Combining Non-Cointegration Tests*

Christian Bayer[†]
Università Commerciale L. Bocconi

Christoph Hanck[‡]
Universiteit Maastricht

First Version: May 2008

March 1, 2009

Preliminary.

Abstract

The local asymptotic power of many popular non-cointegration tests has recently been shown to depend on a certain nuisance parameter. Depending on the value of that parameter, different tests perform best. This paper suggests combination procedures with the aim of providing meta tests that maintain high power across the range of the nuisance parameter. The local asymptotic power of the new meta tests is in general almost as high as that of the more powerful of the underlying tests. When the underlying tests have similar power, the meta tests are even more powerful than the best underlying test. At the same time, our new meta tests avoid the arbitrary decision which test to use if single test results conflict. Moreover it avoids the size distortion inherent in separately applying multiple tests for cointegration to the same data set. We apply our test to 161 data sets from published cointegration studies. There, in one third of all cases single tests give conflicting results whereas our meta tests provide an unambiguous test decision.

Keywords: Cointegration, Meta Test, Multiple Testing
JEL-Codes: C12, C22

*This is a substantially extended version of a manuscript that previously circulated under the title 'Is Double Trouble? How to Combine Cointegration Tests'. Part of the research was carried out when the authors were at Technische Universität Dortmund with support by DFG under Sonderforschungsbereich 475. We are grateful to Carsten Burhop, Carlo Favero, Walter Krämer and Franz Palm as well as conference participants in Neuchâtel and Bonn for valuable comments and suggestions.

[†]IGIER, Via Salasco 5, 20136 Milano, Italy. Tel.: +39 (0)2 5836 3386. email: christian.bayer@unibocconi.it.

[‡]Department of Quantitative Economics, Tongerseweg 53, 6211LM Maastricht, Netherlands. Tel.: +31 (0)43 3883815. email: c.hanck@ke.unimaas.nl.

1 Introduction

Testing for cointegration has become one of the standard tools in applied economic research. Various tests have been suggested for this purpose, most of which are implemented in standard econometric packages and hence are easily available nowadays. Well-known examples include the residual-based test of Engle and Granger (1987), or the system-based tests of Johansen (1988). Error-Correction-based tests have been suggested by Boswijk (1994) and Banerjee *et al.* (1998), while Breitung (2001) covers the nonlinear case, to name just a few. This regularly forces the applied researcher to select from the test decisions of the various applicable procedures. This choice is difficult because, as discussed in e.g. Elliott *et al.* (2005), there exists no uniformly most powerful test, even asymptotically. Often one test rejects the null hypothesis whereas another test does not, making it unclear how to interpret test outcomes then. More generally speaking, the p-values of different tests are typically not perfectly correlated (Gregory *et al.*, 2004).

This imperfect correlation makes it problematic to choose, for example, a testing strategy that relies on the test that achieves the smallest p-value. Such strategy will not control the probability of rejecting a true null hypothesis at some chosen level α because it ignores the multiple testing nature of the problem. Concretely, using the test with the smallest p-value will lead to an oversized test.

The imperfect correlation of different test statistics reflects the fact that the tests are not equivalent, focussing on different statistical characterizations of non-cointegration. This also has implications for their (local) asymptotic power under the alternative. Specifically, Pesavento (2004) shows that the relative power of cointegration tests depends crucially on the squared long-run correlations of error terms driving the variables of the analyzed system. That is, the power ranking of the tests varies by the value of that unknown nuisance parameter.

This suggests that appropriate combinations of tests for no cointegration potentially yield a more robust power performance, and possibly even power gains, relative to applying only a single test. Based on the above-mentioned single cointegration tests, the present paper develops such combination tests. In particular, we propose to combine test statistics in the spirit of Fisher's (1932) famous test. We derive the asymptotic null distribution of our Fisher-type combination test for correlated cointegration test statistics and its local asymptotic power, exploiting Pesavento's (2004) results. Besides successfully tackling the above-mentioned multiple testing nature inherent in combining different test statistics, the combined test indeed enjoys a robust power performance over the range of the squared long-run error correlation. Moreover, we explore a number of alternative combination procedures. For example, Harvey *et al.* (2009) propose a Union-of-Rejections procedure to robustify unit root tests against uncertainty over the initial condition. We generalize their idea and apply the generalized Union-of-Rejections approach to the present testing problem.

Our Fisher-type test turns out to perform very well. It follows closely the power envelope of the underlying single tests, and even exceeds it when the single tests have similar power. In contrast, the Union-of-Rejections procedure is most useful when the underlying tests have strongly different power, in that its power is always close to that of the better underlying test.

Of course, the asymptotic distributions derived here are, as usual, only approximations to the generally analytically intractable finite-sample distributions. Those may or may not be accurate. We therefore additionally propose bootstrap analogs of our combination tests. Specifically, we build on Swensen's (2006) recent bootstrap scheme for cointegrated vector autoregressions.

We conduct extensive finite-sample experiments to investigate the performance our the asymptotic and bootstrap combination tests. The local asymptotic results correctly predict the finite-sample performance. Both the asymptotic and the bootstrap versions successfully control the level α of the test and are at the same time powerful. The bootstrap versions appear to converge to the nominal size somewhat more quickly.

We point out that the above multiple testing problem is pervasive in empirical work and not restricted to testing for cointegration. The meta testing-based solution developed here is rather general and could hence be adopted to other testing problems for which several (imperfectly correlated) tests have been developed. Examples include testing for unit roots or heteroscedasticity.

To check the practical relevance of our proposed tests, we revisit the set of published cointegration studies that Gregory *et al.* (2004) examined for "mixed signals" among cointegration tests, i.e. conflicting test results. Among other things we find that in one third of all cases single tests give conflicting results. In these cases our meta tests are particularly useful. They provide an unambiguous test decision and therefore are a solution to the "mixed signals" problem.

The remainder of this paper is organized as follows: Section 2 provides the setup for the non-cointegration tests. Section 3 derives our combination tests. Section 4 presents local asymptotic power results. Section 5 is devoted to the bootstrap analogs. Section 6 reports Monte Carlo results. Section 7 provides the empirical application of our combination tests. Section 8 concludes. An appendix reports additional results.

The notation to be used is standard. Weak convergence, convergence in probability and in distribution are denoted by $\Rightarrow$, $\to_p$ and $\to_d$. Limits of integration are zero and 1, $\int = \int_0^1$, unless specified otherwise. $[a]$ is the integer part of a. Vectors and matrices are given in boldface. Integrals such as $\int_0^1 \boldsymbol{W}(s)\boldsymbol{W}(s)' \, \mathrm{d}s$ will often be abbreviated as $\int \boldsymbol{W}\boldsymbol{W}'$. When a defines b, we write $b := a$ or $a =: b$.

2 Setup

2.1 Model

We work with the setup studied by Pesavento (2004). Let $\boldsymbol{z}_t := (z_{1t}, \ldots, z_{Kt})' \in \mathbb{R}^K$ be a vector of stochastic variables integrated of order one, $I(1)$. Partition $\boldsymbol{z}_t$ as $\boldsymbol{z}_t = (\boldsymbol{x}_t', \, y_t)'$. The following equations give the model.

$$\begin{aligned} \Delta \boldsymbol{x}_t &= \boldsymbol{\tau}_1 + \boldsymbol{v}_{1t} && \text{(1a)} \\ y_t &= (\mu_2 - \boldsymbol{\theta}'\boldsymbol{\mu}_1) + (\tau_2 - \boldsymbol{\theta}'\boldsymbol{\tau}_1)t + \boldsymbol{\theta}'\boldsymbol{x}_t + u_t && \text{(1b)} \\ u_t &= \rho u_{t-1} + v_{2t} && \text{(1c)} \end{aligned}$$

We make the following assumption on the error vector $\boldsymbol{v}_t := (\boldsymbol{v}_{1t}', v_{2t})'$ from eqs. (1a) and (1c).

Assumption 1. $\{\boldsymbol{v}_t\}$ satisfies a Functional Central Limit Theorem (FCLT), i.e. $T^{-1/2}\sum_{t=1}^{[\lambda T]} \boldsymbol{v}_t \Rightarrow \boldsymbol{\Omega}^{1/2}\boldsymbol{W}(\lambda)$, with $\boldsymbol{\Omega}$ the long-run covariance matrix of $\boldsymbol{v}_t$.

Equation (1a) defines the dynamics of the regressors, while eqs. (1b) and (1c) describe the (single potential) cointegrating relationship. The coefficients $\boldsymbol{\mu} := (\boldsymbol{\mu}_1', \mu_2)'$ and $\boldsymbol{\tau} := (\boldsymbol{\tau}_1', \tau_2)'$ determine the specification of the deterministic components of the model, see Definition 1 below and Pesavento (2004) for details. The vector $\boldsymbol{z}_t$ is said to be cointegrated if there exists at least one $\tilde{\boldsymbol{\theta}} \in \mathbb{R}^K$, $\tilde{\boldsymbol{\theta}} := (-\boldsymbol{\theta}', 1)'$, $\boldsymbol{\theta} \neq \boldsymbol{0}$, such that the stochastic part of $\tilde{\boldsymbol{\theta}}'\boldsymbol{z}_t$ is a stationary $I(0)$ process. In terms of model (1), cointegration therefore obtains if $\rho < 1$. We test the following null hypothesis:

$\mathcal{H}_0$: There exists no cointegrating relationship among the variables in $\boldsymbol{z}_t$.

against the alternative hypothesis

$\mathcal{H}_1$: There exists a $\tilde{\boldsymbol{\theta}} \neq \boldsymbol{0}$ such that the stochastic part of $\tilde{\boldsymbol{\theta}}'\boldsymbol{z}_t$ is $I(0)$.

The literature has suggested various tests to discriminate between $\mathcal{H}_0$ and $\mathcal{H}_1$. We consider the residual-based test of Engle and Granger (1987), a system-based test of Johansen (1988), as well as the error-correction-based tests of Boswijk (1994) and Banerjee *et al.* (1998).

Pesavento (2004) derives the local asymptotic power of these tests. She shows that, under model (1), their power only depends on the local-to-unity parameter $c := T(\rho - 1)$ and the squared correlations of the elements of $\boldsymbol{v}_{1t}$ with v_{2t}. More precisely, partition $\boldsymbol{\Omega}$ conformably with $(\boldsymbol{x}_t', y_t)'$,

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\omega}_{12} \\ \boldsymbol{\omega}_{12}' & \omega_{22} \end{pmatrix}$$

We assume that there are no cointegrating relationships among the variables in $\boldsymbol{x}_t$, i.e.

Assumption 2. $\boldsymbol{\Omega}_{11}$ is invertible.

We can then define the squared correlation as $R^2 := \boldsymbol{\delta}'\boldsymbol{\delta}$, where $\boldsymbol{\delta} := \boldsymbol{\Omega}_{11}^{-1/2}\boldsymbol{\omega}_{12}\omega_{22}^{-1/2}$. Moreover, it is useful to partition $\boldsymbol{W}(\lambda) := [\boldsymbol{W}_1(\lambda)'\ W_2(\lambda)]'$. Define the Ornstein-Uhlenbeck process $J_{12c}(\lambda) := W_{12}(\lambda) + c\int_0^\lambda \mathrm{e}^{(\lambda - s)c} W_{12}(s)\,\mathrm{d}s$, with $W_{12}(\lambda) := \bar{\boldsymbol{\delta}}'\boldsymbol{W}_1(\lambda) + W_2(\lambda)$, where $\bar{\boldsymbol{\delta}}$ is such that $\bar{\boldsymbol{\delta}}'\bar{\boldsymbol{\delta}} = \frac{R^2}{1-R^2}$. Furthermore, we have

Definition 1. *Depending on the assumptions made about the deterministic components, we distinguish the following cases.*

(*i*) $\boldsymbol{W}^d(\lambda) := \boldsymbol{W}(\lambda)$ *and* $J_{12c}^d(\lambda) = J_{12c}(\lambda)$ *if* $\mu_2 - \boldsymbol{\theta}'\boldsymbol{\mu}_1 = 0$, $\boldsymbol{\tau} = \mathbf{0}$ *and no deterministic terms are included in the regressions. We refer to this as case* (*i*).

(*ii*) $\boldsymbol{W}^d(\lambda) := \boldsymbol{W}(\lambda) - \int \boldsymbol{W}(s)\,\mathrm{d}s$ *and* $J_{12c}^d(\lambda) = J_{12c}(\lambda) - \int J_{12c}(s)\,\mathrm{d}s$ *if* $\boldsymbol{\tau} = \mathbf{0}$ *and a constant is included in the regressions. We refer to this as case* (*ii*).

(*iii*) $\boldsymbol{W}^d(\lambda) := \boldsymbol{W}(\lambda) - (4-6\lambda)\int \boldsymbol{W}(s)\,\mathrm{d}s - (12\lambda - 6)\int s\boldsymbol{W}(s)\,\mathrm{d}s$ *and* $J_{12c}^d(\lambda) = J_{12c}(\lambda) - (4 - 6\lambda)\int J_{12c}(s)\,\mathrm{d}s - (12\lambda-6)\int s\boldsymbol{W}(s)\,\mathrm{d}s$ *if there are no restrictions and a constant and trend are included in the regressions. We refer to this as case* (*iii*).

Also, $\boldsymbol{W}_c^d := [\boldsymbol{W}_1^{d\prime}(\lambda) \quad J_{12c}^d(\lambda)]'$ *and* $\boldsymbol{A}_c^d := \int \boldsymbol{W}_c^d \boldsymbol{W}_c^{d\prime}$.

2.2 Single Cointegration Tests

Engle and Granger (1987)

The Engle-Granger test tests the null of no cointegration against the alternative of at least one cointegrating relationship. Suppose we have observations $\boldsymbol{z}_0, \ldots, \boldsymbol{z}_T$. One computes the t-statistic t_γ^{ADF} on γ in the OLS regression

$$\Delta\hat{u}_t = \gamma\hat{u}_{t-1} + \sum_{p=1}^{P-1} \nu_p \Delta\hat{u}_{t-p} + \epsilon_t. \tag{2}$$

Here, $\hat{u}_t$ is the usual residual from a first stage OLS regression of y_t on $\boldsymbol{x}_t$ (and appropriate deterministic terms). The sum $\sum_{p=1}^{P-1} \nu_p \Delta\hat{u}_{t-p}$ captures residual serial correlation.[1] Proposition 1 summarizes the local asymptotic distribution derived by Pesavento (2004).

Proposition 1. *With the terms as in Definition 1, we have*

$$t_\gamma^{\mathrm{ADF}} \Rightarrow c\frac{[\boldsymbol{\eta}_c^{d\prime}\boldsymbol{A}_c^d\boldsymbol{\eta}_c^d]^{1/2}}{[\boldsymbol{\eta}_c^{d\prime}\boldsymbol{D}\boldsymbol{\eta}_c^d]^{1/2}} + \frac{\boldsymbol{\eta}_c^{d\prime}\int \boldsymbol{W}_c^d\,\mathrm{d}\widetilde{\boldsymbol{W}}'\boldsymbol{\eta}_c^d}{[\boldsymbol{\eta}_c^{d\prime}\boldsymbol{A}_c^d\boldsymbol{\eta}_c^d]^{1/2}[\boldsymbol{\eta}_c^{d\prime}\boldsymbol{D}\boldsymbol{\eta}_c^d]^{1/2}}$$

$$\begin{aligned} \text{where} \quad \boldsymbol{\eta}_c^d &:= \left[-\left(\int \boldsymbol{W}_1^{d\prime} J_{12c}^d\right)\left(\int \boldsymbol{W}_1^d \boldsymbol{W}_1^{d\prime}\right)^{-1} \quad 1\right]', \\ \widetilde{\boldsymbol{W}}(\lambda) &:= [\boldsymbol{W}_1'(\lambda) \quad W_{12}(\lambda)]', \\ \boldsymbol{D} &:= \begin{pmatrix} \boldsymbol{I} & \bar{\boldsymbol{\delta}} \\ \bar{\boldsymbol{\delta}}' & 1 + \bar{\boldsymbol{\delta}}'\bar{\boldsymbol{\delta}} \end{pmatrix} \end{aligned}$$

Johansen (1988)

The system-based tests of Johansen (1988) test the presence of h cointegrating relationships by estimating the Vector Error Correction Model (VECM)

$$\Delta\boldsymbol{z}_t = \boldsymbol{\Pi}\boldsymbol{z}_{t-1} + \sum_{p=1}^{P-1} \boldsymbol{\Gamma}_p \Delta\boldsymbol{z}_{t-p} + \boldsymbol{d}_t + \boldsymbol{\varepsilon}_t, \tag{3}$$

[1] Alternatively, one could control for serial correlation by the semiparametric approach of Phillips and Ouliaris (1990).

with $\boldsymbol{d}_t$ appropriate deterministic terms. We employ the $\lambda_{\max}$-test with test statistic

$$\lambda_{\max}(h) = -T \ln(1 - \hat{\pi}_{h+1}) . \tag{4}$$

Here, $\hat{\pi}_j$ denotes the jth largest solution to $|\pi S_{11} - S_{10} S_{00}^{-1} S_{01}| = 0$ (in Johansen's (1995) notation). In view of our $\mathcal{H}_0$, we consider $h = 0$ throughout.

