Nonstationary-Volatility Robust Non-Cointegration Tests with an Application to Cross-Dependent Panels*

Matei Demetrescu
Christian-Albrechts-University of Kiel†

Christoph Hanck
University of Duisburg-Essen‡

December 11, 2015

Abstract

Panel cointegration tests rely to a great extent on the good performance of individual-unit test statistics. Be they from the first or from the second generation, panel tests are therefore distorted too, should the assumptions for single-unit cointegration tests be violated. With cointegration tests in the VAR framework often being oversized under time-varying error variance, is thus becomes possible, if not likely, to confuse error variance nonstationarity with cointegration in heteroskedastic panels. The paper takes an instrumental variable approach to establish individual-unit and panel test statistics for no cointegration that are robust to variance nonstationarity. We first examine a single-unit of a panel whose features also include cross-unit dependence and heterogeneity. The sign of a fitted departure from long-run equilibrium is used as instrument when estimating individual error-correction models. The resulting IV-based test is shown to follow a chi-square limiting null distribution irrespective of the variance pattern of the data generating process. In spite of this, the test proposed here has competitive local power against sequences of local alternatives in $1/T$-neighbourhoods of the null. The standard limiting null distributions enable us to construct robust panel tests of no cointegration by combining p-values from individual units under cross-dependence. The proposed tests perform well in small samples in the presence of cross-sectional correlation and cross-unit cointegration as well as time-varying covariance matrices of the shocks. An application to the term structure theory of interest rates illustrates the dramatic differences between robust and nonrobust tests.

Keywords: Instrument variable, Cauchy instrument, Nonstationary volatility, Cross-unit dependence, Robustness

JEL classification: C12 (Hypothesis Testing), C32 (Time-Series Models), C33 (Models with Panel Data)

*The authors would like to thank Yoosoon Chang for very useful comments and remarks. The authors gratefully acknowledge the support of the German Research Foundation (DFG) through the projects DE 1617/4-1 and HA 6766/2-1.

†Institute for Statistics and Econometrics, Christian-Albrechts-University of Kiel, Olshausenstr. 40-60, D-24118 Kiel, Germany, email: `mdeme@stat-econ.uni-kiel.de`.

‡Faculty of Economics and Business Administration, University of Duisburg-Essen, Universitätsstraße 12, D-45117 Essen, Germany, tel. +49 201 1832263, `christoph.hanck@vwl.uni-due.de`.

1 Introduction

Analyzing long-run relations between stochastically trending variables is not conceivable without cointegration tools. Having nonstandard limiting distributions expressed as functionals of Wiener processes, cointegration tests in the VAR framework are typically not robust to nonstationary volatility: when global homoskedasticity conditions are not fulfilled, the limiting distributions are given in terms of functionals of Gaussian continuous-time processes different from Brownian motions. This makes the limiting distributions depend of nuisance parameters, the so-called (co)variance profiles. Consequently, time-varying covariance matrices of the shocks translate into size distortions even asymptotically; see Cavaliere et al. (2010, 2012).

But data are too often heteroskedastic in the long run, with the variance varying quite persistently in time. The Great Moderation (Stock and Watson, 2002) is the most cited example of such persistent variance changes. Further evidence is provided by Sensier and van Dijk (2004). The finding that the Great Moderation may actually be over and that the volatility of many macroeconomic variables increases back to previous levels (Clark, 2009) only strengthens the point, highlighting that there may be several variance breaks in macro data spanning several decades. Time heteroskedasticity is therefore more than just a theoretical problem, and researchers should resort to robust inferential tools to prevent spurious findings.

Along these lines, Cavaliere et al. (2010) suggest the wild bootstrap (already applied with unit root tests by Cavaliere and Taylor, 2008a) as a way of obtaining correct critical values. See also Swensen (2006), Cavaliere et al. (2012) and Cavaliere et al. (2013) for bootstrapping cointegration rank tests. Other correction methods discussed in the case of unit root testing (Cavaliere and Taylor, 2008b; Beare, 2008; Boswijk, 2005) are less practical in the multivariate setup, requiring nonparametric estimation of either the variance function or the variance profile.

Turning our attention to panel data, note that the corrections mentioned above were developed for individual units. In panels, size distortions from individual units tend to accumulate rather than average out; see Demetrescu and Hanck (2012) or Westerlund (2014). This makes robustification even more important. But (macro) panels are particularly difficult to deal with, among other reasons because of cross-unit dependence in form of cross-correlation or cross-cointegration. One is in general interested in "marginal" cointegration of panel units involving only series from the unit under consideration, irrespective of "joint" long-run dynamics of series from different units. Hence, cross-unit cointegration is seen as nuisance, just like cross-correlation. This poses high demands on any panel bootstrap scheme. Along these lines, Smeekes and Urbain (2014a) discuss modified wild bootstrap procedures in the panel unit root case that can cope in theory with cross-unit dependence including cross-cointegration. They find however that the behavior of the modified wild bootstrap methods in samples of usual size is not reliable in the presence of cross-unit cointegration. In a more parametric approach, the PANIC methodology of Bai and Ng (2004) is able to account for cross-unit correlation and cross-cointegration; see e.g. Gengenbach et al. (2006), Westerlund (2008), Gengenbach et al. (2008) or Bai and Carrion-i Silvestre (2013). Yet a typical requirement in these papers is that the shocks to the idiosyncratic and common components, while allowed to be heteroskedastic in the cross-sectional dimension, must be weakly stationary in the time dimension; see also Westerlund (2015). It is not clear how time heteroskedasticity can be accounted for in the PANIC approach.

The present paper therefore proposes a robust testing procedure based on instrumental variable [IV] estimation in an error-correction model. We work on two levels. First, we examine individual units, where the sign of an estimated equilibrium deviation is used as instruments for estimating the equilibrium adjustment coefficients; see So and Shin (1999) for the use of these so-called Cauchy instruments in the univariate case. The nonlinear IV method has been successfully used in the unit root case: Demetrescu and Hanck (201x, 2012) proved the t-statistic based on the Cauchy estimator to be robust to deterministic heteroskedasticity in the time dimension of arbitrary form and exhibits good local power properties.

On the panel level, we resort to combining significance levels following Demetrescu et al. (2006) and Hanck (2013). The p-values are easily obtained from the individual-unit limiting distributions which turn out to be χ^2, making it straightforward to construct panel test statistics. Moreover, the combination approach allows for unbalanced panels and cross-cointegration.

Concretely, our contributions are as follows. After specifying the panel cointegration model with time-varying volatility and cross-unit dependence in Section 2, we discuss in Section 3 how to use the idea of the Cauchy estimator in such a way that the resulting limiting distributions of the non-cointegration test statistics are χ^2 irrespective of the variance profile. This is not trivial since naively instrumenting the lagged levels by their signs would not lead to the desired invariance to time-varying volatility. We also establish the asymptotic properties of the proposed testing procedure under a sequence of local alternatives in $1/T$-neighbourhoods of the null of no cointegration. In spite of the standard null asymptotics, the proposed test has power in the same neighbourhoods of the null as the Johansen (1988) trace test, and is typically more powerful than the trace test for alternatives close to the null. Section 4 examines the finite-sample properties of the individual-unit and panel tests and confirms the analytical predictions given by the (local) asymptotics. We apply our test to a panel of OECD long and short term interest rates (which appear to be heteroskedastic in the time dimension) to test for the validity of the term structure theory of interest rates in Section 5, and illustrate the need to robust inference and the straightforward construction of panel test statistics.

As regards notation, $\Theta_p(\cdot)$, O_p and o_p have their usual meaning, $\|\cdot\|$ stands for the Euclidean vector norm and the corresponding induced matrix norm, $\|\cdot\|_r = \sqrt[r]{\mathrm{E}\left(\|\cdot\|^r\right)}$ is the L_r norm of a random variable or vector, $\perp$ denotes orthogonal complement w.r.t. $\mathbb{R}^K$, $\Rightarrow$ denotes weak convergence, and $\stackrel{p}{\to}$ and $\stackrel{d}{\to}$ are taken to mean convergence in probability and in distribution.

2 The heteroskedastic panel cointegration model

Denote by $\boldsymbol{y}_{it}$, $i = 1, \ldots, N$ and $t = 1, \ldots, T$, the typical observation of the panel data set to be analyzed. The vector $\boldsymbol{y}_{it}$ is K-dimensional, and the individual series y_{itk} are integrated and may be cointegrated, either within each unit or across the panel. The long-run relations of interest are only those within the units, involving y_{itk} for fixed i. Cross-cointegration, i.e. long-run relations involving variables with different unit index i (as they would appear when similar variables from different units—say, inflation rates in neighbouring countries—have a common trend) is regarded as nuisance. See Wagner and Hlouskova, 2010 for a formal definition of cross-unit cointegration.

Stack together the series in the panel, $\boldsymbol{y}_t = (\boldsymbol{y}'_{1t}, \ldots, \boldsymbol{y}'_{Nt})'$, and let

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{w}_t.$$

We assume the stochastic component for each unit to be accurately described by

Assumption 1 *Let*

$$\Delta \boldsymbol{w}_{it} = \Pi_i \boldsymbol{w}_{it-1} + \sum_{j=1}^{\infty} \Gamma_{ij} \Delta \boldsymbol{w}_{it-j} + \boldsymbol{\varepsilon}_{it}$$

where $\|\Gamma_{ij}\| \leq Ge^{-j}\ \forall i,j$ *for some positive constant* G *and bounded initial conditions.*

This assumption not only allows for ARMA-type dynamics, which we shall approximate by long autoregressions, but also helps deal with cross-unit dynamics; see below.

Since the order of the model is infinite, we will approximate the model by a finite-order (possibly cointegrated) VAR with p_T lagged differences. See e.g. Saikkonen (1992) in the (co)integrated VAR case. The model approximation order p_T should grow to infinity together with the sample size at an appropriate rate to allow for white noise innovations in the limit.

Assumption 2 *Let* $p_T = CT^{\delta}$ *for some* $\delta \in (0, 1/4)$.

We allow $\boldsymbol{\varepsilon}_{it}$ to be conditionally as well as unconditionally heteroskedastic martingale differences. As argued by Cavaliere et al. (2010), long-run relations may be characterized by mean-reversion of departures from equilibrium values rather than their weak or strict stationarity. So cointegration in unit i is still characterized by the matrices Π_i even under time-varying variance. After having conveniently separated the nonstationary dynamics from the nonstationary variance, the discussion proceeds as usually: the process $\boldsymbol{w}_{it}$ is integrated or cointegrated with cointegration rank $r_i = \operatorname{rank} \Pi_i$ under suitable conditions on Γ_{ij} and Π_i, see Assumption 3. Should $\boldsymbol{w}_{it}$ be cointegrated, one has $0 < r_i < K$ and the known factorization holds, $\Pi_i = \boldsymbol{\alpha}_i \boldsymbol{\beta}'_i$, with $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ two $K \times r_i$ matrices of adjustment speed coefficients and of parameters of the long-run relations.

Assumption 3 *Let* Π_i *have rank* $0 \leq r_i < K$ *such that the roots of the characteristic polynomial associated to* $\boldsymbol{w}_t$ *are either* 1 *or have absolute values larger than* 1. *When* $\Pi_i \neq \mathbf{0}$, *let* $\det\left(\boldsymbol{\alpha}'_{i\perp}\left(I - \sum_{j=1}^{\infty}\Gamma_{ij}\right)\boldsymbol{\beta}_{i\perp}\right) \neq 0$.

The following assumption specifies what types of heteroskedasticity we allow for.

Assumption 4 *Let the* N *series of innovations* $\boldsymbol{\varepsilon}_{it}$ *be given as* $\boldsymbol{\varepsilon}_{it} = H_{it}\boldsymbol{\epsilon}_{it}$, *where* $H_{it} = H_i\left(t/T\right)$ *are square matrices of piecewise Lipschitz functions with full rank at all* $s \in [0,1]$. *Further, let* $\boldsymbol{\epsilon}_{it}$

possess the martingale difference property each, with covariance matrix I_K, uniformly bounded moments of order m for some $m > 4$, and uniformly bounded conditional (on $\{\boldsymbol{\epsilon}_{it-1}, \boldsymbol{\epsilon}_{it-2}, \ldots\}$) density functions with real support.

The requirement that the conditional density be uniformly bounded serves for example to prevent a nonzero probability of equilibrium deviations being zero and hence of not having a well-defined sign. This is not an uncommon requirement in the literature on nonlinear transformations of (near-)integrated processes; see e.g. Wang and Phillips (2009).

Assumption 4 implies that $\text{Cov}(\boldsymbol{\varepsilon}_{it}) = H_{it}H'_{it} > 0 \; \forall \, t$. We assume the $H_i(\cdot)$ to be deterministic, but they may be stochastic if they are independent of $\boldsymbol{\epsilon}_{it}$ at all i, t. Also, we have for each unit the weak convergence

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} \boldsymbol{\varepsilon}_{it} \Rightarrow \boldsymbol{M}_i(s) \equiv \int_0^s H_i(r)\, \mathrm{d}\boldsymbol{W}_i(r) \qquad i = 1, \ldots, N \tag{1}$$

as $T \to \infty$, with $\boldsymbol{W}_i$ a K-dimensional vector of independent Wiener processes (see e.g. Cavaliere et al., 2010). The diffusions $\boldsymbol{M}_i$ are only Brownian motions when H_i is constant for all s. The quadratic variation process of $\boldsymbol{M}_i$ gives the covariance profile of the series, and, at $s = 1$, it has the interpretation of an "average" covariance matrix, $\Omega_i = \int_0^1 H_i(r)\, H'_i(r)\, \mathrm{d}r$.

Note that Assumption 4 allows for cross-sectional error dependence of unspecified form. To discuss additional forms of cross-dependence, consider the panel DGP

$$\Delta \boldsymbol{w}_t = \Pi \boldsymbol{w}_{t-1} + \sum_{j=1}^{p} A_j \Delta \boldsymbol{w}_{t-j} + \boldsymbol{\nu}_t \tag{2}$$

for $t = p+1, \ldots, T$. Groen and Kleibergen (2003) discuss the case where Π has block diagonal structure (no cross-cointegration), and Banerjee et al. (2004) analyze the effects of departures from block diagonality on tests assuming such diagonality. They show that imposing block-diagonality of Π when cross-cointegration is actually present can have quite severe consequences such as spurious panel cointegration findings due to heavily oversized tests. See also Hecq et al. (2002) for a discussion of the dynamic properties of such panel models.

To distinguish between unit-specific and cross-unit cointegration we may reformulate the model in (2) for each unit i as

$$\Delta \boldsymbol{w}_{it} = \Pi_i \boldsymbol{w}_{it-1} + \tilde{\Pi}_i \boldsymbol{w}_{t-1} + \sum_{j=1}^{p} A^*_{ij} \Delta \boldsymbol{w}_{t-j} + \boldsymbol{\nu}_{it}$$

where the term $\Pi_i \boldsymbol{w}_{it}$ with Π_i $K \times K$ is unit specific (and our interest), A^*_{ij} contains the corresponding lines of A_j, and $\tilde{\Pi}_i \boldsymbol{w}_t$ with $\tilde{\Pi}_i$ $K \times (NK)$ captures error-correction to long-run relations between variables from different units; cf. Wagner and Hlouskova (2010).

From the perspective of unit i, cross-unit dynamics – i.e. lagged differences from other units and in particular the terms $\tilde{\Pi}_i \boldsymbol{w}_{t-1}$ containing error-correction to long-run relations involving cross-cointegration – play the role of lagged stationary covariates. This motivates us to set up individual-unit error correction models like in Assumption 1, since the covariates are, from the single-unit perspective, nothing else than a form of additive stationary noise. Their effect can be captured (again, in the individual-unit perspective) by *additional* unit-specific lagged differences to ensure white noise errors for the respective unit. When doing so, autoregressive short-run dynamics of infinite order arise; under the relevant null $\Pi_i = 0$, $\Delta \boldsymbol{w}_{it}$ has VAR dynamics overlapping with the error-correction to cross-cointegration, $\tilde{\Pi}_i \boldsymbol{w}_{t-1}$, which makes $\Delta \boldsymbol{w}_{it}$, taken alone, a VARMA process, just like individual series of a VAR process are ARMA in general. The infinite-order VAR assumption in Assumption 1 matches the effect of cross-unit dynamics. Thus, $\boldsymbol{w}_{it}$ is a (co)integrated VAR(∞) process for a given i and we may focus on *unit-specific* cointegration as specified in Assumption 1.

Dealing with cross-unit dynamics in this way implies the residuals $\boldsymbol{\varepsilon}_{it}$ to differ from the shocks $\boldsymbol{\nu}_{it}$. Concretely, $\boldsymbol{\varepsilon}_{it}$ is serially uncorrelated, but only for unit i – while $\boldsymbol{\varepsilon}_t$ stacking $\boldsymbol{\varepsilon}_{it}$ for all i is not necessarily serially uncorrelated; see e.g. Smeekes and Urbain (2014b). So cross-unit dependence of the individual test statistics arises, namely beyond contemporaneously correlated shocks; the panel discussion will have to take all these features into account. Moreover, the cross-sectional correlation structure is left largely unspecified. An implicit condition is however that there exists a common underlying probability space. Hence, individual-unit statistics are dependent in an unspecified manner; this is one of the reasons why we resort to p-value combinations in the panel case. This is different from Demetrescu et al. (2014), who employ so-called integrable instruments and thus achieve asymptotic independence of individual-unit tests, however at the cost of power losses (see the simulations section and also Demetrescu and Hanck, 2013).

We now proceed to building our panel test for the null hypothesis $\Pi_i = 0$.

3 The robust panel no-cointegration test

3.1 Obtaining a single-unit test statistic

A quick way of testing the null $r_i = 0$ for a given unit is to estimate Π_i and test whether it is significantly different from zero. This is nothing else than a Wald test, asymptotically equivalent to Johansen's LR test for $r_i = 0$ against $r_i > 0$. Tempting as it may be, one cannot use Cauchy IV estimation and test directly, because the resulting distribution would not be chi-square distributed, but rather non-standard (as also confirmed by some simulations available upon request from the authors). See Miller (2010) for a case where standard distributions do arise, at the cost of having to use instruments that are integrable nonlinear transformations of the lagged levels.

