261

Reihe Ökonomie

Economics Series

A Simple Panel-CADF Test for Unit Roots

Mauro Costantini, Claudio Lupi

Institut für Höhere Studien
Institute for Advanced Studies
Vienna

261

Reihe Ökonomie

Economics Series

A Simple Panel-CADF Test for Unit Roots

Mauro Costantini, Claudio Lupi

January 2011

Institut für Höhere Studien (IHS), Wien
Institute for Advanced Studies, Vienna

Contact:

Mauro Costantini
Department of Economics
University of Vienna
Bruenner Strasse 72
1210 Vienna, Austria
email: mauro.costantini@univie.ac.at

Claudio Lupi – corresponding author
University of Molise
Faculty of Economics
Dept. SEGeS
Via De Sanctis
I-86100 Campobasso, Italy
Fax +39 0874 311124
☎: +39 0874 404451
email: lupi@unimol.it

Abstract

In this paper we propose a simple extension to the panel case of the covariate-augmented Dickey Fuller (CADF) test for unit roots developed in Hansen (1995). The extension we propose is based on a p-values combination approach that takes into account cross-section dependence. We show that the test is easy to compute, has good size properties and gives power gains with respect to other popular panel approaches. A procedure to compute the asymptotic p-values of Hansen's CADF test is also a side-contribution of the paper. We also complement Hansen (1995) and Caporale and Pittis (1999) with some new theoretical results. Two empirical applications are carried out for illustration purposes on international data to test the PPP hypothesis and the presence of a unit root in international industrial production indices.

Keywords
Unit root, Panel data, Approximate P-values, Monte Carlo

JEL Classification
C22, C23, F31

Contents

1 Introduction

It is well known that standard unit root tests suffer from low power (see e.g. Campbell and Perron, 1991; DeJong *et al.*, 1992; Phillips and Xiao, 1998). Starting from the mid-nineties, it has been suggested that a viable way to increase power in unit root testing is to exploit cross-section variation together with univariate time series dynamics (see Quah, 1994; Levin *et al.*, 2002, among others). Panel unit root tests have become increasingly popular ever since. Of course, potential power gains are not the only reason for using panel tests. A commonly neglected advantage of the panel unit root approach is that it can be useful in avoiding some complications arising from multiple testing, as we will show in this paper. Furthermore, some specific cross-country macroeconomic analyses may fit naturally in the panel framework, in particular when the focus is on testing for the presence of a unit root as an interesting economically interpretable common feature in a whole set of time series. However, the power gain motivation has probably been the dominating one in the majority of theoretical and applied papers and it has been questioned only recently (see e.g. Banerjee *et al.*, 2004, 2005).

In order to obtain more powerful unit root tests, Hansen (1995) adopted a different approach to exploit cross-sectional correlation. Rather than using panel data on a single variable, Hansen (1995) suggested using stationary covariates in an otherwise standard Dickey-Fuller framework, in this way proposing his covariate augmented Dickey-Fuller test (CADF). Indeed, Hansen (1995) and Caporale and Pittis (1999) showed that substantial power gains can be achieved using the CADF test, without incurring severe size distortions.

In this paper we couple the two approaches, extending Hansen's CADF test to small panels. Although Hansen (1995) is the seminal paper concerning covariate-augmented unit root tests, other tests might have been considered. In fact, Elliott and Jansson (2003) show that Hansen's CADF test is not the point optimal test in general, and that feasible point optimal tests based on VAR models can be derived. However, we prefer to use the test proposed in Hansen (1995) for three main reasons. First, simulations reported in Elliott and Jansson (2003) show that the feasible point optimal tests can give power gains at the cost of inferior size performances: this is important in our framework, because Hanck (2008) shows that size distortions tend to cumulate in panel tests of the kind proposed here. Second, Hansen's CADF test is based on the familiar ADF framework, so that it can be more appealing to practitioners once the computational burden related to the computation of the test p-values is eased. Finally, we show that under conditions considered as especially relevant for the panel unit root hypothesis, the CADF test is based on the correct conditional model.

The extension we propose is based on a p-value combination approach advocated independently in Maddala and Wu (1999) and Choi (2001). In this paper we refer mainly to Choi's Z-test, that combines the p-values computed from unit root tests applied to each time series in the panel using an inverse-normal formulation. The method is well grounded in the meta-analytic tradition and its choice is supported by several reasons.

First, provided that we can compute the p-values of the CADF test, the extension to the panel case is straightforward: the panel test is very easy to compute and intuitive and practitioners can track without difficulty what is going on step-by-step in the analysis, from the univariate to the panel case. Second, the asymptotics carries through for the temporal size $T \to \infty$, without requiring also the number of cross-section units $N \to \infty$ as other approaches instead do: in our view, given that allowing $N \to \infty$ in typical macro-panel applications is an implausible hypothesis, this is an extremely important feature of the tests based on Choi (2001). In fact, the test we propose here is especially well suited for small to moderate values of N. Third, we don't need balanced panel data sets, so that individual time series may come in different lengths and span different sample periods: this can be very useful in practice for example when data from many different countries have to be utilized. However, when data come from a balanced macro panel, quite natural stationary covariates can be used for each equation, as suggested in Pesaran (2007) and Chang and Song (2009). Fourth, the test allows for heterogeneous panels: the stochastic as well as the non stochastic components can be different across individual time series. Last, the alternative hypothesis doesn't have to be that all the individual time series are stationary: the alternative that considers that some individual time series have a unit root and others do not can be dealt with by using the tests built upon Choi (2001). Indeed, we deliberately deal with the null hypothesis that all of the series in the panel are $I(1)$ against the alternative that *at least* one of the series is $I(0)$. In fact, this hypothesis is common to most tests for a unit root in panels. Some authors consider this as a disadvantage (see Taylor and Sarno, 1998, among others), but we believe that the extent to which this is a real limitation depends on the specific goal of the analysis.

On the other hand, a potentially serious drawback of the methodology advocated in Choi (2001) is that it is based on the hypothesis that the individual time series are cross-sectionally independent. Indeed, this is a common assumption of many papers dealing with panel unit roots and panel cointegration (see Banerjee, 1999; Baltagi and Kao, 2000; Choi, 2006, for comprehensive surveys). However, it is well known (see e.g. O'Connell, 1998; Maddala and Wu, 1999; Banerjee *et al.*, 2004, 2005; Gengenbach *et al.*, 2006; Lyhagen, 2008; Wagner, 2008) that both short-term and long-run cross-section dependence adversely affects the performance of these panel unit roots tests. Therefore, we extend the approach to cross-sectionally dependent panel units by using the p-value correction method advocated in Hartung (1999) and Demetrescu *et al.* (2006).[1]

Although developed independently, the results reported in the present paper are related to other recent research. Despite some similarities, even in the name, the panel-CADF (pCADF) test presented here should not be confused with the cross-sectionally augmented ADF (CADF) test advocated in Pesaran (2007).[2] The CADF-CIPS test developed by Pesaran is explicitly derived with the aim of addressing directly the problem of

[1] Hartung's correction has been utilized in other recent papers: see, among others, Hassler and Tarcolea (2005) and Westerlund and Costantini (2009).

[2] Notwithstanding the similarity of the names with Pesaran's test, we think that it is fair to refer to the original Hansen's test using the original acronym CADF proposed by Hansen himself. In order to minimize confusion with Pesaran's test, we label our panel extension as pCADF.

cross-sectional dependence. Also Pesaran's test is related to Hansen (1995), but model augmentation takes place using non-stationary covariates. Furthermore, differently from the pCADF test we propose here, in Pesaran (2007) the asymptotic results are derived under $N \to \infty$, either with a fixed T or with $T \to \infty$ sequentially or jointly with N.[3] Chang and Song (2009) also start from the observation that using stationary covariates can greatly improve the power of unit root tests. However, the approach developed in Chang and Song (2009) is rather different from ours: while we use a simple p-value combination approach, Chang and Song (2009) propose a method based on non-linear IV estimation of the autoregressive coefficient, the suggested instruments being non-linear transformations of the lagged levels: this procedure should allow coping with cross-sectional dependencies of unknown form. In fact, Chang and Song (2009) show that the IV-based t-ratios associated with the autoregressive parameters are asymptotically independent even in the presence of cross-sectionally dependent time series. The test is proposed in three variants based on the average, the min, or the max t-ratio, depending on the specific null and alternative hypothesis.[4]

The rest of the paper is organized as follows. Section 2 is devoted to a brief discussion of the test proposed in Hansen (1995). We also illustrate the method we use to obtain the necessary p-values.[5] Indeed, this is a subsidiary, but we believe important, contribution of this paper. In fact, while critical values of Hansen's test are readily available from Hansen (1995), to the best of our knowledge, no other procedure has been proposed so far for the numerical computation of the test p-values. Section 3 is devoted to a brief account of the inverse normal combination method and its modifications to deal with dependent time series. In Section 4 an extensive Monte Carlo analysis of the pCADF test is carried out. The Data Generating Process (DGP) we propose in the paper encompasses other DGPs that are commonly used in the panel unit root literature and it is also related to the DGP used in Hansen (1995). Beside giving us more flexibility in the design of the experiments, our DGP allows us to complement Hansen (1995) and Caporale and Pittis (1999) with new theoretical results and interpretations of the simulations outcomes. The performance of the pCADF test is compared to that of other important panel unit root tests, namely those advocated in Chang and Song (2009), Demetrescu *et al.* (2006) and Moon and Perron (2004). All these tests allow for cross-dependence and share the same null and alternative hypothesis. In Section 5 we show that when the null hypothesis is H_0 : "all of the series are $I(1)$" and the alternative is H_1 : "at least one series is $I(0)$", repeated application of individual unit root tests generates huge size distortions. This is an often neglected reason to prefer panel tests in such circumstances. For the purpose of illustration, in Section 6 we apply our pCADF test to the PPP hypothesis and to interna-

[3]However, Pesaran (2007) shows that satisfactory size and power properties can be obtained even for rather small values of N.

[4]In fact Chang and Song (2009) consider three different formulations of the unit root hypothesis: (A) H_0 : all of the series are $I(1)$ against H_1 : all of the series are $I(0)$; (B) H_0 : all of the series are $I(1)$ against H_1 : at least one of the series is $I(0)$; (C) H_0 : some of the series are $I(1)$ against H_1 : all of the series are $I(0)$.

[5]The algorithm to derive the p-values is described in detail in Appendix A. The R (**?**) package **CADFtest** (Lupi, 2009) that computes Hansen's test and its p-values can be freely downloaded from the Comprehensive R Archive Network (CRAN) at www.cran.r-project.org/package=CADFtest. GAUSS procedures to compute Hansen's test p-values are available from the authors upon request.

tional industrial production indices. The last Section concludes. An Appendix describes the algorithm used to compute the p-values of Hansen's test.

2 The CADF test and the p-values approximation

The CADF test proposed in Hansen (1995) starts from the idea that real economic phenomena are not univariate in general. Therefore, using extra information in unit root testing can make test regressions more efficient, allowing more precise inferences.

Formally, Hansen (1995) assumes that the series y_t to be tested for a unit root can be written as

$$\begin{aligned} y_t &= d_t + s_t & (1)\\ a(L)\Delta s_t &= \delta s_{t-1} + v_t & (2)\\ v_t &= \boldsymbol{b}(L)' (\Delta \boldsymbol{x}_t - \boldsymbol{\mu}_x) + e_t & (3) \end{aligned}$$

where d_t is a deterministic term (usually a constant or a constant and a linear trend), $a(L) := (1 - a_1 L - a_2 L^2 - \ldots - a_p L^p)$ is a polynomial in the lag operator L, $\boldsymbol{x}_t$ is an m-vector, $\boldsymbol{\mu}_x := \mathrm{E}(\Delta \boldsymbol{x})$, $\boldsymbol{b}(L) := (\boldsymbol{b}_{q_2} L^{-q_2} + \ldots + \boldsymbol{b}_{q_1} L^{q_1})$ is a polynomial where both leads and lags are allowed. Furthermore, consider the long-run covariance matrix

$$\boldsymbol{\Omega} := \sum_{k=-\infty}^{\infty} \mathrm{E}\left[\begin{pmatrix} v_t \\ e_t \end{pmatrix} \begin{pmatrix} v_{t-k} & e_{t-k} \end{pmatrix} \right] = \begin{pmatrix} \omega_v^2 & \omega_{ve} \\ \omega_{ve} & \omega_e^2 \end{pmatrix} \quad (4)$$

and define the long-run squared correlation between v_t and e_t as

$$\rho^2 := \frac{\omega_{ve}^2}{\omega_v^2 \, \omega_e^2} \, . \quad (5)$$

When $\Delta \boldsymbol{x}_t$ explains nearly all the zero-frequency variability of v_t, then $\rho^2 \approx 0$. On the contrary, when $\Delta \boldsymbol{x}_t$ has no explicative power on the long-run movement of v_t, then $\rho^2 \approx 1$. Furthermore, as emphasized by Hansen (1995, p. 1151), when e_t is uncorrelated with $\Delta \boldsymbol{x}_{t-k}$ $\forall k$, then $\rho^2 = \omega_e^2 / \omega_v^2$. The case $\rho^2 = 0$ is ruled out (Hansen, 1995, p. 1151), which implies that y_t and $\boldsymbol{x}_t$ cannot be cointegrated.

Similarly to the conventional ADF test, the CADF test is based on three different models representing the "no-constant", "with constant", and "with constant and trend" case, respectively

$$\begin{aligned} a(L)\Delta y_t &= \delta y_{t-1} + \boldsymbol{b}(L)' \Delta \boldsymbol{x}_t + e_t & (6)\\ a(L)\Delta y_t &= \mu + \delta_\mu y_{t-1} + \boldsymbol{b}(L)' \Delta \boldsymbol{x}_t + e_t & (7)\\ a(L)\Delta y_t &= \mu^* + \theta\, t + \delta_\tau y_{t-1} + \boldsymbol{b}(L)' \Delta \boldsymbol{x}_t + e_t & (8) \end{aligned}$$

and is computed as the t-statistic for δ, $\widehat{t(\delta)}$. Hansen (1995, p. 1154) proves that under the unit-root null, the asymptotic distribution of $\widehat{t(\delta)}$ in (6) is

$$\widehat{t(\delta)} \xrightarrow{w} \rho \frac{\int_0^1 W\,dW}{\left(\int_0^1 W^2\right)^{1/2}} + \left(1-\rho^2\right)^{1/2} \mathrm{N}(0,1) \tag{9}$$

where W is a standard Wiener process and $\mathrm{N}(0,1)$ is a standard normal independent of W. Therefore, the asymptotic distribution is a weighted sum of a Dickey-Fuller and a standard normal distribution. As a consequence, if $\rho^2 \neq 1$, conventional ADF critical values would lead to a conservative test.

The asymptotic distribution of the test statistic depends on the nuisance parameter ρ^2 but, provided ρ^2 is given, it can be simulated using standard techniques. The mathematical expression remains unchanged if a model with constant ($\widehat{t(\delta_\mu)}$) or a model with constant and trend ($\widehat{t(\delta_\tau)}$) are considered, except that demeaned and detrended Wiener processes are used instead of the standard Wiener process W.

In order to extend Hansen's CADF unit root test to the panel case using the approach outlined in Choi (2001), we need to compute the p-values of the CADF unit root distribution.

We derive the quantiles of the asymptotic distribution for different values of ρ^2. Given that our goal is the computation of p-values, we simulate the distributions for 40 values of ρ^2 ($\rho^2 = 0.025, 0.05, 0.0725, \ldots, 1$) using $100,000$ replications for each value of ρ^2 and $T = 5,000$ as far as the Wiener functionals are concerned.[6] From the simulated values we derive $1,005$ estimated asymptotic quantiles, (0.00025, 0.00050, 0.00075, 0.001, 0.002, ..., 0.998, 0.999, 0.99925, 0.99950, 0.99975).

Figure 1 reports the estimated asymptotic quantiles for the model with constant, without any smoothing. The surface is extremely regular.[7] Similar considerations carry over for the "no constant" and the "constant plus trend" cases. Therefore we expect that the simulated values can be successfully used to derive asymptotic p-values along lines similar to MacKinnon (1996).

In order to derive p-values from tabulated quantiles of a given distribution, MacKinnon (1996, p. 610) proposed using a local approximation of the kind

$$\Phi^{-1}(p) = \gamma_0 + \gamma_1 \widehat{q(p)} + \gamma_2 \widehat{q(p)}^2 + \gamma_3 \widehat{q(p)}^3 + \nu_p \tag{10}$$

where $\Phi^{-1}(p)$ is the inverse of the cumulative standard normal distribution function evaluated at p and $\widehat{q(p)}$ is the estimated quantile.[8] Equation (10) is not estimated globally (as one would do with a standard response surface). Rather, it is estimated only over a relatively small number of points, in order to obtain a *local* approximation (see MacKinnon, 1996, p. 610, for details).

With respect to MacKinnon (1996), we have the extra difficulty that we have to deal

[6] Simulations have been carried out using R (see **?**).

[7] Figure 1 reports the estimated asymptotic quantiles using a coarser resolution than the one used in the computations.

[8] In MacKinnon (1996) approximate finite sample quantiles are used, instead of the asymptotic ones.

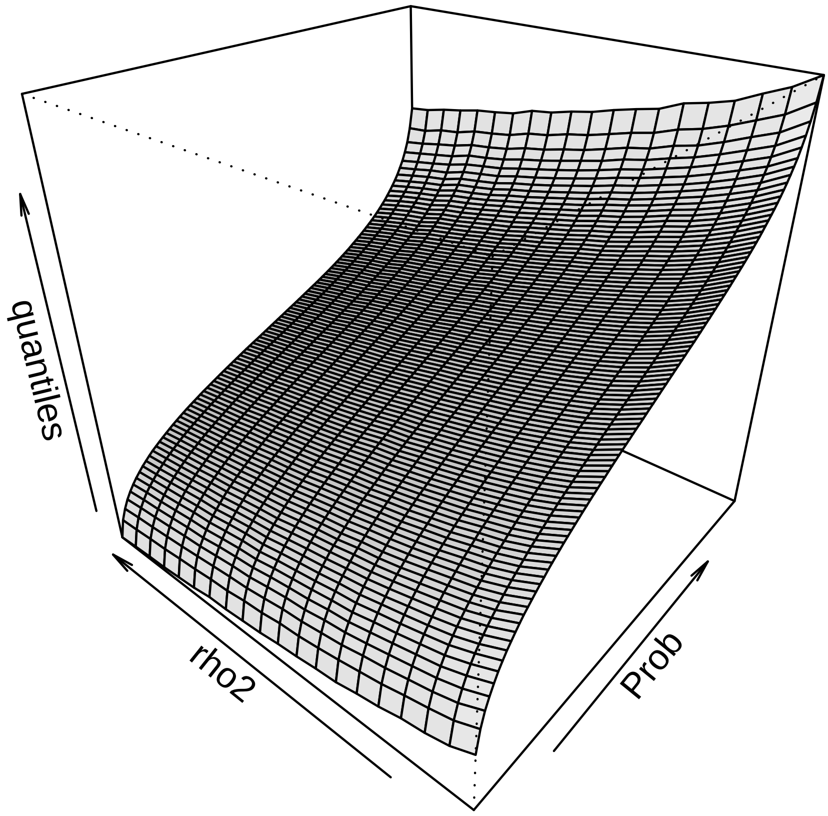

Figure 1 – Estimated asymptotic quantiles of $\widehat{t(\delta_\mu)}$.