Proposition 2. *With the terms as in Definition 1, we have*

$$\begin{aligned}\lambda_{\max} \quad \Rightarrow \quad \max \operatorname{eig} \Big\{ (\boldsymbol{A}_c^d)^{-1} \Big[\int \boldsymbol{W}_c^d \, \mathrm{d}\boldsymbol{W}' \int \mathrm{d}\boldsymbol{W} \boldsymbol{W}_c^{d\prime} + \int \boldsymbol{W}_c^d \, \mathrm{d}\boldsymbol{W}' \boldsymbol{G}_c' \\ + \boldsymbol{G}_c \Big(\int \boldsymbol{W}_c^d \, \mathrm{d}\boldsymbol{W}' \Big)' + \boldsymbol{G}_c \boldsymbol{G}_c' \Big] \Big\},\end{aligned}$$

where $\boldsymbol{G}_c := \int \boldsymbol{W}_c^d J_{12c} [\boldsymbol{0}' \;\; c]$.

Boswijk (1994) and Banerjee et al. *(1998)*

Banerjee *et al.* (1998) and Boswijk (1994) work with the conditional error correction representation of model (1). The equation to be estimated (by OLS) then becomes

$$\Delta y_t = d_t + \boldsymbol{\pi}_{0x}' \Delta \boldsymbol{x}_t + \varphi_0 y_{t-1} + \boldsymbol{\varphi}_1' \boldsymbol{x}_{t-1} + \sum_{p=1}^{P} (\boldsymbol{\pi}_{px}' \Delta \boldsymbol{x}_{t-p} + \pi_{py} \Delta y_{t-p}) + \epsilon_t, \tag{5}$$

with P chosen such that ϵ_t is approximately white noise. Banerjee *et al.*'s test statistic t_γ^{ECR} is the standard t-ratio for the null hypothesis $\mathcal{H}_0 : \varphi_0 = 0$, whereas Boswijk's $\hat{F}$ is the usual Wald statistic for $\mathcal{H}_0 : (\varphi_0, \; \boldsymbol{\varphi}_1')' = \boldsymbol{0}$.

Proposition 3. *With the terms as in Definition 1, we have*

$$\hat{F} \Rightarrow c^2 \int J_{12c}^{d2} + 2c \int J_{12c}^d \, \mathrm{d}W_2 + \int \boldsymbol{W}_c^{d\prime} \, \mathrm{d}W_2 (\boldsymbol{A}_c^d)^{-1} \int \boldsymbol{W}_c^d \, \mathrm{d}W_2$$

$$\begin{aligned} t_\gamma^{\mathrm{ECR}} \quad \Rightarrow \quad & c \Big[\int J_{12c}^{d2} - \int \boldsymbol{W}_1^{d\prime} J_{12c}^d \Big(\int \boldsymbol{W}_1^d \boldsymbol{W}_1^{d\prime} \Big)^{-1} \int \boldsymbol{W}_1^d J_{12c}^d \Big]^{1/2} \\ & + \frac{\int J_{12c}^d \, \mathrm{d}W_2 - \int \boldsymbol{W}_1^{d\prime} J_{12c}^d (\int \boldsymbol{W}_1^d \boldsymbol{W}_1^{d\prime})^{-1} \int \boldsymbol{W}_1^d \, \mathrm{d}W_2}{[\int J_{12c}^{d2} - \int \boldsymbol{W}_1^{d\prime} J_{12c}^d (\int \boldsymbol{W}_1^d \boldsymbol{W}_1^{d\prime})^{-1} \int \boldsymbol{W}_1^d J_{12c}^d]^{1/2}} \end{aligned}$$

For $c = 0$, all quantities in Props. 1-3 reduce to the well-known nuisance-parameter free null distributions. More importantly, all limiting functionals are driven by the same Brownian Motions $\boldsymbol{W}(\lambda)$, such that the propositions allow us to consider the *joint* distribution of the test statistics.

3 Combination Tests

Gregory *et al.* (2004) show that, under $\mathcal{H}_0$, the above test statistics are only weakly correlated, even asymptotically. Also, as pointed out above, Pesavento (2004) demonstrates that the tests differ in their power in different parts of the (c-R^2)-parameter space. In particular, *any* test is the most powerful one in *some* part of the parameter space. As argued in the Introduction, this implies that a more robust, and possibly even more powerful, combination test can in principle be achieved.

Let t_i be the test statistic of cointegration test $i = 1, \dots, N$. We define $\xi_i := t_i$ if test i rejects for large values and let $-\xi_i := t_i$ if test i rejects for small values. Define Ξ_i as one minus the corresponding (typically nonstandard) asymptotic null distribution function, i.e. $\Xi_i(x) := \mathrm{P}(\xi_i \geqslant x)$. The p-values of the tests are then given by $p_i := \Xi_i(\xi_i)$.

3.1 A Fisher-type test

To reach a joint test decision from the different ξ_i, we require a suitable aggregator. One such aggregator is given Fisher's (1932) famous χ^2 test. The following Proposition follows at once from the Continuous Mapping Theorem (CMT).

Proposition 4. *Let $\mathcal{I}$ the index set of the combined single ξ_i. Consider the Fisher-type test statistic*

$$\tilde{\chi}^2_{\mathcal{I}} := -2 \sum_{i \in \mathcal{I}} \ln(p_i). \tag{6}$$

As $T \to \infty$, (a) $\tilde{\chi}^2_{\mathcal{I}} \to_d \mathcal{F}_{\mathcal{I}}$ under $\mathcal{H}_0$, with $\mathcal{F}_{\mathcal{I}}$ some random variable. Further, (b) $\tilde{\chi}^2_{\mathcal{I}} \to \infty$ under $\mathcal{H}_1$ if at least one of the underlying tests is consistent, i.e satisfies $p_i \to_p 0$ under $\mathcal{H}_1$.

Part (a) of Proposition 4 states that $\tilde{\chi}^2_{\mathcal{I}}$ has a well-defined asymptotic null distribution, call it $F_{\mathcal{F}_{\mathcal{I}}}$. The index-set notation $\mathcal{I}$ serves to emphasize that the distribution of the Fisher test depends on which and how many tests are combined. Part (b) establishes the consistency of a test based on $\tilde{\chi}^2_{\mathcal{I}}$. Of course we cannot invoke the conventional $\chi^2(2|\mathcal{I}|)$ (with $|\mathcal{I}|$ the cardinality of $\mathcal{I}$) null distribution for $\tilde{\chi}^2_{\mathcal{I}}$, as independence of the aggregated ξ_i is necessary for this result. However, focussing on the underlying tests from Propositions 1-3, we can straightforwardly infer and simulate their joint distribution. The aggregator $\tilde{\chi}^2_{\mathcal{I}}$ is a continuous function of the t_i, whose null distribution $F_{\mathcal{F}_{\mathcal{I}}}$ can therefore be analogously derived by simulation of the functional (6). Table 1 reports critical values $F^{-1}_{\mathcal{F}_{\mathcal{I}}}(1-\alpha)$ for combinations of the above-mentioned tests, obtained from 100,000 draws from the distributions $F_{\mathcal{F}_{\mathcal{I}}}$. (From Prop. 4, reject if $\tilde{\chi}^2_{\mathcal{I}} > F^{-1}_{\mathcal{F}_{\mathcal{I}}}(1-\alpha)$.) We approximate the Wiener processes with suitably normalized Gaussian random walks of length $T = 1000$ and tabulate 5% critical values for several combinations likely to be relevant in practice (see Appendix A for other levels). Moreover, since the distributions of the underlying cointegration tests depend on $K - 1$ (reported up to 11) as well as the maintained deterministic specification (i)-(iii), that of $\tilde{\chi}^2_{\mathcal{I}}$ will not only depend on $\mathcal{I}$ but also on $K - 1$ and the maintained case.

Table 1: Critical Values for the $\tilde{\chi}^2_{\mathcal{I}}$-tests.

	case											
$K-1$	(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)
$\alpha = 0.05$												
	t_γ^{ADF} and $\lambda_{\max}$			$\hat{F}$ and $\lambda_{\max}$			$\hat{F}$ and t_γ^{ECR}			$\hat{F}$ and t_γ^{ADF}		
1	11.071	11.229	11.269	11.071	11.090	11.068	11.606	11.803	11.862	10.890	11.298	11.507
2	10.838	10.895	10.858	10.701	10.715	10.654	11.556	11.716	11.795	10.794	11.051	11.237
3	10.640	10.637	10.711	10.453	10.459	10.461	11.554	11.683	11.731	10.688	10.880	11.087
4	10.516	10.576	10.532	10.299	10.324	10.318	11.491	11.611	11.696	10.644	10.780	11.000
5	10.406	10.419	10.448	10.237	10.187	10.188	11.478	11.621	11.639	10.635	10.701	10.896
6	10.312	10.352	10.311	10.115	10.167	10.166	11.473	11.611	11.597	10.556	10.670	10.820
7	10.218	10.295	10.222	10.023	10.055	10.033	11.492	11.577	11.621	10.594	10.715	10.813
8	10.185	10.181	10.189	10.041	9.999	10.014	11.511	11.545	11.624	10.591	10.658	10.800
9	10.162	10.154	10.164	10.000	9.978	9.996	11.488	11.590	11.633	10.561	10.738	10.733
10	10.079	10.109	10.070	9.926	9.889	9.870	11.491	11.504	11.565	10.556	10.629	10.703
11	10.057	10.059	10.134	9.928	9.928	9.946	11.450	11.528	11.542	10.548	10.641	10.667
	$\hat{F}$, $\lambda_{\max}$ and t_γ^{ADF}			$\hat{F}$, $\lambda_{\max}$ and t_γ^{ECR}			$\hat{F}$, $\lambda_{\max}$, t_γ^{ADF}, t_γ^{ECR}					
1	16.037	16.363	16.582	16.287	16.572	16.633	21.352	21.931	22.215			
2	15.526	15.732	15.856	15.827	15.927	15.965	20.776	21.106	21.342			
3	15.186	15.294	15.471	15.440	15.512	15.620	20.237	20.486	20.788			
4	14.934	15.025	15.173	15.184	15.291	15.407	19.951	20.143	20.440			
5	14.720	14.825	14.990	15.045	15.092	15.260	19.747	19.888	20.170			
6	14.578	14.685	14.833	14.924	15.056	15.155	19.564	19.761	19.934			
7	14.472	14.612	14.632	14.852	14.964	14.946	19.471	19.688	19.722			
8	14.460	14.427	14.595	14.823	14.825	14.941	19.471	19.447	19.678			
9	14.332	14.405	14.496	14.766	14.801	14.872	19.365	19.492	19.582			
10	14.321	14.322	14.301	14.717	14.733	14.775	19.268	19.365	19.398			
11	14.230	14.300	14.357	14.696	14.773	14.824	19.151	19.345	19.404			

Critical values for combination tests based on $\tilde{\chi}^2$. t_γ^{ADF} is from Engle and Granger (1987), $\lambda_{\max}$ from Johansen (1988), $\hat{F}$ from Boswijk (1994) and t_γ^{ECR} from Banerjee *et al.* (1998).

We find that, for different combinations, the (5%-)critical values cluster around 11 for $|\mathcal{I}| = 2$, and around 15 for $|\mathcal{I}| = 3$. There is little variation across cases. The critical values fall moderately in $K-1$. It is instructive to compare the critical values to those of the $\chi^2(2|\mathcal{I}|)$ distribution. The 5%-critical value is 9.487 for $|\mathcal{I}| = 2$, and 12.591 for $|\mathcal{I}| = 3$. The critical values in Table 1 are uniformly larger. This reflects the fact that the t_i are generally positively correlated, such that a larger critical value is necessary to construct level-α tests based on (6).

Remark 1. The aggregator (6) is only one of many possible choices. Among others, we experiment with an inverse-normal type approach to aggregating p-values, defined by $1/\sqrt{|\mathcal{I}|}\sum_{i\in\mathcal{I}}\Phi^{-1}(p_i)$, where Φ is the distribution function of the standard normal distribution. Its performance was however slightly inferior to the one of the $\tilde{\chi}^2_{\mathcal{I}}$ test, to be reported below. We therefore waive to report detailed results for $1/\sqrt{|\mathcal{I}|}\sum_{i\in\mathcal{I}}\Phi^{-1}(p_i)$, which are available upon request.

Remark 2. We also consider a minimum p-value test, given by $\min_{i\in\mathcal{I}} p_i$. This test can be understood as a direct fix to the 'naive' testing strategy that rejects whenever one of the conducted single tests rejects. The critical values of the distribution of $\min_{i\in\mathcal{I}} p_i$ yield the marginal significance level $\alpha' < \alpha$ at which one needs to test to avoid the oversizedness of the 'naive' approach.

Appendix A provides selected correction factors for the minimum p-value test.

3.2 A Union-of-Rejections test

The latter minimum p-value approach is similar to a recent proposal of Harvey *et al.* (2009), who develop a 'Union-of-Rejections' (UR) approach to combine standard Dickey-Fuller and GLS-demeaned unit root tests. The UR test also rejects whenever one of the two tests rejects, with however a suitable adjustment of the critical values in order to ensure a level-α test. This provides a more robust test as the two single tests are relatively more powerful when the initial condition of the time series is large (small). This situation is analogous to the present one, in that R^2 determines the relative power of the single cointegration tests. We now use and extend the UR approach to the case of cointegration testing considered here.

Denote the single level-α critical value corresponding to the test statistic ξ_i as $cv_{i,\alpha}$. The 'naive' Union-of-Rejections test statistic for $|\mathcal{I}| = 2$ can then be written as

$$UR^{\text{naive}}(\xi_1, \xi_2) := \xi_1 \mathbb{I}\{\xi_1 > cv_{1,\alpha}\} + \xi_2 \mathbb{I}\{\xi_1 \leqslant cv_{1,\alpha}\}, \tag{7}$$

with $\mathbb{I}\{A\}$ the indicator function of event A. The decision rule would be to reject $\mathcal{H}_0$ of non-cointegration if $UR^{\text{naive}}(\xi_1, \xi_2) = \xi_1$ and $UR(\xi_1, \xi_2) > cv_{1,\alpha}$, or if $UR^{\text{naive}}(\xi_1, \xi_2) = \xi_2$ and $UR^{\text{naive}}(\xi_1, \xi_2) > cv_{2,\alpha}$. Of course, the test (7) does not control size.[2] Harvey *et al.* (2009) therefore introduce a scaling constant ψ to modify (7) as follows.

$$UR_{\psi}(\xi_1, \xi_2) := \xi_1 \mathbb{I}\{\xi_1 > \psi cv_{1,\alpha}\} + \xi_2 \mathbb{I}\{\xi_1 \leqslant \psi cv_{1,\alpha}\}, \tag{8}$$

One rejects if $UR_{\psi}(\xi_1, \xi_2) = \xi_1$ and $UR_{\psi}(\xi_1, \xi_2) > \psi cv_{1,\alpha}$ or if $UR_{\psi}(\xi_1, \xi_2) = \xi_2$ and $UR_{\psi}(\xi_1, \xi_2) > \psi cv_{2,\alpha}$. The scaling constant ψ is unique and to be chosen so that $\mathrm{P}(\bigcup_{i=1}^{2} \xi_i > \psi cv_{i,\alpha}) = \alpha$.

However, there is no need to apply the *same* ψ to both critical values $cv_{i,\alpha}$. In fact, there exists a continuum of tuples of scaling constants so as to obtain a level-α Union-of-Rejections test. Define the interval $\mathcal{C} := \mathbb{R} \cap [1, \infty)$ and let $\boldsymbol{\psi} := (\psi_1, \psi_2) \in \mathcal{C} \times \mathcal{C} =: \mathcal{C}^2$. The UR statistic then becomes

$$UR_{\psi_1,\psi_2}(\xi_1, \xi_2) := \xi_1 \mathbb{I}\{\xi_1 > \psi_1 cv_{1,\alpha}\} + \xi_2 \mathbb{I}\{\xi_1 \leqslant \psi_1 cv_{1,\alpha}\}, \tag{9}$$

with a rejection being recorded if $UR_{\psi_1,\psi_2}(\xi_1, \xi_2) = \xi_1$ and $UR_{\psi_1,\psi_2}(\xi_1, \xi_2) > \psi_1 cv_{1,\alpha}$ or if $UR_{\psi_1,\psi_2}(\xi_1, \xi_2) = \xi_2$ and $UR_{\psi_1,\psi_2}(\xi_1, \xi_2) > \psi_2 cv_{2,\alpha}$. Again, the admissible tuples $\boldsymbol{\psi} \in \mathcal{C}^2$ are implicitly defined by

$$\mathrm{P}\left(\bigcup_{i=1}^{2} \xi_i > \psi_i cv_{i,\alpha}\right) = \alpha. \tag{10}$$

[2] Note that $\mathrm{E}\,\mathbb{I}\{\xi_i > cv_{i,\alpha}\} = \mathrm{P}(\xi_i > cv_{i,\alpha})$ gives the rejection probability of test i. Under $\mathcal{H}_0$, $\mathrm{E}\,\mathbb{I}\{\xi_i > cv_{i,\alpha}\} = \alpha$. The size of $UR^{\text{naive}}(\xi_1, \xi_2)$ therefore equals $\mathrm{P}(\bigcup_{i=1}^{2} \xi_i > cv_{i,\alpha}) = \mathrm{P}(\xi_1 > cv_{1,\alpha}) + \mathrm{P}(\xi_2 > cv_{2,\alpha}) - \mathrm{P}(\bigcap_{i=1}^{2} \xi_i > cv_{i,\alpha}) = 2\alpha - \mathrm{P}(\bigcap_{i=1}^{2} \xi_i > cv_{i,\alpha}) \geqslant \alpha$, since $\mathrm{P}(\bigcap_{i=1}^{2} \xi_i > cv_{i,\alpha}) \leqslant \mathrm{P}(\xi_i > cv_{i,\alpha}) = \alpha$.

