

# Detrending bootstrap unit root tests

## Citation for published version (APA):

Smeeke, S. (2013). Detrending bootstrap unit root tests. *Econometric Reviews*, 32(8), 869-891. <https://doi.org/10.1080/07474938.2012.690693>

## Document status and date:

Published: 01/04/2013

## DOI:

[10.1080/07474938.2012.690693](https://doi.org/10.1080/07474938.2012.690693)

## Document Version:

Publisher's PDF, also known as Version of record

## Document license:

Taverne

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.



## Detrending Bootstrap Unit Root Tests

Stephan Smeekes

To cite this article: Stephan Smeekes (2013) Detrending Bootstrap Unit Root Tests, *Econometric Reviews*, 32:8, 869-891, DOI: [10.1080/07474938.2012.690693](https://doi.org/10.1080/07474938.2012.690693)

To link to this article: <https://doi.org/10.1080/07474938.2012.690693>



Published online: 13 Mar 2013.



Submit your article to this journal [↗](#)



Article views: 290



View related articles [↗](#)



Citing articles: 1 View citing articles [↗](#)

## DETRENDING BOOTSTRAP UNIT ROOT TESTS

**Stephan Smeekes**

*Department of Quantitative Economics, Maastricht University, Maastricht,  
The Netherlands*

□ *The role of detrending in bootstrap unit root tests is investigated. When bootstrapping, detrending must not only be done for the construction of the test statistic, but also in the first step of the bootstrap algorithm. It is argued that the two issues should be treated separately. Asymptotic validity of sieve bootstrap augmented Dickey–Fuller (ADF) unit root tests is shown for test statistics based on full sample and recursive ordinary least squares (OLS) and generalized least squares (GLS) detrending. It is also shown that the detrending method in the first step of the bootstrap may differ from the one used in the construction of the test statistic. A simulation study is conducted to analyze the effects of detrending on finite sample performance of the bootstrap test. It is found that full sample OLS detrending should be preferred based on power in the first step of the bootstrap algorithm, and that the decision about the detrending method used to obtain the test statistic should be based on the power properties of the corresponding asymptotic tests.*

**Keywords** Deterministic trends; Sieve bootstrap; Unit root test.

**JEL Classification** C15; C22.

### 1. INTRODUCTION

In recent years we have seen a large number of papers on the application of the bootstrap to nonstationary time series. The good performance of bootstrap methods in stationary time series has led people to adapt the methods to nonstationary settings, and in particular to unit root testing, where finite sample size distortions are known to occur frequently. The literature has focused mainly on how to deal with serial correlation, but it stays relatively silent on how to deal with deterministic

Address correspondence to Stephan Smeekes, Department of Quantitative Economics, Maastricht University, P. O. Box 616, 6200 MD Maastricht, The Netherlands; E-mail: S.Smeekes@maastrichtuniversity.nl

trends. Our aim in this paper is to investigate how the method of detrending impacts the performance of bootstrap unit root tests.

Allowing for deterministic trends is very important in applications. Many economic series such as real gross domestic product (GDP) can be thought of as containing a linear trend, while the inclusion of an intercept is relevant for virtually every economic time series. It is therefore crucial to have tests that can take such trends into account. One way to take a trend into account is to include it in the unit root equation and make it part of the testable hypothesis, such as the  $\Phi$ -tests of Dickey and Fuller (1981). The alternative way, that we consider here, is to perform an initial step of detrending, and then to perform the unit root test on the detrended series.

It is well known in the unit root literature that the method of detrending can have a major impact on the power of the tests. In their seminal work, Elliott et al. (1996) showed that generalized least squares (GLS), or quasi-difference, detrending is optimal in terms of local asymptotic power if the initial condition is equal to zero. Alternatively, Shin and So (2001) and Taylor (2002) propose to detrend the data recursively. Shin and So (2001) show that with recursive demeaning that the bias of the estimate of the autoregressive parameter decreases and correspondingly the power of the test increases.

