

Franz C. Palm, Stephan Smeekes,
Jean-Pierre Urbain

Bootstrap Unit Root Tests: Comparison and
Extensions

RM/06/015

JEL code: C15, C22



Maastricht research school of **E**conomics
of **T**Echnology and **O**Rganizations

Universiteit Maastricht
Faculty of Economics and Business Administration
P.O. Box 616
NL - 6200 MD Maastricht

phone : ++31 43 388 3830
fax : ++31 43 388 4873

BOOTSTRAP UNIT ROOT TESTS: COMPARISON AND EXTENSIONS

Franz C. Palm*, Stephan Smeekes, Jean-Pierre Urbain

Department of Quantitative Economics

Universiteit Maastricht

April 6, 2006

Abstract

In this paper we study and compare the properties of several bootstrap unit root tests recently proposed in the literature. The tests are Dickey-Fuller or Augmented DF-tests, either based on residuals from an autoregression and the use of the block bootstrap (Paparoditis & Politis, 2003) or on first differenced data and the use of the stationary bootstrap (Swensen, 2003a) or sieve bootstrap (Psaradakis, 2001; Chang & Park, 2003). We extend the analysis by interchanging the data transformations (differences versus residuals), the types of bootstrap and the presence or absence of a correction for autocorrelation in the tests. We prove that two sieve bootstrap tests based on residuals remain asymptotically valid, thereby completing the proofs of validity for all the types of DF bootstrap tests.

In contrast to the literature which basically focuses on a comparison of the bootstrap tests with an asymptotic test, we compare the bootstrap tests among them using response surfaces for their size and power in a simulation study. We also investigate how the tests behave when accounting for a deterministic trend, even in the absence of such a trend in the data.

This study leads to the following conclusions: (i) augmented DF-tests are always preferred to standard DF-tests; (ii) the sieve bootstrap performs slightly better than the block bootstrap; (iii) difference-based and residual-based tests behave similarly in terms of size although the latter appear more powerful. The results for the response surfaces allow us to make statements about the behaviour of the bootstrap tests as sample size increases.

Keywords: bootstrap unit root tests, monte carlo, response surface.

JEL Codes: C15, C22

*Corresponding author: Department of Quantitative Economics, Universiteit Maastricht, P.O. Box 616, 6200 MD Maastricht, The Netherlands. Email: f.palm@ke.unimaas.nl. We thank Jeroen van den Berg for his assistance with our simulations and participants at the Econometric Society World Congress, London, August 2005, for helpful comments and suggestions.

1 Introduction

In this paper we study and compare the properties of some bootstrap unit root tests that have recently been proposed in the literature. We also introduce some new tests, prove their first order asymptotic validity and compare them to existing tests.

Due to the good performance of the bootstrap in finite samples for stationary processes, its application to nonstationary series has recently become increasingly popular. This paper deals with the tests proposed by Psaradakis (2001), Chang & Park (2003), Paparoditis & Politis (2003) and Swensen (2003a). These papers consider bootstrap unit root tests, but differ in the way the tests are constructed. Besides showing the asymptotic validity, all these papers compare the finite sample performance of their test(s) to the asymptotic counterpart(s), and the results are overall encouraging. It is however less clear how these tests perform compared to *each other*. This is what we will analyse in this paper. Our goal is to find out which tests perform best under circumstances to be given, and more specifically which aspects of the tests lead to a good or bad finite sample performance. We will analyse and compare the asymptotic properties of these tests, and we will also consider Monte Carlo simulations.

We distinguish three main features of the tests. The first is the actual test statistic. While many different unit root tests are available, most papers restrict their attention to Dickey & Fuller (1979, 1981) type of tests. Some tests use the Dickey-Fuller (DF) test, others the augmented Dickey-Fuller (ADF). As the ADF statistic is asymptotically pivotal, whereas the DF is not, we might expect a bootstrap ADF test to offer asymptotic refinements over the bootstrap DF test and asymptotic tests (Horowitz, 2001). The second feature is which series exactly should be resampled. Bootstrapping a nonstationary series directly is not valid (Basawa et al., 1991). Therefore a stationary series has to be constructed first. Some tests use residuals from a first-order autoregression of the series, others use first-differences of the series. Swensen (2003b) shows that power functions are the same for both cases if the innovations are iid. However as shown by Paparoditis & Politis (2003, 2005), differences lead to poor behaviour of the bootstrap tests under the alternative. The third is the time series bootstrap method that is employed. Some tests that we consider use some form of the *block bootstrap*, in which blocks of (restricted) residuals are resampled. Other tests use the *sieve bootstrap*, that fits an AR model to the (restricted) residuals and resamples the residuals of this AR model. The sieve bootstrap is somewhat easier to use and performs better when valid, but the block bootstrap is valid for more general processes.

Currently, no tests that use the sieve bootstrap and are based on residuals have been shown to be asymptotically valid for the DGPs considered in this paper. We adapt the sieve bootstrap tests by Psaradakis (2001) and Chang & Park (2003) by basing them on residuals instead of differences and show that these new tests are asymptotically valid. As residual-based tests may have better properties under the alternative than difference-based tests, this is an important extension. With these results, all the tests considered in this paper have been

shown to be asymptotically valid.

A word on notation. We denote weak convergence by ‘ \xrightarrow{d} ’, convergence in probability by ‘ \xrightarrow{p} ’ and almost sure convergence by ‘ $\xrightarrow{a.s.}$ ’. $W(r)$ indicates a standard Brownian motion. As usual, we use the superscript ‘*’ to denote bootstrap quantities, both for bootstrap samples and statistics calculated for bootstrap samples. Similarly, ‘ $\xrightarrow{d^*}$ ’ indicates weak convergence of a bootstrap statistic conditional on the original series.

The structure of the paper is as follows. In Section 2, we discuss the bootstrap unit root tests, highlight several features of these tests and prove the asymptotic validity of the new tests proposed. Section 3 contains an extensive Monte Carlo simulation analysis of the various bootstrap unit root tests. The results are summarised using response surfaces. Section 4 concludes.

2 The tests

In this section we discuss several bootstrap unit root tests from a theoretical point of view. We will start with a description of the tests. We will focus on DF-type tests.

2.1 DF sieve bootstrap test

2.1.1 Difference-based DF sieve bootstrap test: Psaradakis (2001)

Psaradakis (2001) considers the following Data Generating Process (DGP) for the time series $y_t, t = 1, \dots, n$:

$$\begin{aligned} y_t &= d_t + v_t \\ v_t &= \rho v_{t-1} + u_t \end{aligned} \tag{1}$$

where d_t consists of deterministic components. Three cases for the deterministic components are considered: the first case is without deterministic components, $d_t = 0$, the second case is with only a constant term, $d_t^\mu = \delta_0$, and the third case is with constant term and linear time trend, $d_t^\tau = \delta_0 + \delta_1 t$. The process u_t is assumed to satisfy the following condition:

Assumption 1

(i) u_t is the linear process

$$u_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \quad (\psi_0 := 1) \tag{2}$$

where ε_t are iid random variables with $E[\varepsilon_1] = 0$, $E[\varepsilon_1^2] = \sigma_\varepsilon^2 > 0$ and $E[\varepsilon_1^4] < \infty$.

(ii) The sequence of constants ψ_j in (2) is such that $\sum_{j=0}^{\infty} j|\psi_j| < \infty$, $\sum_{j=0}^{\infty} \psi_j \neq 0$, and $\sum_{j=0}^{\infty} \psi_j z^j \neq 0$ in $\{z \in \mathbb{C} : |z| \leq 1\}$.

Note that this assumption implies that u_t is an invertible linear process; see Phillips & Solo (1992) for more details. We can rewrite the model (1) into the following form

$$y_t = \rho y_{t-1} + d_t^\dagger + u_t, \quad (3)$$

where $d_t^\dagger = \gamma_0 + \gamma_1 t := (1 - \rho)\delta_0 + \rho\delta_1 + (1 - \rho)\delta_1 t$ (in the first case $\delta_0 = \delta_1 = 0$, in the second case $\delta_1 = 0$). Psaradakis considers the DF coefficient test $n(\hat{\rho} - 1)$ and t-test in equation (3) for testing $\rho = 1$. As stated above, the assumptions on the innovations allow for a sieve bootstrap.

Psaradakis (2001) furthermore needs the following assumption on the order of the autoregression:

Assumption 2 *The order p of the autoregressive approximation is such that $p = p(n) \rightarrow \infty$ as $n \rightarrow \infty$ with $p(n) = o((n/\ln n)^{1/4})$.*

The order of the autoregression can be selected using standard information criteria.

The exact bootstrap procedure can be described as follows.

Bootstrap Test 1 (Psaradakis, 2001)

1. *Fit an AR(p) model to \hat{u}_t , where $\hat{u}_t = \Delta y_t$ if the deterministic part consists of at most a constant term, and $\hat{u}_t = \Delta y_t - n^{-1} \sum_{t=1}^n \Delta y_t$ if the deterministic part contains both a constant term and a linear time trend*

$$\hat{\varepsilon}_t = \hat{u}_t - \sum_{j=1}^p \hat{\phi}_j \hat{u}_{t-j}, \quad t = 1 + p, \dots, n.$$

2. *Generate an iid sample ε_t^* by drawing randomly with replacement from $\hat{\varepsilon}_t - (n-p)^{-1} \sum_{t=1+p}^n \hat{\varepsilon}_t$.*
3. *Construct bootstrap errors by the recursion*

$$u_t^* = \sum_{j=1}^p \hat{\phi}_j u_{t-j}^* + \varepsilon_t^*. \quad (4)$$

4. *The bootstrap sample y_t^* is generated recursively by*

$$y_t^* = y_{t-1}^* + u_t^*$$

in case of no deterministic components or an intercept only, and by

$$y_t^* = n^{-1} \sum_{t=1}^n \Delta y_t + y_{t-1}^* + u_t^*$$

in case of a constant term and a linear trend.

5. Calculate the DF coefficient test and t-test using the bootstrap sample for the previously specified deterministic specification.
6. Repeat steps 2 to 5 B times to find the bootstrap distributions where B denotes the number of bootstrap replications.

Psaradakis (2001) suggests to estimate the AR(p) model in step 1 using the Yule-Walker equations to ensure that the generated innovations u_t^* admit a one-sided MA(∞) representation, but OLS can be used as well, as it is asymptotically equivalent to Yule-Walker. The asymptotic distribution of the bootstrap statistics under the null is shown to be the same as the asymptotic distribution of the original statistic. Note that although the limiting distributions contain nuisance parameters, this does not matter for the bootstrap approach as the critical values for testing are based on the (empirical) distributions of the bootstrap tests that can be approximated by simulation with any accuracy desired.

2.1.2 Residual-based DF sieve bootstrap test: Psaradakis modified

In this section we propose a DF sieve bootstrap test based on residuals. This test is very similar to the test by Psaradakis (2001), except that it is based on residuals. Paparoditis & Politis (2005) have proposed an ADF coefficient test, and we construct our test in the same way as they do. In this section we will show that our test is asymptotically valid when considering the assumptions made by Psaradakis (2001).

We can describe the new algorithm in a very brief way, as the algorithm for the tests by Psaradakis (2001) only needs to be modified in one step:

Bootstrap Test 2 (Residual-based DF sieve bootstrap procedure)

Replace step 1 from Bootstrap Test 1 by

1. Calculate the residuals from the regression

$$\hat{\varepsilon}_t = \tilde{y}_t - \hat{\rho}\tilde{y}_{t-1} - \sum_{j=1}^p \hat{\phi}_j \Delta \tilde{y}_{t-j}, \quad t = 1 + p, \dots, n. \quad (5)$$

where $\tilde{y}_t = y_t$ in the case of a (possibly zero) intercept and $\tilde{y}_t = y_t - \hat{\gamma}_0 - \hat{\gamma}_1 t$ in the case of a linear trend, and $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are the corresponding OLS estimates.

We show that, under the assumptions given above, the bootstrap distributions converge to the same limit distribution as the standard test statistics. First we define

$$\sigma_u^2 = \lim_{n \rightarrow \infty} n^{-1} \text{E}[u_t^2] \quad (6a)$$

$$\sigma^2 = \lim_{n \rightarrow \infty} n^{-1} \text{E}\left[\left(\sum_{t=1}^n u_t\right)^2\right]. \quad (6b)$$

Now we can state the following theorem:

Theorem 1 *Let $n(\hat{\rho}^* - 1)$ and t^* be the coefficient and t-statistic, respectively, that follow from Bootstrap Procedure 2. Under Assumptions 1 and 2, we have that*

$$\begin{aligned} n(\hat{\rho}^* - 1) &\xrightarrow{d^*} \frac{\int_0^1 W(r)dW(r) + (\sigma^2 - \sigma_u^2)/2\sigma^2}{\int_0^1 W(r)^2 dr} \\ t^* &\xrightarrow{d^*} \frac{\int_0^1 W(r)dW(r) + (\sigma^2 - \sigma_u^2)/2\sigma^2}{\left((\sigma_u^2/\sigma^2) \int_0^1 W(r)^2 dr\right)^{1/2}}. \end{aligned}$$

where $W(r)$ is a standard Brownian motion on $[0, 1]$.

Proof: See Appendix A.

We have shown that the DF sieve test as constructed by Psaradakis (2001) remains asymptotically valid if it is based on residuals instead of differences.

2.2 ADF sieve bootstrap test

2.2.1 Difference-based ADF sieve bootstrap test: Chang & Park (2003)

Chang & Park (2003) consider the DGP

$$y_t = \rho y_{t-1} + u_t, \tag{7}$$

where $u_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$. The following assumptions are used:

Assumption 3

(i) u_t is the linear process

$$u_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \quad (\psi_0 := 1) \tag{8}$$

where ε_t is a sequence of iid random variables with $E[\varepsilon_t] = 0$, $E[\varepsilon_t^2] = \sigma_\varepsilon^2$, and $E|\varepsilon_t|^r < \infty$ for some $r \geq 4$,

(ii) Let $\sum_{j=0}^{\infty} \psi_j z^j \neq 0$ for $\{z \in \mathbb{C} : |z| \leq 1\}$, and $\sum_{j=0}^{\infty} |j|^s |\psi_j| < \infty$ for some $s \geq 1$.