Table 2: Correction Factors for the UR_{ψ_1,ψ_2} test

		t_γ^{ADF} and $\lambda_{\max}$			$\hat{F}$ and $\lambda_{\max}$			$\hat{F}$ and t_γ^{ECR}		
$K-1$	case	(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)
		t_γ^{ADF}			$\hat{F}$			$\hat{F}$		
1		1.065	1.050	1.043	1.128	1.104	1.093	1.077	1.042	1.032
2		1.058	1.052	1.044	1.131	1.110	1.095	1.075	1.052	1.038
3		1.055	1.049	1.046	1.122	1.104	1.096	1.070	1.053	1.038
4		1.051	1.045	1.042	1.107	1.099	1.090	1.057	1.053	1.043
5		1.048	1.045	1.041	1.103	1.094	1.088	1.058	1.049	1.043
6		1.046	1.044	1.040	1.096	1.091	1.085	1.060	1.051	1.044
7		1.045	1.042	1.035	1.092	1.082	1.082	1.056	1.055	1.045
8		1.042	1.041	1.039	1.089	1.080	1.081	1.050	1.044	1.044
9		1.040	1.038	1.039	1.085	1.081	1.078	1.049	1.047	1.044
10		1.039	1.035	1.037	1.079	1.008	1.075	1.046	1.041	1.043
11		1.038	1.037	1.035	1.072	1.076	1.071	1.047	1.045	1.041
		$\lambda_{\max}$			$\lambda_{\max}$			t_γ^{ECR}		
1		1.100	1.077	1.065	1.101	1.083	1.070	1.049	1.022	1.018
2		1.080	1.076	1.068	1.084	1.082	1.075	1.046	1.028	1.023
3		1.074	1.063	1.064	1.075	1.067	1.068	1.046	1.033	1.023
4		1.066	1.059	1.056	1.071	1.063	1.061	1.042	1.033	1.028
5		1.061	1.055	1.053	1.063	1.058	1.055	1.040	1.032	1.029
6		1.052	1.051	1.052	1.056	1.052	1.054	1.041	1.034	1.028
7		1.049	1.047	1.054	1.050	1.053	1.049	1.039	1.035	1.029
8		1.045	1.045	1.043	1.047	1.048	1.045	1.036	1.032	1.028
9		1.045	1.042	1.043	1.044	1.042	1.046	1.034	1.032	1.028
10		1.043	1.043	1.038	1.044	1.161	1.039	1.034	1.031	1.030
11		1.040	1.039	1.037	1.043	1.039	1.039	1.035	1.032	1.028

See notes to Table 1.

The $\boldsymbol{\psi}$ are again unique in the sense that, for each $\psi_1 \in \mathcal{C}$, there exists exactly one $\psi_2 \in \mathcal{C}$ such that (10) holds. The solution $\psi = \psi_1 = \psi_2$ considered by Harvey *et al.* (2009) is thus a special case of (10). In contrast, condition (10) defines an entire family of tests.

Remark 3. Notice that searching over $\mathcal{C}^2$ is without loss of generality. Suppose $\psi_1 < 1$. We then have $\mathrm{P}(\xi_1 > \psi_1 cv_{1,\alpha}) = \tilde{\alpha}_1$, say, where $\tilde{\alpha}_1 > \alpha$. Also write $\mathrm{P}(\xi_2 > \psi_2 cv_{2,\alpha}) = \tilde{\alpha}_2$. Similar as in footnote 2, it obtains that $\mathrm{P}\big(\bigcup_{i=1}^2 \xi_i > \psi_i cv_{i,\alpha}\big) = \tilde{\alpha}_1 + \tilde{\alpha}_2 - \mathrm{P}\big(\bigcap_{i=1}^2 \xi_i > \psi_i cv_{i,\alpha}\big) \geqslant \tilde{\alpha}_1 > \alpha$, because $\mathrm{P}\big(\bigcap_{i=1}^2 \xi_i > \psi_i cv_{i,\alpha}\big) \leqslant \tilde{\alpha}_2$. Hence, it is not possible to make one test more liberal and still achieve a level-α UR_{ψ_1,ψ_2} test.

The availability of an entire family of level-α tests, indexed by (ψ_1, ψ_2), of course raises the practical question of which tuple $\boldsymbol{\psi}$ to select. There is no unique uniformly most powerful way to do so. We propose the following rule. Select $\boldsymbol{\psi}$ such that

$$\min_{\psi_1} \left[\frac{\mathrm{P}\big(\bigcap_{i=1}^2 \xi_i > \psi_i cv_{i,\alpha}\big)}{\min\{\mathrm{P}(\xi_1 > \psi_1 cv_{1,\alpha}), \mathrm{P}(\xi_2 > \psi_2 cv_{2,\alpha})\}} \right] \tag{11}$$

Note that it is sufficient to minimize over ψ_1 only, since the corresponding ψ_2 is uniquely determined by (10).[3] We refer to this member of the family of tests as the 'Asymmetric' UR test. The corresponding tuples for the test pairs t_γ^{ADF} and $\lambda_{\max}$, $\hat{F}$ and $\lambda_{\max}$ as well as $\hat{F}$ and t_γ^{ECR} for $K-1$ up to 11 are reported in Table 2. This decision rule can be expected to yield powerful UR_{ψ_1,ψ_2} tests because (11) amounts to minimizing the number of instances where both tests reject under $\mathcal{H}_0$, while still generating a level-α test. That is, the tests are made as 'uncorrelated' as possible, without violating constraint (10). Now, since the behavior of the tests under local alternatives will change continuously from that under $\mathcal{H}_0$, making the tests 'uncorrelated' will produce a high number of correct rejections under $\mathcal{H}_1$.[4]

Remark 4. It turns out that the selection rule (11) satisfies

$$\mathrm{P}(\xi_1 > \psi_1 cv_{1,\alpha}) = \mathrm{P}(\xi_2 > \psi_2 cv_{2,\alpha}) \tag{12}$$

for all combinations considered in Table 2.[5] Under this condition, the UR_{ψ_1,ψ_2} test is equivalent to the min-test described in Remark 2. To show this, we first show that the min-test belongs to the family of UR_{ψ_1,ψ_2} tests. Let $F_{\min}$ be the null distribution function of $\min(p_1,p_2)$. The min-test rejects if $\min(p_1,p_2) < F_{\min}^{-1}(\alpha)$, thus if $p_1 < F_{\min}^{-1}(\alpha) \vee p_2 < F_{\min}^{-1}(\alpha)$. Equivalently, the test rejects if $\Xi_1^{-1}(p_1) > \Xi_1^{-1}\left(F_{\min}^{-1}(\alpha)\right) \vee \Xi_2^{-1}(p_2) > \Xi_2^{-1}\left(F_{\min}^{-1}(\alpha)\right)$ (recall the Ξ_i are defined to be decreasing functions). Since $p_i = \Xi_i(\xi_i)$, this test thus rejects if and only if

$$\xi_1 > \Xi_1^{-1}\left(F_{\min}^{-1}(\alpha)\right) \quad \vee \quad \xi_2 > \Xi_2^{-1}\left(F_{\min}^{-1}(\alpha)\right)$$

or equivalently if

$$\xi_1 > \psi_1 cv_{1,\alpha} \quad \vee \quad \xi_2 > \psi_2 cv_{2,\alpha}.$$

where $\psi_i := \Xi_i^{-1}\left(F_{\min}^{-1}(\alpha)\right)/cv_{i,\alpha}$. We know that, under $\mathcal{H}_0$, $\mathrm{P}(\xi_1 > \Xi_1^{-1}(F_{\min}^{-1}(\alpha)) \;\vee\; \xi_2 > \Xi_2^{-1}(F_{\min}^{-1}(\alpha))) = \alpha$, so that the min-test is a UR_{ψ_1,ψ_2} test. It remains to establish that the min-test is the only UR_{ψ_1,ψ_2} test that satisfies (12). By construction,

$$\mathrm{P}\left(\xi_i > \psi_i cv_{i,\alpha}\right) = \mathrm{P}\left(\xi_i > \Xi_i^{-1}\left(F_{\min}^{-1}(\alpha)\right)\right) = F_{\min}^{-1}(\alpha) \quad i = 1,2. \tag{13}$$

Uniqueness follows from monotonicity of the Ξ_i.

Remark 5. One can furthermore relax another of Harvey *et al.*'s restrictions, viz. that of combining

[3] We add an ϵ to the numerator of (11) to penalize borderline cases in which, due to simulation imprecision of the Wiener integrals, the numerator would otherwise be zero and the denominator very small, but positive.

[4] Unreported experiments with other tuples confirm this conjecture.

[5] To see why, write the numerator of (11) as $\mathrm{P}(\xi_1 > \psi_1 cv_{1,\alpha}) + \mathrm{P}(\xi_2 > \psi_2 cv_{2,\alpha}) - \mathrm{P}\left(\bigcup_{i=1}^2 \xi_i > \psi_i cv_{i,\alpha}\right)$. W.l.o.g. take the denominator to equal $\mathrm{P}(\xi_1 > \psi_1 cv_{1,\alpha})$. Using that $\mathrm{P}\left(\bigcup_{i=1}^2 \xi_i > \psi_i cv_{i,\alpha}\right) = \alpha$ for solutions to (10), (11) equals $\min_{\psi_1}\left[1 + \{\mathrm{P}(\xi_2 > \psi_2 cv_{2,\alpha}) - \alpha\}/\mathrm{P}(\xi_1 > \psi_1 cv_{1,\alpha})\right]$. Taking the derivative w.r.t. $\mathrm{P}(\xi_1 > \psi_1 cv_{1,\alpha})$ yields

$$\frac{\partial \mathrm{P}(\xi_2 > \psi_2 cv_{2,\alpha})/\partial \mathrm{P}(\xi_1 > \psi_1 cv_{1,\alpha}) \mathrm{P}(\xi_1 > \psi_1 cv_{1,\alpha}) - [\mathrm{P}(\xi_2 > \psi_2 cv_{2,\alpha}) - \alpha]}{\mathrm{P}(\xi_1 > \psi_1 cv_{1,\alpha})^2}, \tag{$*$}$$

which has an interior minimum (i.e. $\mathrm{P}(\xi_1 > \psi_1 cv_{1,\alpha}) < \mathrm{P}(\xi_2 > \psi_2 cv_{2,\alpha})$ strictly) if ($*$) equals zero. That is, the 'indifference curves' generated by the solutions $\boldsymbol{\psi}$ to (10) are sufficiently steep to produce the 'corner solution' (12).

$|\mathcal{I}| = 2$ tests. An $|\mathcal{I}|$-dimensional UR test is then, analogously to (9), defined by

$$\mathrm{P}\left(\bigcup_{i=1}^{|\mathcal{I}|} \xi_i > \psi_i cv_{i,\alpha}\right) = \alpha. \tag{14}$$

Of course, the detection of the solution $\boldsymbol{\psi} \in \mathcal{C}^{|\mathcal{I}|}$ then generally becomes numerically more challenging. For the symmetrical solution $\psi = \psi_1 = \psi_2 = \psi_3$ of $|\mathcal{I}| = 3$, where the tests considered are $\hat{F}, \lambda_{\max}$ and t_γ^{ADF}, we find a similar performance to the tests with $|\mathcal{I}| = 2$ discussed above, and therefore do not report detailed results for brevity.

4 Large Sample Results

We now report the large-sample power of the tests discussed in the previous sections. The power functions are computed as the probability that the statistics ξ_i and $\tilde{\chi}^2_{\mathcal{I}}$ exceed their level-α critical value, and the probability that the $UR_{\psi_1,\psi_2}(\xi_1, \xi_2)$ test (9) rejects. Given Propositions 1-3 and the results from Section 3, the asymptotic local power can be approximated by simulating the distributions presented above. We draw 15,000 replications of the functionals, for $T = 1000$. We put $c \in \{-1, -2, -3, \ldots, -30\}$ for the local-to-unity parameter and generate R^2 from $\{0, 0.05, 0.1, \ldots, 0.95\}$. The number of regressors $K - 1$ ranges from 1 to 5.

Table 3, corresponding to case (ii), reports the local asymptotic power of several combination tests as well as the corresponding single tests (see Appendix B for the other cases). Figures 1-2 plot the tests' power against R^2, holding c fixed at -10 and -15, respectively. We report results for $K - 1 = 1$; additional results are available upon request. As regards the single tests, we replicate Pesavento's finding that t_γ^{ECR} is the best test for small R^2. The power of all tests, with the exception of t_γ^{ADF}, increases quite quickly in R^2. The system-based $\lambda_{\max}$ test benefits most from an increase in R^2, fully exploiting the additional information contained in the equations for the $\boldsymbol{x}_t$. The formal similarity of $\hat{F}$ and t_γ^{ECR} translates into very similar local asymptotic power. The combination tests perform very well, in that they track the better of the underlying tests very closely. Their power curves sometimes even lie above that of the underlying tests. This effect is best seen in the lower panels, where the performance of the underlying tests t_γ^{ADF} and $\lambda_{\max}$ differs strongly. The upper panels show that, unsurprisingly, the power of the combination tests differs relatively less from that of either of the underlying tests if these perform similarly. Yet, $UR_{\psi_1,\psi_2}(\hat{F}, t_\gamma^{\mathrm{ECR}})$ and $\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\mathrm{ECR}})$ are again closer to the better of the underlying tests (typically $\hat{F}$) whenever there are discernible differences.

Figures 3-5 plot the tests' power against $-c$, holding R^2 fixed at $0, 0.25$ and 0.7, respectively. The figures confirm that all tests become more powerful as the distance c to the null increases, although the speed differs substantially. For large R^2 and $c = -15$, the power of $\lambda_{\max}$, $\tilde{\chi}^2_{\mathcal{I}}(t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$ and $UR_{\psi_1,\psi_2}(t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$ is more than three times larger than that of t_γ^{ADF}. It is again readily apparent that the combination tests are always close to the better of the two combined single tests. Of

Table 3: Local Asymptotic Power

$-c$	0	5	10	15	20
$R^2 = 0$					
$\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\mathrm{ECR}})$	0.050	0.106	0.240	0.455	0.706
$\tilde{\chi}^2_{\mathcal{I}}(t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	0.050	0.090	0.189	0.365	0.605
$\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\mathrm{ECR}}, t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	0.050	0.107	0.239	0.450	0.699
$UR_{\psi_1,\psi_2}(\hat{F}, t_\gamma^{\mathrm{ECR}})$	0.050	0.102	0.229	0.440	0.690
$UR_{\psi_1,\psi_2}(t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	0.050	0.080	0.171	0.334	0.571
$\hat{F}$	0.050	0.096	0.212	0.408	0.657
t_γ^{ECR}	0.050	0.112	0.255	0.482	0.731
$\lambda_{\max}$	0.050	0.068	0.124	0.239	0.427
t_γ^{ADF}	0.050	0.098	0.221	0.422	0.674
$R^2 = 0.25$					
$\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\mathrm{ECR}})$	0.051	0.116	0.320	0.623	0.858
$\tilde{\chi}^2_{\mathcal{I}}(t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	0.051	0.083	0.198	0.434	0.712
$\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\mathrm{ECR}}, t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	0.053	0.108	0.285	0.580	0.836
$UR_{\psi_1,\psi_2}(\hat{F}, t_\gamma^{\mathrm{ECR}})$	0.051	0.114	0.310	0.609	0.846
$UR_{\psi_1,\psi_2}(t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	0.051	0.081	0.186	0.399	0.661
$\hat{F}$	0.053	0.117	0.317	0.614	0.845
t_γ^{ECR}	0.050	0.114	0.308	0.613	0.853
$\lambda_{\max}$	0.051	0.078	0.185	0.402	0.662
t_γ^{ADF}	0.051	0.081	0.177	0.360	0.603
$R^2 = 0.5$					
$\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\mathrm{ECR}})$	0.052	0.145	0.506	0.832	0.966
$\tilde{\chi}^2_{\mathcal{I}}(t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	0.051	0.080	0.268	0.618	0.897
$\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\mathrm{ECR}}, t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	0.052	0.120	0.434	0.792	0.965
$UR_{\psi_1,\psi_2}(\hat{F}, t_\gamma^{\mathrm{ECR}})$	0.053	0.158	0.517	0.831	0.964
$UR_{\psi_1,\psi_2}(t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	0.051	0.092	0.307	0.639	0.892
$\hat{F}$	0.055	0.171	0.539	0.842	0.966
t_γ^{ECR}	0.050	0.124	0.444	0.792	0.957
$\lambda_{\max}$	0.052	0.109	0.360	0.699	0.922
t_γ^{ADF}	0.051	0.061	0.135	0.292	0.527
$R^2 = 0.75$					
$\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\mathrm{ECR}})$	0.052	0.300	0.834	0.983	0.999
$\tilde{\chi}^2_{\mathcal{I}}(t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	0.054	0.128	0.613	0.954	0.999
$\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\mathrm{ECR}}, t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	0.056	0.238	0.795	0.985	1.000
$UR_{\psi_1,\psi_2}(\hat{F}, t_\gamma^{\mathrm{ECR}})$	0.054	0.365	0.859	0.985	0.999
$UR_{\psi_1,\psi_2}(t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	0.052	0.212	0.738	0.973	1.000
$\hat{F}$	0.056	0.391	0.872	0.987	0.999
t_γ^{ECR}	0.050	0.197	0.718	0.957	0.997
$\lambda_{\max}$	0.053	0.267	0.798	0.984	1.000
t_γ^{ADF}	0.053	0.039	0.083	0.210	0.433

Case (ii). $\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\mathrm{ECR}})$ is our Fisher test (6) based on Boswijk's and Banerjee *et al.*'s tests, and $UR_{\psi_1,\psi_2}(\hat{F}, t_\gamma^{\mathrm{ECR}})$ is the corresponding Union-of-Rejections test (9). The other combination tests are defined analogously. See also notes to Table 1.

