Combining Non-Cointegration Tests*

Christian Bayer† Christoph Hanck‡
Universität Bonn Rijksuniversiteit Groningen

First Version: May 2008
This Version: August 2010

Abstract

The local 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.[1] The local 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 individual test results conflict. Moreover it avoids the size distortion inherent in separately applying multiple tests for cointegration to the same data set. We use the new tests to 286 investigate data sets from published cointegration studies. There, in one third of all cases individual tests give conflicting results whereas our meta tests provide an unambiguous test decision.

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

---

*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 Jörg Breitung, Carsten Burhop, Carlo Favero, Christian Gengenbach, Alain Hecq, Walter Krämer, Franz Palm and Jean-Pierre Urbain as well as conference and seminar participants in Neuchâtel, Bonn, Maastricht, Groningen, Amsterdam, Aachen, Münster and Magdeburg for valuable comments and suggestions.

†IIW, Lennéstr. 37, 53113 Bonn, Germany. Tel.: +49 (0)228 73 4073. email: christian.bayer@uni-bonn.de.

‡Department of Economics and Econometrics, Nettelbosje 2, 9747AE Groningen, Netherlands. Tel.: +31 (0)50 3633836, Fax +31 (0)50 3637337. email: c.h.hanck@rug.nl.

[1] STATA and MATLAB code implementing the procedures suggested in this paper is available at www.rug.nl/staff/c.h.hanck/research.

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 software 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 interpretation of test outcomes unclear. More generally speaking, the p-values of different tests are typically not perfectly correlated (Gregory *et al.*, 2004).

This imperfect correlation rules out relying, for example, 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 p-values reflects that the tests are not equivalent. This also has implications for their behavior under the alternative. Specifically, Pesavento (2004) shows that the power ranking of cointegration tests depends crucially on the value of a single nuisance parameter, viz. the squared long-run correlations of error terms driving the variables of the system.

This suggests that suitable combinations of non-cointegration tests might yield a more robust power performance, and possibly even power gains, relative to applying only an individual test. Using the above-mentioned individual tests, the present paper develops such combination tests. In particular, we 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 power, exploiting Pesavento's (2004) results. Besides solving the above-mentioned multiple testing problem, the combination test indeed enjoys a robust power performance over the range of the squared long-run error correlation. Moreover, we explore several alternative combination procedures. For example, Harvey *et al.* (2009) propose a Union-of-Rejections (UR) test to robustify unit root tests against uncertainty over the initial condition. We generalize their idea and apply the generalized UR test to the present testing problem.

Our Fisher-type test turns out to perform very well. It follows closely the power envelope traced out by the best of the underlying individual tests for different values of the nuisance parameter, and even exceeds it when the individual 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 of the performance our 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 size α of the test and are at the same time powerful. The bootstrap versions converge somewhat more quickly to α.

We point out that the above multiple testing problem is pervasive in empirical work and not restricted to testing for cointegration. The meta testing 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.

We employ the new tests to revisit the set of published studies that Gregory *et al.* (2004) examined for 'mixed signals' among cointegration tests, i.e. conflicting test results. We furthermore update their dataset with publications in the JAE from 2001 to 2010. Among other things we find that in one third of all cases individual 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 some empirical motivation and the setup for the non-cointegration tests. Section 3 derives our combination tests. Section 4 presents local power results. Section 5 is devoted to the bootstrap analogs. Section 6 reports Monte Carlo results. Section 7 revisits the published studies. Section 8 concludes. An appendix in an extended working paper version (available from the authors' websites) reports additional results.

The notation is standard. Weak convergence, convergence in probability and in distribution are denoted by $\Rightarrow$, $\to_p$ and $\to_d$. Limits of integration are 0 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 written as $\int \boldsymbol{W}\boldsymbol{W}'$. When a defines b, we write $b := a$ or $a =: b$.

2 Motivation and Setup

2.1 Motivation

Consider the following situation typical for applied macroeconometric work: a researcher wishes to study whether several individually nonstationary time series are cointegrated, but is unsure about which test to use to investigate the null hypothesis of no cointegration. The conclusion of the researcher may then depend on which test is finally employed. For concreteness, we purposely select some well-known examples from the literature taken from the meta study of Gregory *et al.* (2004) and further discussed in Section 7. These examples show that all kinds of mixed signals are possible—some tests rejecting, some tests not rejecting and no test always being among the rejecting ones:[2]

Clements and Hendry (1995) consider a bivariate system of the (inverse) velocity of circulation v and a learning-adjusted measure of the opportunity cost of holding money R, where $v = m - p - y$ when m, p, and y are the natural logarithms of nominal UK M1, the total final expenditure deflator and real total final expenditure respectively. They use quarterly and seasonally adjusted, data running from 1964:l to 1989:2. They find cointegration using the Johansen (1988) procedure (for some detail on the tests see Section 2.3), which we confirm with our implementation of a $\lambda_{\max}$ p-value of 0.0003. However, had one calculated either the residual-based test of Engle and Granger (1987) or the error-correction-based test of Banerjee *et al.* (1998), the p-values would have been 0.6843 and 0.0883, producing no and only very weak evidence in favor of cointegration. The Boswijk (1994) p-value is 0.0001, such that the split of two rejections and two non-rejections would have produced a mixed signal.

Cooley and Ogaki (1996) re-examine, among other things, the long-run equilibrium relationship between consumption and real wages. They use quarterly seasonally adjusted U.S. data running from 1947:1 to 1990:4, as well as three alternative measures of non-durable consumption: nondurable plus services, non-durable, and food. Wages are average hourly compensation in non-agricultural employment. Real wages were constructed by dividing nominal wages by the implicit deflator of each of the three consumption measures used. Using the variable addition test of Park (1990), they find very little evidence against the null hypothesis of cointegration. For the long-run relationship between the logs of real per capita consumption of non-durables and real wages deflated by non-durables prices, the tests of Johansen (1988), Banerjee *et al.* (1998) and Boswijk (1994) yield the opposite conclusion, not rejecting the null of no cointegration with p-values of 0.0744, 0.5630 and 0.5302, respectively. On the other hand, the Engle and Granger (1987) test is consistent with that of Park (1990), producing a p-value of 0.0142. Again, the practitioner, if he is unsure about the choice of test and hence calculates several test statistics, would observe mixed signals regarding the long-run relationship between real wages and consumption.

As a final example, Martens *et al.* (1998) investigate the cost-of-carry model which, through arbitrage once slight equilibrium deviations are exceeded, predicts cointegration between index and index-futures prices. They employ S&P 500 data from May and November 1993, sampled every 15 seconds. Using both Engle and Granger (1987) and Johansen (1988) tests, they find strong evidence in favor of cointegration for all series. For e.g. the May 1993 equilibrium relationship between futures price and futures and 'theoretical' futures prices (index adjusted for the cost-of-carry), we confirm their results with p-values indistinguishable from zero for both tests. However, the error-correction based tests of Banerjee *et al.* (1998) and Boswijk (1994) would not have pro-

[2] We do *not* intend to suggest that the authors of the studies have been in any way strategic in their choice of which cointegration test to report. In fact, since we impose (see Section 7 for details) a common selection procedure regarding trend, lag length as well as sample size determination in all studies, our results could possibly differ from what the authors would have found. Also, cointegration testing may or may not have been a key concern in any of the applied work studied in this paper.

duced (strong) evidence in favor of cointegration, with p-values of 0.1301 and 0.0764, once more leaving the researcher with a mixed signal.

Overall, we confirm Gregory *et al.* (2004) in that mixed signals can easily be found in applied work. Moreover, no uniformly most powerful choice emerges from applied studies. This motivates the need for a combination procedure for single test results, a task to which we turn next. Section 7 revisits the above examples once the combination procedures have been developed.

2.2 Model

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 $\boldsymbol{z}_t = (\boldsymbol{x}_t',\ y_t)'$. Suppose we observe $\boldsymbol{z}_0, \ldots, \boldsymbol{z}_T$. We work with Pesavento's (2004) model:

$$\Delta \boldsymbol{x}_t = \boldsymbol{\tau}_1 + \boldsymbol{v}_{1t} \tag{1a}$$

$$y_t = (\mu_2 - \boldsymbol{\theta}'\boldsymbol{\mu}_1) + (\tau_2 - \boldsymbol{\theta}'\boldsymbol{\tau}_1)t + \boldsymbol{\theta}'\boldsymbol{x}_t + u_t \tag{1b}$$

$$u_t = \rho u_{t-1} + v_{2t} \tag{1c}$$

Equation (1a) defines the dynamics of the regressors, while eqs. (1b) and (1c) describe the (single potential) cointegrating relationship.[3] 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. Further, define the error vector $\boldsymbol{v}_t := (\boldsymbol{v}_{1t}', v_{2t})'$ from eqs. (1a) and (1c) and let $\boldsymbol{\Omega}$ be the long-run covariance matrix of $\boldsymbol{v}_t$. We assume the following.

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

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 (1), cointegration therefore obtains if $|\rho| < 1$. We test the 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 many tests of $\mathcal{H}_0$ against $\mathcal{H}_1$. We consider the residual-based test of Engle and Granger (1987), a system-based test of Johansen (1988), and the error-correction-based tests of Boswijk (1994) and Banerjee *et al.* (1998). Pesavento (2004) shows that, under (1), the local power of these tests only depends on the local-to-unity parameter $c := T(\rho - 1)$ and the correlations of the elements of $\boldsymbol{v}_{1t}$ with v_{2t}. More precisely, partition $\boldsymbol{\Omega}$ conformably with $(\boldsymbol{x}_t', y_t)'$,

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

Define the squared correlation as $R^2 := \boldsymbol{\delta}'\boldsymbol{\delta}$, where $\boldsymbol{\delta} := \boldsymbol{\Omega}_{11}^{-1/2}\boldsymbol{\omega}_{12}\omega_{22}^{-1/2}$ (Kremers *et al.*'s (1992) 'common factor restriction' is an example for $R^2 = 0$). Moreover, we make

[3] Pesavento (2004) shows that (1) does not generally impose weak exogeneity.

Assumption 2. There are no cointegrating relationships among the variables in $\boldsymbol{x}_t$.

This assumption implies the required invertibility of $\boldsymbol{\Omega}_{11}$. Also, partition $\boldsymbol{W} := (\boldsymbol{W}_1', W_2)'$. 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 $W_{12} := \bar{\boldsymbol{\delta}}'\boldsymbol{W}_1 + W_2$, where $\bar{\boldsymbol{\delta}}'\bar{\boldsymbol{\delta}} = \frac{R^2}{1-R^2}$. Furthermore, we distinguish the following cases.

Definition 1. *Depending on the assumptions made about the deterministic components, we have*

- *(i)* $\boldsymbol{W}^d(\lambda) := \boldsymbol{W}(\lambda)$ *and* $J_{12c}^d(\lambda) = J_{12c}(\lambda)$ *if* $\mu_2 - \boldsymbol{\theta}'\boldsymbol{\mu}_1 = 0$, $\boldsymbol{\tau} = \boldsymbol{0}$ *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* $\boldsymbol{\tau} = \boldsymbol{0}$ *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, $\boldsymbol{W}_c^d := (\boldsymbol{W}_1^{d\prime},\ J_{12c}^d)'$ *and* $\boldsymbol{A}_c^d := \int \boldsymbol{W}_c^d \boldsymbol{W}_c^{d\prime}$.

2.3 Individual Cointegration Tests

Engle and Granger (1987)

The Engle-Granger test tests $\mathcal{H}_0$ against the alternative of at least one cointegrating relationship. One first computes $\hat{u}_t$, the residual from a regression of y_t on $\boldsymbol{x}_t$ (and appropriate deterministics $\boldsymbol{d}_t$), and then the t-statistic t_γ^{ADF} on γ in the regression $\Delta\hat{u}_t = \gamma\hat{u}_{t-1} + \sum_{p=1}^{P-1}\nu_p\Delta\hat{u}_{t-p} + \epsilon_t$, where $\sum_{p=1}^{P-1}\nu_p\Delta\hat{u}_{t-p}$ accounts for serial correlation.[4]

Johansen (1988)

The system-based tests of Johansen (1988) test for h cointegrating relationships. In view of $\mathcal{H}_0$, we consider $h = 0$ throughout. One estimates the Vector Error Correction Model (VECM)

$$\Delta\boldsymbol{z}_t = \boldsymbol{\Pi}\boldsymbol{z}_{t-1} + \sum_{p=1}^{P-1}\boldsymbol{\Gamma}_p\Delta\boldsymbol{z}_{t-p} + \boldsymbol{d}_t + \boldsymbol{\varepsilon}_t \tag{2}$$

We employ the $\lambda_{\max}$ test (one could also use λ_{trace}) with test statistic $\lambda_{\max}(h) = -T\ln(1-\hat{\pi}_1)$. Here, $\hat{\pi}_1$ denotes the largest solution to $|\pi\boldsymbol{S}_{11} - \boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}| = 0$, where the $\boldsymbol{S}_{ij}$ are moment matrices of reduced rank regression residuals (Johansen, 1995).

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

Banerjee *et al.* (1998) and Boswijk (1994) develop error correction-based tests. One estimates (by OLS) the equation $\Delta y_t = d_t + \boldsymbol{\pi}_{0x}'\Delta\boldsymbol{x}_t + \varphi_0 y_{t-1} + \boldsymbol{\varphi}_1'\boldsymbol{x}_{t-1} + \sum_{p=1}^{P}(\boldsymbol{\pi}_{px}'\Delta\boldsymbol{x}_{t-p} + \pi_{py}\Delta y_{t-p}) + \epsilon_t$, with P chosen such that ϵ_t is approximately white noise. Banerjee *et al.*'s test statistic t_γ^{ECR} is the t-ratio for $\mathcal{H}_0 : \varphi_0 = 0$, whereas Boswijk's $\hat{F}$ is the Wald statistic for $\mathcal{H}_0 : (\varphi_0,\ \boldsymbol{\varphi}_1')' = \boldsymbol{0}$.

The following Lemma recalls the local distribution of the above tests.

[4]One could also control for serial correlation by the semiparametric approach of Phillips and Ouliaris (1990).