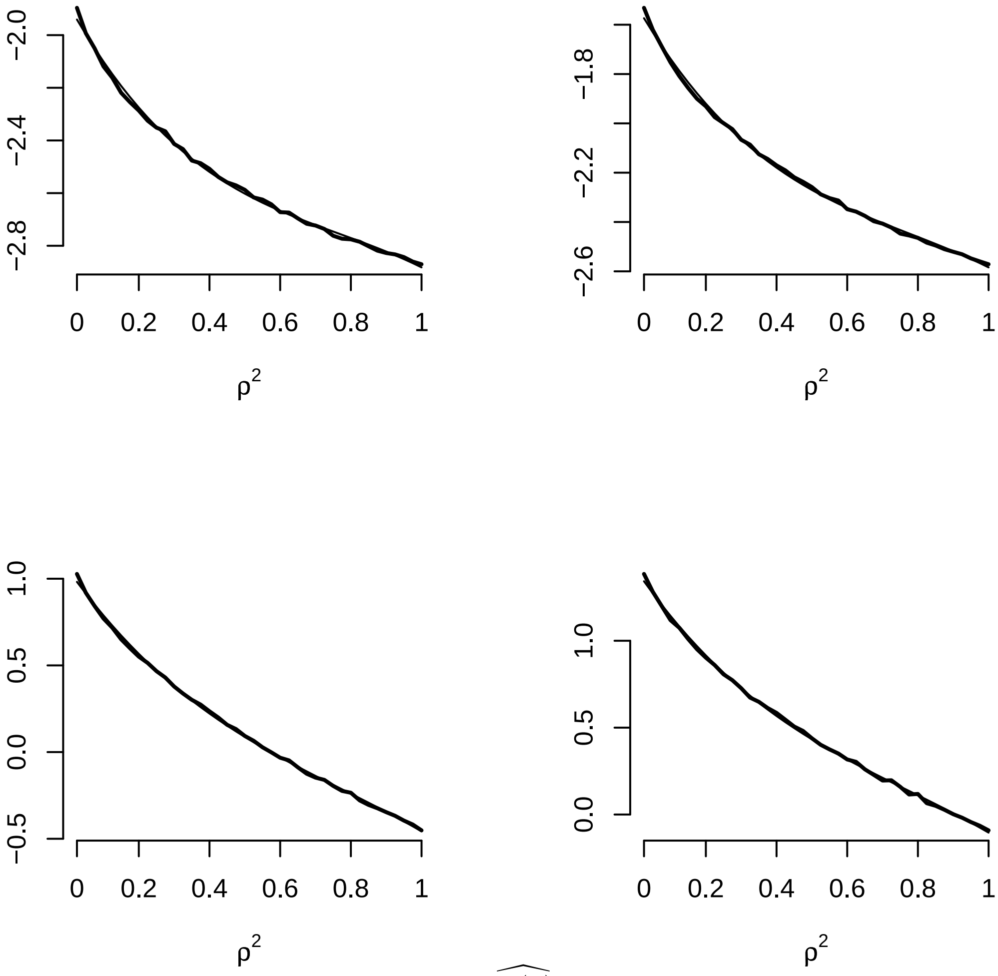

Figure 2 – Interpolation of the quantiles $\widehat{q_\rho(p)}$. From upper-left clockwise: $\alpha = 5\%$, $\alpha = 10\%$, $\alpha = 95\%$, $\alpha = 90\%$. The thick solid lines are the simulated quantiles. The thin lines are the interpolated values.

	Standard			Demeaned			Detrended		
ρ^2	1%	5%	10%	1%	5%	10%	1%	5%	10%
0.05	-2.426	-1.740	-1.380	-2.661	-1.987	-1.626	-2.794	-2.125	-1.767
0.10	-2.450	-1.770	-1.410	-2.760	-2.091	-1.733	-2.937	-2.274	-1.921
0.15	-2.470	-1.795	-1.436	-2.847	-2.183	-1.829	-3.063	-2.408	-2.058
0.20	-2.488	-1.818	-1.460	-2.924	-2.266	-1.915	-3.175	-2.527	-2.181
0.25	-2.503	-1.837	-1.481	-2.990	-2.339	-1.991	-3.274	-2.633	-2.291
0.30	-2.515	-1.854	-1.500	-3.049	-2.403	-2.060	-3.360	-2.727	-2.389
0.35	-2.525	-1.868	-1.517	-3.099	-2.460	-2.121	-3.436	-2.810	-2.476
0.40	-2.534	-1.880	-1.531	-3.142	-2.510	-2.174	-3.502	-2.883	-2.553
0.45	-2.540	-1.890	-1.544	-3.179	-2.554	-2.222	-3.560	-2.947	-2.622
0.50	-2.545	-1.898	-1.555	-3.211	-2.593	-2.265	-3.610	-3.005	-2.683
0.55	-2.550	-1.905	-1.565	-3.239	-2.628	-2.303	-3.654	-3.055	-2.738
0.60	-2.553	-1.911	-1.573	-3.264	-2.658	-2.338	-3.693	-3.101	-2.788
0.65	-2.555	-1.916	-1.581	-3.286	-2.686	-2.369	-3.729	-3.143	-2.834
0.70	-2.558	-1.921	-1.587	-3.307	-2.712	-2.399	-3.762	-3.183	-2.877
0.75	-2.560	-1.925	-1.593	-3.327	-2.737	-2.428	-3.794	-3.220	-2.919
0.80	-2.563	-1.929	-1.598	-3.348	-2.762	-2.456	-3.825	-3.258	-2.960
0.85	-2.566	-1.933	-1.603	-3.371	-2.787	-2.484	-3.858	-3.296	-3.002
0.90	-2.569	-1.938	-1.608	-3.395	-2.814	-2.514	-3.893	-3.336	-3.046
0.95	-2.574	-1.944	-1.613	-3.423	-2.843	-2.545	-3.932	-3.379	-3.093
1.00	-2.580	-1.950	-1.618	-3.455	-2.874	-2.580	-3.975	-3.427	-3.144

Table 1 – Asymptotic critical values of the CADF test.

with the nuisance parameter ρ^2, so that the local approximation must be obtained along two dimensions. However, given that quantiles change fairly smoothly by varying ρ^2, we adopt a rather straightforward two-step procedure. In the first step we interpolate the quantiles $\widehat{q(p)}$ to obtain an approximation for the relevant value of ρ^2. In practice we use

$$\widehat{q_\rho(p)} = \beta_0 + \beta_1 \rho^2 + \beta_2 \rho^4 + \beta_3 \rho^6 + \varepsilon_\rho \tag{11}$$

where we have used the subscript ρ in $\widehat{q_\rho(p)}$ to indicate the dependence of the quantiles on ρ^2. Interpolation is always very good, as can be gathered from Figure 2.

As a by-product of our analysis, we compute a detailed table of asymptotic critical values of the CADF test using equation (11) (see Table 1). Given that these critical values are based on a larger number of replications and on a response surface approach (see e.g. Hendry, 1984), we believe that they can be more accurate than those reported in Hansen (1995).

Finally we apply the procedure advocated in MacKinnon (1996) on the interpolated quantiles to obtain the p-values.[9]

3 The inverse normal combination test

Once computation of the p-values for the distribution (9) is solved, the extension of Hansen's test to the panel case is straightforward. Indeed, Choi (2001) shows that un-

[9]A more detailed description of the procedure is reported in Appendix A.

der some fairly general regularity conditions, if the cross-section units $i = 1, \dots, N$ are independent, under the null

$$Z := \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \widehat{t}_i \xrightarrow{w} \mathrm{N}(0,1) \tag{12}$$

where the $\widehat{t}_i$'s are the *probits* $\widehat{t}_i := \Phi^{-1}(\widehat{p}_i)$, with $\Phi(\cdot)$ the standard normal cumulative distribution function, and $\widehat{p}_i$ the estimated individual p-values for $i = 1, \dots, N$. Convergence in (12) takes place as $T \to \infty$, whereas $N < \infty$ is the number of individual time series. $T \to \infty$ is required for the relevant statistics to converge to a proper continuous distribution, under some regularity conditions. The null hypothesis is $\mathrm{H}_0 : \delta = 0 \; \forall i$, while the alternative is $\mathrm{H}_1 : \delta < 0$ for at least one i, with $i = 1, 2, \dots, N$. This is a different alternative from $\mathrm{H}_1 : \delta < 0 \; \forall i$, used in other tests (see e.g. Levin and Lin, 1993; Levin *et al.*, 2002; Quah, 1994). In fact, we believe that our formulation of the null and of the alternative hypothesis is an advantage, rather than a disadvantage, as some authors claim: the null hypothesis that all of the series are $I(1)$ against the alternative that all of the series are $I(0)$ is not very interesting and informative, given that it can be tested only under the maintained hypothesis that the crucial parameter characterising the presence/absence of the unit root is homogeneous across the individual time series in the panel (see e.g. Levin *et al.*, 2002). Also from the economist's point of view there are instances in which it can be more interesting to test for a unit root collectively over a whole panel of time series because this can be interpreted as a stylized fact that can give stronger support in favour (or against) a particular economic interpretation as compared to the same analysis conducted separately on each single time series. Furthermore, because of the presence of multiple testing the two approaches are not statistically equivalent.

However, the presence of cross-section dependence among the time series complicates substantially the theoretical framework, and the test statistic is no longer asymptotically (with T) standard normal. However, Hartung (1999) suggests that a suitably modified inverse normal combination test can be obtained. The advantage of this solution is that under the null the test statistic has approximately standard normal distribution even in the presence of correlated individual test outcomes. In particular, Hartung (1999) analyses the case where the pairwise correlation across the individual test statistics is constant and equal to ϱ, say. If ϱ were known then, given a set $\lambda_i, \dots, \lambda_N$ of real valued weights such that $\sum_{i=1}^{N} \lambda_i \neq 0$, it would be possible to compute

$$t(\varrho) := \frac{\sum_{i=1}^{N} \lambda_i \widehat{t}_i}{\sqrt{(1-\varrho) \sum_{i=1}^{N} \lambda_i^2 + \varrho \left(\sum_{i=1}^{N} \lambda_i \right)^2}} \tag{13}$$

which under the null would be $\mathrm{N}(0,1)$. When $\varrho = 0$ (no cross-section dependence) and $\lambda_i = 1 \; \forall i$, then (13) collapses into (12).

Of course, ϱ is not known, and the feasible test statistic advocated by Hartung (1999,

p. 851) is

$$t\left(\hat{\varrho}^{*},\kappa\right) := \frac{\sum_{i=1}^{N}\lambda_i\widehat{t_i}}{\sqrt{\sum_{1=1}^{N}\lambda_i^2+\left[\left(\sum_{i=1}^{N}\lambda_i\right)^2-\sum_{i=1}^{N}\lambda_i^2\right]\left(\hat{\varrho}^{*}+\kappa\sqrt{\frac{2}{N+1}}(1-\hat{\varrho}^{*})\right)}} \tag{14}$$

where $\hat{\varrho}^*$ is a consistent estimator of ϱ such that $\hat{\varrho}^* = \max\{-1/(N-1), \hat{\varrho}\}$ with $\hat{\varrho} = 1-(N-1)^{-1}\sum_{i=1}^{N}(\widehat{t_i}-N^{-1}\sum_{i=1}^{N}\widehat{t_i})^2$. $\kappa > 0$ is a parameter that controls the small sample actual significance level. Hartung (1999) shows that under the null $t(\hat{\varrho}^*, \kappa)$ is approximately distributed as $\mathrm{N}(0,1)$. However, the proof offered in Hartung (1999) rests on the assumption that the probits are not only individually $\mathrm{N}(0,1)$, but are also jointly multivariate normal.

Demetrescu *et al.* (2006) generalize Hartung's results in two directions. They first show that the pairwise correlation of the individual test statistics need not be constant for Hartung's results to hold (Demetrescu *et al.*, 2006, Proposition 1, p. 651). Furthermore, they wonder under what conditions does the inverse normal method map the original test statistics to a multivariate normal distribution of the probits and they conclude that the necessary and sufficient condition for $t(\varrho)$ to have a standard normal distribution is that the test statistics from which the probits are derived are such to have the copula of a multivariate normal distribution (Demetrescu *et al.*, 2006, Proposition 2, p. 653). Despite the fact that the augmented Dickey-Fuller test does not satisfy the condition stated in their Proposition 2, Demetrescu *et al.* (2006) suggest that correcting for dependence using (14) may still be a good practice because units cross-correlation is likely to have much stronger adverse effects on inference than deviations from normality of the individual test statistics can have. Indeed, they show by simulation that this is in fact the case.

In this paper we follow the approach suggested by Demetrescu *et al.* (2006) to combine the p-values of the individual CADF unit root tests in the presence of cross-section dependence. We argue that, if the correction proposed in Hartung (1999) works quite nicely in the presence of Dickey-Fuller distributions, it should *a fortiori* work at least as nicely in the presence of distributions that are closer to the standard normal. In other words, given that under the null Hansen's distribution is precisely a weighted sum of a Dickey-Fuller and a standard normal distribution, we expect that the correction for cross-section dependence in our case should be at least as effective as it is in the standard Dickey-Fuller case explored by Demetrescu *et al.* (2006).

4 Monte Carlo simulations

In this Section we compare the performance of the *p*CADF test to that of three unit root tests that are valid under cross-dependence. Specifically, we compare our test with an ADF-based p-values combination test (Demetrescu *et al.*, 2006), with a dynamic factor test (Moon and Perron, 2004) and with a recent IV-based covariate-augmented test (Chang and Song, 2009). For the latter two tests we consider in particular the t_a^* statistic (Moon and Perron, 2004, p. 92) and the *minimum-t* version of the test (see Chang and Song, 2009, pp. 905–906), respectively. All these tests share the same null H_0 : "all of the series are $I(1)$" and the same alternative H_1 : "at least one series is $I(0)$".

We verify the performances of the pCADF test and of the test proposed in Demetrescu *et al.* (2006) using the versions of the tests with constant and with constant and linear trend.[10] The tests advocated in Chang and Song (2009) and Moon and Perron (2004) are examined in both the demeaned and detrended versions.

4.1 Structure of the DGP

In our simulations we consider the following DGP:

$$\Delta \boldsymbol{y}_t = \boldsymbol{\alpha} + \boldsymbol{D}\boldsymbol{y}_{t-1} + \boldsymbol{u}_t \tag{15}$$

$$\begin{pmatrix} \boldsymbol{u}_t \\ \xi_t \end{pmatrix} = \begin{pmatrix} \boldsymbol{B} & \boldsymbol{\gamma} \\ \boldsymbol{0}' & \lambda \end{pmatrix} \begin{pmatrix} \boldsymbol{u}_{t-1} \\ \xi_{t-1} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\eta}_t \\ \varepsilon_t \end{pmatrix} \tag{16}$$

$$\begin{pmatrix} \boldsymbol{\eta}_t \\ \varepsilon_t \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} \boldsymbol{0} \\ 0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\sigma}_{12} \\ \boldsymbol{\sigma}_{12}' & \sigma_{22} \end{pmatrix} \right] \tag{17}$$

where Δ is the usual difference operator, $\boldsymbol{y}_t := (y_{1t}, \ldots, y_{Nt})'$, $\boldsymbol{u}_t := (u_{1t}, \ldots, u_{Nt})'$, $\boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_N)'$, $\boldsymbol{D} := \mathrm{diag}(\delta_1, \ldots, \delta_N)$, $\boldsymbol{B} := \mathrm{diag}(\beta_1, \ldots, \beta_N)$, $\boldsymbol{\gamma} := (\gamma_1, \ldots, \gamma_N)'$ and $\boldsymbol{\eta}_t := (\eta_{1t}, \ldots, \eta_{Nt})'$. Note that (16) defines a VAR(1) which is stationary as long as $|\beta_i| < 1\ \forall i$ and $|\lambda| < 1$.[11] $\delta_i = 0\ \forall i$ under the null, while under the alternative $\delta_i < 0$ for some i.

We believe that the proposed DGP is especially interesting, because it can be viewed as a panel extension of the DGP proposed in Hansen (1995, p. 1161) and at the same time is also a generalization of two DGPs commonly used in the panel unit root literature (see e.g. Chang and Song, 2009; Phillips and Sul, 2003). The two DGPs that are special cases of ours, when $\boldsymbol{\alpha} = \boldsymbol{0}$ share the same equation (15) for $\Delta \boldsymbol{y}_t$, but differ as far as the simulation of the $\boldsymbol{u}_t$'s is concerned:

$$\text{DGP1:} \quad u_{it} = \beta_i u_{i,t-1} + \nu_{it} \tag{18}$$

$$\text{DGP2:} \quad u_{it} = \beta_i u_{i,t-1} + \gamma_i \zeta_t + \nu_{it} \tag{19}$$

where the N-vector $\boldsymbol{\nu}_t$ is i.i.d. $\mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{11})$ with $\boldsymbol{\Sigma}_{11} \neq \boldsymbol{I}$ and ζ_t is a i.i.d. $\mathrm{N}(0, 1)$ common factor independent of $\boldsymbol{\nu}_t$.

It can be seen that, even when $\boldsymbol{\alpha} = \boldsymbol{0}$, our DGP (15)–(17) is more general than both (18) and (19): in fact, in our DGP the "common factor" ξ_t can be autocorrelated and non-zero correlations between the innovations to $u_{i,t}$ and the innovations to ξ_t can be introduced. As a result, the cross-dependence structure is stronger than in either DGP1 or DGP2.

[10] We do not consider the models without deterministic terms that are less relevant in practical applications.

[11] To see this, let's define

$$\boldsymbol{\Phi} := \begin{pmatrix} \boldsymbol{B} & \boldsymbol{\gamma} \\ \boldsymbol{0}' & \lambda \end{pmatrix}$$

and note that $\boldsymbol{\Phi}$ is upper-triangular, so that the eigenvalues of $\boldsymbol{\Phi}$ are given simply by the diagonal elements of $\boldsymbol{\Phi}$, $\mathrm{dg}(\boldsymbol{\Phi})$. Therefore, the VAR(1) is stationary as long as $|\beta_i| < 1\ \forall i$ and $|\lambda| < 1$.

However, DGP2 can be derived as a special case from (15)–(17) when $\lambda = 0$ and $\boldsymbol{\sigma}_{12} = \mathbf{0}$, while DGP1 is retrieved if in addition $\boldsymbol{\gamma} = \mathbf{0}$. In both cases, in general $\boldsymbol{\Sigma}_{11} \neq \boldsymbol{I}$.

Using the DGP (15)–(17) we can determine the form of the model that should be used to test for a unit root in each single y_{it}. For simplicity, assume now $\boldsymbol{\alpha} = \mathbf{0}$. Then, denoting the "past" by $\boldsymbol{Z}_{t-1}$, the correct conditional model for $\Delta y_{i,t}$ is

$$\begin{aligned} \mathrm{E}\left(\Delta y_{i,t} \middle| \xi_t, \boldsymbol{Z}_{t-1}\right) &= \delta_i \left(1-\beta_i\right) y_{i,t-1} + \left(1+\delta_i\right) \beta_i \Delta y_{i,t-1} \\ &+ \frac{\left(\boldsymbol{\sigma}_{12}\right)_i}{\sigma_{22}} \xi_t + \left(\gamma_i - \frac{\left(\boldsymbol{\sigma}_{12}\right)_i}{\sigma_{22}} \lambda\right) \xi_{t-1} \, . \end{aligned} \tag{20}$$

with $\left(\boldsymbol{\sigma}_{12}\right)_i$ the i-th element of $\boldsymbol{\sigma}_{12}$. Note that (20) has the form of a CADF(1,1,0) model. In fact, unless $\boldsymbol{\gamma} = 0$ and $\boldsymbol{\sigma}_{12} = 0$, the standard approach of using a panel combination ADF test in a context where the DGP is supposed to be of the kind of (15)–(17) (which is a fairly standard situation in the panel unit root literature) is bound to be at least inefficient, because the correct models should include ξ_t and/or ξ_{t-1} and the individual tests should be CADF. Even if $\gamma_i = 0$ (i.e., when ξ_t does not Granger-cause u_t), as far as $\left(\boldsymbol{\sigma}_{12}\right)_i \neq 0$ the correct model has the form of a CADF(1,1,0).