While one might expect the power properties of the asymptotic tests to carry over to the bootstrap setting, it might be that the method of detrending in the actual bootstrap procedure has an effect on the size of the bootstrap tests as well. The argument of Shin and So (2001) that the autoregressive parameter is estimated more precisely with recursive detrending, could for example lead one to expect an improvement in size properties of the bootstrap tests. In this paper we investigate the effects of the detrending methods in a bootstrap context.<sup>1</sup>

There is a large array of different bootstrap unit root tests available, see, e.g., Chang and Park (2003), Paparoditis and Politis (2003), Swensen (2003a), Parker et al. (2006), Richard (2007), and Cavaliere and Taylor (2009) among others. The tests that have been proposed in these papers differ in the bootstrap method used (usually sieve, block or wild bootstrap), in the test statistic used, and the tests differ in whether the null is imposed in the estimation for the bootstrap. We choose to focus on one particular test, the residual-based augmented Dickey–Fuller (ADF) sieve bootstrap *t*-test. This test was proposed by Paparoditis and Politis (2005)

<sup>1</sup>It is not our intention in this paper to find an “optimal” unit root test; we focus on the effect of the aforementioned detrending methods in the bootstrap. A thorough investigation of modifications of the basic Dickey–Fuller test, including hybrid forms, has been done in Leybourne et al. (2005).

and Palm et al. (2008) and shown to perform well in the simulation study of Palm et al. (2008) compared to other alternatives.<sup>2</sup>

In this paper we extend the proof of asymptotic validity given in Palm et al. (2008) to a setting with deterministic components in the data generating process (DGP), and allowing for a range of detrending methods that includes full sample and recursive ordinary least squares (OLS) and GLS detrending. A simulation study investigates the impact of the method of detrending on the performance of the bootstrap unit root test. By allowing for a different method of detrending in the first step of the bootstrap procedure than in the calculation of the test statistic, we can analyze the two points separately.

An interesting question is when to apply the tests with just an intercept, and when to include both an intercept and a trend. As analyzed by, among others, Harvey et al. (2009), estimating the model with trend in the absence of a trend in the DGP leads to a significant loss of power compared to the model with just an intercept. On the other hand, the tests with intercept only are not invariant to the presence of a trend in the DGP and should therefore not be applied in this setting. We will not analyze this issue explicitly in combination with the bootstrap; for the tests considered in this paper the problem is essentially the same whether one uses the bootstrap or not. As such, the conclusions of Harvey et al. (2009) remain relevant with the application of the bootstrap as well. The approach to deal with uncertainty regarding the deterministic trend and the initial condition by a union of tests as proposed by Harvey et al. (2009) and Harvey et al. (2012) has been extended to a bootstrap setting by Smeekes and Taylor (2012).

The outline of the paper is as follows. Section 2 will describe the model used for the theoretical analysis. The tests and their limit distributions are discussed in Section 3. In Section 4 a simulation study will be undertaken. Section 5 concludes. All proofs are contained in the Appendix. Finally, a word on notation.  $[x]$  is the largest integer smaller than or equal to  $x$ . Convergence in distribution (probability) is denoted by  $\xrightarrow{d}$  ( $\xrightarrow{p}$ ). Bootstrap quantities (conditional on the original sample) are indicated by appending a superscript  $*$  to the standard notation. Convergence in distribution (probability) of bootstrap statistics is denoted  $\xrightarrow{d^*}$  ( $\xrightarrow{p^*}$ ), where the bootstrap convergence holds in probability.  $W(r)$  denotes a univariate standard Brownian motion.

<sup>2</sup>Wild bootstrap tests such as those of Cavaliere and Taylor (2008) and Cavaliere and Taylor (2009) have also been shown to perform well. However, in our setting of homoskedasticity they are very similar to the sieve bootstrap test we consider.

## 2. THE MODEL WITH DETERMINISTIC TRENDS

We consider the following DGP for  $y_t$  ( $t = 1, \dots, T$ ):

$$\begin{aligned} y_t &= x_t + \beta' z_t, & x_t &= \rho x_{t-1} + u_t, \\ u_t &= \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \psi(L) \varepsilon_t. \end{aligned} \quad (1)$$

The process  $z_t$  is a deterministic process. In particular, we consider  $z_t = 1$  and  $z_t = (1, t)'$ . In the remainder of the paper we will focus on the case with linear trend, but it is clear that all results will also hold for the intercept only case. The null hypothesis  $H_0 : \rho = 1$  corresponds to a unit root, possibly in the presence of a deterministic trend. Under the alternative  $H_1 : |\rho| < 1$ , with the following conditions on the linear process  $u_t$ , the process is integrated of order zero.