A two-step Engle-Granger procedure would alternatively be implementable along the lines of Chang and Nguyen (2012), i.e. by recursive computation of the first-stage residuals. One reason not to pursue this route in more detail here are the negative effects of common factor-type restrictions on the power of such tests; see e.g. Pesavento (2004).

So our preferred solution is to test for error-correction. To arrive at a test statistic, first assume $\boldsymbol{w}_{it}$ to be directly observable; we deal with the deterministic component below. Assuming that the cointegration rank is 1 under the alternative (see Banerjee et al., 1998 for a discussion of this assumption in the error-correction testing framework), denote by $e_{it-1} = \boldsymbol{\beta}_i' \boldsymbol{w}_{it-1}$ the (hypothesized) equilibrium error at time $t-1$ and consider the K seemingly unrelated regressions

$$\Delta \boldsymbol{w}_{it} = \boldsymbol{\alpha}_i e_{it-1} + \sum_{j=1}^{\infty} \Gamma_{ij} \Delta \boldsymbol{w}_{it-j} + \boldsymbol{\varepsilon}_{it}. \tag{3}$$

(As argued in Section 2, we would work with a truncated version of the model containing p_T lagged differences, and let $p_T \to \infty$ together with $T \to \infty$ at a suitable rate for the asymptotic analysis.)

Under the null of no cointegration, one has $\boldsymbol{\alpha}_i = \mathbf{0}$ while $\boldsymbol{\alpha}_i \neq \mathbf{0}$ under the alternative. So we could try to estimate $\boldsymbol{\alpha}_i$ in (3) by instrumenting e_{it-1} and conduct a corresponding test. This approach would also offer information about which of the K series y_{itk} error-correct, which is a major advantage over the Engle-Granger procedure. But, of course, e_{it-1} is not observed.

Assume for the moment that the cointegrating vector $\boldsymbol{\beta}_i$ is pre-specified. Then $e_{it-1} = \boldsymbol{\beta}_i' \boldsymbol{w}_{it-1}$ is readily obtained, and nonlinear IV estimation can be conducted Cauchy-style for each of the K equations. (Under the null, e_{it-1} would be integrated, so simple OLS leads to nonstandard distributions.) Moreover, the joint distribution of the resulting K t-statistics would be multivariate normal under the null (along the lines of Corollary 1 below) and a test for no cointegration is immediately built. The consequences of misspecifying the cointegrating vector are severe, however (see Zivot, 2000), and the cases where the economic theory implies specific values for $\boldsymbol{\beta}_i$ (say, purchasing power parity relations) are rare.

We hence resort to an estimate of the equilibrium error. The estimator however needs to satisfy certain requirements. In particular, its sign needs to be an adapted sequence for a md central limit theorem to apply in this setup (see So and Shin, 1999, for the univariate case) leading to χ^2-distributions asymptotically. The solution we propose is to use a recursive form of estimating the equilibrium error e_{it-1}: our estimate, denoted by $\tilde{e}_{it-1}$, is taken to be the tth residual from a regression of some element of $\boldsymbol{w}_{it}$ on the other $K-1$ using only the first t observations in the sample. This resembles the proposal of Chang and Nguyen (2012), yet has important differences: Chang and Nguyen test (recursively fitted) residuals from a static regression for a unit root, while we test for no error correction (in response to a recursively fitted equilibrium error) in the full system.[1] Moreover, the procedure of Chang and Nguyen (2012) is not robust to time-varying variance; see Demetrescu et al. (2014).

Should $\boldsymbol{\mu}$ be unknown and $\boldsymbol{w}_{it}$ not be observed directly, the regression must account for non-zero means unless the long-run equilibrium is known to have zero expectation. Equivalently to using a regression with intercept for each subsample $j = 1, \ldots, t$, we may use the recursively demeaned

[1]Another neat consequence of testing error correction in each equation in response to the filtered equilibrium deviation $\tilde{e}_{it-1}$ is that each equation of this system is exactly identified. Hence equation-by-equation IV is equivalent to a system estimation taking cross-equation correlation into account (Hayashi, 2000, Sec. 4.4).

the elements of $\boldsymbol{y}_{it}$ directly,

$$\tilde{\boldsymbol{y}}_{it} = \boldsymbol{y}_{it} - \frac{1}{t}\sum_{j=1}^{t} \boldsymbol{y}_{ij}, \quad t = 1, \ldots, T;$$

our arguments hold for both variants. Note that $\tilde{\boldsymbol{y}}_{it} = \tilde{\boldsymbol{w}}_{it}$ and $\Delta \boldsymbol{y}_{it} = \Delta \boldsymbol{w}_{it}$. Should the direction of the deterministic component be orthogonal to the cointegration space, one may actually skip the recursive demeaning step. Knowledge of such orthogonality is not always certain, however, and the power losses due to recursive demeaning turn out to be negligible (see Subsection 3.3), so we recommend to always demean.

Partition $\boldsymbol{y}_{it}$ w.l.o.g. as $\boldsymbol{y}_{it} = (y_{it1}, \boldsymbol{y}'_{it2})'$.[2] The recursively filtered equilibrium deviation is given by

$$\tilde{e}_{it-1} = \tilde{y}_{it-11} - \tilde{\boldsymbol{y}}'_{it-12}\left(\sum_{j=1}^{t-1} \tilde{\boldsymbol{y}}_{ij2}\tilde{\boldsymbol{y}}'_{ij2}\right)^{-1}\left(\sum_{j=1}^{t-1} \tilde{\boldsymbol{y}}_{ij2}\tilde{y}_{ij1}\right)$$

with $\tilde{e}_{it-1} = 0$ for $t = 2, \ldots, K+1$. The use of $\tilde{e}_{it-1}$ for IV estimation and testing will be shown to lead to the desired robustness properties.

Let hats denote "Cauchy" IV estimates of each of the K equations, which use the sign of $\tilde{e}_{it-1}$ as an IV for $\tilde{e}_{it-1}$, and with $\Delta \boldsymbol{y}_{it-j}$ instrumenting $\Delta \boldsymbol{y}_{it-j}$:

$$\Delta y_{itk} = \hat{\alpha}_{ik}\tilde{e}_{it-1} + \sum_{j=1}^{p_T} \hat{\boldsymbol{\gamma}}'_{ijk}\Delta \boldsymbol{y}_{it-j} + \hat{\varepsilon}_{itk}^{(p_T)}, \tag{4}$$

where the notation $\varepsilon_{itk}^{(p_T)}$ indicates that the errors in this regression are only approximately serially uncorrelated due to the truncation of the AR(∞) dynamics. More precisely,

$$\varepsilon_{itk}^{(p_T)} = \varepsilon_{itk} + \sum_{j=p_T+1}^{\infty} \boldsymbol{\gamma}_{ijk}\Delta \boldsymbol{w}_{it-j}$$

with $\boldsymbol{\gamma}_{ijk}$ the corresponding rows of Γ_{ij}. Let $\boldsymbol{x}_{it-1} = \left(\Delta \boldsymbol{y}'_{it-1}, \ldots, \Delta \boldsymbol{y}'_{it-p_T}\right)'$. The IV estimator of α_{ik} is then given by

$$\hat{\alpha}_{ik} = \frac{\sum_{t=p_T+2}^{T} \operatorname{sgn}\left(\tilde{e}_{it-1}\right)\Delta y_{itk} - \sum \operatorname{sgn}\left(\tilde{e}_{it-1}\right)\boldsymbol{x}'_{it-1}\left(\sum \boldsymbol{x}_{it-1}\boldsymbol{x}'_{it-1}\right)^{-1}\sum \boldsymbol{x}_{it-1}\Delta y_{itk}}{\sum \left|\tilde{e}_{it-1}\right| - \sum \operatorname{sgn}\left(\tilde{e}_{it-1}\right)\boldsymbol{x}'_{it-1}\left(\sum \boldsymbol{x}_{it-1}\boldsymbol{x}'_{it-1}\right)^{-1}\sum \boldsymbol{x}_{it-1}\tilde{e}_{it-1}}$$

The usual IV standard error of $\hat{\alpha}_{ik}$ is given by

$$s.e.\left(\hat{\alpha}_{ik}\right) = \hat{\sigma}_{ik}\sqrt{\frac{\sum \operatorname{sgn}^2\left(\tilde{e}_{it-1}\right) - \sum \operatorname{sgn}\left(\tilde{e}_{it-1}\right)\boldsymbol{x}'_{it-1}\left(\sum \boldsymbol{x}_{it-1}\boldsymbol{x}'_{it-1}\right)^{-1}\sum \boldsymbol{x}_{it-1}\operatorname{sgn}\left(\tilde{e}_{it-1}\right)}{\left(\sum \left|\tilde{e}_{it-1}\right| - \sum \operatorname{sgn}\left(\tilde{e}_{it-1}\right)\boldsymbol{x}'_{it-1}\left(\sum \boldsymbol{x}_{it-1}\boldsymbol{x}'_{it-1}\right)^{-1}\sum \boldsymbol{x}_{it-1}\tilde{e}_{it-1}\right)^2}}$$

[2]Under the null, this is inconsequential asymptotically, but, under the alternative, y_{it1} should be part of the cointegration relation. Typically, empirical applications however pre-selected variables using economic arguments anyhow.

with $\hat{\sigma}_{ik}^2 = \frac{1}{T}\sum\left(\hat{\varepsilon}_{itk}^{(p_T)}\right)^2$. Denote by $\boldsymbol{t}_i = (t_{\alpha_{i1}}, \ldots, t_{\alpha_{iK}})$ the vector of K equation-specific t-statistics for the null $\alpha_{ik} = 0$, $t_{\alpha_{ik}} = \frac{\hat{\alpha}_{ik}}{s.e.(\hat{\alpha}_{ik})}$. We prove in the following subsection that $t_{\alpha_{ik}}$ follows a standard normal distribution under the null of no error correction. Moreover, a test for no cointegration can be built on the joint test for $\boldsymbol{\alpha}_i = 0$. Unlike the case of integrable instruments (Demetrescu et al., 2014), the individual t-statistics are not asymptotically independent. But we prove $\boldsymbol{t}_i$ to follow a multivariate normal distribution with covariance matrix

$$\Xi_i = (\operatorname{diag}\Omega_i)^{-0.5}\,\Omega_i\,(\operatorname{diag}\Omega_i)^{-0.5}$$

under the null, where Ω_i can be consistently estimated by the sample covariance matrix of the residuals $\hat{\boldsymbol{\varepsilon}}_{it}^{(p_T)}$; see the proof of Proposition 1 (in particular Equation (8)) in the Appendix. The results hold irrespective of the functions H_i, and no heteroskedasticity-consistent standard errors are required, unlike for the case of integrable instruments (Demetrescu et al., 2014).

One may also base a test on the estimates $\hat{\boldsymbol{\alpha}}_i$ directly. Similarly, $\hat{\boldsymbol{\alpha}}_i'\widehat{\operatorname{Cov}(\hat{\boldsymbol{\alpha}}_i)}^{-1}\hat{\boldsymbol{\alpha}}_i$ has an asymptotic $\chi^2(K)$ null distribution; in fact it is shown in Corollary 2 below that $\hat{\boldsymbol{\alpha}}_i'\widehat{\operatorname{Cov}(\hat{\boldsymbol{\alpha}}_i)}^{-1}\hat{\boldsymbol{\alpha}}_i$ and $\boldsymbol{t}_i'\hat{\Xi}_i^{-1}\boldsymbol{t}_i$ are actually asymptotically equivalent. Concretely, the usual (in the seemingly unrelated regressions framework) estimate of the covariance matrix of $\hat{\boldsymbol{\alpha}}_i$ is calculated as follows. Define $\hat{\boldsymbol{\phi}}_i = (\hat{\alpha}_{i1}, \hat{\boldsymbol{\gamma}}_{i11}', \ldots, \hat{\boldsymbol{\gamma}}_{ip_T1}', \ldots, \hat{\alpha}_{iK}, \hat{\boldsymbol{\gamma}}_{i1K}', \ldots, \hat{\boldsymbol{\gamma}}_{ip_TK}')'$. Then

$$\widehat{\operatorname{Cov}(\hat{\boldsymbol{\phi}}_i)} = \hat{\Omega}_i \otimes \left(P_i' R_i^{-1} P_i\right)^{-1}$$

with

$$\hat{\Omega}_i = \frac{1}{T}\sum \hat{\boldsymbol{\varepsilon}}_{it}^{(p_T)}\left(\hat{\boldsymbol{\varepsilon}}_{it}^{(p_T)}\right)'$$

and

$$\begin{aligned} P_i &= \begin{pmatrix} \sum|\tilde{e}_{it-1}| & \sum\operatorname{sgn}(\tilde{e}_{it-1})\,\boldsymbol{x}_{it-1}' \\ \sum\boldsymbol{x}_{it-1}\tilde{e}_{it-1} & \sum\boldsymbol{x}_{it-1}\boldsymbol{x}_{it-1}' \end{pmatrix} \\ R_i &= \begin{pmatrix} \sum\operatorname{sgn}^2(\tilde{e}_{it-1}) & \sum\operatorname{sgn}(\tilde{e}_{it-1})\,\boldsymbol{x}_{it-1}' \\ \sum\boldsymbol{x}_{it-1}\operatorname{sgn}(\tilde{e}_{it-1}) & \sum\boldsymbol{x}_{it-1}\boldsymbol{x}_{it-1}' \end{pmatrix}. \end{aligned}$$

Define the $K \times K(1+Kp)$ selector matrix $\mathcal{S}$ with ones in positions $k, 1+(k-1)(pK+1)$, $k = 1, \ldots, K$, and zeros else. Finally,

$$\widehat{\operatorname{Cov}(\hat{\boldsymbol{\alpha}}_i)} = \mathcal{S}\left(\hat{\Omega}_i \otimes \left(P_i' R_i^{-1} P_i\right)^{-1}\right)\mathcal{S}'.$$

Remark: While we allow for a non-zero mean of $\boldsymbol{y}_{it}$, a linear trend is more difficult to account for. The reason is that adjusting the differences for a non-zero mean induces distortions in the asymptotic null distribution, making it dependent on the variance profile. See Demetrescu and Hanck (201x) for the unit root case (to allow the use of the Cauchy IV method in the case with a trend, they adopt a wild bootstrap correction following Cavaliere and Taylor, 2008a, but the correction is computationally costly).

3.2 Asymptotic behavior

We begin by recalling a useful result. The levels behave under a local alternative as follows

Lemma 1 *Under Assumptions 1, 3 and 4, we have for local alternatives of the form* $\Pi_{i,T} = \frac{1}{T}\boldsymbol{\alpha}_i\boldsymbol{\beta}_i'$ *that*

$$\frac{1}{\sqrt{T}}\boldsymbol{w}_{i[sT]} \Rightarrow \boldsymbol{J}_i(s)$$

as $T \to \infty$*, where*

$$\mathrm{d}\boldsymbol{J}_i(s) = \boldsymbol{\alpha}_i\boldsymbol{\beta}_i'\boldsymbol{J}_i(s)\,\mathrm{d}s + \Gamma_i^{-1}\mathrm{d}\boldsymbol{M}_i(s)$$

with $\Gamma_i = I - \sum_{j=1}^{\infty}\Gamma_{ij}$.

Proof: *follows from weak convergence in* (1).

We are now in a position to state and prove the main result.

Proposition 1 *Under Assumptions 1, 2, 3 and 4 and local alternatives of the form* $\Pi_{i,T} = \frac{1}{T}\boldsymbol{\alpha}_i\boldsymbol{\beta}_i'$*, it holds as* $T \to \infty$ *that*

$$\begin{aligned}\boldsymbol{t}_i \ &\overset{d}{\to}\ \int_0^1 \operatorname{sgn}\left(\tilde{E}_i(s)\right)(\operatorname{diag}\Omega_i)^{-0.5}\,\mathrm{d}\boldsymbol{M}_i(s) - \ (\operatorname{diag}\Omega_i)^{-0.5}\,\boldsymbol{\alpha}_i\boldsymbol{\beta}_i'\int_0^1 \operatorname{sgn}\left(\tilde{E}_i(s)\right)\boldsymbol{J}_i(s)\,\mathrm{d}s\\ &\equiv\ \mathcal{T}_{0i} - \mathcal{T}_{1i}\end{aligned}$$

where $\tilde{E}_i(0) = 0$ *a.s. and, for* $s > 0$,

$$\tilde{E}_i(s) = \tilde{J}_{i1}(s) - \tilde{\boldsymbol{J}}_{i2}'(s)\left(\int_0^s \tilde{\boldsymbol{J}}_{i2}(r)\,\tilde{\boldsymbol{J}}_{i2}'(r)\,\mathrm{d}r\right)^{-1}\left(\int_0^s \tilde{\boldsymbol{J}}_{i2}(r)\,\tilde{J}_{i1}(r)\,\mathrm{d}r\right)$$

with corresponding partition $\tilde{\boldsymbol{J}}_i = \left(\tilde{J}_{i1}, \tilde{\boldsymbol{J}}_{i2}'\right)'$*, and* $\tilde{\boldsymbol{J}}_i(s) = \boldsymbol{J}_i(s) - \frac{1}{s}\int_0^s \boldsymbol{J}_i(r)\,\mathrm{d}r$ *with* $\tilde{\boldsymbol{J}}_i(0) = \boldsymbol{0}$ *a.s. denoting the recursively demeaned version of* $\boldsymbol{J}_i(s)$.

Proof: *see the Appendix.*

Corollary 1 *Under the null hypothesis of no cointegration,*

$$\boldsymbol{t}_i \overset{d}{\to} \mathcal{N}(0, \Xi_i).$$

In particular,

$$t_{\alpha_{ik}} \overset{d}{\to} \mathcal{N}(0, 1).$$

Proof: *see the Appendix.*

Under the null, the joint asymptotic distribution of $\boldsymbol{t}_i$ is conveniently multivariate normal, where Ξ_i is the "average" correlation matrix of the $\boldsymbol{\varepsilon}_i$s. The matrix Ξ_i is can be consistently estimated via the correlation matrix of the residuals $\hat{\boldsymbol{\varepsilon}}_{it}^{(p_T)}$; see the proof of Proposition 1.