Expression (20) is very similar to an expression derived in Caporale and Pittis (1999, p. 586, equation 11) and some special cases can be of interest. Under DGP2 ($\lambda = 0$ and $\boldsymbol{\sigma}_{12} = \mathbf{0}$) the correct conditional model becomes

$$\mathrm{E}\left(\Delta y_{i,t} \middle| \xi_t, \boldsymbol{Z}_{t-1}\right) = \delta_i \left(1-\beta_i\right) y_{i,t-1} + \left(1+\delta_i\right) \beta_i \Delta y_{i,t-1} + \gamma_i \xi_{t-1} \tag{21}$$

and we should expect the pCADF test to have a better performance than the tests based on the conventional ADF. Of course, the same conditional model (21) holds for the i-th unit if only $\left(\boldsymbol{\sigma}_{12}\right)_i = 0$, while if $\lambda = 0$ and $\left(\boldsymbol{\sigma}_{12}\right)_i \neq 0$ we have

$$\begin{aligned} \mathrm{E}\left(\Delta y_{i,t} \middle| \xi_t, \boldsymbol{Z}_{t-1}\right) &= \delta_i \left(1-\beta_i\right) y_{i,t-1} + \left(1+\delta_i\right) \beta_i \Delta y_{i,t-1} \\ &+ \frac{\left(\boldsymbol{\sigma}_{12}\right)_i}{\sigma_{22}} \xi_t + \gamma_i \xi_{t-1} \, . \end{aligned} \tag{22}$$

On the other hand, under DGP1 ($\lambda = 0$, $\boldsymbol{\sigma}_{12} = \mathbf{0}$, $\boldsymbol{\gamma} = \mathbf{0}$), the correct conditional model is simply

$$\mathrm{E}\left(\Delta y_{i,t} \middle| \xi_t, \boldsymbol{Z}_{t-1}\right) = \delta_i \left(1-\beta_i\right) y_{i,t-1} + \left(1+\delta_i\right) \beta_i \Delta y_{i,t-1} \tag{23}$$

which has the form of an ordinary ADF(1) test equation, so that in this case the pCADF test has no advantage on p-values combination tests based on the ADF test.

From the discussion in Section 2, we know that the power of the CADF test depends crucially on the nuisance parameter ρ^2. Therefore, the power of the pCADF tests will depend on the values of this parameter for each unit in the panel, ρ_i^2. Using the DGP (15)–(17) we can derive analytically the theoretical value of ρ_i^2 under the DGP.[12] This

[12] Hansen (1995) derives ρ^2 by simulation using different models. None of the models used by Hansen (1995, Table 3, p. 1162) correspond to the correctly specified CADF(1,1,0), so that all the models are either over- or under-parameterized. Using the theoretical results derived below jointly with Hansen's results, we

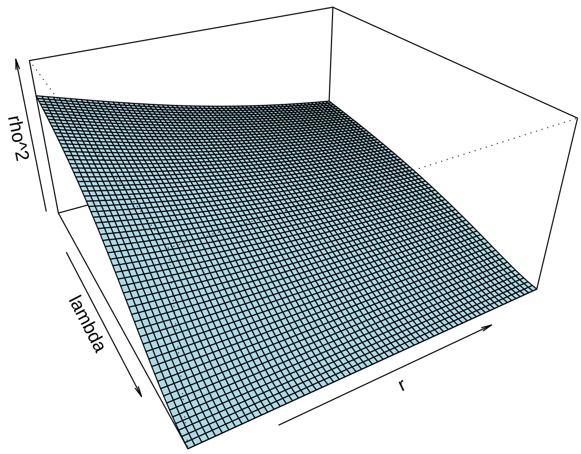

Figure 3 – Values of ρ^2 for varying values of $0 < \lambda < 1$ and $0 < r_i < 1$, $\gamma_i = 0.5$, $\sigma_{e_i} = \sigma_\varepsilon = 1$. ρ^2 is plotted on a $0-1$ scale.

result gives important insights to better investigate the performance of the test in the Monte Carlo experiments.

Consider the residual $e_{i,t}$ from the correct conditional model (20)

$$\begin{aligned} e_{i,t} &= \Delta y_{i,t} - \delta_i (1-\beta_i) y_{i,t-1} - (1+\delta_i) \beta_i \Delta y_{i,t-1} \\ &- \frac{(\boldsymbol{\sigma}_{12})_i}{\sigma_{22}} \xi_t - \left(\gamma_i - \frac{(\boldsymbol{\sigma}_{12})_i}{\sigma_{22}} \lambda \right) \xi_{t-1} . \end{aligned} \quad (24)$$

Given that $e_{i,t}$ is the residual from the correct conditional model, it must be an innovation uncorrelated with ξ_{t-k} $\forall k$. As discussed in Hansen (1995, p. 1151), in this case $\rho_i^2 = \omega_{e_i}^2 / \omega_{v_i}^2$ with ω_h^2 the long-run variance of h, that is the zero-frequency spectral density of h ($h \in \{e_i, v_i\}$). Given that $e_{i,t}$ is an innovation, its long-run variance is just the variance of $e_{i,t}$, apart from the normalizing factor $(2\pi)^{-1}$.

Now consider

$$v_{i,t} = \frac{(\boldsymbol{\sigma}_{12})_i}{\sigma_{22}} \xi_t + \left(\gamma_i - \frac{(\boldsymbol{\sigma}_{12})_i}{\sigma_{22}} \lambda \right) \xi_{t-1} + e_{i,t} . \quad (25)$$

In order to compute the long-run variance of $v_{i,t}$, $\omega_{v_i}^2$, from (16) note that $\xi_t = (1 - \lambda L)^{-1} \varepsilon_t$ and define $r_i := (\boldsymbol{\sigma}_{12})_i / \sigma_{22}$. Then, rewrite (25) as

$$\begin{aligned} v_{i,t} &= [r_i + (\gamma_i - r_i \lambda) L] \xi_t + e_{i,t} \\ &= \frac{r_i + (\gamma_i - r_i \lambda) L}{1 - \lambda L} \varepsilon_t + e_{i,t} . \end{aligned} \quad (26)$$

can show that under-parameterization can result in biased estimates of ρ^2, with adverse effects on inference. This is particularly evident with respect to Hansen's experiments 11 and 15, where the simulation-based estimates of ρ^2 from the CADF(2,0,1) model are equal to 0.87 and 0.90, respectively, while the true values under the DGP are 0.67 and 0.50.

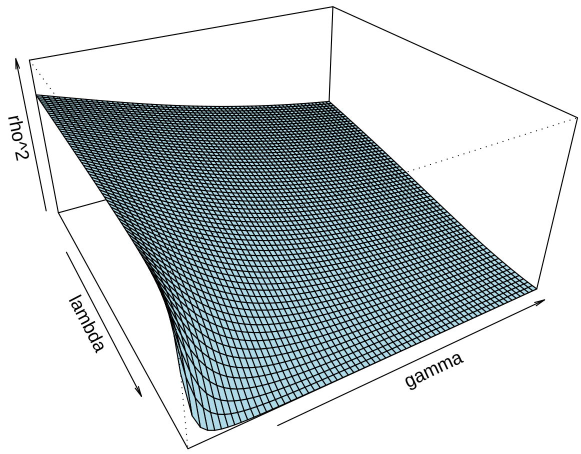

Figure 4 – Values of ρ^2 for varying values of $0 < \lambda < 1$ and $0 < \gamma_i < 1$, $r = 0.5$, $\sigma_{e_i} = \sigma_\varepsilon = 1$. ρ^2 is plotted on a $0 - 1$ scale.

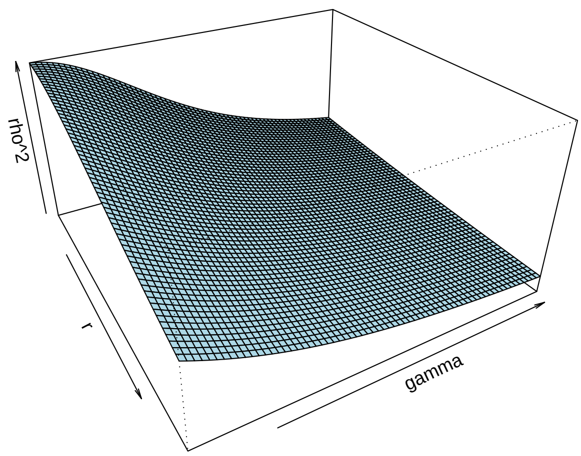

Figure 5 – Values of ρ^2 for varying values of $0 < r < 1$ and $0 < \gamma_i < 1$, $\lambda = 0.5$, $\sigma_{e_i} = \sigma_\varepsilon = 1$. ρ^2 is plotted on a $0 - 1$ scale.

The spectral density of $v_{i,t}$ at frequency ω is

$$f_{v_i}(\omega) \propto \frac{\left|r_i + (\gamma_i - r_i\lambda)\, e^{-i\omega}\right|^2}{\left|1 - \lambda e^{-i\omega}\right|^2}\, \sigma_\varepsilon^2 + \sigma_{e_i}^2 \tag{27}$$

so that the long-run variance of $v_{i,t}$, $\omega_{v_i}^2$, is

$$\omega_{v_i}^2 := f_{v_i}(0) \propto \frac{\left[\gamma_i + (1-\lambda)\, r_i\right]^2}{(1-\lambda)^2}\, \sigma_\varepsilon^2 + \sigma_{e_i}^2 \,. \tag{28}$$

Finally, ρ_i^2 is given by

$$\rho_i^2 = \frac{\omega_{e_i}^2}{\omega_{v_i}^2} = \frac{\sigma_{e_i}^2}{\frac{[\gamma_i + (1-\lambda) r_i]^2}{(1-\lambda)^2}\, \sigma_\varepsilon^2 + \sigma_{e_i}^2} \,. \tag{29}$$

The value of ρ_i^2 is a nonlinear function of $(\boldsymbol{\sigma}_{12})_i$, σ_{22}, γ_i and λ. Contrary to what is suggested in Hansen (1995, p. 1161), we find that the value of λ is crucial in determining the value of the nuisance parameter ρ^2, also when the VAR(1) (16) is stationary. Of course, when $\lambda \to 1$, then $\omega_{v_i}^2 \to \infty$ and $\rho^2 \to 0$: this is an expected result, because if $\lambda = 1$, ξ_t has a unit root and is cointegrated with $y_{i,t}$. Conversely, if $\gamma_i = 0$ and $r_i = 0$, then $\rho_i^2 = 1$: in this case there would be no advantage in using individual CADF tests instead of standard ADF tests. Under DGP2, given that $\lambda = 0$ and $r_i = 0$, ρ_i^2 simply varies inversely with γ_i. Under DGP1, where it is also $\gamma_i = 0\ \forall i$, then $\rho_i^2 = 1\ \forall i$ and the power of the pCADF test is substantially the same as the power of the test based on Demetrescu *et al.* (2006), consistently with what already pointed out while discussing the conditional model.

In (29) the larger are either λ, γ_i or r_i, the smaller is ρ_i^2. Given that the power of the CADF test is higher the smaller is the value of ρ_i^2, this in turn defines the regions where the test is expected to perform better. A graphical summary of the relation between ρ_i^2 and the values of λ, γ_i and r_i is offered in Figures 3–5.

4.2 Parameters setting and experimental design

Some care must be exerted in simulating the DGP (15)–(17), especially as far as the simulation of $(\boldsymbol{\eta}_t', \varepsilon_t)'$ is concerned. From (17), $(\boldsymbol{\eta}_t', \varepsilon_t)' \sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$, with

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\sigma}_{12} \\ \boldsymbol{\sigma}_{12}' & \sigma_{22} \end{pmatrix} . \tag{30}$$

We assume $\mathrm{diag}(\boldsymbol{\Sigma}) := \boldsymbol{\imath}$, with $\boldsymbol{\imath} := (1, \ldots, 1)$ so that the generic element of $\boldsymbol{\sigma}_{12}$, $(\boldsymbol{\sigma}_{12})_i$, coincides with r_i. However, we have to distinguish two different settings for $\boldsymbol{\Sigma}_{11}$, depending on $\boldsymbol{\sigma}_{12} = \mathbf{0}$ or $\boldsymbol{\sigma}_{12} \neq \mathbf{0}$.

When $\boldsymbol{\sigma}_{12} = \mathbf{0}$ (e.g. under DGP1 and DGP2), then we must generate the correlation matrix $\boldsymbol{\Sigma}_{11}$ in a way that is as flexible and unrestricted as possible. At the same time we want to introduce fairly strong dependence. Therefore, we start by generating a symmetric matrix $\boldsymbol{\Sigma}^*$ whose diagonal elements are equal to 1 and whose non-diagonal elements are randomly drawn from $\mathrm{U}_{(0,0.8)}$. Of course, although symmetric, $\boldsymbol{\Sigma}^*$ is not in general positive definite. Therefore, we find a positive definite symmetric matrix $\boldsymbol{\Sigma}^\dagger$ that is "close" to $\boldsymbol{\Sigma}^*$ by computing $\boldsymbol{\Sigma}^\dagger = \boldsymbol{V}^*\boldsymbol{\Lambda}^\dagger\boldsymbol{V}^{*\prime}$ where the matrix $\boldsymbol{V}^*$ is derived from the singular value decomposition of $\boldsymbol{\Sigma}^*$ and $\boldsymbol{\Lambda}^\dagger$ is the diagonal matrix of the eigenvalues of $\boldsymbol{\Sigma}^*$, after substituting the negative eigenvalues with very small but positive values. Finally, the positive definite covariance matrix obtained in this way (the diagonal elements are not exactly equal to one) is transformed into the required correlation matrix $\boldsymbol{\Sigma}_{11}$ by normalization.[13] The resulting symmetric positive definite matrix $\boldsymbol{\Sigma}_{11}$ is such that most of the simulated correlations are positive, as we probably would expect in many empirical macro panel settings, and the average correlation is larger than the one simulated using the method proposed by Chang (2002) and Chang and Song (2009).[14] Furthermore, the simulated $\boldsymbol{\Sigma}_{11}$ is likely to satisfy Proposition 1 in Demetrescu *et al.* (2006).

On the other hand, when $\boldsymbol{\sigma}_{12} \neq \mathbf{0}$ the parameters $r_i := (\boldsymbol{\sigma}_{12})_i$ enter the expression for ρ_i^2 and are therefore important design parameters that we want to control precisely. In this case we want to simulate a correlation matrix $\boldsymbol{\Sigma}$ whose last column is a given vector $(\boldsymbol{\sigma}_{12}', 1)'$. Furthermore, given the vector of correlations $\boldsymbol{\sigma}_{12}$, it is reasonable to consider $\boldsymbol{\Sigma}_{11} \neq \boldsymbol{I}$. However, $\boldsymbol{\Sigma}_{11}$ in this case must be consistent with the given $\boldsymbol{\sigma}_{12}$. Therefore, we introduce a minimal structure in $\boldsymbol{\Sigma}_{11}$ by assuming that its generic off-diagonal element is $(\boldsymbol{\Sigma}_{11})_{ij} := (\boldsymbol{\sigma}_{12})_i (\boldsymbol{\sigma}_{12})_j$ (with $i \neq j$) and $\mathrm{diag}(\boldsymbol{\Sigma}_{11}) := \boldsymbol{\imath}$. This structure essentially states that the more η_{it} is correlated with ε_t and η_{jt} is correlated with ε_t, the more η_{it} is correlated with η_{jt}, that is what we should expect in the usual case. Simulating such a $\boldsymbol{\Sigma}$ is very easy: just draw the elements of $\boldsymbol{\sigma}_{12}$ from a specified distribution, $\mathrm{U}_{(r_{\min}, r_{\max})}$, say, and compute $\boldsymbol{S} = \boldsymbol{\sigma}_{12}\boldsymbol{\sigma}_{12}'$. Set $\mathrm{diag}(\boldsymbol{S}) := \boldsymbol{\imath}$ and call $\boldsymbol{\Sigma}_{11}$ the resulting matrix. Then, build the correlation matrix $\boldsymbol{\Sigma}$ as in (30). The matrix $\boldsymbol{\Sigma}$ simulated in this way is symmetric positive definite.[15]

As already pointed out in the previous subsection, we expect the nuisance parameter ρ^2 to influence the performance of our test. Therefore, rather than embarking in a full factorial design, we concentrate on just a few experiments carefully selected in such a way that they differ in the underlying value of ρ^2 (see Table 2).

[13] The proposed algorithm is essentially equivalent to the procedure advocated in Rebonato and Jäckel (1999, Section 3).

[14] In a pilot simulation carried out using 50,000 replications we found that the average non-diagonal element of a 20×20 simulated correlation matrix was about 0.34 with the simulated correlations spanning the interval $(-0.30, 0.96)$. We also used the procedure outlined in Demetrescu *et al.* (2006, p. 659). The results are very similar to those reported here and are available from the authors upon request.

[15] To see this, note that $\boldsymbol{\Sigma}$ is real symmetric by construction. Then there exists a matrix $\boldsymbol{P}$ such that $\boldsymbol{P}'\boldsymbol{\Sigma}\boldsymbol{P} = \boldsymbol{\Lambda}$, with $\boldsymbol{\Lambda}$ the diagonal matrix of the eigenvalues of $\boldsymbol{\Sigma}$. $\boldsymbol{P}$ and $\boldsymbol{\Lambda}$ can be found using the Schur canonical form:

$$\begin{pmatrix} \boldsymbol{I} & -\boldsymbol{\sigma}_{12} \\ \mathbf{0}' & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\sigma}_{12} \\ \boldsymbol{\sigma}_{12}' & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{I} & \mathbf{0} \\ -\boldsymbol{\sigma}_{12}' & 1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} - \boldsymbol{\sigma}_{12}\boldsymbol{\sigma}_{12}' & \mathbf{0} \\ \mathbf{0}' & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1-(\boldsymbol{\sigma}_{12})_1^2 & 0 & \dots & 0 & 0 \\ 0 & 1-(\boldsymbol{\sigma}_{12})_2^2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1-(\boldsymbol{\sigma}_{12})_N^2 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix} = \boldsymbol{\Lambda}.$$

Since all the eigenvalues of $\boldsymbol{\Sigma}$ are positive, $\boldsymbol{\Sigma}$ is positive definite.

Experiment	λ	γ	r	ρ^2
1	0.0	0.0	0.0	1.000
2	0.0	$U_{(0.7,0.9)}$	0.0	0.610
3	0.2	$U_{(0.7,0.9)}$	$U_{(0.1,0.3)}$	0.410
4	0.5	$U_{(0.1,0.3)}$	$U_{(0.7,0.9)}$	0.410
5	0.2	$U_{(0.7,0.9)}$	$U_{(0.7,0.9)}$	0.236
6	0.5	$U_{(0.7,0.9)}$	$U_{(0.7,0.9)}$	0.148

Table 2 – Parameters setting. The values of ρ^2 are computed using the means of the Uniform distributions.