Lemma 1 (Pesavento, 2004). *With the terms as in Definition 1, we have*

i.

$$\begin{aligned} t_\gamma^{\text{ADF}} &\Rightarrow c\,\frac{(\boldsymbol{\eta}_c^{d\prime}\boldsymbol{A}_c^d\boldsymbol{\eta}_c^d)^{1/2}}{(\boldsymbol{\eta}_c^{d\prime}\boldsymbol{D}\boldsymbol{\eta}_c^d)^{1/2}} + \frac{\boldsymbol{\eta}_c^{d\prime}\int \boldsymbol{W}_c^d\,\mathrm{d}\widetilde{\boldsymbol{W}}'\boldsymbol{\eta}_c^d}{(\boldsymbol{\eta}_c^{d\prime}\boldsymbol{A}_c^d\boldsymbol{\eta}_c^d)^{1/2}(\boldsymbol{\eta}_c^{d\prime}\boldsymbol{D}\boldsymbol{\eta}_c^d)^{1/2}} \\ \textit{where}\quad \boldsymbol{\eta}_c^d &:= \left[-\left(\int \boldsymbol{W}_1^{d\prime}J_{12c}^d\right)\left(\int \boldsymbol{W}_1^d\boldsymbol{W}_1^{d\prime}\right)^{-1},\quad 1\right]', \\ \widetilde{\boldsymbol{W}}(\lambda) &:= (\boldsymbol{W}_1'(\lambda),\quad W_{12}(\lambda))', \qquad \boldsymbol{D} := \begin{pmatrix} \boldsymbol{I} & \bar{\boldsymbol{\delta}} \\ \bar{\boldsymbol{\delta}}' & 1+\bar{\boldsymbol{\delta}}'\bar{\boldsymbol{\delta}} \end{pmatrix} \end{aligned}$$

ii. With $\boldsymbol{G}_c := \int \boldsymbol{W}_c^d J_{12c}(\boldsymbol{0}',\ c)$*,*

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

iii.

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

For $c = 0$, all quantities in Lemma 1 reduce to the well-known nuisance-parameter free null distributions. More importantly, all limiting functionals are driven by the same Brownian Motions $\boldsymbol{W}$, such that the lemma allows us to consider the *joint* distribution of the test statistics. Lemma 1 further shows that the different statistics are non-equivalent functionals of $\boldsymbol{W}$, and differentially affected by nuisance parameters under $c < 0$. Hence, as formalized by Pesavento (2004) and further discussed in Section 4, we can expect different tests to be powerful for different values of the nuisance parameter. This forms the basis of the combination procedures presented next.

3 Combination Tests

Under $\mathcal{H}_0$, many of the above statistics are only weakly correlated, even asymptotically (Gregory *et al.*, 2004). Further, Pesavento (2004) shows that the tests differ in their power in different parts of the (c-R^2)-parameter space. In particular, different tests are most powerful in different parts of the parameter space. Thus, a more robust, and possibly even more powerful, combination test can in principle be achieved. To this end, let t_i be the test statistic of cointegration test

$i \in \mathcal{N} := \{1, \dots, N\}$. Take $\xi_i := t_i$ if test i rejects for large values and $-\xi_i = t_i$ if test i rejects for small values. Also, $\Xi_i(x) := \mathrm{P}(\xi_i \geqslant x)$, i.e. one minus test i's asymptotic null distribution function, with P the probability under $\mathcal{H}_0$. The p-value of test i is 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 need a suitable aggregator. One such aggregator is given by Fisher's (1932) famous χ^2 test. Let $\mathcal{I}$, $\mathcal{I} \subseteq \mathcal{N}$, the index set of the individual ξ_i to be aggregated. We then have the following

Proposition 1. *Consider the test statistic*

$$\tilde{\chi}^2_{\mathcal{I}} := -2 \sum_{i \in \mathcal{I}} \ln(p_i). \tag{3}$$

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) $\tilde{\chi}^2_{\mathcal{I}} \to_{\mathrm{p}} \infty$ under $\mathcal{H}_1$ if at least one of the underlying tests is consistent.

Proof. This follows from the continuous mapping theorem (see also White (2000, Prop. 2.2)), for details see Appendix A. □

Part (a) states that the $\tilde{\chi}^2_{\mathcal{I}}$ have well-defined asymptotic null distributions, call them $F_{\mathcal{F}_{\mathcal{I}}}$. These are nuisance-parameter free because of (i) the single ξ_i are nuisance parameter free (cf. e.g. Appendix A) and (ii) the $F_{\mathcal{F}_{\mathcal{I}}}$ take the cross-relation between the ξ_i fully into account. The index-set notation $\mathcal{I}$ serves to emphasize that the $F_{\mathcal{F}_{\mathcal{I}}}$ depend on which and how many tests are combined. Part (b) establishes the consistency of the $\tilde{\chi}^2_{\mathcal{I}}$ tests. 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 ξ_i, $i \in \mathcal{I}$, would be necessary.

Clearly, it would be nice to express the limiting random variable of $\tilde{\chi}^2_{\mathcal{I}}$ under $\mathcal{H}_0$ as an explicit functional of $\boldsymbol{W}$. We conjecture this to be overwhelmingly difficult analytically, bearing in mind that finding closed-form representations is complicated and only possible in special cases even for sums of standard and independent random variables (e.g. Bierens, 2005). Here, the test statistics ξ_i are nonstandard and dependent in a complicated way. However, we can straightforwardly infer and simulate the *joint* distribution of the underlying tests from Lemma 1. This is a standard procedure to find critical values for any single unit root or cointegration test statistic, including the ones combined here. The aggregator $\tilde{\chi}^2_{\mathcal{I}}$ is a continuous function of the t_i, whose null distribution $F_{\mathcal{F}_{\mathcal{I}}}$ can hence be derived by simulation of the functional (3). Table 1 reports 5%-critical values $cv_{\mathcal{I},0.05} := F^{-1}_{\mathcal{F}_{\mathcal{I}}}(0.95)$ for several combinations likely to be relevant in practice (see Table B.2 for other levels).[5] From Prop. 1, reject if $\tilde{\chi}^2_{\mathcal{I}} > F^{-1}_{\mathcal{F}_{\mathcal{I}}}(1-\alpha)$. Since the distributions of the

[5]These are obtained from 100,000 draws from the $F_{\mathcal{F}_{\mathcal{I}}}$, approximating the Wiener processes with suitably normalized Gaussian random walks of length $T = 1,000$.

Table 1: 5%-critical values $cv_{\mathcal{I},0.05}$ for the $\tilde{\chi}^2_{\mathcal{I}}$ tests

	case											
$K-1$	(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)
	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			

5%-critical values for combination tests based on $\tilde{\chi}^2_{\mathcal{I}}$. 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).

underlying cointegration tests depend on $K-1$ as well as the maintained deterministic case (i)-(iii) (cf. Def. 1), that of $\tilde{\chi}^2_{\mathcal{I}}$ will not only depend on $\mathcal{I}$ but also on $K-1$ (reported up to 11) and the maintained case.

For different combinations, the $cv_{\mathcal{I},0.05}$ cluster around 11 for $|\mathcal{I}|=2$, and around 15 for $|\mathcal{I}|=3$. There is little variation across cases. The $cv_{\mathcal{I},0.05}$ fall moderately in $K-1$. It is instructive to compare the $cv_{\mathcal{I},0.05}$ to the $\chi^2(2|\mathcal{I}|)$ critical values. The 5%-critical value is 9.487 for $|\mathcal{I}|=2$, and 12.591 for $|\mathcal{I}|=3$. The $cv_{\mathcal{I},0.05}$ in Table 1 are uniformly larger. This reflects that the ξ_i are generally positively correlated, such that larger $cv_{\mathcal{I},0.05}$ are necessary to construct level-α tests based on (3). Moreover, for each version of $\tilde{\chi}^2_{\mathcal{I}}$, the $cv_{\mathcal{I},0.05}$ are smaller than $-2\sum_{i\in\mathcal{I}}\ln(0.05)$ (which e.g. equals 11.983 for $|\mathcal{I}|=2$). Hence, $\tilde{\chi}^2_{\mathcal{I}}$ rejects whenever all individual tests reject. Moreover, $\tilde{\chi}^2_{\mathcal{I}}$ may reject even if none of the individual tests reject at level α. For example, if $K-1=1$, case (iii) and the p-values of all four tests equal 0.0622, we have $-2\cdot 4\cdot\ln(0.0622)=22.215$ and therefore a rejection using $\tilde{\chi}^2_{\mathcal{I}}$.

Remark 1. The aggregator (3) is only one of many possible choices. Among others, we tried an inverse-normal approach, defined by $1/\sqrt{|\mathcal{I}|}\sum_{i\in\mathcal{I}}\Phi^{-1}(p_i)$, with Φ^{-1} the quantile function of the standard normal distribution. Its performance was however slightly inferior to that of the $\tilde{\chi}^2_{\mathcal{I}}$

tests, to be reported below. The superiority of $\tilde{\chi}^2_{\mathcal{I}}$ may not be surprising in that known optimality results under independence (Littell and Folks, 1971) appear to carry over to the dependent case. Detailed results are available upon request.

3.2 Union-of-Rejections tests

The latter $\min_{i\in\mathcal{I}} p_i$ test is similar to a recent proposal of Harvey *et al.* (2009), who develop 'Union-of-Rejections' (UR) tests to combine standard Dickey-Fuller and GLS-demeaned unit root tests. The UR test also rejects whenever one of the two tests rejects, however suitably adjusting the critical values to ensure a level-α test. The UR test has robust power as the two individual 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 individual cointegration tests. We now use and extend the UR approach to the case of cointegration testing. Denote the individual level-α critical value of test i as $cv_{i,\alpha}$, e.g., $cv_{i,0.05} = |-2.763|$ for t_γ^{ADF}, $K = 2$ and case (i). The 'naive' UR test statistic for $|\mathcal{I}| = 2$ can be written as

$$UR^{\text{naive}}(\xi_1, \xi_2) := \mathbb{I}\{\xi_1 > cv_{1,\alpha}\} + \mathbb{I}\{\xi_1 \leqslant cv_{1,\alpha}\}\mathbb{I}\{\xi_2 > cv_{2,\alpha}\}, \tag{4}$$

with $\mathbb{I}\{A\}$ the indicator function of event A. One would reject $\mathcal{H}_0$ if $UR^{\text{naive}}(\xi_1, \xi_2) = 1$. Of course, the test (4) does not control size.[6] Harvey *et al.* (2009) therefore introduce a scaling constant ψ to modify (4) as follows.

$$UR_\psi(\xi_1, \xi_2) := \mathbb{I}\{\xi_1 > \psi cv_{1,\alpha}\} + \mathbb{I}\{\xi_1 \leqslant \psi cv_{1,\alpha}\}\mathbb{I}\{\xi_2 > \psi cv_{2,\alpha}\}, \tag{5}$$

One rejects if $UR_\psi(\xi_1, \xi_2) = 1$, where 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 UR test. Define the interval $\mathcal{C} := \mathbb{R} \cap [1, \infty)$ and let $\widetilde{\boldsymbol{\psi}} := (\widetilde{\psi}_1, \widetilde{\psi}_2) \in \mathcal{C} \times \mathcal{C} =: \mathcal{C}^2$. The UR statistic then becomes

$$UR_{\psi_{\mathcal{I}}}(\xi_1, \xi_2) := \mathbb{I}\{\xi_1 > \widetilde{\psi}_1 cv_{1,\alpha}\} + \mathbb{I}\{\xi_1 \leqslant \widetilde{\psi}_1 cv_{1,\alpha}\}\mathbb{I}\{\xi_2 > \widetilde{\psi}_2 cv_{2,\alpha}\} \tag{6}$$

One rejects if $UR_{\psi_{\mathcal{I}}}(\xi_1, \xi_2) = 1$. The admissible tuples $\widetilde{\boldsymbol{\psi}}$, denoted $\boldsymbol{\psi}$, are implicitly defined by

$$\mathrm{P}\left(\bigcup_{i=1}^{2} \xi_i > \psi_i cv_{i,\alpha}\right) = \alpha, \tag{7}$$

yielding an entire family of tests. The $\boldsymbol{\psi}$ are identified as, for each $\psi_1 \in \mathcal{C}$, there is exactly one $\psi_2 \in \mathcal{C}$ such that (7) holds. Harvey *et al.*'s (2009) solution $\psi = \psi_1 = \psi_2$ is a special case of the more general approach (7).

Remark 2. Searching over $\mathcal{C}^2$ is without loss of generality. Suppose $\widetilde{\psi}_1 < 1$. We then have $\mathrm{P}(\xi_1 > \widetilde{\psi}_1 cv_{1,\alpha}) =: \widetilde{\alpha}_1 > \alpha$. Also write $\mathrm{P}(\xi_2 > \widetilde{\psi}_2 cv_{2,\alpha}) =: \widetilde{\alpha}_2$. It obtains that (cf. fn. 6)

[6] The null rejection probability of test i is $\mathrm{E}\,\mathbb{I}\{\xi_i > cv_{i,\alpha}\} = \mathrm{P}(\xi_i > cv_{i,\alpha}) = \alpha$. 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 some $UR_{\psi_{\mathcal{I}}}$ tests

$K-1$	case	t_γ^{ADF} and $\lambda_{\max}$ (i)	(ii)	(iii)	$\hat{F}$ and $\lambda_{\max}$ (i)	(ii)	(iii)	$\hat{F}$ and t_γ^{ECR} (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.

$\mathrm{P}\big(\bigcup_{i=1}^{2} \xi_i > \widetilde{\psi}_i cv_{i,\alpha}\big) = \widetilde{\alpha}_1 + \widetilde{\alpha}_2 - \mathrm{P}\big(\bigcap_{i=1}^{2} \xi_i > \widetilde{\psi}_i cv_{i,\alpha}\big) \geqslant \widetilde{\alpha}_1 > \alpha$, because $\mathrm{P}\big(\bigcap_{i=1}^{2} \xi_i > \widetilde{\psi}_i cv_{i,\alpha}\big) \leqslant \widetilde{\alpha}_2$. Hence, one cannot make one test more liberal and still achieve a level-α $UR_{\psi_{\mathcal{I}}}$ test.

The availability of a family of level-α tests raises the practical question of which $\boldsymbol{\psi}$ to select. There is no uniformly most powerful choice. We propose to select $\boldsymbol{\psi}$ such that, subject to (7),

$$\psi_1 = \underset{\widetilde{\psi}_1 \in \mathcal{C}}{\arg\min} \left\{ \frac{\mathrm{P}\big(\xi_1 > \widetilde{\psi}_1 cv_{1,\alpha} \cap \xi_2 > \psi_2 cv_{2,\alpha}\big)}{\min\{\mathrm{P}(\xi_1 > \widetilde{\psi}_1 cv_{1,\alpha}), \mathrm{P}(\xi_2 > \psi_2 cv_{2,\alpha})\}} \right\} \tag{8}$$

It is sufficient to minimize over ψ_1 only, since the corresponding ψ_2 is uniquely determined by (7).[7] We refer to this member of the family of tests as the 'asymmetric' UR test. The tuples $\boldsymbol{\psi}$ 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 $UR_{\psi_{\mathcal{I}}}$ tests as the availability of an entire family of tests provides the opportunity to optimally select a tuple $\boldsymbol{\psi}$, where Harvey *et al.* (2009) impose a restriction, viz. $\psi = \psi_1 = \psi_2$. Further, and more importantly, (8) minimizes 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 (7). Now, since the behavior of the tests under local alternatives changes continuously from that under $\mathcal{H}_0$, making the tests

[7] We add an ϵ to the numerator of (8) 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.

