A Simple Panel-CADF Test for Unit Roots[*]

Mauro Costantini† and Claudio Lupi‡

† *Department of Economics and Finance, Brunel University.*
Kingston Lane, Uxbridge, Middlesex UB8 3PH (UK)
(e-mail: Mauro.Costantini@brunel.ac.uk)
‡*Department of Economics, Management, and Social Sciences, University of Molise.*
Via De Sanctis, I-86100 Campobasso (Italy).
(e-mail: 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 panel test we propose is based on a p values combination approach that takes into account cross-section dependence. We show that the test has good size properties and gives power gains with respect to other popular panel approaches. An empirical application is carried out for illustration purposes on international data to test the PPP hypothesis.

I.   Introduction

In order to obtain more powerful unit root tests, Hansen (1995) suggests using covariate augmented Dickey-Fuller (CADF) tests, i.e. unit root tests that exploit stationary covariates in an otherwise standard Dickey-Fuller framework. In this paper we extend Hansen's CADF test to small panels. Although the CADF test is not the covariate augmented point optimal test in general, we decided to use it for three main reasons. First, simulations reported in Elliott and Jansson (2003) show that the feasible point optimal test 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 panel CADF (pCADF, for short) test we propose is specifically designed for macro-panels, where the time dimension T is large and the number of panel units N is typically fairly small. The test is based on the inverse-normal p value combination advocated in Choi (2001) and extended in Demetrescu *et al.* (2006) to cope with dependence across the panel units. The advantages of this approach are fourfold. First, provided that we can compute the p values of the CADF test, the extension to the panel case is straightforward. Second, the asymptotics carries through for $T \to \infty$, without requiring $N \to \infty$. Third,

---

we do not need balanced panels, so that individual time series may come in different lengths and span different sample periods. Fourth, the test allows for the stochastic as well as the non stochastic components to be different across individual time series. The null hypothesis of the test is that all the series have a unit root, while the alternative is that at least one time series is stationary. Some authors consider this as a disadvantage, but we believe that the extent to which this is a real limitation depends on the specific goal of the analysis. In fact, from the economist's point of view there are instances in which it is especially interesting to test for the presence of a unit root collectively over a whole panel of time series precisely because the presence of a unit root in *all* the series can be interpreted as a stylized fact that can give stronger support in favour (or against) a particular economic interpretation.

Although developed independently, the results reported in the present paper are related to other recent research on covariate augmented panel tests. Despite some similarities, even in the name, the panel-CADF test presented here should not be confused with the cross-sectionally augmented ADF (CADF) test advocated in Pesaran (2007). Chang and Song (2009) also start from the observation that using stationary covariates can greatly improve the power of unit root tests. However, their approach is completely 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 rest of the paper is organized as follows. Section II is devoted to a brief discussion of the test proposed in Hansen (1995) and illustrates the method we use to obtain the necessary p values. Section III offers a brief account of the inverse-normal combination method and its modifications to deal with cross-dependent time series. In Section IV an extensive Monte Carlo analysis of the pCADF test is carried out. For the purpose of illustration, in Section V we apply our pCADF test to the PPP hypothesis. The last Section concludes.

II. The CADF test and the p values approximation

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

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

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, $\boldsymbol{x}_t \sim I(1)$ is an m-vector such that $\Delta \boldsymbol{x}_t \sim I(0)$, $\boldsymbol{\mu}_x := \mathrm{E}(\Delta \boldsymbol{x})$, $\boldsymbol{b}(L) := (\boldsymbol{b}_{q_2} L^{-q_2} + \ldots + \boldsymbol{b}_{q_1} L^{q_1})$ is a polynomial where both leads and lags are allowed. Furthermore, denote by ρ^2 the long-run squared correlation between v_t and e_t. 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 $\rho^2 \approx 1$. The case $\rho^2 = 0$ is ruled out, 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

$$a_0(L)\Delta y_t = \delta_0 y_{t-1} + \boldsymbol{b}_0(L)'\Delta \boldsymbol{x}_t + e_{0t} \tag{4}$$
$$a_\mu(L)\Delta y_t = \mu_\mu + \delta_\mu y_{t-1} + \boldsymbol{b}_\mu(L)'\Delta \boldsymbol{x}_t + e_{\mu t} \tag{5}$$
$$a_\tau(L)\Delta y_t = \mu_\tau + \theta_\tau t + \delta_\tau y_{t-1} + \boldsymbol{b}_\tau(L)'\Delta \boldsymbol{x}_t + e_{\tau t} \tag{6}$$

and is computed as the t-statistic for δ_m, $\widehat{t(\delta_m)}$ (with $m \in \{0, \mu, \tau\}$). Hansen (1995, p. 1154) proves that under the unit root null, if some mild regularity conditions are satisfied, the asymptotic distribution of $\widehat{t(\delta_0)}$ in (4) is

$$\widehat{t(\delta_0)} \xrightarrow{w} \rho \frac{\int_0^1 W \, dW}{\left(\int_0^1 W^2\right)^{1/2}} + \left(1 - \rho^2\right)^{1/2} \mathrm{N}(0,1) \tag{7}$$

where W is a standard Wiener process and $\mathrm{N}(0,1)$ is a standard normal independent of W. The mathematical expression remains unchanged if models (5) and (6) 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) and Demetrescu *et al.* (2006), we need to compute the p values of the CADF unit root distribution (7). Notice that the asymptotic distribution (7) depends on the nuisance parameter ρ^2 but, provided ρ^2 is given, it can be simulated using standard techniques. Therefore, we first 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 ($\rho^2 = 0.025, 0.05, 0.0725, \dots, 1$) using $100,000$ replications for each value of ρ^2 and $T = 5,000$ as far as the Wiener functionals are concerned. From the simulated values we derive $1,005$ estimated asymptotic quantiles, $(0.00025, 0.00050, 0.00075, 0.001, 0.002, \dots, 0.998, 0.999, 0.99925, 0.99950, 0.99975)$. We then use the asymptotic quantiles to compute the p values. To this aim, we follow MacKinnon (1996, p. 610) that proposes using a local approximation of the kind

$$\Phi^{-1}(p) = \gamma_0 + \gamma_1 \widehat{q(p)} + \gamma_2 \widehat{q(p)}^2 + \gamma_3 \widehat{q(p)}^3 + \nu_p \tag{8}$$

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. Equation (8) is estimated only over a relatively small number of points, in order to obtain a *local* approximation.

With respect to MacKinnon (1996), we have the extra difficulty that we have to deal with the nuisance parameter ρ^2. However, given that quantiles change fairly smoothly with ρ^2, we adopt a 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

$$\widehat{q_\rho(p)} = \beta_0 + \beta_1 \rho^2 + \beta_2 \rho^4 + \beta_3 \rho^6 + \varepsilon_\rho \tag{9}$$

where we have used the subscript ρ in $\widehat{q_\rho(p)}$ to indicate the dependence of the quantiles on ρ^2. Finally we apply the procedure advocated in MacKinnon (1996) on the interpolated quantiles to obtain the desired p values.[1]

III.   The inverse-normal combination test

Once the goal of the computation of the p values for the distribution (7) is achieved, the extension of Hansen's test to the panel case is straightforward. Indeed, Choi (2001) shows that under some fairly general regularity conditions, if the cross-section units $i = 1, \dots, N$ are independent, under the null the test statistic $N^{-1/2} \sum_{i=1}^N \widehat{t_i} \xrightarrow{w} \mathrm{N}(0,1)$, 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

---

[1]A more detailed description of the procedure is reported in Lupi (2009) and in the extended discussion paper version of this article (Costantini and Lupi, 2011). As a by-product of our analysis, we can compute a detailed table of asymptotic critical values of the CADF test: see Costantini and Lupi (2011, Table 1).

function, and $\widehat{p}_i$ $(i = 1, \ldots, N)$ the estimated individual p values from standard ADF tests. Convergence takes place as $T \to \infty$, whereas $N < \infty$ is the number of individual time series. The null hypothesis is that all the series have a unit root, while the alternative is that at least one series is stationary. In the presence of cross-section dependence among the time series Choi's test statistic is no longer asymptotically standard normal, but Hartung (1999) suggests that, under the assumption that the pairwise correlation among the individual test statistics ϱ is constant and that the probits are jointly multivariate normal, an asymptotically standard normal combination test can be obtained as

$$
t\left(\hat{\varrho}^*, \kappa\right) := \frac{\sum_{i=1}^{N} \lambda_i \widehat{t}_i}{\sqrt{\sum_{1=1}^{N} \lambda_i^2 + \left[\left(\sum_{i=1}^{N} \lambda_i\right)^2 - \sum_{i=1}^{N} \lambda_i^2\right]\left(\hat{\varrho}^* + \kappa\sqrt{\frac{2}{N+1}}(1 - \hat{\varrho}^*)\right)}} \tag{10}
$$

where $\hat{\varrho}^*$ is a consistent estimator of ϱ and $\kappa > 0$ is a parameter that controls the small sample actual significance level. 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. Furthermore, they show that the necessary and sufficient condition for (10) to have a limiting standard normal distribution is that the individual test statistics from which the probits are derived are such to have the copula of a multivariate normal distribution. Despite the fact that the augmented Dickey-Fuller test does not satisfy this condition, Demetrescu *et al.* (2006) suggest that correcting for cross-dependence using (10) may still be a good practice because the presence of cross-dependence 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, given that under the null Hansen's distribution is a weighted sum of a Dickey-Fuller and a standard normal distribution, 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).