The other parameters of the DGP are generated as in Chang and Song (2009): in particular, $\beta_i \sim U_{(0.2,0.4)}$ and $\gamma_i \sim U_{(0.5,3)}$ (with $i = 1, \ldots, N$). Under the null $\delta_i = 0\ \forall i$, under the alternative $\delta_i \sim U_{(-0.2,-0.01)}$ for the stationary units. In order to highlight the power of the tests when only a few series are stationary, the number of stationary units under the alternative is limited to 2.

Given that our DGP allows for a non-zero drift α_i, we run the experiments first using $\alpha_i = 0\ \forall i$ and then using $\alpha_i \sim U_{(0.7,0.9)}$.

Finally, the experiments are carried out using 2,500 replications with $T \in \{100, 300\}$ and $N \in \{10, 20\}$ that are fairly typical values in macro-panel applications.

Since the use of the pCADF test implies a sequence of decisions, we use a pseudo-real setting that aims at replicating the way these decisions might be taken in practice. Therefore, the choice to correct or not to correct for cross-unit dependence is based on a test for the presence of cross-unit correlation (Pesaran, 2004). When the test rejects the absence of correlation among the cross-section units, the panel test is performed by using the modified weighted inverse normal combination (14), otherwise standard inverse normal combination (12) is utilized. When the modified version (14) is used, consistently with Hartung (1999) and Demetrescu *et al.* (2006), in our experiments we use $\lambda_i = 1\ \forall i$ and $\kappa = 0.2$. Furthermore, the selection of the lags structure for the lagged differences of both the dependent variable and the covariate in the pCADF test equations (6)-(8) is based on the BIC separately for each equation. The choice of the variable to be used as the stationary covariate in testing the unit root for the i-th series in the panel is determined using three different criteria. First, the "true" covariate is used; second, we consider as the stationary covariate the average of the differenced series Δy_{jt} ($\forall j \neq i$) related to the other units in the panel, as in Chang and Song (2009); third, we use the first difference of the first principal component of the series. A word of caution is in order here. It could be argued that selecting the stationary covariate using the average of the other Δy_{jt} or the differences of the first principal component of the series may overlook the problem that the derived covariate might be non-invertible. In fact, Hansen (1995) showed that over-differencing the covariates raises theoretical problems and can have some adverse effects on the size and power of the test. However, for this to be the case it would be necessary that all the series are $I(0)$. In this instance the test would have high power anyway.

The panel-ADF test is carried out in the version proposed by Demetrescu *et al.* (2006), that exploits the correction for cross-section dependence introduced by Hartung (1999).

The number of lags is selected also in this case by using the BIC and Hartung's correction is applied after pre-testing for cross-dependence as for the p-CADF test. If no cross-dependence is detected, then the test is applied as in Choi (2001).

The test developed in Moon and Perron (2004) deals directly with cross-unit correlation by using an approximate linear factor model. We set the maximum number of factors to 4 and select the actual number of factors to be used in the test by the BIC_3 criterion, as suggested in Moon and Perron (2004, p. 94).

Finally, as far as the test proposed by Chang and Song (2009) is concerned, the procedure that we use in our Monte Carlo simulations amounts to the selection of the lag order of the lagged differences and of the covariate for each cross-section unit using the BIC and the selection of the appropriate covariate to be used by selecting the one that has the highest correlations with the error process (see, on this, Chang and Song, 2009, footnote 9).

4.3 Simulation results

The simulation results are presented using the graphical approach proposed in Davidson and MacKinnon (1993, 1998). Let's denote by $\hat{F}(x_i)$ the estimated empirical distribution of the p-values at any point $x_i \in (0,1)$. Under the null, the p-values are uniformly distributed, so that it should be true that $\hat{F}(x_i) \approx x_i$. A useful way to investigate the size properties of a test is therefore to plot $\hat{F}(x_i) - x_i$ against x_i. This is what Davidson and MacKinnon call a *p-value discrepancy plot*. The statistical significance of the discrepancies $\hat{F}(x_i) - x_i$ can be approximately assessed by using the Kolmogorov-Smirnov distribution.[16] Using the p-value discrepancy plots it is possible to investigate the size properties of the tests not only in correspondence of a couple of selected points, but along all the p-values distribution. However, given that we are mostly interested in the left tail of the distribution, we confine our attention to the nominal size up to 30%.

In order to analyse the power of the tests, we plot the power against the actual size. Although we must be careful in the presence of heavy size distortions or in the presence of biased tests, nevertheless plotting the power against the actual size makes it very easy to compare the power of different tests even in the presence of moderate size distortions. Davidson and MacKinnon (1998) call these plots *size-power curves*.[17] By plotting the power on the vertical axis and the actual size on the horizontal one, we have a graphical representation of the power for any desired size of the test. A 45° line is also plotted that is equivalent to the size-power curve of a hypothetical test whose power is always equal to the size. Of course, for a test to be of any value, we should expect its size-power curve to lie always well above the 45° line. Ideally, the curve should be very close to 1 for any actual size. Depending on the size and power properties of each test, the corresponding size-power curves based on the actual size may cross each other (see e.g.

[16] Other statistics and distributions could in principle be used that give more weight to the left tail of the p-values distribution (see, e.g., Delicado and Placencia, 2001) but the Kolmogorov-Smirnov statistic proposed by Davidson and MacKinnon (1998) fits perfectly in the graphical framework adopted here.

[17] Size-power curves were first introduced by Wilk and Gnanadesikan (1968) as a specialized example of the use of P-P plots .

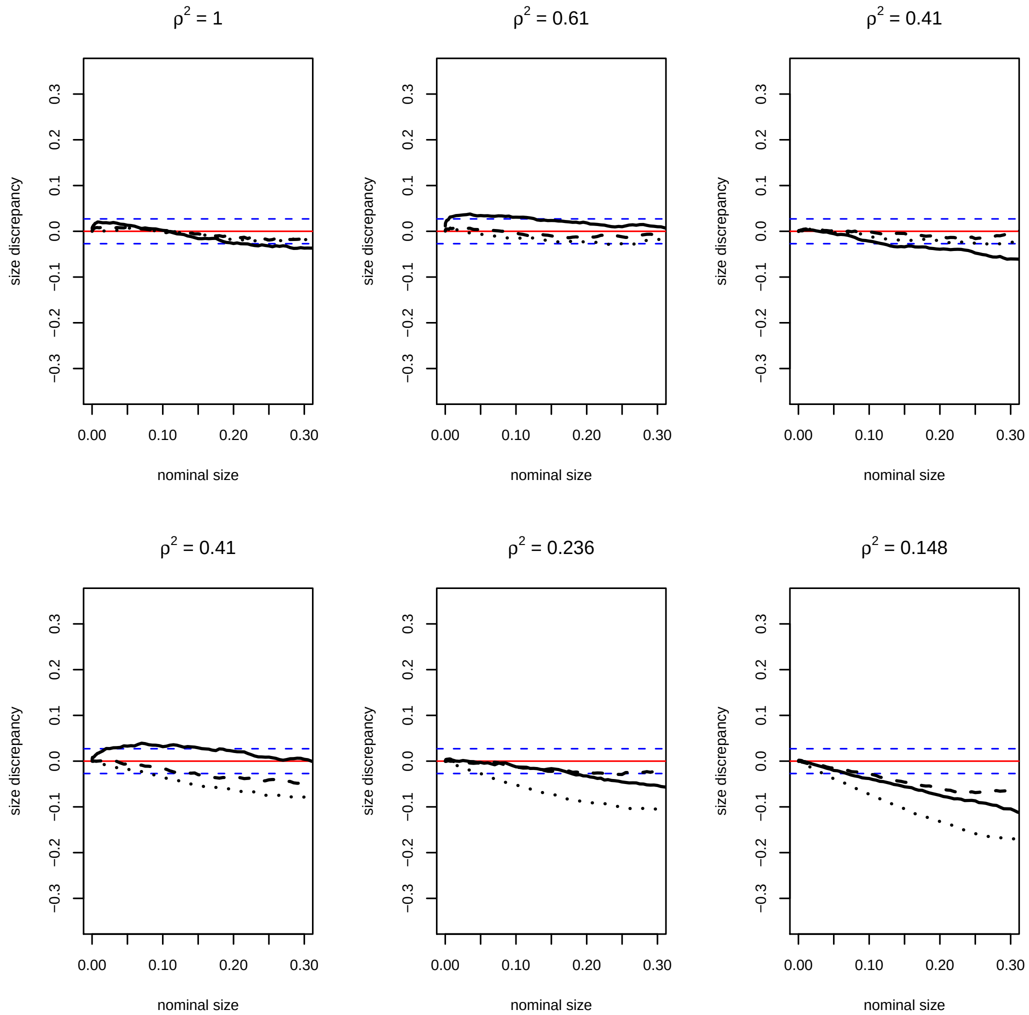

Figure 6 – Size discrepancy plots of the p-CADF test. The first row refers to experiments 1 to 3, the second to experiments 4 to 6. DGP with no drift, model with no trend. $T = 100$, $N = 10$. Solid line, true covariate; dashed, average Δy_{jt} $(j \neq i)$ as the stationary covariate; dotted, first difference of the first principal component as the stationary covariate. The horizontal dashed lines represent 5% Kolmogorov-Smirnov critical values.

Wilk and Gnanadesikan, 1968; Davidson and MacKinnon, 1998, for examples of crossing size-power curves).

While it is customary to report simulation results only with respect to the percentage of rejections obtained in correspondence with conventional significance levels (5% and 10%, say), we offer for the first time a detailed analysis of the whole empirical distribution function of the p-values of different panel unit root tests under cross-dependence. We can do this because we can compute the p-values of each test and not just the critical values. Of course, this greater detail comes at the cost of some extra computational burden.

All the figures presented in this Section are produced using the same scale in order to ease comparison among the tests and across the experiments.

We start the analysis by considering experiments 1–6 of Table 2 with $\boldsymbol{\alpha} = \mathbf{0}$ in the DGP and no trend in the model. The size discrepancies of the tests are reported in Fig-

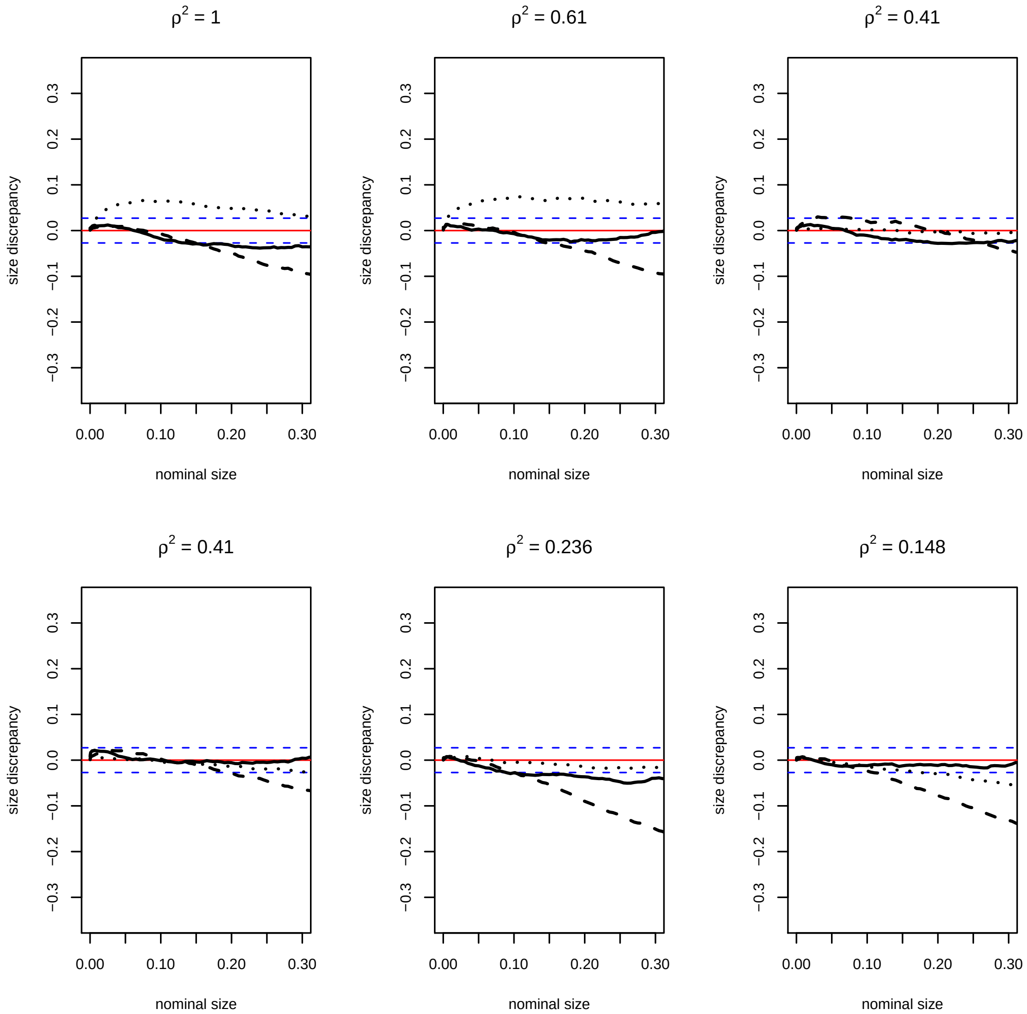

Figure 7 – Size discrepancy plots. The first row refers to experiments 1 to 3, the second to experiments 4 to 6. DGP with no drift, model with no trend. $T = 100$, $N = 10$. Solid line, Demetrescu *et al.* (2006); dashed, Chang and Song (2009); dotted, Moon and Perron (2004). The horizontal dashed lines represent 5% Kolmogorov-Smirnov critical values.

ures 6 and 7. The test proposed by Demetrescu *et al.* (2006) has the best overall size properties across experiments. However, the pCADF test performs quite well, with no large size discrepancies in correspondence of the usual size levels. It tends to under-reject especially in experiment 6, when the first principal component is used to derive the stationary covariate. On the contrary, the test advocated by Moon and Perron (2004) tends to over-reject in experiments 1 and 2, where the factor structure is weaker. In all the other experiments it performs remarkably well in terms of size. Finally, the test developed by Chang and Song (2009) does not display significant discrepancies in correspondence with the usual size levels, but shows a general tendency towards under-rejecting, especially for experiments 5 and 6.

The size-power curves for the same experiments are reported in Figures 8 and 9. The power of the pCADF test increases significantly with decreasing values of ρ^2, as expected. Indeed, when $\rho^2 < 0.5$, the pCADF correctly rejects the null more often than the other tests when the true covariate is used and, for somewhat smaller values of ρ^2 also when the estimated covariates are used as well. The covariate-augmented test proposed by Chang and Song (2009) shows a rather stable rejection rate across experiments and performs better that the p-CADF only for relatively high values of ρ^2. However, it should be reminded that the p-CADF is equivalent to the panel ADF test when $\rho^2 = 1$. When $\rho^2 < 1$, the power gain obtained by using stationary covariates is substantial. The power of Moon and Perron's test is rather disappointing, being virtually identical to the size for most experiments.

Let's now turn to the analysis of the performance of the tests with trend (p-CADF and Demetrescu *et al.*'s) or detrended (Chang and Song's and Moon and Perron's) over the same DGP as above. The size discrepancies are plotted in Figures 10 and 11.

Demetrescu *et al.*'s ADF-based test ranks first, as in the previous case. The p-CADF test has approximately correct size in the usual size ranges. It is again slightly conservative in experiment 6, especially when the difference of the first principal component is used as the stationary covariate. Under-rejection of Chang and Song's test is now more evident. Indeed this test tends to be conservative across all the experiments. On the other hand, Moon and Perron's test tends to over-reject substantially.

The size-power curves are plotted in Figures 12 and 13. The presence of the trend in the model tends to reduce the power of all the tests. As far as the ADF test is concerned, this is a known result. The p-CADF test behaves quite well also in this case, even if the rejections do not increase monotonically when ρ^2 decreases. In fact, the same kind of behaviour is mirrored, on a different scale, by Demetrescu *et al.*'s test. However, the power gain deriving from using the covariate is again substantial, especially when the correct covariate or a good proxy for it is used, but the power of the p-CADF declines together with the power of the pure ADF-based test. Despite being conservative, Chang and Song's test has good power and the rejections remain fairly stable across experiments, as in the no-trend case. Nevertheless, the pCADF test still compares well with Chang and Song's, above all when the correct covariate is considered. At any rate, the power of the pCADF test is significantly higher than Demetrescu *et al.*'s and Moon and

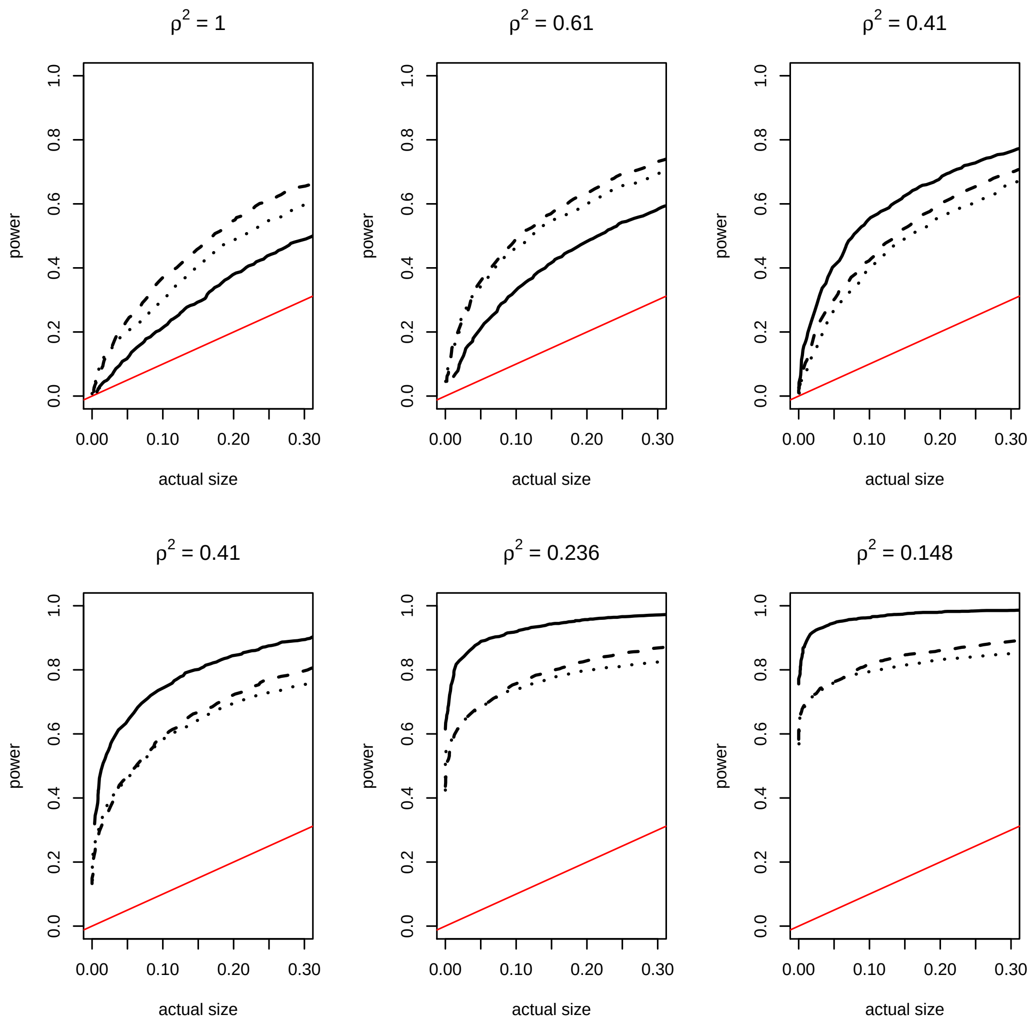

Figure 8 – Size-power plots of the *p*-CADF test. The first row refers to experiments 1 to 3, the second to experiments 4 to 6. DGP with no drift, model with no trend. $T = 100$, $N = 10$, 2 series are stationary. Solid line, true covariate; dashed, average Δy_{jt} ($j \neq i$) as the stationary covariate; dotted, first difference of the first principal component as the stationary covariate.