Corollary 2 *It holds that*

$$\hat{\boldsymbol{\alpha}}_i' \widehat{\text{Cov}\,(\hat{\boldsymbol{\alpha}}_i)}^{-1} \hat{\boldsymbol{\alpha}}_i - \boldsymbol{t}_i' \hat{\Xi}_i^{-1} \boldsymbol{t}_i \xrightarrow{p} 0.$$

Proof: *see the Appendix.*

Under a fixed alternative, $\tilde{e}_{it-1}$ consistently estimates the equilibrium error (although not for fixed t) so $\hat{\boldsymbol{\alpha}}_i$ is easily shown to be consistent for a non-zero vector. Then, the t-statistics diverge and the test itself is consistent. Moreover, Proposition 1 implies that the test has power against local alternatives in $1/T$-neighbourhoods of the null.

3.3 Numerical evidence

Proposition 1 implies a noncentral χ^2-type of distribution for $\hat{Q}_i := \boldsymbol{t}_i' \widehat{\Xi}_i^{-1} \boldsymbol{t}_i$,

$$\hat{Q}_i := \boldsymbol{t}_i' \widehat{\Xi}_i^{-1} \boldsymbol{t}_i \quad \xrightarrow{d} \quad \mathcal{T}_{0i}' \Xi_i^{-1} \mathcal{T}_{0i} - 2\mathcal{T}_{0i}' \Xi_i^{-1} \mathcal{T}_{1i} + \mathcal{T}_{1i}' \Xi_i^{-1} \mathcal{T}_{1i}.$$

This being a nonstandard distribution, we resort to a numerical evaluation of the local power of the individual-unit test. As a benchmark, we consider the local power of the trace test for cointegration, λ_{tr}, which has been studied by Johansen (1995) and Saikkonen and Lütkepohl (1999). More specifically, let us look at the special case of interest here, viz. testing the non-cointegration null against the local alternative of a single cointegrating relation. For the numerical results below, we further focus on the case of no serial correlation. With $\boldsymbol{N}_i(s) = \Omega_i^{-1/2} \boldsymbol{J}_i(s)$,

$$\lambda_{\text{tr}} \quad \xrightarrow{d} \quad \text{tr}\left[\left(\int_0^1 \boldsymbol{N}_i^\star \mathrm{d}\boldsymbol{N}_i'\right)' \left(\int_0^1 \boldsymbol{N}_i^\star \boldsymbol{N}_i^{\star\prime} \mathrm{d}s\right)^{-1} \left(\int_0^1 \boldsymbol{N}_i^\star \mathrm{d}\boldsymbol{N}_i'\right)\right]$$

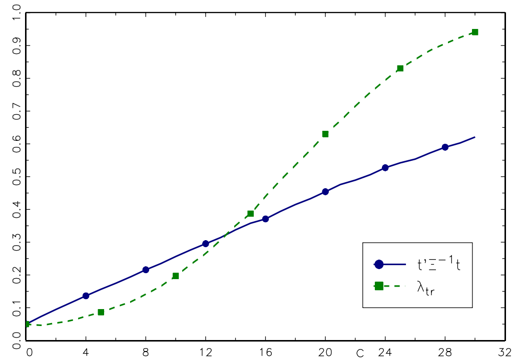
with

$$\boldsymbol{N}_i^\star(s) = \begin{cases} \boldsymbol{N}_i(s) & \text{zero-mean case} \\ [\boldsymbol{N}_i(s)' : 1]' & \text{non-zero mean.} \end{cases}$$

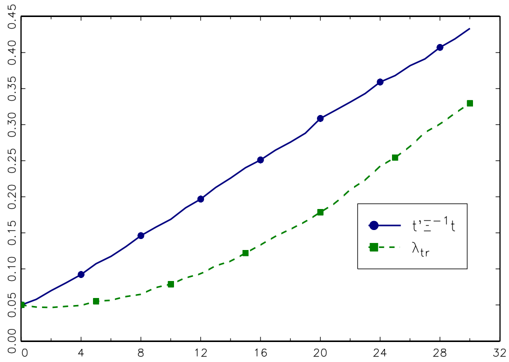
Johansen (1995) shows that the local power of λ_{tr} only depends on the parameters of the process through $f_i = \boldsymbol{\beta}_i' \boldsymbol{\alpha}_i$ and $g_i^2 = \boldsymbol{\alpha}_i' \Omega_i^{-1} \boldsymbol{\alpha}_i \boldsymbol{\beta}_i' \Omega_i \boldsymbol{\beta}_i - f_i^2$.

Figures 1 and 2 report results for local alternatives of the form $\Pi_i = \frac{1}{T} c \boldsymbol{\alpha}_i \boldsymbol{\beta}_i'$ where $\boldsymbol{\alpha}_i' = -(1 \quad 0)$, $\boldsymbol{\beta}_i' = (1 \quad 0)$ and $c = 0, 1, \ldots, 30$. Further, $\Omega_i = \boldsymbol{I}_2$. Hence, $f_i = -c$ and $g = 0$. For $H_t \neq I$, the local power curves of λ_{tr} are size-adjusted to a nominal size of 5%. This is of interest because the curves will then give the local power of the wild bootstrap version of the trace test developed by Cavaliere et al. (2010, 2012); see Cavaliere and Taylor (2008a, p. 54) for the corresponding justification in the unit root case.

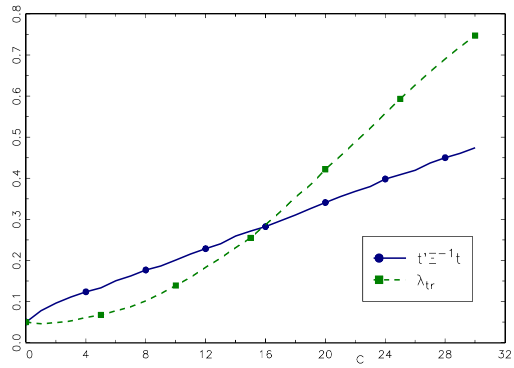
The results show that, for several empirically plausible variance profiles including homoskedasticity, neither test has uniformly better power in c. The local power of $\hat{Q}_i$ increases roughly linearly for small to moderate c, while that of λ_{tr} exhibits an S-type shape. Moreover, $\hat{Q}_i$ has



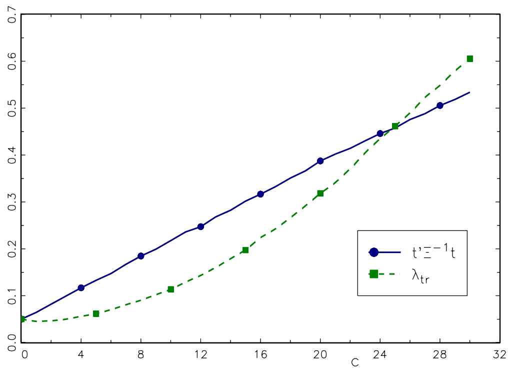
$H_i(s) = I$.



Late Upward Break, $H_i(s) = \text{diag}(1\mathbb{I}(s < 0.9) + 5\mathbb{I}(s \geqslant 0.9))$.



Negatively Trending Variance, $H_i(s) = \text{diag}(1 - s)$.



Trending Variance, $H_i(s) = \text{diag}(s)$.

Zero-Mean case, $\boldsymbol{\beta}_i' = (1 \quad 0)$, $\boldsymbol{\alpha}_i' = -c(1 \quad 0)$, $c = 0, 1, 2, \ldots, 30$ ($f = -c$, $g = 0$).

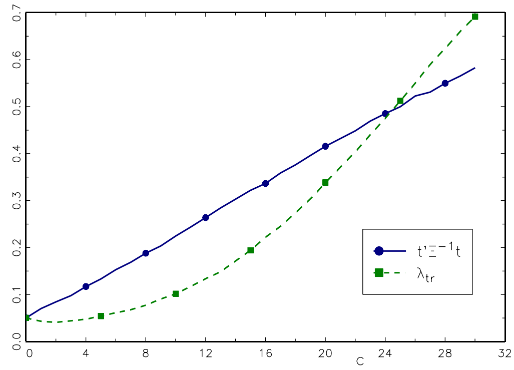
Figure 1: (Size-adjusted) Local power vs. c, zero mean

higher local power than λ_{tr} for smaller c.[3] This ranking reverses for larger c for most, but not all, variance profiles. We consider these results to be intuitive in the sense that a sign IV entails little loss of information for alternatives close to the null, but more heavily discounts the signal for more distant alternatives. Interestingly, while λ_{tr} exhibits the familiar pattern of lower power when accounting for deterministics, this does not appear to be the case for $\hat{Q}_i$.[4] Thus, $\hat{Q}_i$ performs relatively less well in the case without adjustment for deterministic terms. This behavior is reminiscent of that of the GLS-based unit root test of Elliott et al. (1996) which is known to follow the zero-mean Dickey-Fuller distribution even when accounting for an intercept term; see Leybourne et al. (2005) for a similar finding for unit root testing.

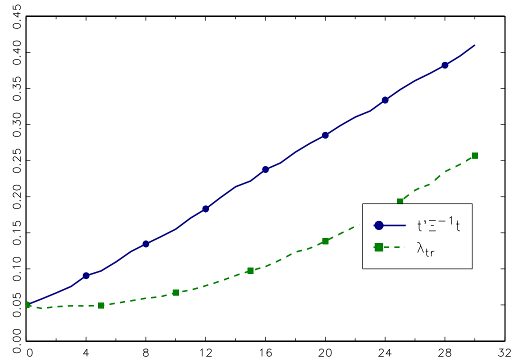
Figures 3 and 4 report analogous figures for the case in which no variable is weakly exogenous. Specifically, $\boldsymbol{\alpha}_i' = -c\,(1 \quad 0.5)$, so that $g = 0, 0.5, 1, \ldots, 15$. All in all, the results are qualitatively similar. Of course, both tests are now somewhat more powerful, as there is error correction to be detected for both variables. There does however not seem to a be a systematic pattern suggesting that λ_{tr} would do relatively better than $\hat{Q}_i$ when both variables error-correct. Under homoskedasticity, λ_{tr} does have higher local power $\hat{Q}_i$ already for smaller c now, but for instance under negatively trending variances, λ_{tr} even seems to be biased and does not achieve the same

[3] Such behavior has already been documented by So and Shin (1999) in the unit root testing case with iid shocks; see also Demetrescu and Hanck (201x) for the heteroskedastic extension.

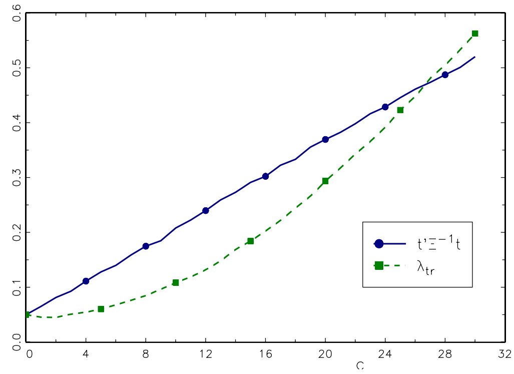
[4] The version of $\hat{Q}_i$ not accounting for deterministics simply skips the recursive demeaning step.



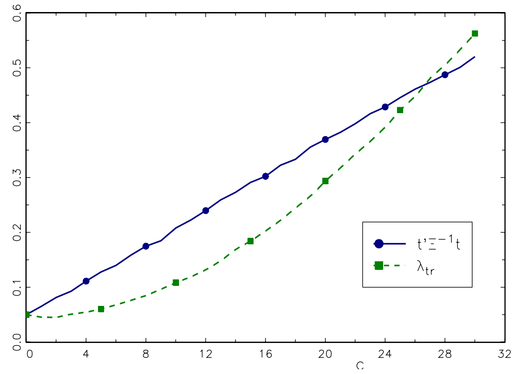
$H_i(s) = I$.



Late Upward Break, $H_i(s) = \text{diag}(1\mathbb{I}(s < 0.9) + 5\mathbb{I}(s \geqslant 0.9))$.



Negatively Trending Variance, $H_i(s) = \text{diag}(1 - s)$.



Trending Variance, $H_i(s) = \text{diag}(s)$.

Mean case, $\boldsymbol{\beta}_i' = (1 \quad 0)$, $\boldsymbol{\alpha}_i' = -c(1 \quad 0)$, $c = 0, 1, 2, \ldots, 30$ ($f = -c$, $g = 0$).

Figure 2: (Size-adjusted) Local power vs. c, constant mean

local power as $\hat{Q}_i$ for the range of c considered here.

3.4 Combining p-values

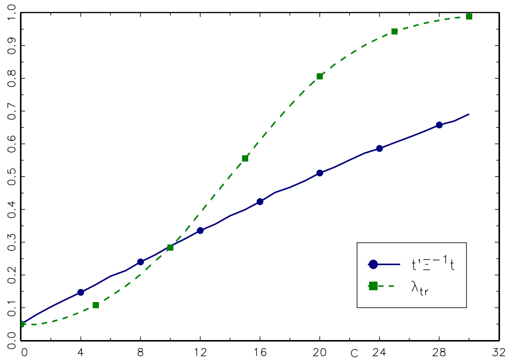
When moving from the single-unit level to the panel level, note that the approach of Shin and Kang (2006), which only addresses cross-unit correlation in dealing with the Cauchy estimator, is not likely to be robust to generic forms of cross-dependence.[5] Cross-cointegration is the reason why we do not pool $\boldsymbol{\alpha}$ anyway. So it may be more promising to resort to procedures that are known to be less restricted, at least heuristically. Denoting by p_i the p-values for $\hat{Q}_i$—readily available due to standard asymptotics—and letting $\mathcal{C}_i = \Phi^{-1}(p_i)$, we resort to

1. Hartung's (1999) correction: let $\hat{\xi}^* = \max(-1/(N-1), \hat{\xi})$, with $\hat{\xi} = 1 - 1/(N-1)\sum_{i=1}^N \big(\mathcal{C}_i - \sum_{i=1}^N \mathcal{C}_i/N\big)^2$ and

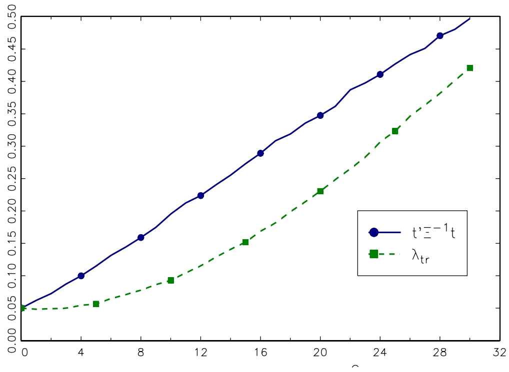
$$t_{\hat{\xi}^*,\kappa} = \frac{\sum_{i=1}^N \mathcal{C}_i}{\sqrt{N + (N^2 - N)\left(\hat{\xi}^* + 0.2\sqrt{\frac{2}{N+1}}(1 - \hat{\xi}^*)\right)}}$$

   We take $t_{\hat{\xi}^*,\kappa} \xrightarrow{d}_{H_0} \mathcal{N}(0,1)$.

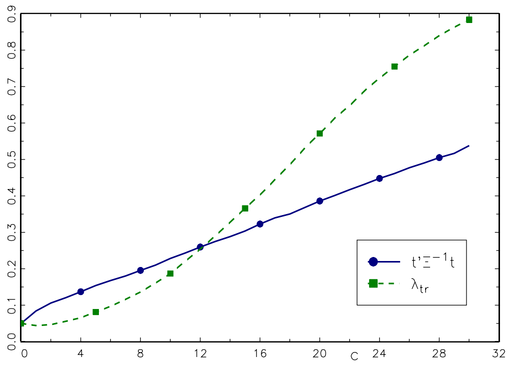
[5] For the homoskedastic panel unit root case, Chang and Song (2009) suggest an alternative IV-based approach to deal with cross-unit cointegration.



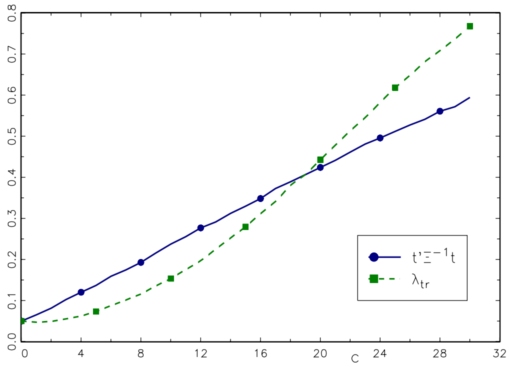
$H_i(s) = I$.



Late Upward Break, $H_i(s) = \mathrm{diag}(1\mathbb{I}(s < 0.9) + 5\mathbb{I}(s \geqslant 0.9))$.



Negatively Trending Variance, $H_i(s) = \mathrm{diag}(1 - s)$.



Trending Variance, $H_i(s) = \mathrm{diag}(s)$.

Zero-Mean case, $\boldsymbol{\beta}_i' = (1 \quad 0)$, $\boldsymbol{\alpha}_i' = -c(1 \quad 0.5)$, $c = 0, 1, 2, \ldots, 30$ ($f = -c$, $g = 0, 0.5, 1, \ldots, 15$).