IV. Monte Carlo simulations

In this Section we compare the performance of the pCADF test to that of the tests proposed by Demetrescu *et al.* (2006), Moon and Perron (2004), and 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 and alternative hypothesis.

Structure of the DGP

In our simulations we consider the following DGP:

$$
\Delta \boldsymbol{y}_t = \boldsymbol{\alpha} + \boldsymbol{D}\boldsymbol{y}_{t-1} + \boldsymbol{u}_t \tag{11}
$$

$$
\begin{pmatrix} \boldsymbol{u}_t \\ \xi_t \end{pmatrix} = \begin{pmatrix} \boldsymbol{B} & \boldsymbol{\gamma} \\ \boldsymbol{0}' & \lambda \end{pmatrix} \begin{pmatrix} \boldsymbol{u}_{t-1} \\ \xi_{t-1} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\eta}_t \\ \varepsilon_t \end{pmatrix} \tag{12}
$$

$$
\begin{pmatrix} \boldsymbol{\eta}_t \\ \varepsilon_t \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} \boldsymbol{0} \\ 0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\sigma}_{12} \\ \boldsymbol{\sigma}_{12}' & \sigma_{22} \end{pmatrix}\right] \tag{13}
$$

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 (12) defines a VAR(1) which is stationary as long as $|\beta_i| < 1 \ \forall i$ and $|\lambda| < 1.$[2] $\delta_i = 0 \ \forall i$ 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. The two DGPs that are special cases of ours share the same equation (11) for $\Delta \boldsymbol{y}_t$ when $\boldsymbol{\alpha} = \boldsymbol{0}$, but differ as far as the simulation of the $\boldsymbol{u}_t$'s is concerned:

$$\text{DGP1:} \quad u_{it} \ = \ \beta_i u_{i,t-1} + \nu_{it} \tag{14}$$

$$\text{DGP2:} \quad u_{it} \ = \ \beta_i u_{i,t-1} + \gamma_i \zeta_t + \nu_{it} \tag{15}$$

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

It can be seen that, even when $\boldsymbol{\alpha} = \boldsymbol{0}$, our DGP (11)–(13) is more general than both (14) and (15): 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. However, DGP2 can be derived as a special case from (11)–(13) when $\lambda = 0$ and $\boldsymbol{\sigma}_{12} = \boldsymbol{0}$, while DGP1 is retrieved if in addition $\boldsymbol{\gamma} = \boldsymbol{0}$. In both cases, in general $\boldsymbol{\Sigma}_{11} \neq \boldsymbol{I}$.

Using the DGP (11)–(13) 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 $\boldsymbol{\alpha} = \boldsymbol{0}$. Then, denoting the "past" by $\boldsymbol{Z}_{t-1}$, the correct conditional model for $\Delta y_{i,t}$ is

$$\begin{aligned} \text{E} \left(\Delta y_{i,t} | \xi_t, \boldsymbol{Z}_{t-1} \right) \ &= \ \delta_i \left(1 - \beta_i \right) y_{i,t-1} + \left(1 + \delta_i \right) \beta_i \Delta y_{i,t-1} \\ &+ \ \frac{(\boldsymbol{\sigma}_{12})_i}{\sigma_{22}} \xi_t + \left(\gamma_i - \frac{(\boldsymbol{\sigma}_{12})_i}{\sigma_{22}} \lambda \right) \xi_{t-1} \end{aligned} \tag{16}$$

with $(\boldsymbol{\sigma}_{12})_i$ the i-th element of $\boldsymbol{\sigma}_{12}$. Note that (16) 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 (11)–(13) (which is a fairly standard situation in the panel unit root literature) is bound to be 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 $(\boldsymbol{\sigma}_{12})_i \neq 0$ the correct model has the form of a CADF(1,1,0).

Equation (16) is very similar to an expression derived in Caporale and Pittis (1999, p. 586, eq. 11) and some special cases can be of interest. Under DGP2 ($\lambda = 0$ and $\boldsymbol{\sigma}_{12} = \boldsymbol{0}$) the correct conditional model becomes

$$\text{E} \left(\Delta y_{i,t} | \xi_t, \boldsymbol{Z}_{t-1} \right) = \delta_i \left(1 - \beta_i \right) y_{i,t-1} + \left(1 + \delta_i \right) \beta_i \Delta y_{i,t-1} + \gamma_i \xi_{t-1} \tag{17}$$

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 (17) holds for the i-th unit if only $(\boldsymbol{\sigma}_{12})_i = 0$, while if $\lambda = 0$ and $(\boldsymbol{\sigma}_{12})_i \neq 0$ we have

$$\begin{aligned} \text{E} \left(\Delta y_{i,t} | \xi_t, \boldsymbol{Z}_{t-1} \right) \ &= \ \delta_i \left(1 - \beta_i \right) y_{i,t-1} + \left(1 + \delta_i \right) \beta_i \Delta y_{i,t-1} \\ &+ \ \frac{(\boldsymbol{\sigma}_{12})_i}{\sigma_{22}} \xi_t + \gamma_i \xi_{t-1} \,. \end{aligned} \tag{18}$$

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

$$\text{E} \left(\Delta y_{i,t} | \xi_t, \boldsymbol{Z}_{t-1} \right) \ = \ \delta_i \left(1 - \beta_i \right) y_{i,t-1} + \left(1 + \delta_i \right) \beta_i \Delta y_{i,t-1} \tag{19}$$

---

[2]See Costantini and Lupi (2011) for details.

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.

The power of Hansen's 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 (11)–(13) we can derive analytically the theoretical value of ρ_i^2 under the DGP.

Consider the residual $e_{i,t}$ from the correct conditional model (16)

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

Given that $e_{i,t}$ is the residual from the correct conditional model, it must be an innovation uncorrelated with $\xi_{t-k} \ \forall k$. 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 (where $h \in \{e_i, v_i\}$). 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

$$
v_{i,t} = \frac{\left(\boldsymbol{\sigma}_{12}\right)_i}{\sigma_{22}} \xi_t + \left(\gamma_i - \frac{\left(\boldsymbol{\sigma}_{12}\right)_i}{\sigma_{22}} \lambda\right) \xi_{t-1} + e_{i,t} \, .
\tag{21}
$$

In order to compute the long-run variance of $v_{i,t}$, $\omega_{v_i}^2$, from (12) note that $\xi_t = (1 - \lambda L)^{-1} \varepsilon_t$ and define $r_i := \left(\boldsymbol{\sigma}_{12}\right)_i / \sigma_{22}$. Then, rewrite (21) as

$$
\begin{aligned}
v_{i,t} = & \ \left[r_i + \left(\gamma_i - r_i \lambda\right) L\right] \xi_t + e_{i,t} \\
= & \ \frac{r_i + \left(\gamma_i - r_i \lambda\right) L}{1 - \lambda L} \varepsilon_t + e_{i,t} \, .
\end{aligned}
\tag{22}
$$

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

$$
f_{v_i}(\omega) \propto \frac{\left|r_i + \left(\gamma_i - r_i \lambda\right) e^{-i\omega}\right|^2}{\left|1 - \lambda e^{-i\omega}\right|^2} \sigma_\varepsilon^2 + \sigma_{e_i}^2
\tag{23}
$$

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

$$
\omega_{v_i}^2 := f_{v_i}(0) \propto \frac{\left[\gamma_i + (1 - \lambda) r_i\right]^2}{(1 - \lambda)^2} \sigma_\varepsilon^2 + \sigma_{e_i}^2 \, .
\tag{24}
$$

Finally, ρ_i^2 is given by

$$
\rho_i^2 = \frac{\omega_{e_i}^2}{\omega_{v_i}^2} = \frac{\sigma_{e_i}^2}{\frac{\left[\gamma_i + (1 - \lambda) r_i\right]^2}{(1 - \lambda)^2} \sigma_\varepsilon^2 + \sigma_{e_i}^2} \, .
\tag{25}
$$

The value of ρ_i^2 is a nonlinear function of $\left(\boldsymbol{\sigma}_{12}\right)_i$, 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, even when the VAR(1) (12) 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 (25) 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.