For the order of the autoregressive approximation, Chang & Park (2003) consider two different assumptions:

Assumption 4 *Let $p(n) \rightarrow \infty$ and $p(n) = o(n^\kappa)$ with $\kappa < \frac{1}{2}$ as $n \rightarrow \infty$.*

The following assumption is stronger.

Assumption 5 *Let $p(n) = cn^\kappa$ for some constant c and $1/rs < \kappa < \frac{1}{2}$.*

These assumptions are very similar to the assumptions Psaradakis (2001) uses. Like Psaradakis, Chang & Park (2003) assume that u_t follows an invertible linear process. The main difference is the rate at which the lag length p is allowed to grow. Furthermore there is a difference if $r > 4$ and/or $s > 1$, which means stronger assumptions.¹

The bootstrap procedure of Chang & Park (2003) is very similar to the procedure of Psaradakis (2001):

Bootstrap Test 3 (Chang and Park, 2003)

Follow the same steps as in the procedure for the Psaradakis test, but only for the deterministic specification $d_t = 0$. Replace step 5 by

5. Calculate the ADF coefficient statistic $(1 - \sum_{j=1}^p \hat{\phi}_j)^{-1}n(\hat{\rho}^* - 1)$ and the corresponding t -statistic² from the ADF regression

$$y_t^* = \rho^* y_{t-1}^* + \sum_{j=1}^p \phi_j^* \Delta y_{t-j}^* + \varepsilon_t^*.$$

Chang & Park (2003) show that their bootstrap tests converge to the same asymptotic distributions under the null as the asymptotic tests. The convergence is shown to hold almost surely under the strong assumptions, and in probability under the weaker assumptions. They claim that their tests are also valid when applied to demeaned or detrended data, but they do not provide any further analysis. As these distributions are asymptotically pivotal, meaning that the limiting distribution does not depend on nuisance parameters, one might expect that the ADF bootstrap test offers asymptotic refinements.

2.2.2 Residual-based ADF sieve bootstrap test: Chang and Park modified

Similar to the previous section, we construct a residual-based test that is based on the test by Chang & Park (2003) and resembles the residual-based ADF test of Paparoditis & Politis (2005) strongly.

Bootstrap Test 4 (Residual-based ADF sieve bootstrap test) *Replace step 1 from Bootstrap Test 3 by*

1. Calculate the residuals from an ADF regression as in the following equation

$$\hat{\varepsilon}_t = y_t - \tilde{\rho} y_{t-1} - \sum_{j=1}^p \hat{\phi}_j \Delta y_{t-j}, \quad t = 1 + p, \dots, n. \tag{9}$$

¹If $r > 4$ higher moments exist; if $s > 1$ the dependence dies out more quickly.

²Chang and Park suggest using $\hat{\sigma}_\varepsilon^2$ for the t-test instead of $\hat{\sigma}_\varepsilon^{*2}$, although both are appropriate. Similarly, it is possible to use $1 - \sum_{j=1}^p \hat{\phi}_j^*$ for the coefficient test.

The next theorem shows that, under the assumptions given above, the bootstrap distributions converge to the same limit distributions as the asymptotic test statistics.

Theorem 2 *Let $n(\hat{\rho}^* - 1)$ and t^* be the bootstrap coefficient statistic and t -statistic, respectively, that follow from Bootstrap Test 4. Let Assumption 3 and 4 hold. Then*

$$\begin{aligned} n(\hat{\rho}^* - 1) &\xrightarrow{d^*} \frac{\int_0^1 W(r)dW(r)}{\int_0^1 W(r)^2 dr}, \\ t^* &\xrightarrow{d^*} \frac{\int_0^1 W(r)dW(r)}{\left(\int_0^1 W(r)^2 dr\right)^{1/2}}. \end{aligned}$$

Proof: See Appendix A.

We have shown that the ADF sieve test as constructed by Chang & Park (2003) is also asymptotically valid if it is based on residuals, although in Theorem 2 we have obtained convergence in distribution whereas Chang & Park (2003) proved a.s. convergence for their strong assumptions. By imposing the unit root restriction difference-based tests rely on stationary series for which a.s. convergence holds. Although not imposing the unit root when applying the sieve bootstrap is certainly a drawback, we still feel our result is worthwhile as it does provide justification for using a residual-based sieve bootstrap, even if it is not the same justification as Chang & Park (2003) provide for their tests.

2.3 (A)DF block bootstrap test

2.3.1 Residual-based (A)DF block bootstrap test: Paparoditis & Politis (2003)

Paparoditis & Politis (2003) propose a block bootstrap method to test for unit roots. Their method, the residual-based block bootstrap (RBB), is a block bootstrap method applied to residuals of a regression of the series y_t on its first lag. We first state the assumptions under which the RBB is appropriate. Two sets of assumptions are considered, such that one of these should be satisfied by the process y_t to validate the use of the RBB. Paparoditis & Politis (2003) consider the process

$$y_t = \alpha + \rho y_{t-1} + u_t \tag{10}$$

where if $\alpha \neq 0$ there is a drift under the null of $\rho = 1$.

For u_t assumption 6 (i) and (ii)(A) must hold:

Assumption 6

(i) the process u_t is generated by

$$u_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \tag{11}$$

with ε_t a sequence of iid random variables with $E[\varepsilon_t] = 0$, $E[\varepsilon_t^2] = \sigma_\varepsilon^2 > 0$ and $E[\varepsilon_t^4] < \infty$.

(ii) (A) Let $\psi_0 = 1$, $\sum_{j=1}^{\infty} j|\psi_j| < \infty$ and $\sum_{j=0}^{\infty} \psi_j \neq 0$.

(B) $\sum_{j=0}^{\infty} \psi_j z^j$ is bounded, and bounded away from zero for $\{z \in \mathbb{C} : |z| \leq 1\}$.

This condition simply states that u_t is a stationary linear process under the null while under the alternative y_t is stationary and linear, and under the additional assumption (ii)(B) it is invertible as well. This assumption is similar to those Psaradakis and Chang and Park employ.

The second condition states that u_t should satisfy a mixing condition. In contrast to the condition needed for the sieve bootstrap, one should note that the generating process of u_t does not have to belong to the class of linear processes to satisfy this condition. Hence, we see here a class of processes (possibly non-linear) for which the block bootstrap is valid but the sieve bootstrap is not.

Assumption 7 For each value of ρ , the series u_t is strong mixing and satisfies the following conditions: $E[u_t] = 0$, $E|u_t|^r < \infty$ for some $r > 2$, $f_u(0) > 0$, where f_u denotes the spectral density of u_t , i.e., $f_u(\lambda) = \sum_{h=-\infty}^{\infty} \gamma_u(h) \exp(i\lambda h)$ and $\gamma_u(h) = E[u_t u_{t+h}]$. Furthermore, $\sum_{k=0}^{\infty} \alpha(k)^{1-2/r} < \infty$, where $\alpha(\cdot)$ denotes the strong mixing coefficient of u_t .

Their procedure can be described as follows.

Bootstrap Test 5 (Paparoditis and Politis, 2003)

1. Calculate the centred residuals

$$\tilde{u}_t = \hat{u}_t - \frac{1}{n-1} \sum_{j=2}^n \hat{u}_t = (y_t - \tilde{\rho} y_{t-1}) - \frac{1}{n-1} \sum_{j=2}^n (y_t - \tilde{\rho} y_{t-1})$$

where $\tilde{\rho}$ is a consistent estimator of ρ .

2. Choose the block length b , and draw points i_0, i_1, \dots, i_{k-1} , where $k = \lfloor (n-1)/b \rfloor$,³ from the uniform distribution on the set $\{1, 2, \dots, n-b\}$. These points will serve as the beginning points of the blocks of centred residuals:

$$y_t^* = \begin{cases} y_1 & \text{for } t = 1 \\ \hat{\alpha} + y_{t-1}^* + \tilde{u}_{i_m+s} & \text{for } t = 2, 3, \dots, n \end{cases} \quad (12)$$

where $m = \lfloor (t-2)/b \rfloor$, $s = t - mb - 1$, and $\hat{\alpha}$ is a drift parameter that is either set equal to zero or it is a consistent estimator of α .

3. From the bootstrap series y_t^* compute the desired statistics.

4. Repeat steps 2 to 3 B times to find the bootstrap distribution.

³The bootstrap sample y_t^* will have total length $l = kb + 1$.

Although most bootstrap unit root tests are based on differences, Paparoditis & Politis (2003) show that this may lead to a loss of power. We will give a short intuition of why this is so. This is adapted from Paparoditis & Politis (2001). For simplicity we neglect for the moment that the \hat{u}_t 's might have non-zero mean and that the residuals have to be recentred, and we set $\alpha = 0$. Then $\tilde{u}_t = y_t - \tilde{\rho}y_{t-1}$. As the goal of the bootstrap is to approximate the distribution, we want to generate samples that are $I(1)$, which is accomplished through the equation $y_t^* = y_{t-1}^* + \tilde{u}_{i_m+s}$, see (12). Combining these two equations yields

$$\begin{aligned} y_t^* &= y_{t-1}^* + y_{i_m+s} - \tilde{\rho}y_{i_m+s-1} \\ &= y_{t-1}^* + y_{i_m+s} - \rho y_{i_m+s-1} + \rho y_{i_m+s-1} - \tilde{\rho}y_{i_m+s-1} \\ &= y_{t-1}^* + u_{i_m+s} + (\rho - \tilde{\rho})y_{i_m+s-1}. \end{aligned} \tag{13}$$

As the bootstrap should generate unit root samples for all values of ρ , the term $(\rho - \tilde{\rho})y_{i_m+s-1}$ should be negligible for all ρ . Therefore, choosing $\tilde{\rho} = 1$ is not appropriate if in reality $\rho \neq 1$ and $\tilde{\rho}$ has to be a consistent estimator of ρ . Hence choosing $\tilde{\rho} = 1$ will lead to problems under the alternative. Paparoditis & Politis (2003) formally show that using differences instead of residuals leads to loss of power.

As said above, $\tilde{\rho}$ should be a consistent estimator of ρ . Furthermore, it is required that $\tilde{\rho}$ converges at rate $o_p(1)$ if $\rho \neq 1$, at rate $O_p(n^{-1})$ if $\rho = 1$ and $\alpha = 0$, and converges at rate $O_p(n^{-3/2})$ if $\rho = 1$ and $\alpha \neq 0$. Many estimators satisfy these conditions. Paparoditis & Politis (2003) focus on what they call the least squares (LS) estimator, which is just the DF estimator, and the ADF estimator (they call this the DF estimator). For the validity of the ADF estimator the additional condition iiB is needed to ensure invertibility.

They prove the consistency of the RBB for the DF coefficient test and the ADF coefficient test. For models where $\alpha = 0$, they recommend to use the OLS estimator of ρ in

$$y_t = \alpha + \rho y_{t-1} + u_t \tag{14}$$

or the ADF equivalent⁴ as $\tilde{\rho}$, which is used to construct the residuals. In the second step $\hat{\alpha}$ is set to zero, as there should be no drift. They also recommend for the RBB ADF test to use the block bootstrap⁵ of $y_t - y_{t-1}$ directly as lagged differences instead of $y_t^* - y_{t-1}^*$. For both the tests without deterministic components as the tests with a constant included, the consistency of the DF and ADF RBB tests is proved.

For the case of nonzero mean, $\alpha \neq 0$, Paparoditis & Politis (2003) recommend using the same estimator for $\tilde{\rho}$ as before but set $\hat{\alpha} = \tilde{\alpha}$ where $\tilde{\alpha}$ is the estimator of α in (14). They prove the consistency of the DF coefficient test with constant and trend and claim the consistency of the corresponding ADF test can be established similarly.

⁴Depending on which unit root test is performed.

⁵Using the same blocks as for the residuals.

2.3.2 Difference-based (A)DF block bootstrap test: Paparoditis & Politis (2003)

Similar to the previous two tests we also consider an alternative version of the tests by Paparoditis & Politis (2003). However, the situation here is the reverse of the two tests above, as the original tests are based on residuals, so the modified tests will be based on differences. The new procedure simply replaces $\tilde{\rho}$ by 1. Paparoditis & Politis (2003) already showed the asymptotic validity of these alternative tests.

2.4 DF stationary bootstrap test

2.4.1 Difference-based DF stationary bootstrap test: Swensen (2003a)

Swensen (2003a) considers a unit root test without deterministic components based on the stationary bootstrap of Politis & Romano (1994). He assumes the DGP

$$y_t = \rho y_{t-1} + u_t \tag{15}$$

with the following assumptions on u_t .

Assumption 8

(i) The process u_t is strictly stationary with $E[u_t] = 0$ for all t .

(ii) If $\gamma(k) = E[u_t u_{t+k}]$, then

$$\gamma_0 + \sum_{r=0}^{\infty} |r\gamma(r)| < \infty.$$

(iii) $\sum_{r,s,t} \kappa_4(r, s, t) = K < \infty$ where $\kappa_4(r, s, t)$ is the fourth cumulant of the distribution of $(u_j, u_{j+r}, u_{j+r+s}, u_{j+r+s+t})$.

Assumption (iii) is used to ensure that the variance of $\frac{1}{n} \sum_t u_t^2$ tends to zero and implies that σ^2 can be consistently estimated. Under these conditions Swensen (2003a) proves the consistency of the DF tests without deterministic components based on the stationary bootstrap. Note that the conditions needed are significantly weaker than those needed for the sieve bootstrap.

The algorithm can be described as follows:

Bootstrap Test 6 (Swensen, 2003)

1. Compute centred differences

$$\tilde{u}_t = \Delta y_t - (n-1)^{-1} \sum_{j=2}^n \Delta y_j.$$

2. Apply the stationary bootstrap of Politis & Romano (1994) to the centred residuals to obtain bootstrap errors u_t^* :

(a) Draw the index of the starting points of the blocks, i_1, i_2, \dots , from the uniform distribution $P(t_1 = t) = \frac{1}{n}, t = 1, \dots, n$. Let p_L be a fixed number between 0 and 1. Draw the length of the blocks b_1, b_2, \dots from the geometric distribution $P(b_1 = l) = (1 - p_L)^{l-1} p_L$. The expected block length is $1/p_L$.

