldsymbol{\nu}_{it} \in \mathbb{R}^{K-1}$  are defined implicitly from Assumption 1. Not including contemporaneous differences  $\Delta \mathbf{w}_{e,it}$  in Equation (1), as Banerjee, Dolado and Mestre (1998) do, is compensated for by having allowed for correlated innovations, see Assumption 3. We aim to test either no error correction in a single equation ( $\alpha_i = 0$ ) or no cointegration ( $\boldsymbol{\alpha}_i = \mathbf{0}$ ); see below for a more detailed discussion of the panel hypotheses.

The issue of panel cross-member cointegration can be discussed more easily in the single-equation framework. Namely, it may well happen that regressors, or regressands, from (1) cointegrate across units (as could easily be the case with the same variable in neighboring countries of multi-country studies). This implies a (co-)integrated VAR model for the whole panel with certain restrictions on the cointegrating vectors. Section 3.3 examines such situations as well.

The innovations  $\boldsymbol{\epsilon}_{it} = (\varepsilon_{it}, \boldsymbol{\nu}'_{it})'$  are allowed to correlate both within and across units; the correlation, just like the variance, is allowed to be time-dependent as specified by the following assumption.

**Assumption 3.** *Denote  $\boldsymbol{\epsilon}_t^N = (\boldsymbol{\epsilon}'_{1t}, \dots, \boldsymbol{\epsilon}'_{Nt})'$  and let  $\boldsymbol{\epsilon}_t^N = \Sigma^{0.5}(t/T) \bar{\boldsymbol{\epsilon}}_t^N$  where  $\bar{\boldsymbol{\epsilon}}_t^N \sim iid(0, I_{KN})$  having finite kurtosis, and  $\Sigma(s)$  is a  $NK \times NK$  matrix of piecewise Lipschitz functions, positive definite at any  $s \in [0, 1]$ . Furthermore, assume  $\bar{\boldsymbol{\epsilon}}_t^N$  to have a characteristic function  $\phi(\boldsymbol{\lambda})$  such that  $\int \|\boldsymbol{\lambda}\|^3 |\phi(\boldsymbol{\lambda})| d\boldsymbol{\lambda} < \infty$ .*

With the variance of the innovations depending on the sample size  $T$ , Assumption 3 implies them to actually be a triangular array. To simplify the notation, we drop the additional subscript  $T$ .

Under Assumption 3, an invariance principle holds for the cumulated sums of  $\epsilon_t^N$  even under time-varying variance. The weak limit, however, is not the usual multivariate Brownian motion, but rather the process described by the stochastic integral  $\int_0^s \Sigma^{0.5}(r) d\mathbf{W}(r)$ , where  $\mathbf{W}(s)$  is a vector of  $NK$  independent standard Wiener processes. See Lemma 2 of Cavaliere, Rahbek and Taylor (2010). It is still a Gaussian process, but with a covariance kernel different from the one of the multivariate Brownian motion, implying a quadratic variation depending nonlinearly on time (or a nonlinear variance profile). The homoskedastic case is recovered when  $\Sigma$  is constant and the variance profile is linear as a consequence.

Assumptions 2 and 3 together allow for lack of exogeneity. As will be shown in Section 3.1, this does not affect the asymptotics of the test under the null of no cointegration. This is a nice feature, since the (often encountered) weak exogeneity assumption is one important source of criticism to the single equation approach. The conditions in Assumption 3 are stronger than the typical sets of assumptions under which an invariance principle for cumulated innovations holds, but are needed to establish the asymptotic behavior of the integrable transformations of integrated processes involved in the proofs; see Park and Phillips (1999, 2001), Chang, Park and Phillips (2001), Phillips, Park and Chang (2004), de Jong and Wang (2005) and Wang and Phillips (2009).

Deterministic components (such as non-zero means or linear time trends) can be incorporated in the usual way into cointegrated VARs; see Lütkepohl (2005, Section 6.4) for details. Section 3.2 gives arguments in favor of recursive removal of deterministic components from the levels and shows how it can be implemented for NIV cointegration testing.

In the single-equation case, our test builds on instrumental estimation estimation of the test equation (1) using integrable transformations. More precisely,  $F_i(w_{1,it-1})$  is used as instrument for  $w_{1,it-1}$ , where the function  $F_i(\cdot)$  is restricted as follows:

**Assumption 4.** *Let  $F_i(x)$  be continuous on  $\mathbb{R}$  such that  $\int_{-\infty}^{\infty} |xF_i(x)| dx$  is finite and non-zero.*

In what concerns the other integrated regressors, two possibilities arise. First, we may take them as instruments for themselves. Second, we may take integrable transformations as instruments. We shall call the first case “partial instrumentalization,” and the second will be denoted as “complete instrumentalization.” The completely instrumentalized test is similar in spirit to the trace test of Miller (2010). The derivations of the paper, however, are given only for the case of partial instrumentalization. They are similar in the case of complete instrumentalization so we do not provide the latter to save space.

For individual units, the null hypothesis in the single equation framework is  $\alpha_i = 0$ ; when assuming weak exogeneity,  $\alpha_{ei} = \mathbf{0}$  is implied. Note that, when allowing for error correction to affect the other components of  $\mathbf{w}_t$ , the null of the test is actually absence

of error correction in the studied equation and not lack of cointegration between  $w_{1,it}$  and  $\mathbf{w}_{e,it}$ .<sup>3</sup> We will show in the following section that the NIV test for  $\alpha_i = 0$  is not directly affected by the presence of error-correction terms in any of the  $K - 1$  equations in (2). We are hence able to directly test the following hypothesis:

Null hypothesis (no single-equation error correction):  $\alpha_i = 0$ .

Under the alternative,  $\alpha_i$  needs to be negative if error-correction is present only in Equation (1). Otherwise,  $\alpha_i$  may also be positive (for an example see Johansen, 1995, p. 54). Thus, we test against

Alternative hypothesis:  $\alpha_i \neq 0$ .

Since, as shall be seen in Section 3, there is a simple way to combine evidence from  $K$  single-equation tests in the NIV framework, we are able to consider a test for cointegration as well:

Null hypothesis (no cointegration):  $\boldsymbol{\alpha}_i = \mathbf{0}$

with the corresponding

Alternative hypothesis:  $\boldsymbol{\alpha}_i \neq \mathbf{0}$ .

In the panel case, these null hypotheses should hold for all units,

Null hypothesis (no panel error correction):  $\alpha_i = 0 \forall i = 1, \dots, N$ ,

Null hypothesis (no panel cointegration):  $\boldsymbol{\alpha}_i = \mathbf{0} \forall i = 1, \dots, N$ ,

whereas, under the alternative, they are violated for at least one unit:

Alternative hypothesis:  $\exists i$  such that  $\alpha_i \neq 0$

Alternative hypothesis:  $\exists i$  such that  $\boldsymbol{\alpha}_i \neq \mathbf{0}$ .

### 3 Asymptotic results

We begin by addressing the behavior of individual-unit NIV tests: not only do panel results build on individual-unit results, but their properties are interesting in their own right. Deterministic components are dealt with in Subsection 3.2, and the panel case is discussed in Subsection 3.3. To ease the exposition, we drop the index  $i$  for Sections 3.1 and 3.2.

---

<sup>3</sup>This attribute is common to all approaches based on a single equation and not specific to our test.



### 3.1 Single unit tests

Let us first reformulate the test regression (1) to match the usual notation of the single equation framework. Defining

$$\begin{aligned}\mathbf{x}'_{t-1} &= (\Delta w_{1,t-1}, \dots, \Delta w_{1,t-p}, \Delta \mathbf{w}'_{e,t-1}, \dots, \Delta \mathbf{w}'_{e,t-p}, \mathbf{w}'_{e,t-1}) \\ &= (\mathbf{x}'_{t-1,0}, \mathbf{x}'_{t-1,1}),\end{aligned}$$

the I(1) variables  $\mathbf{x}_{t-1,1} = \mathbf{w}_{e,t-1}$  are separated from the I(0) ones,  $\mathbf{x}_{t-1,0}$ . The single-equation model becomes, with  $y_t = w_{1,t}$ ,

$$\Delta y_t = \alpha y_{t-1} + \boldsymbol{\beta}' \mathbf{x}_{t-1} + \varepsilon_t, \quad t = 1, \dots, T, \quad (3)$$

and we assume the  $p$  starting values for  $t = -p + 1, \dots, 0$  to be zero for the case without deterministic. The new parameter vector  $\boldsymbol{\beta}$  is given by

$$\boldsymbol{\beta}' = (\delta_1, \dots, \delta_p, \gamma'_1, \dots, \gamma'_p, \alpha \boldsymbol{\theta}'),$$

where  $\delta_j$  and  $\gamma_j$ ,  $j = 1, 2, \dots, p$ , are the respective coefficients of the lag polynomials from Equation (1). It is convenient to write  $\boldsymbol{\beta}' = (\boldsymbol{\beta}'_0, \boldsymbol{\beta}'_1)$ , with  $\boldsymbol{\beta}'_0 = (\delta_1, \dots, \delta_p, \gamma'_1, \dots, \gamma'_p)$  and  $\boldsymbol{\beta}'_1 = \alpha \boldsymbol{\theta}'$ , in accordance with the partition  $\mathbf{x}'_{t-1} = (\mathbf{x}'_{t-1,0}, \mathbf{x}'_{t-1,1})$ .

The  $t$ -type statistic of the estimated parameter  $\hat{\alpha}$  remains the natural choice as a test statistic for the null  $\alpha = 0$ , even with IV estimation. For the case of partial instrumentalization, one obtains with the help of standard regression algebra that

$$\hat{\alpha} - \alpha = Q^{-1}M,$$

where

$$M = \sum_{t=1}^T F(y_{t-1}) \varepsilon_t - \sum_{t=1}^T F(y_{t-1}) \mathbf{x}'_{t-1} \left( \sum_{t=1}^T \mathbf{x}_{t-1} \mathbf{x}'_{t-1} \right)^{-1} \sum_{t=1}^T \mathbf{x}_{t-1} \varepsilon_t$$

and

$$Q = \sum_{t=1}^T F(y_{t-1}) y_{t-1} - \sum_{t=1}^T F(y_{t-1}) \mathbf{x}'_{t-1} \left( \sum_{t=1}^T \mathbf{x}_{t-1} \mathbf{x}'_{t-1} \right)^{-1} \sum_{t=1}^T \mathbf{x}_{t-1} y_{t-1}.$$

For the  $t$  statistic, it holds under the null hypothesis  $\alpha = 0$

$$t_{\hat{\alpha}} = \frac{\hat{\alpha}}{\hat{\sigma}_{\hat{\alpha}}},$$

with  $\hat{\sigma}_{\hat{\alpha}}$  the estimated standard deviation of  $\hat{\alpha}$ . Assuming homoskedasticity,  $\hat{\sigma}_{\hat{\alpha}}$  is given by

$$\hat{\sigma}_{\hat{\alpha}}^2 = \hat{\sigma}_{\varepsilon}^2 Q^{-2} P,$$

where

$$P = \sum_{t=1}^T F(y_{t-1})^2 - \sum_{t=1}^T F(y_{t-1}) \mathbf{x}'_{t-1} \left( \sum_{t=1}^T \mathbf{x}_{t-1} \mathbf{x}'_{t-1} \right)^{-1} \sum_{t=1}^T \mathbf{x}_{t-1} F(y_{t-1}),$$

and  $\widehat{\sigma}_\varepsilon^2$  is a consistent estimator of the residual variance.