'uncorrelated' leads to many rejections under $\mathcal{H}_1$. (Unreported experiments with other tuples confirm this conjecture. In particular, the power of $UR_{\psi_{\mathcal{I}}}$ then is markedly higher than that of UR_{ψ}.) As for $\tilde{\chi}^2_{\mathcal{I}}$, any correlation between test statistics is automatically taken care of through the respective ψ_i. E.g., the formal similarity of $\hat{F}$ and t_{γ}^{ECR} translates into strong positive correlation. Hence $\hat{F}$ and t_{γ}^{ECR} will seldom disagree. Therefore, only small ψ_i are necessary to satisfy (7). For instance, for case (i) and $K-1=1$ Table 2 reports that $\psi_1+\psi_2=1.077+1.049=2.126$, whereas $\psi_1+\psi_2=1.128+1.101=2.229$ for the apparently much more weakly correlated $\lambda_{\max}$ and $\hat{F}$.

Remark 3. It turns out that the selection rule (8) satisfies

$$\mathrm{P}(\xi_1 > \psi_1 cv_{1,\alpha}) = \mathrm{P}(\xi_2 > \psi_2 cv_{2,\alpha}) \tag{9}$$

for all combinations considered in Table 2.[8] Under (9), the $UR_{\psi_{\mathcal{I}}}$ test is equivalent to a minimum p-value test, defined by $\min_{i\in\mathcal{I}} p_i$. This test is a direct fix to the 'naive' strategy that rejects whenever one of the individual tests rejects. The critical values of the $\min_{i\in\mathcal{I}} p_i$ test yield the level $\alpha' < \alpha$ at which one needs to test to avoid the oversizedness of the 'naive' approach. Table B.1 provides critical values for the $\min_{i\in\mathcal{I}} p_i$ test. (Incidentally, we find $\alpha' \gg \alpha/|\mathcal{I}|$ so that $\min_{i\in\mathcal{I}} p_i$ is more powerful than a Bonferroni-type multiple test.)

To show that $UR_{\psi_{\mathcal{I}}}$ and $\min_{i\in\mathcal{I}} p_i$ are indeed equivalent, we first show that the min-test belongs to the family of $UR_{\psi_{\mathcal{I}}}$ 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 $\Xi_i(x)$ are decreasing in x). 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, for $\psi_i := \Xi_i^{-1}\left(F_{\min}^{-1}(\alpha)\right)/cv_{i,\alpha}$,

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

Under $\mathcal{H}_0$, we have $\mathrm{P}(\xi_1 > \Xi_1^{-1}(F_{\min}^{-1}(\alpha)) \vee \xi_2 > \Xi_2^{-1}(F_{\min}^{-1}(\alpha))) = \alpha$. Thus, the min-test is a $UR_{\psi_{\mathcal{I}}}$ test. It remains to establish that it is the *only* $UR_{\psi_{\mathcal{I}}}$ test satisfying (9). By construction,

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

Uniqueness follows from monotonicity of the Ξ_i.

Remark 4. One can also relax Harvey *et al.*'s restriction to combine $|\mathcal{I}| = 2$ tests. An $|\mathcal{I}|$-dimensional UR test is, analogously to (6), defined by $\mathrm{P}\left(\bigcup_{i=1}^{|\mathcal{I}|} \xi_i > \psi_i cv_{i,\alpha}\right) = \alpha$. Of course,

[8] To see why, write the numerator of (8) 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 (7), (8) 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 (7) are sufficiently steep to produce the 'corner solution' (9).

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

4 Large Sample Results

We now report the large-sample power of the tests discussed in Sections 2 and 3. As for single cointegration tests, the local power functions of $\tilde{\chi}^2_{\mathcal{I}}$ and $UR_{\psi_{\mathcal{I}}}(\xi_1, \xi_2)$ are not available in closed form. Following Pesavento (2004), these functions are therefore approximated by simulating the distributions given in Lemma 1 and Section 3. They give the probability that ξ_i and $\tilde{\chi}^2_{\mathcal{I}}$ exceed their level-α critical values, and the probability that $UR_{\psi_{\mathcal{I}}}(\xi_1, \xi_2) = 1$ (cf. (6)). We draw 25,000 replications of the functionals, for $T = 1{,}000$. We consider $c \in \{0, -1, -2, \ldots, -30\}$, $R^2 \in \{0, 0.05, 0.1, \ldots, 0.95\}$ and $K - 1 \in \{1, \ldots, 5\}$.

Table 3 reports the local power of several combination tests as well as the corresponding individual tests for case (ii) (cf. Appendix C for cases (i) and (iii)).[9] Figure 1 plots the tests' power against R^2, for $c = -15$ and $K - 1 = 1$; additional results are available. We replicate Pesavento's finding that t_γ^{ECR} is the best individual test for small R^2. The power of all tests but t_γ^{ADF} increases 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 similar local power. The combination tests (we initially focus on the case $|\mathcal{I}| = 2$ for expositional clarity) perform very well, tracking the better test very closely. Their power curves sometimes even lie above that of the underlying tests. This is best seen in the lower panel, where the performance of the underlying tests t_γ^{ADF} and $\lambda_{\max}$ differs strongly. The upper panel shows that, unsurprisingly, the power of the combination tests differs relatively less from that of either underlying test if these perform similarly. Yet, $UR_{\psi_{\mathcal{I}}}(\hat{F}, t_\gamma^{\text{ECR}})$ and $\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\text{ECR}})$ are again closer to the better underlying test (typically $\hat{F}$) whenever there are discernible differences.

Figures 2-3 plot the tests' power against $-c$, for $R^2 = 0.25, 0.7$. All tests become more powerful as the distance c to $\mathcal{H}_0$ increases, although the speed differs substantially. For large R^2 and $c = -15$, the power of $\lambda_{\max}$, $\tilde{\chi}^2_{\mathcal{I}}(t_\gamma^{\text{ADF}}, \lambda_{\max})$ and $UR_{\psi_{\mathcal{I}}}(t_\gamma^{\text{ADF}}, \lambda_{\max})$ is more than three times larger than that of t_γ^{ADF}. The combination tests are again close to the better of the individual tests. Of course, when the difference between the individual tests is large, as in the lower panel of Figure 3, the power distance to the best individual test is somewhat larger—but still a lot smaller than that to the worse individual test. Thus, the combination tests cheaply insure against selecting an inferior test, in that one never sacrifices much power, and potentially gains a lot. Moreover, for $R^2 = 0.25$, both $\tilde{\chi}^2_{\mathcal{I}}(t_\gamma^{\text{ADF}}, \lambda_{\max})$ and $UR_{\psi_{\mathcal{I}}}(t_\gamma^{\text{ADF}}, \lambda_{\max})$ even outperform both constituent tests. Note from Figure 1 that the power curves of the constituent tests t_γ^{ADF} and $\lambda_{\max}$ intersect at $R^2 \approx 0.25$. Thus, combination tests appear to outperform the constituent tests when the latter

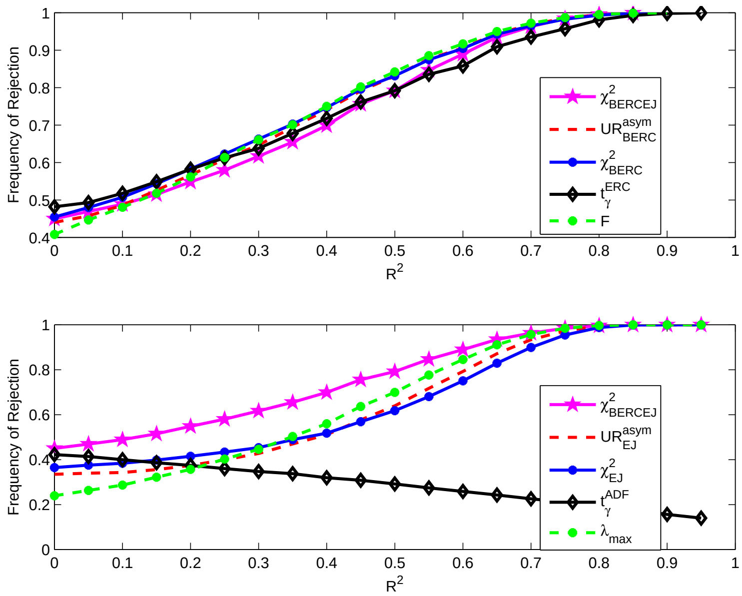
[9] Critical values are obtained under $R^2 = 0$. Thus, the slight deviations from α under $R^2 \neq 0$ and $c = 0$ are due to simulation variability.

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
$UR_{\psi_{\mathcal{I}}}(\hat{F}, t_\gamma^{\mathrm{ECR}})$	0.050	0.102	0.229	0.440	0.690
$UR_{\psi_{\mathcal{I}}}(t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	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
$UR_{\psi_{\mathcal{I}}}(\hat{F}, t_\gamma^{\mathrm{ECR}})$	0.051	0.114	0.310	0.609	0.846
$UR_{\psi_{\mathcal{I}}}(t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	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}	0.051	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
$UR_{\psi_{\mathcal{I}}}(\hat{F}, t_\gamma^{\mathrm{ECR}})$	0.053	0.158	0.517	0.831	0.964
$UR_{\psi_{\mathcal{I}}}(t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	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}	0.052	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
$UR_{\psi_{\mathcal{I}}}(\hat{F}, t_\gamma^{\mathrm{ECR}})$	0.054	0.365	0.859	0.985	0.999
$UR_{\psi_{\mathcal{I}}}(t_\gamma^{\mathrm{ADF}}, \lambda_{\max})$	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}	0.052	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

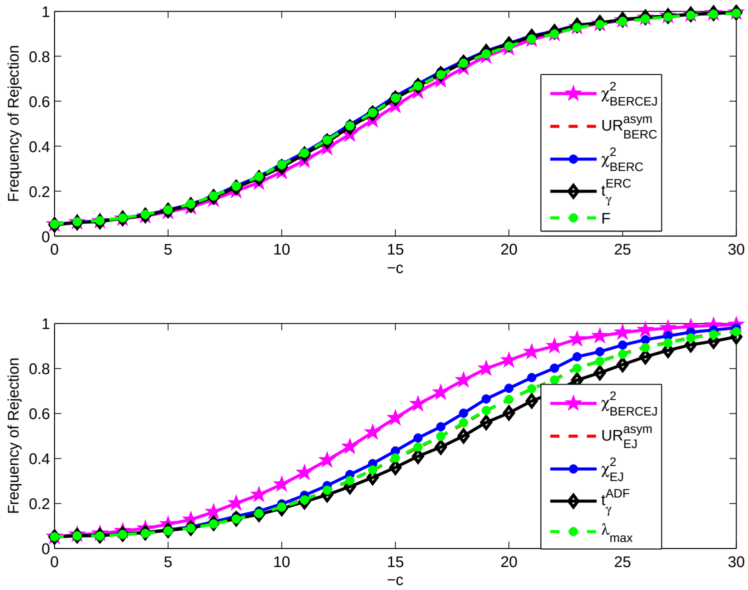
Case (*ii*). $\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\mathrm{ECR}})$ is our Fisher test (3) based on Boswijk's and Banerjee *et al.*'s tests, and $UR_{\psi_{\mathcal{I}}}(\hat{F}, t_\gamma^{\mathrm{ECR}})$ is the corresponding Union-of-Rejections test (6). The other combination tests are defined analogously. See also notes to Table 1.

Figure 1: Local asymptotic power as a function of R^2, $c = -15$



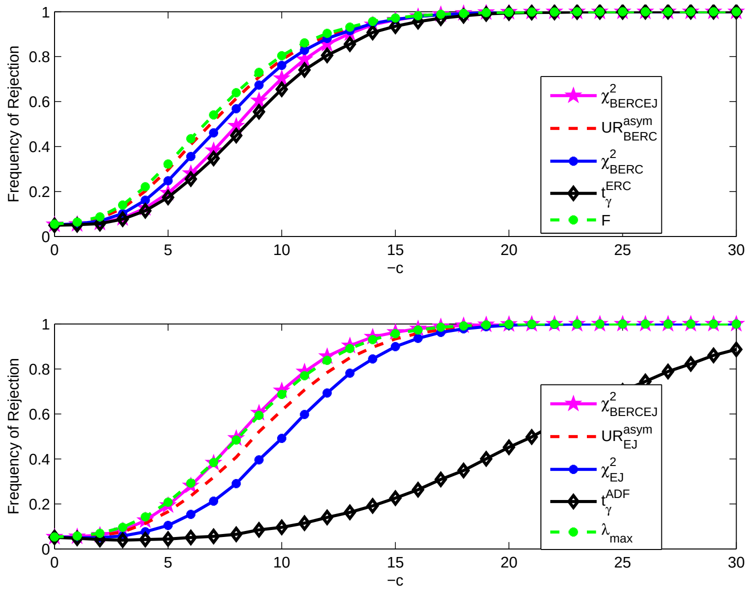
Results are for the demeaned case (ii). χ^2_{BERC} is our Fisher test (3) based on Boswijk's and Banerjee *et al.*'s tests. χ^2_{EJ} is based on Engle and Granger's and Johansen's tests. χ^2_{BERCEJ} combines all four tests. $UR^{\text{asym}}_{\text{BERC}}$ and $UR^{\text{asym}}_{\text{EJ}}$ are the corresponding asymmetric $UR_{\psi_{\mathcal{I}}}$ tests (6). The individual tests' power curves are for comparison.

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



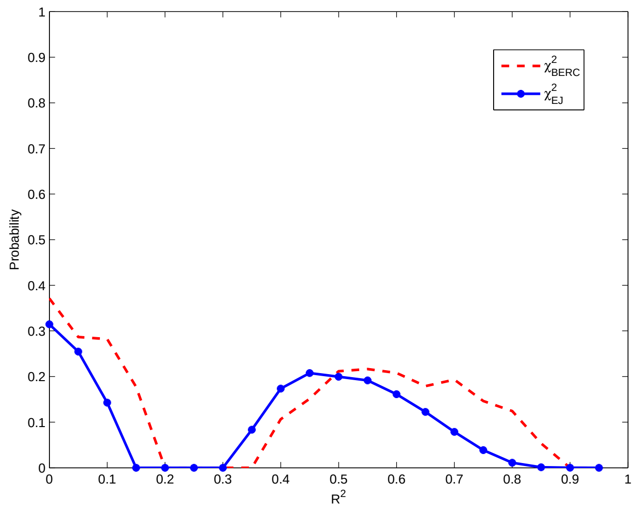
See notes to Figure 1.

Figure 3: Local asymptotic power as a function of $-c$, $R^2 = 0.7$



See notes to Figure 1.

Figure 4: 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 $c = -15$.

are equally powerful. Intuitively, this is because the ξ_i will then often be individually just too small to discard $\mathcal{H}_0$, but the evidence from the two taken together suffices to reject. This effect becomes more pronounced with $K-1$, cf. Figure 2 with C.3.