Parameters setting and experimental design

Some care must be exerted in simulating the DGP (11)–(13), especially as far as the simulation of $(\boldsymbol{\eta}_t', \varepsilon_t)'$ is concerned. From (13), $(\boldsymbol{\eta}_t', \varepsilon_t)' \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$, with

$$\boldsymbol{\Sigma} = \left(\begin{array}{cc} \boldsymbol{\Sigma}_{11} & \boldsymbol{\sigma}_{12} \\ \boldsymbol{\sigma}_{12}' & \sigma_{22} \end{array} \right) . \tag{26}$$

We assume $\mathrm{diag}(\boldsymbol{\Sigma}) := \boldsymbol{\imath}$, with $\boldsymbol{\imath} := (1, \dots, 1)$ 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 is transformed into the required correlation matrix $\boldsymbol{\Sigma}_{11}$ by normalization. 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 e.g. by Chang and Song (2009).

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 (26). The matrix $\boldsymbol{\Sigma}$ simulated in this way is symmetric positive definite.[3]

The other parameters of the DGP are generated as in Chang and Song (2009): in particular, $\beta_i \sim \mathrm{U}_{(0.2,0.4)}$ and $\gamma_i \sim \mathrm{U}_{(0.5,3)}$ (with $i = 1, \dots, N$). Under the null $\delta_i = 0 \; \forall i$, under the alternative $\delta_i \sim \mathrm{U}_{(-0.2,-0.01)}$ 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 fixed to 2 in all experiments dealing with power. 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 $\alpha_i \sim \mathrm{U}_{(0.7,0.9)}$.

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. Given that we expect the nuisance parameter ρ^2 to influence the performance of our test, 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 1).

---

[3]See Costantini and Lupi (2011) for a proof.

TABLE 1
Parameters setting. The values of ρ^2 are computed using the means of the
Uniform distributions

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

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 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, the panel test is performed by using the modified weighted inverse-normal combination (10), otherwise Choi's standard inverse-normal combination is utilized. When (10) is used, consistently with Hartung (1999) and Demetrescu *et al.* (2006), in our experiments we set $\lambda_i = 1 \ \forall i$ and $\kappa = 0.2$. The selection of the lags structure for the lagged differences of both the dependent variable and the covariate in the pCADF test equations (4)-(6) is based on the BIC separately for each time series. 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, ξ_t is used as the stationary covariate; second, we consider as the stationary covariate the average of the other differenced series Δy_{jt} ($\forall j \neq i$), as in Chang and Song (2009); third, we use the differenced first principal component of $\Delta \boldsymbol{y}_t$. 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. 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. Furthermore, one could wonder if using a covariate different from ξ_t would ensure convergence of the test statistic to the correct asymptotic distribution. In fact, in Hansen (1995) there is no "true" covariate to be used, and convergence to (7) holds for *any* stationary covariate satisfying Assumption 1 (Hansen, 1995, p. 1151), which in turn is more likely to be satisfied if models (4)-(6) include appropriate lag polynomials. However, while the choice of the stationary covariate(s) does not influence the size of the test (at least asymptotically), it can nevertheless have a significant impact on its power so that the choice of "good" covariates is essential to reach the potential power gains offered by the CADF and pCADF tests.

As far as Demetrescu *et al.*'s panel-ADF test is is concerned, the number of lags is selected also in this case by using the BIC and, differently from Demetrescu *et al.* (2006), Hartung's correction is applied after pre-testing for cross-dependence: if no cross-dependence is detected, then the test is applied as in Choi (2001). For Moon and Perron's test 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 (see Moon and Perron, 2004, p. 94). Finally, as far as the test proposed by Chang and Song (2009) is concerned, for each time series in the panel we select the lag orders of the differences and of the covariate by using the BIC; the covariates are determined by selecting the ones that have the highest correlations with the error processes (see, on this, Chang and Song, 2009, footnote 9).
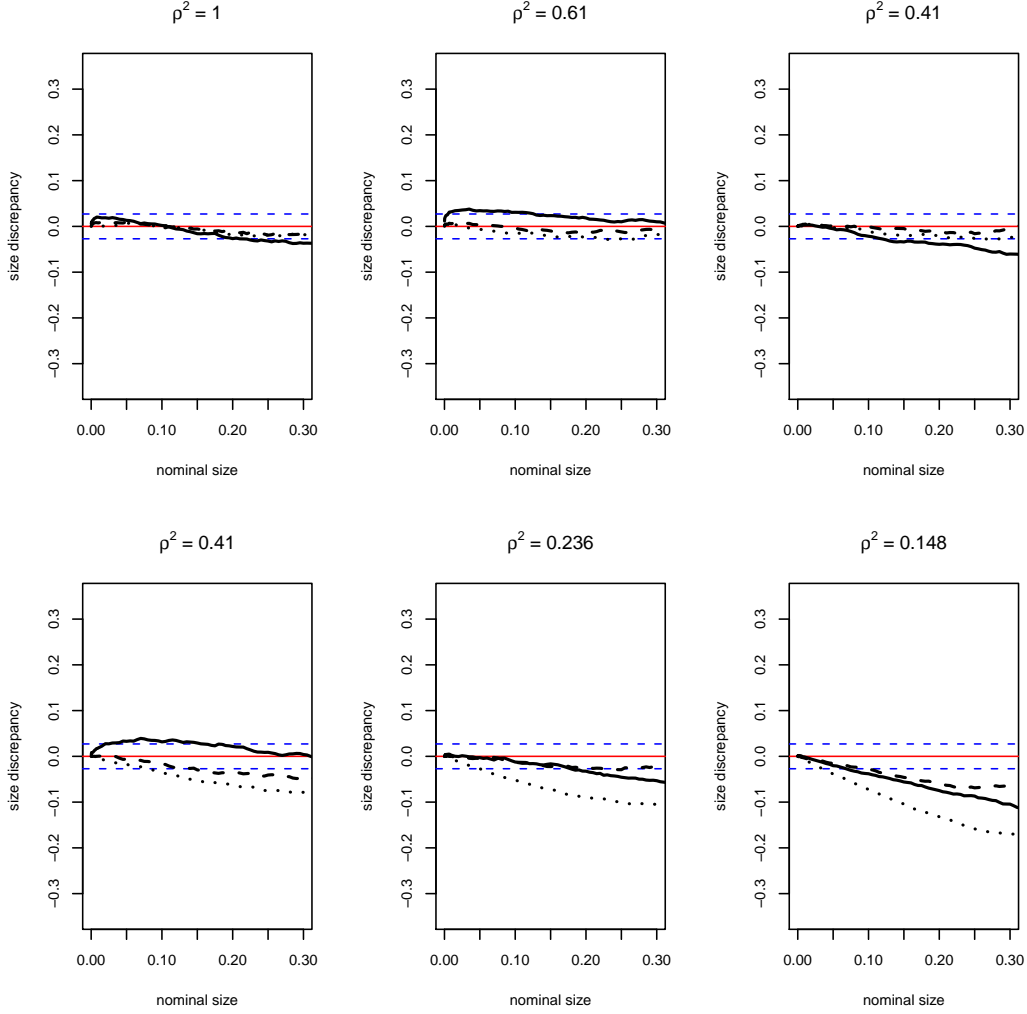
Figure 1. Size discrepancy plots of the pCADF test. The first row refers to experiments 1 to 3, the second to experiments 4 to 6. DGP with no drift, models with no trend. $T = 100$, $N = 10$. Solid line, ξ_t as the stationary covariate; dashed, average Δy_{jt} $(j \neq i)$ as the stationary covariate; dotted, first difference of the first principal component as the stationary covariate. The horizontal dashed lines represent 5% Kolmogorov-Smirnov critical values