(b) Form blocks using the drawn starting points and block lengths. For block $m + 1$ we have

$$u_t^* = \tilde{u}_{i_{m+1} + t - b(m) - 1} \quad (16)$$

where $t = b(m) + 1, \dots, b(m) + b_{m+1}$ and $b(m) = \sum_{j=1}^m b_j$.

(c) Stop after generating B blocks if $l_B = \sum_{j=1}^B b_j \geq n$. Lay the blocks end-to-end in the order sampled, and cut off the resulting series $u_1^*, \dots, u_{l_B}^*$ at u_n^* if $l_B > n$.

3. Construct the bootstrap sample y_t^* with the recursion $y_t^* = y_{t-1}^* + u_t^*$.

4. Compute the bootstrap DF coefficient and t -statistic.

5. Repeat steps 2 to 4 B times to find the bootstrap distribution.

2.4.2 Residual-based DF stationary bootstrap test: Parker, Paparoditis & Politis (2006)

Again, we also consider a modified version of these tests. Here we base the modified version on residuals instead of differences. Instead of the centred differences we calculate in step 1 centred residuals as in the bootstrap procedure of Paparoditis & Politis (2003). This test has recently been considered by Parker et al. (2006) and they also show the asymptotic validity of this test.

2.5 Comparison

As mentioned before, we will compare the tests on different aspects. We will mainly focus on three aspects: whether differences or residuals are used, the bootstrap method and the test statistic.

We have two tests that use the sieve bootstrap, and two that use (a form of) the block bootstrap. Theory for stationary series might lead us to expect a slightly better performance based on this aspect of the sieve bootstrap tests of Psaradakis (2001) and Chang & Park (2003).

Two tests use the DF statistic, one uses the ADF statistic and one uses both. Considering that the ADF test is asymptotically pivotal, a better performance of the Chang & Park (2003) test and the ADF test of Paparoditis & Politis (2003) based on this aspect seems credible.

Because we construct the modified tests, all tests are both based on residuals as well as differences. These modified tests allow us to look at each aspect separately. We have seen above that for the original tests differences are often combined with the sieve bootstrap. In fact, there is no sieve bootstrap test that is based on residuals. To be able to analyse the effects of the bootstrap method and the construction of the $I(0)$ series separately, we have to use the new tests.

The authors of the papers discussed above of course also performed simulations. Psaradakis (2001) performs some Monte Carlo simulations that show that for a model with a error term with a large negative MA term the bootstrap tests have better finite sample size than all asymptotic tests considered. In terms of power the bootstrap performs similar to the asymptotic tests. Monte Carlo simulations by Chang & Park (2003) show that bootstrap tests have better empirical size in finite samples than their asymptotic equivalent in the presence of a large negative MA-root. It is also shown that bootstrapping improves the coefficient test more than the t-test. The Monte Carlo simulation results reported by Paparoditis & Politis (2003) are qualitatively similar to those discussed earlier: the size of the tests is closer to the nominal level than the size of the asymptotic tests. The problems with the negative MA-root however cannot be solved with the RBB test. Since the RBB-ADF test appears to be less sensitive to the lag length p than for example the sieve bootstrap, increasing p does not improve performance much. They also compare the performance of the RBB test with a block bootstrap test based on differences and with a sieve bootstrap. From the simulations it seems that the power of the RBB test is higher than the power of both the block bootstrap test based on differences as the sieve bootstrap test, as well as higher than the power of the asymptotic tests. Swensen (2003a) also did simulations on the performance of his test. He found that in general the stationary bootstrap test performs well, although it cannot fully remove the size problems linked to a large negative MA-root. The simulations show that the stationary bootstrap is very sensitive to the choice of the expected block length. Furthermore, when the errors have AR components, the stationary bootstrap suffers from a lack of power regardless of the choice of the expected block length.

It is clear from these simulations that in finite samples the bootstrap tests perform better than the asymptotic tests in terms of size, and similar in terms of power. It is however not clear from these simulations, how the tests compare to each other.

Table 1 summarises all the test statistics and their main points. A note on the notation: we use τ for a coefficient test and t for a t-test. The first subscript indicates the bootstrap method, so S stands for sieve bootstrap, B for block bootstrap, and St for stationary bootstrap; the second subscript indicates whether a test is based on differences (d) or residuals (r). A superscript a states that the test is an augmented DF test.

INSERT TABLE 1 ABOUT HERE

3 Finite sample performance: Monte Carlo results

We will analyse the finite sample behaviour of the various tests, and focus in particular on the above mentioned points, by Monte Carlo simulations.

3.1 Monte Carlo setup

Let us first describe the DGPs that are used. We generate a series $y_t, t = 1, \dots, n$ according to the recursion

$$y_t = \rho y_{t-1} + u_t, \quad y_0 = 0, \quad (17)$$

where different values for ρ are used: 1, 0.99, 0.95, 0.9 and 0.8. Furthermore, we let u_t be generated by different processes. As sample sizes we consider $n = 50$, $n = 100$ and $n = 250$. We use three different significance levels: 0.01, 0.05 and 0.10. All experiments will be based on 5000 simulations and 999 bootstrap replications. All simulations are performed using GAUSS 6.0.

We use an ARMA(1,1) process to generate u_t :

$$u_t = \phi u_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}, \quad (18)$$

where $\varepsilon_t \sim IN(0, 1)$. The values used for ϕ and θ vary from -0.8 to 0.8.⁶

We perform two sets of simulations with these models. The first set considers the tests based on models without deterministic components. In the second set of simulation we will consider the tests with a constant and a trend (but not in the DGP). These extensions are not discussed in all papers, so that not all tests we consider have been shown to be theoretically valid. For the ADF test of Paparoditis & Politis (2003), we follow their (clear) instructions on how to handle the model with trend. Chang & Park (2003) on the other hand indicate that their tests can be applied in the model with trend by applying the bootstrap test to the detrended data. We hence follow their advise; we first detrend the series (by OLS), then we apply the bootstrap tests to this detrended series as if this was our original series in the first place. For the test proposed by Swensen (2003a), we add deterministic components in the same way as Psaradakis (2001) does this.

We select the lag length for the sieve bootstrap by *AIC*, so it is applied to differenced data. We estimate the AR(p) models by OLS. For the lag length in the ADF test of Paparoditis and Politis we use the modified AIC by Ng & Perron (2001). For the block length we choose fixed numbers: 5 for $n = 50$, 8 for $n = 100$ and 15 for $n = 250$. The fact that there is no easy way to estimate block lengths remains a problem.

The large number of DGP's and tests statistics in our simulations leads to a huge number of results that is rather hard to analyse in standard tables. We circumvent this problem by

⁶Specific values used for (ϕ, θ) are: (0, 0), (-0.8, 0), (-0.4, 0), (0.4, 0), (0.8, 0), (0, -0.8), (0, -0.4), (0, 0.4), (0, 0.8), (0.4, 0.4), (-0.4, -0.4).

estimating different response surfaces for the rejection frequencies observed in our simulations for each of the test statistics.

Because the empirical rejection frequency \hat{P} lies always between 0 and 1, we use the following transformation:

$$L(P) = \ln\left(\frac{P}{1-P}\right). \quad (19)$$

As dependent variable we use $L(\hat{P})$, as explanatory variables we consider several functions of the nominal level and the parameters in the underlying DGP. We will provide more details below. The specific form of the response surfaces is test specific. To avoid lengthy specification searches, we rely on *PcGets* (Hendry & Krolzig, 2001) to select the most appropriate specification using a large set of possible variables. The reported standard errors are White heteroscedasticity consistent standard errors. Apart from the coefficient estimates and their standard errors, the R^2 of the regression is also reported.

3.2 Results

3.2.1 The tests without deterministic components

Size Tables 2 and 3 give the parameter estimates and standard errors for the response surfaces for the size. We consider the following response surface for the size:

$$L(\hat{P}) = \beta_1 L(P_a) + \beta_2' \mathbf{f}(L(P_a), \phi, \theta, n) + \nu, \quad (20)$$

where P_a is the nominal size, or significance level, of the test and $\mathbf{f}(L(P_a), \phi, \theta, n)$ is a vector of functions of $L(P_a)$, the ARMA parameters ϕ and θ and the sample size n . These functions are stated in the leftmost column of the tables.

INSERT TABLE 2 ABOUT HERE

INSERT TABLE 3 ABOUT HERE

Several things can be seen from the response surfaces. The term $\beta_2' \mathbf{f}(\cdot)$ captures the deviations of the actual size from the nominal size as a function of the parameters of the DGP and sample size. $\beta_1 L(P_a)$ gives an indication of the asymptotic size of the tests. When β_1 is equal to 1, the empirical size of the test is equal to the nominal size for large n . We see that for some tests β_1 is significantly different from 1, although for most it is close to it. Since the terms in $\beta_2' \mathbf{f}(\cdot)$ are either of order $O(n^{-1/2})$ or $O(n^{-1})$, this means that for most tests the size distortions decrease when n increases. Results for the tests by Psaradakis (2001), the tests by Swensen (2003a) and, to a lesser degree, the DF test by Paparoditis & Politis (2003) and the tests by Parker et al. (2006) are somewhat worrying. Note that these are all DF tests, as opposed to the ADF tests for which β_1 is much closer to 1. For the block tests, the value

for β_1 is higher for the difference-based version than the residual-based version, which may indicate that in general the residual-based tests give slightly higher rejection frequencies than difference-based tests, which is something we also see in the “raw” results of the simulations.

Next, we see that the response surfaces for block tests in general require more parameters than for sieve tests. Specifically, we see that functions of ϕ play a larger role for block tests than for sieve tests. As ϕ is the autoregressive parameter, this is not so strange. Because the sieve bootstrap filters out the dependence using an $\text{AR}(p)$ model, it filters ϕ out directly, therefore it will not have much influence on the size. Basically we will be resampling an approximately iid series ε_t (if there is no MA dependence). In the block bootstrap however, there is no filtering out of the dependence, so it is more likely that ϕ will play a role here.

We do not only see this difference when we compare block bootstrap tests to sieve bootstrap tests, but also when we compare residual-based tests to difference-based tests. The AR parameter ϕ seems to be more important for the size of residual-based tests than of difference-based tests. Again this may be understood by noting that the AR parameter has an influence through the estimator of ρ when estimating residuals, whereas it does not when using differences. As said before, the difference-based test might perform somewhat better under the null because we impose it.

Another point we notice is the fact that the MA parameter θ plays a role for all tests. This is due to the size distortions that all tests exhibit when $\theta = -0.8$, a well-known fact for all unit root tests. The size distortions are not the same for all tests, though. The sieve tests in general have lower size distortions for finite samples than the block tests, which is reflected in the response surfaces by the fact that the coefficients for functions of θ in general are lower in absolute value for sieve tests than for block tests. Note that of the block tests, the size of the ADF test by Paparoditis & Politis (2003) ($\tau_{B,r}^a$) is the least affected by θ . Again this shows that the sieve bootstrap can also filter out a large part of the dependence created through θ , which might lead to its better performance. Similarly, the lags in the ADF test also remove most of the dynamics created through θ . This result is slightly overstated as we fix the block length whereas we select the lag length using information criteria. Therefore the lag length can adapt to the large dependence in the errors. However, simulations not reported here show that although increasing the block length will somewhat reduce the problem, the size distortions remain worse than for the sieve bootstrap. Besides, as there currently exist no methods to select the optimal block length we have no other option than to fix the block length.

Again we also see a difference between residual- and difference-based tests. Residual-based tests have higher size distortions than the same tests based on differences, as reflected through the (absolute) larger coefficients in the response surfaces.

As an illustration we provide graphical representations of all the coefficient tests for two specific DGPs: the pure random walk ($\phi = 0, \theta = 0$) and an MA-process with a large negative MA-root ($\phi = 0, \theta = -0.8$), for which unit root tests are known to perform badly. We use PP and PP discrepancy plots: The PP plot graphs the empirical rejection frequency \hat{P} against the nominal level P_a , whereas the PP discrepancy plot graphs the deviation of the empirical rejection frequency from the nominal level against the nominal level (Davidson & MacKinnon, 1998). Hence ideally the graph of a test would lie on the 45° line in the PP plot and on a horizontal line at 0 in the PP discrepancy plot.

Figures 1 and 2 present, respectively, the PP plots and PP discrepancy plots of both cases. We see that in the first case ($\phi = 0$ and $\theta = 0$) in the PP plot all tests are very close to the 45° line, and although the PP discrepancy plot shows deviations from 0, these deviations are quite small in magnitude. We can conclude from these figures that all tests perform well in terms of size for this DGP.

Both the PP plot and the PP discrepancy plot of the second case ($\phi = 0$ and $\theta = -0.8$) contrast heavily with the first case. Both graphs clearly show that there are large size distortions for all tests, although some are worse than others. The size of the $\tau_{B,r}$ and $\tau_{St,r}$ test is particularly dramatic. Even for a nominal level of 0.05, the empirical size already exceeds 0.9. In general the difference-based tests have lower size distortions than their residual-based counterparts. The sieve tests (based on both differences as residuals) perform best in this case, together with the $\tau_{B,r}^a$ test, even though this test is somewhat undersized for high nominal levels (which are not interesting in practice). The $\tau_{B,d}^a$ test is undersized for all sizes, which sharply contrasts with the other tests.

INSERT FIGURE 1 ABOUT HERE

INSERT FIGURE 2 ABOUT HERE

Power Tables 4 and 5 give the response surfaces for the power. We choose to report only unadjusted power as we feel this is the most relevant, because this is what matters in practice. We now estimate the following response surface:

$$L(\hat{P}) = \beta_0 + \beta_1 L(P_a) + \beta_2' \mathbf{f}(\rho, L(P_a), \phi, \theta, n) + \beta_3(\rho - 1) + \nu, \quad (21)$$

where again all variables in $\mathbf{f}(\rho, L(P_a), \phi, \theta, n)$ are either of order $O(n^{-1/2})$ or $O(n^{-1})$. So in this case the asymptotic behaviour can be deduced from $\beta^a = (\beta_0, \beta_1, \beta_3)'$. As we do not have an asymptotic benchmark as explanatory variable, it is difficult to analyse the coefficients β^a like we did for the size.