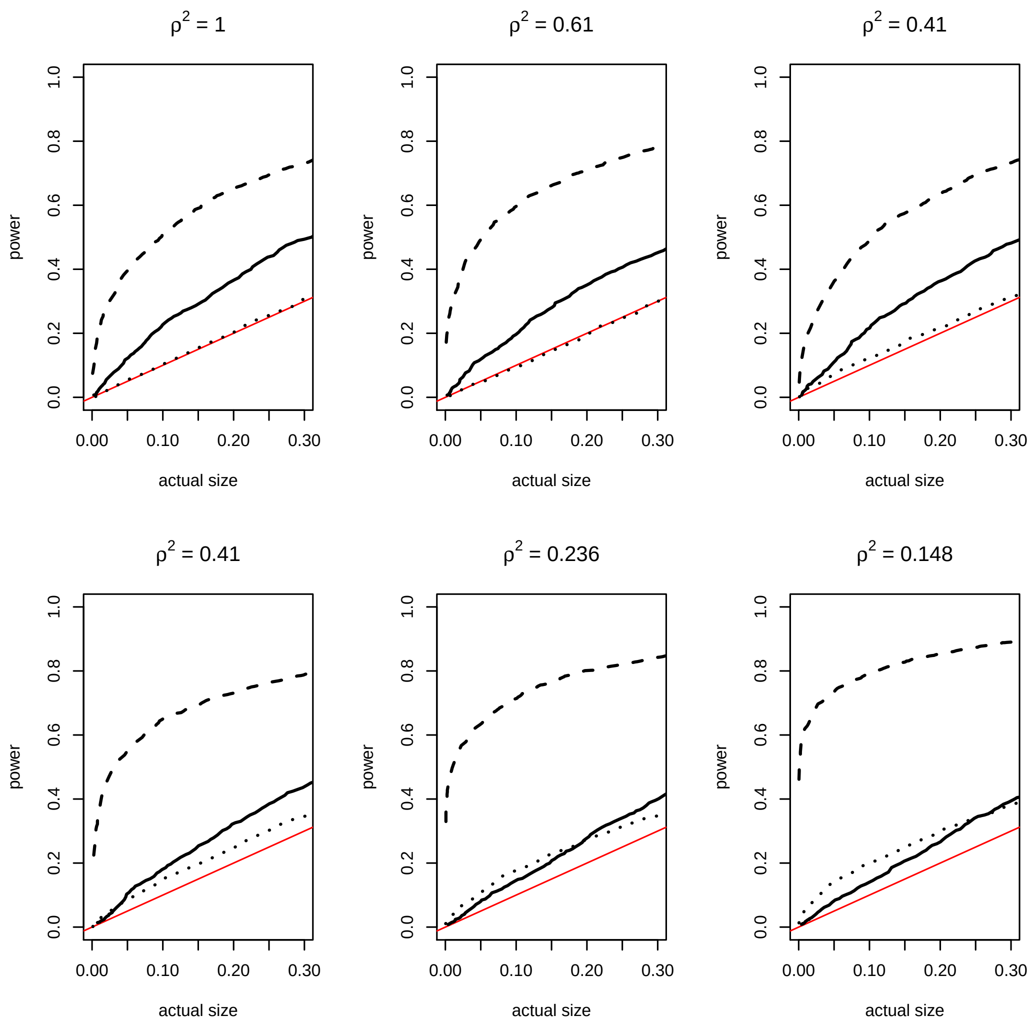

Figure 9 – Size-power plots. The first row refers to experiments 1 to 3, the second to experiments 4 to 6. DGP with no drift, model with no trend. $T = 100$, $N = 10$, 2 series are stationary. Solid line, Demetrescu *et al.* (2006); dashed, Chang and Song (2009); dotted, Moon and Perron (2004).

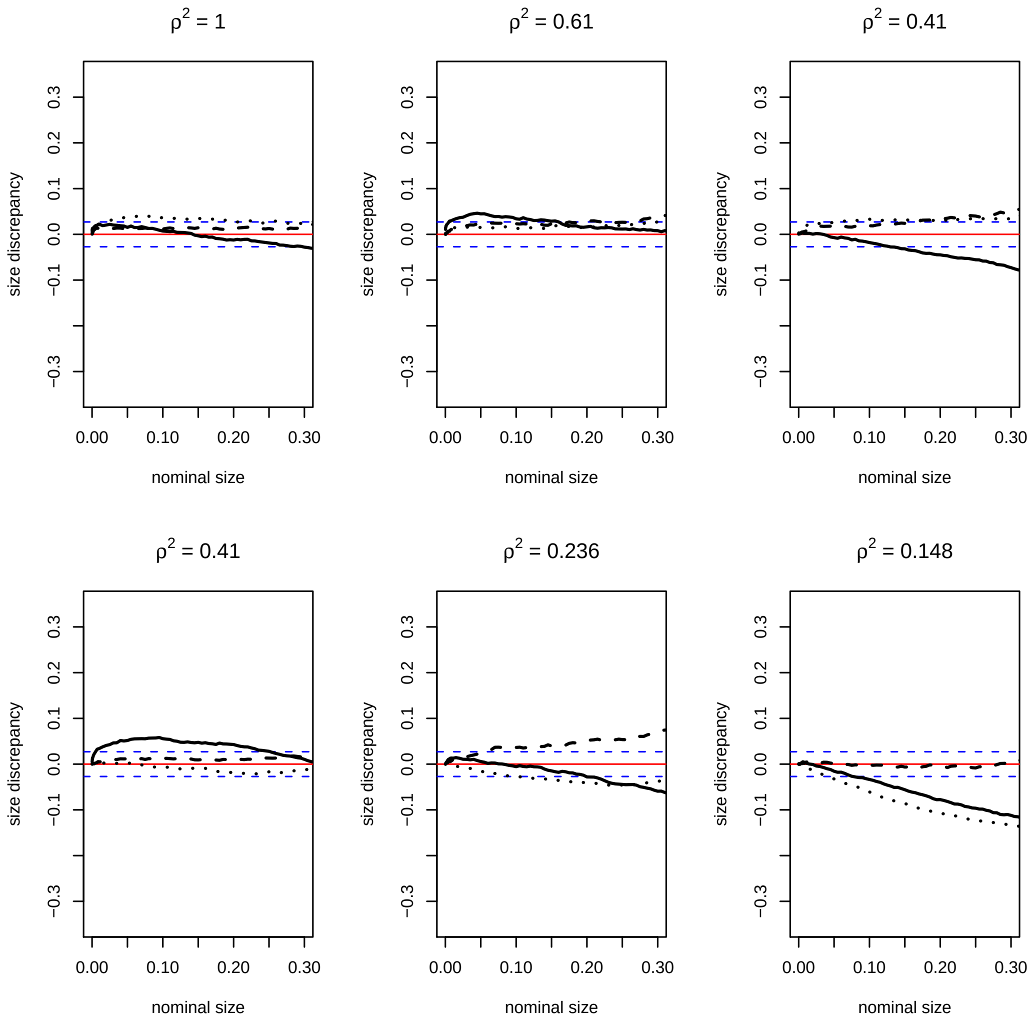

Figure 10 – Size discrepancy plots of the p-CADF test. The first row refers to experiments 1 to 3, the second to experiments 4 to 6. DGP with no drift, model with trend. $T = 100$, $N = 10$. Solid line, true covariate; dashed, average Δy_{jt} ($j \neq i$) as the stationary covariate; dotted, first difference of the first principal component as the stationary covariate. The horizontal dashed lines represent 5% Kolmogorov-Smirnov critical values.

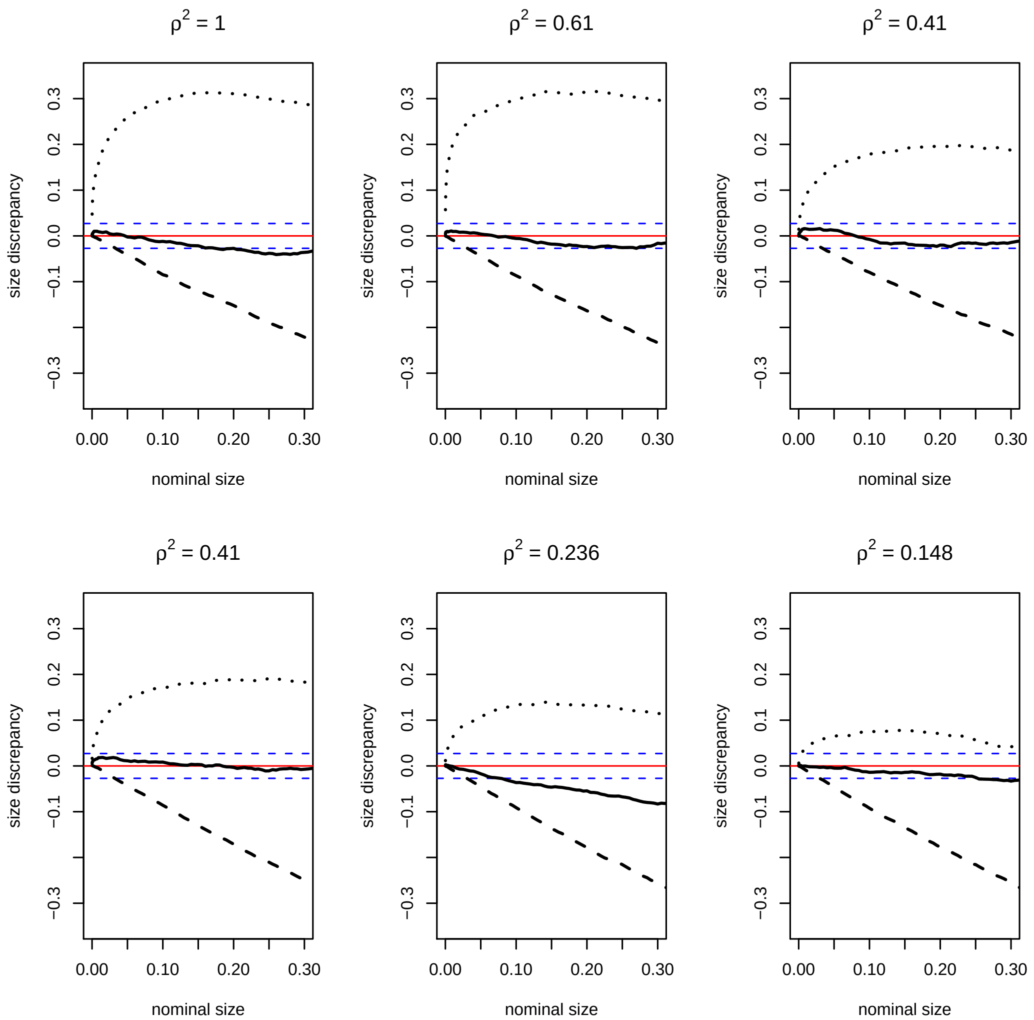

Figure 11 – Size discrepancy plots. The first row refers to experiments 1 to 3, the second to experiments 4 to 6. DGP with no drift, model with trend. $T = 100$, $N = 10$. Solid line, Demetrescu *et al.* (2006); dashed, Chang and Song (2009); dotted, Moon and Perron (2004). The horizontal dashed lines represent 5% Kolmogorov-Smirnov critical values.

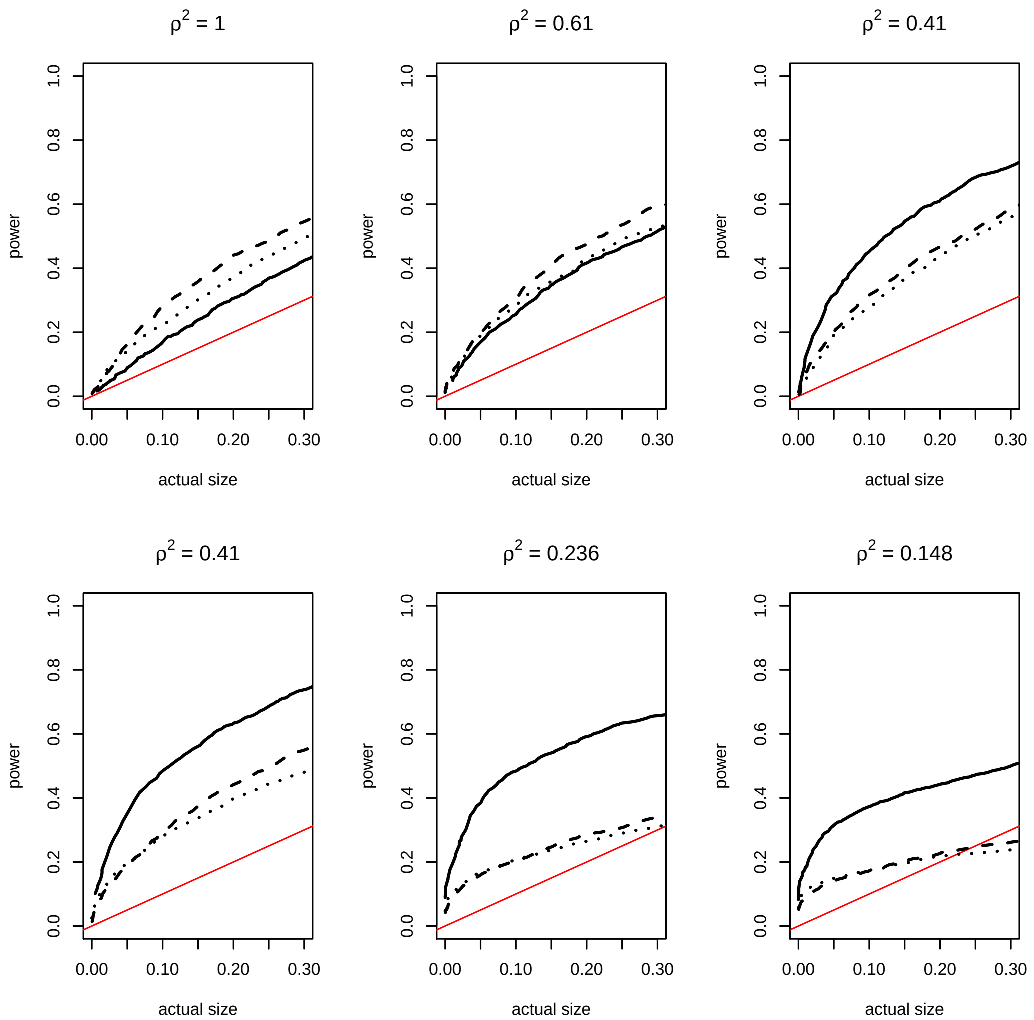

Figure 12 – Size-power plots of the p-CADF test. The first row refers to experiments 1 to 3, the second to experiments 4 to 6. DGP with no drift, model with trend. $T = 100$, $N = 10$, 2 series are stationary. Solid line, true covariate; dashed, average Δy_{jt} $(j \neq i)$ as the stationary covariate; dotted, first difference of the first principal component as the stationary covariate.

Perron's. In fact, the latter test has virtually no power at all.

We now extend our analysis also to cover the case where the DGP includes a drift term $\boldsymbol{\alpha} \neq \mathbf{0}$. In particular, in our simulations we consider $\alpha_i \sim \mathrm{U}_{(0.7,0.9)}$ (with $i = 1, \ldots, N$). Given the presence of a drift, in this case we only consider the tests based on models including the deterministic trend (or the detrended versions of the tests) and avoid carrying out the simulations for the no-trend (or the demeanded) cases.

When we allow for a non-zero drift in the DGP, the behaviour of the p-CADF test and of Demetrescu *et al.*'s test remains substantially unchanged in terms of size. The presence of the drift adversely affects the size of Chang and Song's test that becomes very conservative. On the contrary, Moon and Perron's test rejects much too often (see Figures 14 and 15).

However, the most dramatic changes happen when the power is considered. In order to show the difference with respect to the previous cases, in Figures 16 and 17 we

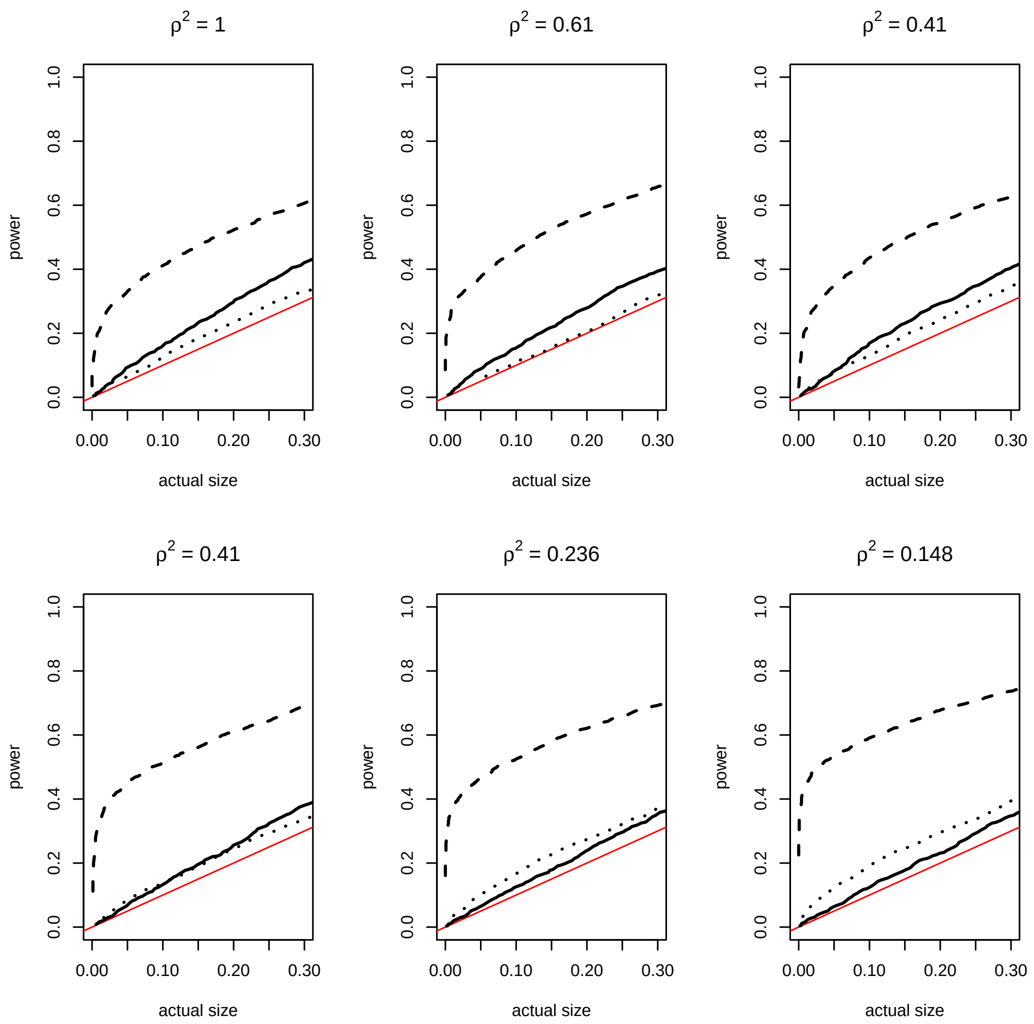

Figure 13 – Size-power plots. The first row refers to experiments 1 to 3, the second to experiments 4 to 6. DGP with no drift, model with trend. $T = 100$, $N = 10$, 2 series are stationary. Solid line, Demetrescu *et al.* (2006); dashed, Chang and Song (2009); dotted, Moon and Perron (2004).