**Assumption 1.** (i) Let  $\varepsilon_t$  be i.i.d. with  $E\varepsilon_t = 0$ ,  $E\varepsilon_t^2 = \sigma^2$ , and  $E\varepsilon_t^4 < \infty$ . (ii)  $\psi(z) \neq 0$  for all  $|z| \leq 1$ , and  $\sum_{j=0}^{\infty} j|\psi_j| < \infty$ .

These assumptions, which are comparable to those found in the literature (cf. Phillips and Solo, 1992; Chang and Park, 2002), are sufficient for the derivation of the asymptotic distribution of the test statistic and its bootstrap counterpart. For the derivation of the asymptotic distribution of the original test statistic, Assumption 1(i) could be relaxed to allow  $\varepsilon_t$  to be a martingale difference series, but for the sieve bootstrap one needs the independent and identically distributed (i.i.d.) assumption (Chang and Park, 2003).

The treatment of the deterministic components is comparable to Elliott et al. (1996). Moreover, as in Elliott et al. (1996), we assume that the initial condition is zero, i.e.,  $x_0 = 0$ . While this is an innocuous assumption under the null hypothesis as  $x_0$  cannot be identified if a constant is included in the model, this is a crucial assumption under the alternative for the optimality of the approach of Elliott et al. (1996), as discussed by Müller and Elliott (2003), Elliott and Müller (2006), and Harvey et al. (2009) among others. A theoretical discussion on the role of the initial condition for the optimality of the tests is beyond the scope of this paper, but we will return to the point in the simulation study in Section 4.

## 3. BOOTSTRAP ADF TESTS WITH DETRENDING

In this section the (bootstrap) ADF tests with detrending are discussed and their limit distributions are derived. We describe the methods in a general framework, of which OLS, GLS, and recursive detrending are

special cases.<sup>3</sup> In the following we will focus on the Dickey–Fuller  $t$ -statistic, as this is the most popular in practice. We do not explicitly consider the ADF coefficient test, which has been discussed by Xiao and Phillips (1998) with GLS detrending. However, all results derived here also apply to the ADF coefficient test, although a slightly stronger assumption on the lag length in the ADF regression is needed (cf. Chang and Park, 2002).

### 3.1. Detrended ADF Statistics

We define the detrended series  $y_{t,\gamma,\lambda}^d$  as

$$y_{t,\gamma,\lambda}^d = y_t - \hat{\beta}'_{t,\gamma,\lambda} z_t, \quad \text{where} \quad \hat{\beta}_{t,\gamma,\lambda} = \left( \sum_{t=1}^{\lambda^*} z_{c_{T,\gamma},t} z'_{c_{T,\gamma},t} \right)^{-1} \left( \sum_{t=1}^{\lambda^*} z_{c_{T,\gamma},t} y_{c_{T,\gamma},t} \right). \tag{2}$$

Here  $\lambda^* = \max(t, \lfloor T\lambda \rfloor)$  has the same meaning as in Taylor (2002). It indicates if and how recursive detrending is used, as only observations up to  $\lambda^*$  are used. If  $\lambda = 0$ ,  $\lambda^* = t$  and “full” recursive detrending is used. If  $\lambda = 1$ ,  $\lambda^* = T$  and the full sample is always used to detrend. If  $\lambda$  is between 0 and 1, recursive detrending is used, but a minimum proportion of the sample is always used in estimating  $\beta$ .

We consider the variant of recursive detrending of Taylor (2002). It is easier to apply than the Shin and So (2001) method and does not require the adjustment of Sul (2009), which is necessary for the Shin and So (2001) method as their approach is not invariant to the trend parameter (Rodrigues, 2006). Moreover, it directly lends itself to be put into the framework described above. The main argument for recursive detrending is to avoid using an explanatory variable (the first lag) that is correlated with the error term, which happens for full sample OLS demeaning through the subtraction of the overall mean estimate. Shin and So (2001) showed using simulations that the first order autoregressive estimator under recursive demeaning is less biased than under full sample demeaning, and as a consequence, unit root tests based on recursive demeaning are more powerful.