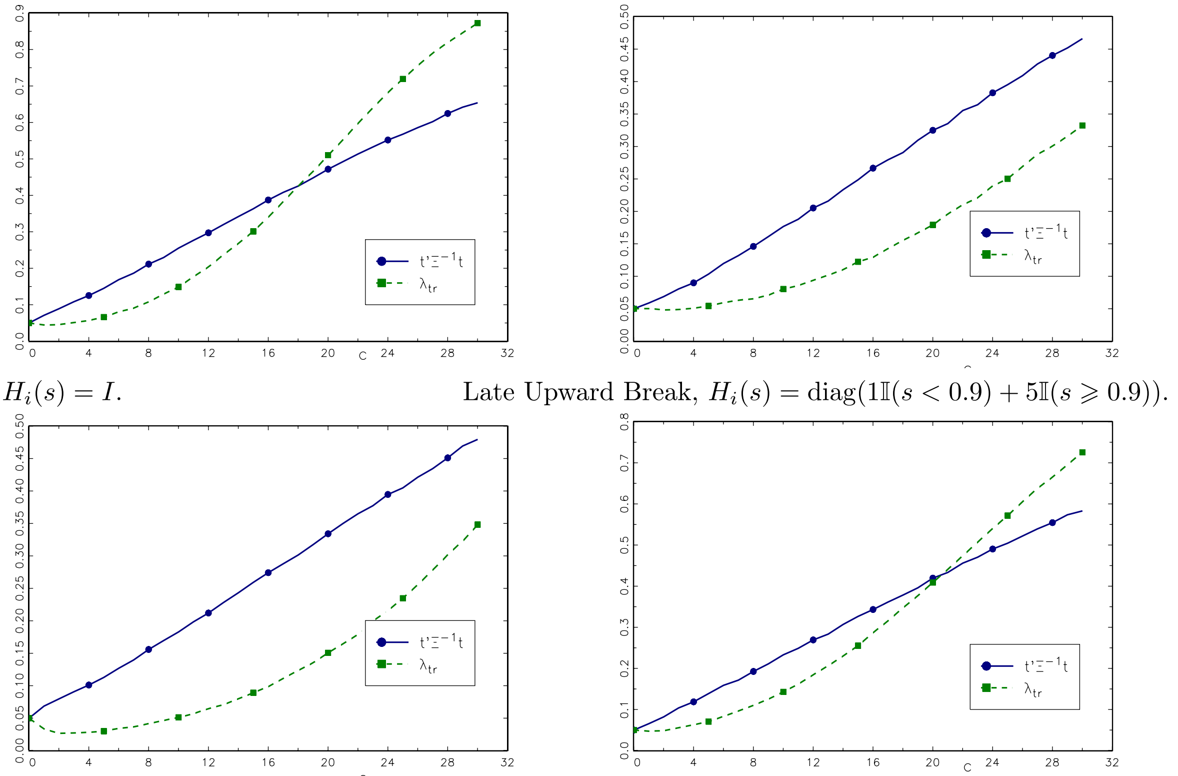
Figure 3: (Size-adjusted) Local power vs. c, zero mean, no weak exogeneity

2. Simes (1986): reject the panel null at level α if

$$p_{(j)} \leqslant j \cdot \alpha / N \quad \text{for some } \ j \in \{1, \ldots, N\}.$$

The advantages of combining p-values are manifold in our setup. For instance, it is not required to balance an unbalanced panel data set, which usually means chopping off the individual series to the shortest length and thus losing sample information. Heterogeneity across the panel is naturally accounted for, and the cross-sectional dependence need not be explicitly modelled, which, as shown in Section 2, may be quite intricate.

It is worth emphasizing that these panel statistics are heuristic in the sense that verifying whether the types of dependence structures typically encountered in nonstationary panels yield level-α tests based on these statistics is not straightforward. Some analytical results for Hartung's and Simes' tests are available in Demetrescu et al. (2006) and Sarkar (1998), respectively. It is, however, not obvious to extend these results to, say, cross-unit cointegrated panels. Following promising results in, e.g., Hanck (2013), we study the effectiveness of these panel approaches in Section 4 below.



$H_i(s) = I$. Late Upward Break, $H_i(s) = \mathrm{diag}(1\mathbb{I}(s < 0.9) + 5\mathbb{I}(s \geqslant 0.9))$.

Negatively Trending Variance, $H_i(s) = \mathrm{diag}(1 - s)$. Trending Variance, $H_i(s) = \mathrm{diag}(s)$.

Mean case, $\boldsymbol{\beta}_i' = (1 \quad 0)$, $\boldsymbol{\alpha}_i' = -c(1 \quad 0.5)$, $c = 0, 1, 2, \ldots, 30$ ($f = -c$, $g = 0, 0.5, 1, \ldots, 15$).

Figure 4: (Size-adjusted) Local power vs. c, constant mean, no weak exogeneity

4 Finite-sample properties

The present section provides some simulation results for the procedures introduced above, comparing them to other pertinent testing procedures for cointegration and panel cointegration. In so doing, we elaborate on the findings provided in Demetrescu et al. (2014). We initially study time series cointegration tests to then move on to the panel case.

4.1 Time-series evidence

Concretely, we compare $\hat{Q}_i$ to the robust integrable IV-based test $Q_{\boldsymbol{\alpha}}$ (we use the version with contemporaneous differences) from Demetrescu et al. (2014), to the trace test of Johansen (1988) (λ_{tr}), the residual-based test of Engle and Granger (1987) (EG), the NIV test $\tilde{\tau}_i$ of Chang and Nguyen (2012) (see their eq. (16)) and the wild bootstrap test of Cavaliere et al. (2010, 2012, 2014, CRT).[6] The heteroskedastic DGP is developed in Cavaliere et al. (2010, 2014):

$$\Delta \boldsymbol{w}_{it} = \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{w}_{i,t-1} + \boldsymbol{\varepsilon}_{it}, \quad t = 1, \ldots, T,$$

[6] For $Q_{\boldsymbol{\alpha}}$ and $\tilde{\tau}_i$, we use Chang's nonlinear IV $F(x) = x \exp(-|x|)$ with $x = (4/\widehat{\sigma}_{\Delta y})y_{t-1}$ as in Demetrescu et al. (2014).

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

δ	T	EG	$\tilde{\tau}_i$	λ_{tr}	CRT	$\hat{Q}_i$	$Q_{\boldsymbol{\alpha}}$
		break fraction $\tau = 1/5$					
.33	100	.236	.075	.478	.060	.041	.038
	200	.233	.071	.477	.051	.045	.045
	500	.230	.078	.471	.055	.051	.041
	1000	.230	.078	.474	.048	.046	.046
1	100	.045	.050	.057	.048	.051	.046
	200	.051	.056	.055	.046	.052	.049
	500	.047	.051	.058	.052	.051	.053
	1000	.047	.052	.051	.050	.053	.046
5	100	.030	.033	.034	.049	.062	.042
	200	.031	.016	.031	.050	.045	.041
	500	.038	.012	.032	.055	.052	.047
	1000	.036	.009	.026	.046	.053	.049

$M_K = \boldsymbol{I}$. δ: break direction. EG: Engle/Granger (Ecma 1987), $\tilde{\tau}_i$: Chang and Nguyen (JoE 2012), λ_{tr}: Johansen (1995), CRT: Cavaliere et al. (JoE 2010), $Q_{\boldsymbol{\alpha}}$: Demetrescu et al. (JTSA 2014). $\hat{Q}_i := \boldsymbol{t}_i'\widehat{\Xi}_i^{-1}\boldsymbol{t}_i$.

with $\boldsymbol{\varepsilon}_{it} = H_t\boldsymbol{\epsilon}_{it}$ and $\boldsymbol{\epsilon}_{it} \sim iid\mathcal{N}(0, M_K)$, where $M_K = \xi\iota\iota' + (1-\xi)I_K$ for some ξ. H_t is a time-varying diagonal matrix with $H_0 = I_K$ and $H_t = \delta I_K$ for $t = \lfloor\tau T\rfloor, \lfloor\tau T\rfloor + 1, \ldots, T$ and $T = \{100, 200, 500, 1000\}$. We consider $K = 2$, 'early' to 'late' breaks $\tau \in \{1/5, 1/3, 2/3, 4/5\}$ and negative to positive shifts $\delta \in \{1/5, 1/3, 1, 3, 5\}$ and present a subset of these results below. When $\delta = 1$, we simply study the homoskedastic case. We study size by setting $\boldsymbol{\alpha} = \boldsymbol{0}$. As our focus is on the effect of heteroskedasticity, we omit further complications such as short-run dynamics and set $p = 0$. As usual, these complications would certainly not improve the small-sample performance of any of the tests studied here, but we see no strong reason that they should affect what is our main concern here, the relative ranking in terms of size and power. Moreover, we consider cross-cointegration below, which implies the need to include lagged differences anyhow. All results are for the 5% level and are based on 5,000 replications.

Table 1 reveals that, as in Demetrescu et al. (2014), EG, $\tilde{\tau}_i$ and λ_{tr} either do not control size in the presence of heteroskedastic disturbances or may be extremely undersized, with obvious detrimental consequences for their power ($\tilde{\tau}_i$). This is, in line with the inappropriateness of the standard limiting null distributions under heteroskedasticity, not a small-sample phenomenon, as evidenced by the results for $T = 1000$. This is not surprising as they were not designed with robustness to such departures from the standard case in mind. The robust tests $\hat{Q}_i$, $Q_{\boldsymbol{\alpha}}$ and CRT, in turn, are all level α.

Tables 2 and 3 provide some power results. To complement the local analysis from Section 3.3, we consider fixed alternatives here. We set here $\boldsymbol{\beta} = (1 \quad -1)'$. In Table 2, both variables error-correct, while in Table 3 only the first does, but more strongly so. We omit results for the non-robust tests when $\delta \neq 1$ in view of their size distortions.[7] When $\delta = 1$, EG performs best when only one unit error-corrects, in line with the findings of e.g. Pesavento (2004). The

[7] Clearly, one could report size-adjusted power, with, however, questionable value for empirical practice. At any rate, the wild bootstrap test CRT can, for instance, be seen as a size-corrected version of λ_{tr}.

Table 2: Power of Time Series Cointegration Tests under Nonstationary Volatility

δ	T	EG	$\tilde{\tau}_i$	λ_{tr}	CRT	$\hat{Q}_i$	$Q_{\boldsymbol{\alpha}}$
		break fraction $\tau = 1/5$					
.33	100				.069	.072	.034
	200				.117	.136	.043
	500				.455	.532	.051
	1000				.953	.944	.058
1	100	.088	.069	.129	.124	.115	.051
	200	.190	.072	.357	.335	.259	.049
	500	.830	.079	.974	.966	.821	.051
	1000	1.00	.101	1.00	1.00	.999	.060
5	100				.152	.115	.045
	200				.405	.232	.048
	500				.979	.768	.061
	1000				1.00	.994	.069

$\boldsymbol{\alpha} = -.05 \cdot \mathbf{1}$, $\boldsymbol{\beta} = (1 \quad -1)'$, $M_K = \boldsymbol{I}$. δ: break direction. EG: Engle/Granger (Ecma 1987), $\tilde{\tau}_i$: Chang and Nguyen (JoE 2012), λ_{tr}: Johansen (1995), CRT: Cavaliere et al. (JoE 2010), $Q_{\boldsymbol{\alpha}}$: Demetrescu et al. (JTSA 2014). $\hat{Q}_i := \boldsymbol{t}_i'\widehat{\Xi}_i^{-1}\boldsymbol{t}_i$.

tests that, with possibly slight abuse of terminology, could be called system-based, viz. λ_{tr}, CRT and $\hat{Q}_i$ perform relatively better when both variables contribute to the error-correction mechanism. Clearly, we notice that for alternatives as close to the null as those considered here, the tests based on integrable instruments, $\tilde{\tau}_i$ and $Q_{\boldsymbol{\alpha}}$ are strongly outperformed. Again, this is not implausible as the integrability condition on the IVs requires to discount large values to zero. This reduces the variability of the signal and hence costs power. In fact, unreported calculations indicate that $Q_{\boldsymbol{\alpha}}$ does not even local have power in neighborhoods of the same size as the other tests considered here; see also Demetrescu and Hanck (2013).

In line with Proposition 1 and the results in CRT, the power of the tests is affected by the variance pattern governing the $\boldsymbol{\varepsilon}_{it}$. In particular, in line with the above local power results, no clear ranking emerges from the comparison between CRT and $\hat{Q}_i$. The latter is more powerful for for instance early downward breaks, and the former for late downward ones. Bearing in mind the results from Figures 1-4, we however caution that the present results deal with one fixed distance from the null. They should hence not be read as definitive recommendations to choose a certain test for particular variance patters (which one might estimate from the data), but rather as supportive finite-sample evidence for the above local power results.

4.2 Panel evidence

Let us now augment the time series DGP to

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

We generate cross-sectional dependence through a factor structure: $\varepsilon_{k,it} = \lambda_i \nu_t + \tilde{\varepsilon}_{k,it}$, $k = 1, \ldots, K$, where $\tilde{\boldsymbol{\varepsilon}}_{it} = H_t \boldsymbol{\epsilon}_{it}$ and $\boldsymbol{\epsilon}_{it} \sim iid\mathcal{N}(0, M_K)$ with variances of 1 and covariance 0.25.

Table 3: Power of Time Series Cointegration Tests under Nonstationary Volatility

δ	T	EG	$\tilde{\tau}_i$	λ_{tr}	CRT	$\hat{Q}_i$	$Q_{\boldsymbol{\alpha}}$
		break fraction $\tau = 1/3$					
.33	100				.043	.093	.047
	200				.109	.215	.059
	500				.677	.737	.089
	1000				.999	.996	.159
1	100	.216	.095	.117	.105	.156	.056
	200	.659	.108	.356	.336	.383	.060
	500	1.00	.153	.994	.991	.948	.083
	1000	1.00	.214	1.00	1.00	1.00	.116
3	100				.135	.151	.057
	200				.391	.369	.072
	500				.988	.913	.104
	1000				1.00	1.00	.149

$\boldsymbol{\alpha} = -(0.1\ 0)'$, $\boldsymbol{\beta} = (1\ \ -1)'$, $M_K = \boldsymbol{I}$. δ: break direction. EG: Engle/Granger (Ecma 1987), $\tilde{\tau}_i$: Chang and Nguyen (JoE 2012), λ_{tr}: Johansen (1995), CRT: Cavaliere et al. (JoE 2010), $Q_{\boldsymbol{\alpha}}$: Demetrescu et al. (JTSA 2014). $\hat{Q}_i := \boldsymbol{t}_i'\widehat{\Xi}_i^{-1}\boldsymbol{t}_i$.

The loadings are distributed as $\lambda_i \sim \text{Uniform}[-1, 2]$ and $\nu_t \sim \mathcal{N}(0, 1)$. The panel dimensions considered are $T = \{100, 200, 500\}$ and $N = \{20, 50, 100, 200\}$, and $\boldsymbol{\beta}_i = (1\ \ -1)'$ unless specified otherwise.

As there is no straightforward way to construct a panel test based on the wild bootstrap tests from CRT, in what follows, we only consider the remaining robust tests from Demetrescu et al. (2014) and those suggested here. Demetrescu et al. (2014) propose both a test statistic based on the sum of the individual $Q_{\boldsymbol{\alpha}}$, denoted $\tilde{Q}$, and a pooled statistic $\tilde{Q}^{\text{P}}$. The panel statistics resulting from applying Simes' and Hartung's approach to the N single-unit statistics $\hat{Q}_i$ are denoted by S and $t_{\hat{\xi}^*,\kappa}$.

Table 4 provides some selected (but representative) size results. We observe that both variants from Demetrescu et al. (2014) as well as the Simes-based sign-IV panel cointegration test studied here exhaust and generally control nominal size. The only exceptions concern very wide ($N = 200$) panels in which the fixed-N asymptotics are relatively less informative. In such cases, small time-series size distortions may add up to produce slightly more heavily size distorted panel tests. In turn, the Hartung variant $t_{\hat{\xi}^*,\kappa}$ is fairly undersized.

Table 5 reports some power results. Unsurprisingly, the pooled statistic $\tilde{Q}^{\text{P}}$ performs better than $\tilde{Q}$ for such a homogenous DGP. It is noteworthy that, despite being undersized, $t_{\hat{\xi}^*,\kappa}$ still is more powerful than $\tilde{Q}$. Bearing in mind the single-unit results, it may appear surprising that $\tilde{Q}^{\text{P}}$ is competitive with S although the former is based on a single-unit test statistics $Q_{\boldsymbol{\alpha}}$ with substantially lower power. The explanation we put forward for this finding is that, first, the single-unit test statistics based on integrable IVs are asymptotically independent. Hence, the sum of these, $\tilde{Q}^{\text{P}}$, may be compared to the critical values of an exact (in N) distribution, as the sum of independent χ^2-r.v.s is, of course, also χ^2. Hence, $\tilde{Q}^{\text{P}}$ efficiently handles the dependence structure of the single-unit statistics. In turn, the $\hat{Q}_i$ are not independent across i so that we need

Table 4: Panel Size, $\tau = 1/5$

N	T	$\tilde{Q}$	$\tilde{Q}^{\text{P}}$	S	$t_{\hat{\xi}^*,\kappa}$	$\tilde{Q}$	$\tilde{Q}^{\text{P}}$	S	$t_{\hat{\xi}^*,\kappa}$
		$\delta = 1$				$\delta = 1/5$			
20	100	.053	.053	.056	.024	.038	.070	.037	.013
	200	.042	.054	.054	.020	.039	.051	.040	.019
	500	.046	.054	.052	.016	.039	.049	.056	.016
50	100	.054	.067	.069	.009	.068	.056	.029	.008
	200	.053	.058	.060	.007	.040	.049	.033	.010
	500	.046	.054	.046	.009	.046	.052	.043	.014
100	100	.068	.066	.057	.003	.075	.064	.028	.012
	200	.057	.060	.046	.004	.057	.057	.041	.010
	500	.045	.057	.047	.004	.041	.062	.041	.010
200	100	.096	.083	.080	.001	.118	.066	.028	.016
	200	.064	.070	.042	.002	.066	.065	.040	.018
	500	.071	.064	.045	.000	.047	.056	.045	.015

$\boldsymbol{\alpha}_i = \mathbf{0}$. δ: break direction. $\tilde{Q}$ and $\tilde{Q}^{\text{P}}$: Demetrescu et al. (JTSA 2014). S / $t_{\hat{\xi}^*,\kappa}$: Simes / Hartung based on $\hat{Q}_i$.

to resort to a panel statistic such as Simes' that is robust to such cross-sectional dependence. Ensuring against such dependence comes at a certain price in terms of power.