INSERT TABLE 4 ABOUT HERE

INSERT TABLE 5 ABOUT HERE

Instead, we calculate the asymptotic power implied by β^a for different combinations of ρ and the significance level. The results can be found in Table 6. Although for most combinations the asymptotic power is 1 (as it should be for a consistent test), there are still some clear differences. Except for the Chang & Park (2003) tests, the residual-based tests have higher power than the difference-based tests. It also seems that the block tests in general have somewhat higher power than the sieve tests, although here the higher power is often combined with higher size distortions.

INSERT TABLE 6 ABOUT HERE

By far the worst test in terms of power is the ADF test based on differences by Paparoditis & Politis (2003). This result corresponds to what Paparoditis & Politis (2003) show in their paper. The Psaradakis (2001) tests also seem to suffer from a lack of power as compared to the other tests. The Swensen (2003a) tests and the residual-based DF test by Paparoditis & Politis (2003) seem to have highest power, but they also have the highest size distortions, so this is not necessarily a positive finding. When we compare the (difference-based) Chang & Park (2003) test to the (residual-based) ADF test of Paparoditis & Politis (2003), whose size behaviour is quite similar and therefore allows a direct comparison, we see that the Paparoditis and Politis test has higher power, which confirms the idea that residual-based tests and block tests have higher power than difference-based tests and sieve tests.

The differences in the coefficients for the functions of the ARMA parameters between sieve and block tests, and difference-based and residual-based tests, are less pronounced here. It seems that the ARMA parameters of the DGP in general have more impact on the power of the tests than on the size.

Concluding, it seems the tests by Chang & Park (2003) and the ADF test by Paparoditis & Politis (2003) perform best. The Chang and Park tests have advantage of depending less on the AR coefficient, and the Paparoditis and Politis test has the advantage of having higher power.

3.2.2 The tests with intercept and linear trend

The simulations for the tests with trends give in general very similar results to the ones discussed above. Again the size of the sieve tests and difference-based tests is more robust to the ARMA parameters than the size of the block tests and residual-based tests, respectively.

The power of the tests is in general lower. Similar to the case without deterministic components, the higher power of the residual-based tests is again confirmed. This is again in line with what Paparoditis & Politis (2005) claim. The (slight) advantage of the block tests is not apparent anymore, therefore we cannot conclude in general that block tests have a higher power than sieve tests. The coefficient test by Psaradakis (2001) seems to perform better relative to the previous section. In general we also see a lack of power for the t-tests.

One remarkable thing we notice is that for the Chang & Park (2003) tests we have to include a positive intercept, meaning that these tests have a certain size independent of the nominal level, which obviously makes the tests invalid. This shows that the handling of deterministic components is not as straightforward as some authors, including Chang and Park, claim it to be.

We can explain this as follows. Chang & Park (2003) claim in their paper that their tests can be applied to demeaned and detrended series, when deterministic components are present. This is exactly what we do: we first detrend the data, and then apply exactly the same test as before to this detrended data. The problem here is that in the bootstrap procedure, the deterministic trends are not taken into account anymore, as unit root tests without deterministic trends are performed on the resampled series. Therefore the bootstrap distribution will mimic the Dickey-Fuller distribution for tests without deterministic trends instead of the distribution for the test with a constant and a trend, which is in this case the relevant distribution. One way to get the correct bootstrap distribution would be to detrend the resampled series as well before performing the test on the resampled series.

So far we have only considered a linear DGP for the errors. As the sieve bootstrap is invalid for nonlinear models, but the block bootstrap is not, provided mixing conditions are satisfied, this would be a very interesting extension to consider. We have indeed performed some simulations where we use a nonlinear DGP for the innovations instead of an ARMA process. Specifically, we use a Markov Switching model for the errors. Our results indicate that the difference between the sieve bootstrap and the block bootstrap is rather small for such models. Therefore, we might expect the sieve bootstrap to work reasonably well also for nonlinear models, in particular for nonlinear processes which have a Wold representation (which can be approximated by an ARMA model).

4 Conclusion

We have shown that the sieve bootstrap procedures proposed by Psaradakis (2001) and Chang & Park (2003) remain asymptotically valid if they are based on residuals instead of differences. With these results we have completed the set of bootstrap Dickey-Fuller tests.

In finite samples, for the tests without deterministic trends, we can draw several conclusions. First, ADF tests clearly perform better than DF tests, which is what we expected from our discussion about asymptotically pivotal statistics. Second, it seems that sieve tests perform slightly better in terms of size than block tests for ARMA models, which is in line with the results for stationary series. We also see that the stationary bootstrap test performs somewhat worse in terms of size than the overlapping blocks bootstrap. In the case of a nonlinear generating process for the innovations, we do not really see the expected superior

performance of the block bootstrap.

For the applied researcher there is also a practical reason to use the sieve bootstrap. The selection of the lag length can be done quite easily, and appears to work if based on an information criterium. On the other hand, choosing the block length on the basis of intuition is difficult, and there exist no satisfactory methods for it. Hence this means when one wishes to apply the block bootstrap to test for unit roots, one has to use some trial-and-error methods; choosing different block lengths and then regarding how much the results change. Taking all this into account, the sieve bootstrap has a slight advantage over the block bootstrap.

The choice between difference-based tests and residual-based tests is not obvious. While the residual-based tests have higher power than the difference-based tests, these tests also have higher size distortions, although for the residual-based ADF tests, the size distortions are less serious than for residual-based DF tests. Hence, in combination with an ADF test, using residuals is preferable to using differences.

Summarising, the main point is to use an ADF test instead of a DF test. When we use an ADF test, the differences between sieve and block tests and between difference-based and residual-based tests become small, although residual-based tests still have higher power. The Chang & Park (2003) tests, our ADF sieve tests based on residuals and the ADF test of Paparoditis & Politis (2003) all perform well. When we investigate the behaviour of the test statistics with deterministic components (for DGPs without deterministic trends), we see that both size and power is lower in general, but the main conclusions still hold.

References

- Basawa, I. V., Mallik, A. K., McCormick, W. P., Reeves, J. H., & Taylor, R. L. (1991). Bootstrapping unstable first-order autoregressive processes. *Annals of Statistics*, *19*, 1098–1101.
- Bühlmann, P. (1995). Moving-average representations of autoregressive approximations. *Stochastic Processes and their Applications*, *60*, 331–342.
- Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli*, *3*, 123–148.
- Chang, Y. & Park, J. Y. (2002). On the asymptotics of ADF tests for unit roots. *Econometric Reviews*, *21*, 431–447.
- Chang, Y. & Park, J. Y. (2003). A sieve bootstrap for the test of a unit root. *Journal of Time Series Analysis*, *24*, 379–400.
- Davidson, R. & MacKinnon, J. G. (1998). Graphical methods for investigating the size and power of test statistics. *The Manchester School*, *66*, 1–26.

- Dickey, D. A. & Fuller, W. A. (1979). Distribution of estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*, 427–431.
- Dickey, D. A. & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, *49*, 1057–1072.
- Hendry, D. F. & Krolzig, H.-M. (2001). *Automatic Econometric Model Selection*. London: Timberlake Consultants Press.
- Horowitz, J. L. (2001). The bootstrap. In J. J. Heckman & E. E. Leamer (Eds.), *Handbook of Econometrics*, volume 5 chapter 52, (pp. 3159–3228). Amsterdam: North Holland Publishing.
- Ng, S. & Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica*, *69*, 1519–1554.
- Paparoditis, E. & Politis, D. N. (2001). Unit root testing via the continuous-path block bootstrap. Discussion Paper 2001-06, Dept. of Economics, University of California.
- Paparoditis, E. & Politis, D. N. (2003). Residual-based block bootstrap for unit root testing. *Econometrica*, *71*, 813–855.
- Paparoditis, E. & Politis, D. N. (2005). Bootstrapping unit root tests for autoregressive time series. *Journal of the American Statistical Association*, *100*, 545–553.
- Park, J. Y. (2002). An invariance principle for sieve bootstrap in time series. *Econometric Theory*, *18*, 469–490.
- Parker, C., Paparoditis, E., & Politis, D. N. (2006). Unit root testing via the stationary bootstrap. *Journal of Econometrics*. forthcoming.
- Phillips, P. C. B. & Solo, V. (1992). Asymptotics for linear processes. *Annals of Statistics*, *20*, 971–1001.
- Politis, D. N. & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, *89*, 1303–1313.
- Psaradakis, Z. (2001). Bootstrap tests for an autoregressive unit root in the presence of weakly dependent errors. *Journal of Time Series Analysis*, *22*, 577–594.
- Swensen, A. R. (2003a). Bootstrapping unit root tests for integrated processes. *Journal of Time Series Analysis*, *24*, 99–126.
- Swensen, A. R. (2003b). A note on the power of bootstrap unit root tests. *Econometric Theory*, *19*, 32–48.

A Proofs

As the tests we have proposed are very similar to the existing tests of Psaradakis (2001) and Chang & Park (2003) we do not offer a complete proof of the validity of our tests, but instead we adapt the already existing proofs of Psaradakis (2001) and Chang & Park (2003) to our tests. Therefore in the following sections we will only present those elements of the proofs that differ from those for the tests by Psaradakis (2001) and Chang & Park (2003). These differences result from the fact that we use residuals where they use differences. To be specific, the residuals that are resampled in our tests are constructed as

$$\hat{\varepsilon}_t = y_t - \hat{\rho}y_{t-1} - \sum_{j=1}^p \hat{\phi}_j \Delta y_{t-j}, \quad (22)$$

where $\hat{\rho}, \hat{\phi}_1, \dots, \hat{\phi}_p$ are the OLS estimates from the (augmented Dickey-Fuller) regression of y_t on $y_{t-1}, \Delta y_{t-1}, \dots, \Delta y_{t-p}$.

The residuals that are resampled in the tests of Psaradakis (2001) and Chang & Park (2003) are constructed as

$$\tilde{\varepsilon}_t = \Delta y_t - \sum_{j=1}^p \tilde{\phi}_j \Delta y_{t-j}, \quad (23)$$

where $\tilde{\phi}_1, \dots, \tilde{\phi}_p$ are the OLS (or Yule-Walker) estimates from the regression of Δy_t on $\Delta y_{t-1}, \dots, \Delta y_{t-p}$.

One effect of not imposing the unit root restriction is that the estimation of ρ means that we are only able to obtain some results in probability that can be obtained a.s. when imposing the unit root.

A.1 Proof of Theorem 1

A small difference between our tests and the tests by Psaradakis (2001) is that we use OLS instead of Yule-Walker. However, as both methods are asymptotically equivalent, the results for Psaradakis also hold for OLS. We first have to prove Lemma 2 of Psaradakis (2001); the version for our test with residuals is stated below.

Lemma 1 *Suppose assumptions 1 and 2 hold. Let $\hat{\phi}_{j,p}$ be Yule-Walker estimates. Then as $n \rightarrow \infty$:*

- (a) *there exists a random variable n_0 such that $\sup_{n \geq n_0} \sum_{j=0}^{\infty} j |\hat{\psi}_{j,n}| < \infty$ in probability;*
- (b) $\sup_{0 \leq j \leq \infty} |\hat{\psi}_{j,n} - \psi_j| = o_p(1)$;
- (c) $\text{Var}^*[\varepsilon_t^*] - \sigma_\varepsilon^2 = o_p(1)$
- (d) $\text{Var}^*[n^{-1/2} \sum_{t=1}^n \varepsilon_t^*] - \sigma^2 = o_p(1)$.

As mentioned above, we can only show parts (a) and (b) in probability, while Psaradakis (2001) has these almost sure. This will not affect the end result as only convergence in probability is needed.

Proof of (a) Psaradakis (2001) refers to Bühlmann (1995) for his proof. Following Bühlmann (1995), we have to show that

$$\sum_{j=0}^{\infty} j |\hat{\phi}_{j,n} - \phi_j| = o_p(1). \quad (24)$$

Let $\tilde{\phi}_{j,n}$ be the OLS-estimates from the regression

$$\Delta y_t = \sum_{j=1}^p \phi_{j,n} \Delta y_{t-j} + \varepsilon_t. \quad (25)$$

We have that $\hat{\phi}_{j,n} = \tilde{\phi}_{j,n} + O_p(n^{-1})$ from Chang & Park (2002).

We can write

$$\begin{aligned} \sum_{j=0}^{\infty} j |\hat{\phi}_{j,n} - \phi_j| &= \sum_{j=0}^{\infty} j |\hat{\phi}_{j,n} - \tilde{\phi}_{j,n} + \tilde{\phi}_{j,n} - \phi_j| \\ &\leq \sum_{j=0}^{\infty} j |\hat{\phi}_{j,n} - \tilde{\phi}_{j,n}| + \sum_{j=0}^{\infty} j |\tilde{\phi}_{j,n} - \phi_j|. \end{aligned} \quad (26)$$

Bühlmann (1995) shows that $\sum_{j=0}^{\infty} j |\tilde{\phi}_{j,n} - \phi_j| = o(1)$ a.s. and furthermore we have that

$$\sum_{j=0}^{\infty} j |\hat{\phi}_{j,n} - \tilde{\phi}_{j,n}| \leq p^2 \max_{1 \leq j \leq p} |\hat{\phi}_{j,n} - \tilde{\phi}_{j,n}| = o_p(1). \quad (27)$$

Hence, $\sum_{j=0}^{\infty} j |\hat{\phi}_{j,n} - \phi_j| = o_p(1)$ and the proof of part (a) is completed.

Proof of (b) Bühlmann (1995) (Lemma 3.2) proves this for the Yule-Walker estimates, but, as OLS is asymptotically equivalent to Yule-Walker, it holds for our tests as well.

Proof of (c) and (d) For part (c) and (d) we need that

$$\mathbb{E}^*[(\varepsilon_t^*)^{2w}] = \mathbb{E}[(\varepsilon_t)^{2w}] + o_p(1). \quad (28)$$

Following Bühlmann (1997)

$$\mathbb{E}^*[(\varepsilon_t^*)^{2w}] = (n-p)^{-1} \sum_{t=p+1}^n (\hat{\varepsilon}_{t,n} - \hat{\varepsilon}_n^{(\cdot)})^{2w}, \quad (29)$$

where $\hat{\varepsilon}_n^{(\cdot)} = (n-p)^{-1} \sum_{t=p+1}^n \hat{\varepsilon}_{t,n}$.