Next,  $z_{c_{T,\gamma},t} = z_t - (1 - c_{T,\gamma})z_{t-1}$  for  $t \geq 2$  and  $z_{c_{T,\gamma},1} = z_1$ . We specify  $c_{T,\gamma}$  as  $c_{T,\gamma} = \bar{c}^\gamma T^{-\gamma}$ . If  $\gamma = 0$ , this is OLS detrending as  $c_{T,0} = 1$  and hence  $z_{c_{T,0},t} = z_t$ . If  $\gamma = 1$ , this is the GLS detrending of Elliott et al. (1996) as  $c_{T,1} = \bar{c} T^{-1}$  and hence  $z_{c_{T,1},t} = z_t - (1 - \bar{c} T^{-1})z_{t-1}$ .  $y_{c_{T,\gamma}}$  is defined accordingly. Elliott et al. (1996) consider the construction of unit root tests that are point optimal against a local alternative  $\rho = 1 - \bar{c} T^{-1}$ .

<sup>3</sup>A general framework that nests all these options was presented by Broda et al. (2009). Ours slightly deviates from theirs as our objectives are different.

The parameter  $\bar{c}$  has to be selected by the user. Elliott et al. (1996) recommend using  $\bar{c} = 7$  for the intercept only case and  $\bar{c} = 13.5$  for the linear trend case, as the power functions of the DF-GLS test are very close to the power envelope for these values. As these values are commonly accepted we will use them as well later in our simulation study.

To lighten the notational load, we will not explicitly mention the dependence on  $\gamma$  and  $\lambda$  when no confusion can arise. Hence, we usually write  $y_t^d$  and  $\hat{\beta}_t$  when the context is clear. The ADF  $t$ -statistic  $ADF_{\gamma,\lambda}$  is then the usual regression  $t$ -statistic of significance on  $\delta$  in the augmented DF regression

$$\Delta y_t^d = \delta y_{t-1}^d + \sum_{j=1}^p \phi_j \Delta y_{t-j}^d + \varepsilon_{p,t}^d.$$

We need the following assumption on the lag length  $p$  in the ADF regression.

**Assumption 2.** *Let  $p \rightarrow \infty$  and  $p = o(T^{1/2})$  as  $T \rightarrow \infty$ .*

The limiting distribution of the ADF  $t$ -statistic is given below. Note that the asymptotic distributions given below reduce to the standard distributions found in Elliott et al. (1996) and Chang and Park (2002) when  $\lambda = 1$ .

**Theorem 1.** *Let  $y_t$  be generated by (1) with  $\rho = 1$ , and let Assumptions 1 and 2 hold. Let  $\gamma = 0, 1$  and  $\lambda \in [0, 1]$ . Then, as  $T \rightarrow \infty$ , we have that*

$$ADF_{\gamma,\lambda} \xrightarrow{d} \frac{W_\gamma(1, \lambda)^2 - W_\gamma(0, \lambda)^2 - 1}{2 \left( \int_0^1 W_\gamma(r, \lambda)^2 dr \right)^{1/2}},$$

where

$$W_0(r, \lambda) = W(r) - 2\bar{r}^{-2}(2\bar{r} - 3r) \int_0^{\bar{r}} W(s) ds - 6\bar{r}^{-3}(2r - \bar{r}) \int_0^{\bar{r}} sW(s) ds,$$

$$W_1(r, \lambda) = W(r) - r\bar{r}^{-1} \left( 1 + \bar{c}\bar{r} + \frac{1}{3}\bar{c}^2\bar{r}^2 \right)^{-1} \left[ (1 + \bar{c}\bar{r})W(\bar{r}) + \bar{c}^2 \int_0^{\bar{r}} sW(s) ds \right],$$

and  $\bar{r} = \max(r, \lambda)$ .

**Remark 1.** Under the local alternative  $\rho = 1 - cT^{-1}$ , the limit distribution will remain the same as in Theorem 1, but with  $W(r)$  replaced by  $W_c(r) = \int_0^r e^{-(r-s)c} dW(s)$  in the expressions for  $W_\gamma(r, \lambda)$ . This can be shown straightforwardly, though tediously, using standard results regarding the invariance principle (cf. Phillips and Perron, 1988) and our proofs in the appendix.



### 3.2. Bootstrap ADF Statistics and Their Asymptotic Properties

The bootstrap algorithm we consider is an extension of Bootstrap Test 4 given in Palm et al. (2008). The extension is Step 1, on the treatment of deterministic components.