We shall prove in the following that a limiting null distribution not depending on nuisance parameters such as the variance profile arises in the globally heteroskedastic case, provided that White heteroskedasticity-robust variance estimators are employed:

$$\widehat{\sigma}_{\widehat{\alpha}, W}^2 = Q^{-2} P_W,$$

where  $P_W$  is given by

$$\begin{aligned} & \sum_{t=1}^T F(y_{t-1})^2 \widehat{\varepsilon}_t^2 - 2 \sum_{t=1}^T F(y_{t-1}) \mathbf{x}'_{t-1} \left( \sum_{t=1}^T \mathbf{x}_{t-1} \mathbf{x}'_{t-1} \right)^{-1} \sum_{t=1}^T \mathbf{x}_{t-1} F(y_{t-1}) \widehat{\varepsilon}_t^2 \\ & + \sum_{t=1}^T F(y_{t-1}) \mathbf{x}'_{t-1} \left( \sum_{t=1}^T \mathbf{x}_{t-1} \mathbf{x}'_{t-1} \right)^{-1} \sum_{t=1}^T \mathbf{x}_{t-1} \mathbf{x}'_{t-1} \widehat{\varepsilon}_t^2 \left( \sum_{t=1}^T \mathbf{x}_{t-1} \mathbf{x}'_{t-1} \right)^{-1} \sum_{t=1}^T \mathbf{x}_{t-1} F(y_{t-1}), \end{aligned}$$

and  $\widehat{\varepsilon}_t$  are the regression residuals,

$$\widehat{\varepsilon}_t = \Delta y_t - \widehat{\alpha} y_{t-1} - \widehat{\beta}_1 \mathbf{x}_{t-1,1} - \widehat{\beta}_0 \mathbf{x}_{t-1,0}. \quad (4)$$

In order for  $\widehat{\varepsilon}_t$  to be consistently estimated, certain minimal convergence rates for the estimators associated to the integrated regressors  $\mathbf{x}_{t-1,1}$  are required, beyond consistency. They are, however, not always given in the NIV framework; see Proposition 1 below and also Demetrescu (2009) for the univariate case.

Moreover, it is argued in the proof of Proposition 2 analyzing the  $t$ -type statistic of  $\alpha$  that usual standard errors do not lead to a robust statistic. So the remainder of the paper employs White standard errors.<sup>4</sup>

Under the null  $\alpha = 0$ , using White standard errors leads to

$$t_{\widehat{\alpha}} = \frac{M}{\sqrt{P_W}}.$$

In order to establish its asymptotic behavior under the null and the alternative, it will be

---

<sup>4</sup>White standard errors are also shown by Boswijk (2010) to robustify inference on cointegration parameters against breaks in the variance.

more convenient to study the pivotal quantity

$$t^* = \frac{\hat{\alpha} - \alpha}{\widehat{\sigma}_{\hat{\alpha}, W}},$$

irrespective of the value of  $\alpha$ . Under the null  $\alpha = 0$ , we obviously have that  $t_{\hat{\alpha}} = t^* = \frac{M}{\sqrt{P_W}}$ ; under the alternative, it holds that  $t_{\hat{\alpha}} = t^* + \alpha/\widehat{\sigma}_{\hat{\alpha}, W}$ .

The following proposition establishes convergence properties of the NIV estimators. They are needed to establish the (lack of) consistency of the residuals  $\widehat{\varepsilon}_t$ .

**Proposition 1.** *For partially instrumentalized NIV estimation of test equation (3), it holds as  $T \rightarrow \infty$  under Assumptions 1 through 4 and  $\alpha = 0$  that*

a) if  $\alpha_e = \mathbf{0}$

$$\hat{\alpha} - \alpha = O_p(T^{-0.75}), \quad (5)$$

$$\widehat{\beta}_0 - \beta_0 = O_p(T^{-0.25}) \quad (6)$$

and

$$\widehat{\beta}_1 - \beta_1 = O_p(T^{-0.75}) \quad (7)$$

b) if  $\alpha_e \neq \mathbf{0}$ ,

$$\widehat{\beta}_0 - \beta_0 = O_p(T^{-0.25}).$$

However, the convergence rates of  $\hat{\alpha}$  or  $\widehat{\beta}_1$  to the true values  $\alpha$  and  $\beta_1$  can be of an order as low as  $T^{-0.25}$  (instead of  $T^{-0.75}$ ), depending on the cointegrating vector  $(1, \boldsymbol{\theta}')$ .

**Proof:** See the Appendix.

The direct consequence of Proposition 1 is that one should not use NIV residuals to compute  $\widehat{\sigma}_{\hat{\alpha}, W}$ , since the convergence rate of  $\widehat{\beta}_1$  may not be high enough to ensure consistent residuals, given that the regressors corresponding to  $\widehat{\beta}_1$  are integrated. A simple solution is to use the OLS estimators of  $\alpha$  and  $\beta_1$  for computing residuals; their convergence rates are known to be high enough to meet the requirements of Lemma 1 establishing the behavior of the pivotal quantity  $t^*$ .

**Lemma 1.** *If using  $\widehat{\varepsilon}_t$  from (4) with  $\hat{\alpha} - \alpha = o_p(1)$ ,  $\widehat{\beta}_1 - \beta_1 = o_p(T^{-0.5})$  and  $\widehat{\beta}_0 - \beta_0 = o_p(1)$ , it holds as  $T \rightarrow \infty$  under Assumptions 1 through 4 that*

$$t^* = \frac{T^{-0.25} \sum_{t=1}^T F(y_{t-1}) \varepsilon_t}{\sqrt{T^{-0.5} \sum_{t=1}^T F(y_{t-1})^2 \varepsilon_t^2}} + o_p(1),$$

for any  $\alpha$ , regardless of whether  $\alpha_e = \mathbf{0}$  or  $\alpha_e \neq \mathbf{0}$ .

**Proof:** See the Appendix.

It now becomes clear why no restrictive weak exogeneity assumptions have to be made: Lemma 1 shows that no regressor except for the lagged dependent variable  $y_{t-1}$  influences the test statistic in the limit. In contrast to that, OLS estimation of the test equation requires either weak exogeneity or inclusion of leads to account for second-order bias (see Banerjee, Dolado and Mestre, 1998, for details); and OLS estimation leads to test statistics affected by nonstationary volatility.

The following proposition summarizes the asymptotic behavior of the proposed test statistic for the null  $\alpha = 0$  under both null and alternative hypotheses.

**Proposition 2.** *Under the assumptions of Lemma 1, it holds as  $T \rightarrow \infty$ :*

a) *if  $\alpha = 0$ , then*

$$t_{\hat{\alpha}} \xrightarrow{d} \mathcal{N}(0, 1);$$

b) *if  $\alpha \neq 0$ , then*

$$|t_{\hat{\alpha}}| \xrightarrow{p} \infty.$$

**Proof:** See the Appendix.

**Remark 1.** Confidence intervals for the parameter  $\alpha$  are straightforward to build, since the pivotal statistic  $t^*$  has asymptotic standard normal distribution whatever the true value of  $\alpha$  is, as long as consistent residuals are available.

Loosely speaking, the transformed variable  $F(y_{t-1})$  does not quite behave like an integrated process (e.g. it has uniformly bounded variance for any  $t$ ), which eases the application of a central limit theorem with random normalization. This leads to asymptotic normality of the statistic  $t^*$  (and thus of  $t_{\hat{\alpha}}$  under the null). Integrable instruments are actually not the only instruments leading to asymptotic normality. For panel unit root testing, Shin and Kang (2006) use so-called Huber instruments instead of Chang's (2002) integrable instruments; but they do not lead to asymptotic independence of individual test statistics, while integrable ones do, see Chang (2002). Still, (panel) unit root tests based on Huber-type instruments are robust to unconditional heteroskedasticity; see Demetrescu and Hanck (2011a,b).

The asymptotic independence derived by Chang (2002) in the panel unit root case yields a test for no cointegration, i.e. for the null  $\boldsymbol{\alpha} = \mathbf{0}$ . Namely, if individual test statistics for no error correction are asymptotically independent across units in the presence of cross-unit correlation (see Proposition 5 below), why would they not be so for the  $K$  equations *within* one unit? Denoting  $t_{\hat{\alpha}k}$ ,  $k = 1, \dots, K$ , the test statistics for no error-correction in each of

the  $K$  equations of a given unit, the test statistic for no cointegration would then be

$$Q_{\alpha} = \sum_{k=1}^K t_{\hat{\alpha}k}^2, \quad (8)$$

which would follow a chi-square distribution with  $K$  degrees of freedom asymptotically. The intuition is indeed true, as stated in the following proposition.

**Proposition 3.** *Under the assumptions of Lemma 1, it holds as  $T \rightarrow \infty$ :*

a) if  $\alpha = \mathbf{0}$ , then

$$Q_{\alpha} \xrightarrow{d} \chi^2(K);$$

b) if  $\alpha \neq \mathbf{0}$ , then

$$Q_{\alpha} \xrightarrow{p} \infty.$$

**Proof:** See the Appendix.

Hence, a test statistic against the null of no cointegration is simply built by summing the squared  $t$  statistics for each of the  $K$  equations; reject for (too) large values with critical values from the  $\chi^2(K)$  distribution. The proof builds on the asymptotics employed by Chang (2002) and Chang and Nguyen (2011) too, but has to take into account the time varying volatility of the data generating process.

**Remark 2.** In the case of a no cointegration test, one could use the null restrictions  $\alpha = 0$  and  $\beta_1 = \mathbf{0}$  to compute the residuals  $\hat{\varepsilon}_t$  from the regression of  $\Delta y_t$  on lags of  $\Delta y_t$  and of  $\Delta \mathbf{w}_{e,t}$ .

**Remark 3.** Propositions 2 and 3 could also be established for the case where  $\Delta \mathbf{w}_t$  is a general linear process with coefficients satisfying a weak summability condition, if an autoregressive approximation of order growing to infinity, but slower than  $T$ , is used. See Saikkonen and Lütkepohl (1996) for the analysis of (co-)integrated VAR( $\infty$ ) processes and Demetrescu (2009) for the discussion of the NIV panel unit root test with autoregressive approximations. Moreover, Demetrescu (2011) shows that Chang's unit root test retains its asymptotic properties if using  $F\left(\frac{y_t-1}{T^\eta}\right)$  as instrument, as long as  $0 \leq \eta < 0.5$ . Extending this paper's results in these direction is tedious, yet straightforward and we omit the details.