Second, the Simes (1986) test may be expected to be relatively most powerful in panels in which only a small fraction of units violates the single-unit null. Such is sufficient to imply a rejection of the panel null, which states that none of the units in the panel exhibits a cointegrating relation, whose negation, and hence alternative, results if at least one unit cointegrates. Simes' test then is effective because one p-value not exceeding its threshold is sufficient to reject the null. Hence, we may reject the null already if the smallest p-value is less than α/N, irrespective of how large the remaining p-values are. In a homogenous DGP as the above, each of the units' relations may not be strong enough to provide such a small p-value, although the aggregate signal suffices to reject when using a sum-type statistic such as $\tilde{Q}^{\text{P}}$ or $\tilde{Q}$.

To shed some further light on this, Table 6 studies the case in which only a single unit cointegrates with $\boldsymbol{\alpha}_1 = (-0.5 \;\; 0)'$. In line with the above reasoning, S now clearly is most powerful. Plausibly, $\tilde{Q}$ is now also more powerful than $\tilde{Q}^{\text{P}}$.

We focus in the remainder of the section on the case in which there exist cross-unit cointegration relations. More specifically, the panel DGP is as follows (see also Banerjee et al., 2005):

$$\Delta \boldsymbol{w}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{w}_{t-1} + \boldsymbol{\varepsilon}_t \tag{5}$$

with $\boldsymbol{\alpha} = -0.1 \cdot I_{2N}$ and

$$\boldsymbol{\beta}' = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & B \end{pmatrix} \tag{6}$$

Table 5: Panel Power, $\tau = 1/5$

N	T	$\tilde{Q}$	$\tilde{Q}^{\mathrm{P}}$	S	$t_{\hat{\xi}^*,\kappa}$	$\tilde{Q}$	$\tilde{Q}^{\mathrm{P}}$	S	$t_{\hat{\xi}^*,\kappa}$
		$\delta = 1$				$\delta = 1/5$			
20	100	.086	.231	.221	.408	.088	.166	.109	.138
	200	.115	.399	.748	.940	.142	.378	.595	.814
	500	.261	.790	1.00	1.00	.485	.899	1.00	1.00
50	100	.118	.496	.255	.633	.128	.343	.120	.194
	200	.193	.783	.875	.995	.246	.722	.702	.934
	500	.490	.996	1.00	1.00	.780	.996	1.00	1.00
100	100	.181	.792	.274	.828	.206	.564	.122	.255
	200	.275	.978	.928	1.00	.376	.933	.758	.972
	500	.734	1.00	1.00	1.00	.935	1.00	1.00	1.00
200	100	.291	.973	.307	.938	.351	.802	.118	.353
	200	.443	1.00	.976	1.00	.575	.986	.807	.988
	500	.928	1.00	1.00	1.00	.989	1.00	1.00	1.00

$\boldsymbol{\alpha}_i = -(0.1\ \ 0)'$, $\boldsymbol{\beta}_i = (1\ \ -1)'$. δ: break direction. $\tilde{Q}$ and $\tilde{Q}^{\mathrm{P}}$: Demetrescu et al. (JTSA 2014). S / $t_{\hat{\xi}^*,\kappa}$: Simes / Hartung based on $\hat{Q}_i$.

Table 6: Panel Power, heterogeneous case

N	T	$\tilde{Q}$	$\tilde{Q}^{\mathrm{P}}$	S	$t_{\hat{\xi}^*,\kappa}$	$\tilde{Q}$	$\tilde{Q}^{\mathrm{P}}$	S	$t_{\hat{\xi}^*,\kappa}$
		$\delta = 1$				$\delta = 1/5$			
20	100	.271	.255	.957	.439	.485	.298	.836	.309
	200	.581	.573	1.00	.794	.828	.646	.997	.743
	500	.974	.970	1.00	.999	.998	.982	1.00	.988
50	100	.199	.158	.914	.126	.335	.150	.741	.090
	200	.380	.311	1.00	.290	.673	.348	.992	.255
	500	.891	.796	1.00	.737	.989	.857	1.00	.710
100	100	.163	.112	.846	.044	.296	.119	.695	.040
	200	.282	.191	1.00	.067	.524	.209	.988	.088
	500	.718	.558	1.00	.260	.962	.628	1.00	.294
200	100	.170	.115	.793	.012	.286	.106	.633	.022
	200	.209	.135	1.00	.014	.426	.150	.986	.051
	500	.547	.296	1.00	.052	.868	.372	1.00	.108

$\boldsymbol{\alpha}_1 = (-0.5\ \ 0)'$ and zero otherwise, $\boldsymbol{\beta}_1 = (1\ \ -1)'$. $\tau = 1/5$. δ: break direction. $\tilde{Q}$ and $\tilde{Q}^{\mathrm{P}}$: Demetrescu et al. (JTSA 2014). S / $t_{\hat{\xi}^*,\kappa}$: Simes / Hartung based on $\hat{Q}_i$.

where

$$B_{(N\times N)} = \begin{pmatrix} 1 & -1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & \cdots & \cdots & & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \cdots & & 0 \end{pmatrix} \tag{7}$$

That is, we simulate a KN-dimensional vector error correction model with, as before, $K = 2$,

Table 7: Panel Size under Cross-Unit Cointegration, $\tau = 1/5$

δ	T	S	$t_{\hat{\xi}^*,\kappa}$	S	$t_{\hat{\xi}^*,\kappa}$	S	$t_{\hat{\xi}^*,\kappa}$
		$N = 10$		$N = 20$		$N = 50$	
0.2	100	.063	.055	.071	.046	.085	.070
	200	.041	.031	.042	.024	.048	.025
	500	.039	.022	.056	.026	.056	.022
	1000	.050	.029	.055	.024	.062	.024
1	100	.071	.058	.081	.059	.073	.045
	200	.058	.042	.066	.029	.064	.024
	500	.058	.036	.055	.023	.058	.009
	1000	.048	.026	.056	.018	.057	.011
5	100	.069	.047	.055	.029	.052	.021
	200	.060	.035	.055	.021	.049	.010
	500	.052	.028	.050	.016	.041	.006
	1000	.049	.025	.050	.019	.048	.005

$p_T = [4(T/100)^{1/4}]$. δ: break direction. S / $t_{\hat{\xi}^*,\kappa}$: Simes / Hartung based on $\hat{Q}_i$.

in which we are interested in testing for cointegration relations between series i and series $i + N$ for all $i = 1, \ldots, N$. (Think of the first N series as—see the empirical application in Section 5—different countries' short-run and the remaining series $N + 1, \ldots, 2N$ as long-run interest rates.) Writing out the single equations of (5) however reveals that such relations however do not exist, so that the DGP (5)-(7) allows us to study the size of our tests for the null of interest of no within-unit cointegration. This size might be affected by the fact that, as evident from the matrix B, series $N + 1$ cointegrates with series $N + 2$ with cointegration vector $(1 \quad -1)$, series $N + 2$ with series $N + 3$, etc. For example, the long-run rates of the US and the UK might cointegrate. There are $N/5$ cross-unit cointegrating relations.[8]

Tables 7-9 report results. We approximate the resulting single-unit VECM(∞)-structure (recall equation (3)) by the deterministic lag-length selection rule $p_T = [4(T/100)^{1/4}]$ (other choices led to only moderately different results). Both S and $t_{\hat{\xi}^*,\kappa}$ control size, with the latter even being closer to exhaust the nominal size than without such cross-unit cointegrating relations. In particular, Tables 8 and 9 suggest that neither the strength nor the number of cross-unit cointegrating relations seem to affect this robustness property.

5 A panel analysis of long- and short-term interest rates

This section illustrates the use of the cointegration tests investigated above by means of an application to long- and short-term interest rates. As already recognized in the seminal contribution of Engle and Granger (1987), the term structure theory of interest rates (Cox et al., 1985) predicts cointegration between these rates (provided they are individually nonstationary, see Lanne, 2000) as possible extra returns compensating the additional risks from long-term investments are

[8] We do not perform a power study of the tests for brevity as there is no reason to believe that cross-unit cointegration should affects the tests' capability to detect within-unit cointegration, so that we omit this case for brevity.

Table 8: Panel Size under Cross-Unit Cointegration, $\tau = 1/5$, $\boldsymbol{\alpha} = -0.5I_{2N}$

δ	T	S	$t_{\hat{\xi}^*,\kappa}$	S	$t_{\hat{\xi}^*,\kappa}$	S	$t_{\hat{\xi}^*,\kappa}$
		$N=10$		$N=20$		$N=50$	
0.2	100	.054	.047	.058	.050	.068	.048
	200	.032	.020	.036	.025	.045	.035
	500	.031	.018	.036	.023	.064	.022
	1000	.030	.020	.041	.021	.065	.026
1	100	.068	.052	.063	.049	.072	.043
	200	.070	.040	.065	.029	.061	.013
	500	.059	.034	.055	.022	.058	.012
	1000	.048	.030	.062	.024	.058	.010
5	100	.056	.048	.055	.041	.046	.019
	200	.048	.034	.052	.017	.053	.010
	500	.057	.032	.056	.022	.058	.006
	1000	.043	.028	.050	.018	.056	.004

$p_T = [4(T/100)^{1/4}]$. δ: break direction. S / $t_{\hat{\xi}^*,\kappa}$: Simes / Hartung based on $\hat{Q}_i$.

Table 9: Panel Size under Cross-Unit Cointegration, $\tau = 1/5$, $\boldsymbol{\alpha} = -0.5I_{2N}$, $N/2$ CUCs

δ	T	S	$t_{\hat{\xi}^*,\kappa}$	S	$t_{\hat{\xi}^*,\kappa}$	S	$t_{\hat{\xi}^*,\kappa}$
		$N=10$		$N=20$		$N=50$	
0.2	100	.062	.050	.074	.059	.080	.066
	200	.046	.039	.069	.045	.069	.046
	500	.040	.027	.083	.050	.088	.044
	1000	.036	.026	.101	.052	.105	.049
1	100	.079	.064	.070	.044	.068	.044
	200	.056	.037	.074	.038	.067	.024
	500	.053	.037	.058	.029	.054	.017
	1000	.056	.034	.060	.025	.062	.014
5	100	.052	.048	.061	.037	.046	.016
	200	.054	.037	.057	.022	.044	.013
	500	.048	.028	.056	.016	.049	.010
	1000	.045	.024	.046	.015	.053	.008

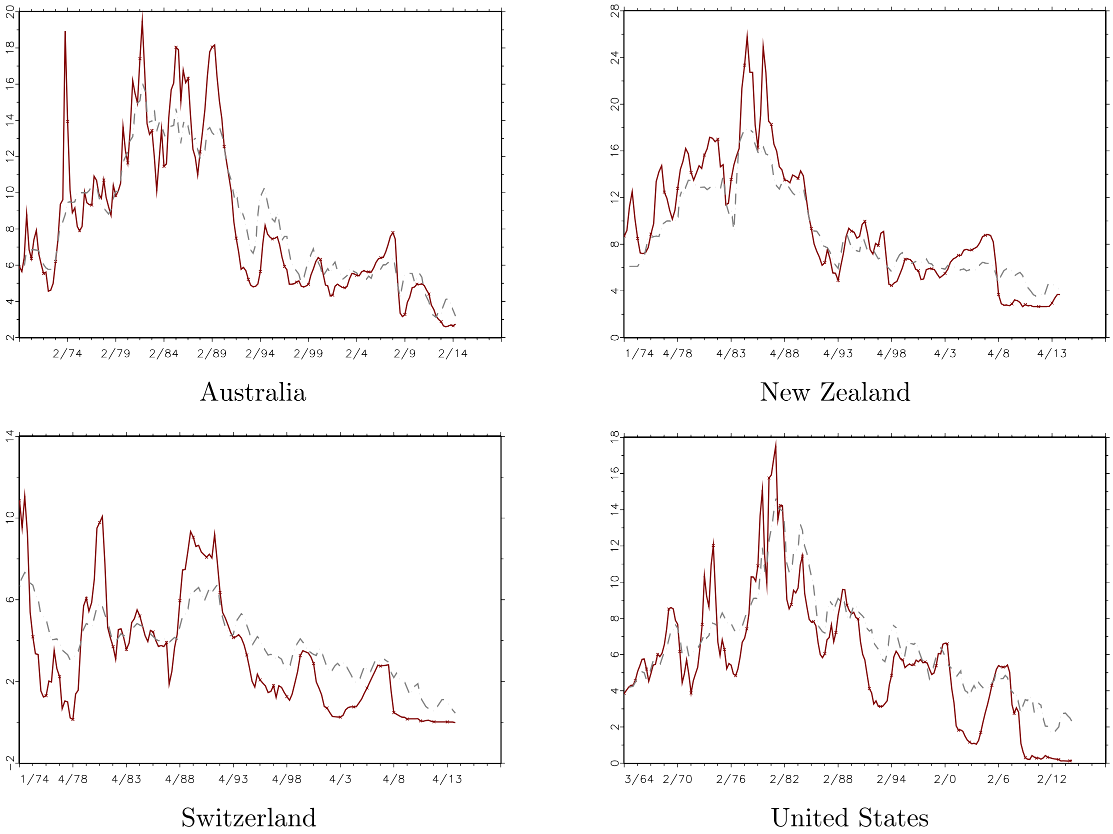
$p_T = [4(T/100)^{1/4}]$. δ: break direction. S / $t_{\hat{\xi}^*,\kappa}$: Simes / Hartung based on $\hat{Q}_i$.

unlikely to deviate too much from the short term rates. This is because the larger the gap, the higher the relative demand for longer-term investments, bringing down the long-term rate back towards its equilibrium value. The following simple error-correction summarizes the argument,

$$\begin{aligned} \Delta r_{St} &= \alpha_S(r_{Lt-1} - \beta r_{St-1}) + \epsilon_{St} \\ \Delta r_{Lt} &= \alpha_L(r_{Lt-1} - \beta r_{St-1}) + \epsilon_{Lt} \end{aligned}$$

where r_{Lt} and r_{St} denote the long term (here 10 years) and short term (here 3 months) interest rates at time t.

Despite a tremendous body of empirical evidence, a large portion of which employs cointegration methods (see, e.g., Campbell and Shiller, 1987; Hall et al., 1992; Taylor, 1992; Shea, 1992, for some



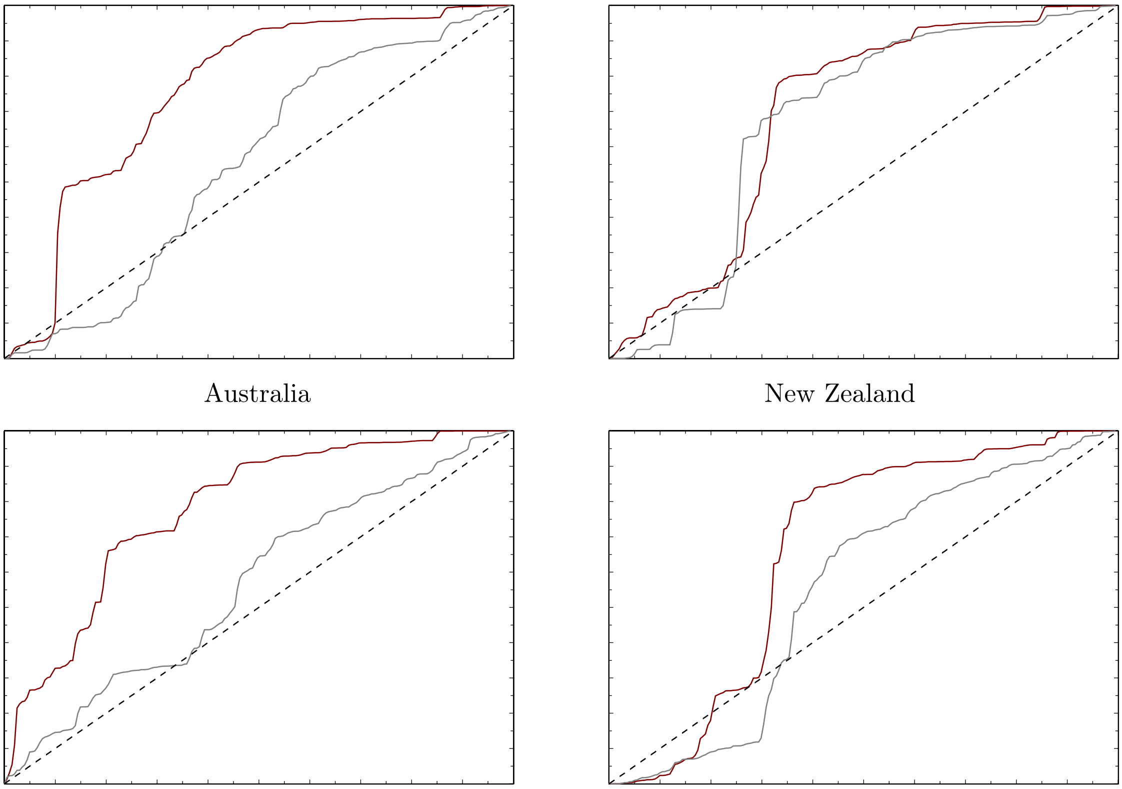
Solid red: Short term rate. Grey dash/dots: Long term rate

Figure 5: Selected interest rate series

seminal contributions), the evidence regarding the term structure theory still is mixed. Jondeau and Ricart (1999) emphasize sensitivity of the results to the particular country under study. This might point at the need to allow for flexible error structures, as, for instance, a homoskedasticity assumption may be compatible with the data in some countries but not in others. As with any time series application, the results may moreover be sensitive to the particular sampling period. Hence, an update using data encompassing the recent financial crisis seems worthwhile.

We use OECD data for $N = 34$ countries starting, for some countries, in 1953Q2 and running until 2014Q4.[9] The time series lengths vary from $T_i = 51$ (Japan) to $T_i = 240$ (Canada) for the short rates and from $T_i = 52$ (Slovenia) to $T_i = 247$ (USA) for the long rates. Within each country, we balance the two series to the shorter length. Some representative series are shown in Figure 5. We note that, possibly due to monetary policy innovations leading to more stable series, there seems to be a decrease in the volatility of the series. This is confirmed by the estimated variance profiles (see Cavaliere and Taylor, 2008b, eq. (12)) reported in Figure 6. These show what fraction of the total variance in the series has accrued after some fraction of the time series length s, $s \in [0, 1]$. All estimated profiles, and intuitively in particular those for

[9]The series can be accessed at OECD (2015a) and OECD (2015b). We drop the series for Chile, Hungary, Mexico and Luxemburg due to too many missing observations. For Japan, Portugal and Slovakia, there is a missing observation in one of the series which we interpolate linearly so as not to lose the information contained in these series.