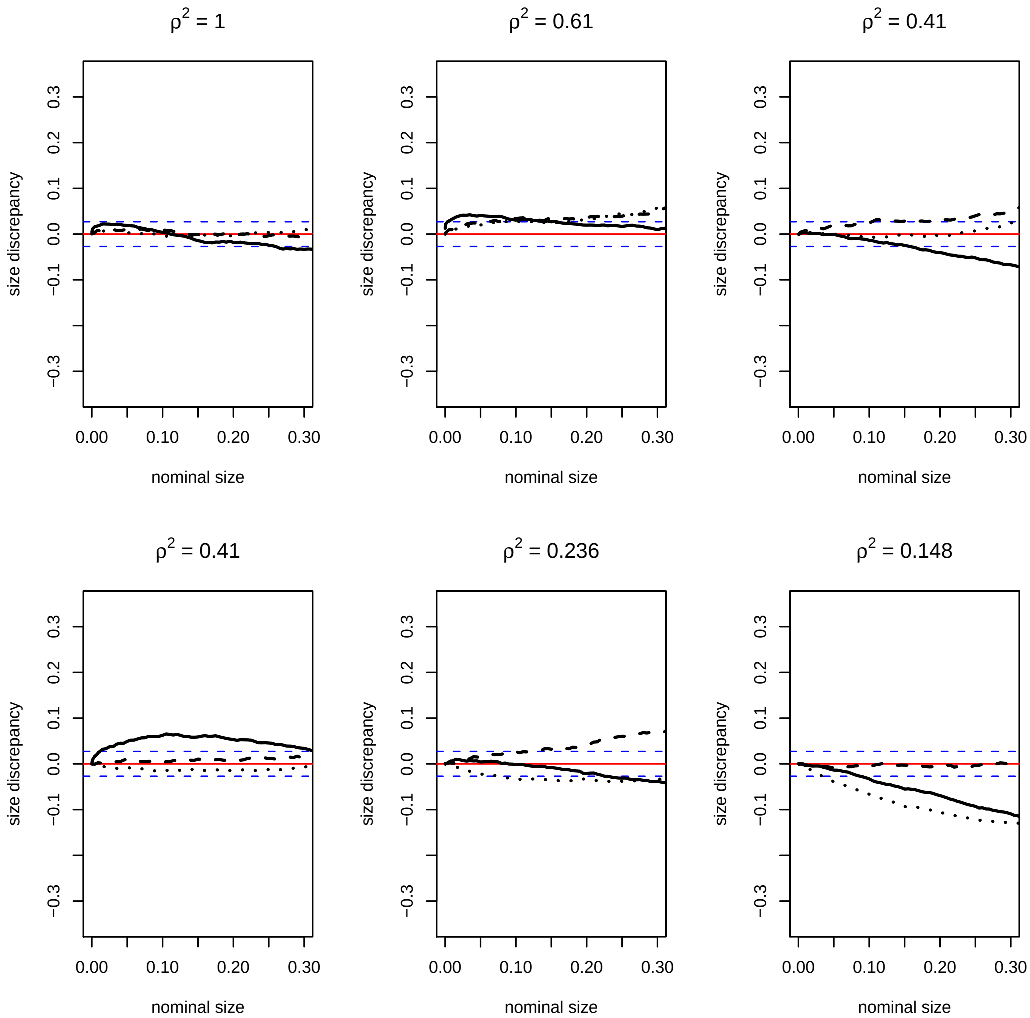

Figure 14 – Size discrepancy plots of the p-CADF test. The first row refers to experiments 1 to 3, the second to experiments 4 to 6. DGP with non-zero drift, model with trend. $T = 100$, $N = 10$. Solid line, true covariate; dashed, average Δy_{jt} ($j \neq i$) as the stationary covariate; dotted, first difference of the first principal component as the stationary covariate. The horizontal dashed lines represent 5% Kolmogorov-Smirnov critical values.

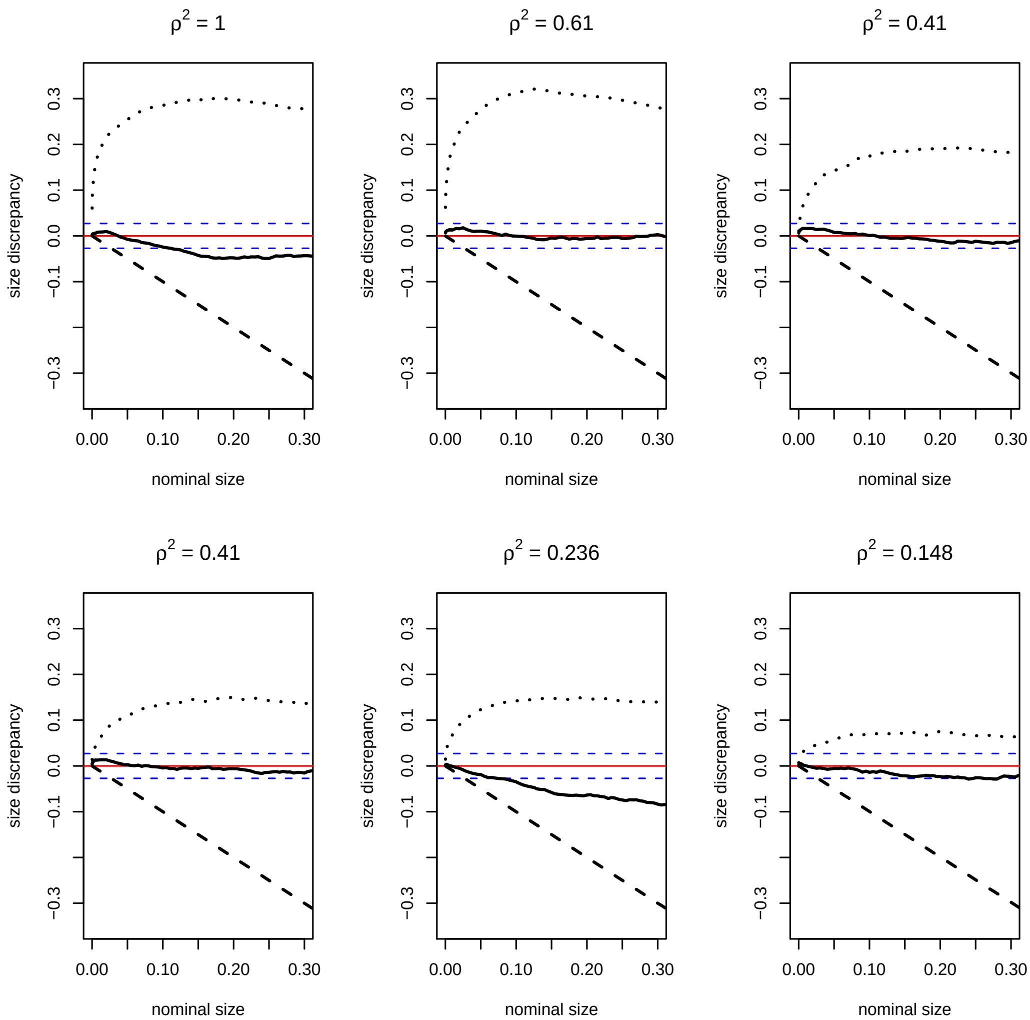

Figure 15 – Size discrepancy plots. The first row refers to experiments 1 to 3, the second to experiments 4 to 6. DGP with non-zero drift, model with trend. $T = 100$, $N = 10$. Solid line, Demetrescu *et al.* (2006); dashed, Chang and Song (2009); dotted, Moon and Perron (2004). The horizontal dashed lines represent 5% Kolmogorov-Smirnov critical values.

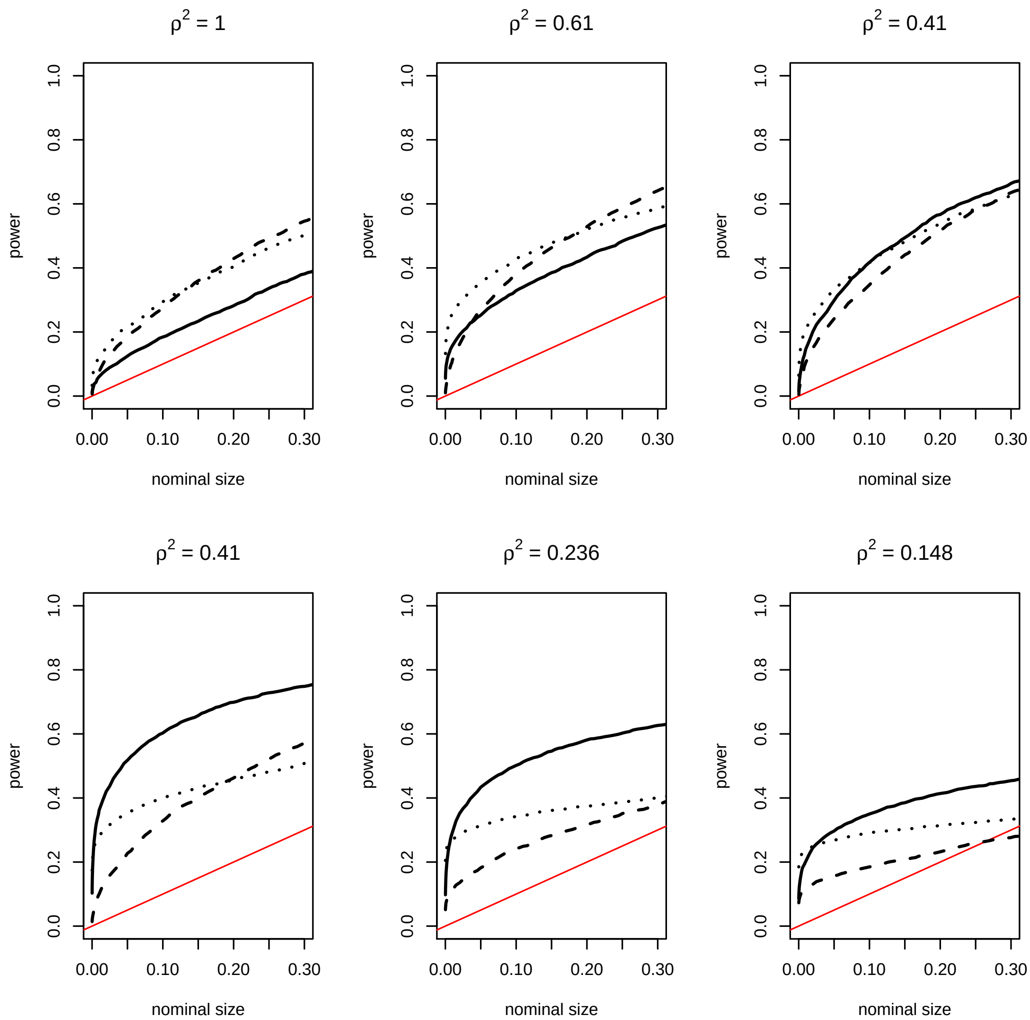

Figure 16 – Power of the p-CADF test against nominal size. The first row refers to experiments 1 to 3, the second to experiments 4 to 6. DGP with non-zero drift, model with trend. $T = 100$, $N = 10$, 2 series are stationary. Solid line, true covariate; dashed, average Δy_{jt} $(j \neq i)$ as the stationary covariate; dotted, first difference of the first principal component as the stationary covariate.

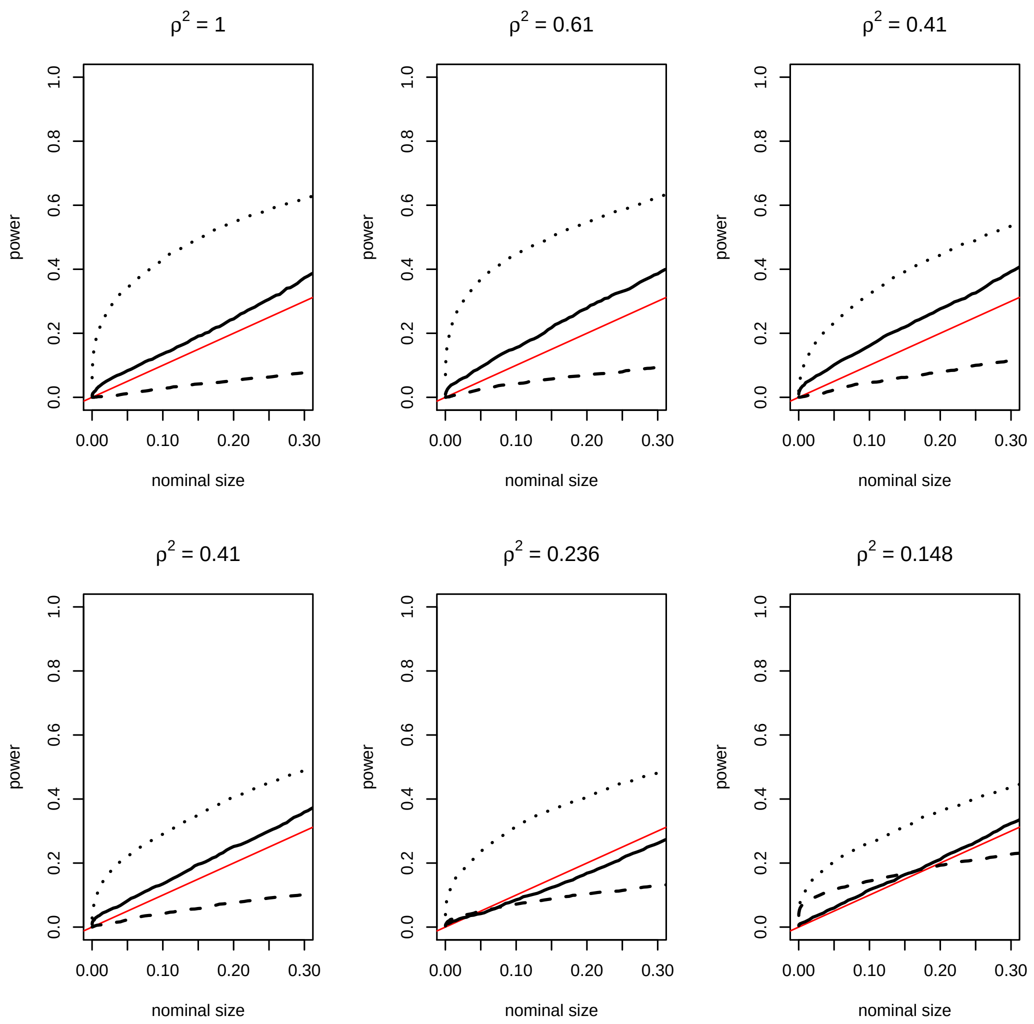

Figure 17 – Power against nominal size. The first row refers to experiments 1 to 3, the second to experiments 4 to 6. DGP with non-zero drift, model with trend. $T = 100$, $N = 10$, 2 series are stationary. Solid line, Demetrescu *et al.* (2006); dashed, Chang and Song (2009); dotted, Moon and Perron (2004).

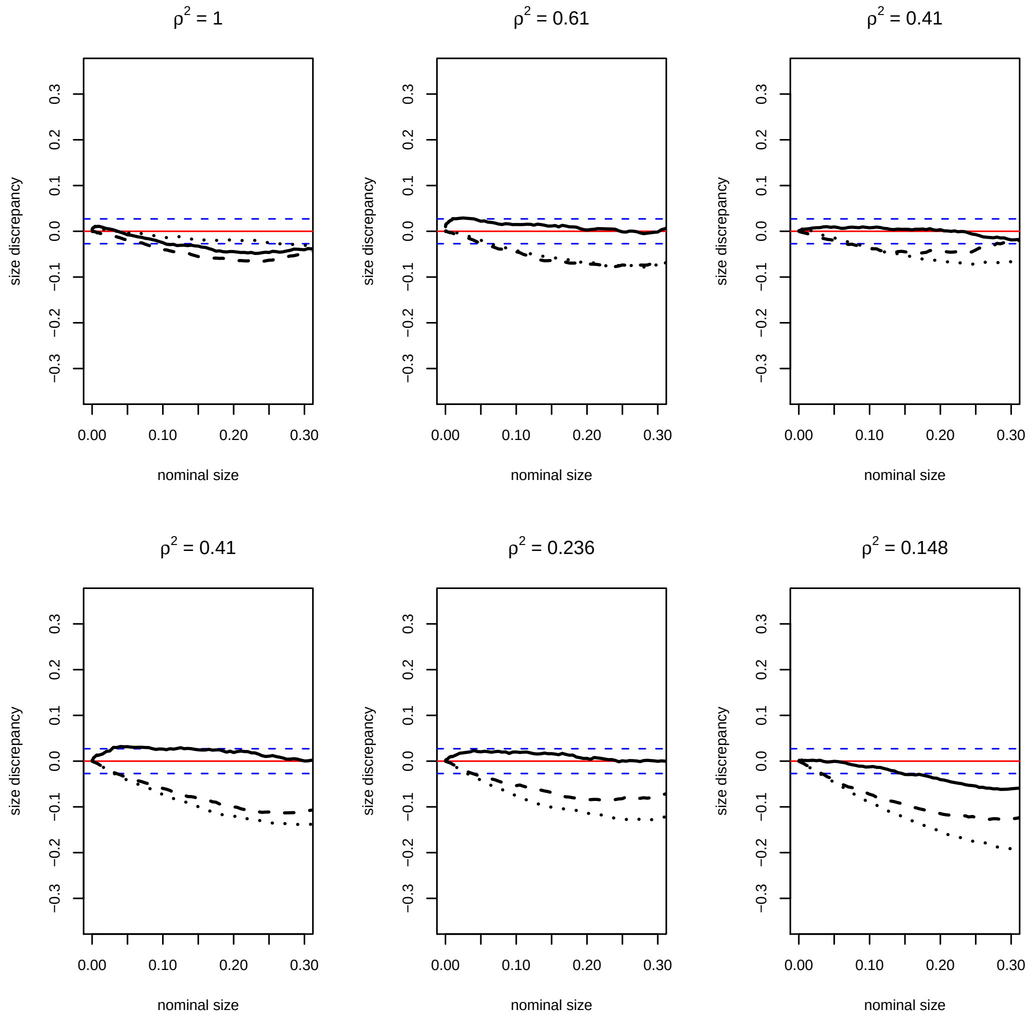

Figure 18 – Size discrepancy plots of the p-CADF test. The first row refers to experiments 1 to 3, the second to experiments 4 to 6. DGP with no drift, model with no trend. $T = 300$, $N = 20$. Solid line, true covariate; dashed, average Δy_{jt} $(j \neq i)$ as the stationary covariate; dotted, first difference of the first principal component as the stationary covariate. The horizontal dashed lines represent 5% Kolmogorov-Smirnov critical values.

plot power against nominal size. While the performance of the p-CADF test improves somewhat with respect to the trend case without drift,[18] Chang and Song's test becomes heavily biased with rejections well below the nominal size. On the other hand, the rejections of Demetrescu *et al.*'s and Moon and Perron's tests are very similar to the previous case without drift. Given the bias in Chang and Song's test, the comparison of the size-power curves would be misleading in this case.