### 3.2 Accounting for deterministic components

When accounting for deterministic components such as nonzero starting values or linear deterministic trends, the lagged differences are either not affected by a constant non-zero mean in levels, or can be easily demeaned. For the levels  $y_{t-1}$ , one must make sure that the

product of instrument and innovation possesses the martingale property after removing the deterministic component, purpose to which we follow Chang (2002) and resort to recursive (adaptive) schemes of demeaning or detrending of  $y_{t-1}$ . See Born and Demetrescu (2011) for the LR trace test for the cointegration rank with recursive adjustment for deterministic.

For a non-zero mean, this means that the NIV cointegration test has to be carried out in following test equation:

$$\Delta y_t = \alpha y_{t-1}^\mu + \beta' \mathbf{x}_{t-1}^\mu + \varepsilon_t, \quad (9)$$

where the recursively demeaned lagged level  $y_{t-1}^\mu$  is given for  $t \geq 2$  by

$$y_{t-1}^\mu = y_{t-1} - \frac{1}{t-1} \sum_{j=1}^{t-1} y_j.$$

The stationary regressors, being differences, need no adjustment, but the integrated regressors  $\mathbf{x}_{t-1,1}$  may also require demeaning, hence the notation  $\mathbf{x}_{t-1}^\mu$  in (9). Usual projection on a constant is allowed for the integrated regressors, in contrast to the case of the lagged dependent variable. Moreover, it can be shown that  $\mathbf{x}_{t-1,1}$  themselves may be recursively demeaned without affecting the asymptotics.

For a linear trend, one correspondingly uses as test equation

$$\Delta y_t - \overline{\Delta y} = \alpha y_{t-1}^\tau + \beta' \mathbf{x}_{t-1}^\tau + \varepsilon_t, \quad (10)$$

where the recursively detrended lagged level  $y_{t-1}^\tau$  is given for  $t \geq 2$  by

$$y_{t-1}^\tau = y_{t-1} + \frac{2}{t-1} \sum_{j=1}^{t-1} y_j - \frac{6}{t(t-1)} \sum_{j=1}^{t-1} j y_j,$$

and the integrated regressors may be detrended the usual way. The stationary regressors  $\mathbf{x}_{t-1,0}$  and the regressand  $\Delta y_t$  only require usual demeaning. Being under the null of no cointegration, deterministic components orthogonal to the cointegrating vectors are not an issue.

Then, one uses as instruments  $F(y_{t-1}^\mu)$  or  $F(y_{t-1}^\tau)$ . For the case of the test equations (9) and (10), the results analogous to Proposition 1, Lemma 1 and Proposition 2 can be shown to hold true.

**Proposition 4.** *Under the assumptions of Lemma 1 and recursive demeaning or detrending, it holds for the  $t$  statistics from test equations (9) or (10) as  $T \rightarrow \infty$ :*

a) if  $\alpha = 0$ , then

$$t_{\hat{\alpha}} \xrightarrow{d} \mathcal{N}(0, 1);$$

b) if  $\alpha \neq 0$ , then

$$|t_{\hat{\alpha}}| \xrightarrow{p} \infty.$$

**Proof:** See the Appendix.

The extension of Proposition 3 to the case of recursive demeaning and detrending holds as well.

### 3.3 Panel test

We now turn our attention to the panel case and assume the number of cross-sections to be finite. For macroeconomic panels, this is not a shortcoming, but an advantage:  $N$ -asymptotics are usually used to get an approximation for the finite- $N$  distribution of (panel unit root or panel cointegration) test statistics. Here, the test statistics are asymptotically normal and asymptotically independent as  $T \rightarrow \infty$  (cf. Propositions 2 and 3) without requiring  $N$  to be large, as is for instance needed for approximate factor models.

Chang's (2002) work, together with Lemma 1, suggests that panels may be unbalanced in the sense of her Assumption 4.1:

**Assumption 5.** Let  $T_{\min}$  be the smallest and  $T_{\max}$  the largest of the unit lengths  $T_i$ , and assume that

$$T_{\min}^{-1} + T_{\min}^{-0.75} T_{\max}^{0.25} \log T_{\max} \rightarrow 0.$$

The panel test statistic for the null of no error-correction is built the same way as the individual statistic for no cointegration. Namely, add (this time across the panel) individual test statistics, which, by the same mechanism exploited by the proof of Proposition 3, will be asymptotically independent:

$$\tilde{X} = \sum_{i=1}^N t_{\hat{\alpha}i}^2. \quad (11)$$

The single test statistics  $t_{\hat{\alpha}i}$  may be computed with recursive demeaning or detrending. In building  $\tilde{t}$ , there's the issue of which of the  $K$  equations in each unit should be examined for error-correction, unless economic theory dictates which the dependent variable should be. If there are no prior arguments in favor of a certain choice, one might be better off using the no cointegration test (i.e. sum the squared  $t$  statistics over  $K$  and  $N$ ); see Proposition 6 below.

The asymptotic properties of the panel test statistic (11) are stated in the following proposition.

**Proposition 5.** Under the assumptions of Proposition 2 together with Assumption 5, it holds for  $\tilde{X}$  from (11) as  $T \rightarrow \infty$

a) if  $\alpha_i = 0 \forall i \in \{1, 2, \dots, N\}$

$$\tilde{X} \xrightarrow{d} \chi^2(N),$$

b) if  $\exists i$  such that  $\alpha_i \neq 0$

$$\tilde{X} \xrightarrow{p} \infty.$$

**Proof:** Follows with the arguments used in the proof of Proposition 3 above and in the proof of Chang's Lemma 4.2 (2002).

Note that the proof of Proposition 3 (and thus the proof of Proposition 5) holds for any time-varying covariance matrix of the innovations; the panel test is robust to cross-unit correlation. Moreover, although we do not consider cross-unit dynamics to ease the exposition, it is easily seen from the proof of Proposition 3 that augmenting the test regressions with lagged differences from other cross-sectional units can be allowed for without affecting the asymptotic distributions. The augmentation might even be desirable, since, if ignoring cross-unit dynamics when they are present, each unit  $\Delta \mathbf{w}_{it}$  follows (marginally) an ARMA process, which requires approximation by means of an autoregressive process of order growing to infinity; see also Remark 3.

In what concerns cross-member cointegration, Lemma 1 shows that the terms containing integrated regressors and lagged differences are asymptotically negligible. Thus, the elements of  $\mathbf{w}_{e,it-1}$  may cointegrate across units. Should the lagged dependent variables  $y_{it-1}$  cointegrate across units, or cointegrate with  $\mathbf{w}_{e,jt-1}$ ,  $i \neq j$ , asymptotic independence is no longer guaranteed. But Chang and Song (2009) show that independence of single test statistics holds, if the instrument generating functions  $F_i$  satisfy certain orthogonality conditions. They suggest the use of Hermite polynomials, and point out that these need rescaling before using them as instrument generating functions (see Chang and Song, 2009, for a complete discussion). Clearly, the Hermite-based instrument generating functions can be used in our case as well.

**Remark 4.** If one knows the adjustment coefficients to be, for instance, negative under the alternative, one-sided testing would lead to more powerful procedures. A simple way to build the one-sided panel test is to take the standardized sum of single test statistics,

$$\tilde{t}^- = \frac{1}{\sqrt{N}} \sum_{i=1}^N t_{\hat{\alpha}_i} \xrightarrow{d} \mathcal{N}(0, 1). \quad (12)$$

For the case of the panel test of no cointegration, the discussion is analogous: sum the



$N$  individual no cointegration statistics  $Q_{\alpha_i}$ ,

$$\tilde{Q} = \sum_{i=1}^N Q_{\alpha_i}. \quad (13)$$

A chi-square distribution with  $KN$  degrees of freedom results asymptotically for  $\tilde{Q}$ , and the test is consistent against alternatives exhibiting at least one cointegrated unit; see the following proposition.

**Proposition 6.** *Under the assumptions of Proposition 2 together with Assumption 5, it holds for  $\tilde{Q}$  from (13) as  $T \rightarrow \infty$*

a) if  $\alpha_i = \mathbf{0} \forall i \in \{1, 2, \dots, N\}$

$$\tilde{Q} \xrightarrow{d} \chi^2(KN),$$

b) if  $\exists i$  such that  $\alpha_i \neq \mathbf{0}$

$$\tilde{Q} \xrightarrow{p} \infty.$$

**Proof:** Obvious and omitted.

**Remark 5.** As  $N$  increases, one can make use of the standard normal approximation for chi-squared distributions with a large number of degrees of freedom, leading to

$$\frac{\tilde{Q} - NK}{\sqrt{2NK}} \underset{\text{approx}}{\sim} \mathcal{N}(0, 1).$$

## 4 Small-sample behavior

### 4.1 Time series evidence

We now present some simulation evidence of the behavior of the tests suggested here relative to other popular procedures. In order to motivate the need for robust tests we investigate if nonstationary volatility is actually detrimental in practice to the performance of some widely used non-robust tests. All existing panel cointegration tests are in some way or another suitable combinations of underlying time series cointegration tests. Hence, it is without loss of generality, but helpful to sharpen the focus on nonstationary volatility, if we initially consider time series cointegration tests. Section 4.2 below provides results for the panel case.

Concretely, we compare  $t_{\hat{\alpha}}$  and  $Q_{\alpha}$  to the trace test of Johansen (1995) ( $\lambda_{\text{tr}}$ ), the residual-based test of Engle and Granger (1987) (EG) and to the nonlinear IV-based test

$\tilde{\tau}_i$  of Chang and Nguyen (2011) (see their Equation (16)). (The statistic  $t_{\hat{\alpha}}$  is computed for each of the  $K = 2$  equations of the system.)

The heteroskedastic cointegration DGP is taken from Cavaliere, Rahbek and Taylor (2010):

$$\Delta \mathbf{w}_t = \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{w}_{t-1} + \boldsymbol{\epsilon}_t, \quad t = 1, \dots, T,$$

where  $\boldsymbol{\epsilon}_t = \Sigma_t^{0.5} \bar{\boldsymbol{\epsilon}}_t$  and  $\bar{\boldsymbol{\epsilon}}_t \sim iid \mathcal{N}(0, I_K)$ , with  $\Sigma_t^{0.5}$  a time-varying  $K \times K$  diagonal matrix initialized at  $\Sigma_0^{0.5} = I_K$ . Furthermore,  $\Sigma_t^{0.5} = \delta I_K$  for  $t = \lfloor \tau T \rfloor, \lfloor \tau T \rfloor + 1, \dots, T$  and  $T = \{100, 200, 500\}$ . We take  $K = 2$  and consider ‘early’ to ‘late’ break fractions  $\tau \in \{1/5, 1/3, 2/3, 4/5\}$  and large negative to large positive variance shifts  $\delta \in \{1/5, 1/3, 1, 3, 5\}$ . The case  $\delta = 1$  covers the benchmark homoskedastic scenario. The experiments use 5,000 replications, and all results are for a nominal 5% level.

As our aim is to demonstrate that traditional time series cointegration tests are not level- $\alpha$  under nonstationary volatility, we set  $\boldsymbol{\alpha} = \mathbf{0}$ .