Figure 1: Local asymptotic power as a function of R^2, $c = -10$

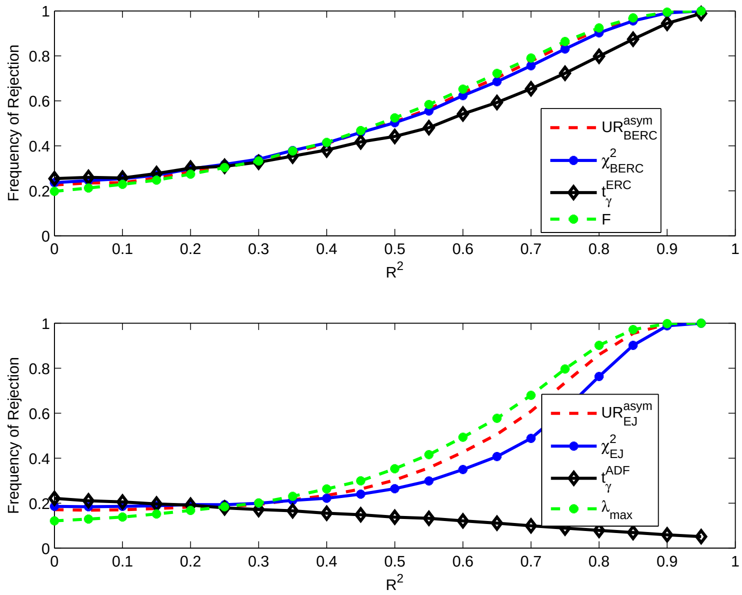
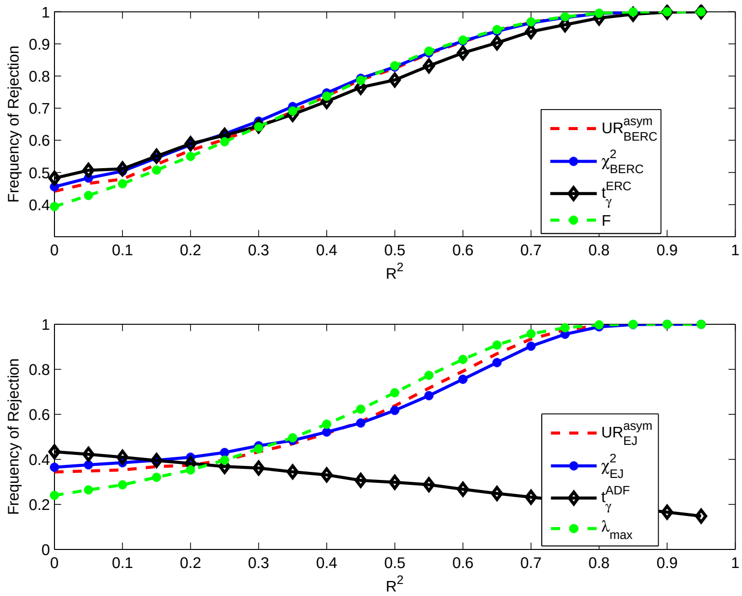


Figure 2: Local asymptotic power as a function of R^2, $c = -15$



Results are for the demeaned case (*ii*). χ^2_{BERC} is our Fisher test (6) based on Boswijk's and Banerjee *et al.*'s tests. χ^2_{EJ} is based on Engle and Granger's and Johansen's tests. $UR^{\text{asym}}_{\text{BERC}}$ and $UR^{\text{asym}}_{\text{EJ}}$ are the corresponding asymmetric UR_{ψ_1,ψ_2} tests (9). The single tests' power curves are for comparison.

Figure 3: Local asymptotic power as a function of $-c$, $R^2 = 0$

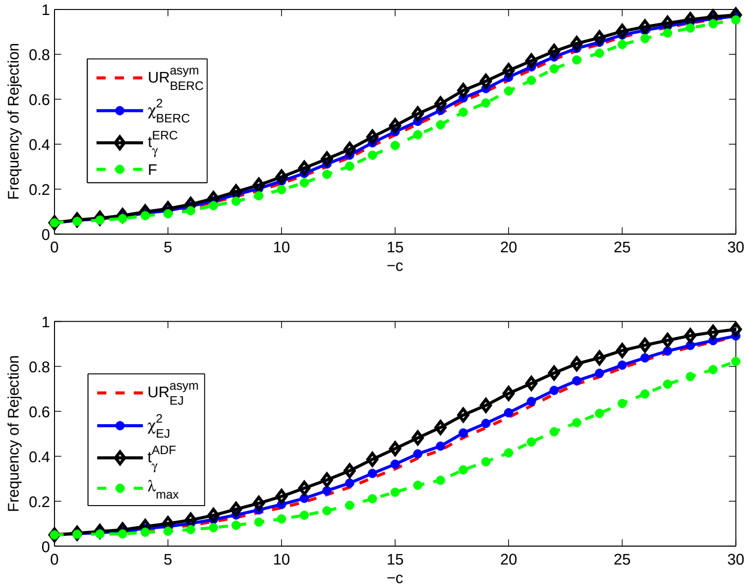
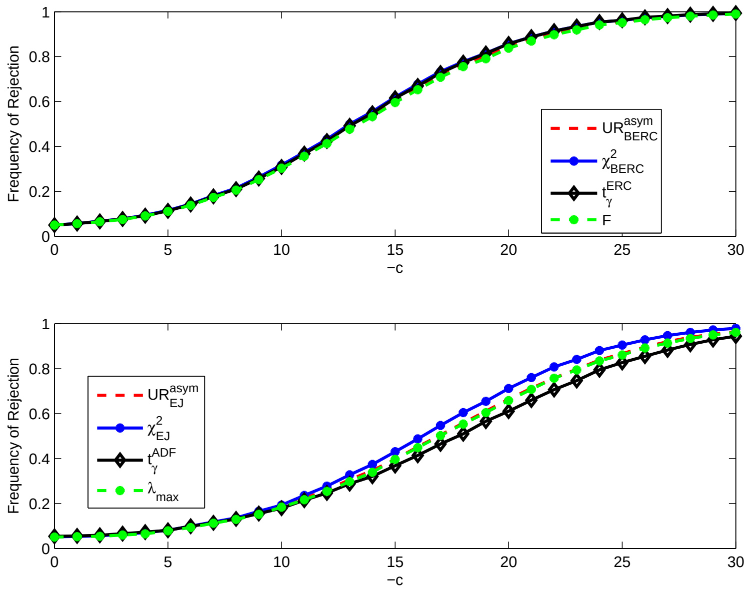


Figure 4: Local asymptotic power as a function of $-c$, $R^2 = 0.25$



See notes to Figure 1.

Figure 5: Local asymptotic power as a function of $-c$, $R^2 = 0.7$

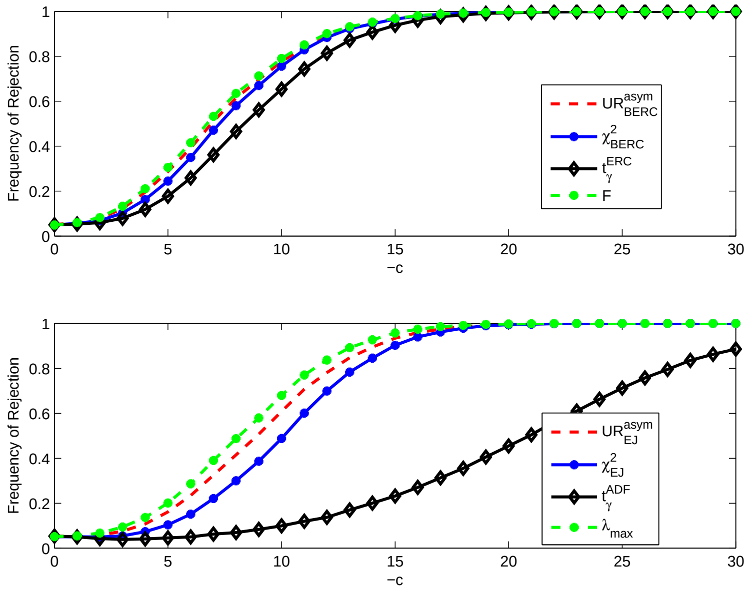
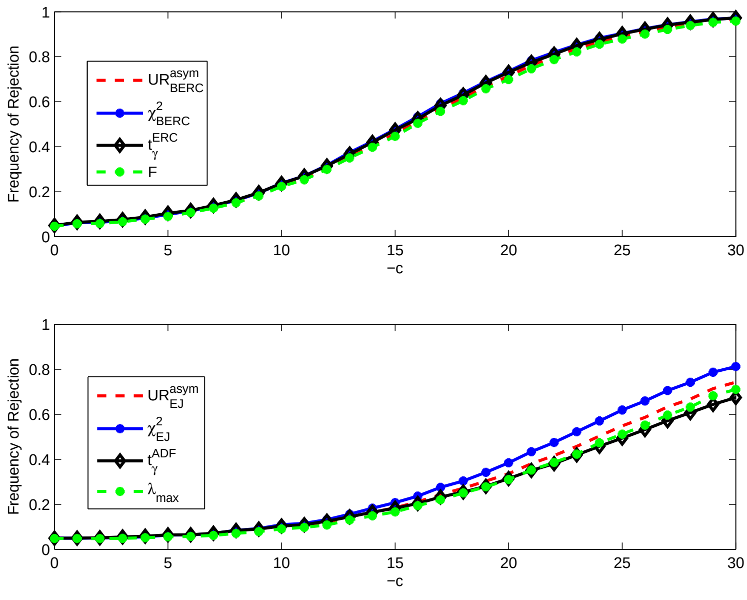
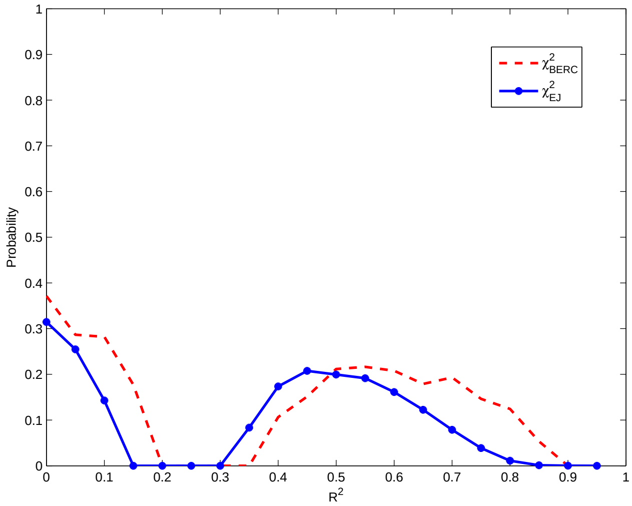


Figure 6: Local asymptotic power as a function of c, $R^2 = 0.35$, $K - 1 = 3$



See notes to Figure 1.

Figure 7: Cutoff probability q



The probability q, with which a pretest using the underlying tests (t_γ^{ADF} and $\lambda_{\max}$ for $\tilde{\chi}^2_{\mathcal{I}}(t_\gamma^{\text{ADF}}, \lambda_{\max})$, denoted χ^2_{EJ} in the plot; and analogously for $\hat{F}$, t_γ^{ECR} and χ^2_{BERC}) needs to select the weaker test for our Fisher test to be at least as powerful as the pretest, is plotted against R^2. $K-1=1$ and $c=-15$.

course, when the difference between the single tests is large, as in the lower panel of Figure 5, the power distance of the combination tests to the best single tests is somewhat larger. Note, however, that the combination tests' power is much closer to that of the better single test. Thus, a suitable combination test effectively offers a cheap insurance against selecting an inferior test, in that one never sacrifices much power, and potentially gains a lot. Moreover, it is noteworthy that, for $R^2 = 0.25$, both the Fisher test $\tilde{\chi}^2_{\mathcal{I}}(t_\gamma^{\text{ADF}}, \lambda_{\max})$ and the corresponding UR_{ψ_1,ψ_2} test even outperform both constituent single tests. Note from Figures 1-2 (the effect is more apparent in Fig. 2) that the power curves of the constituent tests t_γ^{ADF} and $\lambda_{\max}$ intersect at roughly $R^2 = 0.25$. Thus, combination tests appear to outperform the constituent tests when the latter are equally powerful. This effect generally becomes more pronounced with increasing $K-1$, cf. Figures 6 and 4.

Comparing the relative performance of the Fisher and UR_{ψ_1,ψ_2} tests, we find that the former are generally somewhat more powerful when both constituent tests have relatively high power. The UR_{ψ_1,ψ_2} tests outperform the Fisher tests when there is a large difference in power between the single tests, in particular if the weaker test has low absolute power. This is intuitive as the UR_{ψ_1,ψ_2} test looks for (at least) one single test indicating that the alternative $\mathcal{H}_1$ holds, effectively ignoring the less powerful test once the more powerful underlying test rejects. On the other hand, the Fisher test combines evidence from both tests, such that one test with relatively little power can tilt the overall decision of the Fisher test towards a non-rejection of $\mathcal{H}_0$. If both tests are at least moderately powerful, the Fisher test will combine that evidence to produce a rejection of $\mathcal{H}_0$.

Remark 6. As discussed above, some single tests are most powerful when R^2 is low, and others when R^2 is large. This might, alternatively to the approach discussed here, suggest a pretest strategy where one first estimates R^2 and then selects the most powerful cointegration test given the estimate $\hat{R}^2$. However, as pointed out by Pesavento (2007), because (unlike in Elliott *et al.*, 2005) $\boldsymbol{\theta}$ is assumed unknown and several key quantities are not consistently estimable in the present local-to-unity framework, it is not clear whether such an estimator $\hat{R}^2$ is feasible at all. Moreover, the above results show that the combination tests are never much less powerful than the best single test, and sometimes even more powerful. They are generally a lot more powerful than the worst test. Thus, even if an estimator $\hat{R}^2$ was available, it would not, certainly not in small samples, estimate R^2 without error, such that this pretest strategy would sometimes select the *less* powerful test. A pretest-based approach would therefore likely be less powerful than the strategies advocated here. Some rough calculations may help to illustrate this point. Let q denote the probability that the inferior test is selected. As an example, consider from Table 3 $\lambda_{\max}$, t_γ^{ADF} and $\tilde{\chi}^2_{\mathcal{I}}(t_\gamma^{\text{ADF}}, \lambda_{\max})$ for $R^2 = 0.75$ and $c = -15$. A pretest, if available, would need to indicate the use of the inferior test t_γ^{ADF} in only $q = [0.954 - 0.984]/[0.210 - 0.984] \times 100 \approx 4\%$ of the cases for the pretest-based strategy to be inferior to the combination test. For $\tilde{\chi}^2_{\mathcal{I}}(t_\gamma^{\text{ADF}}, \lambda_{\max})$, $\tilde{\chi}^2_{\mathcal{I}}(t_\gamma^{\text{ECR}}, \hat{F})$, $c = -15$ and $K - 1 = 1$, Figure 7 plots q against R^2 for that case. We see that q never exceeds 0.3. We find q to even equal 0 for $R^2 \in [0.15, 0.3] \cup (0.85, 1)$ (in the case of $\tilde{\chi}^2_{\mathcal{I}}(t_\gamma^{\text{ADF}}, \lambda_{\max})$), reflecting that the Fisher test is sometimes as or *more* powerful than a perfect pretest could be.

5 Bootstrap Analogs

The previous results rely entirely on asymptotic theory. In particular, the combination tests cannot be expected not to share small-sample deficiencies of the underlying single cointegration tests. The small-sample behavior of cointegration tests has, among many others, been analyzed by Haug (1996), who finds the tests to be somewhat sensitive to short-run dynamics in the errors. In particular, the finite-sample size of the tests depends on the choice of estimation method for these nuisance parameters. Thus, the local asymptotic power curves presented above are effectively approximations to the tests' finite-sample power curves. The bootstrap has recently been successfully employed to improve the small-sample behavior of single cointegration tests (Swensen, 2006; Palm *et al.*, 2009). We therefore now introduce bootstrap analogs of the combination tests to provide potentially more reliable inference in small samples.

Recall the aggregator of p-values from the Fisher test,

$$\tilde{\chi}^2_{\mathcal{I}} = -2 \sum_{i=1}^{|\mathcal{I}|} \ln(p_i).$$

To bootstrap the distribution of $\tilde{\chi}^2_{\mathcal{I}}$, we require a method to bootstrap cointegration tests. A suitable procedure has recently been proposed by Swensen (2006). In brief, Swensen's procedure resamples residuals from an estimated VECM representation of the data-generating process (DGP) to then generate integrated but non-cointegrated time series.