Australia New Zealand

Switzerland United States

red: Short term rate. Grey: Long term rate. Dashes: $45°$-degree line. Both axes range from 0 to 1.

Figure 6: Estimated variance profiles

the short term rates, show marked deviations from the $45°$-degree line to which they should be close if the series were homoskedastic. The estimated profiles are generally located above the $45°$-degree line, demonstrating the late downward shifts in volatility.

We first present single-country evidence. We fix a common deterministic lag length of $[4 \cdot (T_i/100)^{1/4}]$ for all tests (but found all results to be robust for other choices). As a pretest for stationarity, we employ the nonstationary-volatility robust unit root test of Demetrescu and Hanck (201x). We only find rejections of the unit root null in at least one the series for Germany and South Africa and discard these from the cointegration analysis.

Table 10 presents the p-values of the $\hat{Q}_i$ test, the integrable IV-based test $Q_{\boldsymbol{\alpha}}$ from Demetrescu et al. (2014), the trace test of Johansen (1988) (λ_{tr}), the residual-based test of Engle and Granger (1987) (EG), the NIV test $\tilde{\tau}_i$ of Chang and Nguyen (2012) and the wild bootstrap test of Cavaliere et al. (2010, 2012, 2014) (CRT, with $B = 5000$ bootstrap replications), also studied in Section 4. Overall, the robust tests only find relatively weak evidence for cointegration. There are several individually strongly significant p-values for CRT, and a few around 0.05 for $\hat{Q}_i$ and $Q_{\boldsymbol{\alpha}}$. The average, however, of the p-values reported in the final row of Table 10 is comparatively close to 0.5, the theoretical mean if the p-values were generated under the non-cointegration null. While the p-values for $\hat{Q}_i$ and $Q_{\boldsymbol{\alpha}}$ correlate fairly strongly, those for CRT and $\hat{Q}_i$ do not. Such findings are not uncommon in the cointegration literature, see e.g. Gregory et al. (2004) and Pesavento

Table 10: Single-country p-values

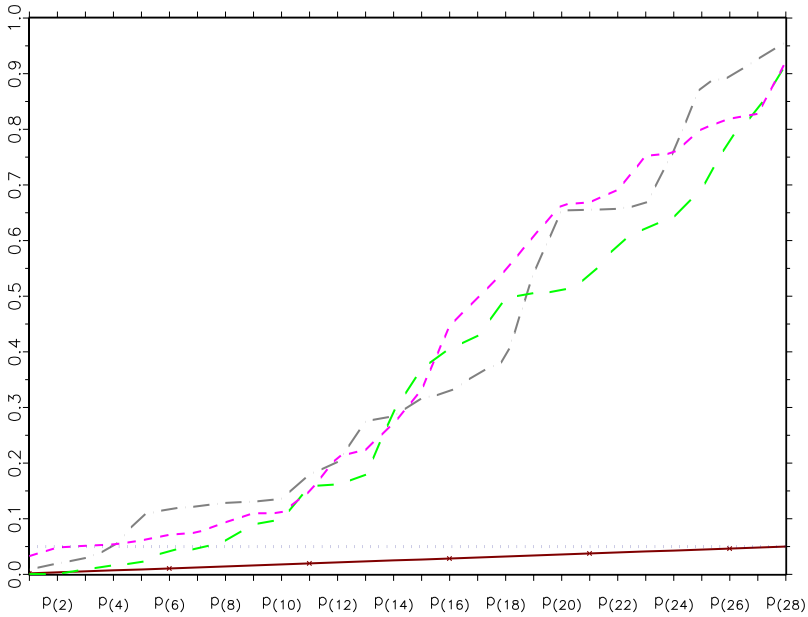
country	EG	$\tilde{\tau}_i$	λ_{tr}	CRT	$\hat{Q}_i$	$Q_{\boldsymbol{\alpha}}$
Australia	0.000	0.139	0.000	0.001	0.094	0.118
Austria	0.082	0.628	0.050	0.090	0.212	0.048
Belgium	0.000	0.593	0.000	0.001	0.075	0.893
Canada	0.002	0.589	0.001	0.009	0.061	0.655
Czech Republic	0.241	0.609	0.552	0.783	0.224	0.957
Denmark	0.026	0.587	0.030	0.158	0.608	0.665
Finland	0.297	0.893	0.208	0.595	0.691	0.122
France	0.000	0.378	0.017	0.045	0.052	0.316
Greece	0.319	0.338	0.501	0.693	0.271	0.654
Iceland	0.349	0.859	0.583	0.642	0.447	0.284
Ireland	0.146	0.511	0.119	0.370	0.669	0.542
Israel	0.228	1.000	0.159	0.534	0.828	0.179
Italy	0.306	0.739	0.430	0.835	0.801	0.136
Japan	0.609	0.966	0.539	0.917	0.053	0.275
Korea	0.063	0.351	0.094	0.177	0.033	0.029
Netherlands	0.057	0.226	0.202	0.405	0.048	0.131
New Zealand	0.001	0.622	0.002	0.045	0.332	0.361
Norway	0.017	0.593	0.017	0.061	0.757	0.657
Poland	0.062	0.046	0.366	0.622	0.071	0.128
Portugal	0.132	0.684	0.006	0.098	0.665	0.926
Russia	0.131	0.646	0.000	0.016	0.110	0.201
Slovak Republic	0.410	0.976	0.249	0.505	0.497	0.107
Slovenia	0.633	0.269	0.267	0.508	0.819	0.327
Spain	0.041	0.869	0.158	0.496	0.753	0.385
Sweden	0.065	0.595	0.152	0.428	0.922	0.883
Switzerland	0.060	0.276	0.052	0.162	0.547	0.761
United Kingdom	0.009	0.346	0.021	0.293	0.148	0.009
United States	0.007	0.212	0.011	0.021	0.110	0.019
average p-values	0.153	0.555	0.171	0.340	0.389	0.385

EG: Engle/Granger (Ecma 1987), $\tilde{\tau}_i$: Chang and Nguyen (JoE 2012), λ_{tr}: Johansen (1995), CRT: Cavaliere et al. (JoE 2010), $Q_{\boldsymbol{\alpha}}$: Demetrescu et al. (JTSA 2014). $\hat{Q}_i := \boldsymbol{t}_i' \widehat{\Xi}_i^{-1} \boldsymbol{t}_i$.

(2004).

The average p-values of the nonrobust EG and λ_{tr} tests are substantially smaller. We do not see this as evidence of higher power but rather as a reflection of their lack of robustness to nonstationary volatility. Note that the relatively early downward shift in volatility present in our data roughly agrees with the scenario investigated in the upper panel of Table 1. There, both tests were fairly strongly upward size distorted. Again in line with the conservativeness found in the simulations, the p-values of $\tilde{\tau}_i$ tend to be substantially larger than those of the other tests.

Next, we turn to the panel analysis. Figure 7 compares the above robust p-values to the Simes (1986) rejection curve. For CRT (this result can be sensitive to the particular bootstrap distribution), we would obtain a rejection of the panel null, although it is worth noticing that it would result from one or at most two small p-values. This is of course in line with the alternative of a test in heterogenous panels, viz. the existence of at least one cointegrated pair of series, but should serve to recall that such a rejection of a panel null may or may not be informative. For $Q_{\boldsymbol{\alpha}}$, we find marginal results with a test statistic of 81.75, quite close to the critical value of



Solid brown: Simes' cutoff line. Blue dots: 0.05. Long green dashes: p-values Cavaliere et al. (2010, 2012, 2014). Short magenta dashes: p-values $Q_{\boldsymbol{\alpha}}$ (Demetrescu et al., 2014). Grey dash/dots: p-values $\hat{Q}_i$.

Figure 7: Panel Results for Robust Tests

74.47. Given the small p-values of EG and λ_{tr}, a Simes (1986)-based panel version of these tests would of course also reject, but in view of the non-robustness of these tests, we would view such a rejection as spurious. The weak evidence in favor of panel cointegration might be explained by the time variation of the term premium; cf. Doh (2013).

Bearing in mind that long-run relations may need a longer time span to unfold, we perform a robustness check by eliminating the series which are not observed prior of the 80s and have mainly been subject to declining interest rates. The results support the same conclusion, and, together with the list of considered countries, are available upon request.

6 Concluding remarks

We propose a robust test for no cointegration. The individual-unit test is based on an IV test for no error correction in any of the equations of the system. In spite of standard asymptotics under the null, it has local power that typically exceeds that of the Johansen test when the alternative is close to the null. For aggregating N individual-unit statistics to obtain panel evidence, we resorted to multiple testing techniques and combinations of p-values.

In simulations, we demonstrate that traditional cointegration tests may be heavily size-distorted under nonstationary volatility. We find the tests developed here to be competitive with bootstrap-based alternatives that have been suggested in the literature.

In line with the Monte Carlo evidence, the application to the term structure theory of interest rates evidence shows that properly accounting for nonstationary volatility has an important impact on results in applied work.

Appendix

Before proceeding to the proofs, we state and prove a useful lemma. Note: sums run from p_T+2 to T unless specified otherwise.

Lemma 2 *Under the assumptions of Proposition 1, we have as $T \to \infty$ that*

1. $\frac{1}{\sqrt{T}} \tilde{e}_{i[sT]} \Rightarrow \tilde{E}(s)$;
2. $\Pr(|\tilde{e}_{it-1} - \tilde{e}_{it-l}| > |\tilde{e}_{it-l}|) \leq C \Pr(\|\boldsymbol{w}_{it-1} - \boldsymbol{w}_{it-l}\| > |\tilde{e}_{it-l}|) + o(1)$ *uniformly for all l such that $p_T/l + l/\sqrt{T} \to 0$;*
3. $\left\| \sum_{t=p_T+2}^{T} \operatorname{sgn}(\tilde{e}_{it-1}) \boldsymbol{x}_{it-1} \right\| = o_p\left(\sqrt{p_T} T^{0.75}\right)$;
4. $\left\| \sum_{t=p_T+2}^{T} \boldsymbol{x}_{it-1} \boldsymbol{w}_{it}' \right\| = O_p\left(p_T^{0.5} T\right)$;
5. $\left\| \sum_{t=p_T+2}^{T} \boldsymbol{x}_{it-1} \tilde{e}_{it-1} \right\| = O_p\left(p_T^{0.5} T\right)$;
6. $\left\| \sum_{t=p_T+2}^{T} \boldsymbol{x}_{it-1} \boldsymbol{x}_{it-1}' \right\| = \Theta_p(T)$;
7. $\frac{1}{T^{0.5}} \sum \operatorname{sgn}(\tilde{e}_{it-1}) \boldsymbol{\varepsilon}_{tk}^{(p_T)} \Rightarrow \int_0^1 \operatorname{sgn}\left(\tilde{E}_i(s)\right) \mathrm{d}\boldsymbol{M}_i(s)$.
8. $\left\| \sum_{t=p_T+2}^{T} \boldsymbol{x}_{it-1} \boldsymbol{\varepsilon}_{it}^{(p_T)} \right\| = O_p\left(p_T^{0.5} T^{0.5}\right)$;
9. $\frac{1}{T^{1.5}} \sum |\tilde{e}_{t-1}| \Rightarrow \int_0^1 \left|\tilde{E}(s)\right| \mathrm{d}s$;
10. $\frac{1}{T\sqrt{T}} \sum_{t=p_T+2}^{T} \operatorname{sgn}(\tilde{e}_{it-1}) \boldsymbol{w}_{it-1} \Rightarrow \int_0^1 \operatorname{sgn}\left(\tilde{E}_i(s)\right) \boldsymbol{J}_i(s) \mathrm{d}s$.

where the convergence is joint with that in Equation (1).

Proof of Lemma 2

1. With obvious notation $\tilde{\boldsymbol{w}}_{it} = (\tilde{w}_{it1}, \tilde{\boldsymbol{w}}_{it2}')'$, note first that

$$\frac{1}{\sqrt{T}} \tilde{e}_{it-1} = \frac{1}{\sqrt{T}} \tilde{w}_{it-11} - \frac{1}{\sqrt{T}} \tilde{\boldsymbol{w}}_{it-12}' \left(\frac{1}{T^2} \sum_{j=1}^{t-1} \tilde{\boldsymbol{w}}_{ij2} \tilde{\boldsymbol{w}}_{ij2}' \right)^{-1} \left(\frac{1}{T^2} \sum_{j=1}^{t-1} \tilde{\boldsymbol{w}}_{ij2} \tilde{w}_{ij1} \right),$$

leading thanks e.g. to Proposition 2 in Born and Demetrescu (2015) and the continuous mapping theorem [CMT] to the desired result. (Note also that $\tilde{E}_i(s)$ is pathwise continuous and as such uniformly bounded in probability on $[0,1]$, and so is $\frac{1}{\sqrt{T}} \tilde{e}_{it-1}$ thanks to the weak convergence.)

2. We show first that $|\tilde{e}_{it-1} - \tilde{e}_{it-l}| \leq C_T \|\boldsymbol{w}_{it-1} - \boldsymbol{w}_{it-l}\| + o_p(1)$ where $C_T = O_p(1)$ and the o_p and O_p terms are uniform in l, t. To this end, we discuss the bivariate case only to ease notation; the extension to $K > 2$ is straightforward. Let first $\Delta_{l-1} = 1 - L^{l-1}$ and $\hat{\beta}_{it-1} = \frac{\sum_{j=1}^{t-1} w_{ij1} w_{ij2}}{\sum_{j=1}^{t-1} w_{ij2}^2}$. Then,

$$\begin{aligned}\tilde{e}_{it-1} - \tilde{e}_{it-l} &= \Delta_{l-1}\tilde{e}_{it-1} = \Delta_{l-1} w_{it-11} - \Delta_{l-1}\left(w_{it-12}\hat{\beta}_{it-1}\right) \\ &= \Delta_{l-1} w_{it-11} - \hat{\beta}_{it-1}\Delta_{l-1} w_{it-12} - \frac{w_{it-l2}}{\sqrt{T}}\sqrt{T}\Delta_{l-1}\hat{\beta}_{it-1}\end{aligned}$$

where

$$\begin{aligned}\sqrt{T}\Delta_{l-1}\hat{\beta}_{it-1} &= \sqrt{T}\frac{\left(\sum_{j=1}^{t-1} w_{ij1}w_{ij2}\right)\left(\sum_{j=1}^{t-l} w_{ij2}^2\right) - \left(\sum_{j=1}^{t-l} w_{ij1}w_{ij2}\right)\left(\sum_{j=1}^{t-1} w_{ij2}^2\right)}{\left(\sum_{j=1}^{t-1} w_{ij2}^2\right)\left(\sum_{j=1}^{t-l} w_{ij2}^2\right)} \\ &= \frac{\left(\frac{1}{T^2}\sum_{j=1}^{t-l} w_{ij2}^2\right)}{\left(\frac{1}{T^2}\sum_{j=1}^{t-1} w_{ij2}^2\right)\left(\frac{1}{T^2}\sum_{j=1}^{t-l} w_{ij2}^2\right)}\left(\frac{1}{T^{3/2}}\sum_{j=t-l+1}^{t-1} w_{ij1}w_{ij2}\right) \\ &\quad - \frac{\left(\frac{1}{T^2}\sum_{j=1}^{t-l} w_{ij1}w_{ij2}\right)}{\left(\frac{1}{T^2}\sum_{j=1}^{t-1} w_{ij2}^2\right)\left(\frac{1}{T^2}\sum_{j=1}^{t-l} w_{ij2}^2\right)}\left(\frac{1}{T^{3/2}}\sum_{j=t-l+1}^{t-1} w_{ij2}^2\right).\end{aligned}$$

Now,

$$\max_{t,l}\frac{1}{T}\left|\sum_{j=t-l+1}^{t-1} w_{ij1}w_{ij2}\right| \leq \sum_{j=t-l+1}^{t-1}\max_{t,l}\left|\frac{w_{ij1}}{\sqrt{T}}\frac{w_{ij2}}{\sqrt{T}}\right| \leq Cl\max_t\left\|\frac{1}{\sqrt{T}}\boldsymbol{w}_{it-1}\right\|^2$$

and correspondingly

$$\max_{t,l}\frac{1}{T}\left|\sum_{j=t-l+1}^{t-1} w_{ij2}^2\right| \leq Cl\max_t\left\|\frac{1}{\sqrt{T}}\boldsymbol{w}_{it-1}\right\|^2.$$

This, together with $\left(\frac{1}{T^2}\sum_{j=1}^{t-l} w_{ij2}^2\right)$ being bounded away from zero w.p.1 (the innovations have uniformly bounded density function, such that the probability that $w_{ijk} = 0$ is zero), implies that

$$\frac{\left(\frac{1}{T^2}\sum_{j=1}^{t-l} w_{ij2}^2\right)}{\left(\frac{1}{T^2}\sum_{j=1}^{t-1} w_{ij2}^2\right)\left(\frac{1}{T^2}\sum_{j=1}^{t-l} w_{ij2}^2\right)} = O_p(1)$$

and

$$\frac{\left(\frac{1}{T^2}\sum_{j=1}^{t-l} w_{ij1}w_{ij2}\right)}{\left(\frac{1}{T^2}\sum_{j=1}^{t-1} w_{ij2}^2\right)\left(\frac{1}{T^2}\sum_{j=1}^{t-l} w_{ij2}^2\right)} = O_p(1)$$

uniformly. Hence

$$\sup_t\left|\sqrt{T}\Delta_{l-1}\hat{\beta}_{it-1}\right| = o_p(1),$$

from which the stated upper bound for $|\tilde{e}_{it-1} - \tilde{e}_{it-l}|$ follows given that $\sup_t \left|\hat{\beta}_{it-1}\right| = O_p(1)$. The desired result is obtained by integrating out C_T.