In order to check the performance of the tests for larger values of T and N, we repeat the experiments of Table 2 with $T = 300$ and $N = 20$. Power is investigated again using only 2 (out of 20) stationary series. The results essentially confirm the tendencies already highlighted using $T = 100$ and $N = 10$; to save space we report only the results for the models with constant (or demeaned data). Similar conclusions carry over for the other

[18]This is true also when the size-power curves are considered.

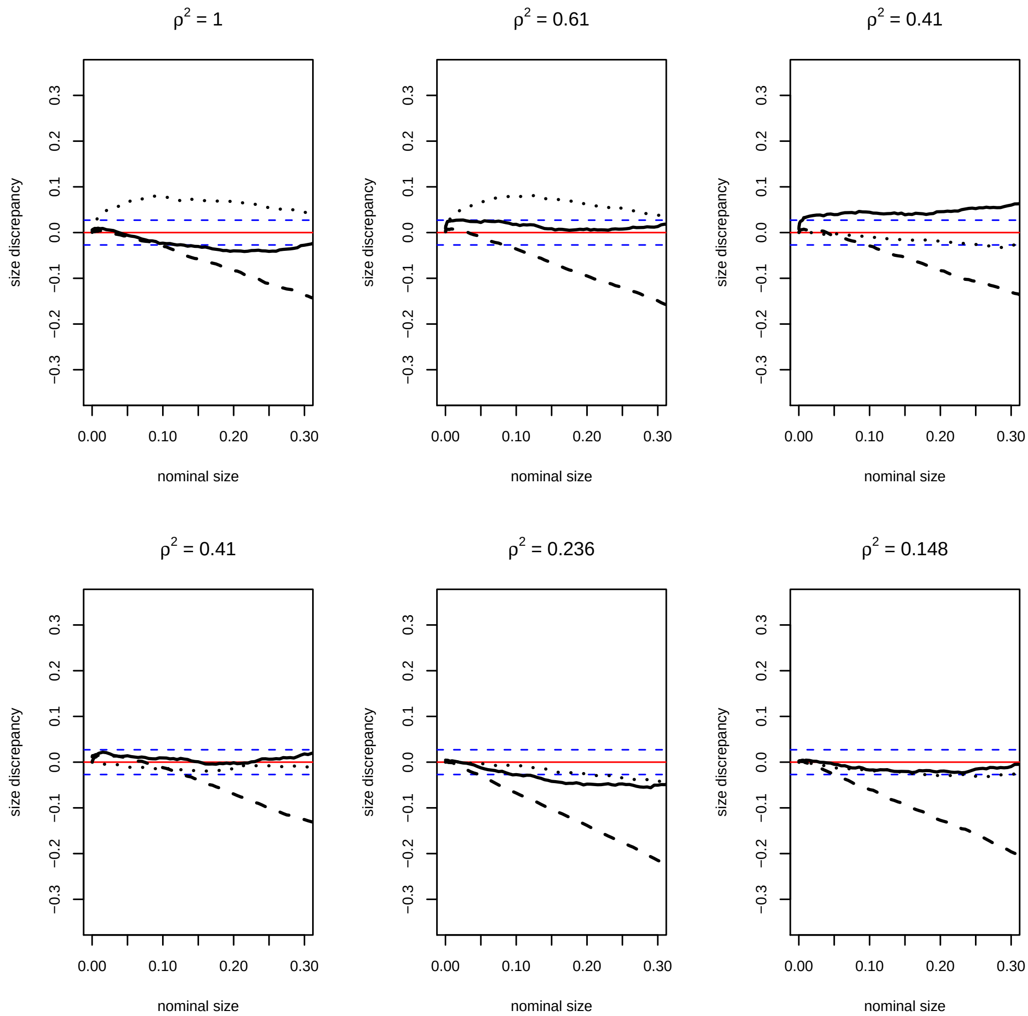

Figure 19 – Size discrepancy plots. The first row refers to experiments 1 to 3, the second to experiments 4 to 6. DGP with no drift, model with no trend. $T = 300$, $N = 20$. Solid line, Demetrescu *et al.* (2006); dashed, Chang and Song (2009); dotted, Moon and Perron (2004). The horizontal dashed lines represent 5% Kolmogorov-Smirnov critical values.

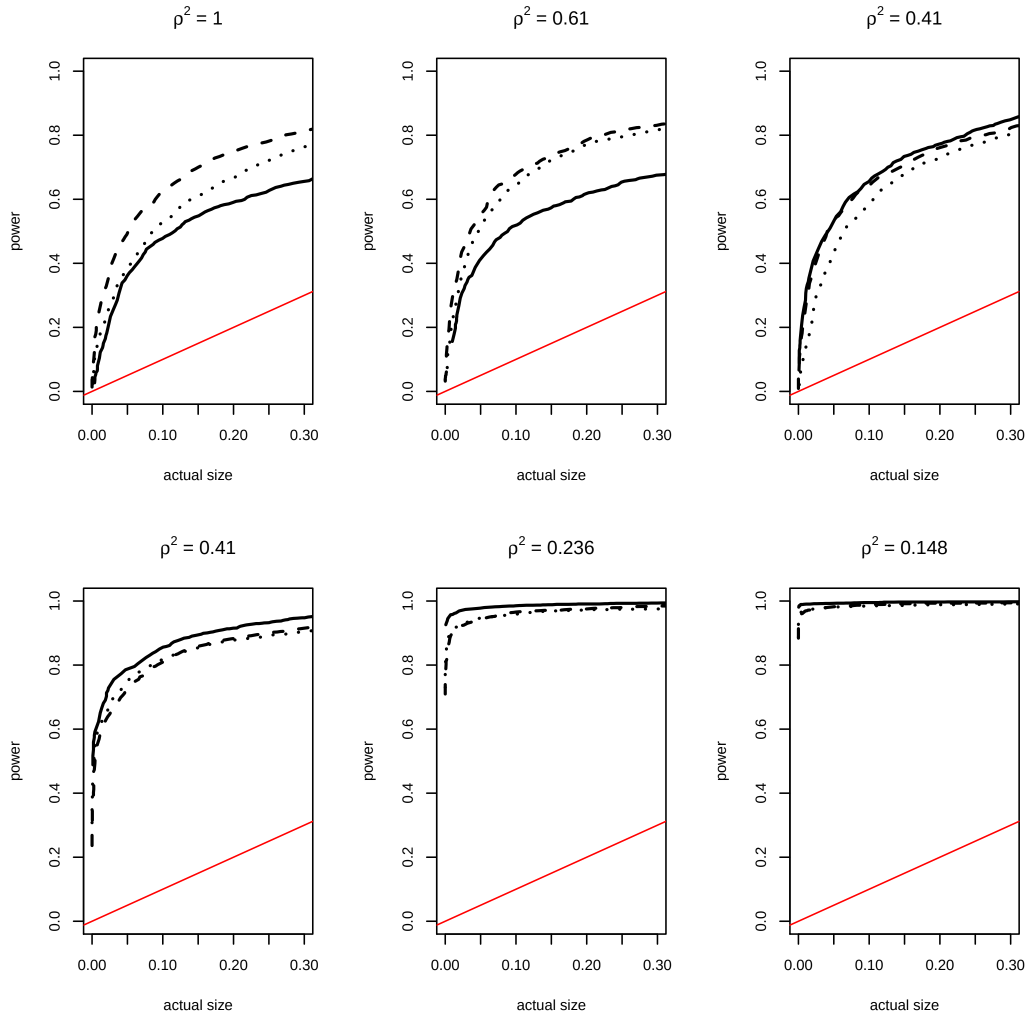

Figure 20 – Size-power plots of the p-CADF test. The first row refers to experiments 1 to 3, the second to experiments 4 to 6. DGP with no drift, model with trend. $T = 300$, $N = 20$, 2 series are stationary. Solid line, true covariate; dashed, average Δy_{jt} $(j \neq i)$ as the stationary covariate; dotted, first difference of the first principal component as the stationary covariate.

cases.[19]

When the true covariate is used, the pCADF test has approximately correct size for all experiments. On the other hand, it tends to be slightly conservative when estimated stationary covariates are used (see Figure 18). The ADF-based test proposed by Demetrescu *et al.* (2006) has again good size. The performance of Moon and Perron's test is also very similar to the case with $T = 100$ and $N = 10$ and tends to over-reject in the presence of a weak factor structure. Quite on the contrary, the tendency towards under-rejection of the test advocated by Chang and Song (2009) is now even more pronounced than in the $T = 100$, $N = 10$ case (Figure 19).

As far as power is concerned, the size-power curves plotted in Figure 20 show that power of the pCADF increases with decreasing values of ρ^2 and the test virtually always

[19] Detailed results are available upon request.

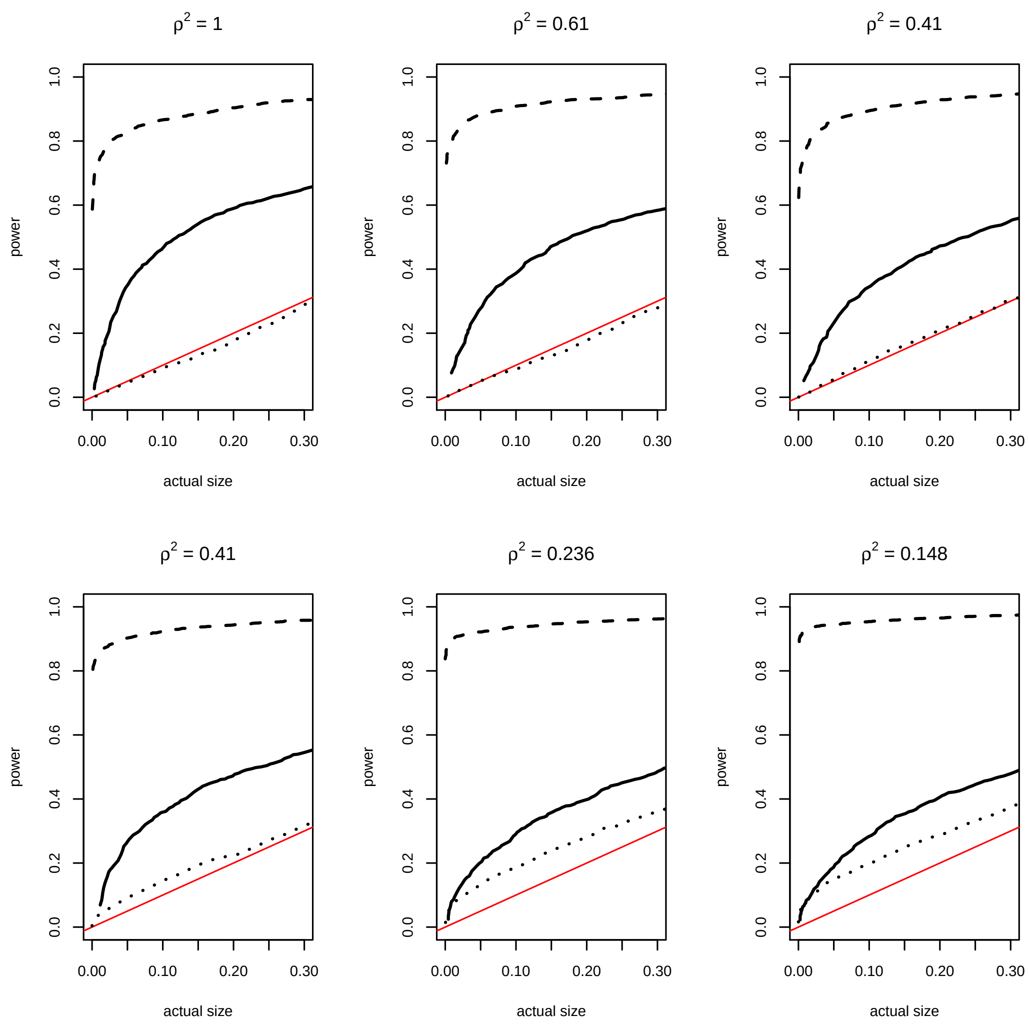

Figure 21 – Size-power plots. The first row refers to experiments 1 to 3, the second to experiments 4 to 6. DGP with no drift, model with trend. $T = 300$, $N = 20$, 2 series are stationary. Solid line, Demetrescu *et al.* (2006); dashed, Chang and Song (2009); dotted, Moon and Perron (2004).

reject when ρ^2 is small, despite having only 2 (out of 20) stationary series. The use of estimated stationary covariates in this case gives excellent results, very close to those that can be obtained using the true covariate. The comparison with the performance of the test proposed by Demetrescu *et al.* (2006) (see Figure 21) gives a measure of the gain that can be obtained by using the stationary covariates within the panel test. Moon and Perron's test has again virtually no power at all. On the contrary, the test advocated by Chang and Song (2009) has the best performance for high values of ρ^2, while its power is slightly worse than the pCADF's for small values of the nuisance parameter.

5 The size of panel vs individual tests

The fact that in some instances we are using covariates that are derived from the same series for which we want to test the presence of a unit root may induce in the reader the erroneous opinion that a better solution might be that of simply considering the unit root tests on the individual time series, instead of the panel test. At first sight this procedure uses the same data and gives more information. However, remember that our null of interest is H_0 : "all of the series are $I(1)$", against the alternative H_1 : "at least one series is $I(0)$".

We have already highlighted in the Introduction that a commonly neglected reason to use the panel approach is that it avoids the complications arising from multiple testing. In this Section we show that using individual tests to investigate our null of interest does indeed imply strong size distortions and incorrect inference. Surprisingly, this aspect is usually ignored in the empirical literature. In fact, when individual tests based on the single time series are used to investigate the null hypothesis, inference is based on a sequence of dependent tests. It is fairly well known that such a procedure is likely to produce severe over-rejections (see e.g. Shaffer, 1995).

In order to give a flavour of the size distortions implied by the individual testing approach, it is sufficient to look at Figure 22, where the size discrepancies of the individual-based ADF and CADF tests are plotted using the same simulations utilized to produce Figures 6 and 7. Contrary to what happens with the panel tests where size distortions are fairly small, using the individual tests gives rise to terrific size distortions so that we can falsely reject the null 90% of the times for a 10% nominal size. A possible reply to this criticism is that a Bonferroni-like method could be used to adjust the p-values. However, these methods are in general designed to work with *independent* tests, while here we are mainly interested on *dependent* ones. For this reason we should expect fairly large size distortions also from the application of Bonferroni-like procedures in our setting. This is in fact confirmed by our simulations (not reported here to save space). Furthermore, Maddala and Wu (1999) showed that application of the Bonferroni correction leads to tests with low power.[20]

Quite surprisingly, the empirical applications based on single time series approaches

[20] Other, more sophisticated, methods could in theory be developed to deal with dependent tests, but this is at the moment out of the scope of the present work.

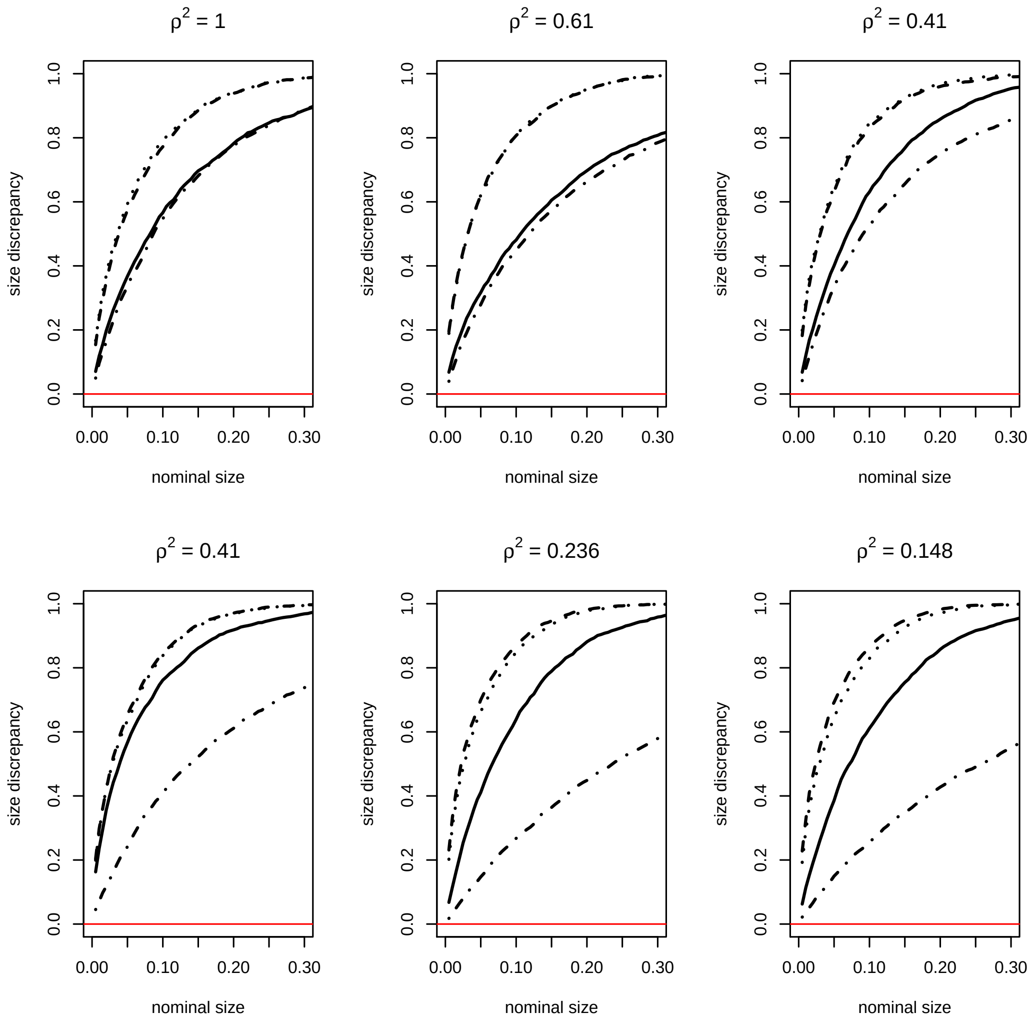

Figure 22 – Size discrepancy plots based on the individual tests. The first row refers to experiments 1–3, the second to experiments 4–6. DGP with no drift, model with no trend. $T = 100$, $N = 10$. Solid line, pCADF with true covariate; dashed, pCADF with average Δy_{jt} ($j \neq i$) as the stationary covariate; dotted, pCADF with first difference of the first principal component as the stationary covariate; dot-dash, Demetrescu *et al.* (2006). Same experiments used to produce Figures 6 and 7.

typically do not address the multiple testing issue at all so that they are likely to overstate substantially the importance of rejections.

6 Applications

For the sake of illustration, in this Section we offer two different applications developed using macro-panel data. We first consider the PPP hypothesis. This is a well-known example in many panel unit root papers. Then we consider the issue of the existence of a unit root in the industrial production indices. In this paper we are using these topics merely as illustrative examples of application of the pCADF test. Conclusive answers on the validity of the underlying economic theories would require more structured empirical analyses and are out of the scope of the present work.

In all applications we use exactly the same procedure adopted in the Monte Carlo analysis, with automatic model selection and correction for cross-dependence based upon the outcome of the test proposed in Pesaran (2004). Furthermore, we apply all the tests considered in the Monte Carlo to the actual data. In addition, in carrying out the pCADF tests we use stationary covariates chosen on theoretical grounds.