A constant is removed for all tests. We purposely avoid additional complications such as short-run dynamics and corresponding lag-length determination in order to isolate the effect of nonstationary volatility. All qualitative findings remain intact if we augment the DGP to also take these features into account.<sup>5</sup> For the implementation of  $t_{\hat{\alpha}}$  and  $Q_{\boldsymbol{\alpha}}$ , the instrument generating function [IGF] is picked as in Chang (2002),  $F(x) = x \exp(-|x|)$ . We follow her suggestion to improve the finite-sample performance of the tests by taking  $x = C y_{t-1}$ . Concretely, we take  $C = 4/\hat{\sigma}_{\Delta y}$ ; and also do so for  $\tilde{\tau}_i$ .<sup>6</sup>

Results are reported in Table 1. As expected, all traditional tests effectively handle the homoskedastic case  $\delta = 1$  (which we only report once to avoid redundancies). We observe however, in line with Cavaliere, Rahbek and Taylor (2010), that the Johansen (1995) test often exhibits large upward size distortions when  $\delta \neq 1$ . Concretely, early negative and late positive breaks lead to severe rates of overrejection. Table 1 reveals a similar, although slightly less extreme, picture for the Engle and Granger (1987) test. In line with the inappropriateness of the standard limiting distributions under nonstationary volatility, the distortions also do not vanish as  $T$  increases; and hence are, as expected, not a finite-sample issue. In turn, the nonlinear IV test of Chang and Nguyen (2011) surprisingly often controls size reasonably well even under  $\delta \neq 1$ . Yet, that it is not a robust test is evidenced by its strong undersizedness for large positive breaks, see in particular  $\delta = 5$ . In fact, it seems that rejection rates tend to zero as  $T$  grows.

The performance of the tests proposed here, in turn, is robust over all configurations of  $\tau$  and  $\delta$ . This is true for both the error-correction tests,  $t_{\hat{\alpha},1}$  and  $t_{\hat{\alpha},2}$ , as well as for the

---

<sup>5</sup>We have moreover experimented with additional features of the DGP of Cavaliere, Rahbek and Taylor (2010), such as variance breaks in only a subset of the  $K$  series, or time-varying correlations. Again, the qualitative conclusions are entirely unaffected.

<sup>6</sup>Unreported, but available, results show that this has a minor effect on  $t_{\hat{\alpha}}$  and  $Q_{\boldsymbol{\alpha}}$  in the time series case and does not affect the conclusions about  $\tilde{\tau}_i$ 's lack of robustness under time-varying volatility.

Table 1: Size of Time Series Cointegration Tests under Nonstationary Volatility

$\delta$	$T$	$t_{\hat{\alpha},1}$	$t_{\hat{\alpha},2}$	$Q_{\alpha}$	EG	$\tilde{\tau}_i$	$\lambda_{tr}$	$t_{\hat{\alpha},1}$	$t_{\hat{\alpha},2}$	$Q_{\alpha}$	EG	$\tilde{\tau}_i$	$\lambda_{tr}$
		$\tau = 1/5$						$\tau = 2/3$					
1	100	.048	.043	.041	.045	.054	.056						
	200	.047	.048	.046	.048	.054	.056						
	500	.049	.046	.048	.045	.053	.059						
0.2	100	.029	.034	.029	.407	.067	.745	.042	.040	.039	.132	.069	.206
	200	.038	.034	.028	.413	.061	.730	.049	.043	.042	.137	.064	.203
	500	.042	.042	.037	.432	.074	.755	.045	.047	.041	.136	.066	.201
0.33	100	.035	.040	.033	.225	.073	.477	.041	.046	.038	.107	.074	.178
	200	.039	.042	.037	.230	.070	.477	.049	.048	.042	.110	.056	.162
	500	.045	.045	.040	.238	.076	.472	.046	.050	.047	.117	.064	.176
3	100	.050	.049	.047	.036	.033	.036	.044	.049	.043	.079	.022	.131
	200	.049	.047	.047	.033	.023	.032	.048	.047	.045	.078	.016	.129
	500	.045	.049	.045	.033	.016	.030	.045	.046	.042	.079	.012	.130
5	100	.052	.044	.044	.039	.032	.033	.043	.047	.042	.141	.013	.240
	200	.042	.045	.038	.033	.018	.037	.054	.048	.051	.152	.010	.245
	500	.049	.052	.047	.037	.011	.035	.047	.051	.048	.145	.004	.245
		$\tau = 1/3$						$\tau = 4/5$					
0.2	100	.030	.036	.030	.345	.072	.637	.050	.047	.048	.084	.064	.118
	200	.038	.040	.035	.333	.072	.615	.048	.046	.044	.087	.063	.120
	500	.043	.045	.039	.347	.065	.612	.052	.051	.050	.089	.059	.116
0.33	100	.039	.041	.034	.201	.074	.429	.044	.040	.037	.081	.060	.106
	200	.040	.041	.037	.186	.065	.410	.053	.045	.042	.082	.058	.110
	500	.045	.050	.043	.208	.075	.402	.050	.051	.048	.083	.057	.104
3	100	.048	.045	.042	.049	.027	.049	.041	.050	.043	.083	.026	.164
	200	.050	.049	.047	.048	.016	.051	.051	.049	.047	.080	.018	.155
	500	.048	.051	.046	.050	.016	.048	.044	.045	.043	.079	.016	.151
5	100	.046	.050	.046	.053	.017	.063	.048	.048	.046	.154	.012	.340
	200	.044	.047	.040	.050	.010	.055	.047	.046	.044	.166	.006	.353
	500	.046	.053	.047	.051	.007	.056	.051	.049	.046	.170	.007	.343

Note:  $\delta$  gives the break direction of the errors,  $\tau$  the break fraction.  $t_{\hat{\alpha},i}$  is defined in Prop. 2,  $Q_{\alpha}$  in Prop. 3. EG is by Engle and Granger (1987),  $\tilde{\tau}_i$  by Chang and Nguyen (2011) and  $\lambda_{tr}$  is by Johansen (1995). 5,000 replications, 5% nominal size.

cointegration test  $Q_{\alpha}$ . There are some mild upward distortions under negative breaks for  $T = 100$  that however vanish as  $T$  increases. The tests also handle the benchmark case  $\delta = 1$ . We conclude that only  $t_{\hat{\alpha}}$  and  $Q_{\alpha}$  are promising candidates to construct a panel cointegration statistic under time-varying volatility, and therefore waive to consider panel offsprings of the other tests considered above such as Pedroni (2004), Larsson, Lyhagen and Löthgren (2001) or Chang and Nguyen (2011) in what follows.

## 4.2 Panel Evidence

We augment the time series DGP as follows to investigate the performance of the new tests:

$$\Delta \mathbf{w}_{it} = \boldsymbol{\alpha}_i \boldsymbol{\beta}'_i \mathbf{w}_{it-1} + \boldsymbol{\epsilon}_{it}, \quad t = 1, \dots, T, \quad i = 1, \dots, N.$$

In order to investigate the effect of cross-sectional dependence on the tests, we consider the following factor structure for the disturbances:  $\epsilon_{k,it} = \lambda_i \nu_t + \tilde{\epsilon}_{k,it}$ ,  $k = 1, \dots, K$ , where  $\tilde{\epsilon}_{it} = \Sigma_{it}^{0.5} \bar{\boldsymbol{\epsilon}}_t$  and  $\bar{\boldsymbol{\epsilon}}_t \sim iid\mathcal{N}(0, I_K)$ . Here,  $\Sigma_{it}^{0.5}$  is a time-varying  $K \times K$  diagonal matrix as in Section 4.1. The  $\lambda_i$  are drawn from a Uniform $[-1, 2]$  distribution, while  $\nu_t \sim \mathcal{N}(0, 1)$ . We again take  $K = 2$  and consider early and late break fractions  $\tau \in \{1/3, 2/3\}$  and large negative to large positive variance shifts  $\delta \in \{1/5, 1/3, 1, 3, 5\}$ . The case  $\delta = 1$  again covers the homoskedastic scenario. Size results obtain for  $\boldsymbol{\alpha}_i = \mathbf{0}$ ,  $i = 1, \dots, N$ . The power study uses and  $\boldsymbol{\beta}'_i = (1 \ 0)$  and  $\boldsymbol{\alpha}_i = (-.2 \ 0)'$ . The panel dimensions are  $T = \{100, 200, 500\}$  and  $N = \{20, 50, 100\}$ . The experiments use 5,000 replications, and all results are for a nominal 5% level.

Table 2: Size of the Robust Panel Cointegration Test  $\tilde{Q}$  under Nonstationary Volatility

$\delta$	$\tau$	$N$	0.2			0.33			0.66			0.8		
			20	50	100	20	50	100	20	50	100	20	50	100
0.2	100		.066	.096	.137	.054	.078	.117	.054	.071	.087	.049	.062	.088
	200		.049	.066	.089	.046	.063	.080	.048	.055	.068	.045	.053	.068
	500		.042	.056	.056	.044	.051	.060	.047	.051	.054	.046	.047	.056
0.33	100		.057	.080	.109	.055	.071	.103	.046	.071	.084	.050	.064	.088
	200		.049	.058	.078	.045	.052	.073	.049	.047	.061	.043	.052	.068
	500		.050	.049	.055	.050	.052	.057	.049	.043	.053	.042	.049	.056
1	100		.044	.064	.088	.053	.060	.085	.051	.070	.090	.051	.064	.084
	200		.049	.055	.066	.052	.058	.067	.048	.055	.066	.048	.049	.068
	500		.047	.054	.055	.045	.047	.056	.043	.048	.058	.044	.052	.054
3	100		.056	.071	.090	.061	.066	.096	.053	.073	.101	.058	.066	.091
	200		.055	.063	.079	.050	.059	.083	.048	.061	.072	.045	.059	.078
	500		.047	.058	.069	.047	.057	.057	.045	.047	.065	.046	.053	.059
5	100		.052	.074	.092	.061	.080	.106	.058	.076	.109	.053	.084	.094
	200		.052	.060	.080	.059	.065	.082	.049	.057	.082	.051	.059	.079
	500		.047	.056	.064	.048	.060	.071	.049	.056	.067	.050	.059	.060

Note:  $\delta$  gives the break direction of the errors,  $\tau$  the break fraction.  $\tilde{Q}$  is defined in Prop. 6. 5,000 replications, 5% nominal size.