We propose the following Algorithm to estimate the finite-sample distribution of $\tilde{\chi}^2_{\mathcal{I}}$ to account for dependency among the test statistics.

Algorithm 1.

1. *Estimate the unrestricted VAR*

$$\boldsymbol{z}_t = \sum_{p=1}^{P} \boldsymbol{\Phi}_p \boldsymbol{z}_{t-p} + \boldsymbol{d}_t + \boldsymbol{\varepsilon}_t \tag{15}$$

to obtain coefficient estimates $\hat{\boldsymbol{d}}_t, \hat{\boldsymbol{\Phi}}_p$ and residuals $\hat{\boldsymbol{\varepsilon}}_t$. Transform $\hat{\boldsymbol{\Phi}}_p$, $p = 1, \ldots, P$, to $\hat{\boldsymbol{\Gamma}}_p$, $p = 1, \ldots, P-1$, as in representation (3). (See e.g. Hamilton (1994, Eq. 19.1.38) for the procedure.)[6]

2. *Check whether the system has no explosive root, i.e. whether $\|z\| > 1$, by solving $\det[\hat{\boldsymbol{B}}(z)] = 0$, where*

$$\hat{\boldsymbol{B}}(z) := \boldsymbol{I}_K - \hat{\boldsymbol{\Gamma}}_1 z - \cdots - \hat{\boldsymbol{\Gamma}}_{P-1} z^{P-1}.\text{[7]} \tag{16}$$

3. *If so, draw B series of pseudo errors $\{\boldsymbol{\varepsilon}^*_{t,b}\}^{b=1,\ldots,B}_{t=P,\ldots,T}$ by resampling non-parametrically with replacement from the residuals $\{\hat{\boldsymbol{\varepsilon}}_t\}_{t=P,\ldots,T}$.*

4. *With these pseudo errors, construct B series of pseudo observations $\boldsymbol{z}^*_{t,b}$ from*

$$\Delta \boldsymbol{z}^*_{t,b} = \sum_{p=1}^{P-1} \hat{\boldsymbol{\Gamma}}_p \Delta \boldsymbol{z}^*_{t-p,b} + \hat{\boldsymbol{d}}_t + \boldsymbol{\varepsilon}^*_{t,b}.$$

For the initial observations, set $\boldsymbol{z}^*_{t,b} = \boldsymbol{z}_t, t = 0, \ldots, P-1$.[8]

5. *Compute the test statistics $\xi^*_{i,b}$ for all pseudo samples $b = 1, \ldots, B$ and all cointegration tests that are to be combined, $i = 1, \ldots, |\mathcal{I}|$.*

6. *Estimate the distribution function of the test statistic of each test as*

$$\frac{\#\{\xi^*_{i,h} \le x | h = 1, \ldots, B\}}{B} =: 1 - \Xi^*_i(x)$$

and calculate the corresponding p-values $p^*_{i,b} := \Xi^*_i(\xi^*_{i,b})$. Correspondingly, calculate the p-values, p^*_i, of the test statistics on the original data by evaluating $\Xi^*_i(\xi_i)$.

[6] As pointed out by Swensen (2006) one could alternatively estimate a restricted VAR in first differences $\Delta \boldsymbol{z}_t$ and impose the null of no cointegration. However, as Paparoditis and Politis (2003) show for unit-root tests, imposing such a restriction may lead to a power loss.

[7] See Swensen (2006, Remark 1) and Johansen (1995, p. 71) for a discussion of this technical requirement. Note that under $h = 0$, $\hat{\boldsymbol{\alpha}}\hat{\boldsymbol{\beta}}' = \boldsymbol{0}$ in Swensen's notation, such that we have $\hat{A}(z) = (1-z)\hat{\boldsymbol{B}}(z)$, with the l.h.s. in Swensen's notation again. Thus his condition (iii) is equivalent to (16) in our context.

[8] Since we require pseudo observations that are integrated but non-cointegrated, $\boldsymbol{\Pi} = \boldsymbol{0}$ is imposed.

7. *Obtain the corresponding aggregate* $\tilde{\chi}^2_{\mathcal{I}}$ *test statistic*

$$\tilde{\chi}^{2,*}_{\mathcal{I},b} = -2 \sum_{i=1}^{|\mathcal{I}|} \ln \left(p^*_{i,b}\right).$$

8. *Estimate the cumulative distribution function* Θ *of the* $\tilde{\chi}^{2,*}_{\mathcal{I},b}$ *by*

$$\Theta^*(x) := \frac{\#\{\tilde{\chi}^{2,*}_{\mathcal{I},h} \leq x | h = 1, \ldots, B\}}{B}.$$

This provides us with a dependency robust version of the Fisher test, where the bootstrap p-values p^*_i of the underlying tests are obtained as in step 6 of Algorithm 1,

$$\tilde{\chi}^{2,*}_{\mathcal{I}} = -2 \sum_{i=1}^{|\mathcal{I}|} \ln \left(p^*_i\right)$$

and then reject $\mathcal{H}_0$ at level α if $\tilde{\chi}^{2,*}_{\mathcal{I}}$ exceeds the $(1-\alpha)$-quantile of Θ^*.

Heuristically, the method can be expected to work as follows. Swensen (2006) analytically proves that his bootstrap procedure for the Johansen λ_{trace} test (i.e. steps 1-6 in Algorithm 1) delivers a consistent estimate $1 - \Xi^*_i$ of the distribution of the test statistic under $\mathcal{H}_0$. It hence yields consistent estimates of p-values. The key element in Swensen's (2006) proposition is that the above bootstrap algorithm yields pseudo observations which have a representation asymptotically equivalent to the true DGP. Therefore, we can expect Swensen's proposition to carry over to other tests for cointegration (in particular the ones we mention above), as these essentially also rely on the availability of suitable pseudo-observations $\boldsymbol{z}^*_{t,b}$. The CMT with $\boldsymbol{\xi} := (\xi_1, \ldots, \xi_{|\mathcal{I}|})'$ as functions of the observations $\boldsymbol{z}_{i,t}$, for which an invariance principle holds, then also ensures a well-defined *joint* distribution of the test statistics $\boldsymbol{\xi}$. That joint distribution will then be consistently estimated with Algorithm 1 under fairly weak regularity conditions (Horowitz, 2001). We provide extensive numerical support for this heuristic argument in Section 6.[9]

Remark 7. In view of the equivalence of the UR_{ψ_1,ψ_2} and min-test established in Remark 4, we can also provide bootstrap analogs of the UR_{ψ_1,ψ_2} tests by bootstrapping the distribution of $\min_{i\in\mathcal{I}} p_i$, using Algorithm 1. We reject if the minimum of the p-values is less than the α-quantile of the bootstrap distribution $F^*_{\min}$.

[9] Appendix C describes an alternative bootstrap test that we found to have slightly higher power in unreported simulations. That approach requires stronger theoretical assumptions. For robustness reasons, we therefore advocate using the Fisher test described above.

6 Monte Carlo Experiments

6.1 Setup

We now study the properties of the proposed tests in a series of Monte Carlo experiments. As emphasized in the Introduction, different tests for cointegration are likely to differ in their power against different points in the (c-R^2)-space of the alternative hypothesis. Further, Johansen's $\lambda_{\max}$ test can be expected to be relatively more powerful if the data is indeed generated by a finite order VECM with uncorrelated errors. Since our test combines information from tests that are powerful in different directions, a likely advantage of our testing strategy is more robust power across different DGPs.

We consider the following DGPs:

$$\begin{aligned} \text{DGP(A):} \quad & \Delta x_t = v_{1t} \\ & y_t = \theta x_t + u_t \\ & u_t = \rho_T u_{t-1} + v_{2t}, \end{aligned}$$

where $\theta = 1$. The autoregressive coefficient $\rho_T = 1 + c/T$. $\mathcal{H}_0$ is obtained when $c = 0$, whereas we parameterize $\mathcal{H}_1$ by $c = -15$.[10] The errors $\boldsymbol{v}_t$ are drawn from

$$\boldsymbol{v}_t = \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}),$$

where

$$\boldsymbol{\Omega} = \begin{pmatrix} 1 & \delta \\ \delta & 1 \end{pmatrix}$$

For $R^2 = \delta^2$, we select $R^2 = \{0, 0.25, 0.5, 0.75\}$. DGP(A) closely follows Pesavento's model (1). To investigate the generality of her setup we additionally investigate the following DGPs.

$$\begin{aligned} \text{DGP(B):} \quad & \Delta \boldsymbol{z}_t = \boldsymbol{\Pi} \boldsymbol{z}_{t-1} + \boldsymbol{\Gamma} \Delta \boldsymbol{z}_{t-1} + \boldsymbol{u}_t \\ & \boldsymbol{\Gamma} = 0.2 \boldsymbol{I}_2. \end{aligned}$$

$$\begin{aligned} \text{DGP(C):} \quad & y_t + \eta x_t = a_{1t}, \quad y_t + \theta x_t = a_{2t} \\ & \theta = -1, \; \eta = -1/2 \\ & a_{1t} = a_{1t-1} + u_{1t}, \; a_{2t} = \rho_T a_{2t-1} + u_{2t}. \end{aligned}$$

In DGPs (B) and (C) we set $\boldsymbol{u}_t = (u_{1t}, u_{2t})' \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I}_2)$. For DGP(B) $\mathcal{H}_0$ is obtained when $\boldsymbol{\Pi} = \mathbf{0}$, whereas we parameterize the alternative hypothesis of cointegration by $\boldsymbol{\Pi} = (1 \quad 0)' \, (.15 \quad -.15)$. For DGP(C), $\mathcal{H}_0$ and $\mathcal{H}_1$ are parameterized as in DGP(A).[11] DGPs (A) and (C) are local, such

[10] Power results for other c are given in Appendix D.

[11] Of course, Granger's representation theorem would allow us to write DGP(C) in a VECM form. However, error

that power ought to remain roughly constant when increasing T, while power should increase for DGP(B).

These designs are widely used in Monte Carlo studies of cointegration tests. See for instance Pesavento (2004, 2007) or Elliott *et al.* (2005) for DGP(A), Swensen (2006) for (B), or Engle and Granger (1987), Haug (1996) and Gregory *et al.* (2004) for (C).

For each DGP, we draw 5,000 replications under $\mathcal{H}_0$ and $\mathcal{H}_1$. We choose $T \in \{50, 75, 100, 150, 200\}$. These time-series lengths correspond to typical sample sizes encountered in applied macroeconometric work, e.g. when using quarterly data. To mitigate the effect of initial conditions under $\mathcal{H}_1$, we simulate each DGP for $T + 30$ time periods and discard the first 30 observations. For each replication, we compute the UR^* and the $\tilde{\chi}^{2,*}_{\mathcal{I}}$ tests based on $B = 10,000$ bootstrap replications. To keep the setup simple, we initially combine $|\mathcal{I}| = 2$ underlying tests. In particular, we select Johansen's (1988) $\lambda_{\max}$ test and the augmented Engle and Granger (1987) residual-based test (t^{ADF}_{γ}). We opt for this pair of tests as they are all widely used in applied research. Moreover, Section 4 establishes that these tests have high power for different values of the nuisance parameter R^2, such that combining them seems promising. For comparison, we also combine Boswijk's (1994) $\hat{F}$ test and Banerjee *et al.*'s t^{ECR}_{γ} test

To investigate the relative performance of the new tests, we compare them to following alternative possibilities to test for cointegration: First, the standard augmented t^{ADF}_{γ}, $\lambda_{\max}$, t^{ECR}_{γ} and $\hat{F}$ tests, where we reject the null hypothesis if the test statistics fall short of (respectively exceed) the level α critical value computed from the appropriate distribution of the tests.[12] Second, we investigate bootstrap versions of the tests (denoted in the following by $t^{\text{ADF},*}_{\gamma}$, $\lambda^*_{\max}$, $t^{\text{ECR},*}_{\gamma}$ and $\hat{F}^*$), which are by-products of our UR^* and $\tilde{\chi}^{2,*}_{\mathcal{I}}$ tests. Third, we compute a 'naive' meta test based on the bootstrapped versions of the two underlying tests. This test rejects whenever at least one of the tests rejects. We call this test 'naive' because it ignores the multiple-testing nature of the problem. Studying this test hence reveals the size distortion incurred by selecting the most rejective test from a set of cointegration tests.

Implementation of the cointegration tests requires to select an order $\hat{P}$ of lagged differences to account for auto-correlation. In practice this is often done via some lag-length selection criterion, see e.g. Lütkepohl (2005). To reduce the computational burden we waive this option and use the correct lag order throughout. All tests are based on case (*iii*).

6.2 Results

Table 4 reports the small sample size of the tests based on $\lambda_{\max}$ and t^{ADF}_{γ} at the 5% level. Results for DGP(A) are based on $R^2 = 0.25$.[13] The main findings may be summarized as follows. As

terms would be correlated, the matrix $\boldsymbol{\Pi}$ would have no rows of zeros under the alternative and $\boldsymbol{\Gamma}$ would equal $\mathbf{0}$.

[12] In the case of the t^{ADF}_{γ} test we follow the standard practice of using MacKinnon (1996)-type critical values, which approximate those of the unknown finite-sample distributions.

[13] Appendix D reports results for other values of R^2. Furthermore, we ran all simulations described above at the 1% and 10% level, with qualitatively similar results. We also experimented with a version of DGP(C) with AR(1)

Table 4: Small-sample size based on $\lambda_{\max}$ and $t_{\gamma}^{\mathrm{ADF}}$

		Bootstrap tests					asymptotic tests				
DGP	T	$\tilde{\chi}^{2,*}_{\mathcal{I}}$	UR^*	naive	$\lambda^*_{\max}$	$t_{\gamma}^{\mathrm{ADF},*}$	$\lambda_{\max}$	$t_{\gamma}^{\mathrm{ADF}}$	naive	$\tilde{\chi}^2_{\mathcal{I}}$	UR_{ψ_1,ψ_2}
(A)	50						0.054	0.080	0.113	0.062	0.084
	75						0.055	0.077	0.110	0.059	0.080
	100			TO BE ADDED			0.054	0.075	0.111	0.056	0.072
	150						0.054	0.063	0.099	0.049	0.069
	200						0.048	0.058	0.090	0.047	0.059
(B)	50	0.068		0.102	0.068	0.053	0.067	0.069	0.108	0.063	0.077
	75	0.057		0.088	0.053	0.053	0.060	0.062	0.098	0.060	0.065
	100	0.052		0.082	0.055	0.049	0.061	0.059	0.093	0.060	0.066
	150	0.050		0.075	0.049	0.046	0.057	0.060	0.090	0.057	0.061
	200						0.057	0.063	0.092	0.063	0.063
(C)	50						0.053	0.081	0.114	0.060	0.081
	75						0.055	0.076	0.110	0.055	0.077
	100			TO BE ADDED			0.054	0.069	0.103	0.054	0.072
	150						0.054	0.064	0.099	0.049	0.070
	200						0.048	0.058	0.089	0.044	0.060

Average rejection rates at nominal level of 5%. 5,000 replications and 10,000 bootstrap replications. $t_{\gamma}^{\mathrm{ADF}}$ and $\lambda_{\max}$ refer to Engle and Granger (1987) and Johansen (1988) tests, $t_{\gamma}^{\mathrm{ADF},*}$ and $\lambda^*_{\max}$ are their bootstrap counterparts. naive rejects when $t_{\gamma}^{\mathrm{ADF},*}$ or $\lambda^*_{\max}$ or both reject. UR_{ψ_1,ψ_2} is the test defined by (9) and (11) and and UR^* is the bootstrap counterpart. $\tilde{\chi}^2_{\mathcal{I}}$ is the Fisher test (6) and $\tilde{\chi}^{2,*}_{\mathcal{I}}$ is its bootstrap counterpart. (UR^* and $\tilde{\chi}^{2,*}_{\mathcal{I}}$ are described in Algorithm 1.)

expected, the 'naive' test is oversized and its size exceeds that of the single tests by approximately 3 - 4 percentage points.[14] All other bootstrap tests control size reasonably well. The UR_{ψ_1,ψ_2} test (and to a lesser extent also the $\tilde{\chi}^2_{\mathcal{I}}$ test) exhibits a slight upward size distortion for small T, partly due to a distortion of $t_{\gamma}^{\mathrm{ADF},*}$ for small T. However, this size distortion vanishes for $T \geqslant 100$. The bootstrap versions (where available) of the tests seem to approach the nominal size somewhat more quickly, which reflects the fact that the bootstrap distribution generated in Algorithm 1 generally is a somewhat more accurate approximation to the unknown-finite sample distribution than the asymptotic one.

Table 5 reports the small sample power of the $\lambda_{\max}$ and $t_{\gamma}^{\mathrm{ADF}}$-based tests at the level α of 5%. For DGP(A), we find that the local asymptotic results from Section 4 predict the finite-sample results rather well, in that $t_{\gamma}^{\mathrm{ADF}}$ and $\lambda_{\max}$ again have similar power for this R^2. Moreover, the combination tests $\tilde{\chi}^2_{\mathcal{I}}$ and UR_{ψ_1,ψ_2} again outperform both single tests. As expected, power increases in T for all tests for DGP(B). While of the single tests the $t_{\gamma}^{\mathrm{ADF}}$ test is the most powerful single test for DGP(C), the $\lambda_{\max}$ and $\lambda^*_{\max}$ tests are most powerful for DGP(B). This result may not be entirely surprising, as both tests were originally designed having DGPs of type (B) and (C) respectively in mind. For those DGPs, the meta tests $\tilde{\chi}^2_{\mathcal{I}}$ and UR_{ψ_1,ψ_2} again both perform similarly and well, in that their power is again close or superior to that of the better of the two constituent tests.

error terms instead of white noise $\boldsymbol{u}_t$. Again, results are qualitatively similar. Tables with the additional results are available upon request.