Simulation results

The simulation results are presented using the graphical approach proposed in Davidson and MacKinnon (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. Using the p value discrepancy plots it is possible to investigate the size properties of the tests not only in correspondence with 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. Davidson and MacKinnon (1998) call these plots *size-power curves*. By plotting the power on the vertical axis and the actual size on the
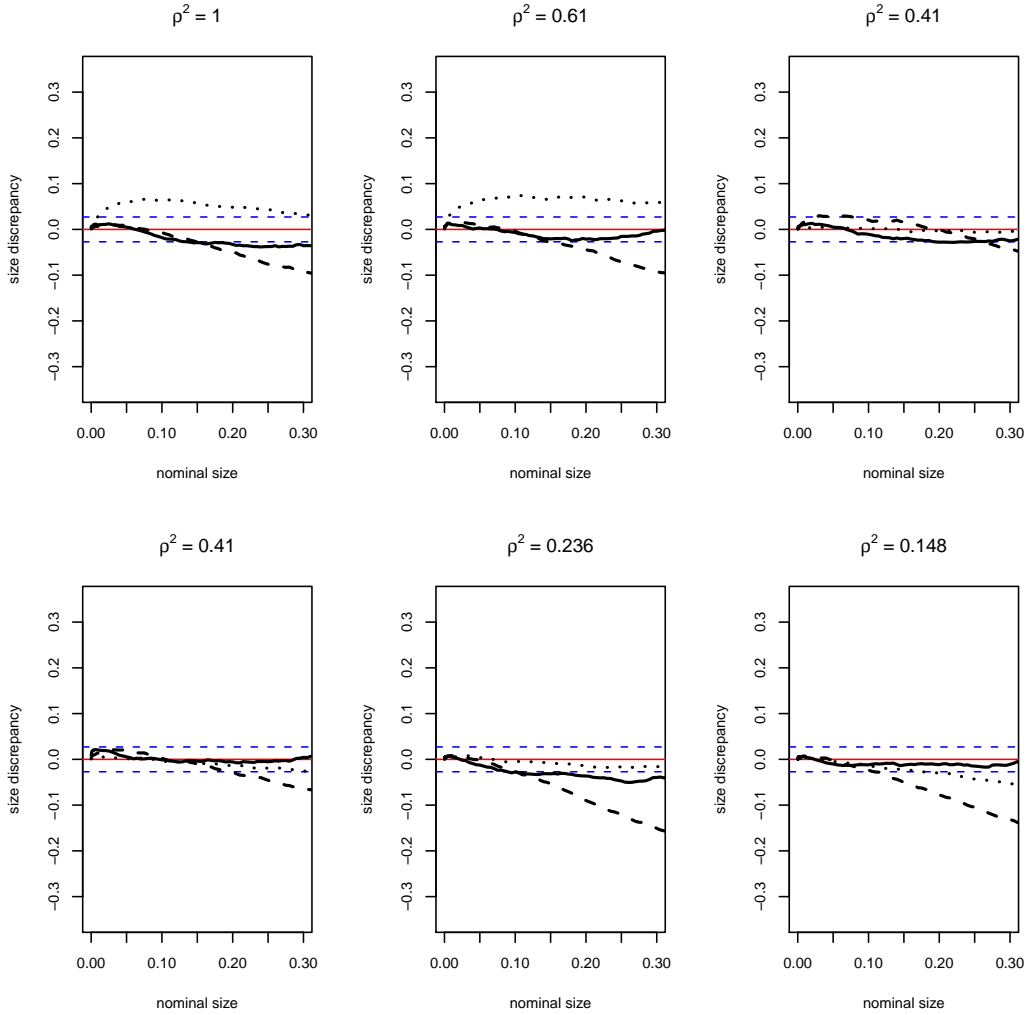
Figure 2. Size discrepancy plots. The first row refers to experiments 1 to 3, the second to experiments 4 to 6. DGP with no drift, models with no trend. $T = 100$, $N = 10$. Solid line, Demetrescu *et al.* (2006); dashed, Chang and Song (2009); dotted, Moon and Perron (2004). The horizontal dashed lines represent 5% Kolmogorov-Smirnov critical values

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. Size-power curves are related to the receiver operating characteristic (ROC) curves (see e.g. Lloyd, 2005). In fact, a plot of the power against the size is the ROC curve of the test. It is worth emphasizing that any point on the estimated ROC (size-power) curve represents the estimated power of the test when the *correct* (as opposed to the *nominal*) critical value for a given test size is utilized. In other words, the ROC curve is a graphical representation of the *intrinsic* (size-adjusted) power of the test (Lloyd, 2005). 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 1 with $\boldsymbol{\alpha} = \boldsymbol{0}$ in the DGP and no trend in the model. The size discrepancies of the tests are reported in Figures 1 and 2. The test proposed by Demetrescu *et al.* (2006) has the best overall size properties across experiments. The pCADF test performs quite well, with no large size discrepancies in correspondence with the usual size levels. However, it tends to be
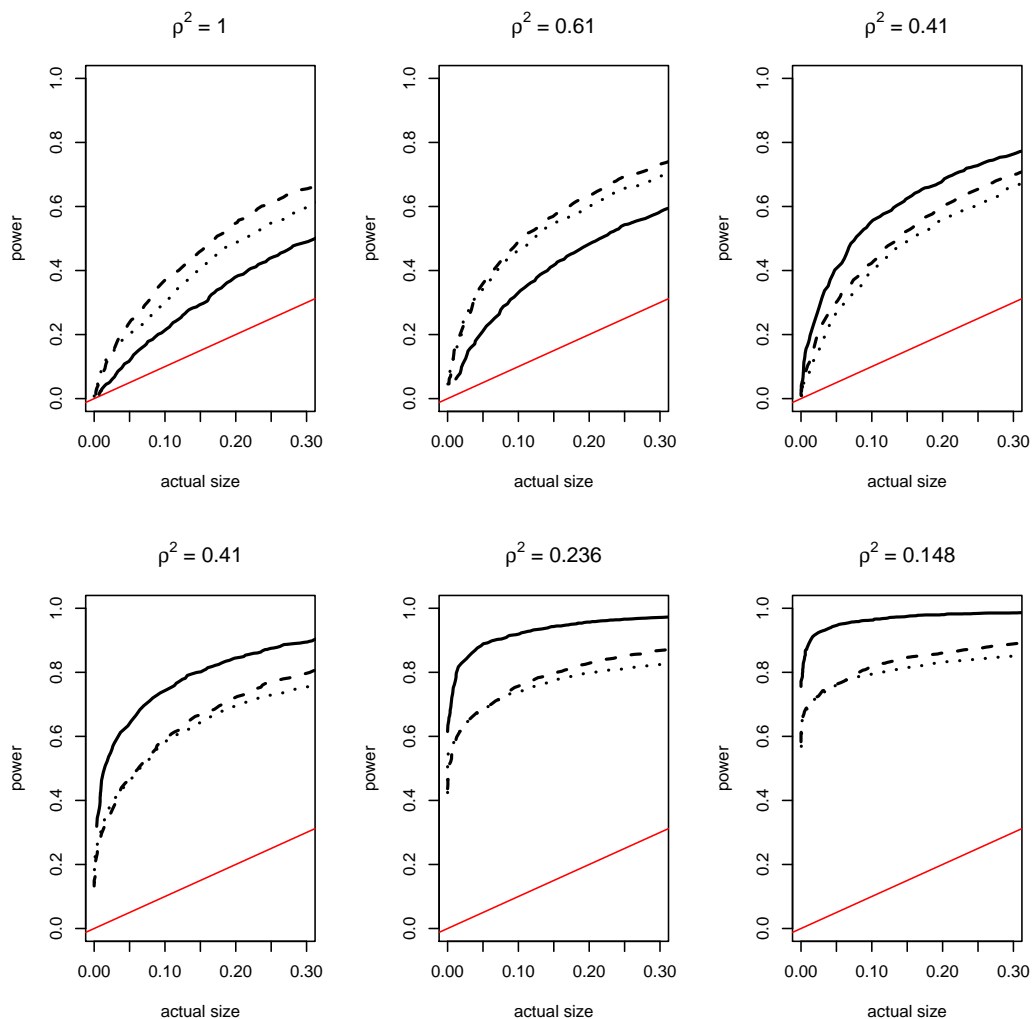
Figure 3. Size-power plots of the pCADF test. The first row refers to experiments 1 to 3, the second to experiments 4 to 6. DGP with no drift, models with no trend. $T = 100$, $N = 10$, 2 series are stationary. Solid line, ξ_t as the stationary covariate; dashed, average Δy_{jt} $(j \neq i)$ as the stationary covariate; dotted, first difference of the first principal component as the stationary covariate

conservative in experiment 6, especially 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-rejection, especially in experiments 5 and 6.