We first look at $\hat{\varepsilon}_n^{(\cdot)}$ and prove that it is $o_p(1)$. Note that

$$\varepsilon_t = \Delta y_t - \sum_{j=1}^{\infty} \phi_j \Delta y_{t-j} \quad (30a)$$

$$\hat{\varepsilon}_{t,n} = y_t - \hat{\rho} y_{t-1} - \sum_{j=1}^p \hat{\phi}_j \Delta y_{t-j}. \quad (30b)$$

Then we write

$$\begin{aligned} \hat{\varepsilon}_n^{(\cdot)} &= (n-p)^{-1} \sum_{t=p+1}^n (\varepsilon_t - \varepsilon_t + \hat{\varepsilon}_{t,n}) \\ &= (n-p)^{-1} \sum_{t=p+1}^n \left[\varepsilon_t - (\Delta y_t - \sum_{j=1}^{\infty} \phi_j \Delta y_{t-j}) + (y_t - \hat{\rho} y_{t-1} - \sum_{j=1}^p \hat{\phi}_{j,n} \Delta y_{t-j}) \right] \\ &= (n-p)^{-1} \sum_{t=p+1}^n \left[\varepsilon_t - (\Delta y_t - (y_t - \hat{\rho} y_{t-1})) \right. \\ &\quad \left. - \sum_{j=1}^p (\hat{\phi}_{j,n} - \phi_j) \Delta y_{t-j} + \sum_{j=p+1}^{\infty} \phi_j \Delta y_{t-j} \right] \\ &= (n-p)^{-1} \sum_{t=p+1}^n (A_t + B_t + C_t + D_t) \end{aligned} \quad (31)$$

Hence it has to be shown that $(n-p)^{-1} \sum_{t=p+1}^n A_t$, $(n-p)^{-1} \sum_{t=p+1}^n B_t$, $(n-p)^{-1} \sum_{t=p+1}^n C_t$, $(n-p)^{-1} \sum_{t=p+1}^n D_t = o_p(1)$.

It is trivial that $(n-p)^{-1} \sum_{t=p+1}^n A_t$ and $(n-p)^{-1} \sum_{t=p+1}^n D_t$ are $o_p(1)$. Next we turn to B_t :

$$\begin{aligned} B_t &= -\Delta y_t + y_t - \hat{\rho} y_{t-1} \\ &= (1 - \hat{\rho}) y_{t-1} \end{aligned} \quad (32)$$

As $1 - \hat{\rho} = O_p(n^{-1})$ (Chang & Park, 2002), we have that

$$(n-p)^{-1} \sum_{t=p+1}^n (1 - \hat{\rho}) y_{t-1} = (1 - \hat{\rho}) (n-p)^{-1} \sum_{t=p+1}^n y_{t-1} = o_p(1). \quad (33)$$

Finally, we look at C_t . By the Cauchy-Schwartz inequality,

$$\begin{aligned} & (n-p)^{-1} \sum_{t=p+1}^n \sum_{j=1}^p (\hat{\phi}_{j,n} - \phi_j) \Delta y_{t-j} \\ & \leq \left[\sum_{j=1}^p (\hat{\phi}_{j,n} - \phi_j)^2 \right]^{1/2} \left[(n-p)^{-1} \sum_{t=p+1}^n \sum_{j=1}^p (\Delta y_{t-j})^2 \right]^{1/2}. \end{aligned} \quad (34)$$

Then, using that $\hat{\phi}_{j,n} - \phi_j = O_p((\ln n/n)^{1/2}) + o(p^{-1})$ and $p(n) = o((n/\ln n)^{1/2})$, we have that

$$\begin{aligned} \left[\sum_{j=1}^p (\hat{\phi}_{j,n} - \phi_j)^2 \right]^{1/2} &= \left[\sum_{j=1}^p \left\{ O_p((\ln n/n)^{1/2}) + o(p^{-1}) \right\}^2 \right]^{1/2} \\ &= \left[\sum_{j=1}^p \left\{ O_p((\ln n/n)^{1/4}) + o_p(p^{-1}(\ln n/n)^{1/2}) + o(p^{-2}) \right\}^2 \right]^{1/2} \\ &= \left[O_p(p(\ln n/n)^{1/4}) + o_p((\ln n/n)^{1/2}) + o(p^{-1}) \right]^{1/2} \\ &= \left[o_p((\ln n/n)^{1/2}) + o(p^{-1}) \right]^{1/2} \end{aligned} \quad (35)$$

Therefore,

$$(n-p)^{-1} \sum_{t=p+1}^n C_t = \left[o_p((\ln n/n)^{1/2}) + o(p^{-1}) \right]^{1/2} O_p(p^{1/2}) = o_p(1) \quad (36)$$

and the proof is complete.

After having proven that $\hat{\varepsilon}_n^{(\cdot)}$ is $o_p(1)$, we turn to $(n-p)^{-1} \sum_{t=p+1}^n (\hat{\varepsilon}_{t,n})^{2w}$ and show that it is equal to $E[(\varepsilon_t)^{2w}] + o_p(1)$.

Following Bühlmann (1997), we now have to show that

$$(n-p)^{-1} \sum_{t=p+1}^n |B_t|^{2w} = o_p(1), \quad (37a)$$

$$(n-p)^{-1} \sum_{t=p+1}^n |C_t|^{2w} = o_p(1), \quad (37b)$$

$$(n-p)^{-1} \sum_{t=p+1}^n |D_t|^{2w} = o_p(1). \quad (37c)$$

These proofs follow straightforwardly by adapting the proofs used in the previous part. Hence we have shown that $(n-p)^{-1} \sum_{t=p+1}^n (\hat{\varepsilon}_{t,n})^{2w} = E[(\varepsilon_t)^{2w}] + o_p(1)$. Now the result in (28) follows from the proof of Bühlmann (1997).

The proof of Lemma 1 is the main element in the proof that is different from Psaradakis (2001) if we use residuals instead of differences. The rest of the proof of Theorem 1 therefore simply follows from the proof in Psaradakis (2001) by setting the deterministic components to zero.

A.2 Proof of Theorem 2

Next we prove the validity of our sieve bootstrap ADF test based on residuals. As this test is very similar to the test of Chang & Park (2003), we follow the same strategy as in the previous section and adapt the proofs from Chang & Park (2003) when necessary.

Chang & Park (2003) rely on the results in Park (2002), hence we first focus on this paper. Note that all definitions are the same as before, unless mentioned.

As we mentioned above, we cannot prove all the results from Park (2002) a.s. for our case, but only in probability, due to the use of residuals and the corresponding estimation of ρ that is involved.

We first give our equivalent of Lemma 3.1 in Park (2002):

Lemma 2 *Let assumption 3 hold and let $p(n) = o((n/\ln n)^{1/2})$. Then it follows that*

$$\max_{1 \leq j \leq p} |\hat{\phi}_{j,n} - \phi_j| = O_p((\ln n/n)^{1/2}) + o(p^{-s})$$

for large n . Moreover, we have

$$\begin{aligned} \hat{\sigma}_n^2 &= \sigma^2 + O_p((\ln n/n)^{1/2}) + o(p^{-s}) \\ \sum_{j=1}^p \hat{\phi}_{j,n} &= \sum_{j=1}^{\infty} \phi_j + O_p(p(\ln n/n)^{1/2}) + o(p^{-s}) \end{aligned}$$

as $n \rightarrow \infty$.

Proof The first and the third statement of Lemma 2 follow directly from Lemma 3.5 of Chang & Park (2002). The second statement follows from Bühlmann (1995), the proof of theorem 3.2.

Next we will give our equivalent of Lemma 3.2 in Park (2002).

Lemma 3 *Let assumption 3 hold and let $p(n) = o((n/\ln n)^{1/2})$. Then $n^{1-r/2} \mathbf{E}^* |\varepsilon_t^*|^r \xrightarrow{p} 0$ and*

$$W_n^*(i) = \frac{1}{\hat{\sigma}_n \sqrt{n}} \sum_{k=1}^{[ni]} \varepsilon_k^* \xrightarrow{d^*} W$$

as $n \rightarrow \infty$.

Proof To show Lemma 3 we need to show that

$$n^{1-r/2} \mathbf{E}^* |\varepsilon_t^*|^r = n^{1-r/2} \left(\frac{1}{n} \sum_{t=1}^n \left| \hat{\varepsilon}_{t,n} - \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_{t,n} \right|^r \right) \xrightarrow{p} 0. \quad (38)$$

Following Park (2002), we have that

$$\frac{1}{n} \sum_{t=1}^n \left| \hat{\varepsilon}_{t,n} - \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_{t,n} \right|^r \leq c(A_n + B_n + C_n + D_n) \quad (39)$$

where

$$\begin{aligned} A_n &= \frac{1}{n} \sum_{t=1}^n |\varepsilon_t|^r, \\ B_n &= \frac{1}{n} \sum_{t=1}^n |\varepsilon_{t,n} - \varepsilon_t|^r, \\ C_n &= \frac{1}{n} \sum_{t=1}^n |\hat{\varepsilon}_{t,n} - \varepsilon_{t,n}|^r, \\ D_n &= \left| \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_{t,n} \right|^r. \end{aligned} \quad (40)$$

Hence we have to show that $n^{1-r/2} A_n \xrightarrow{p} 0$ and similarly for the other parts. Note that the only term that is different from the definition in Park (2002) because our test is based on residuals, is $\hat{\varepsilon}_{t,n}$. As A_n and B_n do not involve any $\hat{\varepsilon}_{t,n}$ terms we can just follow his proof in this.

Next we turn to C_n . We write

$$\begin{aligned} \hat{\varepsilon}_{t,n} &= y_t - \hat{\rho} y_{t-1} - \sum_{j=1}^p \hat{\phi}_{j,n} \Delta y_{t-j} \\ &= (y_t - \hat{\rho} y_{t-1} - \Delta y_t) + (\Delta y_t - \sum_{j=1}^p \hat{\phi}_{j,n} \Delta y_{t-j}) \\ &= -(\hat{\rho} - 1) y_{t-1} + (\varepsilon_{t,n} - \sum_{j=1}^p (\hat{\phi}_{j,n} - \phi_{j,n}) \Delta y_{t-j} - \sum_{j=1}^p (\phi_{j,n} - \phi_j) \Delta y_{t-j}) \end{aligned} \quad (41)$$

where the last line follows from Park (2002). Then it follows that

$$|\hat{\varepsilon}_{t,n} - \varepsilon_{t,n}|^r \leq c \left(|(\hat{\rho} - 1) y_{t-1}|^r + \left| \sum_{j=1}^p (\hat{\phi}_{j,n} - \phi_{j,n}) \Delta y_{t-j} \right|^r + \left| \sum_{j=1}^p (\phi_{j,n} - \phi_j) \Delta y_{t-j} \right|^r \right). \quad (42)$$

where $c = 3^{r-1}$. Then we define

$$\begin{aligned}
C_{0n} &= \frac{1}{n} \sum_{t=1}^n |(\hat{\rho} - 1)y_{t-1}|^r \\
C_{1n} &= \frac{1}{n} \sum_{t=1}^n \left| \sum_{j=1}^p (\hat{\phi}_{j,n} - \phi_{j,n}) \Delta y_{t-j} \right|^r \\
C_{2n} &= \frac{1}{n} \sum_{t=1}^n \left| \sum_{j=1}^p (\phi_{j,n} - \phi_j) \Delta y_{t-j} \right|^r
\end{aligned} \tag{43}$$

and hence it has to be shown that $n^{1-r/2}C_{in} \xrightarrow{a.s.} 0$ for $i = 0, 1, 2$. Again, we follow the proof of Lemma 3.2 in Park (2002) to show this for C_{1n} and C_{2n} . Our C_{2n} is exactly the same as C_{2n} in Park (2002) so this result follows straightforwardly. C_{1n} is majorised by

$$\begin{aligned}
& \left(\max_{1 \leq j \leq p} |\hat{\phi}_{j,n} - \phi_{j,n}|^r \right) \frac{1}{n} \sum_{t=1}^n \sum_{j=1}^p |\Delta y_{t-j}|^r \\
& \leq \left(\max_{1 \leq j \leq p} |\hat{\phi}_{j,n} - \phi_{j,n}|^r \right) \frac{p}{n} \left(\sum_{t=0}^{n-1} |\Delta y_t|^r + \sum_{t=-1}^{1-p} |\Delta y_t|^r \right) \\
& \leq c \left(\max_{1 \leq j \leq p} (|\tilde{\phi}_{j,n} - \phi_{j,n}|^r + |\hat{\phi}_{j,n} - \tilde{\phi}_{j,n}|^r) \right) \frac{p}{n} \left(\sum_{t=0}^{n-1} |\Delta y_t|^r + \sum_{t=-1}^{1-p} |\Delta y_t|^r \right) \\
& = [O((\ln n/n)^r) + O_p(n^{-r})] (p/n)O(n) = O(p(\ln n/n)^r) \\
& = o_p((\ln n/n)^{r-1/2}).
\end{aligned} \tag{44}$$

Hence, $C_{1n} \xrightarrow{p} 0$.

Next we consider C_{0n} . Note that we can write this as

$$\frac{1}{n} \sum_{t=1}^n |(\hat{\rho} - 1)y_{t-1}|^r = |\hat{\rho} - 1|^r \frac{1}{n} \sum_{t=1}^n |y_{t-1}|^r = o_p(1). \tag{45}$$

This proves that $n^{1-r/2}C_n \xrightarrow{p} 0$.