[14]This size distortion is very close to the one that can be inferred from Table I in Gregory *et al.* (2004).

Table 5: Small-sample power based on $\lambda_{\max}$ and $t_{\gamma}^{\mathrm{ADF}}$

		Bootstrap tests					asymptotic tests				
DGP	T	$\tilde{\chi}_{\mathcal{I}}^{2,*}$	UR^*	naive	$\lambda_{\max}^*$	$t_{\gamma}^{\mathrm{ADF},*}$	$\lambda_{\max}$	$t_{\gamma}^{\mathrm{ADF}}$	naive	$\tilde{\chi}_{\mathcal{I}}^{2}$	UR_{ψ_1,ψ_2}
(A)	50						0.288	0.362	0.462	0.359	0.374
	75						0.290	0.320	0.440	0.343	0.344
	100						0.279	0.296	0.413	0.307	0.318
	150						0.279	0.270	0.394	0.301	0.302
	200						0.275	0.258	0.386	0.284	0.290
(B)	50						0.099	0.104	0.161	0.100	0.118
	75						0.170	0.143	0.238	0.157	0.172
	100			TO BE ADDED			0.293	0.211	0.372	0.269	0.283
	150						0.623	0.402	0.691	0.591	0.593
	200						0.888	0.646	0.918	0.880	0.869
(C)	50						0.194	0.372	0.413	0.310	0.329
	75						0.193	0.342	0.384	0.285	0.297
	100						0.177	0.316	0.358	0.268	0.271
	150						0.178	0.299	0.344	0.258	0.260
	200						0.173	0.277	0.327	0.239	0.246

See notes to Table 4. For DGP(A), $R^2 = 0.25$ and for (A) and (C), $c = -15$.

Tables 6 and 7 reports analogous results for the tests based on $\hat{F}$ and $t_{\gamma}^{\mathrm{ECR}}$. Once more, all tests have a slight upward size distortion for small T, which vanishes as T increases. The performance of the single $\hat{F}$ and $t_{\gamma}^{\mathrm{ECR}}$ tests is again similar, as predicted by Section 4. It is therefore not surprising that the performance of the meta tests $\tilde{\chi}_{\mathcal{I}}^{2}$ and UR_{ψ_1,ψ_2} is also very similar to that of the single tests. Comparing Tables 5 and 7, we find that $t_{\gamma}^{\mathrm{ADF}}$ and $\lambda_{\max}$ outperform either $\hat{F}$ or $t_{\gamma}^{\mathrm{ECR}}$ for DGP(C) and (B), respectively, which again reflects that the former tests were designed having such DGPs in mind. This also implies that the superior local asymptotic power properties of $\hat{F}$ and $t_{\gamma}^{\mathrm{ECR}}$ found by Pesavento (2004) may be somewhat model-specific, in that these results do not carry over to other parameterizations of cointegrated systems such as DGPs (B) and (C). Hence, it would be premature to recommend routine application of either the $\hat{F}$ or $t_{\gamma}^{\mathrm{ECR}}$ test in practice. Indeed, our meta tests are attractive because they not only offer a robust insurance against wrong test choice given the nuisance parameter R^2, but effectively also robustness when there is uncertainty over the form of the DGP, as is the case in practice.

6.3 Extension to more than two tests

We combined the $t_{\gamma}^{\mathrm{ADF}}$ and $\lambda_{\max}$ as well as the $\hat{F}$ and $t_{\gamma}^{\mathrm{ECR}}$ tests to illustrate our approach with widely applied cointegration tests. Of course, as the discussion in Section 3 makes clear, our approach is not restricted to combining $|\mathcal{I}| = 2$ tests. The procedures can accommodate other and more tests as well. Potentially, this could yield further gains in power if the additional tests have high power for the given nuisance parameter value.

We therefore run extra simulations, where we combine all four tests considered in the previous subsection (denoted $\tilde{\chi}_{\mathcal{I}}^{2}(4)$) and compare its performance to the combination tests based on $\lambda_{\max}$

Table 6: Small-sample size based on $\hat{F}$ and t_γ^{ECR}

		Bootstrap tests					asymptotic tests				
DGP	T	$\tilde{\chi}^{2,*}_{\mathcal{I}}$	UR^*	naive	$\hat{F}^*$	$t_\gamma^{\mathrm{ECR},*}$	$\hat{F}$	t_γ^{ECR}	naive	$\tilde{\chi}^2_{\mathcal{I}}$	UR_{ψ_1,ψ_2}
(A)	50						0.084	0.077	0.093	0.079	0.082
	75						0.076	0.072	0.086	0.075	0.076
	100						0.073	0.073	0.084	0.074	0.073
	150						0.065	0.062	0.073	0.065	0.066
	200						0.057	0.053	0.063	0.054	0.057
(B)	50						0.069	0.068	0.079	0.070	0.069
	75						0.067	0.064	0.076	0.065	0.065
	100			TO BE ADDED			0.063	0.060	0.072	0.061	0.063
	150						0.060	0.057	0.069	0.058	0.058
	200						0.064	0.063	0.071	0.062	0.063
(C)	50						0.083	0.076	0.091	0.079	0.082
	75						0.071	0.069	0.081	0.070	0.070
	100						0.068	0.064	0.075	0.067	0.067
	150						0.068	0.065	0.076	0.068	0.067
	200						0.057	0.058	0.066	0.058	0.059

See notes to Table 4. $\hat{F}$ and t_γ^{ECR} are from Boswijk (1994) and Banerjee *et al.* (1998), respectively. Starred tests are bootstrap counterparts.

and t_γ^{ADF}, denoted $\tilde{\chi}^2_{\mathcal{I}}(2)$. In view of the qualitatively similar performance of bootstrap and asymptotic tests we restrict ourselves to the latter for brevity. We find that for the case at hand the more general $\tilde{\chi}^2_{\mathcal{I}}(4)$ test outperforms its simple counterpart $\tilde{\chi}^2_{\mathcal{I}}(2)$ rather markedly. Of course, the asymptotic results from Section 4 predict that this is a setting where t_γ^{ADF} is less powerful, such that one might exclude it from the meta tests. Yet, bearing Remark 6 in mind, such knowledge about the DGP will rarely be available in practice. Indeed, we view it as implausible that researchers should feel the need to conduct statistical inference about a key feature of the time series at hand—cointegration versus non-cointegration—while at the same time having accurate knowledge about some nuisance parameter. Hence, the extra robustness that can be gained from combining $|\mathcal{I}| = 4$ tests may well be attractive for practitioners.

To summarize, both UR_{ψ_1,ψ_2} and $\tilde{\chi}^2_{\mathcal{I}}$ control the size of the test and yet provide a robust, powerful and flexible alternative to traditional cointegration tests.

7 Empirical Application

7.1 Setup

Naturally we are interested in the applicability and the relevance of our testing strategy in practice. To shed light on this question, we revisit the studies which Gregory *et al.* (2004) investigated for 'mixed signals', i.e. conflicting test results from cointegration tests. Gregory *et al.* (2004) analyze the cointegration tests reported in 34 studies dealing with cointegration which were published in

Table 7: Small-sample power based on $\hat{F}$ and $t_{\gamma}^{\mathrm{ECR}}$

		Bootstrap tests					asymptotic tests				
DGP	T	$\tilde{\chi}_{\mathcal{I}}^{2,*}$	UR^{*}	naive	$\hat{F}^{*}$	$t_{\gamma}^{\mathrm{ECR},*}$	$\hat{F}$	$t_{\gamma}^{\mathrm{ECR}}$	naive	$\tilde{\chi}_{\mathcal{I}}^{2}$	UR_{ψ_1,ψ_2}
(A)	50						0.553	0.517	0.578	0.542	0.542
	75						0.528	0.491	0.553	0.517	0.519
	100						0.496	0.463	0.526	0.487	0.488
	150						0.474	0.435	0.500	0.463	0.463
	200						0.457	0.413	0.478	0.440	0.445
(B)	50						0.133	0.116	0.146	0.122	0.129
	75						0.193	0.157	0.207	0.176	0.186
	100			TO BE ADDED			0.265	0.223	0.281	0.244	0.256
	150						0.460	0.389	0.472	0.424	0.443
	200						0.660	0.572	0.671	0.621	0.636
(C)	50						0.297	0.321	0.336	0.315	0.306
	75						0.281	0.300	0.313	0.294	0.288
	100						0.254	0.278	0.290	0.272	0.261
	150						0.246	0.270	0.282	0.264	0.256
	200						0.232	0.259	0.269	0.248	0.240

See notes to Table 4. $\hat{F}$ and $t_{\gamma}^{\mathrm{ECR}}$ are from Boswijk (1994) and Banerjee *et al.* (1998), respectively. Starred tests are bootstrap counterparts. For DGP(A), $R^2 = 0.25$ and for (A) and (C), $c = -15$.

the Journal of Applied Econometrics from 1994 to March/April 2001.[15] From these studies we construct 161 data sets in which we test for cointegration. The data sets exhibit large differences in sample size, which ranges from 27 to 7693 with a median size of 73. Similarly the number of variables differs across studies and ranges from 2 to 11.

Our goal is to document the extent to which conflicting test results arise in actual applications and how our proposed meta test is able to heal this problem. As Gregory *et al.* (2004), we do not intend to suggest that the authors of the original studies have been in any way strategic in their choice of which test for cointegration to apply. Most applied researchers tend to view the different tests as rather interchangeable, with the choice more dependent on the nature of the investigation.

We follow Gregory *et al.* (2004) closely in their setup. The original published studies employ different methods to test their specifications. To make the results comparable, we impose a unifying but standard methodology. For the residual-based tests where a dependent variable is required, we follow the choice in the original paper if possible. If there is no obvious dependent variable, we choose it on the basis of the highest coefficient of determination of first-stage regressions. Additionally we need to allow for variation in lag lengths across data sets. The literature discusses a number of different methods for choosing the number of lags. We have chosen a fairly standard one and determine the lag length $\hat{P}$ for the VECM estimation of our algorithm using a Schwarz Information Criterion (BIC) as described e.g. in Lütkepohl (2005, Sections 4.3.2 and 8.1). We

[15] The raw data are available online at http://qed.econ.queensu.ca/jae/2004-v19.1/gregory-haug-lomuto/. Our modified data sets are available upon request.

Table 8: Rejection rates when combining $|\mathcal{I}| > 2$ tests

		Size		Power	
DGP	T	$\tilde{\chi}^2_{\mathcal{I}}(2)$	$\tilde{\chi}^2_{\mathcal{I}}(4)$	$\tilde{\chi}^2_{\mathcal{I}}(2)$	$\tilde{\chi}^2_{\mathcal{I}}(4)$
(A)	50	0.062	0.071	0.801	0.936
	75	0.059	0.068	0.801	0.935
	100	0.056	0.063	0.803	0.926
	150	0.049	0.057	0.808	0.921
	200	0.047	0.047	0.813	0.919
(B)	50	0.063	0.069	0.100	0.114
	75	0.060	0.063	0.157	0.171
	100	0.060	0.060	0.269	0.267
	150	0.057	0.055	0.591	0.531
	200	0.063	0.062	0.880	0.810
(C)	50	0.060	0.069	0.310	0.330
	75	0.055	0.061	0.285	0.309
	100	0.054	0.060	0.268	0.281
	150	0.049	0.059	0.258	0.271
	200	0.044	0.052	0.239	0.255

Average rejection rates at nominal level of 5%. 5,000 replications. $UR_{\psi_1,\psi_2}(|\mathcal{I}|)$ and $\tilde{\chi}^2_{\mathcal{I}}(|\mathcal{I}|)$ combine the $|\mathcal{I}|$ tests described in the text. For DGP(A), results are based on $R^2 = 0.75$.

search over the range $1 \leq \hat{P} \leq \min\left(8\left(\frac{T}{100}\right)^{1/5}, \frac{T-2}{2(K+2)}\right)$, and impose the same number of lags for all tests. Our qualitative conclusions would not be different if alternative selection methods were employed. All tests include a constant and a trend.

7.2 Results

We compare the test results of $\lambda_{\max}$, t_γ^{ADF}, t_γ^{ECR} and $\hat{F}$ tests as underlying tests with the $UR_{\psi_1,\psi_2}(\lambda_{\max}, t_\gamma^{\text{ADF}})$, $UR_{\psi_1,\psi_2}(t_\gamma^{\text{ECR}}, \hat{F})$, $\tilde{\chi}^2_{\mathcal{I}}(\lambda_{\max}, t_\gamma^{\text{ADF}})$, $\tilde{\chi}^2_{\mathcal{I}}(t_\gamma^{\text{ECR}}, \hat{F})$, and $\tilde{\chi}^2_{\mathcal{I}}(\lambda_{\max}, t_\gamma^{\text{ADF}}, t_\gamma^{\text{ECR}}, \hat{F})$ tests. Specifically, we proceed as follows. We first check whether all single tests agree or not in their testing decision at the 5% level, see left panel of Table 9. In those cases where conflicting test results occurred we check what the test used in the original paper had suggested as a test result (more precisely what would have been the outcome of our version with the chosen lag-length criterion), see the right panel of Table 9. In all cases we compare the results to that of the $\tilde{\chi}^2_{\mathcal{I}}(\lambda_{\max}, t_\gamma^{\text{ADF}}, t_\gamma^{\text{ECR}}, \hat{F})$ test.[16]

Table 9 reports the frequencies for all possible pairs of outcomes. We see that when all tests do not reject $\mathcal{H}_0$, the meta test does not reject either. However, such cases of agreeing tests make up only 64% $(= (52 + 51)/161)$ of all data sets (tests).

For the remaining 36% of data sets we have conflicting single tests and here our test turns out to be most useful. It allows the researcher to arrive at a definite conclusion. We find in 47%

[16] See Appendix E for results based on $\tilde{\chi}^2_{\mathcal{I}}(\lambda_{\max}, t_\gamma^{\text{ADF}})$; results for the other combination tests and bootstrap based tests are available upon request.

Table 9: Test results in applied studies and the $\tilde{\chi}^2_{\mathcal{I}}$ test

number of cases in which...	single test results... agree r	$\neg r$	conflict	$\sum$	... in case of conflicting results: 'preferred' test†	r	$\neg r$	$\sum$
$\tilde{\chi}^2_{\mathcal{I}}(4) : r$	50	0	31	81	$\tilde{\chi}^2_{\mathcal{I}}(4) : r$	20	9	29
$\tilde{\chi}^2_{\mathcal{I}}(4) : \neg r$	2	51	27	80	$\tilde{\chi}^2_{\mathcal{I}}(4) : \neg r$	17	7	24
$\sum$	52	51	58	161	$\sum$	37	16	53

$\tilde{\chi}^2_{\mathcal{I}}(4)$ abbreviates $\tilde{\chi}^2_{\mathcal{I}}(\lambda_{\max}, t_\gamma^{\text{ADF}}, t_\gamma^{\text{ECR}}, \hat{F})$.
r : test rejects; $\neg r$: test does not reject
† : Test type on which conclusions in the original study were based (see fn. 17).
Absolute frequencies of cointegration-test results for data from Gregory *et al.* (2004). Single tests include Engle-Granger, Boswijk, Banerjee *et al.* and Johansen tests. The $\tilde{\chi}^2_{\mathcal{I}}(4)$ combines these tests as described in Section 3.

(= 27/58) of the conflicting cases that the meta test does not reject the null. In the remaining 53% of the conflicting cases, however, the $\tilde{\chi}^2_{\mathcal{I}}$ test leads to a rejection of the null of no cointegration. Moreover, we note the following.

First, rejecting whenever at least one (but not all) of the tests rejected would have lead to a substantial overstatement of cointegration (58 vs. 31 cases according to the $\tilde{\chi}^2_{\mathcal{I}}$ test). Similarly, not rejecting whenever one test did not reject would have lead to an understatement of cointegration.

Second, the tests that have been 'preferred' in the actual studies tend to be more rejective than our meta test (37 vs. 29 rejections in 53 tests).[17] This suggests that the evidence in favor of cointegration would have been somewhat less pronounced if the studies could have relied on a suitable meta test for cointegration. (Note that the preferred test being more rejective than the meta test here does not contradict the favorable power properties of the meta test found in Section 6, as the latter can, and should, of course only be shown to be powerful in a class of level-α tests. Whether or not the way researchers identify their 'preferred' test leads to a level-α test or suffers from data-mining is impossible to say without knowledge of the decision process.)

Third, whether or not the preferred test rejected the null does not seem to be informative on whether or not $\tilde{\chi}^2_{\mathcal{I}}$ rejects conditional on observing conflicting test results. This is reflected by very similar conditional probabilities: $27/58 \simeq 17/37 \simeq 7/16 \approx 0.45$. In other words, we cannot conclude from a published test result what the $\tilde{\chi}^2_{\mathcal{I}}$ test would indicate, conditional on the fact that a further single test leads to a conflicting test result.