3. Let B_{ij} denote the coefficient matrices of the inverse of the matrix lag polynomial $I_K - \sum_{j\geq 1} \Gamma_{ij} L^j$. Note that, for all $1 \leq j \leq p_T$,

$$\sum \operatorname{sgn}(\tilde{e}_{it-1}) \Delta \boldsymbol{w}_{it-j} = \sum \operatorname{sgn}(\tilde{e}_{it-1}) \sum_{k\geq 0} B_{ik} \boldsymbol{\varepsilon}_{it-j-k} + O\left(\max_{1\leq t\leq T} \|\boldsymbol{w}_{it}\|\right)$$

where $T^{-1/2}\boldsymbol{w}_{it}$ is uniformly bounded in probability such that $\max_{1\leq t\leq T} \|\boldsymbol{w}_{it}\| = O_p\left(\sqrt{T}\right)$ uniformly in j. Since $m > 4$, we may pick some $l = T^\lambda$ with $\lambda \in \left(\delta, \frac{1}{2} - \frac{1}{2(m-1)}\right)$ and write

$$\begin{aligned}
&\sum \operatorname{sgn}(\tilde{e}_{it-1}) \sum_{k\geq 0} B_{ik} \boldsymbol{\varepsilon}_{it-j-k} = \\
&\quad = \sum_{t=p_T+2}^{l} \operatorname{sgn}(\tilde{e}_{it-1}) \sum_{k\geq 0} B_{ik} \boldsymbol{\varepsilon}_{it-j-k} + \sum_{t=l+1}^{T} \operatorname{sgn}(\tilde{e}_{it-l}) \sum_{k=0}^{l-j-1} B_{ik} \boldsymbol{\varepsilon}_{it-j-k} \\
&\qquad + \sum_{t=l+1}^{T} \operatorname{sgn}(\tilde{e}_{it-l}) \sum_{k\geq l-j} B_{ik} \boldsymbol{\varepsilon}_{it-j-k} + \sum_{t=l+1}^{T} (\operatorname{sgn}(\tilde{e}_{it-1}) - \operatorname{sgn}(\tilde{e}_{it-l})) \sum_{k\geq 0} B_{ik} \boldsymbol{\varepsilon}_{it-j-k}. \\
&\quad = A_{T,j} + B_{T,j} + C_{T,j} + D_{T,j}.
\end{aligned}$$

Now, we obviously have that $\|A_{T,j}\|_m \leq \sum_{t=p_T+2}^{l} \left\|\sum_{k\geq 0} B_{ik} \boldsymbol{\varepsilon}_{it-j-k}\right\|_m \leq Cl$ such that $\frac{1}{l} A_{T,j}$ is uniformly L_m-bounded for $m > 4$ and hence

$$\max_{1\leq j\leq p_T} \left|\frac{1}{l} A_{T,j}\right| = o_p\left(p_T^{1/4}\right)$$

where $l p_T^{1/4} = o\left(T^{3/4}\right)$. Then, for all j, $\operatorname{Var}(B_{T,j}) \leq CT$ since the term can be written as a sum of md terms having uniformly bounded variance, and we have

$$\max_{1\leq j\leq p_T} \left|\frac{1}{\sqrt{T}} B_{T,j}\right| = O_p\left(p_T^{1/2+\epsilon}\right)$$

for any $\epsilon > 0$, so ϵ can be chosen such that $\sqrt{T} p_T^{1/2+\epsilon} = o\left(T^{3/4}\right)$. Furthermore, $\|C_{T,j}\|_m = O\big(T \sum_{k\geq l-j} \|B_{ik}\|\big)$ or $O\left(Te^{-(l-j)}\right)$ due to the fact that the coefficient matrices B_{ik} have exponential decay, and

$$\max_{1\leq j\leq p_T} \left|\frac{1}{Te^{-(l-j)}} C_{T,j}\right| = o_p\left(p_T^{1/4}\right)$$

where $p_T^{1/4} Te^{-(l-j)} \to 0$ at any polynomial rate and hence also 3/4 since the exponential dominates. Summing up, $A_{T,j}$, $B_{T,j}$, $C_{T,j}$ are dominated and the result follows if

$$\max_{1\leq j\leq p_T} |D_{T,j}| = o_p\left(T^{3/4}\right).$$

To show this, apply Hölder's inequality with $1 = (m-1)/m + 1/m$ to obtain

$$|D_{T,j}| \leq \sqrt[\frac{m}{m-1}]{\sum_{t=l+1}^{T} |\mathrm{sgn}\,(\tilde{e}_{it-1}) - \mathrm{sgn}\,(\tilde{e}_{it-l})|^{\frac{m}{m-1}}} \sqrt[m]{\sum_{t=l+1}^{T} \left\| \sum_{k\geq 0} B_{ik} \boldsymbol{\varepsilon}_{it-j-k} \right\|^{m}}.$$

Given the finiteness of mth order moments of $\boldsymbol{\varepsilon}_t$, the second term of the r.h.s. is of order $O_p\left(T^{1/m}\right)$, where it is tedious, yet straightforward to shown this to occur uniformly in $1 \leq j \leq p_T$, and we only have to examine the behavior of the first term, which does not depend on j. Note that, for suitable C,

$$\begin{aligned}
\mathrm{E}\left(\sum_{t=l+1}^{T} |\mathrm{sgn}\,(\tilde{e}_{it-1}) - \mathrm{sgn}\,(\tilde{e}_{it-l})|^{m/(m-1)}\right) &= C \sum_{t=s+1}^{T} \mathrm{E}\big(|\mathrm{sgn}\,(\tilde{e}_{it-1}) - \mathrm{sgn}\,(\tilde{e}_{it-l})|\big) \\
&= C \sum_{t=s+1}^{T} \Pr\left(|\mathrm{sgn}\,(\tilde{e}_{it-1}) - \mathrm{sgn}\,(\tilde{e}_{it-l})| = 2\right).
\end{aligned}$$

Each probability on the r.h.s. is given by the (unconditional) probability of a change of sign from $t-l$ to t,

$$\begin{aligned}
\Pr\left(|\mathrm{sgn}\,(\tilde{e}_{it-1}) - \mathrm{sgn}\,(\tilde{e}_{it-l})| = 2\right) &\leq 2 \Pr\left(|\tilde{e}_{it-1} - \tilde{e}_{it-l}| > |\tilde{e}_{it-l}|\right) \\
&\leq C \Pr\left(\|\boldsymbol{w}_{it-1} - \boldsymbol{w}_{it-l}\| > |\tilde{e}_{it-l}|\right)
\end{aligned}$$

thanks to item 2 of this lemma. The quantities $W_{T,t} = \frac{1}{\sqrt{l}} \|\boldsymbol{w}_{it-1} - \boldsymbol{w}_{it-l}\|$ and $E_{T,t} = \frac{1}{\sqrt{T}} \tilde{e}_{it-l}$ have a well-defined joint distribution; because of the weak convergence in Lemma 1 and item 1 of this lemma, and of the assumption on the conditional densities of the shocks, the density function, say $f_{T,t}(w,e)$, exists and is nonzero a.e.; hence

$$\begin{aligned}
\Pr\left(\|\boldsymbol{w}_{it-1} - \boldsymbol{w}_{it-l}\| > |\tilde{e}_{it-l}|\right) &= \Pr\left(W_{T,t} > \frac{\sqrt{T}}{\sqrt{l}} |E_{T,t}|\right) = \int\!\!\int_{w > \frac{\sqrt{T}}{\sqrt{l}}|e|} f_{T,t}(w,e)\, \mathrm{d}w \mathrm{d}e \\
&= \sqrt{\frac{l}{T}} \int\!\!\int_{w>|u|} f_{T,t}\left(w, \sqrt{\frac{l}{T}} u\right) \mathrm{d}w \mathrm{d}u \\
&\leq C \sqrt{\frac{l}{T}}.
\end{aligned}$$

Thus,

$$\sum_{t=l+1}^{T} \Pr\left(|\mathrm{sgn}\,(\tilde{e}_{it-1}) - \mathrm{sgn}\,(\tilde{e}_{it-l})| = 2\right) \leq C T^{1/2} \sqrt{l}$$

such that, uniformly in $1 \leq j \leq p_T$,

$$\left| \sum_{t=s+1}^{T} \left(\mathrm{sgn}\,(\tilde{e}_{it-1}) - \mathrm{sgn}\,(\tilde{e}_{it-l})\right) \sum_{k\geq 0} B_{ik} \boldsymbol{\varepsilon}_{it-j-k} \right| = O_p\left(\left(T^{1/2}\sqrt{l}\right)^{(m-1)/m} T^{1/m}\right).$$

The result follows, since $\left(T^{1/2}\sqrt{l}\right)^{(m-1)/m} T^{1/m} = o\left(T^{3/4}\right)$ for $\delta < \lambda < \frac{1}{2} - \frac{1}{2(m-1)}$ under Assumption 2.

4. This was proved in the univariate homoskedastic case by Chang and Park (2002), Lemma 3.2(b); the multivariate extension to time-varying variance is straightforward and omitted.

5. Begin by noting that

$$\sum_{t=p_T+2}^{T} \boldsymbol{x}_{it-1}\tilde{e}_{it-1} = \sum_{t=p_T+2}^{T} \boldsymbol{x}_{it-1}\tilde{w}_{it-11} - \sum_{t=p_T+2}^{T} \boldsymbol{x}_{it-1}\tilde{\boldsymbol{w}}'_{it-12}\left(\sum_{j=1}^{t-1}\tilde{\boldsymbol{w}}_{ij2}\tilde{\boldsymbol{w}}'_{ij2}\right)^{-1}\left(\sum_{j=1}^{t-1}\tilde{\boldsymbol{w}}_{ij2}\tilde{w}_{ij1}\right),$$

where $\left(\sum_{j=1}^{t-1}\tilde{\boldsymbol{w}}_{ij2}\tilde{\boldsymbol{w}}'_{ij2}\right)^{-1}\left(\sum_{j=1}^{t-1}\tilde{\boldsymbol{w}}_{ij2}\tilde{w}_{ij1}\right)$ is uniformly bounded in probability thanks to the weak convergence in item 1 of this lemma. Then, $\sum \boldsymbol{x}_{it-1}\tilde{w}_{it-11}$ and $\sum \boldsymbol{x}_{it-1}\tilde{\boldsymbol{w}}'_{it-12}$ are dealt with like in Lemma A.1.F in Demetrescu and Hanck (201x), and the result follows.

6. This is a straightforward multivariate extension of e.g. Chang and Park (2002), Lemma 3.2(a) and we omit the details.

7. Begin by noting that $\|\Delta\boldsymbol{w}_{it}\|_4 \leq C_1 + C_2\|\boldsymbol{w}_{it-1}\|_4$ thanks to the exponential decay of B_{ij} and the uniform L_4-boundedness of $\boldsymbol{\varepsilon}_{it}$; then, $\|\boldsymbol{w}_{it-1}\|_4 \leq \sum_{j=1}^{t-1}\|\Delta\boldsymbol{w}_{ij}\|_4$, such that if $\|\Delta\boldsymbol{w}_{i1}\|_4 < \infty$ it is easily shown by induction that $\|\Delta\boldsymbol{w}_{it}\|_4$ is uniformly L_4 bounded. Let us then examine

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{t=p_T+2}^{T}\boldsymbol{\varepsilon}_{it}^{(p_T)}\operatorname{sgn}(\tilde{e}_{it-1})\\
&= \frac{1}{\sqrt{T}}\sum_{t=p_T+2}^{T}\boldsymbol{\varepsilon}_{it}\operatorname{sgn}(\tilde{e}_{it-1}) + \frac{1}{\sqrt{T}}\sum_{t=p_T+2}^{T}\sum_{j=p_T+1}^{\infty}\Gamma_{ij}\Delta\boldsymbol{w}_{it-j}\operatorname{sgn}(\tilde{e}_{it-1})\\
&= \frac{1}{\sqrt{T}}\sum_{t=p_T+2}^{T}\boldsymbol{\varepsilon}_{it}\operatorname{sgn}(\tilde{e}_{it-1}) + \sum_{j=p_T+1}^{\infty}\Gamma_{ij}\left(\frac{1}{\sqrt{T}}\sum_{t=p_T+2}^{T}\Delta\boldsymbol{w}_{it-j}\operatorname{sgn}(\tilde{e}_{it-1})\right).
\end{aligned}$$

Thanks to $\Delta\boldsymbol{w}_{it-j}\operatorname{sgn}(\tilde{e}_{it-1})$ being, just like $\Delta\boldsymbol{w}_{it}$, uniformly L_2-bounded, we have that

$$\sup_{j>p_T}\left\|\frac{1}{\sqrt{T}}\sum_{t=p_T+2}^{T}\Delta\boldsymbol{w}_{it-j}\operatorname{sgn}(\tilde{e}_{it-1})\right\|_2 \leq \frac{1}{\sqrt{T}}\sum_{t=p_T+2}^{T}\sup_{j>p_T,1\leq t\leq T}\|\Delta\boldsymbol{w}_{it-j}\operatorname{sgn}(\tilde{e}_{it-1})\|_2 = \Theta\left(\sqrt{T}\right).$$

Recall that $\|\Gamma_{ij}\| = O\left(e^{-j}\right)$ so $\sum_{j=p_T+1}^{\infty}\|\Gamma_{ij}\| = O\left(e^{-p_T}\right)$ with $p_T = CT^{\delta}$; it follows that

$$\left\|\sum_{j=p_T+1}^{\infty}\Gamma_{ij}\left(\frac{1}{\sqrt{T}}\sum_{t=p_T+2}^{T}\Delta\boldsymbol{w}_{it-j}\operatorname{sgn}(\tilde{e}_{it-1})\right)\right\|_2 \leq C\sqrt{T}\sum_{j=p_T+1}^{\infty}\|\Gamma_{ij}\| = O\left(e^{-p_T}\sqrt{T}\right) = o(1)$$

and Chebyschev's inequality takes us to

$$\frac{1}{\sqrt{T}}\sum_{t=p_T+2}^{T}\boldsymbol{\varepsilon}_{it}^{(p_T)}\operatorname{sgn}\left(\tilde{e}_{it-1}\right)=\frac{1}{\sqrt{T}}\sum_{t=p_T+2}^{T}\boldsymbol{\varepsilon}_{it}\operatorname{sgn}\left(\tilde{e}_{it-1}\right)+o_p\left(1\right).$$

The sign of $\tilde{e}_{i[sT]-1}$ converges weakly to $\operatorname{sgn}\left(\tilde{E}_i\left(s\right)\right)$, see Christopeit (2009), and the weak convergence to the Ito-type integral then follows with the arguments used by So and Shin (1999) in the proof of their Theorem 1(ii).

8. We have

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=p_T+2}^{T}\boldsymbol{x}_{it-1}\boldsymbol{\varepsilon}_{it}^{(p_T)} &= \frac{1}{\sqrt{T}}\sum_{t=p_T+2}^{T}\boldsymbol{x}_{it-1}\boldsymbol{\varepsilon}_{it}+\sum_{j=p_T+1}^{\infty}\Gamma_{ij}\left(\frac{1}{\sqrt{T}}\sum_{t=p_T+2}^{T}\boldsymbol{x}_{it-1}\Delta\boldsymbol{w}_{it-j}\right)\\ &= \Theta_p\left(1\right)\end{aligned}$$

because $\left\|\frac{1}{\sqrt{T}}\sum_{t=p_T+2}^{T}\boldsymbol{x}_{it-1}\Delta\boldsymbol{w}_{it-j}\right\|_2\leq Cp_T^{0.5}\sqrt{T}$ for all $j\in\mathbb{N}$ and $\sum_{j=p_T+1}^{\infty}\left\|\Gamma_{ij}\right\|=O\left(e^{-p_T}\right)$ as in the proof of item 7.