#### Algorithm 1.

1. Calculate  $\tilde{y}_t^d = y_t - \tilde{\beta}'_t z_t$ , where  $\tilde{\beta}_t = \hat{\beta}_{t, \tilde{\gamma}, \tilde{\lambda}}$  is defined in (2), but it is not necessary that  $\tilde{\gamma} = \gamma$  and  $\tilde{\lambda} = \lambda$ .
2. Estimate an ADF regression of order  $q$  for  $\tilde{y}_t^d$  by OLS and calculate the residuals

$$\hat{\varepsilon}_{q,t}^d = \Delta \tilde{y}_t^d - \hat{\delta} \tilde{y}_{t-1}^d - \sum_{j=1}^q \hat{\phi}_j \Delta \tilde{y}_{t-j}^d. \tag{3}$$

3. Resample with replacement from the recentered residuals  $\left( \hat{\varepsilon}_{q,t}^d - \bar{\varepsilon}_{q,t}^d \right)$  to obtain bootstrap errors  $\varepsilon_t^*$ .
4. Build  $u_t^*$  recursively as  $u_t^* = \sum_{j=1}^q \hat{\phi}_j u_{t-j}^* + \varepsilon_t^*$ , using the estimated parameters  $\hat{\phi}_j$  from Step 2, and build  $x_t^*$  as  $x_t^* = x_{t-1}^* + u_t^*$ . Finally, let  $y_t^* = x_t^* + \beta^* z_t$ . See Remark 2 for the choice of  $\beta^*$ .
5. Using the bootstrap sample  $y_t^*$ , apply the same method of detrending as applied to the original sample to obtain the detrended bootstrap series  $y_t^{*d} = y_{t, \gamma, \lambda}^{*d}$ . Calculate  $ADF_{\gamma, \lambda}^{*d}$  as the  $t$ -statistic of significance of  $\delta^*$  in the ADF regression of order  $p^*$

$$\Delta y_t^{*d} = \delta^* y_{t-1}^{*d} + \sum_{j=1}^{p^*} \phi_j^* \Delta y_{t-j}^{*d} + \varepsilon_{p^*, t}^{*d}.$$

6. Repeat Steps 3 to 5  $B$  times, obtaining bootstrap test statistics  $ADF_{\gamma, \lambda}^{*b}$  for  $b = 1, \dots, B$ , and select the bootstrap critical value  $c_\alpha^*$  as  $c_\alpha^* = \max\{c : \sum_{b=1}^B I(ADF_{\gamma, \lambda}^{*b} < c) \leq \alpha\}$ , or equivalently as the  $\alpha$ -quantile of the ordered  $ADF_{\gamma, \lambda}^{*b}$  statistics. Reject the null of a unit root if  $ADF_{\gamma, \lambda}^{*d}$  is smaller than  $c_\alpha^*$ , where  $\alpha$  is the nominal level of the test.

As can be seen from the algorithm above, we allow for a different lag length in the sieve bootstrap ( $q$ ) than in the calculation of the test statistic ( $p$ ). Moreover, we allow for a different lag length in the calculation of the bootstrap test statistic ( $p^*$ ). It might seem a logical choice to set  $q = p$ , as both are based on an ADF regression. However we do not wish to impose this because, if the methods of detrending differ, the ADF regressions are not the same, and one might obtain a different  $p$  and  $q$  if the choice is based on an information criterion.

It is also important to allow for lag length selection of  $p^*$  within the bootstrap, as this will improve the finite sample properties of the test. In the following we will simply denote  $p^*$  by  $p$  to lighten the notational load. This is a harmless simplification as we require  $p^*$  to satisfy Assumption 2 as well, and moreover  $p$  and  $p^*$  will never be in the same part of the proof anyway. The finite sample performance of the tests might improve by imposing certain restrictions on the relation between  $p$  and  $p^*$ ; see Richard (2009) for more details. We will not explore this here any further. We need the following assumption on the lag lengths.

**Assumption 3.** (i) Let  $q \rightarrow \infty$  and  $q = o((n/\ln n)^{1/3})$  as  $n \rightarrow \infty$ . (ii) Let  $p/q \rightarrow \kappa > 1$  as  $T \rightarrow \infty$ , where  $\kappa$  may be infinite.

The second part of the assumption essentially states that, for large  $T$ ,  $p$  should be at least as large as  $q$ .

**Remark 2.** It is unnecessary to include deterministic components in Step 4 of the bootstrap algorithm, as the tests we consider are invariant with respect to the true deterministic components in the (bootstrap) DGP. Therefore, we recommend setting  $\beta^* = 0$  for simplicity. While our results would continue to hold for other values of  $\beta^*$ , it would not be valid to set  $y_t^* = x_t^* + \beta_t^* z_t$ , with  $\beta_t^* = \hat{\beta}_t$ , as this would mean that the parameters of the deterministic trends are time-varying, which is not the case in the original sample.