The size-power curves for the same experiments are reported in Figures 3 and 4. In particular, Figure 3 shows that 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 ξ_t is used as the stationary covariate and, for somewhat smaller values of ρ^2 also when the other covariates are used as well. The covariate-augmented test proposed by Chang and Song (2009) shows a rather stable rejection rate across experiments (see Figure 4), with only a fairly small increase for low values of ρ^2. A direct comparison with the pCADF test is offered in Figure 5 that shows
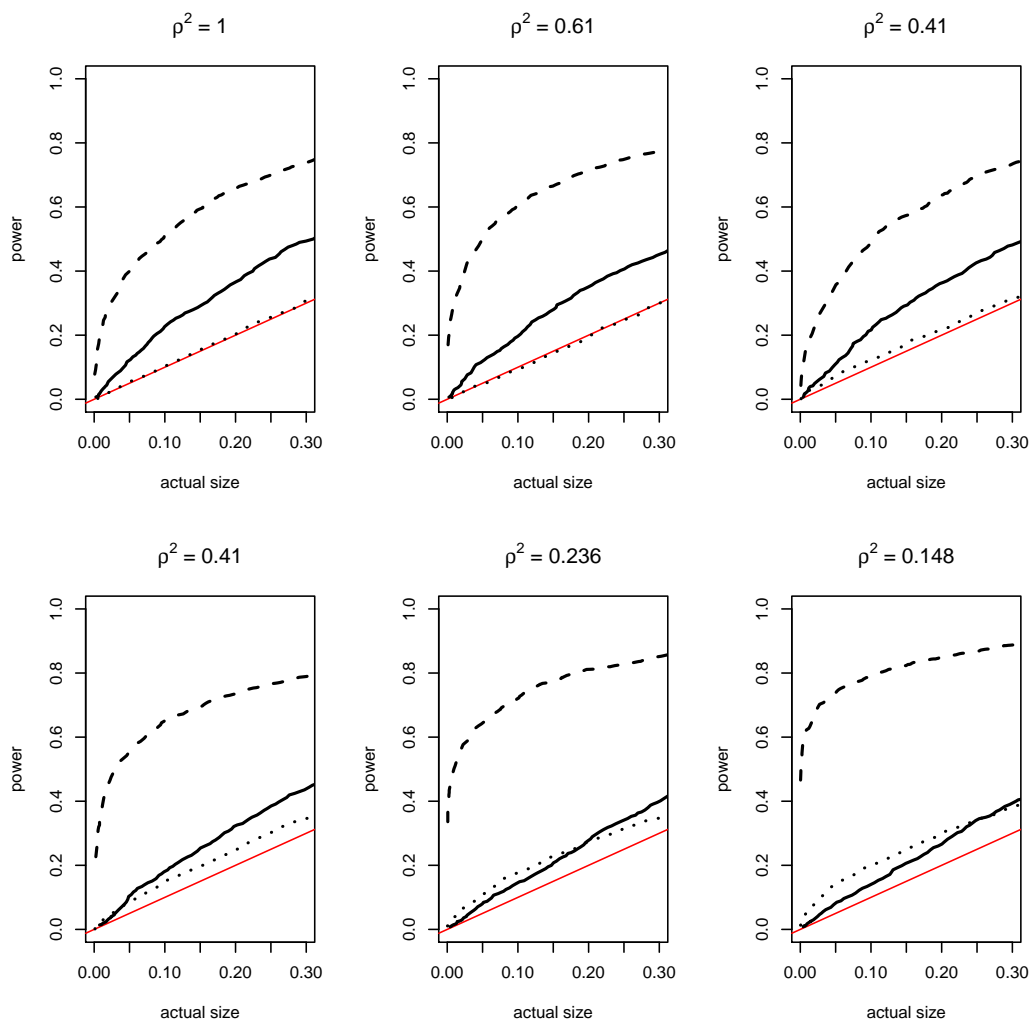
Figure 4. Size-power plots. The first row refers to experiments 1 to 3, the second to experiments 4 to 6. DGP with no drift, models with no trend. $T = 100$, $N = 10$, 2 series are stationary. Solid line, Demetrescu *et al.* (2006); dashed, Chang and Song (2009); dotted, Moon and Perron (2004)
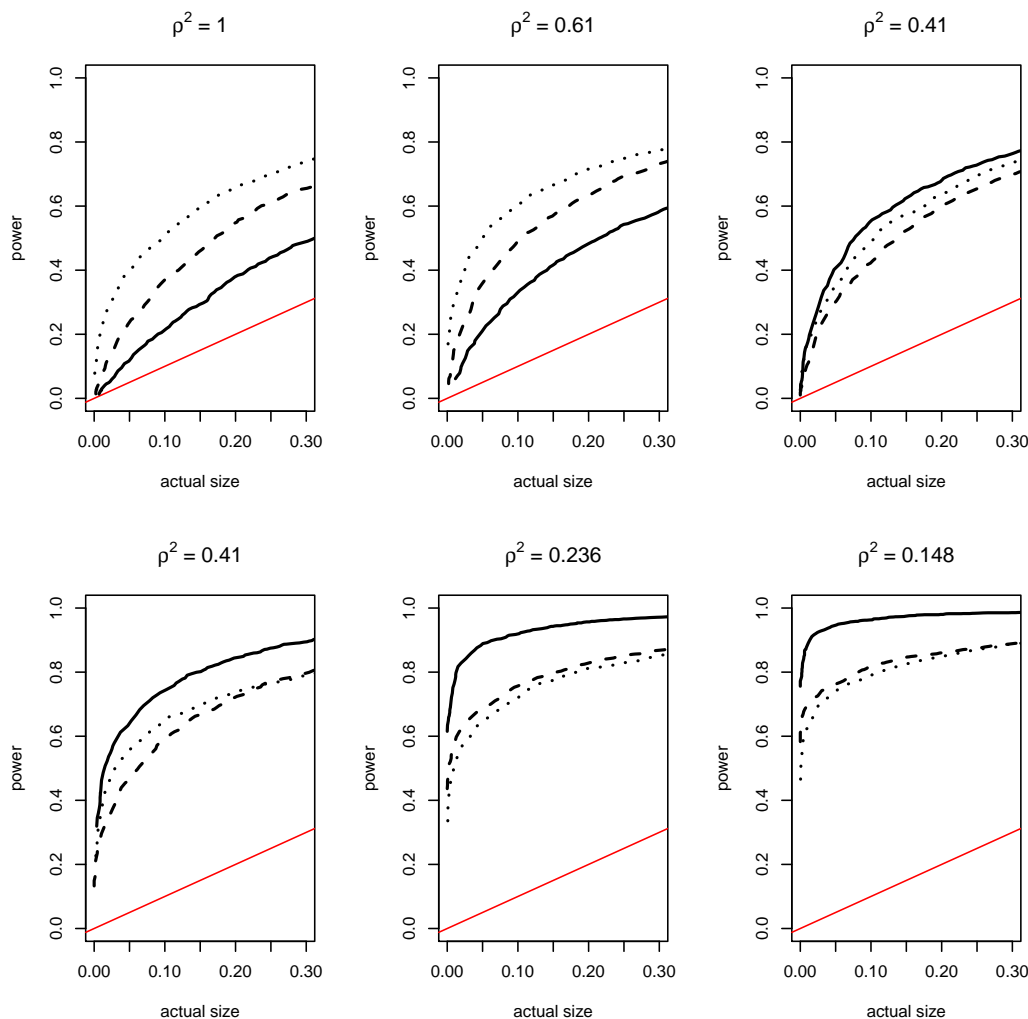
Figure 5. Size-power plots. The first row refers to experiments 1 to 3, the second to experiments 4 to 6. DGP with no drift, models with no trend. $T = 100$, $N = 10$, 2 series are stationary. Solid line, pCADF with ξ_t as the stationary covariate; dashed, pCADF with average Δy_{jt} $(j \neq i)$ as the stationary covariate; dotted, Chang and Song (2009)

that the test proposed by Chang and Song (2009) performs better than the pCADF only for relatively high values of ρ^2. However, it should be reminded that the pCADF is equivalent to the panel ADF test when $\rho^2 = 1$ while, when $\rho^2 < 1$, the power gain obtained by using stationary covariates can be substantial. In fact, the power of Chang and Song's test is still higher than the power of the pCADF test for $\rho^2 = 0.61$. This is due to the fact that, although the power of the pCADF test increases as ρ^2 decreases, nevertheless the relation between the power and ρ^2 is not linear, and larger power gains are expected for fixed decrements of ρ^2 when the value of ρ^2 is small. In fact, simulation results are consistent with the behaviour of the asymptotic power envelope of the (ordinary) CADF test (see Hansen, 1995, p. 1153). Therefore, it is reasonable that the pCADF test becomes more powerful than Chang and Song's only for values of ρ^2 that are below some threshold. Finally, inspection of Figure 4 suggests that the power of Moon and Perron's test is instead rather disappointing, being virtually identical to the size in most experiments.