9. Follows with the CMT.

10. Follows with the CMT.

Proof of Proposition 1

We begin by showing that the estimators $\hat{\alpha}_k$ are superconsistent. We have with

$$\Delta w_{itk}=\alpha_i\boldsymbol{\beta}_i\boldsymbol{w}_{it-1}+\boldsymbol{\gamma}_{ip_T}'\boldsymbol{x}_{it-1}+\varepsilon_{itk}^{(p_T)},$$

where $\boldsymbol{\gamma}_{ip_T}$ stacks $\boldsymbol{\gamma}_{ij}$ for $j=1,\ldots,p_T$, that

$$\begin{aligned}T\hat{\alpha}_{ik} &= \frac{\frac{1}{\sqrt{T}}\sum\operatorname{sgn}\left(\tilde{e}_{it-1}\right)\varepsilon_{itk}^{(p_T)}-\frac{1}{T}\sum\operatorname{sgn}\left(\tilde{e}_{it-1}\right)\boldsymbol{x}_{it-1}'\left(\frac{1}{T}\sum\boldsymbol{x}_{it-1}\boldsymbol{x}_{it-1}'\right)^{-1}\frac{1}{\sqrt{T}}\sum\boldsymbol{x}_{it-1}\varepsilon_{itk}^{(p_T)}}{\frac{1}{T^{1.5}}\sum\left|\tilde{e}_{it-1}\right|-\frac{1}{T}\sum\operatorname{sgn}\left(\tilde{e}_{it-1}\right)\boldsymbol{x}_{it-1}'\left(\frac{1}{T}\sum\boldsymbol{x}_{it-1}\boldsymbol{x}_{it-1}'\right)^{-1}\frac{1}{T^{1.5}}\sum\boldsymbol{x}_{it-1}\tilde{e}_{it-1}}\\ &\quad+\alpha_{ik}\frac{\frac{1}{T^{1.5}}\sum\operatorname{sgn}\left(\tilde{e}_{it-1}\right)\boldsymbol{w}_{it-1}'-\frac{1}{T}\sum\operatorname{sgn}\left(\tilde{e}_{it-1}\right)\boldsymbol{x}_{it-1}'\left(\frac{1}{T}\sum\boldsymbol{x}_{it-1}\boldsymbol{x}_{it-1}'\right)^{-1}\frac{1}{T^{1.5}}\sum\boldsymbol{x}_{it-1}\boldsymbol{w}_{it-1}'}{\frac{1}{T^{1.5}}\sum\left|\tilde{e}_{it-1}\right|-\frac{1}{T}\sum\operatorname{sgn}\left(\tilde{e}_{it-1}\right)\boldsymbol{x}_{it-1}'\left(\frac{1}{T}\sum\boldsymbol{x}_{it-1}\boldsymbol{x}_{it-1}'\right)^{-1}\frac{1}{T^{1.5}}\sum\boldsymbol{x}_{it-1}\tilde{e}_{it-1}}\boldsymbol{\beta}_i.\end{aligned}$$

With Lemma 2 it follows that

$$\begin{aligned}&\left\|\frac{1}{T}\sum\operatorname{sgn}\left(\tilde{e}_{it-1}\right)\boldsymbol{x}_{it-1}'\left(\frac{1}{T}\sum\boldsymbol{x}_{it-1}\boldsymbol{x}_{it-1}'\right)^{-1}\frac{1}{\sqrt{T}}\sum\boldsymbol{x}_{it-1}\varepsilon_{itk}^{(p_T)}\right\|\\ &\quad\leq\left\|\frac{1}{T}\sum\operatorname{sgn}\left(\tilde{e}_{it-1}\right)\boldsymbol{x}_{it-1}'\right\|\left\|\left(\frac{1}{T}\sum\boldsymbol{x}_{it-1}\boldsymbol{x}_{it-1}'\right)^{-1}\right\|\left\|\frac{1}{\sqrt{T}}\sum\boldsymbol{x}_{it-1}\varepsilon_{itk}^{(p_T)}\right\|\\ &\quad\overset{p}{\to}0,\end{aligned}$$

$$
\begin{aligned}
&\left\| \frac{1}{T} \sum \operatorname{sgn}\left(\tilde{e}_{it-1}\right) \boldsymbol{x}_{it-1}' \left(\frac{1}{T} \sum \boldsymbol{x}_{it-1} \boldsymbol{x}_{it-1}' \right)^{-1} \frac{1}{T^{1.5}} \sum \boldsymbol{x}_{it-1} \tilde{e}_{it-1} \right\| \\
&\quad \leq \left\| \frac{1}{T} \sum \operatorname{sgn}\left(\tilde{e}_{it-1}\right) \boldsymbol{x}_{it-1}' \right\| \left\| \left(\frac{1}{T} \sum \boldsymbol{x}_{it-1} \boldsymbol{x}_{it-1}' \right)^{-1} \right\| \left\| \frac{1}{T^{1.5}} \sum \boldsymbol{x}_{it-1} \tilde{e}_{it-1} \right\| \\
&\quad \overset{p}{\to} 0
\end{aligned}
$$

and

$$
\begin{aligned}
&\left\| \frac{1}{T} \sum \operatorname{sgn}\left(\tilde{e}_{it-1}\right) \boldsymbol{x}_{it-1}' \left(\frac{1}{T} \sum \boldsymbol{x}_{it-1} \boldsymbol{x}_{it-1}' \right)^{-1} \frac{1}{T^{1.5}} \sum \boldsymbol{x}_{it-1} \boldsymbol{w}_{it-1}' \right\| \\
&\quad \leq \left\| \frac{1}{T} \sum \operatorname{sgn}\left(\tilde{e}_{it-1}\right) \boldsymbol{x}_{it-1}' \right\| \left\| \left(\frac{1}{T} \sum \boldsymbol{x}_{it-1} \boldsymbol{x}_{it-1}' \right)^{-1} \right\| \left\| \frac{1}{T^{1.5}} \sum \boldsymbol{x}_{it-1} \boldsymbol{w}_{it-1}' \right\| \\
&\quad \overset{p}{\to} 0.
\end{aligned}
$$

Summing up, Lemma 2 and the CMT take us to

$$
T \hat{\alpha}_k \overset{d}{\to} \frac{\int_0^1 \operatorname{sgn}\left(\tilde{E}_i(s)\right) \mathrm{d}M_{ik}}{\int_0^1 \left|\tilde{E}_i(s)\right| \mathrm{d}s} + \alpha_{k1} \boldsymbol{\beta}_1' \frac{\int_0^1 \operatorname{sgn}\left(\tilde{E}(s)\right) \boldsymbol{J}_i(s) \, \mathrm{d}s}{\int_0^1 \left|\tilde{E}_i(s)\right| \mathrm{d}s}.
$$

It can also be checked that $\hat{\Gamma}_{ij}$ are $\sqrt{T}$-consistent. Then, the superconsistency of $\hat{\alpha}_k$ under the null and the local alternative implies in turn that the residuals $\hat{\varepsilon}_{itk}^{(p_T)}$ are consistent, $\sup_t \left|\hat{\varepsilon}_{itk}^{(p_T)} - \varepsilon_{itk}\right| = o_p(1)$, since $\sup_t |\boldsymbol{y}_{it}| = \Theta_p\left(\sqrt{T}\right)$ while $\Delta \boldsymbol{w}_{it}$, being uniformly L_4-bounded (see item 7 of Lemma 2), satisfies $\sup_t \|\Delta \boldsymbol{w}_{it}\| = O_p\left(T^{1/4}\right)$.

For the residual variance estimators convergence also occurs, but to the average covariance ω_{ikl}:

$$
\frac{1}{T} \sum \hat{\varepsilon}_{itk}^{(p_T)} \hat{\varepsilon}_{itl}^{(p_T)} = \frac{1}{T} \sum \varepsilon_{itk} \varepsilon_{itl} + o_p(1).
$$

Since

$$
\frac{1}{T} \sum \varepsilon_{itk} \varepsilon_{itl} \overset{p}{\to} \omega_{ikl} \tag{8}
$$

along the lines of Corollary A.1 in Cavaliere and Taylor (2009), consistency of the residual covariance matrix for Ω_i follows (and also of residual correlation matrix for Ξ_i). This implies consistency of the i,kth residual variance estimator for ω_{ik}.

Moving on to the standard errors, note that Lemma 2 gives

$$
\begin{aligned}
&\left\| \frac{1}{T} \sum \operatorname{sgn}\left(\tilde{e}_{it-1}\right) \boldsymbol{x}_{it-1}' \left(\frac{1}{T} \sum \boldsymbol{x}_{it-1} \boldsymbol{x}_{it-1}' \right)^{-1} \frac{1}{T} \sum \boldsymbol{x}_{it-1} \operatorname{sgn}\left(\tilde{e}_{it-1}\right) \right\| \\
&\quad \leq \left\| \frac{1}{T} \sum \operatorname{sgn}\left(\tilde{e}_{it-1}\right) \boldsymbol{x}_{it-1}' \right\| \left\| \left(\frac{1}{T} \sum \boldsymbol{x}_{it-1} \boldsymbol{x}_{it-1}' \right)^{-1} \right\| \left\| \frac{1}{T} \sum \boldsymbol{x}_{it-1} \operatorname{sgn}\left(\tilde{e}_{it-1}\right) \right\| \\
&\quad \overset{p}{\to} 0
\end{aligned}
$$

such that

$$\frac{1}{T}\sum \operatorname{sgn}^2\left(\tilde{e}_{it-1}\right)-\frac{1}{T}\sum \operatorname{sgn}\left(\tilde{e}_{it-1}\right)\boldsymbol{x}'_{it-1}\left(\frac{1}{T}\sum \boldsymbol{x}_{it-1}\boldsymbol{x}'_{it-1}\right)^{-1}\frac{1}{T}\sum \boldsymbol{x}_{it-1}\operatorname{sgn}\left(\tilde{e}_{it-1}\right)=1+o_p\left(1\right).$$

We hence obtain as required that

$$t_{\alpha_{ik}}=\frac{\sum \operatorname{sgn}\left(\tilde{e}_{it-1}\right)\varepsilon_{kt}}{\omega_{ik}\sqrt{T}}+\alpha_{ik}\boldsymbol{\beta}'_i\frac{\sum \operatorname{sgn}\left(\tilde{e}_{it-1}\right)\boldsymbol{w}_{it-1}}{\omega_{ik}T^{1.5}}+o_p\left(1\right).$$

Proof of Corollary 1

The joint null distribution of $t_{\alpha_{ik}}$ follows with a CLT for md sequences, cf. Theorem 24.3 in Davidson (1994): finite moments of 4th order, and

$$\frac{1}{T}\sum \operatorname{sgn}^2\left(\tilde{e}_{t-1}\right)\boldsymbol{\varepsilon}_{it}\boldsymbol{\varepsilon}'_{it}\stackrel{p}{\to}\Omega,$$

see Equation (8) above.

Proof of Corollary 2

The suitably normalized matrices P_i and R_i are block-diagonal in the limit; see the proof of Proposition 1. Hence $P'_iR_iP_i$ has as first diagonal element the required quantity and the result follows using the properties of the Kronecker product.

References

Bai, J. and J. L. Carrion-i Silvestre (2013). Testing panel cointegration with unobservable dynamic common factors that are correlated with the regressors. *The Econometrics Journal 16*(2), 222–249.

Bai, J. and S. Ng (2004). A panic attack on unit roots and cointegration. *Econometrica 72*(4), 1127–1177.

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*(3), 267–283.

Banerjee, A., M. Marcellino, and C. Osbat (2004). Some cautions on the use of panel methods for integrated series of macroeconomic data. *The Econometrics Journal 7*(2), 322–340.

Banerjee, A., M. Marcellino, and C. Osbat (2005). Testing for PPP: Should we use panel methods? *Empirical Economics 30*(1), 77–91.

Beare, B. C. (2008). Unit root testing with unstable volatility. Technical Report 2008-W06, Nuffield College, University of Oxford.

Born, B. and M. Demetrescu (2015). Recursive adjustment for general deterministic components and improved tests for the cointegration rank. *Journal of Time Series Econometrics 7*(2), 143–179.

Boswijk, H. P. (2005). Adaptive testing for a unit root with nonstationary volatility. UvA-Econometrics Discussion Paper 07, Universiteit van Amsterdam.

Campbell, J. and R. Shiller (1987). Cointegration and tests of present value models. *Journal of Political Economy 95*(5), 1062–88.

Cavaliere, G., A. Rahbek, and A. R. Taylor (2010). Testing for co-integration in vector autoregressions with non-stationary volatility. *Journal of Econometrics 158*(1), 7–24.

Cavaliere, G., A. Rahbek, and R. Taylor (2012). Bootstrap determination of the co-integration rank in vector autoregressive models. *Econometrica 80*(4), 1721–1740.

Cavaliere, G., A. Rahbek, and R. Taylor (2014). Bootstrap determination of the co-integration rank in heteroskedastic VAR models. *Econometric Reviews 33*(5-6), 606–650.

Cavaliere, G. and A. M. R. Taylor (2008a). Bootstrap unit root tests for time series with nonstationary volatility. *Econometric Theory 24*(1), 43–71.

Cavaliere, G. and A. M. R. Taylor (2008b). Time-transformed unit root tests for models with non-stationary volatility. *Journal of Time Series Analysis 29*(2), 300–330.

Cavaliere, G. and A. M. R. Taylor (2009). Heteroskedastic time series with a unit root. *Econometric Theory 25*(5), 1228–1276.

Cavaliere, G., R. A. M. Taylor, and C. Trenkler (2013). Bootstrap co-integration rank testing: the role of deterministic variables and initial values in the bootstrap recursion. *Econometric Reviews 32*(7), 814–847.

Chang, Y. and C. M. Nguyen (2012). Residual based tests for cointegration in dependent panels. *Journal of Econometrics 167*(2), 504–520.

Chang, Y. and J. Y. Park (2002). On the asymptotics of ADF tests for unit roots. *Econometric Reviews 21*(4), 431–447.

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*(3), 903–935.

Christopeit, N. (2009). Weak Convergence of Nonlinear Transformations of Integrated Processes: The Multivariate Case. *Econometric Theory 25*(5), 1180–1207.

Clark, T. E. (2009). Is the Great Moderation over? An empirical analysis. *Economic Review 4*, 5–42.

Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985). A theory of the term structure of interest rates. *Econometrica 53*(2), 385–407.

Davidson, J. (1994). *Stochastic Limit Theory*. Oxford: Oxford University Press.

Demetrescu, M. and C. Hanck (2012). Unit root testing in heteroskedastic panels using the Cauchy estimator. *Journal of Business & Economic Statistics 30*(2), 256–264.

Demetrescu, M. and C. Hanck (2013). Nonlinear iv panel unit root testing under structural breaks in the error variance. *Statistical Papers 54*(4), 1043–1066.

Demetrescu, M. and C. Hanck (201x). Robust inference for near-unit root processes with time-varying error variances. *Econometric Reviews*.

Demetrescu, M., C. Hanck, and A. I. Tarcolea (2014). IV-based cointegration testing in dependent panels with time-varying variance. *Journal of Time Series Analysis 35*(5), 393–406.

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*(5), 647–663.

Doh, T. (2013). Long-run risks in the term structure of interest rates: Estimation. *Journal of Applied Econometrics 28*(3), 478–497.

Elliott, G., T. J. Rothenberg, and J. H. Stock (1996). Efficient tests for an autoregressive unit root. *Econometrica 64*(4), 813–836.

Engle, R. F. and C. W. Granger (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica 55*(2), 251–76.

Gengenbach, C., F. C. Palm, and J.-P. Urbain (2006). Cointegration testing in panels with common factors. *Oxford Bulletin of Economics and Statistics 68*(S1), 683–719.

Gengenbach, C., J.-P. Urbain, and J. Westerlund (2008). Error correction testing in panels with global stochastic trends. *METEOR Research Memorandum 08*, 1–51.

Gregory, A. W., A. A. Haug, and N. Lomuto (2004). Mixed signals among tests for cointegration. *Journal of Applied Econometrics 19*(1), 89–98.

Groen, J. and F. Kleibergen (2003). Likelihood-based cointegration analysis in panels of vector error correction models. *Journal of Business & Economic Statistics 21*(2), 295–317.

Hall, A. D., H. M. Anderson, and C. W. J. Granger (1992). A cointegration analysis of treasury bill yields. *The Review of Economics and Statistics 74*(1), 116–126.

Hanck, C. (2013). An intersection test for panel unit roots. *Econometric Reviews 32*(2), 183–203.

Hartung, J. (1999). A note on combining dependent tests of significance. *Biometrical Journal 41*(7), 849–855.

Hayashi, F. (2000). *Econometrics*. Princeton: Princeton University Press.

Hecq, A., F. C. Palm, and J. P. Urbain (2002). Separation, weak exogeneity, and PT decomposition in cointegrated VAR systems with common features. *Econometric Reviews 21*(3), 273–307.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control 12*(2–3), 231–254.

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

Jondeau, E. and R. Ricart (1999). The expectations hypothesis of the term structure: tests on US, German, French, and UK euro-rates. *Journal of International Money and Finance 18*(5), 725–750.

Lanne, M. (2000). Near unit roots, cointegration, and the term structure of interest rates. *Journal of Applied Econometrics 15*(5), 513–529.

Leybourne, S., T. H. Kim, and P. Newbold (2005). Examination of some more powerful modifications of the Dickey-Fuller test. *Journal of Time Series Analysis 26*(3), 355–369.

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).

OECD (2015a). Long-term interest rates (indicator), doi: 10.1787/662d712c-enbc. Technical report.

OECD (2015b). Short-term interest rates (indicator), doi: 10.1787/2cc37d77-en. Technical report.

Pesavento, E. (2004). Analytical evaluation of the power of tests for the absence of cointegration. *Journal of Econometrics 122*(2), 349–384.

Saikkonen, P. (1992). Estimation and testing of cointegrated systems by an autoregressive approximation. *Econometric Theory 8*(1), 1–27.

Saikkonen, P. and H. Lütkepohl (1999). Local power of likelihood ratio tests for the cointegrating rank of a VAR process. *Econometric Theory 15*(1), 50–78.

Sarkar, S. K. (1998). Probability inequalities for ordered MTP_2 random variables: A proof of the Simes conjecture. *The Annals of Statistics 26*(2), 494–504.

Sensier, M. and D. van Dijk (2004). Testing for volatility changes in U.S. macroeconomic time series. *The Review of Economics and Statistics 86*(3), 833–839.

Shea, G. S. (1992). Benchmarking the expectations hypothesis of the interest-rate term structure: An analysis of cointegration vectors. *Journal of Business & Economic Statistics 10*(3), 347–366.

Shin, D. W. and S. Kang (2006). An Instrumental Variable Approach for Panel Unit Root Tests under Cross-Sectional Dependence. *Journal of Econometrics 134*(1), 215–234.

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika 73*(3), 751–754.

Smeekes, S. and J.-P. Urbain (2014a). A multivariate invariance principle for modified wild bootstrap methods with an application to unit root testing. *Maastricht University GSBE Research Memorandum* (008).

Smeekes, S. and J.-P. Urbain (2014b). On the applicability of the sieve bootstrap in time series panels. *Oxford Bulletin of Economics and Statistics 76*(1), 139–151.

So, B. S. and D. W. Shin (1999). Cauchy estimators for autoregressive processes with applications to unit root tests and confidence intervals. *Econometric Theory 15*(2), 165–176.

Stock, J. H. and M. W. Watson (2002). Has the business cycle changed and why? *NBER Macroeconomics Annual 17*(1), 159–218.

Swensen, A. R. (2006). Bootstrap algorithms for testing and determining the cointegration rank in VAR models. *Econometrica 74*(6), 1699–1714.

Taylor, M. P. (1992). Modelling the yield curve. *The Economic Journal 102*(412), 524–537.

Wagner, M. and J. Hlouskova (2010). The performance of panel cointegration methods: Results from a large scale simulation study. *Econometric Reviews 29*(2), 182–223.

Wang, Q. and P. Phillips (2009). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory 25*(3), 710–738.

Westerlund, J. (2008). Panel cointegration tests of the Fisher effect. *Journal of Applied Econometrics 23*(2), 193–233.

Westerlund, J. (2014). Heteroscedasticity robust panel unit root tests. *Journal of Business & Economic Statistics 32*(1), 112–135.

Westerlund, J. (2015). The power of PANIC. *Journal of Econometrics 185*(2), 495–509.

Zivot, E. (2000). The power of single equation tests for cointegration when the cointegrating vector is prespecified. *Econometric Theory 16*(3), 407–439.