6.1 Testing the PPP hypothesis

It is well known that a necessary condition for the PPP to hold is that the real exchange rate must be *mean-reverting* (for a recent survey see Taylor and Taylor, 2004). This of course excludes the possibility that the real exchange rate can have a trending behaviour or a unit root. For this reason, a number of influential papers on panel unit root testing, including Choi (2001) and Chang and Song (2009), consider the same empirical application. Other papers employ instead covariate-augmented tests in the time series framework. In particular, Amara and Papell (2006) use the tests developed in both Hansen (1995) and Elliott and Jansson (2003), Elliott and Pesavento (2006) employ Elliott and Jansson's feasible point optimal test and Lee and Tsong (2009) utilize Hansen's CADF test with a stationary factor-based covariate selection.

It should be noticed that we are deliberately *not* dealing with the (alternative) hypothesis that the PPP is valid in general. Here we are interested in testing the null that the PPP is *not* valid in general. This implies that, if the null is rejected, this means only that the data are consistent with the PPP hypothesis in *at least* one case (country). On the other hand, if the null is not rejected, one could seriously wonder about the validity of the PPP hypothesis.

For greater comparability, we use quarterly data from Chang and Song (2009) covering the period 1973q1–1998q4.[21] The same countries over the same period are used also in other papers (see e.g. Amara and Papell, 2006). Given that under the PPP hypothesis the

[21] We warmly thank Yoosoon Chang and Wonho Song for having provided their data. The original sources are the International Monetary Fund's International Financial Statistics and cover 20 countries (Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom).

Test	test statistic	p-value
Demetrescu *et al.* (2006)	-0.383	0.351
Moon and Perron (2004)	-1.134	0.128
Chang and Song (2009)	-0.634	0.998
pCADF (principal component)	-0.672	0.251
pCADF (nominal exchange rate)	-4.210	0.000

Table 3 – Panel tests of the PPP hypothesis ($T = 103$, $N = 20$).

real exchange rate should not exhibit trends of any kind, in developing our application of the pCADF test, we focus specifically on the test without deterministic trends. This model specification is consistent with, e.g., Choi (2001, p. 269) and Amara and Papell (2006, p. 32). Consistently with Elliott and Pesavento (2006, pp. 1412–1413), we apply the pCADF test also using the first differences of the nominal exchange rate as the stationary covariate. Since the covariate should not cointegrate with the dependent variable, in order to verify that the nominal exchange rate is not cointegrated with the variable of interest, we apply the group mean cointegration tests proposed in Westerlund (2007). The null hypothesis of these tests is no cointergation for all the panel units, while the alternative is that cointegration is present in at least a panel unit. The p-values of Westerlund's G_τ and G_α tests are equal to 0.325 and 0.757, respectively, supporting the validity of the nominal exchange rate as a potential covariate.

The empirical results are summarized in Table 3. Here we also replicate Chang and Song (2009), so our results are identical to theirs.

As far as the PPP hypothesis is concerned, the only test that rejects the null is the pCADF when the differenced nominal exchange rate is used as the stationary covariate, consistently with Elliott and Pesavento (2006) that reject the null for most countries when the nominal exchange rate is used as the stationary covariate. The empirical results suggest that the choice of the covariate can influence the outcome of the test. Indeed, this feature of the covariate-augmented tests is already well known and documented in other papers (see, e.g, Elliott and Pesavento, 2006; Lee and Tsong, 2009).

6.2 Unit roots in international industrial production indices

We offer a second application that deals with checking for the presence of a unit root in industrial production indices in 9 OECD countries. As before, the pCADF test is performed using the differenced first principal component as the stationary covariate. In addition, we use the differences of real GDPs as stationary covariates. In this context we may interpret real GDP as a measure of demand. The variable of interest is quarterly seasonally adjusted industrial production index (total industry, 2005 = 100) over the period 1983q1–2008q3. The covariate is quarterly seasonally adjusted real GDP (chained volume estimates). Both industrial production and GDP are log-transformed.[22] We consider here

[22]The considered countries are: Australia, Canada, France, Italy, Japan, Norway, Switzerland, United Kingdom, United States. Industrial production is from the OECD Main Economic Indicators data base. Real GDP is from the OECD Quarterly National Accounts data base. The data sample is truncated to 2008q3 to avoid potential complications arising from the deep fall of industrial activity after the international financial

Test	test statistic	p-value
Demetrescu *et al.* (2006)	1.309	0.905
Moon and Perron (2004)	1.234	0.891
Chang and Song (2009)	-0.089	0.996
*p*CADF (principal component)	2.128	0.983
*p*CADF (real GDP)	-2.839	0.002

Table 4 – Panel unit root tests for industrial production ($T = 103$, $N = 9$).

the versions of the tests that include a constant and a deterministic trend, or the equivalent versions of the tests on detrended series. In order to avoid using a covariate whose levels are cointegrated with the variable of interest, we test again the null of lack of cointegration among industrial production in all countries against the alternative that there is cointegration in at least one country using the group mean tests proposed in Westerlund (2007). Indeed, there are theoretical reasons that induce to anticipate the absence of cointegration between industrial production and real GDP and in fact the empirical results are strongly supportive of the null (the p-values of the G_α and G_τ tests are 1.000 and 0.993, respecyively).

Indeed, the results reported in Table 4 indicate that the *p*CADF test is again the only one to reject the null, while all the other tests are very far from rejecting.

7 Concluding remarks

A simple covariate augmented Dickey-Fuller (CADF) test for unbalanced heterogeneous panels is proposed. The test, that we label panel-CADF (*p*CADF), is a generalization of the CADF test proposed in Hansen (1995) and is developed along the lines suggested in Choi (2001). This allows us to be very general in the specification of the individual unit root tests. Thanks to the application of a correction originally due to Hartung (1999), the proposed test can be used in the presence of cross-dependent time series and, given that the asymptotics used in Choi (2001) does not require $N \to \infty$, it is especially well suited to deal with macroeconomic panels where the cross-section dimension is typically rather small.

Given that the *p*CADF test is based on a (possibly modified) inverse normal p-value combination, the p-values of the individual CADF tests have to be obtained. For this reason, a procedure to compute the asymptotic p-values of Hansen's CADF test is also proposed.

The size and power properties of the *p*CADF test are investigated using an extensive Monte Carlo with cross-dependent DGPs. Simulation results are reported using the graphical approach suggested in Davidson and MacKinnon (1998) that allows us to obtain detailed and readily interpretable results. The performance of the *p*CADF test is compared with that of the panel unit-roots tests proposed in Moon and Perron (2004), Demetrescu *et al.* (2006) and Chang and Song (2009). It is shown that the *p*CADF test in

crisis. The choice of the countries was somewhat forced by the availability of the data.

general does not suffer from important size distortions and can offer important power gains. In particular, it is shown that the power of the *p*CADF test is inversely related to the nuisance parameter ρ^2 and that, when ρ^2 is small, the test can be more powerful than the covariate augmented test proposed by Chang and Song (2009). In all the experiments analysed in the paper, the power of the *p*CADF test is significantly higher than the power of the tests advocated by Moon and Perron (2004) and Demetrescu *et al.* (2006). When a drift is present in the DGP, the *p*CADF test has the best performance in terms of power, among all the examined tests.

A section of the paper is also dedicated to the comparison of single equation Vs panel tests. It is shown that panel tests offer the advantage of avoiding the complications arising in the presence of multiple testing.

In order to show that the test is viable, we consider two empirical applications dealing with the PPP hypothesis and with industrial production indices, respectively. It is shown that the test is easy to implement and offers advantages over other popular panel unit root tests.

Acknowledgements

We would like to thank Stephan Popp for his contributions in the early stages of this work. Comments from audiences at the University of Oxford, at the Institute for Advanced Studies at Vienna, at the Third Italian Congress of Econometrics and Empirical Economics at Ancona are gratefully acknowledged. We are especially grateful to Roy Cerqueti, Juan Dolado, Luciano Gutierrez, Robert Kunst, Paresh Kumar Narayan, Paolo Paruolo, Martin Wagner, Joakim Westerlund, and Dmitri Zaykin for comments and discussion on the paper or specific parts of it. We are very pleased to thank Yoosoon Chang and Wonho Song for having provided their data and GAUSS code (Chang and Song, 2009). None of them is responsible for any remaining error.

References

Amara J, Papell DH (2006). "Testing for Purchasing Power Parity Using Stationary Covariates." *Applied Financial Economics*, **16**(1-2), 29–39.

Baltagi BH, Kao C (2000). "Nonstationary Panels, Cointegration in Panels and Dynamic Panels: A Survey." In BH Baltagi (ed.), *Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, volume 15 of *Advances in Econometrics*, chapter 1, pp. 7–51. Elsevier Science, New York.

Banerjee A (1999). "Panel Data Unit Roots and Cointegration: An Overview." *Oxford Bulletin of Economics and Statistics*, **61**(Supplement 1), 607–630.

Banerjee A, Marcellino M, Osbat C (2004). "Some Cautions on the Use of Panel Methods for Integrated Series of Macroeconomic Data." *Econometrics Journal*, **7**(2), 322–340.

Banerjee A, Marcellino M, Osbat C (2005). "Testing for PPP: Should We Use Panel Methods?" *Empirical Economics*, **30**(1), 77–91.

Campbell JY, Perron P (1991). "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots." In O Blanchard, S Fischer (eds.), *NBER Macroeconomics Annual*, volume 6, pp. 141–201. University of Chicago Press, Chicago.

Caporale GM, Pittis N (1999). "Unit Root Testing Using Covariates: Some Theory and Evidence." *Oxford Bulletin of Economics and Statistics*, **61**(4), 583–595.

Chang Y (2002). "Nonlinear IV Unit Root Tests in Panels with Cross-Sectional Dependency." *Journal of Econometrics*, **110**(2), 261–292.

Chang Y, Song W (2009). "Test for Unit Roots in Small Panels with Short-run and Long-run Cross-sectional Dependencies." *Review of Economic Studies*, **76**(3), 903–935.

Choi I (2001). "Unit Root Tests for Panel Data." *Journal of International Money and Finance*, **20**(2), 249–272.

Choi I (2006). "Nonstationary Panels." In TC Mills, K Patterson (eds.), *Econometric Theory*, volume 1 of *Palgrave Handbook of Econometrics*, chapter 13, pp. 511–539. Palgrave MacMillan, New York.

Davidson R, MacKinnon JG (1993). *Estimation and Inference in Econometrics*. Oxford University Press.

Davidson R, MacKinnon JG (1998). "Graphical Methods for Investigating The Size and Power of Hypothesis Tests." *The Manchester School*, **66**(1), 1–26.

DeJong DN, Nankervis JC, Savin NE, Whiteman CH (1992). "The Power Problems of Unit Root Tests in Time Series with Autoregressive Errors." *Journal of Econometrics*, **53**(1-3), 323–343.

Delicado P, Placencia I (2001). "Comparing Empirical Distributions of P-values from Simulations." *Communications in Statistics – Simulation and Computation*, **30**(2), 403–422.

Demetrescu M, Hassler U, Tarcolea AI (2006). "Combining Significance of Correlated Statistics with Application to Panel Data." *Oxford Bulletin of Economics and Statistics*, **68**(5), 647–663.

Elliott G, Jansson M (2003). "Testing for Unit Roots With Stationary Covariates." *Journal of Econometrics*, **115**(1), 75–89.

Elliott G, Pesavento E (2006). "On the Failure of Purchasing Power Parity for Bilateral Exchange Rates after 1973." *Journal of Money, Credit, and Banking*, **38**(6), 1405–1430.

Gengenbach C, Palm FC, Urbain JP (2006). "Cointegration Testing in Panels with Common Factors." *Oxford Bulletin of Economics and Statistics*, **68**(Supplement), 683–719.

Hanck C (2008). "The Error-in-Rejection Probability of Meta-Analytic Panel Tests." *Economics Letters*, **101**(1), 27–30.

Hansen BE (1995). "Rethinking the Univariate Approach to Unit Root Testing: Using Covariates to Increase Power." *Econometric Theory*, **11**(5), 1148–1171.

Hartung J (1999). "A Note on Combining Dependent Tests of Significance." *Biometrical Journal*, **41**(7), 849–855.

Hassler U, Tarcolea AI (2005). "Combining Multi-Country Evidence on Unit Roots: The Case of Long-Term Interest Rates." *Applied Economics Quarterly*, **51**(2), 181–190.

Hatanaka M (1996). *Time-Series-Based Econometrics: Unit Roots and Cointegration*. Advanced Texts in Econometrics. Oxford University Press, Oxford.

Hendry DF (1984). "Monte Carlo Experimentation in Econometrics." In Z Griliches, M Intriligator (eds.), *Handbook of Econometrics*, volume 2, chapter 16, pp. 937–976. Elsevier Science Publishers, Amsterdam.

Lee CF, Tsong CC (2009). "Covariate Selection for Testing Purchasing Power Parity." *Applied Economics*. Forthcoming.

Levin A, Lin CF (1993). "Unit Root Tests in Panel Data: New Results." *Working Paper 56*, Department of Economics, University of California at San Diego.

Levin A, Lin CF, Chu CSJ (2002). "Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties." *Journal of Econometrics*, **108**(1), 1–24.

Lupi C (2009). "Unit Root CADF Testing with R." *Journal of Statistical Software*, **32**(2), 1–19.

Lyhagen J (2008). "Why not Use Standard Panel Unit Root Test for Testing PPP." *Economics Bulletin*, **3**(26), 1–11.

MacKinnon JG (1994). "Approximate Asymptotic Distribution Functions for Unit-Root and Cointegration Tests." *Journal of Business and Economic Statistics*, **12**(2), 167–176.

MacKinnon JG (1996). "Numerical Distribution Functions for Unit Root and Cointegration Tests." *Journal of Applied Econometrics*, **11**(6), 601–618.

Maddala GS, Wu S (1999). "A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test." *Oxford Bulletin of Economics and Statistics*, **61**(Supplement 1), 631–652.

Moon HR, Perron B (2004). "Testing for a Unit Root in Panels with Dynamic Factors." *Journal of Econometrics*, **122**(1), 81–126.

O'Connell PGJ (1998). "The Overvaluation of Purchasing Power Parity." *Journal of International Economics*, **44**(1), 1–19.

Pesaran MH (2004). "General Diagnostic Tests for Cross Section Dependence in Panels." *Cambridge Working Papers in Economics 0435*, Department of Applied Economics, University of Cambridge.

Pesaran MH (2007). "A Simple Panel Unit Root Test in the Presence of Cross-Section Dependence." *Journal of Applied Econometrics*, **22**(2), 265–312.

Phillips PCB, Sul D (2003). "Dynamic Panel Estimation and Homogeneity Testing under Cross Section Dependence." *Econometrics Journal*, **6**(1), 217–259.

Phillips PCB, Xiao Z (1998). "A Primer on Unit Root Tesing." *Journal of Economic Surveys*, **12**(5), 423–470.

Quah D (1994). "Exploiting Cross-Section Variation for Unit Root Inference in Dynamic Data." *Economics Letters*, **44**(1-2), 9–19.

Rebonato R, Jäckel P (1999). "The Most General Methodology to Create a Valid Correlation Matrix for Risk Management and Option Pricing Purposes." *Journal of Risk*, **2**(2).

Shaffer JP (1995). "Multiple Hypothesis Testing." *Annual Review of Psichology*, **46**, 561–584.

Taylor AM, Taylor MP (2004). "The Purchasing Power Parity Debate." *Journal of Economic Perspectives*, **18**(4), 135–158.

Taylor MP, Sarno L (1998). "Behavior of Real Exchange Rates During the Post–Bretton Woods Period." *Journal of International Economics*, **46**(2), 281–312.

Wagner M (2008). "On PPP, Unit Roots and Panels." *Empirical Economics*, **35**(2), 229–249.

Westerlund J (2007). "Testing for Error Correction in Panel Data." *Oxford Bulletin of Economics and Statistics*, **69**(6), 709–748.

Westerlund J, Costantini M (2009). "Panel Cointegration and the Neutrality of Money." *Empirical Economics*, **36**(1), 1–26.

Wilk MB, Gnanadesikan R (1968). "Probability Plotting Methods for the Analysis of Data." *Biometrika*, **55**(1), 1–17.

Appendix A:
The algorithm to compute the p-values of the CADF test

This paper proposes a way to compute the p-values of the covariate augmented Dickey-Fuller (CADF) test developed in Hansen (1995). The procedure is based on a response surface approach (see e.g. Hendry, 1984). We believe that this is a side, but important, original contribution of our paper. The computer routines have been developed under both GAUSS and R. The R package **CADFtest** that computes Hansen's test and its p-values can be freely downloaded from the Comprehensive R Archive Network at www.cran.r-project.org/package=CADFtest. The use of the package and its main features are illustrated in detail in Lupi (2009). The GAUSS procedures are available upon request from the authors.

In this Appendix we give a detailed account of the algorithm described in Section 2. In order to set up from scratch a procedure that computes the p-values of Hansen's distribution (9), the following steps can be followed:

1. Simulate the asymptotic distribution (9) over a grid of values for $\rho^2 \in (0, 1]$. In the paper we use 40 distinct values. Once ρ^2 is fixed, the asymptotic distribution can be simulated using standard techniques (see e.g. Hatanaka, 1996). In this paper we use $100,000$ replications and $T = 5,000$ for the simulation of the Wiener functionals. The asymptotic distribution must be simulated separately for the "no constant", "constant" and "constant plus trend" case.

2. Derive the simulated quantiles of (9) over a grid of desired probabilities. We use a strict grid of $1,005$ probability values ranging from 0.00025 to 0.99975. Save the results in a table. In our case we have a $1,005 \times 40$ table.

3. For each probability value p considered in the table (i.e., for each row of the table) estimate
$$q_\rho(p) = \beta_0 + \beta_1\rho^2 + \beta_2\left(\rho^2\right)^2 + \beta_3\left(\rho^2\right)^3 + \varepsilon_\rho$$
and save the estimated parameters in a table. In our case we have a $1,005 \times 4$ table of estimated parameters.

4. Use the estimated parameters to derive fitted values $\widehat{q_{\rho_0}(p)}$ $(\forall p)$ of the quantiles for any value of ρ_0^2 you are interested in. $\widehat{q_{\rho_0}(p)}$ is a vector.

5. Plug $\widehat{q_{\rho_0}(p)}$ in (10) following the procedure proposed in MacKinnon (1994, p. 172) and MacKinnon (1996, p.610), that is:
   - Find the fitted quantile that is closest to the sample statistic;
   - Interpolate locally (we used 11 observations) by means of (10);
   - Derive the fitted p-value.

Note that it is not necessary to repeat steps 1-3 each time you want to compute a p-value. Once the β's have been estimated and saved in the relevant tables (for the three cases “no constant”, “constant” and “constant plus trend”), the task is reduced to solving steps 4 and 5 above. In fact, the routines we make available read the estimated β's and solve only steps 4 and 5.

Authors: Mauro Costantini, Claudio Lupi

Title: A Simple Panel-CADF Test for Unit Roots

Reihe Ökonomie / Economics Series 261

Editor: Robert M. Kunst (Econometrics)
Associate Editors: Walter Fisher (Macroeconomics), Klaus Ritzberger (Microeconomics)

ISSN: 1605-7996

Stumpergasse 56, A-1060 Vienna • ☎ +43 1 59991-0 • Fax +43 1 59991-555 • http://www.ihs.ac.at

ISSN: 1605-7996