[17] For this purpose, we categorize the studies according to whether they use a residual- (i.e. those by Engle and Granger (1987) or Phillips and Ouliaris (1990)) or Johansen (1988) system-based test. That is, we identify all Johansen tests with $\lambda_{\max}$ and all residual-based tests with t_γ^{ADF}. In view of the highly positive correlation within classes of tests established by Gregory *et al.* (2004), this approximation is accurate. In five (58 − 53) cases of conflicting test results, the original study did not report a cointegration test but was rather concerned with e.g. estimating cointegration vectors.

8 Conclusion

This paper proposes meta tests that combine information from different underlying tests for cointegration. The tests take into account the multiple testing nature of running more than one underlying test and hence control size. The meta tests are constructed by deriving the distribution of suitable aggregators of the underlying tests (e.g., Fisher's), by appropriately modifying the critical values of the underlying tests, as well as by using corresponding bootstrap methods. By contrast, running more than one test and then simply inferring about the hypothesis from the most rejective test does not achieve this goal but leads to a significantly oversized test, as we have shown. While controlling size, the proposed meta tests are powerful, and certainly more powerful than traditional methods to account for multiplicity like for example the Bonferroni method.

Extensive asymptotic and Monte Carlo results demonstrate the effectiveness of our approach. An application of our test to a set of cointegration studies confirms its practical value. It allows the applied researcher to arrive at an unambiguous test decision in cases of conflicting single test results.

The setup we put forward is fairly general and hence can be adopted to other testing problems for which several (imperfectly correlated) tests have been developed. Examples include testing for unit roots or heteroscedasticity. Essentially, what is needed is either the distribution of some suitable aggregator or a bootstrap method suitable for the phenomenon of interest. For the above mentioned testing problems such bootstrap methods would be the sieve and the wild bootstrap, respectively.

In practice, a major advantage of our proposed test should be that it relieves the applied researcher from the discretionary and sometimes arbitrary choice of the cointegration test(s) she wants to rely on to reach a test decision.

References

Banerjee A, Dolado JJ, Mestre R. 1998. Error-correction mechanism tests for cointegration in a single-equation framework. *Journal of Time Series Analysis* **19**: 267–283.

Boswijk HP. 1994. Testing for an unstable root in conditional and unconditional error correction models. *Journal of Econometrics* **63**: 37–60.

Breitung J. 2001. Rank tests for nonlinear cointegration. *Journal of Business & Economic Statistics* **19**: 331–340.

Demetrescu M, Hassler U, Tarcolea AI. 2006. Combining significance of correlated statistics with application to panel data. *Oxford Bulletin of Economics and Statistics* **68**: 647–663.

Elliott G, Jansson M, Pesavento E. 2005. Optimal power for testing potential cointegrating vectors with known parameters for nonstationarity. *Journal of Business & Economic Statistics* **23**: 34–48.

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

Fisher R. 1932. *Statistical Methods for Research Workers*. London: Oliver and Boyd.

Gregory AW, Haug AA, Lomuto N. 2004. Mixed signals among tests for cointegration. *Journal of Applied Econometrics* **19**: 89–98.

Hamilton JD. 1994. *Time Series Analysis*. Princeton: Princeton University Press.

Hartung J. 1999. A note on combining dependent tests of significance. *Biometrical Journal* **41**: 849–855.

Harvey DI, Leybourne SJ, Taylor AMR. 2009. Unit root testing in practice: Dealing with uncertainty over the trend and initial condition. *Econometric Theory* **forthcoming**.

Haug AA. 1996. Tests for cointegration: A Monte Carlo comparison. *Journal of Econometrics* **71**: 89–115.

Horowitz JL. 2001. The bootstrap. In Heckman JJ, Leamer EE (eds.) *Handbook of Econometrics*, vol. 5, chap. 52, Amsterdam: Elsevier, pages 3159–3228.

Johansen S. 1988. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* **12**: 231–254.

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

Lütkepohl H. 2005. *New Introduction to Multiple Time Series Analysis*. Berlin: Springer.

MacKinnon JG. 1996. Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics* **11**: 601–618.

Palm FC, Smeekes S, Urbain JP. 2009. A sieve bootstrap test for cointegration in a conditional error correction model. *Econometric Theory* **to appear**.

Paparoditis E, Politis DN. 2003. Residual-based block bootstrap for unit root testing. *Econometrica* **71**: 813–855.

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

Pesavento E. 2007. Residuals-based tests for the null of no-cointegration: An analytical comparison. *Journal of Time Series Analysis* **28**: 111–137.

Phillips PCB, Ouliaris S. 1990. Asymptotic properties of residual based tests for cointegration. *Econometrica* **58**: 165–193.

Swensen AR. 2006. Bootstrap algorithms for testing and determining the cointegration rank in VAR models. *Econometrica* **74**: 1699–1714.

Appendix A Further critical values and correction factors

Table A.1: Critical Values for the $\tilde{\chi}^2$-test.

	case											
$K-1$	(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)
$\alpha = 0.01$												
	t_γ^{ADF} and $\lambda_{\max}$			$\hat{F}$ and $\lambda_{\max}$			$\hat{F}$ and t_γ^{ECR}			$\hat{F}$ and t_γ^{ADF}		
1	16.948	17.304	17.289	17.077	17.175	17.066	17.827	18.201	18.230	16.551	17.390	17.572
2	16.651	16.679	16.720	16.443	16.355	16.227	17.888	18.051	18.176	16.361	16.686	17.078
3	16.236	16.259	16.263	15.787	15.814	15.777	17.831	17.951	18.069	16.137	16.430	16.795
4	15.871	15.845	15.973	15.384	15.497	15.430	17.763	17.912	18.017	16.074	16.396	16.493
5	15.626	15.701	15.666	15.241	15.143	15.202	17.889	17.813	17.937	16.011	16.201	16.295
6	15.412	15.348	15.467	15.015	15.038	14.995	17.773	17.710	17.937	15.858	15.997	16.326
	$\hat{F}$, $\lambda_{\max}$ and t_γ^{ADF}			$\hat{F}$, $\lambda_{\max}$ and t_γ^{ECR}			$\hat{F}$, $\lambda_{\max}$, t_γ^{ADF}, t_γ^{ECR}					
1	24.174	25.263	25.420	25.151	25.718	25.726	32.713	33.969	34.334			
2	23.595	23.855	24.091	24.369	24.501	24.623	31.793	32.077	32.601			
3	22.685	23.026	23.446	23.485	23.731	23.936	30.651	31.169	31.742			
4	22.256	22.498	22.681	23.144	23.344	23.461	30.088	30.774	30.836			
5	21.924	22.020	22.058	22.799	22.974	23.003	29.800	29.850	30.113			
6	21.686	21.729	21.887	22.633	22.548	22.677	29.222	29.544	29.962			
$\alpha = 0.1$												
	t_γ^{ADF} and $\lambda_{\max}$			$\hat{F}$ and $\lambda_{\max}$			$\hat{F}$ and t_γ^{ECR}			$\hat{F}$ and t_γ^{ADF}		
1	8.612	8.678	8.686	8.614	8.596	8.588	8.895	9.085	9.120	8.478	8.739	8.892
2	8.457	8.479	8.451	8.368	8.390	8.351	8.907	9.031	9.062	8.434	8.607	8.702
3	8.350	8.363	8.352	8.251	8.241	8.254	8.868	8.980	9.049	8.370	8.494	8.611
4	8.290	8.301	8.272	8.199	8.151	8.167	8.915	8.957	9.015	8.346	8.478	8.555
5	8.221	8.242	8.276	8.150	8.105	8.127	8.887	8.939	9.009	8.353	8.440	8.563
6	8.165	8.200	8.199	8.094	8.093	8.076	8.892	8.899	8.973	8.366	8.456	8.507
	$\hat{F}$, $\lambda_{\max}$ and t_γ^{ADF}			$\hat{F}$, $\lambda_{\max}$ and t_γ^{ECR}			$\hat{F}$, $\lambda_{\max}$, t_γ^{ADF}, t_γ^{ECR}					
1	12.570	12.761	12.855	12.542	12.748	12.863	16.593	16.964	17.187			
2	12.218	12.378	12.374	12.265	12.379	12.358	16.171	16.444	16.507			
3	12.008	12.075	12.177	12.031	12.175	12.244	15.920	16.097	16.239			
4	11.873	11.962	12.008	12.007	12.059	12.108	15.776	15.938	16.086			
5	11.807	11.857	11.915	11.971	11.999	12.044	15.681	15.804	15.989			
6	11.711	11.773	11.826	11.880	11.970	11.995	15.644	15.746	15.872			

1% and 10% Critical values for combination tests based on $\tilde{\chi}^2$. t_γ^{ADF} is from Engle and Granger (1987), $\lambda_{\max}$ from Johansen (1988), $\hat{F}$ from Boswijk (1994) and t_γ^{ECR} from Banerjee *et al.* (1998).

Table A.2: Correction Factors for the minimum p-value test.

$K-1$	case (i)	(ii)	(iii)	(i)	(ii)	(iii)
$\alpha = 0.01$						
	t_γ^{ADF} and $\lambda_{\max}$			$\hat{F}$ and t_γ^{ECR}		
1	0.006	0.006	0.006	0.007	0.008	0.008
2	0.006	0.006	0.006	0.007	0.008	0.008
3	0.006	0.006	0.006	0.007	0.007	0.008
4	0.005	0.005	0.005	0.007	0.007	0.007
5	0.005	0.005	0.005	0.007	0.007	0.007
6	0.005	0.005	0.005	0.007	0.007	0.007
$\alpha = 0.05$						
	t_γ^{ADF} and $\lambda_{\max}$			$\hat{F}$ and t_γ^{ECR}		
1	0.031	0.033	0.033	0.038	0.041	0.043
2	0.030	0.030	0.030	0.037	0.038	0.040
3	0.029	0.029	0.029	0.036	0.038	0.039
4	0.028	0.028	0.028	0.036	0.037	0.038
5	0.028	0.028	0.028	0.035	0.036	0.037
6	0.027	0.027	0.028	0.035	0.035	0.037
$\alpha = 0.1$						
	t_γ^{ADF} and $\lambda_{\max}$			$\hat{F}$ and t_γ^{ECR}		
1	0.064	0.067	0.067	0.077	0.083	0.086
2	0.061	0.062	0.062	0.075	0.079	0.081
3	0.059	0.059	0.060	0.074	0.076	0.079
4	0.058	0.058	0.058	0.072	0.075	0.077
5	0.057	0.057	0.057	0.072	0.074	0.075
6	0.056	0.056	0.056	0.071	0.073	0.075

Correction Factors for the minimum p-value test.

Appendix B Local Asymptotic Power, further results

Table B.1: Local Asymptotic Power

$-c$	0	5	10	15	20
$R^2 = 0$					
$\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\mathrm{ECR}})$	0.050	0.153	0.404	0.716	0.917
$\tilde{\chi}^2_{\mathcal{I}}(t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	0.050	0.120	0.311	0.595	0.841
$\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\mathrm{ECR}}, t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	0.050	0.153	0.403	0.709	0.913
$UR_{\psi_1,\psi_2}(\hat{F}, t_\gamma^{\mathrm{ECR}})$	0.049	0.137	0.372	0.682	0.898
$UR_{\psi_1,\psi_2}(t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	0.050	0.103	0.280	0.555	0.813
$\hat{F}$	0.050	0.114	0.319	0.616	0.861
t_γ^{ECR}	0.050	0.175	0.450	0.762	0.939
$\lambda_{\max}$	0.050	0.076	0.187	0.391	0.641
t_γ^{ADF}	0.050	0.134	0.364	0.669	0.892
$R^2 = 0.25$					
$\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\mathrm{ECR}})$	0.049	0.196	0.561	0.862	0.974
$\tilde{\chi}^2_{\mathcal{I}}(t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	0.049	0.126	0.377	0.714	0.933
$\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\mathrm{ECR}}, t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	0.050	0.179	0.523	0.847	0.975
$UR_{\psi_1,\psi_2}(\hat{F}, t_\gamma^{\mathrm{ECR}})$	0.049	0.172	0.511	0.827	0.965
$UR_{\psi_1,\psi_2}(t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	0.046	0.116	0.337	0.647	0.891
$\hat{F}$	0.049	0.174	0.513	0.819	0.958
t_γ^{ECR}	0.050	0.198	0.558	0.864	0.976
$\lambda_{\max}$	0.047	0.105	0.312	0.614	0.867
t_γ^{ADF}	0.048	0.120	0.331	0.625	0.871
$R^2 = 0.5$					
$\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\mathrm{ECR}})$	0.050	0.293	0.757	0.954	0.995
$\tilde{\chi}^2_{\mathcal{I}}(t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	0.053	0.157	0.541	0.893	0.991
$\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\mathrm{ECR}}, t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	0.053	0.254	0.723	0.958	0.997
$UR_{\psi_1,\psi_2}(\hat{F}, t_\gamma^{\mathrm{ECR}})$	0.049	0.288	0.729	0.942	0.993
$UR_{\psi_1,\psi_2}(t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	0.051	0.172	0.532	0.861	0.982
$\hat{F}$	0.052	0.328	0.763	0.949	0.994
t_γ^{ECR}	0.050	0.230	0.689	0.938	0.993
$\lambda_{\max}$	0.049	0.192	0.578	0.888	0.988
t_γ^{ADF}	0.054	0.106	0.284	0.581	0.842
$R^2 = 0.75$					
$\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\mathrm{ECR}})$	0.052	0.573	0.954	0.997	1.000
$\tilde{\chi}^2_{\mathcal{I}}(t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	0.051	0.344	0.898	0.997	1.000
$\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\mathrm{ECR}}, t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	0.051	0.516	0.955	0.999	1.000
$UR_{\psi_1,\psi_2}(\hat{F}, t_\gamma^{\mathrm{ECR}})$	0.052	0.616	0.953	0.997	1.000
$UR_{\psi_1,\psi_2}(t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	0.050	0.431	0.914	0.997	1.000
$\hat{F}$	0.052	0.659	0.963	0.997	1.000
t_γ^{ECR}	0.050	0.369	0.892	0.992	1.000
$\lambda_{\max}$	0.050	0.495	0.942	0.998	1.000
t_γ^{ADF}	0.051	0.079	0.235	0.523	0.805

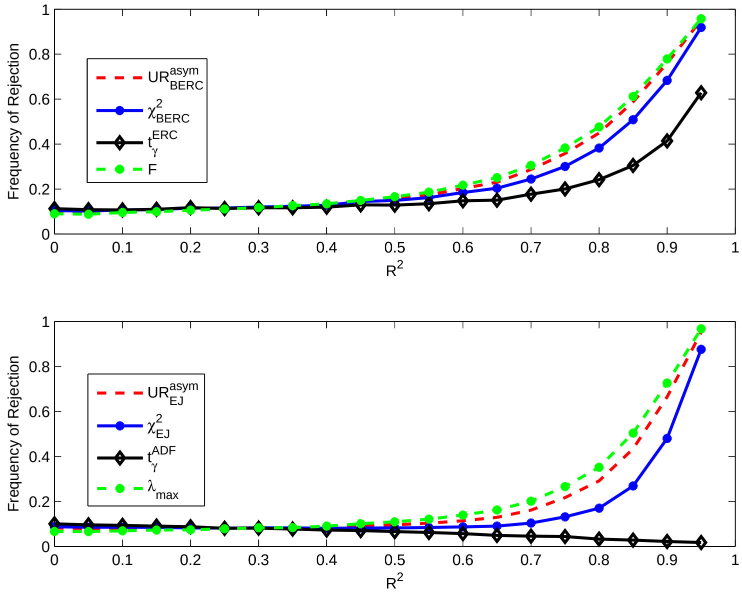
Case (i). See notes to Table 3.