When the size of the tests with trend (pCADF and Demetrescu *et al.*'s) or detrended (Chang and Song's and Moon and Perron's) over the same DGP as above are considered,

Demetrescu *et al.*'s ADF-based test ranks first, as in the previous case.[4] The pCADF 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, while Chang and Song's test is now very conservative across all the experiments. On the other hand, Moon and Perron's test tends to over-reject substantially. Furthermore, 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 well known result. Despite the observed moderate power reduction, the pCADF test continues to behave quite well, even if 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, comparison with the latter test shows that the power gain deriving from using the stationary covariates is again substantial. Chang and Song's test has good intrinsic power and the rejections remain fairly stable across experiments, as in the no-trend case. The pCADF test still compares well with Chang and Song's, above all when the correct covariate is considered. Finally, Moon and Perron's test has virtually no power at all.

We now extend our analysis 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).

When we allow for a non-zero drift in the DGP, the behaviour of the pCADF test and of Demetrescu *et al.*'s test remains substantially unchanged and fairly good in terms of size (see Figures 6 and 7). On the contrary, Chang and Song's detrended test is so conservative that it hardly rejects even in correspondence with quite high nominal size levels, while Moon and Perron's test rejects much too often (see Figure 7).

The power of the pCADF test (see Figure 8) improves somewhat with respect to the trend case without drift and is very good, compared to Demetrescu *et al.*'s and Moon and Perron's tests (see Figure 9), whose power is very similar to the case without drift. Chang and Song's test maintains good intrinsic power, but it should be emphasised that the correct critical values that ensure that the test has correct size are very different from the theoretical ones so that it is difficult to imagine that the test can be really useful in practice under these circumstances.[5]

In order to check the performance of the tests for larger values of T and N, we repeat the experiments of Table 1 with $T = 300$ and $N = 20$. Power is investigated again using only 2 stationary series. The results essentially confirm the tendencies already highlighted using $T = 100$ and $N = 10$.[6] In particular, when the size of the tests is examined, the pCADF test has approximately the same behaviour as in the $T = 100$ and $N = 10$ case, being slightly conservative especially for low values of ρ^2. 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 corresponding DGP 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 more pronounced than in the $T = 100$, $N = 10$ case. As far as power is concerned, the simulations show that the power of the pCADF test increases with decreasing values of ρ^2 and the test virtually always reject when ρ^2 is small, despite being in the presence of only 2 out of 20 stationary series. In other words, even if the fraction of series under the alternative is smaller than in the previous experiments conducted with $T = 100$ and

---

[4]In order to save space, the figures are not reported for these experiments. The detailed results are available in Costantini and Lupi (2011).

[5]If power is plotted against *nominal* size, it becomes apparent that under this DGP Chang and Song's detrended test is heavily biased, with the empirical rejections being well below the nominal size.

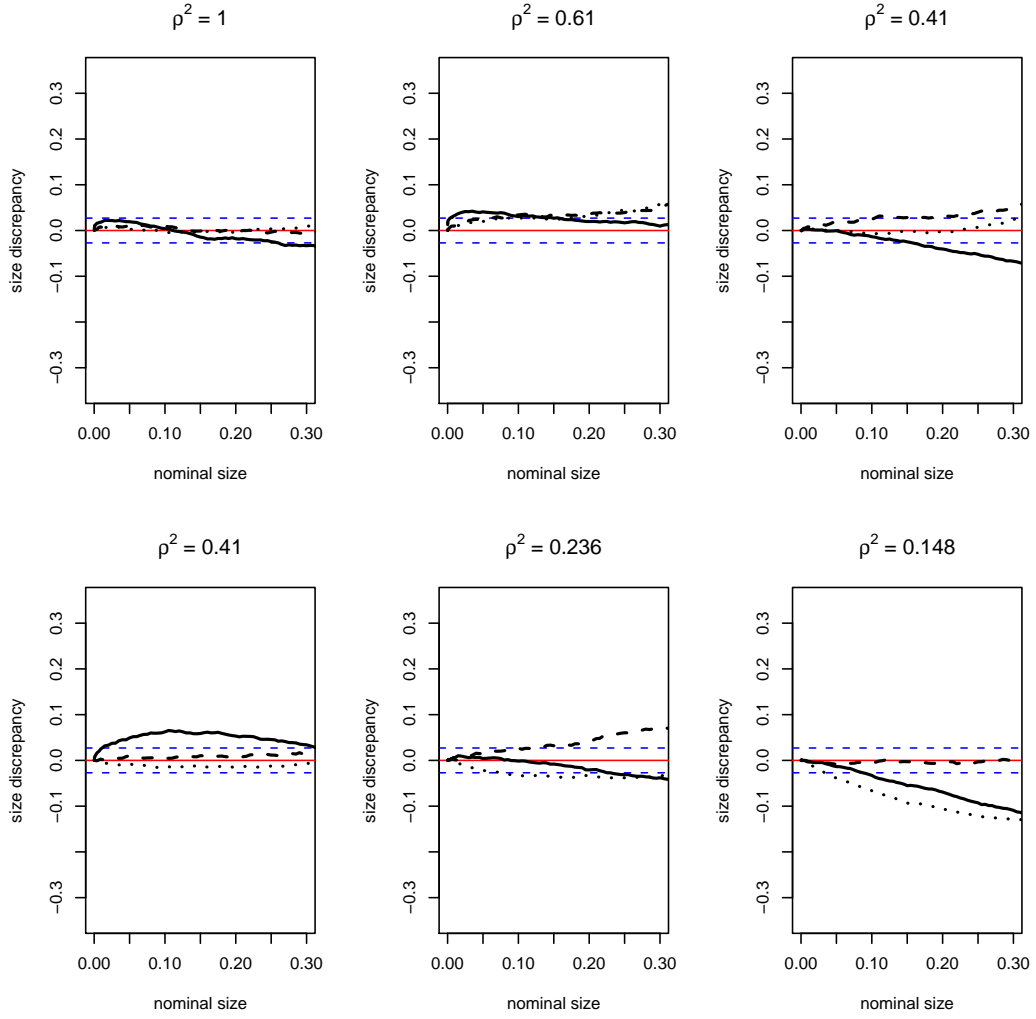[6]To save space we refer the readers to (Costantini and Lupi, 2011) for the detailed results.