Table 2 reports size results for  $\tilde{Q}$ . In general, the empirical size using the asymptotic critical values is close to the nominal size. Size is somewhat less accurate in short, but wide ( $N = 100$ ) panels. This is expected as the time series size distortions then tend to accumulate in the panel statistic. These distortions however vanish as  $T \rightarrow \infty$ . There seems

Table 3: Power of the Robust Panel Cointegration Test  $\tilde{Q}$  under Nonstationary Volatility

$\delta$	$\tau$	$N$	0.2			0.33			0.66			0.8		
			20	50	100	20	50	100	20	50	100	20	50	100
0.2	100		.191	.283	.365	.201	.300	.395	.211	.342	.479	.224	.369	.519
	200		.392	.536	.644	.406	.585	.705	.429	.674	.813	.466	.711	.855
	500		.855	.954	.988	.883	.972	.996	.922	.986	.999	.931	.993	.999
0.33	100		.192	.313	.409	.190	.314	.445	.211	.358	.530	.230	.375	.555
	200		.392	.604	.746	.443	.647	.787	.450	.719	.865	.479	.741	.892
	500		.899	.983	.998	.918	.991	.999	.938	.995	1.00	.947	.998	1.00
1	100		.222	.420	.609	.240	.406	.599	.229	.402	.592	.235	.401	.587
	200		.482	.777	.921	.476	.756	.932	.484	.770	.931	.479	.766	.920
	500		.948	.999	1.00	.948	.998	1.00	.951	.999	1.00	.953	.999	1.00
3	100		.314	.541	.772	.334	.581	.784	.375	.615	.816	.339	.585	.780
	200		.606	.889	.987	.645	.912	.989	.715	.944	.993	.701	.938	.991
	500		.982	1.00	1.00	.990	1.00	1.00	.996	1.00	1.00	.996	1.00	1.00
5	100		.302	.536	.764	.335	.577	.797	.466	.744	.899	.468	.759	.908
	200		.595	.890	.990	.668	.926	.993	.842	.983	.999	.874	.990	.999
	500		.982	1.00	1.00	.994	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Note: See notes to Table 2.

to be no discernible pattern to describe the distortions as functions of  $\delta$  or  $\tau$ , suggesting that the type of break (early vs. late and downward vs. upward) matters comparatively little.

Table 3 reports power results. As expected from Proposition 6b, power grows in  $T$ . It also grows quite quickly in  $N$ , justifying the use of panel data. The rejection rate of  $\tilde{Q}$  also increases in  $\delta$  and (moderately) in  $\tau$ , so that  $\tilde{Q}$  is most effective in detecting panel cointegration under late upward breaks in the variance.

## 5 Concluding remarks

We proposed single-unit and panel tests for no error correction and for no cointegration robust to unconditional heteroskedasticity in the time dimension as well as to cross-unit correlation or cross-unit cointegration.

Our tests are based on the error correction representation of Banerjee, Dolado and Mestre (1998), to which the nonlinear instrumental variable method of Chang (2002) was applied. The proposed test statistics were shown to follow standard distributions asymptotically and to require no exogeneity assumptions. Moreover, the asymptotic null distributions are not affected by nonstationary volatility if White standard errors are used. We found, however, that the residuals should not be computed using nonlinear IV estimators of the parameters, since the IV estimators may not converge fast enough to ensure the consis-

tency of the residuals. Instead, one should estimate the residuals under the null hypothesis or use OLS estimators.

In cross-correlated as well as in cross-cointegrated panels with unconditional heteroskedasticity in the time dimension, individual test statistics were shown to be asymptotically independent; thus, panel tests for error-correction or panel cointegration robust to cross-dependence can be built. Panels may also be unbalanced up to a certain degree, and no  $N$  asymptotics are required. The appeal of using the nonlinear IV methodology for panel cointegration testing is that no correction for cross-unit dependence and global heteroskedasticity is required. The tests performed reliably in finite-dimensional panels with time-varying volatility and factor-driven shocks.

## Appendix

Assume throughout the appendix that enough initial values are available so that all sums run from  $t = 1$  to  $T$ . The proofs of the propositions stated in the paper require the following lemma.

**Lemma A** Let  $X(s)$  be some continuous Gaussian process, and denote by  $\mathcal{L}_X(s, x)$  the chronological local time of  $X$ ,

$$\mathcal{L}_X(s, x) = \lim_{\epsilon \rightarrow 0} \frac{1}{2\epsilon} \int_0^s \mathbf{1}(|X(r) - x| < \epsilon) dr$$

and by  $\tilde{\mathcal{L}}_X^g(s, x)$  a weighted version of the local time,

$$\tilde{\mathcal{L}}_X^g(s, x) = \lim_{\epsilon \rightarrow 0} \frac{1}{2\epsilon} \int_0^s \mathbf{1}(|X(r) - x| < \epsilon) dg(r),$$

with  $g$  a piecewise continuous function and  $\mathbf{1}(\cdot)$  the usual indicator function,  $\mathbf{1}(A) = 1$ , if proposition  $A$  is true, and 0, otherwise.<sup>7</sup>

Denote by  $\omega_i^2(s)$  the first diagonal element of the  $K \times K$  matrix

$$\Omega_i = \left( I_K - \sum_{j=1}^p A_{ij} \right)^{-1} \int_0^s \Sigma_i(r) dr \left( I_K - \sum_{j=1}^p A'_{ij} \right)^{-1}$$

where  $\Sigma_i(\cdot)$  is the  $i$ th of the  $N$   $K \times K$  diagonal blocks of  $\Sigma(\cdot)$  and by  $\eta_i^2(s)$  the (unscaled) variance profile

$$\eta_i^2(s) = \int_0^s \sigma_i^2(r) dr$$

---

<sup>7</sup>If  $g(s)$  is the quadratic variation of  $X$ ,  $g(s) = \int_0^s (dX(r))^2$ ,  $\tilde{\mathcal{L}}_X^g(s, x)$  is the local time given in terms of its quadratic variation.

with  $\sigma_i^2(\cdot)$  the  $(N(i-1)+1)$ st diagonal element of  $\Sigma(\cdot)$ .

Finally, let  $W(s)$  and  $\widetilde{W}(s)$  be two independent copies of the standard Wiener processes. Under the assumptions of Proposition 1, it holds as  $T \rightarrow \infty$  that:

A.0

$$t^{0.5} E |F(y_{t-1})| < C < \infty \quad \forall t > 0;$$

A.1

$$\begin{aligned} \frac{1}{T^{0.5}} \sum_{t=1}^T F^2(y_{t-1}) &\xrightarrow{d} \mathcal{L}_{W(\omega(\cdot))}(1, 0) \int_{-\infty}^{\infty} F^2(x) dx; \\ \frac{1}{T^{0.5}} \sum_{t=1}^T y_{t-1} F(y_{t-1}) &\xrightarrow{d} \mathcal{L}_{W(\omega(\cdot))}(1, 0) \int_{-\infty}^{\infty} x F(x) dx; \end{aligned}$$

A.2

$$\frac{1}{T^{0.5}} \sum_{t=1}^T F^2(y_{t-1}) \widehat{\varepsilon}_t^2 \xrightarrow{d} \widetilde{\mathcal{L}}_{W(\omega(\cdot))}^{\eta^2}(1, 0) \int_{-\infty}^{\infty} F^2(x) dx;$$

A.3

$$\frac{1}{T^{0.25}} \sum_{t=1}^T F(y_{t-1}) \varepsilon_t \xrightarrow{d} \sqrt{\widetilde{\mathcal{L}}_{W(\omega(\cdot))}^{\eta^2}(1, 0) \int_{-\infty}^{\infty} F^2(x) dx} \cdot \widetilde{W}(1);$$

A.4

$$\frac{1}{T^{0.5}} \sum_{t=1}^T F(y_{t-1}) \mathbf{x}'_{t-1,0} = O_p(1);$$

A.5

$$\frac{1}{T^{0.75}} \sum_{t=1}^T \widehat{\varepsilon}_t^2 F(y_{t-1}) \mathbf{x}'_{t-1,0} = O_p(1);$$

A.6

$$\frac{1}{T} \sum_{t=1}^T \mathbf{x}_{t-1,0} \mathbf{x}'_{t-1,0} \widehat{\varepsilon}_t^2 = O_p(1);$$

A.7

$$\frac{1}{T} \sum_{t=1}^T \mathbf{x}_{t-1,0} \mathbf{x}'_{t-1,1} \widehat{\varepsilon}_t^2 = O_p(1);$$

A.8

$$\frac{1}{T^2} \sum_{t=1}^T \mathbf{x}_{t-1,1} \mathbf{x}'_{t-1,1} \widehat{\varepsilon}_t^2 = O_p(1);$$

If, additionally,  $\boldsymbol{\alpha}_e = \mathbf{0}$ , it holds that

A.9

$$\frac{1}{T} \sum_{t=1}^T F(y_{t-1}) \mathbf{x}'_{t-1,1} = O_p(1);$$

$$\frac{1}{T^{1.25}} \sum_{t=1}^T \widehat{\varepsilon}_t^2 F(y_{t-1}) \mathbf{x}'_{t-1,1} = O_p(1);$$

if  $\boldsymbol{\alpha}_e \neq \mathbf{0}$ , a different behavior of this sample cross-moment can emerge, depending on  $\boldsymbol{\theta}$ :

A.9\*

$$\frac{1}{T} \sum_{t=1}^T F(y_{t-1}) \mathbf{x}'_{t-1,1} = O_p(T^{-0.5}).$$

$$\frac{1}{T^{0.75}} \sum_{t=1}^T \widehat{\varepsilon}_t^2 F(y_{t-1}) \mathbf{x}'_{t-1,1} = O_p(1).$$

### Proof of Lemma A

A.0 Follows from Jeganathan (2008, Lemma 8), his assumption on the characteristic function being fulfilled, and we assume stronger moment conditions.

A.1 Follows directly from Wang and Phillips (2009), Corollary 2.2(ii), since finiteness of  $\int xF(x)dx$  together with continuity of  $F(\cdot)$  implies finiteness of  $\int F(x)dx$  and thus of  $\int F^2(x)dx$  (recall that continuity implies that  $F(\cdot)$  has no poles).

A.2 We first show that

$$\frac{1}{T^{0.5}} \sum_{t=1}^T F^2(y_{t-1}) \widehat{\varepsilon}_t^2 = \frac{1}{T^{0.5}} \sum_{t=1}^T F^2(y_{t-1}) \varepsilon_t^2 + o_p(1) \quad (14)$$

as follows. From (4), we have that

$$\begin{aligned} \frac{1}{T^{0.5}} \sum_{t=1}^T F^2(y_{t-1}) \widehat{\varepsilon}_t^2 &= \frac{1}{T^{0.5}} \sum_{t=1}^T F^2(y_{t-1}) \varepsilon_t^2 \\ &\quad - \frac{2}{T^{0.5}} \sum_{t=1}^T F^2(y_{t-1}) \left( (\widehat{\alpha} - \alpha) y_{t-1} + (\widehat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0) \mathbf{x}_{t-1,0} + (\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) \mathbf{x}_{t-1,1} \right) \\ &\quad + \frac{1}{T^{0.5}} \sum_{t=1}^T F^2(y_{t-1}) \left( (\widehat{\alpha} - \alpha) y_{t-1} + (\widehat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0) \mathbf{x}_{t-1,0} + (\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) \mathbf{x}_{t-1,1} \right)^2. \end{aligned}$$

The second term on the r.h.s. is given by

$$\frac{2(\widehat{\alpha} - \alpha)}{T^{0.5}} \sum_{t=1}^T F^2(y_{t-1}) y_{t-1} + \frac{2(\widehat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0)}{T^{0.5}} \sum_{t=1}^T F^2(y_{t-1}) \mathbf{x}_{t-1,0} + \frac{2(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1)}{T^{0.5}} \sum_{t=1}^T F^2(y_{t-1}) \mathbf{x}_{t-1,1}$$



and is thus  $o_p(1)$  thanks to A.1, A.4 and A.9 (or A.9\*) and to noting that the odd function  $x F^2(x)$  satisfies the conditions of Wang and Phillips (2011). Similar arguments apply for the third term and (14) is established.