For D_n we have to prove that

$$\frac{1}{n} \sum_{t=1}^n \sum_{j=p+1}^{\infty} \phi_j \Delta y_{t-j} \xrightarrow{p} 0, \quad (46)$$

$$\frac{1}{n} \sum_{t=1}^n \sum_{j=p+1}^{\infty} (\phi_{j,n} - \phi_j) \Delta y_{t-j} \xrightarrow{p} 0, \quad (47)$$

$$\frac{1}{n} \sum_{t=1}^n \sum_{j=p+1}^{\infty} (\hat{\phi}_{j,n} - \phi_{j,n}) \Delta y_{t-j} \xrightarrow{p} 0, \quad (48)$$

$$\frac{1}{n} \sum_{t=1}^p (1 - \hat{\rho}) y_{t-1} \xrightarrow{p} 0, \quad (49)$$

where (46) and (47) just follow from Park (2002). For (48) we define

$$N_n = \sum_{j=1}^p (\hat{\phi}_{j,n} - \phi_{j,n}) \sum_{t=1}^n \Delta y_{t-j} \quad (50)$$

and

$$Q_n = \sum_{j=1}^p \left| \sum_{t=1}^n \Delta y_{t-j} \right|. \quad (51)$$

Then N_n is dominated by

$$Q_n \max_{1 \leq j \leq p} |\hat{\phi}_{j,n} - \phi_{j,n}|. \quad (52)$$

Park (2002) shows that $Q_n = o(pn^{1/2}(\ln n)^{1/r}(\ln \ln n)^{(1+\delta)/r})$ a.s. for any $\delta > 0$.

Furthermore,

$$\max_{1 \leq j \leq p} |\hat{\phi}_{j,n} - \phi_{j,n}| \leq \max_{1 \leq j \leq p} (|\tilde{\phi}_{j,n} - \phi_{j,n}| + |\hat{\phi}_{j,n} - \tilde{\phi}_{j,n}|) = O\left((\ln n/n)^{1/2}\right) \text{ a.s.} + O_p(n^{-1}). \quad (53)$$

Then following Park (2002) we find that $N_n = o_p(n)$, which proves the result.

We have seen (49) several times before, and it is easy to see that this is true. This completes the proof of D_n and hence of Lemma 3.

Our versions of Theorem 3.3 and Lemma 4.1 of Park (2002) can easily be seen to hold given that we have shown before that $\sum_{j=0}^{\infty} |\hat{\psi}_{j,n}| < \infty$.

The proof of Theorem 2 then simply follows by straightforward adaptation of the proofs of Chang & Park (2003) for their Theorem 2, using the lemmas above.

Table 1: Main features of the tests.

Test ^a	Bootstrap Method	Based on	Test Statistic
$\tau_{S,d}$	Sieve	Differences	DF τ
$t_{S,d}$	Sieve	Differences	DF t
$\tau_{S,r}$	Sieve	Residuals	DF τ
$t_{S,r}$	Sieve	Residuals	DF t
$\tau_{S,d}^a$	Sieve	Differences	ADF τ
$t_{S,d}^a$	Sieve	Differences	ADF t
$\tau_{S,r}^a$	Sieve	Residuals	ADF τ
$t_{S,r}^a$	Sieve	Residuals	ADF t
$\tau_{B,r}$	Block	Residuals	DF τ
$\tau_{B,d}$	Block	Differences	DF τ
$\tau_{B,r}^a$	Block	Residuals	ADF τ
$\tau_{B,d}^a$	Block	Differences	ADF τ
$\tau_{St,d}$	Stationary	Differences	DF τ
$t_{St,d}$	Stationary	Differences	DF t
$\tau_{St,r}$	Stationary	Residuals	DF τ
$t_{St,r}$	Stationary	Residuals	DF t

^aWe use τ for a coefficient test and t for a t-test. The first subscript indicates the bootstrap method, so S stands for sieve bootstrap, B for block bootstrap, and St for stationary bootstrap; the second subscript indicates whether a test is based on differences (d) or residuals (r). A superscript a states that the test is an augmented DF test.

	$\tau_{S,d}$	$t_{S,d}$	$\tau_{S,r}$	$t_{S,r}$	$\tau_{S,d}^a$	$t_{S,d}^a$	$\tau_{S,r}^a$	$t_{S,r}^a$
$L(p_a)$	0.76 (0.05)	0.71 (0.05)	0.98 (0.02)	0.98 (0.02)	0.97 (0.01)	1.00 (0.01)	1.04 (0.01)	1.05 (0.01)
$L(p_a)/n^{1/2}$	5.96 (0.83)	6.44 (0.89)						
$L(p_a)/n$					-8.01 (1.06)	-4.37 (0.90)		
$1/n^{1/2}$			-1.82 (0.66)	-2.43 (0.69)	-3.38 (0.48)	-0.84 (0.45)		
$1/n$	65.91 (13.00)	65.44 (11.76)					-11.87 (3.36)	
$\phi/n^{1/2}$								
ϕ/n			-39.82 (5.42)	-39.37 (5.00)			-8.38 (3.16)	
$\theta/n^{1/2}$	-6.72 (2.00)	-7.11 (2.12)	-6.87 (1.38)	-6.92 (1.39)	-10.12 (1.95)	-8.88 (1.75)	-10.71 (2.64)	-4.44 (0.85)
θ/n					53.28 (15.52)	50.62 (13.97)	40.45 (21.25)	
$\phi\theta/n^{1/2}$								
$\phi\theta/n$			83.61 (32.37)	80.07 (32.84)				
$\phi^2/n^{1/2}$				2.65 (1.13)				
ϕ^2/n	-34.91 (9.09)	-38.71 (13.47)				-9.05 (3.69)		
$\theta^2/n^{1/2}$	16.63 (6.84)	4.34 (2.00)	23.04 (5.20)	24.11 (5.18)	23.27 (3.56)	17.10 (3.08)	18.87 (4.49)	6.99 (1.11)
θ^2/n	-117.06 (71.69)		-80.24 (44.38)	-77.16 (43.59)	-78.19 (28.84)	-53.67 (24.76)	-68.04 (38.09)	
$\phi^3/n^{1/2}$								
ϕ^3/n		24.36 (12.34)						
$\theta^3/n^{1/2}$			-7.22 (3.15)	-7.44 (3.13)	-5.92 (1.58)	-8.00 (1.42)	-38.54 (2.55)	-5.84 (2.17)
θ^3/n	-91.50 (45.56)	-68.82 (36.63)						
R^2	0.96	0.95	0.95	0.96	0.98	0.99	0.97	0.97

Table 2: Response surfaces of size - part I.

	$\tau_{B,r}$	$\tau_{B,d}$	$\tau_{B,r}^a$	$\tau_{B,d}^a$	$\tau_{St,d}$	$t_{St,d}$	$\tau_{St,r}$	$t_{St,r}$
$L(p_a)$	0.90 (0.03)	1.18 (0.02)	0.99 (0.02)	1.33 (0.04)	1.26 (0.06)	1.22 (0.05)	0.93 (0.03)	0.89 (0.03)
$L(p_a)/n^{1/2}$					1.98 (0.84)	3.60 (0.69)		
$L(p_a)/n$								
$1/n^{1/2}$				3.24 (1.26)	4.85 (1.93)	8.28 (1.68)		
$1/n$								
$\phi/n^{1/2}$	-51.89 (7.25)	-12.25 (2.14)	33.35 (9.34)	74.41 (11.34)	-19.18 (1.37)		-70.49 (7.65)	-67.44 (6.98)
ϕ/n	225.32 (54.71)		-294.83 (76.20)	-498.93 (85.48)		-75.31 (17.69)	293.62 (63.56)	279.38 (55.94)
$\theta/n^{1/2}$	-26.09 (3.09)	-14.13 (2.51)		76.18 (13.21)	-14.00 (3.17)	-13.90 (2.46)	-26.51 (3.90)	-26.13 (3.89)
θ/n				-542.53 (101.86)				
$\phi\theta/n^{1/2}$				-153.00 (37.90)				
$\phi\theta/n$				1087.20 (278.55)				
$\phi^2/n^{1/2}$	36.98 (10.04)		-74.65 (9.61)	-93.26 (8.81)			36.31 (11.70)	34.46 (10.62)
ϕ^2/n	-269.90 (82.59)		406.08 (68.37)	525.91 (63.57)			-257.66 (95.39)	-226.92 (81.52)
$\theta^2/n^{1/2}$	98.44 (7.67)	49.95 (7.64)	8.61 (1.66)		51.07 (9.50)	46.63 (9.78)	107.41 (8.73)	102.29 (8.60)
θ^2/n	-547.67 (57.67)	-230.32 (60.16)		71.03 (17.51)	-239.39 (78.71)	-208.68 (77.20)	-612.56 (65.91)	-549.34 (64.20)
$\phi^3/n^{1/2}$	-15.37 (6.38)		-58.50 (21.88)	-71.84 (22.43)		-12.71 (4.39)		
ϕ^3/n		-112.35 (41.71)	407.19 (167.54)	421.52 (166.53)				
$\theta^3/n^{1/2}$	-113.28 (10.61)	-39.43 (10.62)	-19.30 (1.87)	-74.88 (23.69)	-40.56 (12.54)	-44.91 (12.69)	-130.69 (12.37)	-127.68 (12.17)
θ^3/n	814.52 (70.43)	229.05 (76.36)		453.27 (185.98)	225.81 (93.47)	271.90 (92.94)	923.28 (81.38)	884.00 (77.66)
R^2	0.96	0.95	0.92	0.94	0.94	0.95	0.95	0.95

Table 3: Response surfaces of size - part II.

	$\tau_{S,d}$	$t_{S,d}$	$\tau_{S,r}$	$t_{S,r}$	$\tau_{S,d}^a$	$t_{S,d}^a$	$\tau_{S,r}^a$	$t_{S,r}^a$
$\rho - 1$	-97.88 (8.31)	-96.68 (8.52)	-109.45 (9.90)	-107.29 (10.61)	-51.80 (2.99)	-95.68 (7.30)	-111.41 (8.02)	-96.54 (7.82)
$(\rho - 1)/n^{1/2}$	1174.43 (152.66)	1174.56 (156.53)	1197.36 (198.18)	1162.06 (212.58)	267.40 (23.64)	1116.76 (134.40)	1340.08 (151.11)	1164.23 (143.70)
$(\rho - 1)/n$	-4122.47 (674.16)	-4169.84 (691.11)	-3837.90 (923.62)	-3693.95 (989.83)		-3831.94 (593.50)	-4638.88 (680.29)	-4093.96 (635.53)
$L(p_a)$	3.36 (0.57)	3.39 (0.58)	1.08 (0.04)	1.05 (0.05)	1.37 (0.20)	2.64 (0.51)	2.79 (0.53)	2.79 (0.52)
$L(p_a)/n^{1/2}$	-37.42 (10.56)	-38.64 (10.80)			-3.66 (1.58)	-25.63 (9.45)	-28.14 (9.94)	-28.13 (9.72)
$L(p_a)/n$	174.67 (46.95)	181.28 (48.13)				97.43 (41.82)	112.18 (44.75)	109.73 (43.42)
$1/n^{1/2}$	-188.39 (41.15)	-185.96 (41.42)	-77.60 (18.85)	-75.66 (19.82)	-142.66 (14.79)	-153.09 (38.30)	-166.58 (39.75)	-158.35 (37.70)
$1/n$	772.32 (181.72)	752.63 (183.12)	271.43 (88.68)	263.20 (93.52)	520.50 (55.76)	560.42 (168.72)	640.27 (178.22)	579.36 (167.26)
$\phi/n^{1/2}$		-18.76 (4.96)	-8.01 (1.62)	-8.29 (1.59)	-5.62 (0.67)	-13.40 (4.23)	-7.20 (1.26)	-6.66 (1.35)
ϕ/n		117.50 (37.46)				76.60 (31.88)		
$\theta/n^{1/2}$	-5.21 (1.27)		-7.35 (1.62)	-7.86 (1.62)				
θ/n								
$\phi\theta/n^{1/2}$								
$\phi\theta/n$								
$\phi^2/n^{1/2}$	-8.21 (1.25)	-27.03 (9.31)	-18.37 (10.34)	-21.58 (11.05)		-18.88 (7.84)	-3.57 (1.27)	-18.22 (7.96)
ϕ^2/n		149.82 (70.47)	157.17 (78.16)	182.86 (83.31)		128.68 (59.03)		126.49 (60.20)
$\theta^2/n^{1/2}$			40.03 (12.79)	42.57 (13.35)		10.00 (1.18)	35.12 (9.58)	7.62 (1.26)
θ^2/n			-147.12 (96.54)	-162.76 (100.14)	93.40 (10.08)		-182.42 (71.84)	
$\phi^3/n^{1/2}$	-29.85 (9.25)		-40.23 (10.80)	-47.05 (11.76)			-18.93 (9.01)	
ϕ^3/n	178.49 (70.38)		286.04 (77.40)	340.42 (84.28)			179.53 (65.22)	50.83 (17.35)
$\theta^3/n^{1/2}$	-36.07 (10.38)	-44.95 (9.72)	-53.91 (13.91)	-59.30 (14.70)		-49.04 (8.60)	-50.24 (10.45)	-43.99 (8.72)
θ^3/n	180.36 (76.78)	195.35 (73.91)	286.95 (100.00)	327.89 (104.94)	-116.66 (11.30)	248.75 (64.42)	232.23 (78.48)	211.05 (65.77)
Constant	12.77 (2.24)	12.76 (2.25)	5.61 (0.97)	5.43 (1.00)	9.75 (1.08)	10.76 (2.08)	10.92 (2.13)	10.93 (2.04)
R^2	0.95	0.94	0.94	0.94	0.89	0.95	0.95	0.94

Table 4: Response surfaces of power - part I.