Figure 6. Size discrepancy plots of the pCADF test. The first row refers to experiments 1 to 3, the second to experiments 4 to 6. DGP with non-zero drift, models with trend. $T = 100$, $N = 10$. Solid line, ξ_t as the stationary 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
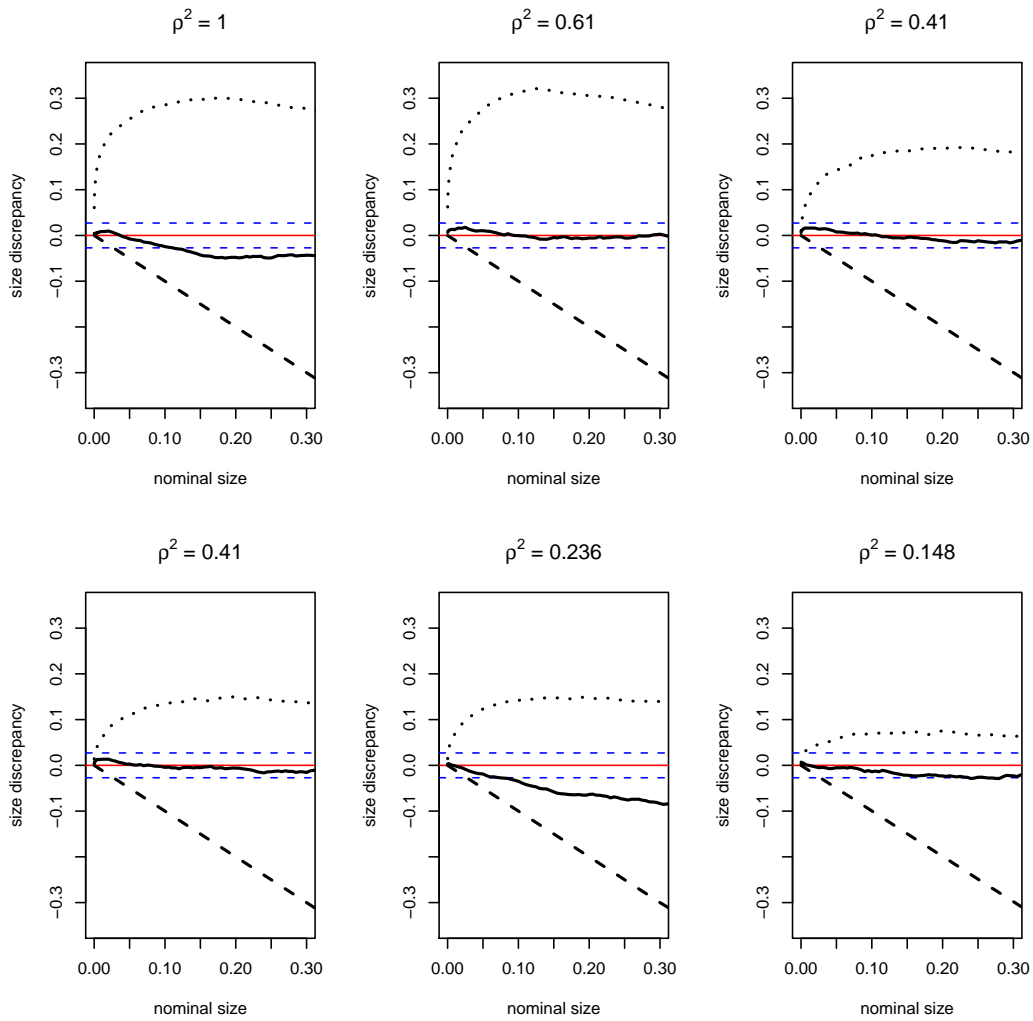
Figure 7. Size discrepancy plots. The first row refers to experiments 1 to 3, the second to experiments 4 to 6. DGP with non-zero drift, models with trend. $T = 100$, $N = 10$. Solid line, Demetrescu *et al.* (2006); dashed, Chang and Song (2009); dotted, Moon and Perron (2004). The horizontal dashed lines represent 5% Kolmogorov-Smirnov critical values
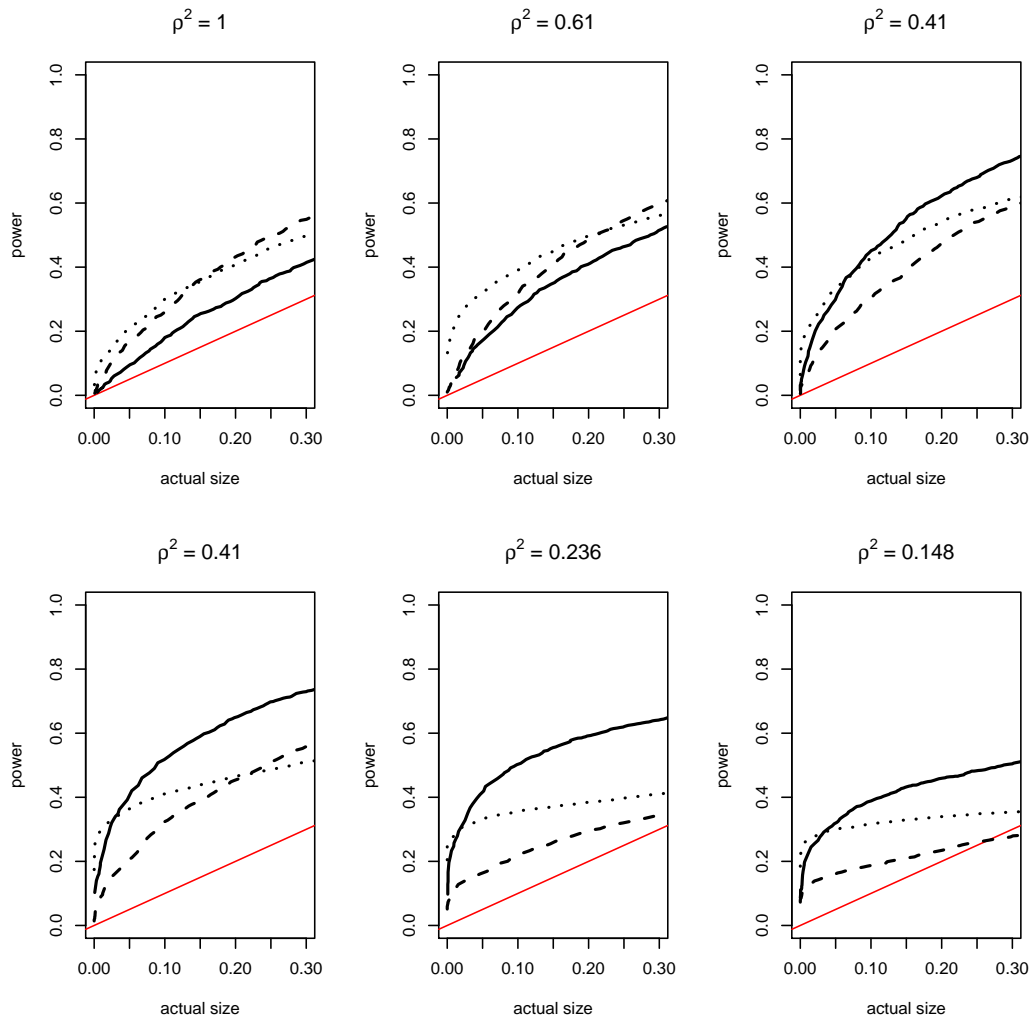
Figure 8. Size-power plots of the pCADF test. The first row refers to experiments 1 to 3, the second to experiments 4 to 6. DGP with non-zero drift, models with trend. $T = 100$, $N = 10$, 2 series are stationary. Solid line, ξ_t as the stationary covariate; dashed, average Δy_{jt} $(j \neq i)$ as the stationary covariate; dotted, first difference of the first principal component as the stationary covariate
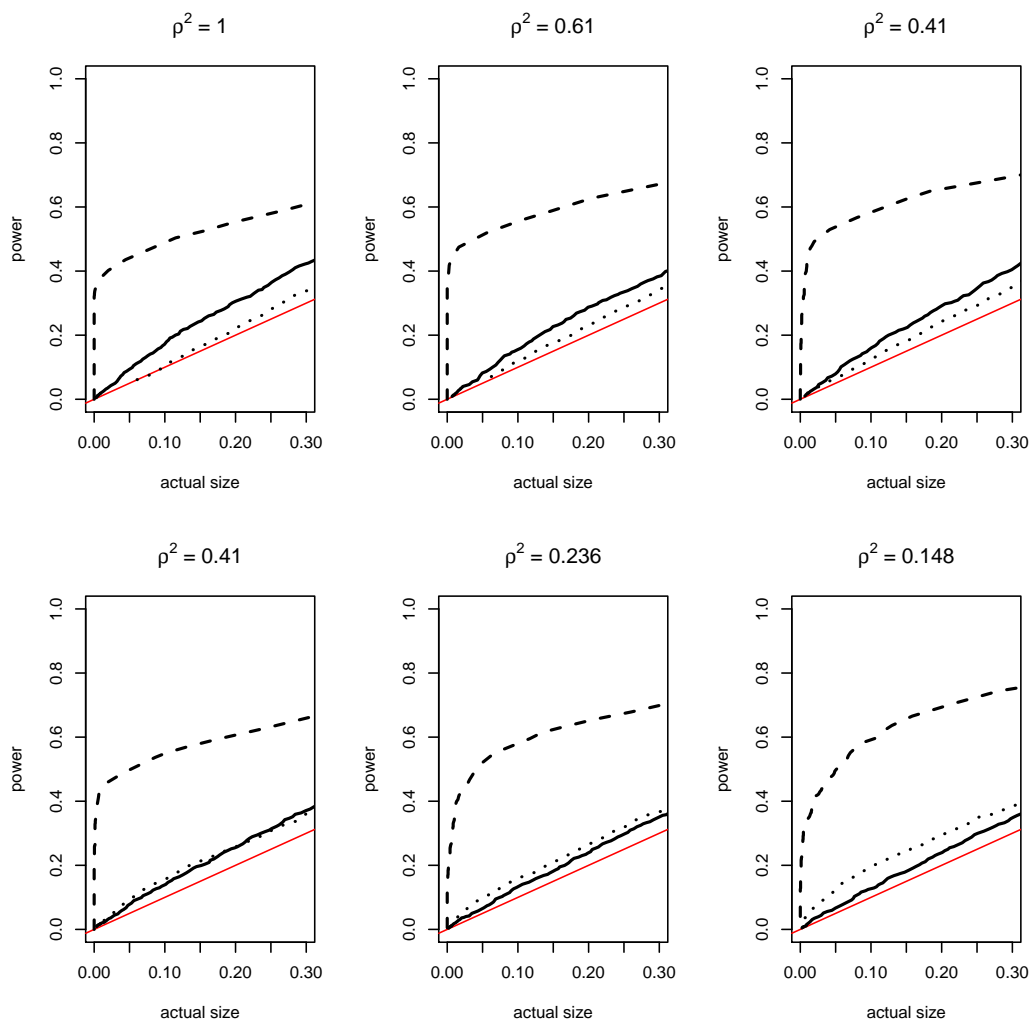
Figure 9. Size-power plots. The first row refers to experiments 1 to 3, the second to experiments 4 to 6. DGP with non-zero drift, models with trend. $T = 100$, $N = 10$, 2 series are stationary. Solid line, Demetrescu *et al.* (2006); dashed, Chang and Song (2009); dotted, Moon and Perron (2004)