Second, we show that

$$\frac{1}{T^{0.5}} \sum_{t=1}^T F^2(y_{t-1}) \varepsilon_t^2 = \frac{1}{T^{0.5}} \sum_{t=1}^T F^2(y_{t-1}) \sigma_t^2 + o_p(1)$$

with  $\sigma_t^2 = \text{Var}(\varepsilon_t^2)$ . Since  $F^2(y_{t-1})(\varepsilon_t^2 - \sigma_t^2)$  is a martingale difference sequence and  $\varepsilon_t$  an independent sequence,

$$\text{Var} \left( \frac{1}{T^{0.5}} \sum_{t=1}^T F^2(y_{t-1}) (\varepsilon_t^2 - \sigma_t^2) \right) = \frac{1}{T} \sum_{t=1}^T \text{Var}(F^2(y_{t-1})) \text{Var}((\varepsilon_t^2 - \sigma_t^2)),$$

and the result follows with A.0. It remains to be proved that

$$\frac{1}{T^{0.5}} \sum_{t=1}^T F^2(y_{t-1}) \sigma_t^2 \xrightarrow{d} \tilde{\mathcal{L}}_{W(\omega^2(\cdot))}^{\eta^2}(1, 0),$$

which follows with a tedious, yet straightforward adaptation of the arguments of Wang and Phillips (2009).

A.3 The key result required for the indicated mixed Gaussian distribution to arise in the limit is the independence of the Wiener process  $\widetilde{W}$  and the mixing variable given in terms of the weighted local time of a time-transformed Wiener process. It has been extensively studied for the homoskedastic case, see e.g. Phillips, Park and Chang (2004) or Jeganathan (2008). But the time transformation is only a scaling operation which does not affect independence and the result follows thanks to A.2.

A.4 is a direct consequence of Lemma 1, item *c*), in Demetrescu (2009), his assumptions being fulfilled under ours.

A.5 The arguments used in the proof of A.2 indicate that

$$\frac{1}{T^{0.75}} \sum_{t=1}^T \widehat{\varepsilon}_t^2 F(y_{t-1}) \mathbf{x}'_{t-1,0} = \frac{1}{T^{0.75}} \sum_{t=1}^T \varepsilon_t^2 F(y_{t-1}) \mathbf{x}'_{t-1,0} + o_p(1);$$

the result follows with an application of the Cauchy-Schwarz inequality and the finiteness of 4th order moments of  $\varepsilon_t$ .

A.6 Obvious and omitted.

A.7 Follows with arguments similar to those used in the proof of Proposition 1 in Demetrescu (2010).

A.8 Follows with the same arguments used to establish A.7.

A.9 The first relation is shown to hold true under no cointegration (i.e.  $\alpha = 0$  and  $\boldsymbol{\alpha}_e = \mathbf{0}$ ) by Chang, Park and Phillips (2001, Lemma 5), and is easily shown to hold under our moment restrictions. The second relation follows with arguments similar to A.5.

A.9\* Under cointegration of  $y_{t-1}$  and  $\mathbf{x}_{t-1,1}$  (as implied by  $\boldsymbol{\alpha}_e \neq \mathbf{0}$ ), each element of  $\mathbf{x}_{t-1,1}$  can be expressed as a linear combination of I(1) variables that are either cointegrated with  $y_{t-1}$  or not. When at least one of the I(1) variables that are not cointegrated with  $y_{t-1}$  is present in the linear combinations, the cross-moment is of magnitude  $O_p(1)$  due to A.9; otherwise, when elements of  $\mathbf{x}_{t-1,1}$  equal  $y_{t-1}$  plus I(0) noise, the  $O_p(T^{-0.5})$  magnitude order emerges due to A.1, second relation, and A.4.

The second relation follows along the lines of the proof of A.9 and is omitted.

### Proof of Proposition 1

a) Recall,  $\hat{\alpha} - \alpha = Q^{-1}M$ , with  $M$  and  $Q$  defined in the text. Let us now examine the behavior of  $M$ . The second term on the right-hand side of the equation defining  $M$  can be written as

$$\left( \sum_{t=1}^T F(y_{t-1}) \mathbf{x}'_{t-1} \right) D_T^{-1} D_T \left( \sum_{t=1}^T \mathbf{x}_{t-1} \mathbf{x}'_{t-1} \right)^{-1} D_T D_T^{-1} \left( \sum_{t=1}^T \mathbf{x}_{t-1} \varepsilon_t \right)$$

with  $D_T$  a  $K(p+1) \times K(p+1)$  diagonal matrix partitioned according to the stationary and integrated components of  $\mathbf{x}_{t-1}$ :

$$D_T = \begin{pmatrix} T^{0.5} I_{Kp} & 0 \\ 0 & T I_K \end{pmatrix}, \quad D_T^{-1} = \begin{pmatrix} T^{-0.5} & 0 \\ 0 & T^{-1} \end{pmatrix}.$$

It follows from A.4 and A.9 that

$$\begin{aligned} & \left( \sum_{t=1}^T F(y_{t-1}) \mathbf{x}'_{t-1} \right) \begin{pmatrix} T^{-0.5} & 0 \\ 0 & T^{-1} \end{pmatrix} \\ &= \left( T^{-0.5} \sum_{t=1}^T F(y_{t-1}) \mathbf{x}'_{t-1,0}, T^{-1} \sum_{t=1}^T F(y_{t-1}) \mathbf{x}'_{t-1,1} \right) \\ &= (O_p(1), O_p(1)) \end{aligned}$$

and

$$\begin{pmatrix} T^{-0.5} & 0 \\ 0 & T^{-1} \end{pmatrix} \begin{pmatrix} T \\ \sum_{t=1}^T \mathbf{x}_{t-1} \varepsilon_t \end{pmatrix} = \begin{pmatrix} T^{-0.5} \sum_{t=1}^T \mathbf{x}_{t-1,0} \varepsilon_t \\ T^{-1} \sum_{t=1}^T \mathbf{x}_{t-1,1} \varepsilon_t \end{pmatrix} = \begin{pmatrix} O_p(1) \\ O_p(1) \end{pmatrix},$$

Then,

$$D_T \left( \sum_{t=1}^T \mathbf{x}_{t-1} \mathbf{x}'_{t-1} \right)^{-1} D_T = \left( D_T^{-1} \left( \sum_{t=1}^T \mathbf{x}_{t-1} \mathbf{x}'_{t-1} \right) D_T^{-1} \right)^{-1}$$

which equals

$$\begin{aligned} & \left( \begin{pmatrix} T^{-0.5} & 0 \\ 0 & T^{-1} \end{pmatrix} \begin{pmatrix} \sum \mathbf{x}_{t-1,0} \mathbf{x}'_{t-1,0} & \sum \mathbf{x}_{t-1,0} \mathbf{x}'_{t-1,1} \\ \sum \mathbf{x}_{t-1,1} \mathbf{x}'_{t-1,0} & \sum \mathbf{x}_{t-1,1} \mathbf{x}'_{t-1,1} \end{pmatrix} \begin{pmatrix} T^{-0.5} & 0 \\ 0 & T^{-1} \end{pmatrix} \right)^{-1} \\ &= \begin{pmatrix} T^{-1} \sum \mathbf{x}_{t-1,0} \mathbf{x}'_{t-1,0} & T^{-1.5} \sum \mathbf{x}_{t-1,0} \mathbf{x}'_{t-1,1} \\ T^{-1.5} \sum \mathbf{x}_{t-1,1} \mathbf{x}'_{t-1,0} & T^{-2} \sum \mathbf{x}_{t-1,1} \mathbf{x}'_{t-1,1} \end{pmatrix}^{-1} \\ &= \begin{pmatrix} O_p(1) & O_p(T^{-0.5}) \\ O_p(T^{-0.5}) & O_p(1) \end{pmatrix}^{-1} = \begin{pmatrix} O_p(1) & O_p(T^{-0.5}) \\ O_p(T^{-0.5}) & O_p(1) \end{pmatrix}, \end{aligned}$$

due to continuity of matrix inversion and nonsingularity. Hence,

$$\begin{aligned} M &= \sum_{t=1}^T F(y_{t-1}) \varepsilon_t - (O_p(1), O_p(1)) \begin{pmatrix} O_p(1) & O_p(T^{-0.5}) \\ O_p(T^{-0.5}) & O_p(1) \end{pmatrix} \begin{pmatrix} O_p(1) \\ O_p(1) \end{pmatrix} \\ &= \sum_{t=1}^T F(y_{t-1}) \varepsilon_t - O_p(1) = O_p(T^{0.25}). \end{aligned}$$

For  $Q$ , we only need to examine

$$\begin{pmatrix} T^{-0.5} & 0 \\ 0 & T^{-1} \end{pmatrix} \begin{pmatrix} T \\ \sum_{t=1}^T \mathbf{x}_{t-1} y_{t-1} \end{pmatrix} = \begin{pmatrix} T^{0.5} \left( T^{-1} \sum_{t=1}^T \mathbf{x}_{t-1,0} y_{t-1} \right) \\ T \left( T^{-2} \sum_{t=1}^T \mathbf{x}_{t-1,1} y_{t-1} \right) \end{pmatrix} = \begin{pmatrix} O_p(T^{0.5}) \\ O_p(T) \end{pmatrix},$$

which, under no cointegration, leads to

$$\begin{aligned} Q &= \sum_{t=1}^T F(y_{t-1}) y_{t-1} - (O_p(1), O_p(1)) \begin{pmatrix} O_p(1) & O_p(T^{-0.5}) \\ O_p(T^{-0.5}) & O_p(1) \end{pmatrix} \begin{pmatrix} O_p(T^{0.5}) \\ O_p(T) \end{pmatrix} \\ &= O_p(T); \end{aligned}$$

note that it is an exact order of magnitude. The convergence rate for  $\hat{\alpha}$  follows directly,

$$\hat{\alpha} - \alpha = O_p(T^{-1}) O_p(T^{0.25}) = O_p(T^{-0.75}).$$

For  $\beta$ , we have

$$\hat{\beta} - \beta = J^{-1} R$$

with  $R$  a column vector:

$$R = \sum_{t=1}^T \mathbf{x}_{t-1} \varepsilon_t - \sum_{t=1}^T \mathbf{x}_{t-1} y_{t-1} \left( \sum_{t=1}^T F(y_{t-1}) y_{t-1} \right)^{-1} \sum_{t=1}^T F(y_{t-1}) \varepsilon_t$$

and  $J$  a matrix

$$J = \sum_{t=1}^T \mathbf{x}_{t-1} \mathbf{x}'_{t-1} - \sum_{t=1}^T \mathbf{x}_{t-1} y_{t-1} \left( \sum_{t=1}^T F(y_{t-1}) y_{t-1} \right)^{-1} \sum_{t=1}^T F(y_{t-1}) \mathbf{x}'_{t-1}.$$

Partition  $R$  and  $J$  corresponding to the stationary and integrated regressors. Then, it is straightforward to check that for  $R = (R'_0, R'_1)'$  it holds

$$R_0 = O_p(T^{0.75}) \quad \text{and} \quad R_1 = O_p(T^{1.75}).$$

Then, for

$$J = \begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

it holds due to Lemma A that  $A = O_p(T)$ ,  $B = O_p(T^{1.5})$ ,  $C = O_p(T^2)$  and  $D = O_p(T^{2.5})$ . Using formulae for inverting partitioned matrices (see Lütkepohl, 1996, p. 147), one obtains, after some algebra,

$$J^{-1} = \begin{pmatrix} O_p(T^{-1}) & O_p(T^{-2}) \\ O_p(T^{-1.5}) & O_p(T^{-2.5}) \end{pmatrix},$$

from which the desired convergence rates follow.