Table 3 shows that $\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t^{\text{ECR}}_{\gamma}, t^{\text{ADF}}_{\gamma}, \lambda_{\max})$ outperforms $\tilde{\chi}^2_{\mathcal{I}}(t^{\text{ADF}}_{\gamma}, \lambda_{\max})$, but is (slightly) outperformed by $\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t^{\text{ECR}}_{\gamma})$. This is not surprising as $\hat{F}$ and t^{ECR}_{γ} generally perform best under (1). Section 6 studies other relevant DGPs under which $\lambda_{\max}$ and t^{ADF}_{γ} outperform $\hat{F}$ and t^{ECR}_{γ}. Consequently $\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t^{\text{ECR}}_{\gamma}, t^{\text{ADF}}_{\gamma}, \lambda_{\max})$ then outperforms $\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t^{\text{ECR}}_{\gamma})$. As such, it would be wrong to recommend routine use of $\hat{F}$ and t^{ECR}_{γ} or $\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t^{\text{ECR}}_{\gamma})$. Overall, this suggests that the transparent strategy to combine all available tests can be recommended for empirical practice. On the other hand, t^{ADF}_{γ} and $\lambda_{\max}$ are still the most widely used tests, such that providing a detailed discussion how to combine the two likely is relevant for practitioners.

Comparing the performance of $\tilde{\chi}^2_{\mathcal{I}}$ and $UR_{\psi_{\mathcal{I}}}$, we find that the former are somewhat more powerful when both constituent tests have relatively high power. The $UR_{\psi_{\mathcal{I}}}$ tests outperform the $\tilde{\chi}^2_{\mathcal{I}}$ tests when there is a large difference in power between the individual tests, in particular if the weaker one has low absolute power. This is intuitive as $UR_{\psi_{\mathcal{I}}}$ looks for (at least) one individual test indicating that $\mathcal{H}_1$ holds, effectively ignoring the less powerful test once the more powerful one rejects. On the other hand, $\tilde{\chi}^2_{\mathcal{I}}$ combines evidence from both tests, such that a test with low power can tilt $\tilde{\chi}^2_{\mathcal{I}}$ towards a non-rejection of $\mathcal{H}_0$. If both tests are at least moderately powerful, $\tilde{\chi}^2_{\mathcal{I}}$ will combine that evidence to produce a rejection of $\mathcal{H}_0$.

Remark 5. As discussed above, some individual 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 test given the estimate $\hat{R}^2$. However, as (unlike in Elliott *et al.*, 2005) $\boldsymbol{\theta}$ is assumed unknown and several 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 (Pesavento, 2007). Moreover, the above results show that the combination tests are never much less, and sometimes even more, powerful than the best individual test. 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 for T finite, estimate R^2 without error. Hence, a pretest would sometimes select the *less* powerful test. A pretest would therefore likely have less power than the strategies advocated here. To further illustrate this point, let q denote the probability to select the inferior test. As an example, consider from Table 3 $\lambda_{\max}$, t^{ADF}_{γ} and $\tilde{\chi}^2_{\mathcal{I}}(t^{\text{ADF}}_{\gamma}, \lambda_{\max})$ for $R^2 = 0.75$ and $c = -15$. A pretest, if available, would need to select the worse test (t^{ADF}_{γ}) in only $q = (0.954 - 0.984)/(0.210 - 0.984) \times 100 \approx 4\%$ of the cases for it to be inferior to $\tilde{\chi}^2_{\mathcal{I}}(t^{\text{ADF}}_{\gamma}, \lambda_{\max})$. For $\tilde{\chi}^2_{\mathcal{I}}(t^{\text{ADF}}_{\gamma}, \lambda_{\max})$, $\tilde{\chi}^2_{\mathcal{I}}(t^{\text{ECR}}_{\gamma}, \hat{F})$, Figure 4 plots q against R^2 (for $c = -15$ and $K - 1 = 1$). We see that q never exceeds 0.3, and even find $q = 0$ for $R^2 \in [0.15, 0.3] \cup (0.85, 1)$ (for $\tilde{\chi}^2_{\mathcal{I}}(t^{\text{ADF}}_{\gamma}, \lambda_{\max})$), reflecting that $\tilde{\chi}^2_{\mathcal{I}}$ is sometimes as or *more* powerful than even a perfect pretest. Moreover, q is always substantially smaller than 0.5, implying that the $\tilde{\chi}^2_{\mathcal{I}}$ tests uniformly outperform the strategy of randomly selecting one of the underlying tests.

Remark 6. It is also tempting to develop 'R^2-weighted' versions of the meta tests. Consider e.g. $\tilde{\chi}^2_{\mathcal{I},R^2} := -2\sum_{i\in\mathcal{I}} \varpi_i(R^2)\ln(p_i)$, where ϖ_i is a weight function such that $\sum_{i\in\mathcal{I}} \varpi_i(R^2) = |\mathcal{I}|$ (in (3), each i implicitly has $\varpi_i(R^2) = 1$). Again, an estimator $\hat{R}^2$ would be necessary. Moreover, if the weights ϖ_i depend on R^2, so would the null distribution of a weighted meta test like $\tilde{\chi}^2_{\mathcal{I},R^2}$. Hence, $\tilde{\chi}^2_{\mathcal{I},R^2}$ would no longer be nuisance-parameter free, making such an approach unattractive.

5 Bootstrap Analogs

The previous results rely entirely on asymptotic theory. The combination tests cannot be expected not to share small-sample deficiencies of the underlying 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 above local power curves are effectively approximations to the tests' finite-sample power curves. The bootstrap has recently been successfully employed to improve the small-sample behavior of cointegration tests (Swensen, 2006; Palm *et al.*, 2010). 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 $\tilde{\chi}^2_{\mathcal{I}}$, 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 DGP to then generate integrated but non-cointegrated time series. We propose the following Algorithm to estimate the finite-sample distribution of $\tilde{\chi}^2_{\mathcal{I}}$.

Algorithm 1.

1. *Estimate the unrestricted VAR* $\boldsymbol{z}_t = \sum_{p=1}^{P}\boldsymbol{\Phi}_p\boldsymbol{z}_{t-p} + \boldsymbol{d}_t + \boldsymbol{\varepsilon}_t$ *to obtain 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* (2) *(see e.g. Lütkepohl (2005, p. 247) for the procedure).*[10]
2. *Check that the system has no explosive root, i.e.* $\|z\| > 1$, *by solving* $\det\{\hat{\boldsymbol{B}}(z)\} = 0$, *where*
$$\hat{\boldsymbol{B}}(z) := \boldsymbol{I}_K - \hat{\boldsymbol{\Gamma}}_1 z - \cdots - \hat{\boldsymbol{\Gamma}}_{P-1}z^{P-1}.\text{[11]} \tag{11}$$
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* $\{\boldsymbol{\varepsilon}^*_{t,b}\}^{b=1,\ldots,B}_{t=P,\ldots,T}$, *construct* B *series of pseudo observations* $\boldsymbol{z}^*_{t,b}$ *from* $\Delta\boldsymbol{z}^*_{t,b} = \hat{\boldsymbol{d}}_t + \sum_{p=1}^{P-1}\hat{\boldsymbol{\Gamma}}_p\Delta\boldsymbol{z}^*_{t-p,b} + \boldsymbol{\varepsilon}^*_{t,b}$. *For the initial observations, set* $\boldsymbol{z}^*_{t,b} = \boldsymbol{z}_t, t = 0,\ldots,P-1$.[12]

[10] One could alternatively estimate a VAR for $\Delta\boldsymbol{z}_t$, imposing $\mathcal{H}_0$ (cf. Swensen, 2006). However, as Paparoditis and Politis (2003) show for unit-root tests, imposing $\mathcal{H}_0$ may lead to a power loss.

[11] See Swensen (2006, Remark 1) and Johansen (1995, p. 71) for a discussion of this condition. 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 (11) in our context.

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

5. *Compute the vector of test statistics* $\boldsymbol{\xi}_b^* := (\xi_{1,b}^*, \ldots, \xi_{|\mathcal{I}|,b}^*)'$*, for each* $b = 1, \ldots, B$*.*
6. *Estimate the distribution function of each test statistic as* $B^{-1} \sum_{h=1}^{B} \mathbb{I}\{\xi_{i,h}^* \leq x\} =: 1 - \Xi_i^*(x)$ *and calculate the corresponding p-values* $p_{i,b}^* := \Xi_i^*(\xi_{i,b}^*)$*. Correspondingly, calculate the p-values of the test statistics* ξ_i *on the original data* $\boldsymbol{z}_{i,t}$ *by* $p_i^* := \Xi_i^*(\xi_i)$*.*
7. *Obtain the corresponding aggregate* $\tilde{\chi}_{\mathcal{I}}^2$ *test statistic* $\tilde{\chi}_{\mathcal{I},b}^{2,*} = -2 \sum_{i=1}^{|\mathcal{I}|} \ln\big(p_{i,b}^*\big)$*.*
8. *Estimate the distribution function* $F_{\mathcal{F}_{\mathcal{I}}^*}$ *of the* $\tilde{\chi}_{\mathcal{I},b}^{2,*}$ *by* $\hat{F}_{\mathcal{F}_{\mathcal{I}}^*}(x) := B^{-1} \sum_{h=1}^{B} \mathbb{I}\big\{\tilde{\chi}_{\mathcal{I},h}^{2,*} \leq x\big\}$*.*

This provides us with a bootstrap version of the $\tilde{\chi}_{\mathcal{I}}^2$ test, $\tilde{\chi}_{\mathcal{I}}^{2,*} = -2 \sum_{i=1}^{|\mathcal{I}|} \ln(p_i^*)$, where we reject $\mathcal{H}_0$ at level α if $\tilde{\chi}_{\mathcal{I}}^{2,*}$ exceeds the $(1-\alpha)$-quantile of $\hat{F}_{\mathcal{F}_{\mathcal{I}}^*}$.

Heuristically, the method can be expected to work as follows. Swensen (2006) analytically proves that his bootstrap procedure (i.e. steps 1-4 in Algorithm 1) yield pseudo-observations $\boldsymbol{z}_{t,b}^*$ which have a representation asymptotically equivalent to the true DGP. Moreover, he proves that steps 5 and 6 consistently estimate the null distribution of the Johansen λ_{trace} test, hence yielding consistent estimates of p-values. Therefore, we can expect the proposition to carry over to the cointegration tests mentioned above, as these essentially also rely on the availability of suitable $\boldsymbol{z}_{t,b}^*$. 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, ensures a well-defined *joint* distribution of the statistics $\boldsymbol{\xi}$. That joint distribution can be consistently estimated with Algorithm 1 under fairly weak regularity conditions (Horowitz, 2001). We provide extensive numerical support for this argument in Section 6.[13]

Remark 7. Algorithm 1 is only about as computationally demanding as Swensen's (2006). It also requires resampling B pseudo-observations, and no double bootstrapping. The difference to Swensen's algorithm is that $|\mathcal{I}|$ instead of one statistic (λ_{trace}) need to be calculated for each b.

Remark 8. In view of the equivalence of the $UR_{\psi_{\mathcal{I}}}$ and min-test established in Remark 3, a version of Algorithm 1 also provides bootstrap $UR_{\psi_{\mathcal{I}}}$ tests by bootstrapping the distribution of $\min_{i \in \mathcal{I}} p_i$. We reject $\mathcal{H}_0$ if $\min_{i \in \mathcal{I}} p_i < \hat{F}_{\min}^{*,-1}(\alpha)$, the α-quantile of the bootstrap distribution $\hat{F}_{\min}^*$.

6 Monte Carlo Experiments

6.1 Setup

We now study the finite-sample properties of the tests in a series of Monte Carlo experiments.As shown above, different tests for cointegration differ in their power against different points in the (c-R^2)-space of the alternative hypothesis. Further, e.g. Johansen's $\lambda_{\max}$ test can be expected to be relatively more powerful if $\Delta \boldsymbol{z}_t$ is indeed generated by a finite order VECM. Since our tests combine 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:

[13] Appendix D describes an alternative bootstrap test that we found to have slightly higher power in unreported simulations. As that approach requires stronger theoretical assumptions, we advocate using $\tilde{\chi}_{\mathcal{I}}^{2,*}$.

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

The autoregressive coefficient $\rho_T = 1 + c/T$. $\mathcal{H}_0$ is obtained when $c = 0$, whereas we parameterize $\mathcal{H}_1$ by $c = -15$.[14] The errors $\boldsymbol{v}_t$ are drawn from

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

For $R^2 = \delta^2$, we select $R^2 = \{0, 0.25, 0.5, 0.75\}$. DGP(A) closely follows model (1). We consider two additional DGPs which are not special cases of model (1) under $\mathcal{H}_1$ in order to investigate the generality of this setup. In particular, we are interested to see whether the favorable asymptotic results for t_γ^{ECR}, $\hat{F}$ and $\tilde{\chi}^2_{\mathcal{I}}(\hat{F}, t_\gamma^{\text{ECR}})$ from Section 4 carry over to other parameterizations. First, we consider

$$\text{DGP(B):} \quad \Delta \boldsymbol{z}_t = \boldsymbol{\Pi}_T \boldsymbol{z}_{t-1} + \boldsymbol{\Gamma} \Delta \boldsymbol{z}_{t-1} + \boldsymbol{u}_t, \quad \text{where } \boldsymbol{\Gamma} = 0.2\boldsymbol{I}_2 \text{ and } \boldsymbol{u}_t = (u_{1t}, u_{2t})' \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I}_2)$$

For (B) $\mathcal{H}_0$ is obtained when $\boldsymbol{\Pi}_T = \mathbf{0}$, whereas we parameterize $\mathcal{H}_1$ by $\boldsymbol{\Pi}_T = \frac{c}{T}(0 \quad 1)'(1 \quad -1)$. Elliott *et al.* (2005) show that variants of DGP(A) and (B) are closely related, yet they differ in how short-run dynamics enter the DGPs. DGP(B) can be written as

$$[(\boldsymbol{I} - \boldsymbol{\Gamma} L)(1 - L) - (\rho_T - 1)\boldsymbol{\Pi}_T L]\boldsymbol{z}_t = \boldsymbol{u}_t \tag{12}$$

whereas an equivalent way of writing model (1) for the corresponding case of a VAR(2) is

$$[(\boldsymbol{I} - \boldsymbol{\Phi} L)(1 - L) - (\boldsymbol{I} - \boldsymbol{\Phi} L)(\rho_T - 1)\boldsymbol{\Pi}_T L]\boldsymbol{z}_t = \boldsymbol{u}_t \tag{13}$$

(see Pesavento's eq. (2.1)). Because $\boldsymbol{I} - \boldsymbol{\Phi} L$ also affects the error-correction term in (13) it is not possible to find a $\boldsymbol{\Phi}$ such that (12) and (13) imply the same dynamics in our parametrization. This also implies that it is no longer directly possible to infer the R^2s associated with DGP(B).