Table B.2: Local Asymptotic Power

$-c$	0	5	10	15	20
$R^2 = 0$					
$\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\text{ECR}})$	0.050	0.073	0.148	0.290	0.487
$\tilde{\chi}^2_{\mathcal{I}}(t_\gamma^{\text{ADF}}, \lambda_{\max})$	0.050	0.069	0.132	0.253	0.423
$\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\text{ECR}}, t_\gamma^{\text{ADF}}, \lambda_{\max})$	0.050	0.074	0.151	0.294	0.490
$UR_{\psi_1,\psi_2}(\hat{F}, t_\gamma^{\text{ECR}})$	0.049	0.070	0.142	0.279	0.471
$UR_{\psi_1,\psi_2}(t_\gamma^{\text{ADF}}, \lambda_{\max})$	0.051	0.064	0.116	0.230	0.392
$\hat{F}$	0.050	0.070	0.138	0.271	0.457
t_γ^{ECR}	0.050	0.076	0.155	0.305	0.508
$\lambda_{\max}$	0.050	0.054	0.092	0.165	0.283
t_γ^{ADF}	0.050	0.074	0.150	0.290	0.486
$R^2 = 0.25$					
$\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\text{ECR}})$	0.048	0.081	0.191	0.405	0.668
$\tilde{\chi}^2_{\mathcal{I}}(t_\gamma^{\text{ADF}}, \lambda_{\max})$	0.050	0.072	0.127	0.267	0.495
$\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\text{ECR}}, t_\gamma^{\text{ADF}}, \lambda_{\max})$	0.049	0.084	0.194	0.406	0.664
$UR_{\psi_1,\psi_2}(\hat{F}, t_\gamma^{\text{ECR}})$	0.051	0.069	0.121	0.247	0.456
$UR_{\psi_1,\psi_2}(t_\gamma^{\text{ADF}}, \lambda_{\max})$	0.050	0.079	0.171	0.364	0.626
$\hat{F}$	0.047	0.083	0.199	0.412	0.668
t_γ^{ECR}	0.050	0.083	0.183	0.388	0.652
$\lambda_{\max}$	0.050	0.067	0.123	0.261	0.471
t_γ^{ADF}	0.050	0.070	0.115	0.222	0.398
$R^2 = 0.5$					
$\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\text{ECR}})$	0.049	0.089	0.285	0.621	0.874
$\tilde{\chi}^2_{\mathcal{I}}(t_\gamma^{\text{ADF}}, \lambda_{\max})$	0.050	0.063	0.146	0.386	0.699
$\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\text{ECR}}, t_\gamma^{\text{ADF}}, \lambda_{\max})$	0.049	0.080	0.231	0.552	0.840
$UR_{\psi_1,\psi_2}(\hat{F}, t_\gamma^{\text{ECR}})$	0.049	0.102	0.318	0.648	0.882
$UR_{\psi_1,\psi_2}(t_\gamma^{\text{ADF}}, \lambda_{\max})$	0.049	0.069	0.179	0.439	0.734
$\hat{F}$	0.048	0.108	0.339	0.669	0.891
t_γ^{ECR}	0.050	0.079	0.228	0.537	0.823
$\lambda_{\max}$	0.048	0.078	0.221	0.511	0.794
t_γ^{ADF}	0.050	0.052	0.077	0.151	0.292
$R^2 = 0.75$					
$\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\text{ECR}})$	0.051	0.134	0.596	0.923	0.993
$\tilde{\chi}^2_{\mathcal{I}}(t_\gamma^{\text{ADF}}, \lambda_{\max})$	0.054	0.069	0.356	0.811	0.983
$\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\text{ECR}}, t_\gamma^{\text{ADF}}, \lambda_{\max})$	0.053	0.107	0.524	0.906	0.993
$UR_{\psi_1,\psi_2}(\hat{F}, t_\gamma^{\text{ECR}})$	0.050	0.196	0.689	0.946	0.995
$UR_{\psi_1,\psi_2}(t_\gamma^{\text{ADF}}, \lambda_{\max})$	0.053	0.117	0.531	0.907	0.993
$\hat{F}$	0.052	0.216	0.714	0.952	0.996
t_γ^{ECR}	0.050	0.077	0.385	0.801	0.970
$\lambda_{\max}$	0.051	0.153	0.607	0.937	0.996
t_γ^{ADF}	0.054	0.029	0.035	0.071	0.166

Case (*iii*). See notes to Table B.1.

Figure B.1: Local asymptotic power as a function of R^2, $c = -5$



Results are for the demeaned case (ii). χ^2_{BERC} is our Fisher test (6) based on Boswijk's and Banerjee *et al.*'s tests. χ^2_{EJ} is based on Engle and Granger's and Johansen's tests. $UR^{\text{asym}}_{\text{BERC}}$ and $UR^{\text{asym}}_{\text{EJ}}$ are the corresponding asymmetric UR_{ψ_1,ψ_2} test (9). The single tests' power curves are for comparison.

Appendix C Alternative Bootstrap Tests

This Appendix describes an alternative bootstrap approach that makes somewhat stronger assumptions about the joint distribution of the test statistics. Its power was slightly superior to the Fisher-test version in our simulations (detailed results are available). Based on the p-values of the cointegration tests, define a probit representation by $\Phi^{-1}(p_i) =: s_i$, where Φ is the cumulative distribution function of the standard normal distribution.

Let $\boldsymbol{s} = \left(s_1, \ldots, s_{|\mathcal{I}|}\right)'$. Then, asymptotically, the components of $\boldsymbol{s}$ are marginally standard normal under the null. Hartung (1999) additionally assumes that $\boldsymbol{s}$ is jointly normally distributed, denoted $\boldsymbol{s} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$. Under this assumption, we have $\boldsymbol{\iota}'\boldsymbol{s} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\iota}'\boldsymbol{\Sigma}\boldsymbol{\iota})$, where $\boldsymbol{\iota} = (1, \ldots, 1)'$. This leads to a standardized meta test statistic,

$$\tau = \frac{\boldsymbol{\iota}'\boldsymbol{s}}{\sqrt{\boldsymbol{\iota}'\boldsymbol{\Sigma}\boldsymbol{\iota}}}.$$

τ is standard normal under $\mathcal{H}_0$ and joint normality. Fortunately, Demetrescu *et al.* (2006) demonstrate that this assumption is not necessary.

As a practical requirement, we need a feasible consistent estimator of $\boldsymbol{\Sigma}$. If the number of tests $|\mathcal{I}|$ is small, there is no hope to estimate $\boldsymbol{\Sigma}$ meaningfully from the realizations of $\boldsymbol{s}$. We rely on a bootstrap method to estimate $\boldsymbol{\Sigma}$. More specifically, we use the following algorithm.

Algorithm 2.

1. - 6. As in Algorithm 1.

7. Obtain the corresponding probit representation of each test statistic, $s^*_{i,b} = \Phi^{-1}(p^*_{i,b})$*, stacked in* $\boldsymbol{s}^*_b = \left(s^*_{1,b}, \ldots, s^*_{|\mathcal{I}|,b}\right)'$*. Correspondingly, obtain* $s_i = \Phi^{-1}(p_i)$*.*

8. Estimate the covariance matrix $\boldsymbol{\Sigma}$ *of the probits of the tests by*

$$\boldsymbol{\Sigma}^* = \frac{1}{B}\sum_b (\boldsymbol{s}^*_b - \bar{\boldsymbol{s}}^*)(\boldsymbol{s}^*_b - \bar{\boldsymbol{s}}^*)',$$

where $\bar{\boldsymbol{s}}^* = \frac{1}{B}\sum_b \boldsymbol{s}^*_b$*.*

This Algorithm provides a feasible version of the test statistic τ,

$$\tau^* = \frac{\boldsymbol{\iota}'\boldsymbol{s}}{\sqrt{\boldsymbol{\iota}'\boldsymbol{\Sigma}^*\boldsymbol{\iota}}},$$

where $\boldsymbol{s}$ is the probit representation of the bootstrap version of the underlying tests (see step 7 of the above Algorithm). We then reject $\mathcal{H}_0$ at level α if $\tau^* < \Phi^{-1}(\alpha)$.

The following Lemma provides a useful consistency property of the test.

Lemma 5. *If (i)* $\alpha < 1/2$ *and (ii) all underlying tests* s_i *reject at level* α*, then* τ^* *rejects* $\mathcal{H}_0$ *at least at level* α*.*

Proof. Recall that $\Phi^{-1}(\alpha) < 0$ for $\alpha < 1/2$. Then, it follows from (ii) that $s_i < \Phi^{-1}(\alpha) < 0$ for all $i = 1, \ldots, |\mathcal{I}|$. Hence, $\boldsymbol{\iota}'\boldsymbol{s} < 0$. Further, since the entries of the positive semi-definite correlation matrix $\boldsymbol{\Sigma}^*$ are bounded by 1 and -1, we have $\sqrt{\boldsymbol{\iota}'\boldsymbol{\Sigma}^*\boldsymbol{\iota}} \leq |\mathcal{I}|$. Thus,

$$\tau^* = \frac{\boldsymbol{\iota}'\boldsymbol{s}}{\sqrt{\boldsymbol{\iota}'\boldsymbol{\Sigma}^*\boldsymbol{\iota}}} \leq \frac{\boldsymbol{\iota}'\boldsymbol{s}}{|\mathcal{I}|} < \Phi^{-1}(\alpha)$$

and the result follows. □

Appendix D Additional Simulation Results

Table D.1: Small-sample power based on $\lambda_{\max}$ and t_γ^{ADF}, DGP(A), further R^2s

		Bootstrap tests					asymptotic tests				
DGP	T	$\tilde{\chi}_{\mathcal{I}}^{2,*}$	UR^*	naive	$\lambda_{\max}^*$	$t_\gamma^{\mathrm{ADF},*}$	$\lambda_{\max}$	t_γ^{ADF}	naive	$\tilde{\chi}_{\mathcal{I}}^2$	UR_{ψ_1,ψ_2}
$R^2 = 0$	50						0.194	0.440		0.349	0.380
	75						0.192	0.406		0.323	0.341
	100						0.177	0.369		0.300	0.315
	150						0.178	0.335		0.284	0.294
	200						0.173	0.320		0.263	0.275
$R^2 = 0.5$	50						0.528	0.257		0.440	0.501
	75						0.528	0.223		0.435	0.487
	100			TO BE ADDED			0.524	0.207		0.411	0.469
	150						0.522	0.189		0.404	0.468
	200						0.511	0.180		0.389	0.463
$R^2 = 0.75$	50						0.918	0.149		0.801	0.885
	75						0.925	0.121		0.801	0.895
	100						0.925	0.108		0.803	0.895
	150						0.934	0.100		0.808	0.899
	200						0.938	0.095		0.813	0.910

See notes to Table 4.

Table D.2: Small-sample power based on $\hat{F}$ and $t_{\gamma}^{\mathrm{ECR}}$, DGP(A), further R^2s

		Bootstrap tests					asymptotic tests				
DGP	T	$\tilde{\chi}_{\mathcal{I}}^{2,*}$	UR^*	naive	$\hat{F}^*$	$t_{\gamma}^{\mathrm{ECR},*}$	$\hat{F}$	$t_{\gamma}^{\mathrm{ECR}}$	naive	$\tilde{\chi}_{\mathcal{I}}^{2}$	UR_{ψ_1,ψ_2}
$R^2 = 0$	50						0.401	0.427	0.443	0.418	0.411
	75						0.370	0.407	0.418	0.395	0.387
	100						0.343	0.376	0.388	0.364	0.356
	150						0.319	0.353	0.364	0.341	0.329
	200						0.301	0.331	0.341	0.322	0.311
$R^2 = 0.5$	50						0.771	0.663	0.781	0.734	0.762
	75						0.748	0.637	0.759	0.711	0.735
	100			TO BE ADDED			0.739	0.618	0.752	0.700	0.727
	150						0.714	0.594	0.724	0.671	0.696
	200						0.702	0.569	0.711	0.654	0.686
$R^2 = 0.75$	50						0.968	0.882	0.969	0.953	0.965
	75						0.966	0.878	0.967	0.950	0.962
	100						0.959	0.865	0.960	0.941	0.953
	150						0.960	0.853	0.962	0.939	0.955
	200						0.958	0.846	0.960	0.935	0.953

See notes to Table 4. $\hat{F}$ and $t_{\gamma}^{\mathrm{ECR}}$ are from Boswijk (1994) and Banerjee *et al.* (1998), respectively. Starred tests are bootstrap counterparts.

Table D.3: Small-sample power based on $\lambda_{\max}$ and $t_{\gamma}^{\mathrm{ADF}}$, further c

		Bootstrap tests					asymptotic tests				
DGP	T	$\tilde{\chi}_{\mathcal{I}}^{2,*}$	UR^*	naive	$\lambda_{\max}^*$	$t_{\gamma}^{\mathrm{ADF},*}$	$\lambda_{\max}$	$t_{\gamma}^{\mathrm{ADF}}$	naive	$\tilde{\chi}_{\mathcal{I}}^{2}$	UR_{ψ_1,ψ_2}
(A)	50						0.144	0.186	0.257	0.171	0.196
	75						0.141	0.163	0.235	0.153	0.171
	100						0.133	0.147	0.215	0.140	0.161
	150						0.133	0.140	0.217	0.136	0.152
	200						0.131	0.128	0.203	0.129	0.148
(C)	50						0.107	0.197	0.233	0.157	0.179
	75						0.103	0.175	0.216	0.138	0.164
	100						0.098	0.170	0.207	0.135	0.155
	150						0.097	0.159	0.200	0.130	0.147
	200						0.098	0.143	0.187	0.119	0.135

See notes to Table 4. For DGP(A), $R^2 = 0.25$ and for (A) and (C), $c = -10$.

Table D.4: Small-sample power based on $\lambda_{\max}$ and t_γ^{ADF}, further c

		Bootstrap tests					asymptotic tests				
DGP	T	$\tilde{\chi}_{\mathcal{I}}^{2,*}$	UR^*	naive	$\lambda_{\max}^*$	$t_\gamma^{\text{ADF},*}$	$\lambda_{\max}$	t_γ^{ADF}	naive	$\tilde{\chi}_{\mathcal{I}}^2$	UR_{ψ_1,ψ_2}
(A)	50						0.537	0.602	0.718	0.641	0.626
	75						0.528	0.544	0.694	0.609	0.597
	100						0.506	0.485	0.644	0.560	0.543
	150						0.497	0.461	0.629	0.544	0.534
	200						0.487	0.445	0.620	0.531	0.514
(C)	50						0.346	0.600	0.632	0.530	0.533
	75						0.334	0.556	0.599	0.496	0.501
	100						0.306	0.513	0.547	0.458	0.446
	150						0.298	0.481	0.524	0.430	0.421
	200						0.294	0.466	0.510	0.424	0.409

See notes to Table 4. For DGP(A), $R^2 = 0.25$ and for (A) and (C), $c = -20$.

Table D.5: Small-sample power based on $\hat{F}$ and t_γ^{ECR}, further c

		Bootstrap tests					asymptotic tests				
DGP	T	$\tilde{\chi}_{\mathcal{I}}^{2,*}$	UR^*	naive	$\hat{F}^*$	$t_\gamma^{\text{ECR},*}$	$\hat{F}$	t_γ^{ECR}	naive	$\tilde{\chi}_{\mathcal{I}}^2$	UR_{ψ_1,ψ_2}
(A)	50						0.303	0.265	0.320	0.285	0.293
	75						0.264	0.231	0.283	0.249	0.258
	100						0.247	0.214	0.265	0.234	0.241
	150						0.232	0.202	0.252	0.224	0.223
	200						0.219	0.190	0.236	0.203	0.210
(C)	50						0.175	0.184	0.200	0.183	0.179
	75						0.161	0.165	0.181	0.166	0.164
	100						0.153	0.162	0.174	0.161	0.155
	150						0.146	0.154	0.168	0.150	0.147
	200						0.131	0.142	0.152	0.138	0.135

See notes to Table 4. $\hat{F}$ and t_γ^{ECR} are from Boswijk (1994) and Banerjee *et al.* (1998), respectively. Starred tests are bootstrap counterparts. For DGP(A), $R^2 = 0.25$ and for (A) and (C), $c = -15$.

Table D.6: Small-sample power based on $\hat{F}$ and t_γ^{ECR}, further c

		Bootstrap tests					asymptotic tests				
DGP	T	$\tilde{\chi}_{\mathcal{I}}^{2,*}$	UR^*	naive	$\hat{F}^*$	$t_\gamma^{\text{ECR},*}$	$\hat{F}$	t_γ^{ECR}	naive	$\tilde{\chi}_{\mathcal{I}}^2$	UR_{ψ_1,ψ_2}
(A)	50						0.806	0.788	0.827	0.804	0.805
	75						0.782	0.763	0.807	0.784	0.777
	100						0.743	0.720	0.766	0.741	0.738
	150						0.729	0.701	0.754	0.726	0.724
	200						0.710	0.688	0.741	0.710	0.705
(C)	50						0.465	0.493	0.507	0.484	0.474
	75						0.440	0.470	0.482	0.458	0.451
	100						0.407	0.437	0.448	0.429	0.418
	150						0.388	0.420	0.431	0.409	0.398
	200						0.370	0.409	0.420	0.395	0.384

See notes to Table 4. $\hat{F}$ and t_γ^{ECR} are from Boswijk (1994) and Banerjee *et al.* (1998), respectively. Starred tests are bootstrap counterparts. For DGP(A), $R^2 = 0.25$ and for (A) and (C), $c = -20$.

Appendix E Additional Empirical Results

Table E.1: Frequencies of test results in applied studies and the combination tests: combining $\lambda_{\max}$ and t_γ^{ADF}

number of cases in which...

	single test results...				... in case of conflicting results: 'preferred' test†			
	agree		conflict					
	r	$\neg r$		$\sum$		r	$\neg r$	$\sum$
$\tilde{\chi}^2_{\mathcal{I}}(2) : r$	64	0	30	94	$\tilde{\chi}^2_{\mathcal{I}}(2) : r$	15	10	25
$\tilde{\chi}^2_{\mathcal{I}}(2) : \neg r$	2	52	13	67	$\tilde{\chi}^2_{\mathcal{I}}(2) : \neg r$	8	5	13
$\sum$	66	52	43	161	$\sum$	23	15	38

$\tilde{\chi}^2_{\mathcal{I}}(2)$ abbreviates $\tilde{\chi}^2_{\mathcal{I}}(\lambda_{\max}, t_\gamma^{\mathrm{ADF}})$.
r : test rejects; $\neg r$: test does not reject
† : Test type on which conclusions in the original study were based (see fn. 17).
Absolute frequencies of cointegration-test results for data from Gregory *et al.* (2004). Single tests include Engle-Granger's and Johansen's tests. The $\tilde{\chi}^2_{\mathcal{I}}(2)$ combines these tests as described in Section 3.