	$\tau_{B,r}$	$\tau_{B,d}$	$\tau_{B,r}^a$	$\tau_{B,d}^a$	$\tau_{St,d}$	$t_{St,d}$	$\tau_{St,r}$	$t_{St,r}$
$\rho - 1$	-54.29 (5.00)	-72.48 (3.53)	-123.22 (8.13)	-113.16 (7.18)	-73.07 (3.45)	-70.43 (3.42)	-52.30 (5.07)	-49.77 (4.80)
$(\rho - 1)/n^{1/2}$	198.95 (41.76)	360.80 (29.71)	1443.03 (155.78)	1310.81 (127.87)	374.01 (28.91)	364.01 (28.85)	185.13 (42.37)	171.65 (40.12)
$(\rho - 1)/n$			-4911.63 (707.32)	-4452.39 (558.31)				
$L(p_a)$	1.09 (0.08)	1.48 (0.06)	1.52 (0.17)	2.03 (0.19)	1.72 (0.06)	1.87 (0.06)		
$L(p_a)/n^{1/2}$			-3.37 (1.41)	-5.91 (1.52)			20.95 (3.08)	20.11 (2.90)
$L(p_a)/n$							-85.41 (23.02)	-85.34 (21.67)
$1/n^{1/2}$	-120.72 (23.30)	-143.72 (17.30)	-64.73 (18.36)	-31.60 (6.16)	-146.75 (16.96)	-145.19 (16.75)	-36.50 (7.81)	-35.55 (7.41)
$1/n$	360.66 (106.64)	571.92 (81.76)	174.65 (78.74)		595.82 (80.39)	584.13 (78.92)		
$\phi/n^{1/2}$	-82.87 (9.94)	-36.73 (6.98)	9.80 (4.82)	102.56 (15.14)	-17.77 (2.16)	-15.55 (2.04)	-105.78 (9.35)	-98.44 (9.73)
ϕ/n	402.80 (71.13)	172.17 (51.98)	-83.46 (34.25)	-716.76 (113.67)			455.67 (72.01)	484.65 (69.95)
$\theta/n^{1/2}$	-78.54 (10.61)	-18.73 (2.38)	42.50 (10.57)	111.51 (15.94)	-53.24 (8.18)	-55.97 (8.09)	-80.70 (10.55)	-86.21 (9.82)
θ/n	275.59 (80.95)		-350.93 (79.40)	-789.50 (119.22)	270.99 (63.97)	307.44 (62.86)	287.81 (80.41)	315.71 (74.80)
$\phi\theta/n^{1/2}$	37.74 (10.76)	22.89 (7.36)	-64.33 (27.65)	-137.09 (35.47)	23.52 (7.71)	18.93 (7.21)	35.99 (11.52)	
$\phi\theta/n$			445.07 (209.59)	995.37 (258.63)				263.85 (79.52)
$\phi^2/n^{1/2}$			-82.02 (8.80)	-116.62 (10.02)				
ϕ^2/n			449.79 (69.20)	720.04 (74.90)				
$\theta^2/n^{1/2}$	38.01 (3.00)	73.81 (15.99)	38.86 (12.86)	36.50 (11.89)	81.12 (15.44)	83.27 (15.14)	36.95 (2.96)	37.21 (2.81)
θ^2/n		-298.45 (122.94)	-200.33 (96.63)	-159.91 (89.16)	-369.64 (117.71)	-392.68 (116.01)		
$\phi^3/n^{1/2}$	-24.87 (6.90)	-22.99 (4.61)	-15.54 (3.14)	-95.64 (26.28)	-60.67 (14.39)	-58.00 (14.80)		-20.44 (6.89)
ϕ^3/n				602.88 (197.19)	342.66 (105.13)	340.77 (109.28)		
$\theta^3/n^{1/2}$		-72.94 (18.78)	-94.07 (24.06)	-121.47 (29.37)	-21.01 (5.18)	-21.34 (5.04)		
θ^3/n		400.94 (138.75)	577.66 (180.75)	757.92 (219.90)				
Constant	10.30 (1.28)	9.60 (0.93)	6.00 (1.13)	4.41 (0.77)	9.93 (0.91)	10.37 (0.89)	6.08 (0.91)	5.94 (0.85)
R^2	0.89	0.93	0.95	0.94	0.93	0.93	0.89	0.90

Table 5: Response surfaces of power - part II.

level	ρ	$\tau_{S,d}$	$t_{S,d}$	$\tau_{S,r}$	$t_{S,r}$	$\tau_{S,d}^a$	$t_{S,d}^a$	$\tau_{S,r}^a$	$t_{S,r}^a$
0.01	0.99	0.161	0.139	0.854	0.844	0.982	0.399	0.322	0.286
	0.95	0.906	0.885	0.998	0.997	0.998	0.968	0.976	0.950
	0.90	0.999	0.999	1.000	1.000	1.000	1.000	1.000	1.000
	0.80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.05	0.99	0.980	0.977	0.972	0.968	0.998	0.981	0.979	0.975
	0.95	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
	0.90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.10	0.99	0.998	0.998	0.987	0.985	0.999	0.997	0.997	0.997
	0.95	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

level	ρ	$\tau_{B,r}$	$\tau_{B,d}$	$\tau_{B,r}^a$	$\tau_{B,d}^a$	$\tau_{St,d}$	$t_{St,d}$	$\tau_{St,r}$	$t_{St,r}$
0.01	0.99	0.997	0.972	0.568	0.023	0.942	0.924	0.999	0.998
	0.95	1.000	0.998	0.995	0.686	0.997	0.995	1.000	1.000
	0.90	1.000	1.000	1.000	0.998	1.000	1.000	1.000	1.000
	0.80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.05	0.99	1.000	0.997	0.941	0.397	0.996	0.996	0.999	0.998
	0.95	1.000	1.000	1.000	0.984	1.000	1.000	1.000	1.000
	0.90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.10	0.99	1.000	0.999	0.980	0.749	0.999	0.999	0.999	0.998
	0.95	1.000	1.000	1.000	0.996	1.000	1.000	1.000	1.000
	0.90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 6: Implied asymptotic power.

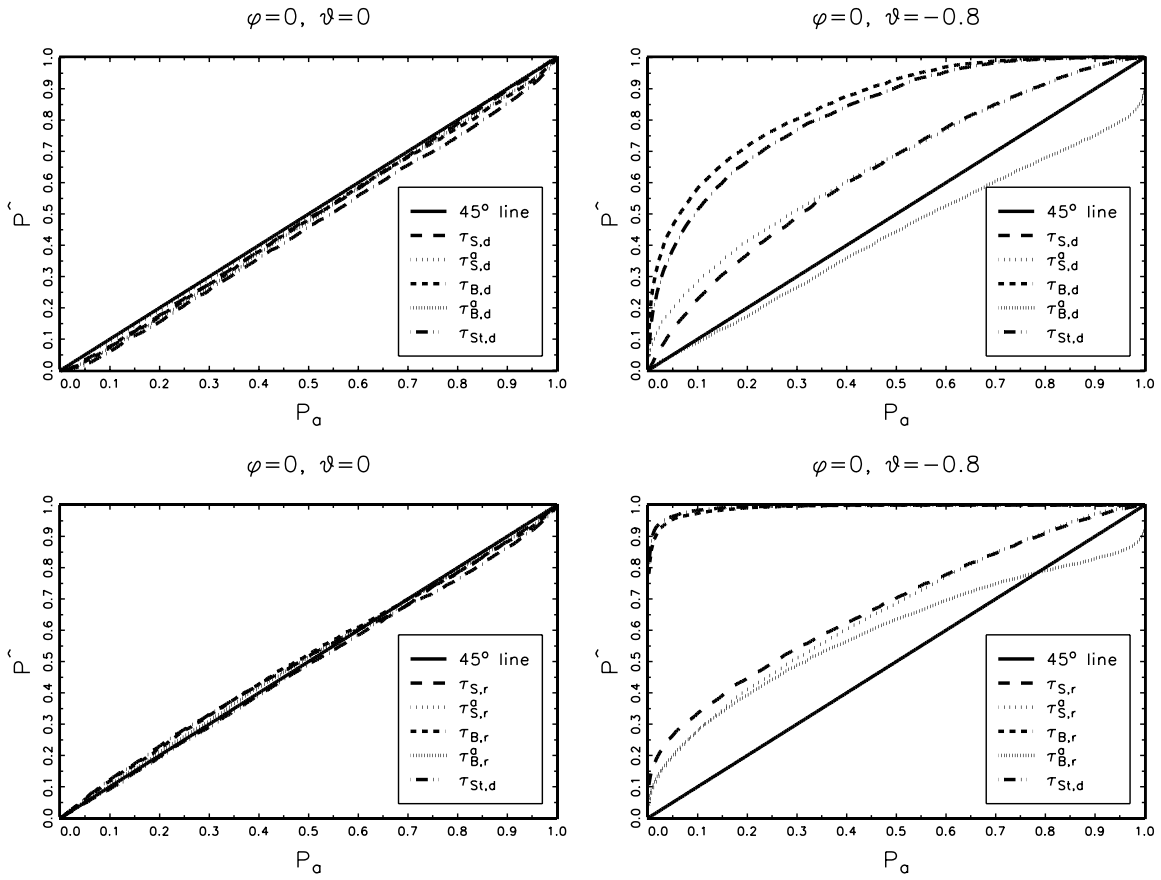


Figure 1: PP plots.

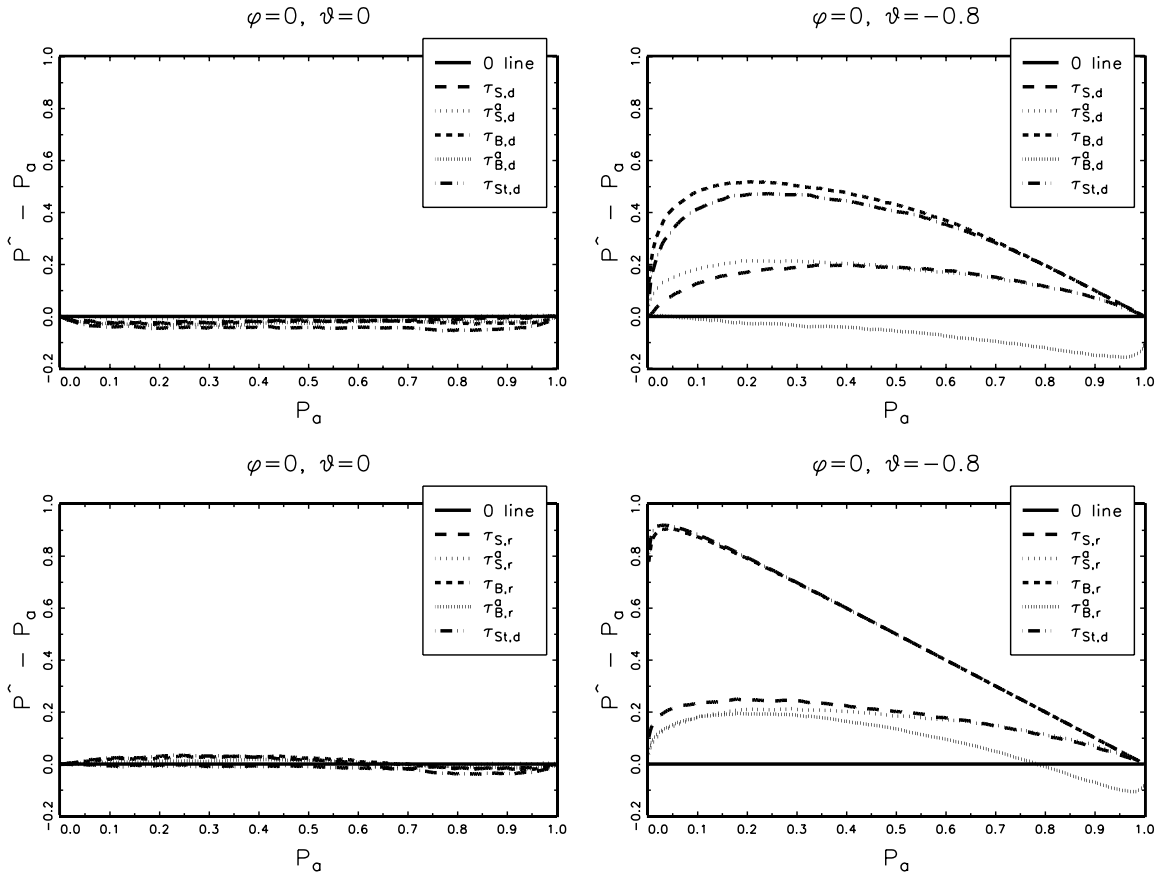


Figure 2: PP discrepancy plots.

Additional results (not for publication)

Response surfaces for the tests with deterministic trends

INSERT TABLE 7 ABOUT HERE

INSERT TABLE 8 ABOUT HERE

INSERT TABLE 9 ABOUT HERE

INSERT TABLE 10 ABOUT HERE

	$\tau_{S,d}$	$t_{S,d}$	$\tau_{S,r}$	$t_{S,r}$	$\tau_{S,d}^a$	$t_{S,d}^a$	$\tau_{S,r}^a$	$t_{S,r}^a$
$L(p_a)$	0.82 (0.05)	0.89 (0.05)	0.92 (0.04)	1.01 (0.03)	0.97 (0.01)	0.89 (0.02)	1.03 (0.02)	0.96 (0.01)
$L(p_a)/n^{1/2}$	7.93 (0.85)	4.04 (0.57)		-1.84 (0.48)		0.65 (0.18)		
$L(p_a)/n$							3.79 (1.45)	
$1/n^{1/2}$	4.64 (1.57)		-6.10 (1.49)	-7.08 (0.96)	5.11 (0.50)	4.05 (0.65)		
$1/n$								-4.86 (2.30)
$\phi/n^{1/2}$		12.78 (1.25)						
ϕ/n			-128.57 (12.59)	-93.11 (15.78)	-6.18 (2.65)		-30.67 (13.12)	-14.85 (9.58)
$\theta/n^{1/2}$			-26.03 (1.13)	-18.45 (2.06)		-1.99 (0.37)		-2.62 (1.02)
θ/n		85.04 (22.91)			-27.19 (5.81)		-28.36 (11.91)	
$\phi\theta/n^{1/2}$						-13.57 (4.20)		
$\phi\theta/n$			244.05 (84.90)	191.25 (42.62)	70.61 (15.59)	131.98 (36.85)	85.23 (29.04)	
$\phi^2/n^{1/2}$			11.89 (2.41)	16.82 (1.15)				
ϕ^2/n	37.27 (17.68)						22.75 (7.94)	
$\theta^2/n^{1/2}$		8.88 (2.03)	59.12 (7.39)	52.49 (5.89)	37.39 (3.13)	16.91 (1.15)	16.23 (0.94)	6.02 (0.62)
θ^2/n			-206.27 (58.47)	-174.11 (45.73)	-193.09 (32.11)	-56.72 (9.76)		
$\phi^3/n^{1/2}$	13.34 (2.19)			-11.60 (3.37)				
ϕ^3/n				126.42 (36.07)			42.48 (22.00)	31.35 (15.50)
$\theta^3/n^{1/2}$	-70.41 (8.15)	-17.85 (4.69)		-32.81 (8.09)	-30.89 (3.49)	-16.95 (1.41)	-11.91 (2.38)	-7.09 (1.82)
θ^3/n	405.17 (62.44)			171.77 (54.70)	197.04 (40.92)	65.24 (10.72)		
Constant					2.97 (0.07)	3.11 (0.06)	3.45 (0.05)	3.40 (0.05)
R^2	0.95	0.91	0.93	0.97	0.98	0.99	0.98	0.99

Table 7: Response surfaces of size (deterministic trends) - part I.