Next, we consider

$$\text{DGP(C):} \quad y_t + \eta x_t = a_{1t}, \quad y_t + x_t = a_{2t}, \quad \text{where} \quad \eta = -1/2 \quad \text{and}$$
$$a_{1t} = a_{1t-1} + u_{1t}, \; a_{2t} = \rho_T a_{2t-1} + u_{2t}, \quad \boldsymbol{u}_t \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I}_2)$$

where ρ_T is as in (A).[15] DGP(C) can be rearranged to (cf. Elliott *et al.*, 2005)

$$\Delta \boldsymbol{z}_t = \frac{2}{3} \begin{pmatrix} \rho - 1 \\ \frac{1}{2}(\rho - 1) \end{pmatrix} (1 \quad 1)\boldsymbol{z}_{t-1} + \tilde{\boldsymbol{v}}_t$$

Hence, DGP(C) does not impose that x_t has an exact unit root under the local alternative and thus is not covered by the assumptions underlying Pesavento's model (1). Hence, the local power

[14] Power results for other c are given in Appendix E.

[15] Of course, Granger's representation theorem would allow us to write DGP(C) in a VECM form. However, error terms would be correlated, the matrix $\boldsymbol{\Pi}$ would have no rows of zeros under $\mathcal{H}_1$ and $\boldsymbol{\Gamma}$ would equal $\mathbf{0}$.

curves derived in Section 4 do not necessarily hold under DGP(C). On the other hand, DGP(C), first considered by Engle and Granger (1987), offers a plausible parametrization of cointegration. It is therefore of interest in its own right, but also as a robustness check on the generality of the findings from Section 4, to study the performance of the single and combination tests under DGP(C).

Remark 9. Appendix E provides additional simulations showing that all qualitative findings remain intact when generating DGPs (B) and (C) with an unrestricted variance-covariance matrix as in (A), $\boldsymbol{\Omega} = \left(\begin{smallmatrix} 1 & \delta \\ \delta & 1 \end{smallmatrix}\right)$. Moreover, we demonstrate that non-diagonality of neither $\boldsymbol{\Pi}$ nor $\boldsymbol{\Gamma}$ affects the conclusions.

All three DGPs are widely used in Monte Carlo studies of cointegration tests. See e.g. Pesavento (2004, 2007) for (A), Swensen (2006) for (B), or Engle and Granger (1987), Haug (1996) and Gregory *et al.* (2004) for (C). The DGPs are local, such that power ought to remain roughly constant in T. 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 $B = 10,000$ resamples. To keep the setup simple, we initially combine $|\mathcal{I}| = 2$ underlying tests (see Section 6.3 for extensions). In particular, we select Johansen's (1988) $\lambda_{\max}$ test and Engle and Granger's (1987) t^{ADF}_{γ} test. We opt for these tests as they are widely used in applied work. 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 (1998) t^{ECR}_{γ} test.

To investigate the performance of the new tests, we compare them to the following cointegration tests: First, the standard t^{ADF}_{γ}, $\lambda_{\max}$, t^{ECR}_{γ} and $\hat{F}$ tests, where we reject $\mathcal{H}_0$ if the test statistics exceed the asymptotic level-α critical value.[16] Second, we investigate bootstrap versions of the tests (denoted $t^{\text{ADF},*}_{\gamma}$, $\lambda^*_{\max}$, $t^{\text{ECR},*}_{\gamma}$ and $\hat{F}^*$), which are by-products of Algorithm 1. Third, we compute a test that rejects whenever at least one of a set of individual tests rejects. We call this test 'naive' as it ignores the multiple-testing nature of the problem. This test reveals the size distortion incurred by selecting the most rejective from a set of cointegration tests.

The tests' implementation requires choosing a lag length $\hat{P}$ to capture autocorrelation. In practice this is often done via selection criteria (e.g. Lütkepohl, 2005). To reduce the computational burden we waive this option and use the correct lag order, i.e. $P = 0$ in (A) and (C) and $P = 1$ in (B).[17] All tests are based on case (*iii*).

[16] In the case of the t^{ADF}_{γ} test we follow the standard practice of using MacKinnon (1996)-type critical values. We also studied Phillips and Ouliaris (1990) and λ_{trace} tests. Since these are very strongly correlated with t^{ADF}_{γ} and $\lambda_{\max}$ resp. (Gregory *et al.*, 2004), adding these to $\tilde{\chi}^2_{\mathcal{I}}$ or $UR_{\psi_{\mathcal{I}}}$ barely affects the latter's performance.

[17] For t^{ADF}_{γ}, we select $P = 1$ under (B) too, as this yields a sufficiently accurate approximation for $\boldsymbol{\Gamma} = 0.2\boldsymbol{I}_2$. For (A) and (C), we take $P = 0$.

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

DGP	T	Bootstrap tests $\lambda^*_{\max}$	$t_{\gamma}^{\text{ADF},*}$	naive*	$\tilde{\chi}^{2,*}_{\mathcal{I}}$	$UR^*_{\psi_{\mathcal{I}}}$	asymptotic tests $\lambda_{\max}$	t_{γ}^{ADF}	naive	$\tilde{\chi}^{2}_{\mathcal{I}}$	$UR_{\psi_{\mathcal{I}}}$
(A)	50	0.051	0.048	0.078	0.051	0.048	0.054	0.080	0.113	0.062	0.084
	75	0.044	0.042	0.072	0.042	0.040	0.055	0.077	0.110	0.059	0.080
	100	0.048	0.048	0.076	0.049	0.046	0.054	0.075	0.111	0.056	0.072
	150	0.046	0.046	0.079	0.045	0.048	0.054	0.063	0.099	0.049	0.069
	200	0.055	0.050	0.086	0.059	0.057	0.048	0.058	0.090	0.047	0.059
(B)	50	0.052	0.050	0.080	0.051	0.049	0.067	0.069	0.108	0.063	0.077
	75	0.050	0.050	0.078	0.049	0.047	0.060	0.062	0.098	0.060	0.065
	100	0.047	0.045	0.075	0.046	0.046	0.061	0.059	0.093	0.060	0.066
	150	0.050	0.047	0.073	0.050	0.046	0.057	0.060	0.090	0.057	0.061
	200	0.050	0.055	0.081	0.053	0.057	0.057	0.063	0.092	0.063	0.063
(C)	50	0.045	0.054	0.083	0.050	0.048	0.053	0.081	0.114	0.060	0.081
	75	0.044	0.043	0.073	0.041	0.041	0.055	0.076	0.110	0.055	0.077
	100	0.046	0.051	0.082	0.048	0.049	0.054	0.069	0.103	0.054	0.072
	150	0.048	0.050	0.082	0.047	0.048	0.054	0.064	0.099	0.049	0.070
	200	0.055	0.051	0.088	0.059	0.055	0.048	0.058	0.089	0.044	0.060

Rejection rates at nominal level of 5%. 5,000 replications and 10,000 bootstrap replications. t_{γ}^{ADF} and $\lambda_{\max}$ refer to Engle and Granger (1987) and Johansen (1988) tests, $t_{\gamma}^{\text{ADF},*}$ and $\lambda^*_{\max}$ are their bootstrap counterparts. naive rejects when $t_{\gamma}^{\text{ADF},*}$ or $\lambda^*_{\max}$ or both reject. $UR_{\psi_{\mathcal{I}}}$ is the test defined by (6) and (8) and and UR^* is the bootstrap counterpart. $\tilde{\chi}^2_{\mathcal{I}}$ is the Fisher test (3) and $\tilde{\chi}^{2,*}_{\mathcal{I}}$ is its bootstrap counterpart. (UR^* and $\tilde{\chi}^{2,*}_{\mathcal{I}}$ are described in Algorithm 1.)

6.2 Results

Table 4 reports the small sample size of the tests based on $\lambda_{\max}$ and t_{γ}^{ADF} at $\alpha = 0.05$. Results for DGP(A) are based on $R^2 = 0.25$.[18] As expected, the 'naive' test is oversized. Its size exceeds that of the individual tests by approximately 3 - 4 percentage points.[19] All other tests control size quite well. Both $UR_{\psi_{\mathcal{I}}}$ and, to a lesser extent, $\tilde{\chi}^2_{\mathcal{I}}$ exhibit a slight upward size distortion for small T, due to a distortion of t_{γ}^{ADF} for small T. However, this size distortion vanishes for larger T. The bootstrap tests approach the nominal size more quickly, which reflects that the bootstrap distributions $\hat{F}_{\mathcal{F}^*_{\mathcal{I}}}$ generally are somewhat more accurate approximations to the unknown-finite sample distributions than the asymptotic ones $F_{\mathcal{F}_{\mathcal{I}}}$.

Table 5 reports the non-size adjusted small sample power of the $\lambda_{\max}$ and t_{γ}^{ADF}-based tests at the level α of 5%. For DGP(A), the local asymptotic results from Section 4 predict the finite-sample results rather well, in that t_{γ}^{ADF} and $\lambda_{\max}$ again have similar power for this R^2. Moreover, the combination tests $\tilde{\chi}^2_{\mathcal{I}}$ and $UR_{\psi_{\mathcal{I}}}$ again outperform both individual tests. While of the individual tests t_{γ}^{ADF} is most powerful for (C), $\lambda_{\max}$ and $\lambda^*_{\max}$ are most powerful for (B). This result may not be entirely surprising, as both tests were designed having DGPs of type (B) and (C) respectively in mind. For those DGPs, $\tilde{\chi}^2_{\mathcal{I}}$ and $UR_{\psi_{\mathcal{I}}}$ 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. The power of

[18] Appendix E reports results for other values of R^2. Furthermore, we ran all simulations at the 1% and 10% level. We also experimented with a version of DGP(C) with AR(1) error terms instead of white noise $\boldsymbol{u}_t$. All results are qualitatively similar; additional results are available upon request.

[19] 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_γ^{ADF}

		Bootstrap tests					asymptotic tests				
DGP	T	$\lambda^*_{\max}$	$t_\gamma^{\text{ADF},*}$	naive*	$\tilde{\chi}^{2,*}_{\mathcal{I}}$	$UR^*_{\psi_{\mathcal{I}}}$	$\lambda_{\max}$	t_γ^{ADF}	naive	$\tilde{\chi}^2_{\mathcal{I}}$	$UR_{\psi_{\mathcal{I}}}$
(A)	50	0.284	0.255	0.389	0.337	0.273	0.288	0.362	0.462	0.359	0.374
	75	0.281	0.246	0.381	0.324	0.264	0.290	0.320	0.440	0.343	0.344
	100	0.269	0.239	0.368	0.317	0.259	0.279	0.296	0.413	0.307	0.318
	150	0.265	0.235	0.366	0.310	0.252	0.279	0.270	0.394	0.301	0.302
	200	0.274	0.233	0.361	0.306	0.257	0.275	0.258	0.386	0.284	0.290
(B)	50	0.366	0.321	0.486	0.405	0.378	0.412	0.388	0.552	0.448	0.460
	75	0.462	0.341	0.554	0.473	0.452	0.518	0.403	0.617	0.521	0.527
	100	0.534	0.367	0.621	0.529	0.510	0.567	0.405	0.655	0.557	0.557
	150	0.604	0.381	0.668	0.580	0.569	0.627	0.417	0.700	0.614	0.609
	200	0.631	0.377	0.696	0.607	0.594	0.656	0.412	0.721	0.623	0.621
(C)	50	0.179	0.271	0.321	0.278	0.223	0.194	0.372	0.413	0.310	0.329
	75	0.170	0.258	0.304	0.259	0.206	0.193	0.342	0.384	0.285	0.297
	100	0.171	0.271	0.319	0.276	0.215	0.177	0.316	0.358	0.268	0.271
	150	0.160	0.252	0.297	0.255	0.202	0.178	0.299	0.344	0.258	0.260
	200	0.178	0.256	0.303	0.263	0.210	0.173	0.277	0.327	0.239	0.246

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

the bootstrap versions is very similar to that of the asymptotic tests throughout, considering the slightly better size of the bootstrap tests (cf. Table 4). The slight upward size distortion of $UR_{\psi_{\mathcal{I}}}$ found in Table 4 explains why $UR_{\psi_{\mathcal{I}}}$ has higher power than $\tilde{\chi}^2_{\mathcal{I}}$ even when $\lambda_{\max}$ and t_γ^{ADF} are roughly equally powerful, unlike what is predicted by the asymptotics (cf. Figure 1).

Tables 6 and 7 reports analogous results for the tests based on $\hat{F}$ and t_γ^{ECR}. Once more, all tests have a slight upward size distortion for small T, which vanishes as T increases. $\hat{F}$ and t_γ^{ECR} again perform similarly, as predicted by Section 4. It is therefore not surprising that the performance of $\tilde{\chi}^2_{\mathcal{I}}$ and $UR_{\psi_{\mathcal{I}}}$ is also very similar to that of the individual tests. Comparing Tables 5 and 7, we find that t_γ^{ADF} and $\lambda_{\max}$ outperform either $\hat{F}$ or t_γ^{ECR} for DGP(C) and (B), respectively, which again reflects that the latter tests were not designed having such DGPs in mind (cf. the discussion below DGP(C)). This also implies that the superior local power of $\hat{F}$ and t_γ^{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 $\hat{F}$ or t_γ^{ECR} 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 other features of the DGP, as is the case in practice.

6.3 Extension to more than two tests

For expositional clarity we so far analyzed combinations of only $|\mathcal{I}| = 2$ tests, combining t_γ^{ADF} and $\lambda_{\max}$ or $\hat{F}$ and t_γ^{ECR} to illustrate our approach. Of course, as discussed in Section 3, our approach can accommodate other and more tests as well. Potentially, this yields further gains in power if the additional tests have high power for the given nuisance parameter value. We therefore

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

		Bootstrap tests					asymptotic tests				
DGP	T	$\hat{F}^*$	$t_{\gamma}^{\text{ECR},*}$	naive*	$\tilde{\chi}_{\mathcal{I}}^{2,*}$	$UR_{\psi_{\mathcal{I}}}^*$	$\hat{F}$	t_{γ}^{ECR}	naive	$\tilde{\chi}_{\mathcal{I}}^2$	$UR_{\psi_{\mathcal{I}}}$
(A)	50	0.050	0.051	0.062	0.051	0.052	0.084	0.077	0.093	0.079	0.082
	75	0.047	0.045	0.055	0.045	0.046	0.076	0.072	0.086	0.075	0.076
	100	0.050	0.053	0.061	0.051	0.052	0.073	0.073	0.084	0.074	0.073
	150	0.045	0.045	0.053	0.045	0.044	0.065	0.062	0.073	0.065	0.066
	200	0.052	0.055	0.062	0.054	0.052	0.057	0.053	0.063	0.054	0.057
(B)	50	0.050	0.056	0.064	0.054	0.053	0.069	0.068	0.079	0.070	0.069
	75	0.051	0.050	0.060	0.050	0.052	0.067	0.064	0.076	0.065	0.065
	100	0.044	0.044	0.052	0.044	0.044	0.063	0.060	0.072	0.061	0.063
	150	0.049	0.047	0.058	0.049	0.050	0.060	0.057	0.069	0.058	0.058
	200	0.054	0.057	0.066	0.056	0.055	0.064	0.063	0.071	0.062	0.063
(C)	50	0.049	0.054	0.061	0.052	0.052	0.083	0.076	0.091	0.079	0.082
	75	0.042	0.044	0.052	0.044	0.044	0.071	0.069	0.081	0.070	0.070
	100	0.051	0.052	0.061	0.051	0.051	0.068	0.064	0.075	0.067	0.067
	150	0.047	0.048	0.055	0.047	0.047	0.068	0.065	0.076	0.068	0.067
	200	0.051	0.053	0.061	0.053	0.052	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). Starred tests are bootstrap counterparts.

now 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 $\lambda_{\max}$ and t_{γ}^{ADF}, denoted $\tilde{\chi}_{\mathcal{I}}^2(2)$. In view of the similar performance of bootstrap and asymptotic tests we focus on the latter for brevity. The more general $\tilde{\chi}_{\mathcal{I}}^2(4)$ test outperforms its simple counterpart $\tilde{\chi}_{\mathcal{I}}^2(2)$ rather markedly. Of course, the asymptotic results from Section 4 predict that this is a setting where t_{γ}^{ADF} and $\lambda_{\max}$ are less powerful than $\hat{F}$ and t_{γ}^{ECR}, such that one might want to choose the latter only. Yet, bearing Remark 5 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—cointegration versus non-cointegration—whilst 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 $UR_{\psi_{\mathcal{I}}}$ and $\tilde{\chi}_{\mathcal{I}}^2$ control size and yet provide a robust, powerful and flexible alternative to traditional cointegration tests.