**b)** The result follows along the same lines, but now A.9\* could hold instead of A.9 (see the proof of Lemma A for details). While  $M$  remains in any case of order  $O_p(T^{0.25})$ ,  $Q$  can be of exact order  $O_p(T^{0.5})$  instead of  $O_p(T)$ , resulting in a convergence order for  $\hat{\alpha}$  of  $O_p(T^{0.25})$ . In what concerns  $\hat{\beta}$ ,  $B$  may be  $O_p(T)$  and  $D$  may be  $O_p(T^2)$ , if A.9\* holds. Thus, a behavior similar to that of  $\hat{\alpha}$  emerges for  $\hat{\beta}_1$ , while the behavior of  $\hat{\beta}_0$  is unaffected.

**Proof of Lemma 1** From the proof of Proposition 1, it follows that

$$M = \frac{1}{T^{0.25}} \sum_{t=1}^T F(y_{t-1}) \varepsilon_t + o_p(1).$$

By using arguments similar to those employed in the proofs of Proposition 1 and of Lemma A.2, we further have that

$$\frac{1}{T^{0.5}} P_W = \frac{1}{T^{0.5}} \sum_{t=1}^T F(y_{t-1})^2 \varepsilon_t^2 + o_p(1),$$

which leads to the desired result, the numerator of the  $t$  statistic being different from zero with probability 1. Note that the result holds irrespective of whether A.9 or A.9\* holds true.

**Proof of Proposition 2** The result follows for a) directly from Lemma 1 by setting  $\alpha = 0$  and employing A.2 and A.3.

If using the usual standard errors,  $P_W$  reduces to  $\hat{\sigma}_\varepsilon^2 P = \hat{\sigma}_\varepsilon^2 \sum_{t=1}^T F^2(y_{t-1}) + o_p(T^{0.5})$  which (properly normalized) has the wrong weak limit (see lemma A.1), and the resulting  $t$  statistic is not pivotal in the presence of time-varying volatility.

For b), note that

$$t_{\hat{\alpha}} = \frac{\alpha + Q^{-1}M}{\sqrt{Q^{-2}P_W}};$$

from the proof of Proposition 1 we know the exact order of magnitude of  $Q$  to be either  $T$  or  $T^{0.5}$ ; with  $\alpha \neq 0$ , its numerator is non-zero w.p.1 and  $O_p(1)$ , hence the  $t$  statistic diverges at a rate of at least  $T^{0.25}$ .

**Proof of Proposition 3** For a), let the number of equations be w.l.o.g.  $K = 2$ . Dealing with all  $K$  equations at the same time, it is more convenient to revert to the notation of Assumption 1, i.e. to use  $\mathbf{w}_t$  instead of  $(y_t, x_t)'$ . Note that, under Assumptions 1 through 3, the normalized levels converge weakly to a Gaussian process with a time-dependent quadratic variation,

$$\frac{1}{\sqrt{T}} \mathbf{w}_{[sT]} \Rightarrow \mathbf{B}(s),$$

where the covariance kernel of  $\mathbf{B}(s) = (B_1(s), B_2(s))'$  depends on  $\Sigma(s)$  and the short-run dynamics of  $\mathbf{w}_t$ . Marginally, but not jointly,  $B_1$  and  $B_2$  are time-transformed Wiener processes.

Given the (mixed) Gaussianity from Proposition 2, we only need to prove that the ratios

$$\frac{\sum_{t=1}^T F_1(w_{1t-1}) \epsilon_{1t}}{\sqrt{\sum_{t=1}^T F_1^2(w_{1t-1}) \epsilon_{1t}^2}} \quad \text{and} \quad \frac{\sum_{t=1}^T F_2(w_{2t-1}) \epsilon_{2t}}{\sqrt{\sum_{t=1}^T F_2^2(w_{2t-1}) \epsilon_{2t}^2}} \quad (15)$$

are asymptotically independent. Following Chang (2002) or Chang and Nguyen (2011), asymptotic independence of the two  $t$  statistics in (15) follows if

$$\int_0^1 F_1(\sqrt{T}B_1(s)) F_2(\sqrt{T}B_2(s)) \sigma_{12}(s) ds = o_p(T^{-0.5}),$$

where  $\sigma_{12}(s)$  is the off-diagonal element of  $\Sigma(s)$ . Chang's derivation of this magnitude order does not apply directly because  $\mathbf{B}(s)$ , while it is a Gaussian process, it is not necessarily a bivariate non-degenerate Brownian motion.

To adapt Chang's argument to our situation, note that there exist a positive definite lower triangular matrix function  $A(s)$  and a vector of independent standard Wiener processes  $\mathbf{W}(s)$  such that

$$\mathbf{B}(s) \stackrel{d}{=} \tilde{\mathbf{B}}(s) = A(s) \mathbf{W}(s);$$

the elements of  $A(s)$  are piecewise Lipschitz and bounded, properties inherited from  $\Sigma(s)$ . Since, as pointed out e.g. by de Jong and Wang (2005), there exists a positive, integrable, continuous and symmetric function  $\tilde{F}(s) > |F(s)|$  monotonically increasing (decreasing) for  $s < 0$  ( $s > 0$ ), with  $\tilde{F}$  fulfilling Assumption 4 as well, we have that

$$\begin{aligned} & \left| \int_0^1 F_1 \left( \sqrt{T} \tilde{B}_1(s) \right) F_2 \left( \sqrt{T} \tilde{B}_2(s) \right) \sigma_{12}(s) ds \right| \\ & \leq \max_{s \in [0,1]} |\sigma_{12}(s)| \int_0^1 \tilde{F} \left( \sqrt{T} |\tilde{B}_1(s)| \right) \tilde{F} \left( \sqrt{T} |\tilde{B}_2(s)| \right) ds. \end{aligned}$$

Now,

$$\begin{aligned} \tilde{F} \left( \sqrt{T} |\tilde{B}_1(s)| \right) &= \tilde{F} \left( \sqrt{T} |a_{11}(s)| |W_1(s)| \right) \\ &\leq \tilde{F} \left( \sqrt{T} |W_1(s)| \min_{s \in [0,1]} |a_{11}(s)| \right) \\ &= \tilde{G}_1 \left( \sqrt{T} |W_1(s)| \right) \end{aligned}$$

since  $\sqrt{T} |a_{11}(s)| |W_1(s)| \geq |W_1(s)| \min_{s \in [0,1]} |a_{11}(s)|$  for all  $s \in [0, 1]$ . The r.h.s. of the last equality only depends on  $W_1$ . Also,

$$\left| \sqrt{T} a_{21}(s) W_1(s) + \sqrt{T} a_{22}(s) W_2(s) \right| \geq \sqrt{T} \left| |a_{21}(s)| |W_1(s)| - |a_{22}(s)| |W_2(s)| \right|,$$

where the r.h.s. is in turn bounded from below by

- $\sqrt{T} \left| |W_1(s)| \min_{s \in [0,1]} |a_{21}(s)| - |W_2(s)| \max_{s \in [0,1]} |a_{22}(s)| \right|$   
if  $|W_2(s)| < |W_1(s)| \frac{\min_{s \in [0,1]} |a_{21}(s)|}{\max_{s \in [0,1]} |a_{22}(s)|}$ ,
- 0 if  $|W_1(s)| \frac{\min_{s \in [0,1]} |a_{21}(s)|}{\max_{s \in [0,1]} |a_{22}(s)|} \leq |W_2(s)| \leq |W_1(s)| \frac{\max_{s \in [0,1]} |a_{21}(s)|}{\min_{s \in [0,1]} |a_{22}(s)|}$ , and
- $\sqrt{T} \left| |W_1(s)| \max_{s \in [0,1]} |a_{21}(s)| - |W_2(s)| \min_{s \in [0,1]} |a_{22}(s)| \right|$   
if  $|W_2(s)| > |W_1(s)| \frac{\max_{s \in [0,1]} |a_{21}(s)|}{\min_{s \in [0,1]} |a_{22}(s)|}$ .

It follows that  $\tilde{F} \left( \sqrt{T} \left| |a_{21}(s)| |W_1(s)| - |a_{22}(s)| |W_2(s)| \right| \right)$  is bounded (from above) by the function  $\tilde{G}_2 \left( \sqrt{T} |W_1(s)|, \sqrt{T} |W_2(s)| \right)$  given by

- $\tilde{F} \left( \sqrt{T} \left| |W_1(s)| \min_{s \in [0,1]} |a_{21}(s)| - |W_2(s)| \max_{s \in [0,1]} |a_{22}(s)| \right| \right)$   
if  $|W_2(s)| < |W_1(s)| \frac{\min_{s \in [0,1]} |a_{21}(s)|}{\max_{s \in [0,1]} |a_{22}(s)|}$ ,

- $\tilde{F}(0)$  if  $|W_1(s)| \frac{\min_{s \in [0,1]} |a_{21}(s)|}{\max_{s \in [0,1]} |a_{22}(s)|} \leq |W_2(s)| \leq |W_1(s)| \frac{\max_{s \in [0,1]} |a_{21}(s)|}{\min_{s \in [0,1]} |a_{22}(s)|}$ , and
- $\tilde{F}\left(\sqrt{T} \left| |W_1(s)| \max_{s \in [0,1]} |a_{21}(s)| - |W_2(s)| \min_{s \in [0,1]} |a_{22}(s)| \right| \right)$   
if  $|W_2(s)| > |W_1(s)| \frac{\max_{s \in [0,1]} |a_{21}(s)|}{\min_{s \in [0,1]} |a_{22}(s)|}$ .