TABLE 2
Panel tests of the PPP hypothesis ($T = 103$, $N = 20$)

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

$N = 10$ (where it was 2 out of 10), nevertheless the pCADF test is now substantially more powerful. Furthermore, using either the average of the differenced series or the differenced first principal component gives in this case excellent results, very close to those that can be obtained using ξ_t as the stationary covariate. Moon and Perron's test has again virtually no intrinsic 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.

V. Empirical application

For the sake of illustration, in this Section we offer an application related to the PPP hypothesis.[7] This is a classical application in the panel unit root literature. In the application 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). In addition, in carrying out the pCADF test we also use stationary covariates chosen on theoretical grounds.

It is well known that a necessary condition for the PPP to hold is that the real exchange rate must be *mean-reverting*. For greater comparability with previous works, in our analysis we use quarterly data from Chang and Song (2009) covering the period 1973q1–1998q4.[8] Data for the same countries over the same period have been used in other papers (see e.g. Amara and Papell, 2006; Papell, 2006). Given that under the PPP hypothesis the real exchange rate should not exhibit trends of any kind, in developing our application of the pCADF test, consistently with the existing literature we focus on tests without deterministic trends. Furthermore, following Elliott and Pesavento (2006, pp. 1412–1413), we apply the pCADF test also using the first difference 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 2. Here we also replicate Chang and Song (2009), so our results are identical to theirs. While the other panel tests in Table 2 do not reject the $I(1)$ null, when the differenced nominal exchange rate is used as the stationary covariate, the pCADF test strongly rejects the unit root null consistently with

---

[7]Interested readers can find a further application dealing with international industrial production indices in Costantini and Lupi (2011).

[8]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).

Elliott and Pesavento (2006) that reject the null for most countries when the same covariate is used in the testing procedure proposed by Elliott and Jansson (2003).[9] This result is also broadly consistent with other papers investigating the same data set: in particular, Amara and Papell (2006) find evidence in favour of the PPP hypothesis in many countries again using the time series approach proposed in Elliott and Jansson (2003), while Papell (2006) reaches the same conclusion using a panel-ADF test based on the homogeneous alternative and parametric bootstrap. Strictly speaking, the outcomes of these papers are not directly comparable because they refer to different null and alternative hypotheses; however, they all point in the same direction.

VI.   Concluding remarks

A simple covariate augmented Dickey-Fuller (CADF) test for unbalanced heterogeneous panels is proposed. The test, that we label panel-CADF (pCADF, for short), is a generalization of the CADF test proposed in Hansen (1995) and is developed along the lines suggested in Choi (2001) and Demetrescu *et al.* (2006). This choice allows us to be very general in the specification of the individual unit root tests and makes the test applicable in the presence of cross-dependent time series. Given that the pCADF test is based on a 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 pCADF test are investigated using an extensive Monte Carlo analysis. The performance of the pCADF test is compared with that of the panel unit root tests proposed in Moon and Perron (2004), Demetrescu *et al.* (2006) and Chang and Song (2009). It is shown that the pCADF test in general does not suffer from important size distortions and can offer significant power gains. In all the experiments analysed in the paper, the power of the pCADF 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 pCADF test has the best performance in terms of power, among all the examined tests.

For the sake of illustration we consider an empirical application dealing with the PPP hypothesis.

References

Amara, J. and Papell, D.H. (2006). 'Testing for purchasing power parity using stationary covariates', *Applied Financial Economics*, Vol. 16, pp. 29–39.

Caporale, G.M. and Pittis, N. (1999). 'Unit root testing using covariates: Some theory and evidence', *Oxford Bulletin of Economics and Statistics*, Vol. 61, pp. 583–595.

Chang, Y. and Song, W. (2009). 'Test for unit roots in small panels with short-run and long-run cross-sectional dependencies', *Review of Economic Studies*, Vol. 76, pp. 903–935.

Choi, I. (2001). 'Unit root tests for panel data', *Journal of International Money and Finance*, Vol. 20, pp. 249–272.

Costantini, M. and Lupi, C. (2011). 'A simple panel-cadf test for unit roots', Economics and Statistics Discussion Paper 62/11, University of Molise, URL http://econpapers.repec.org/paper/molecsdps/esdp11062.htm.

---

[9]The fact that the choice of the covariate can influence the outcome of the test is well known and documented in other papers (see, e.g., Elliott and Pesavento, 2006).

Davidson, R. and MacKinnon, J.G. (1998). 'Graphical methods for investigating the size and power of hypothesis tests', *The Manchester School*, Vol. 66, pp. 1–26.

Demetrescu, M., Hassler, U. and Tarcolea, A.I. (2006). 'Combining significance of correlated statistics with application to panel data', *Oxford Bulletin of Economics and Statistics*, Vol. 68, pp. 647–663.

Elliott, G. and Jansson, M. (2003). 'Testing for unit roots with stationary covariates', *Journal of Econometrics*, Vol. 115, pp. 75–89.

Elliott, G. and Pesavento, E. (2006). 'On the failure of purchasing power parity for bilateral exchange rates after 1973', *Journal of Money, Credit, and Banking*, Vol. 38, pp. 1405–1430.

Hanck, C. (2008). 'The error-in-rejection probability of meta-analytic panel tests', *Economics Letters*, Vol. 101, pp. 27–30.

Hansen, B.E. (1995). 'Rethinking the univariate approach to unit root testing: Using covariates to increase power', *Econometric Theory*, Vol. 11, pp. 1148–1171.

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

Kleiber, C. and Lupi, C. (2011). **punitroots**: *Tests for Unit Roots in Panels of (Economic) Time Series, With and Without Cross-sectional Dependence*, URL https://r-forge.r-project.org/projects/punitroots/, R package version 0.0-1.

Lloyd, C.J. (2005). 'On comparing the accuracy of competing tests of the same hypotheses from simulation data', *Journal of Statistical Planning and Inference*, Vol. 128, pp. 497–508.

Lupi, C. (2009). 'Unit root CADF testing with R', *Journal of Statistical Software*, Vol. 32, pp. 1–19.

MacKinnon, J.G. (1996). 'Numerical distribution functions for unit root and cointegration tests', *Journal of Applied Econometrics*, Vol. 11, pp. 601–618.

Moon, H.R. and Perron, B. (2004). 'Testing for a unit root in panels with dynamic factors', *Journal of Econometrics*, Vol. 122, pp. 81–126.

Papell, D.H. (2006). 'The panel purchasing power parity puzzle', *Journal of Money, Credit and Banking*, Vol. 38, pp. 447–467.

Pesaran, M.H. (2004). 'General diagnostic tests for cross section dependence in panels', Cambridge Working Papers in Economics 0435, Department of Applied Economics, University of Cambridge.

Pesaran, M.H. (2007). 'A simple panel unit root test in the presence of cross-section dependence', *Journal of Applied Econometrics*, Vol. 22, pp. 265–312.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, URL http://R-project.org.

Westerlund, J. (2007). 'Testing for error correction in panel data', *Oxford Bulletin of Economics and Statistics*, Vol. 69, pp. 709–748.