	$\tau_{B,r}$	$\tau_{B,r} (2)$	$\tau_{B,d}$	$\tau_{B,r}^a$	$\tau_{B,d}^a$	$\tau_{St,d}$	$t_{St,d}$	$\tau_{St,r}$	$\tau_{St,r} (2)$	$t_{St,r}$
$L(p_a)$		0.45 (0.35)	0.96 (0.28)	1.73 (0.22)	2.10 (0.36)	1.11 (0.24)	1.53 (0.05)		0.51 (0.27)	0.80 (0.04)
$L(p_a)/n^{1/2}$	25.04 (1.82)	15.63 (7.62)	15.18 (5.99)	-12.82 (4.54)	-18.94 (8.21)	15.33 (4.99)		24.58 (1.52)	13.91 (5.70)	
$L(p_a)/n$	-116.45 (14.11)	-71.79 (37.97)	-72.37 (29.18)	52.71 (22.35)	83.19 (42.92)	-74.69 (24.16)		-114.98 (11.88)	-64.41 (28.17)	
$1/n^{1/2}$					-35.08 (12.72)					-18.76 (3.23)
$1/n$				-37.25 (12.25)	167.45 (95.45)					112.58 (21.29)
$\phi/n^{1/2}$	-72.13 (6.03)	-72.13 (5.98)		40.16 (11.04)	105.32 (24.09)	-18.45 (2.39)	33.35 (8.53)	-71.55 (6.07)	-71.55 (6.12)	-57.15 (3.15)
ϕ/n	251.15 (50.99)	251.15 (50.75)	-187.14 (11.91)	-422.49 (91.65)	-806.36 (178.86)		-203.39 (64.16)	249.73 (51.51)	249.73 (51.86)	270.11 (23.96)
$\theta/n^{1/2}$	-37.53 (4.28)	-37.53 (4.30)	-20.71 (2.77)		84.76 (24.51)	-18.92 (2.37)		-72.63 (15.48)	-72.63 (14.87)	-48.31 (8.85)
θ/n				-75.29 (22.46)	-661.17 (182.67)			291.37 (124.81)	291.37 (120.63)	184.28 (66.26)
$\phi\theta/n^{1/2}$	25.78 (11.77)	25.78 (11.57)	36.67 (9.11)	-77.70 (30.12)	-149.09 (57.44)	41.05 (8.02)	30.49 (10.71)	24.02 (10.66)	24.02 (10.39)	34.78 (3.79)
$\phi\theta/n$				663.47 (251.24)	1281.52 (425.48)					
$\phi^2/n^{1/2}$	98.28 (13.52)	98.28 (13.36)	64.70 (9.77)	-140.32 (8.52)	-144.27 (15.42)	78.34 (9.88)	20.42 (3.53)	100.06 (11.47)	100.06 (11.28)	98.94 (6.68)
ϕ^2/n	-591.96 (105.83)	-591.96 (104.89)	-356.61 (81.28)	877.14 (69.31)	948.82 (113.54)	-425.25 (81.06)		-597.57 (94.67)	-597.57 (93.82)	-497.08 (50.14)
$\theta^2/n^{1/2}$	198.88 (14.52)	198.88 (14.47)	102.24 (13.46)	22.97 (2.24)	33.85 (2.77)	113.34 (10.55)	106.01 (8.89)	199.92 (11.92)	199.92 (11.81)	209.90 (8.77)
θ^2/n	-1081.07 (111.33)	-1081.07 (111.34)	-502.25 (101.79)			-586.63 (79.55)	-590.72 (67.97)	-1092.98 (90.73)	-1092.98 (90.37)	-1141.49 (67.01)
$\phi^3/n^{1/2}$				-73.91 (20.89)	-121.62 (43.20)	39.44 (10.20)				
ϕ^3/n				553.12 (171.90)	860.74 (321.19)	-328.65 (79.17)				
$\theta^3/n^{1/2}$	-248.32 (16.29)	-248.32 (16.33)	-99.97 (16.56)	-19.93 (5.10)	-102.35 (40.83)	-90.60 (12.18)	-78.43 (10.28)	-180.12 (28.33)	-180.12 (27.63)	-195.61 (18.01)
θ^3/n	1647.35 (108.95)	1647.35 (108.82)	585.54 (117.86)		622.70 (306.71)	528.51 (83.67)	452.03 (85.63)	1084.35 (223.74)	1084.35 (218.73)	1155.15 (135.86)
Constant										
R^2	0.97	0.97	0.96	0.95	0.91	0.97	0.87	0.98	0.98	0.99

Table 8: Response surfaces of size (deterministic trends) - part II.

	$\tau_{S,d}$	$t_{S,d}$	$\tau_{S,r}$	$t_{S,r}$	$\tau_{S,d}^a$	$t_{S,d}^a$	$\tau_{S,r}^a$	$t_{S,r}^a$
$\rho - 1$	-116.66 (7.16)	-108.51 (6.37)	-113.67 (6.10)	-86.77 (9.26)	-202.69 (11.11)	-173.12 (5.87)	-190.32 (6.80)	-151.60 (4.09)
$(\rho - 1)/n^{1/2}$	1773.84 (155.18)	1859.60 (133.15)	1436.86 (121.09)	1212.26 (185.80)	2962.31 (203.65)	2492.55 (109.04)	2703.78 (130.10)	2170.68 (70.42)
$(\rho - 1)/n$	-6749.17 (755.18)	-7626.91 (657.44)	-4726.30 (572.14)	-4250.52 (878.69)	-11105.20 (895.76)	-9256.17 (484.71)	-9915.95 (593.25)	-8005.39 (304.12)
$L(p_a)$		1.43 (0.05)	0.93 (0.04)	0.94 (0.04)	2.51 (0.71)	2.20 (0.41)	2.57 (0.45)	1.54 (0.12)
$L(p_a)/n^{1/2}$	26.55 (1.51)				-21.80 (13.03)	-17.49 (7.51)	-22.45 (8.46)	-4.00 (0.95)
$L(p_a)/n$	-94.28 (11.90)				76.11 (57.20)	66.01 (32.90)	88.42 (37.91)	
$1/n^{1/2}$	77.97 (15.90)	24.25 (5.14)	-6.51 (1.26)		-61.99 (46.71)	-52.76 (30.19)	-70.61 (33.63)	-14.37 (4.03)
$1/n$	-412.67 (80.71)	-176.78 (34.48)		-26.38 (10.62)	243.01 (203.24)	210.20 (129.84)	269.18 (148.31)	
$\phi/n^{1/2}$		-49.08 (7.24)		-47.26 (12.75)				
ϕ/n		446.32 (54.95)	-180.40 (19.49)	193.70 (97.59)			-39.72 (11.04)	-20.22 (8.31)
$\theta/n^{1/2}$	-31.71 (3.05)	-40.85 (4.50)		-22.92 (2.33)	-5.30 (1.61)	-3.04 (0.90)		
θ/n	223.62 (23.28)	295.76 (36.08)	-173.57 (18.70)				-40.16 (10.24)	-22.24 (7.70)
$\phi\theta/n^{1/2}$	-55.15 (19.48)							
$\phi\theta/n$	521.90 (160.36)		211.12 (54.18)	126.82 (60.02)	78.19 (23.25)		70.85 (26.13)	
$\phi^2/n^{1/2}$	-42.91 (8.08)	-35.60 (13.07)	-17.87 (9.34)	-45.80 (17.17)	-2.49 (1.38)	-12.52 (6.48)	-17.25 (6.84)	-14.09 (5.74)
ϕ^2/n	361.31 (61.68)	286.87 (99.24)	218.29 (70.50)	471.98 (126.65)		85.42 (48.42)	145.13 (52.24)	104.62 (43.74)
$\theta^2/n^{1/2}$	-33.89 (6.42)		58.17 (8.22)	32.67 (2.33)	31.15 (13.86)	8.58 (0.94)	15.31 (1.24)	
θ^2/n	274.98 (50.38)	47.37 (15.26)	-183.14 (62.05)		-177.63 (103.63)			40.41 (6.53)
$\phi^3/n^{1/2}$	-47.28 (8.14)		-66.17 (11.19)	-45.08 (34.94)	-22.01 (10.35)	-18.00 (6.53)	-18.13 (6.97)	-11.36 (5.83)
ϕ^3/n	432.15 (61.36)		570.63 (88.92)	439.11 (261.85)	142.07 (77.69)	123.42 (48.94)	180.81 (55.22)	115.27 (45.36)
$\theta^3/n^{1/2}$	-16.09 (2.70)		-69.35 (9.21)	-61.27 (12.86)	-27.94 (15.74)	-27.02 (7.23)	-30.58 (9.39)	-21.84 (7.69)
θ^3/n			463.86 (73.62)	391.46 (93.06)	205.62 (114.70)	147.26 (52.50)	153.32 (71.41)	109.59 (58.28)
Constant	-2.58 (0.73)				7.65 (2.59)	7.16 (1.70)	8.25 (1.84)	5.33 (0.50)
R^2	0.94	0.84	0.94	0.88	0.74	0.95	0.95	0.96

Table 9: Response surfaces of power (deterministic trends) - part I.

	$\tau_{B,r}$	$\tau_{B,d}$	$\tau_{B,r}^{\alpha}$	$\tau_{B,d}^{\alpha}$	$\tau_{St,d}$	$t_{St,d}$	$\tau_{St,r}$	$t_{St,r}$
$\rho - 1$	-174.72 (11.52)	-114.61 (12.19)	-132.10 (4.37)	-60.74 (2.60)	-114.33 (8.76)	-75.55 (11.17)	-178.06 (12.10)	-137.01 (12.47)
$(\rho - 1)/n^{1/2}$	2460.92 (215.79)	1556.85 (237.56)	1638.60 (83.20)	401.09 (21.16)	1688.98 (163.17)	1272.13 (208.43)	2523.01 (230.36)	1985.63 (240.50)
$(\rho - 1)/n$	-9312.36 (976.20)	-5521.42 (1090.11)	-5357.49 (386.37)		-6440.94 (732.62)	-5271.59 (942.31)	-9427.97 (1057.31)	-7525.43 (1115.11)
$L(p_a)$	0.89 (0.15)	0.98 (0.36)	1.11 (0.03)	1.15 (0.05)	1.74 (0.05)	1.60 (0.04)	1.16 (0.04)	0.93 (0.04)
$L(p_a)/n^{1/2}$	2.67 (1.20)	13.60 (7.50)						
$L(p_a)/n$		-60.62 (35.62)	-11.10 (2.95)					
$1/n^{1/2}$					-3.30 (1.52)			
$1/n$			-42.64 (8.21)	43.17 (14.18)				
$\phi/n^{1/2}$	-155.92 (9.99)	-23.30 (1.93)	29.08 (6.65)	130.36 (13.24)	-19.34 (2.03)	-34.76 (10.46)	-159.80 (11.03)	-163.92 (12.24)
ϕ/n	835.20 (75.11)		-343.04 (52.70)	-1009.11 (98.88)		273.98 (78.81)	869.73 (82.96)	989.40 (91.30)
$\theta/n^{1/2}$	-106.50 (18.94)	-25.84 (2.02)		92.57 (12.82)	-26.07 (2.03)		-142.74 (9.50)	-140.07 (9.16)
θ/n	497.42 (143.06)		-116.96 (12.50)	-737.51 (96.03)			769.69 (69.04)	798.06 (63.14)
$\phi\theta/n^{1/2}$	163.99 (52.26)	25.50 (6.83)	-70.40 (18.32)	-152.36 (31.41)	36.37 (6.38)	27.38 (9.86)	166.25 (55.78)	159.49 (59.47)
$\phi\theta/n$	-1025.02 (400.57)		557.13 (152.32)	1278.33 (236.53)			-1048.97 (431.12)	-915.36 (460.70)
$\phi^2/n^{1/2}$	62.02 (17.60)	17.83 (2.30)	-129.41 (6.38)	-175.55 (11.33)	25.83 (2.47)	21.17 (2.84)	71.28 (18.88)	64.72 (21.00)
ϕ^2/n	-315.55 (131.88)		798.74 (49.46)	1161.46 (85.46)			-376.24 (139.91)	-233.76 (154.02)
$\theta^2/n^{1/2}$	156.56 (15.35)	121.54 (15.74)	20.18 (1.38)	50.58 (10.09)	123.47 (10.98)	114.35 (13.22)	169.55 (14.66)	146.84 (13.77)
θ^2/n	-770.32 (116.22)	-630.37 (124.89)		-149.14 (75.91)	-626.38 (81.88)	-632.99 (98.20)	-884.57 (110.03)	-675.32 (103.16)
$\phi^3/n^{1/2}$		-42.60 (15.85)	-98.08 (13.57)	-154.33 (25.54)	-43.99 (18.66)			
ϕ^3/n		285.17 (113.29)	736.38 (105.99)	1134.54 (191.75)	299.61 (132.88)			
$\theta^3/n^{1/2}$	-96.13 (36.37)	-108.35 (5.27)	-44.06 (8.30)	-119.01 (23.39)	-108.28 (13.08)	-139.57 (14.53)	-32.03 (6.34)	-35.99 (6.58)
θ^3/n	513.74 (275.10)		249.07 (67.94)	782.04 (175.75)	739.85 (90.59)	859.09 (107.90)		
Constant				-2.04 (0.27)				
R^2	0.93	0.94	0.97	0.93	0.93	0.83	0.93	0.90

Table 10: Response surfaces of power (deterministic trends) - part II.