7 Mixed Signals Revisited

7.1 Setup

Naturally we are interested in the practical applicability and relevance of our approach. To shed light on this question, we revisit the studies which Gregory *et al.* (2004) investigated for 'mixed signals', i.e. conflicting cointegration test results. Gregory *et al.* (2004) analyze 34 studies which were published in the Journal of Applied Econometrics from 1994 to March/April 2001. We additionally perform an analogous exercise for the JAE issues from May/June 2001 through to papers

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

		Bootstrap tests					asymptotic tests				
DGP	T	$\hat{F}^{*}$	$t_{\gamma}^{\text{ECR},*}$	naive*	$\tilde{\chi}_{\mathcal{I}}^{2,*}$	$UR_{\psi_{\mathcal{I}}}^{*}$	$\hat{F}$	t_{γ}^{ECR}	naive	$\tilde{\chi}_{\mathcal{I}}^{2}$	$UR_{\psi_{\mathcal{I}}}$
(A)	50	0.433	0.415	0.467	0.431	0.426	0.553	0.517	0.578	0.542	0.542
	75	0.433	0.409	0.464	0.427	0.426	0.528	0.491	0.553	0.517	0.519
	100	0.423	0.400	0.452	0.418	0.417	0.496	0.463	0.526	0.487	0.488
	150	0.419	0.392	0.450	0.413	0.413	0.474	0.435	0.500	0.463	0.463
	200	0.422	0.387	0.448	0.409	0.411	0.457	0.413	0.478	0.440	0.445
(B)	50	0.267	0.249	0.291	0.259	0.258	0.352	0.300	0.368	0.327	0.341
	75	0.330	0.298	0.353	0.323	0.323	0.403	0.345	0.416	0.375	0.386
	100	0.381	0.336	0.399	0.363	0.370	0.430	0.362	0.447	0.400	0.415
	150	0.415	0.361	0.430	0.387	0.396	0.464	0.394	0.480	0.432	0.445
	200	0.442	0.375	0.461	0.413	0.422	0.474	0.404	0.490	0.441	0.454
(C)	50	0.217	0.247	0.255	0.237	0.227	0.297	0.321	0.336	0.315	0.306
	75	0.210	0.234	0.244	0.226	0.217	0.281	0.300	0.313	0.294	0.288
	100	0.216	0.245	0.255	0.237	0.226	0.254	0.278	0.290	0.272	0.261
	150	0.203	0.227	0.235	0.218	0.209	0.246	0.270	0.282	0.264	0.256
	200	0.212	0.233	0.243	0.226	0.219	0.232	0.259	0.269	0.248	0.240

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

scheduled for publication as of August 2010.[20] From these studies we construct 286 data sets in which we test for cointegration. Of these, 127 are from the period after April 2001, confirming that cointegration continues to receive unabated attention from the econometrics community.[21] When necessary, we perform some preliminary data transformations such as removal of obvious seasonal patterns. We have substantially more tests than studies because, e.g., we can calculate many time-series cointegration tests from investigations using panel data. The data sets exhibit large differences in sample size T, which ranges from 24 to 7693. Similarly the number of variables K 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 tests are able to heal this problem. As Gregory *et al.* (2004), we do not intend to suggest that the authors of the studies have been in any way strategic in their choice of which cointegration test to report. 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. If a test requires a dependent variable y_t, we follow the choice in the original paper if possible. If there is no obvious y_t, we choose it based on the highest coefficient of determination of first-stage regressions. We also need to allow for variation in lag lengths $\hat{P}$ across data sets. We determine $\hat{P}$ using the standard Schwarz Information

[20] We performed a full text search of 'cointegration' and 'cointegrated' on the Wiley Interscience webpage. Of the 34 hits, we excluded 5 papers, e.g. an editorial for a special issue, pure Monte Carlo papers or those using data already used in the set of studies considered by Gregory *et al.* (2004).

[21] The raw 1994-2001 data are available at http://qed.econ.queensu.ca/jae/2004-v19.1/gregory-haug-lomuto/. Our modified and additional 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	0.061	0.071	0.359	0.490
	75	0.060	0.068	0.343	0.464
	100	0.055	0.064	0.307	0.440
	150	0.054	0.056	0.301	0.413
	200	0.044	0.047	0.284	0.391
(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

See notes to Table 4. $UR_{\psi_{\mathcal{I}}}(|\mathcal{I}|)$ and $\tilde{\chi}^2_{\mathcal{I}}(|\mathcal{I}|)$ combine $|\mathcal{I}|$ tests as described in the text. For DGP(A), results are for $R^2 = 0.25$.

Criterion (BIC) as described e.g. in Lütkepohl (2005, Secs. 4.3.2 and 8.1). We 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 for $\hat{P}$ discussed in the literature were employed. All tests include a constant and a trend.

7.2 Results

We compare the results of individually applying $\lambda_{\max}$, t_γ^{ADF}, t_γ^{ECR} and $\hat{F}$ with the meta test $\tilde{\chi}^2_{\mathcal{I}}(\lambda_{\max}, t_\gamma^{\mathrm{ADF}}, t_\gamma^{\mathrm{ECR}}, \hat{F})$. Let us first reconsider the example studies discussed in Section 2.1. First, for the Clements and Hendry (1995) data, $\lambda_{\max}$ and $\hat{F}$ rejected. Second, for the Cooley and Ogaki (1996) data only t_γ^{ADF} rejected. Third, for the Martens *et al.* (1998) data, $\lambda_{\max}$ and t_γ^{ADF} rejected. These patterns of (non-)rejections are noteworthy. The first example shows that t_γ^{ECR} and $\hat{F}$, which are constructed similarly and have similar power properties (see e.g. Section 4), may not agree for the *same* samples. The second example shows that t_γ^{ADF}, which is often thought to be less powerful than many other tests proposed, produces a rejection while the system- and error-correction based tests do not. The third example shows that t_γ^{ECR} and $\hat{F}$ may not reject although $\lambda_{\max}$ does. Overall, the examples show that mixed signals do not stem from a single test always or never rejecting.

How does the meta test $\tilde{\chi}^2_{\mathcal{I}}(\lambda_{\max}, t_\gamma^{\mathrm{ADF}}, t_\gamma^{\mathrm{ECR}}, \hat{F})$ resolve these mixed signals? For the Clements and Hendry (1995) data $\tilde{\chi}^2_{\mathcal{I}}(\lambda_{\max}, t_\gamma^{\mathrm{ADF}}, t_\gamma^{\mathrm{ECR}}, \hat{F}) = 39.870$, clearly exceeding the 5% critical value of 22.215 (cf. Table 1 for $K-1=1$ and case (*iii*)). The meta test hence agrees with $\lambda_{\max}$ and $\hat{F}$ here. On the other hand, $\tilde{\chi}^2_{\mathcal{I}}(\lambda_{\max}, t_\gamma^{\mathrm{ADF}}, t_\gamma^{\mathrm{ECR}}, \hat{F}) = 16.1257 < 22.215$ (and is also smaller than the 10% critical value 17.187, cf. Table B.2) for the Cooley and Ogaki (1996) data, such that the meta tests joins the three non-rejecting $\lambda_{\max}$, t_γ^{ECR} and $\hat{F}$ tests. Apparently, the p-value of t_γ^{ADF} is insufficiently small to lead to a rejection for $\tilde{\chi}^2_{\mathcal{I}}(\lambda_{\max}, t_\gamma^{\text{ADF}}, t_\gamma^{\text{ECR}}, \hat{F})$. The very small p-values for the Martens *et al.* (1998) data of course produce a very large, and therefore rejecting, meta test statistic. Hence, we observe that the meta test aggregates the information from the single tests such that, depending on the relative strengths of rejection and acceptance, either aggregate test result can obtain.

More generally, we check whether all individual tests from the Gregory *et al.* (2004) data and the updated set agree or not in their testing decision at the 5% level, see left panel of Table 9. If there are conflicting test results we check what the test used in the original paper had suggested as a 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.[22] We then 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.

Table 9 thus reports the frequencies for all possible pairs of outcomes.[23] When all tests reject or all tests do not reject $\mathcal{H}_0$, the meta test does so too. However, such cases of agreeing tests make up only 65% (= (56 + 131)/286) of all data sets. For the remaining 35% individual tests conflict. Here our test is most useful, yielding a definite conclusion. In 54% (= 53/99) of the conflicting cases $\tilde{\chi}^2_{\mathcal{I}}$ does not reject $\mathcal{H}_0$. In the remaining conflicting cases $\tilde{\chi}^2_{\mathcal{I}}$ rejects $\mathcal{H}_0$. 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 (99 vs. 46 cases). Similarly, the conservative strategy of only rejecting when all tests reject would have understated the pervasiveness of cointegration.

Second, the tests that have been 'preferred' in the studies are more rejective than our meta test (51 vs. 37 rejections in 77 tests). This suggests that the evidence in favor of cointegration would have been less pronounced if the studies could have relied on a suitable meta test.[24]

Third, whether or not the preferred test rejected $\mathcal{H}_0$ is not informative on whether or not $\tilde{\chi}^2_{\mathcal{I}}$ rejects conditional on observing 'mixed signals'. This is reflected by similar conditional probabilities: $53/99 \simeq 26/51 \simeq 14/26 \approx 1/2$. Thus, we cannot infer from a published result what the $\tilde{\chi}^2_{\mathcal{I}}$ test would indicate, conditional on a further individual test leading to a conflicting test result.

[22] 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 system-based Johansen (1988) test. That is, we identify all Johansen tests with $\lambda_{\max}$ and all residual-based tests with t_γ^{ADF}. Given the highly positive correlation within classes of tests (Gregory *et al.*, 2004), this approximation is accurate. In 22 (99 − 77) cases of conflicting test results, the original studies do not report a cointegration test, being concerned with e.g. estimating cointegration vectors.

[23] Appendix F reports results for $\tilde{\chi}^2_{\mathcal{I}}(\lambda_{\max}, t_\gamma^{\text{ADF}})$; results for other (bootstrap) combination tests are available.

[24] That the preferred test is more rejective than $\tilde{\chi}^2_{\mathcal{I}}$ here does not contradict the favorable power properties of $\tilde{\chi}^2_{\mathcal{I}}$ found in Section 6, as $\tilde{\chi}^2_{\mathcal{I}}$ can, and should, of course only be shown to be powerful in a class of level-α tests. Whether 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.

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

number of cases in which...	...individual test results...				...in case of conflicting results: 'preferred' test†			
	agree		conflict					
	r	$\neg r$		$\sum$		r	$\neg r$	$\sum$
$\tilde{\chi}^2_{\mathcal{I}}(4) : r$	56	0	46	102	$\tilde{\chi}^2_{\mathcal{I}}(4) : r$	25	12	37
$\tilde{\chi}^2_{\mathcal{I}}(4) : \neg r$	0	131	53	184	$\tilde{\chi}^2_{\mathcal{I}}(4) : \neg r$	26	14	40
$\sum$	56	131	99	286	$\sum$	51	26	77

$\tilde{\chi}^2_{\mathcal{I}}(4)$ abbreviates $\tilde{\chi}^2_{\mathcal{I}}(\lambda_{\max}, t^{\text{ADF}}_{\gamma}, t^{\text{ECR}}_{\gamma}, \hat{F})$. r : test rejects; $\neg r$: test does not reject.
† : Test type on which conclusions in the original study were based (see fn. 22).
Absolute frequencies of cointegration-test results for data from Gregory *et al.* (2004). Individual tests include Engle and Granger (1987), Boswijk (1994), Banerjee *et al.* (1998) and Johansen (1988) tests. The $\tilde{\chi}^2_{\mathcal{I}}(4)$ combines these tests as described in Section 3.

8 Conclusion

This paper proposes meta tests that combine information from individual cointegration tests. The tests take into account the multiple testing nature of running more than one individual 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 corresponding bootstrap methods. By contrast, we show that running more than one test and drawing inferences from the most rejective test leads to an oversized test. Asymptotic and Monte Carlo results demonstrate the effectiveness of the proposed meta tests, establishing attractive power properties. An application to a large and up-to-date set of cointegration studies confirms our tests' practical value, yielding an unambiguous test decision in cases of conflicting individual 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.

A major practical advantage of our proposed tests is that they relieve the applied researcher from the discretionary and often arbitrary choice between cointegration tests to reach a 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.

Bierens HJ. 2005. *Introduction to the Mathematical and Statistical Foundations of Econometrics*. Cambridge: Cambridge University Press.

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.

Clements MP, Hendry DF. 1995. Forecasting in cointegrated systems. *Journal of Applied Econometrics* **10**: 127–146.

Cooley TF, Ogaki M. 1996. A time series analysis of real wages, consumption, and asset returns. *Journal of Applied Econometrics* **11**: 119–134.

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 RF, Granger CW. 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.

Harvey DI, Leybourne SJ, Taylor AMR. 2009. Unit root testing in practice: Dealing with uncertainty over the trend and initial condition. *Econometric Theory* **25**: 587–636.

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.

Kremers JJ, Ericsson NR, Dolado JJ. 1992. The power of cointegration tests. *Oxford Bulletin of Economics and Statistics* **54**: 325–348.

Littell RC, Folks JL. 1971. Asymptotic optimality of Fisher's method of combining independent tests. *Journal of the American Statistical Association* **66**: 802–806.