Thus, for some generic constant  $C$ ,

$$\begin{aligned} & \left| \int_0^1 F_1\left(\sqrt{T}\tilde{B}_1(s)\right) F_2\left(\sqrt{T}\tilde{B}_2(s)\right) \sigma_{12}(s) ds \right| \\ & \leq C \int_0^1 \tilde{G}_1\left(\sqrt{T}|W_1(s)|\right) \tilde{G}_2\left(\sqrt{T}|W_1(s)|, \sqrt{T}|W_2(s)|\right) ds. \end{aligned}$$

We know from Kallianpur and Robbins (1953) that the r.h.s. of the inequality is of order  $O_p(T^{-1} \log T)$  if  $(W_1, W_2)'$  is a nondegenerate Brownian motion and  $\tilde{G}_1(x) \tilde{G}_2(x, y)$  is a Borel function which is bounded and integrable in  $\mathbb{R}^2$ ; see also Kasahara and Kotani (1979). We only have to check that  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{G}_1(x) \tilde{G}_2(x, y) dx dy < \infty$ . Due to the continuity of  $\tilde{G}_1(x)$  and  $\tilde{G}_2(x, y)$ , we have that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{G}_1(x) \tilde{G}_2(x, y) dx dy = \int_{-\infty}^{\infty} \tilde{G}_1(x) \left( \int_{-\infty}^{\infty} \tilde{G}_2(x, y) dy \right) dx;$$

let  $I(x) = \int_{-\infty}^{\infty} \tilde{G}_2(x, y) dy$  and recall that  $\tilde{G}_1(x)$  and  $\tilde{G}_2(x, y)$  are nonnegative. Obviously,  $I(x)$  is not uniformly bounded; but it is easily shown that  $I(x) \leq C|x|$  implying that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{G}_1(x) \tilde{G}_2(x, y) dx dy \leq C \int_{-\infty}^{\infty} |x| \tilde{G}_1(x) dx.$$

Recall now that  $\tilde{G}_1(x) = \tilde{F}(x \min_{s \in [0,1]} |a_{11}(s)|)$  with  $\min_{s \in [0,1]} |a_{11}(s)|$  being positive since  $A(s)$  is positive definite for any  $s$ , so the r.h.s. is bounded as required for using the result of Kallianpur and Robbins (1953) given that  $\int_{-\infty}^{\infty} |x| \tilde{F}(x) dx < \infty$ .

Given the distributional equivalence of  $\tilde{\mathbf{B}}(s)$  and  $\mathbf{B}(s)$ , it follows that

$$\int_0^1 F_1\left(\sqrt{T}B_1(s)\right) F_2\left(\sqrt{T}B_2(s)\right) \sigma_{12}(s) ds = O_p\left(\frac{\log T}{T}\right)$$

as required for the result.

For  $b$ ), consistency of the test is obvious given Proposition 1, item  $b$ ).

**Proof of Proposition 4** The result is established by noting that Lemma A holds for these instruments as well, but in terms of recursively demeaned (detrended) time-transformed Brownian motions, which are Gaussian processes with continuous paths and the arguments of Wang and Phillips (2009, Proposition 2.1) apply. The results analog to Propositions 1, 2, and 3 follow.

## References

- Bai, J. and S. Ng (2004), A PANIC Attack on Unit Roots and Cointegration. *Econometrica* 72, 1127-1177.
- Banerjee, A. and J.L. Carrion-i-Silvestre (2006), Cointegration in Panel Data with Breaks and Cross-Section Dependence. *ECB Working Paper* No. 591.
- Banerjee, A., J.J. Dolado and R. Mestre (1998), Error-Correction Mechanism Tests for Cointegration in a Single-Equation Framework. *Journal of Time Series Analysis* 19, 267-283.
- Banerjee, A., M. Marcellino and C. Osbat (2004), Some Cautions on the Use of Panel Methods for Integrated Series of Macroeconomic Data. *Econometrics Journal* 7, 322-340.
- Born, B. and M. Demetrescu (2011), Recursive Adjustment for General Deterministic Components and Improved Cointegration Rank Tests. *Mimeo*.
- Boswijk, H.P. (2010), Nuisance Parameter Free Inference on Cointegration Parameters in the Presence of a Variance Shift. *Economics Letters* 107, 190-193.
- Cavaliere, G., A. Rahbek and A.M.R. Taylor (2010), Testing for Co-Integration in Vector Autoregressions with Non-Stationary Volatility. *Journal of Econometrics* 158, 7-24.
- Cavaliere, G. and A.M.R. Taylor (2009), Heteroskedastic Time Series with a Unit Root. *Econometric Theory* 25, 1228-1276.
- Chang, Y. (2002), Nonlinear IV Unit Root Tests in Panels with Cross-Sectional Dependency. *Journal of Econometrics* 110, 261-292.
- Chang, Y. (2004), Bootstrap Unit Root Tests in Panels with Cross-Sectional Dependency. *Journal of Econometrics* 120, 263-293.
- Chang, Y. and C. M. Nguyen (2011), Residual Based Tests for Cointegration in Dependent Panels. *Journal of Econometrics* forthcoming.
- Chang, Y., J.Y. Park and P.C.B. Phillips (2001), Nonlinear Econometric Models with Cointegrated and Deterministically Trending Regressors. *Econometrics Journal* 4, 1-36.
- Chang, Y. and W. Song (2009), Testing for Unit Roots in Small Panels with Short-run and Long-run Cross-Sectional Dependencies. *Review of Economic Studies* 76, 903-935.
- Demetrescu, M. (2009), Panel Unit Root Testing with Nonlinear Instruments for Infinite-Order Autoregressive Processes. *Journal of Time Series Econometrics* 1 (2), article 3.
- Demetrescu, M. (2010), On the Dickey-Fuller Test with White Standard Errors. *Statistical Papers* 51, 11-25.
- Demetrescu, M. (2011), Integrable IV Estimators in Autoregressions with Time-Varying Volatility. *Mimeo*.



- Demetrescu, M. and C. Hanck (2011a), Robust IV Estimators for Autoregressive Processes under Nonstationary Error Volatility. *Mimeo*.
- Demetrescu, M. and C. Hanck (2011b), Unit Root Testing in Heteroskedastic Panels using the Cauchy Estimator. *Journal of Business & Economic Statistics* forthcoming.
- Demetrescu, M., U. Hassler and A.I. Tarcolea (2006), Combining Significance of Correlated Statistics with Application to Panel Data. *Oxford Bulletin of Economics and Statistics* 68, 647-663.
- Engle, R.F. and C.W.J. Granger (1987), Co-Integration and Error Correction: Representation, estimation and testing. *Econometrica* 55, 251-276.
- Fachin, S. (2005), Long-Run Trends in Internal Migrations in Italy: A study in panel cointegration with dependent units. *Journal of Applied Econometrics* 2, 401-428.
- Gengenbach, C., F. Palm and J.-P. Urbain (2006), Cointegration Testing in Panels with Common Factors. *Oxford Bulletin of Economics and Statistics* 68, 683-719.
- Hanck, C. (2009), Nonstationary-Volatility Robust Panel Unit Root Tests and the Great Moderation. *METEOR Research Memoranda* RM/09/009.
- Hanck, C. (2011), An Intersection Test for Panel Unit Roots. *Econometric Reviews* forthcoming.
- Hartung, J. (1999), A Note on Combining Dependent Tests of Significance. *Biometrical Journal* 41, 849-855.
- Jeganathan, P. (2008) Limit Theorems for Functionals of Sums that Converge to Fractional Brownian and Stable Motions. *Cowles Foundation Discussion Paper* 1649.
- Johansen, S. (1995), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, New York.
- de Jong, R.M. and C.-H. Wang (2005), Further Results on the Asymptotics for Nonlinear Transformations of Integrated Time Series. *Econometric Theory* 21, 413-430.
- Kallianpur, G. and H. Robbins (1953), Ergodic Property of the Brownian Motion Process. *Proceedings of the National Academy of Sciences of the U.S.A.* 39, 525-533.
- Kao, C. (1999), Spurious Regression and Residual-Based Tests for Cointegration in Panel Data. *Journal of Econometrics* 90, 1-44.
- Kasahara, Y. and S. Kotani (1979), On the Limit Process for a Class of Additive Functionals of Recurrent Diffusion Processes. *Zeitschrift für Wahrscheinlichkeitstheorie verw. Gebiete* 49, 133-153.
- Kremers, J.J.M., N.R. Ericsson and J.J. Dolado (1992), The Power of Cointegration Tests. *Oxford Bulletin of Economics and Statistics* 54, 325-348.
- Larsson, R., J. Lyhagen and M. Löthgren (2001), Likelihood-Based Cointegration Tests in Heterogeneous Panels. *The Econometrics Journal* 4, 109-142.

- Lütkepohl, H. (1996), *Handbook of Matrices*. John Wiley and Sons, Chichester.
- Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin.
- Maddala, G.S. and S. Wu (1999), A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test. *Oxford Bulletin of Economics and Statistics*, Special Issue, 631-652.
- Miller, J.I. (2010), A Nonlinear IV Likelihood-Based Rank Test for Multivariate Time Series and Long Panels. *Journal of Time Series Econometrics* 2 (1), article 5.
- O'Connell, P.G.J. (1998), The Overvaluation of Purchasing Power Parity. *Journal of International Economics* 44, 1-19.
- Palm, F.C., S. Smeekes and J.-P. Urbain (2011) Cross-Sectional Dependence Robust Block Bootstrap Panel Unit Root Tests. *Journal of Econometrics* 163, 85-104.
- Park, J.Y. and P.C.B. Phillips (1999), Asymptotics for Nonlinear Transformations of Integrated Time Series. *Econometric Theory* 15, 269-298.
- Park, J.Y. and P.C.B. Phillips (2001), Nonlinear Regressions with Integrated Time Series. *Econometrica* 69, 117-162.
- 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. (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.
- Phillips, P.C.B., J.Y. Park and Y. Chang (2004) Nonlinear Instrumental Variable Estimation of an Autoregression. *Journal of Econometrics* 118, 219-246.
- Saikkonen, P. and H. Lütkepohl (1996), Infinite-Order Cointegrated Vector Autoregressive Processes: Estimation and Inference. *Econometric Theory* 12, 814-844.
- Shin, D.W. and S. Kang (2006), An Instrumental Variable Approach for Panel Unit Root Tests under Cross-Sectional Dependence. *Journal of Econometrics* 134, 215-234.
- Wang, Q. and Phillips, P.C.B. (2009), Asymptotic Theory for Local Time Density Estimation and Nonparametric Cointegrating Regression. *Econometric Theory* 25, 710-738.
- Wang, Q. and Phillips, P.C.B. (2011), Asymptotic Theory for Zero Energy Functionals with Nonparametric Regression Applications. *Econometric Theory* 27, 235-259.
- Wang, S., P. Wang, J. Yang and Z. Li (2010), A Generalized Nonlinear IV Unit Root Test for Panel Data with Cross-Sectional Dependence. *Journal of Econometrics* 157, 101-109.
- Westerlund, J. and R. Larsson (2009), A Note on the Pooling of Individual PANIC Unit Root Tests. *Econometric Theory* 25, 1851-1868.

Westerlund, J. (2007), Testing for Error Correction in Panel Data. *Oxford Bulletin of Economics and Statistics* 69, 709–748.

Westerlund, J. and D.L. Edgerton (2007), A Panel Bootstrap Cointegration Test. *Economics Letters* 97, 185–190.