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.

Martens M, Kofman P, Vorst TCF. 1998. A threshold error-correction model for intraday futures and index returns. *Journal of Applied Econometrics* **13**: 245–263.

Palm FC, Smeekes S, Urbain JP. 2010. A sieve bootstrap test for cointegration in a conditional error correction model. *Econometric Theory* **26**: 647–681.

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

Park JY. 1990. Testing for unit roots and cointegration by variable addition. *Advances in Econornetrics* **8**: 107–133.

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.

Watson MW. 1994. Vector autoregressions and cointegration. In Engle R, McFadden D (eds.) *Handbook of Econometrics*, vol. 4, chap. 47, Amsterdam: Elsevier, pages 2843–2915.

White H. 2000. A reality check for data snooping. *Econometrica* **68**: 1097–1126.

Appendix A Proofs

Proof of Proposition 1. Under $\mathcal{H}_0$, i.e. $c = 0$, the limiting random variables of the ξ_i are nuisance-parameter free functionals of $\boldsymbol{W}$. E.g., for the tests considered in Section 2.2 we obtain

$$
\begin{aligned}
t_\gamma^{\mathrm{ADF}} &\Rightarrow \frac{\boldsymbol{\eta}^{d\prime} \int \boldsymbol{W}^d \,\mathrm{d}\boldsymbol{W}' \boldsymbol{\eta}^d}{(\boldsymbol{\eta}^{d\prime} \boldsymbol{A}^d \boldsymbol{\eta}^d)^{1/2} (\boldsymbol{\eta}^{d\prime} \boldsymbol{\eta}^d)^{1/2}} \\
\text{where} \quad \boldsymbol{\eta}^d &:= \left[-\left(\int \boldsymbol{W}_1^{d\prime} W_2^d \right) \left(\int \boldsymbol{W}_1^d \boldsymbol{W}_1^{d\prime} \right)^{-1}, \quad 1 \right]', \\
\boldsymbol{W}^d &:= (\boldsymbol{W}_1^{d\prime},\ W_2^d)' \qquad \text{and} \quad \boldsymbol{A}^d := \int \boldsymbol{W}^d \boldsymbol{W}^{d\prime} \\
\lambda_{\max} &\Rightarrow \max \operatorname{eig} \left\{ (\boldsymbol{A}^d)^{-1} \int \boldsymbol{W}^d \,\mathrm{d}\boldsymbol{W}' \int \mathrm{d}\boldsymbol{W} \boldsymbol{W}^{d\prime} \right\} \\
\hat{F} &\Rightarrow \int \boldsymbol{W}^{d\prime} \,\mathrm{d}W_2 (\boldsymbol{A}^d)^{-1} \int \boldsymbol{W}^d \,\mathrm{d}W_2 \\
t_\gamma^{\mathrm{ECR}} &\Rightarrow \frac{\int W_2^d \,\mathrm{d}W_2 - \int \boldsymbol{W}_1^{d\prime} W_2^d \left(\int \boldsymbol{W}_1^d \boldsymbol{W}_1^{d\prime} \right)^{-1} \int \boldsymbol{W}_1^d \,\mathrm{d}W_2}{\left[\int W_2^{d2} - \int \boldsymbol{W}_1^{d\prime} W_2^d \left(\int \boldsymbol{W}_1^d \boldsymbol{W}_1^{d\prime} \right)^{-1} \int \boldsymbol{W}_1^d W_2^d \right]^{1/2}}
\end{aligned}
$$

Now, the corresponding limiting cdfs Ξ_i are continuous, such that the quantile transformations $p_i = \Xi_i(\xi_i)$ are uniform on $[0, 1]$ under $\mathcal{H}_0$ as $T \to \infty$. Further, ln as well as $-2 \sum_{i \in \mathcal{I}} f_i$ obviously are continuous functions, such that part (a) follows from the Continuous Mapping Theorem, if the ξ_i converge jointly. (Their *joint* convergence follows from the joint convergence of all sample moments used in the construction of the ξ_i, because the ξ_i are continuous functions of the sample moments themselves. Watson (1994) shows joint convergence of the sample moments.) Part (b) follows because test consistency of test i implies that under $\mathcal{H}_1$, $p_i = o_p(1)$ and hence $\tilde{\chi}^2_{\mathcal{I}} \to_{\mathrm{p}} \infty$ even if $p_j = \mathcal{O}_p(1)$ for $j \neq i$. □

Appendix B Further critical values

Table B.1: Critical values for the minimum p-value test.

$K-1$	case (i)	case (ii)	case (iii)	case (i)	case (ii)	case (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
7	0.005	0.005	0.005	0.007	0.007	0.007
8	0.005	0.005	0.005	0.007	0.007	0.007
9	0.005	0.005	0.005	0.007	0.007	0.007
10	0.005	0.005	0.005	0.007	0.007	0.007
11	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
7	0.027	0.027	0.027	0.035	0.035	0.036
8	0.027	0.027	0.027	0.035	0.035	0.036
9	0.027	0.027	0.027	0.034	0.035	0.035
10	0.027	0.027	0.027	0.034	0.035	0.035
11	0.026	0.027	0.026	0.034	0.034	0.035
$\alpha = 0.1$						
	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
7	0.056	0.056	0.055	0.071	0.072	0.074
8	0.055	0.055	0.055	0.071	0.072	0.073
9	0.055	0.055	0.055	0.070	0.072	0.073
10	0.055	0.054	0.054	0.070	0.071	0.072
11	0.054	0.054	0.054	0.070	0.071	0.072

Critical values for the minimum p-value test.

Table B.2: Critical values for the $\tilde{\chi}^2_{\mathcal{I}}$ 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
7	15.312	15.313	15.184	14.769	14.815	14.839	17.675	17.708	17.837	15.830	15.921	16.176
8	15.183	15.000	15.016	14.670	14.700	14.613	17.696	17.705	17.817	15.830	15.947	16.069
9	14.960	15.007	15.069	14.602	14.604	14.580	17.605	17.763	17.851	15.791	15.941	16.143
10	14.893	14.853	14.788	14.586	14.493	14.483	17.530	17.685	17.692	15.795	15.984	16.080
11	14.690	14.826	14.745	14.358	14.282	14.323	17.554	17.564	17.760	15.670	15.795	16.019
	$\hat{F}$, $\lambda_{\max}$ 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			
7	21.288	21.430	21.572	22.214	22.218	22.336	28.974	29.037	29.440			
8	21.120	21.180	21.163	22.042	22.083	22.203	28.780	28.999	29.084			
9	20.904	20.997	21.118	21.857	22.047	22.045	28.326	28.840	28.875			
10	20.678	20.818	20.901	21.709	21.874	21.774	28.208	28.575	28.577			
11	20.418	20.611	20.769	21.585	21.567	21.688	27.945	28.055	28.518			
$\alpha = 0.1$												
	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
7	8.125	8.169	8.146	8.060	8.054	8.037	8.882	8.909	8.938	8.389	8.449	8.494
8	8.106	8.134	8.146	8.046	8.037	8.010	8.868	8.922	8.949	8.356	8.385	8.454
9	8.067	8.108	8.096	8.019	8.033	8.003	8.864	8.880	8.922	8.370	8.400	8.455
10	8.081	8.067	8.095	7.986	7.988	7.980	8.882	8.885	8.919	8.320	8.395	8.441
11	8.084	8.053	8.084	7.995	7.983	7.974	8.887	8.925	8.921	8.328	8.374	8.431
	$\hat{F}$, $\lambda_{\max}$ 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			
7	11.634	11.763	11.738	11.849	11.956	11.917	15.611	15.731	15.706			
8	11.637	11.643	11.701	11.884	11.885	11.892	15.561	15.591	15.705			
9	11.615	11.631	11.703	11.847	11.880	11.873	15.507	15.528	15.647			
10	11.529	11.567	11.638	11.819	11.833	11.837	15.422	15.476	15.565			
11	11.543	11.581	11.644	11.767	11.856	11.835	15.406	15.476	15.564			

1%- and 10%-critical values for combination tests based on $\tilde{\chi}^2_{\mathcal{I}}$. 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).

Appendix C Local Asymptotic Power, further results

Table C.1: Local asymptotic power

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

Case (i). See notes to Table 3.

Table C.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
$UR_{\psi_{\mathcal{I}}}(\hat{F}, t_\gamma^{\text{ECR}})$	0.049	0.070	0.142	0.279	0.471
$UR_{\psi_{\mathcal{I}}}(t_\gamma^{\text{ADF}}, \lambda_{\max})$	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
$UR_{\psi_{\mathcal{I}}}(\hat{F}, t_\gamma^{\text{ECR}})$	0.051	0.069	0.121	0.247	0.456
$UR_{\psi_{\mathcal{I}}}(t_\gamma^{\text{ADF}}, \lambda_{\max})$	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}	0.049	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
$UR_{\psi_{\mathcal{I}}}(\hat{F}, t_\gamma^{\text{ECR}})$	0.049	0.102	0.318	0.648	0.882
$UR_{\psi_{\mathcal{I}}}(t_\gamma^{\text{ADF}}, \lambda_{\max})$	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}	0.048	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
$UR_{\psi_{\mathcal{I}}}(\hat{F}, t_\gamma^{\text{ECR}})$	0.050	0.196	0.689	0.946	0.995
$UR_{\psi_{\mathcal{I}}}(t_\gamma^{\text{ADF}}, \lambda_{\max})$	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}	0.051	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 3.

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

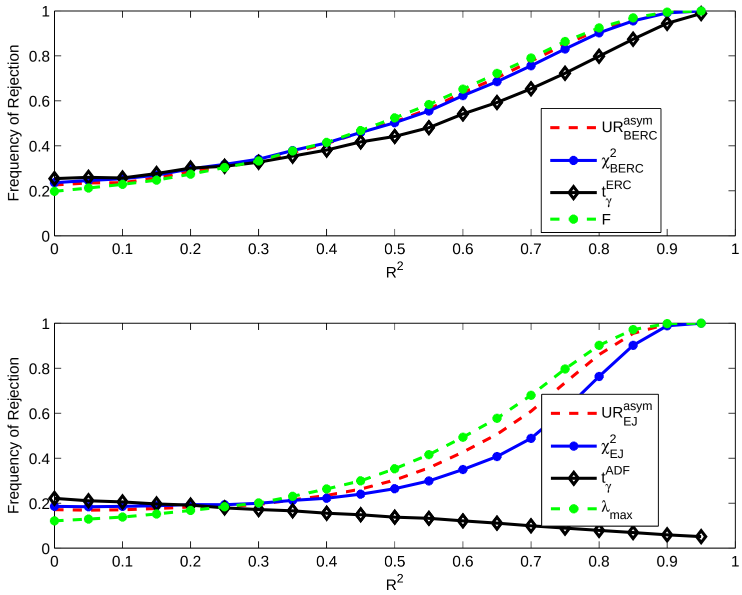
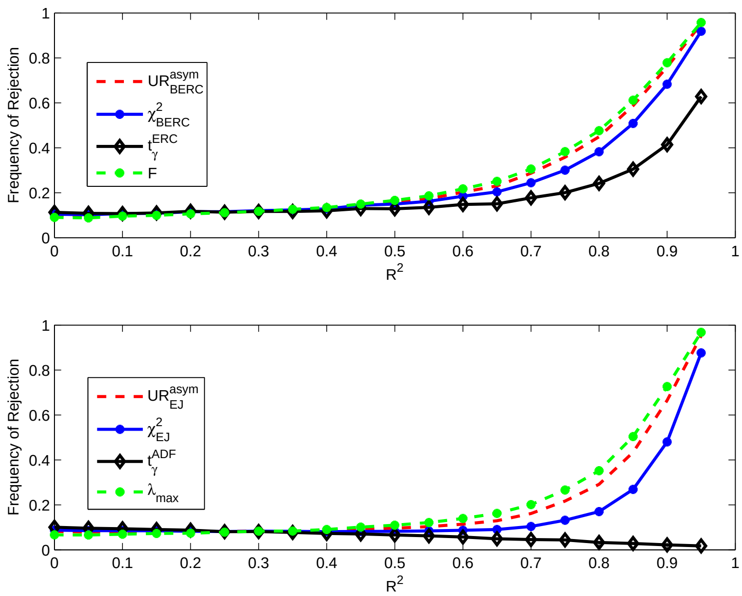
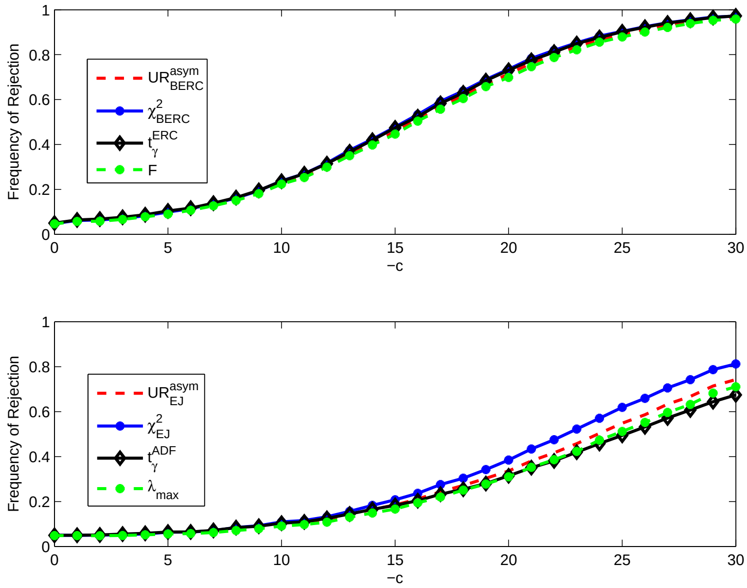


Figure C.2: Local asymptotic power as a function of R^2, $c = -5$



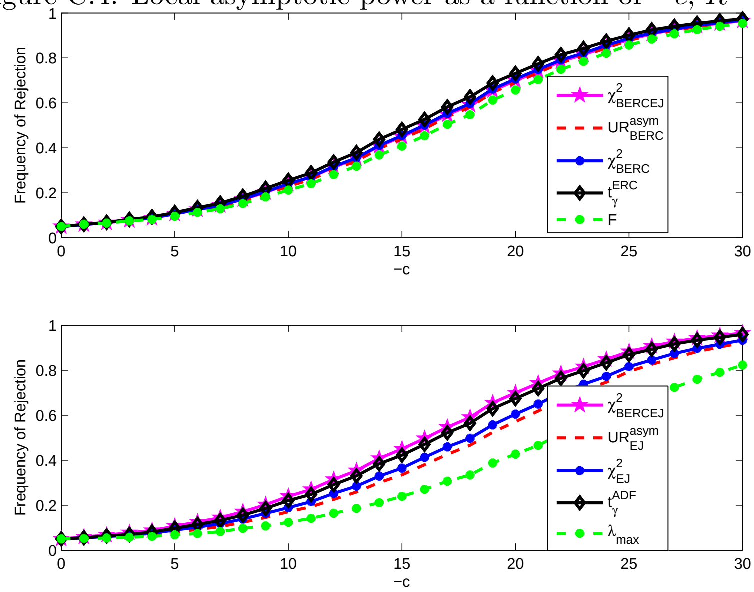
Results are for the demeaned case (*ii*). χ^2_{BERC} is our Fisher test (3) based on Boswijk's and Banerjee *et al.*'s tests. χ^2_{EJ} is based on Engle and Granger's and Johansen's tests. $UR^{\mathrm{asym}}_{\mathrm{BERC}}$ and $UR^{\mathrm{asym}}_{\mathrm{EJ}}$ are the corresponding asymmetric $UR_{\psi_{\mathcal{I}}}$ test (6). The individual tests' power curves are for comparison.

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



See notes to Figure 1.

Figure C.4: Local asymptotic power as a function of $-c$, $R^2 = 0$



See notes to Figure 1.

Appendix D Alternative Bootstrap Tests

This Appendix describes an alternative bootstrap approach that makes somewhat stronger assumptions about the joint distribution of the ξ_i. Its power was slightly superior to the Fisher-test version in our simulations (detailed results are available). Define a probit representation by $\Phi^{-1}(p_i) =: s_i$. Asymptotically, the s_i are marginally standard normal under $\mathcal{H}_0$. Let $\boldsymbol{s} = \left(s_1, \ldots, s_{|\mathcal{I}|}\right)'$. If we additionally assume joint normality for $\boldsymbol{s}$, denoted $\boldsymbol{s} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, we have $\boldsymbol{\iota}'\boldsymbol{s} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\iota}'\boldsymbol{\Sigma}\boldsymbol{\iota})$, where $\boldsymbol{\iota} = (1, \ldots, 1)'$. This leads to a standardized meta test statistic, $\tau = \boldsymbol{\iota}'\boldsymbol{s}/(\boldsymbol{\iota}'\boldsymbol{\Sigma}\boldsymbol{\iota})^{1/2}$. τ is standard normal under $\mathcal{H}_0$ and joint normality. Fortunately, Demetrescu *et al.* (2006) show that this strong assumption is not necessary. As a practical matter, we use the following bootstrap method to provide a feasible estimator of $\boldsymbol{\Sigma}$.

Algorithm 2.

7. *(1.-6. are as in Algorithm 1.) Obtain the probit representation of each* $\xi^*_{i,b}$, $s^*_{i,b} := \Phi^{-1}(p^*_{i,b})$. *Let* $\boldsymbol{s}^*_b := \left(s^*_{1,b}, \ldots, s^*_{|\mathcal{I}|,b}\right)'$. *Correspondingly, obtain* $s^*_i := \Phi^{-1}(p^*_i)$ *and* $\boldsymbol{s}^* := \left(s^*_1, \ldots, s^*_{|\mathcal{I}|}\right)'$.
8. *Letting* $\bar{\boldsymbol{s}}^* := \frac{1}{B}\sum_b \boldsymbol{s}^*_b$, *estimate* $\boldsymbol{\Sigma}$ *by* $\boldsymbol{\Sigma}^* = \frac{1}{B}\sum_b (\boldsymbol{s}^*_b - \bar{\boldsymbol{s}}^*)(\boldsymbol{s}^*_b - \bar{\boldsymbol{s}}^*)'$.

Algorithm 2 provides a feasible version of the test statistic τ, $\tau^* := \boldsymbol{\iota}'\boldsymbol{s}^*/\sqrt{\boldsymbol{\iota}'\boldsymbol{\Sigma}^*\boldsymbol{\iota}}$. We reject $\mathcal{H}_0$ at level α if $\tau^* < \Phi^{-1}(\alpha)$. The following Lemma provides a useful consistency property of the test.

Lemma 2. *If* (i) $\alpha < 1/2$ *and* (ii) *all* 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 $i = 1, \ldots, |\mathcal{I}|$. 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 $\sqrt{\boldsymbol{\iota}'\boldsymbol{\Sigma}^*\boldsymbol{\iota}} \leqslant |\mathcal{I}|$. Thus, $\tau^* \leqslant \boldsymbol{\iota}'\boldsymbol{s}^*/|\mathcal{I}| < \Phi^{-1}(\alpha)$. □

Appendix E Additional Simulation Results

Table E.1: Small-sample results for DGP(B), $\boldsymbol{\Pi} = (-1\ \ 1)'(.1\ \ -.1)$

DGP	T	$\hat{F}$	$t^{\mathrm{ECR}}_{\gamma}$	$\tilde{\chi}^2_{\mathcal{I}}$	$UR_{\psi_{\mathcal{I}}}$	$\lambda_{\max}$	$t^{\mathrm{ADF}}_{\gamma}$	$\tilde{\chi}^2_{\mathcal{I}}$	$UR_{\psi_{\mathcal{I}}}$
Size	50	0.079	0.072	0.076	0.078	0.067	0.072	0.065	0.078
	75	0.070	0.066	0.069	0.070	0.060	0.067	0.061	0.067
	100	0.063	0.060	0.061	0.062	0.058	0.057	0.060	0.066
	150	0.059	0.057	0.057	0.059	0.054	0.055	0.056	0.059
	200	0.054	0.053	0.050	0.054	0.054	0.053	0.053	0.057
Power	50	0.128	0.127	0.130	0.129	0.132	0.148	0.147	0.154
	75	0.180	0.185	0.185	0.184	0.251	0.261	0.279	0.262
	100	0.260	0.279	0.273	0.267	0.412	0.441	0.477	0.439
	150	0.433	0.471	0.459	0.449	0.793	0.815	0.852	0.804
	200	0.632	0.667	0.655	0.645	0.968	0.971	0.981	0.971

See notes to Table 4. We waive to report the analogous bootstrap results.

Table E.2: Small-sample results for DGP(B), $\boldsymbol{\Pi} = (-1 \quad 1)'(.1 \quad -.1)$, $\boldsymbol{\Gamma} = (0.1 \;\; 0.1)'(1 \;\; 1)$, $\boldsymbol{\Omega}$ as in (A), $\delta = 1/2$

DGP	T	$\hat{F}$	t_γ^{ECR}	$\tilde{\chi}^2_{\mathcal{I}}$	$UR_{\psi_{\mathcal{I}}}$	$\lambda_{\max}$	t_γ^{ADF}	$\tilde{\chi}^2_{\mathcal{I}}$	$UR_{\psi_{\mathcal{I}}}$
Size	50	0.073	0.078	0.071	0.072	0.061	0.070	0.062	0.076
	75	0.063	0.068	0.063	0.065	0.052	0.060	0.056	0.068
	100	0.057	0.059	0.058	0.059	0.053	0.059	0.054	0.060
	150	0.056	0.058	0.056	0.057	0.053	0.053	0.052	0.059
	200	0.061	0.059	0.059	0.058	0.059	0.056	0.057	0.059
Power	50	0.490	0.498	0.498	0.495	0.572	0.321	0.505	0.515
	75	0.562	0.569	0.574	0.564	0.660	0.397	0.608	0.603
	100	0.584	0.577	0.586	0.585	0.712	0.424	0.660	0.657
	150	0.639	0.631	0.645	0.638	0.772	0.475	0.725	0.716
	200	0.648	0.634	0.647	0.646	0.798	0.484	0.734	0.741

See notes to Tables 4 and E.1.

Table E.3: Small-sample results for DGP(C), $\boldsymbol{\Omega}$ as in (A), $\delta = 1/2$

DGP	T	$\hat{F}$	t_γ^{ECR}	$\tilde{\chi}^2_{\mathcal{I}}$	$UR_{\psi_{\mathcal{I}}}$	$\lambda_{\max}$	t_γ^{ADF}	$\tilde{\chi}^2_{\mathcal{I}}$	$UR_{\psi_{\mathcal{I}}}$
Size	50	0.082	0.074	0.077	0.081	0.052	0.077	0.058	0.076
	75	0.063	0.061	0.063	0.061	0.044	0.065	0.048	0.063
	100	0.065	0.061	0.062	0.065	0.052	0.067	0.048	0.068
	150	0.062	0.058	0.060	0.061	0.051	0.060	0.045	0.065
	200	0.057	0.058	0.058	0.059	0.052	0.061	0.046	0.064
Power	50	0.217	0.221	0.222	0.219	0.183	0.283	0.261	0.273
	75	0.189	0.196	0.197	0.192	0.173	0.260	0.232	0.236
	100	0.175	0.184	0.181	0.176	0.179	0.238	0.227	0.228
	150	0.173	0.184	0.184	0.176	0.185	0.232	0.222	0.224
	200	0.163	0.176	0.173	0.166	0.170	0.216	0.205	0.201

See notes to Tables 4 and E.1.

Table E.4: Small-sample power, DGP(A), further R^2s

DGP	T	$\hat{F}$	t_γ^{ECR}	$\tilde{\chi}^2_{\mathcal{I}}$	$UR_{\psi_{\mathcal{I}}}$	$\lambda_{\max}$	t_γ^{ADF}	$\tilde{\chi}^2_{\mathcal{I}}$	$UR_{\psi_{\mathcal{I}}}$
$R^2 = 0$	50	0.401	0.427	0.418	0.411	0.194	0.440	0.349	0.380
	75	0.370	0.407	0.395	0.387	0.192	0.406	0.323	0.341
	100	0.343	0.376	0.364	0.356	0.177	0.369	0.300	0.315
	150	0.319	0.353	0.341	0.329	0.178	0.335	0.284	0.294
	200	0.301	0.331	0.322	0.311	0.173	0.320	0.263	0.275
$R^2 = 0.5$	50	0.771	0.663	0.734	0.762	0.528	0.257	0.440	0.501
	75	0.748	0.637	0.711	0.735	0.528	0.223	0.435	0.487
	100	0.739	0.618	0.700	0.727	0.524	0.207	0.411	0.469
	150	0.714	0.594	0.671	0.696	0.522	0.189	0.404	0.468
	200	0.702	0.569	0.654	0.686	0.511	0.180	0.389	0.463
$R^2 = 0.75$	50	0.968	0.882	0.953	0.965	0.918	0.149	0.801	0.885
	75	0.966	0.878	0.950	0.962	0.925	0.121	0.801	0.895
	100	0.959	0.865	0.941	0.953	0.925	0.108	0.803	0.895
	150	0.960	0.853	0.939	0.955	0.934	0.100	0.808	0.899
	200	0.958	0.846	0.935	0.953	0.938	0.095	0.813	0.910

See notes to Tables 4 and E.1.

Table E.5: Small-sample power, further c

DGP	T	$\hat{F}$	t_γ^{ECR}	$\tilde{\chi}^2_{\mathcal{I}}$	$UR_{\psi_{\mathcal{I}}}$	$\lambda_{\max}$	t_γ^{ADF}	$\tilde{\chi}^2_{\mathcal{I}}$	$UR_{\psi_{\mathcal{I}}}$
$c = -5$									
(A)	50	0.126	0.110	0.116	0.121	0.069	0.097	0.072	0.098
	75	0.114	0.102	0.108	0.113	0.067	0.088	0.065	0.086
	100	0.108	0.099	0.102	0.105	0.070	0.087	0.070	0.096
	150	0.105	0.094	0.100	0.101	0.070	0.081	0.067	0.087
	200	0.104	0.092	0.098	0.102	0.072	0.080	0.070	0.089
(B)	50	0.145	0.120	0.130	0.141	0.103	0.107	0.106	0.121
	75	0.129	0.114	0.121	0.127	0.100	0.104	0.102	0.109
	100	0.133	0.108	0.121	0.128	0.105	0.099	0.101	0.111
	150	0.127	0.108	0.115	0.120	0.110	0.094	0.104	0.109
	200	0.126	0.109	0.117	0.121	0.110	0.090	0.101	0.110
(C)	50	0.097	0.095	0.095	0.096	0.059	0.102	0.075	0.097
	75	0.092	0.090	0.090	0.090	0.058	0.097	0.071	0.087
	100	0.089	0.090	0.091	0.090	0.060	0.092	0.069	0.089
	150	0.080	0.081	0.082	0.081	0.065	0.087	0.071	0.083
	200	0.084	0.085	0.085	0.083	0.065	0.090	0.073	0.089
$c = -10$									
(A)	50	0.303	0.265	0.285	0.293	0.144	0.186	0.171	0.196
	75	0.264	0.231	0.249	0.258	0.141	0.163	0.153	0.171
	100	0.247	0.214	0.234	0.241	0.133	0.147	0.140	0.161
	150	0.232	0.202	0.224	0.223	0.133	0.140	0.136	0.152
	200	0.219	0.190	0.203	0.210	0.131	0.128	0.129	0.148
(B)	50	0.252	0.206	0.229	0.240	0.228	0.218	0.244	0.265
	75	0.267	0.227	0.247	0.259	0.272	0.227	0.264	0.280
	100	0.269	0.226	0.247	0.258	0.289	0.209	0.262	0.284
	150	0.284	0.239	0.261	0.272	0.319	0.214	0.286	0.304
	200	0.276	0.234	0.256	0.267	0.327	0.202	0.290	0.306
(C)	50	0.175	0.184	0.183	0.179	0.107	0.197	0.157	0.179
	75	0.161	0.165	0.166	0.164	0.103	0.175	0.138	0.164
	100	0.153	0.162	0.161	0.155	0.098	0.170	0.135	0.155
	150	0.146	0.154	0.150	0.147	0.097	0.159	0.130	0.147
	200	0.131	0.142	0.138	0.135	0.098	0.143	0.119	0.135

See notes to Table 4 and E.1. For DGP(A), $R^2 = 0.25$.

Appendix F Additional Empirical Results

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

number of cases in which...	...individual test results...				...in case of conflicting results: 'preferred' test†			
	agree		conflict					
	r	$\neg r$		$\sum$		r	$\neg r$	$\sum$
$\tilde{\chi}^2_{\mathcal{I}}(2) : r$	70	0	53	123	$\tilde{\chi}^2_{\mathcal{I}}(2) : r$	23	17	40
$\tilde{\chi}^2_{\mathcal{I}}(2) : \neg r$	0	135	28	163	$\tilde{\chi}^2_{\mathcal{I}}(2) : \neg r$	14	6	20
$\sum$	70	135	81	286	$\sum$	37	23	60

$\tilde{\chi}^2_{\mathcal{I}}(2)$ abbreviates $\tilde{\chi}^2_{\mathcal{I}}(\lambda_{\max}, t_{\gamma}^{\text{ADF}})$.
r : test rejects; $\neg r$: test does not reject
† : Test type on which conclusions in the original study were based (see fn. 22).
Absolute frequencies of cointegration-test results for data from Gregory *et al.* (2004). Individual tests include Engle and Granger (1987) and Johansen (1988) tests. The $\tilde{\chi}^2_{\mathcal{I}}(2)$ combines these tests as described in Section 3.