It is important to note that the detrending method in the first step of the bootstrap test using  $\hat{\beta}_{t,\tilde{\gamma},\tilde{\lambda}}$  does not have to be the same as the one performed in the test using  $\hat{\beta}_{t,\tilde{\gamma},\tilde{\lambda}}$ . Specifically, we do not require that  $\tilde{\gamma} = \gamma$  and  $\tilde{\lambda} = \lambda$ ; the properties of the estimated coefficients and residuals are identical asymptotically for any  $\tilde{\gamma}$  and  $\tilde{\lambda}$ . This is formalized in the following lemma.

**Lemma 1.** Define  $\tilde{\phi}_j$ ,  $j = 1, \dots, q$  as the OLS estimators in a regression of  $u_t$  on  $u_{t-1}, \dots, u_{t-q}$ . Let  $\hat{\phi}_j$  be defined as in (3). Let  $\tilde{\beta}_t = \hat{\beta}_{t,\tilde{\gamma},\tilde{\lambda}}$  be defined as in (2) with  $\tilde{\gamma} = 0, 1$  and  $\tilde{\lambda} \in [0, 1]$  and let Assumptions 1 and 3 hold. Then

$$\hat{\phi}_j = \tilde{\phi}_j + O_p(T^{-1}q^{1/2}),$$

uniformly in  $j = 1, \dots, q$ .

Using the above lemma we can use the results on autoregressive approximation and the sieve bootstrap as established by Hannan and Kavalieris (1986) and Bühlmann (1997), used in a unit root setting by Park (2002) and Chang and Park (2003) (also see Remark 4). Given Lemma 1

and the results mentioned above, we can establish the limit distribution of the detrended ADF bootstrap tests.

**Remark 3.** If we restrict ourselves to full sample detrending then one can show that all that is required of  $\tilde{\beta} = (\tilde{\beta}_1, \tilde{\beta}_2)'$  is that it satisfies the conditions  $\tilde{\beta}_1 - \beta_1 = O_p(T^{1/2})$  and  $\tilde{\beta}_2 - \beta_2 = O_p(T^{-1/2})$ , thus allowing for trend estimators beyond the OLS and GLS framework (see Smeekes, 2009a, Chapter 3). We conjecture that a similar result holds for recursive detrending.

**Remark 4.** One might consider using Yule–Walker instead of OLS in the sieve bootstrap to ensure that the estimated autoregression is invertible.<sup>4</sup> In fact, the results of Hannan and Kavalieris (1986) and Bühlmann (1997) are derived for Yule–Walker estimators. However, Theorem 1 of Poskitt (1994) implies that these results are valid for OLS estimation as well.

We can now derive the following theorem on the asymptotic distribution of the bootstrap ADF  $t$ -statistics. Note that, as the limit distribution of the bootstrap statistic is the same as that of its asymptotic counterpart, this theorem establishes the asymptotic validity of the bootstrap ADF test.

**Theorem 2.** *Let  $y_t$  be generated by (1) with  $\rho = 1$ , and let Assumptions 1, 2 and 3 hold. Let  $\gamma = 0, 1$  and  $\lambda \in [0, 1]$ . Then, as  $T \rightarrow \infty$ , we have that*

$$ADF_{\gamma,\lambda}^* \xrightarrow{d^*} \frac{W_\gamma(1, \lambda)^2 - W_\gamma(0, \lambda)^2 - 1}{2 \left( \int_0^1 W_\gamma(r, \lambda)^2 dr \right)^{1/2}} \text{ in probability.}$$

The asymptotic validity of the bootstrap tests that we established is a property of the bootstrap tests under the null hypothesis. We next investigate how the bootstrap performs under the alternative hypothesis. We discern local and fixed alternatives.

Under local alternatives the bootstrap tests must have the same asymptotic distribution as under the null hypothesis, as it is only then that the bootstrap tests will have the same asymptotic local power function as the asymptotic tests. Here this means that Theorem 2 should remain valid under local alternatives. Swensen (2003b) shows that this is true for full sample OLS and GLS detrending under i.i.d. errors. Without going into technical details, it is not difficult to see that under local alternatives the result of Theorem 2 will continue to hold. Under local alternatives

<sup>4</sup>The disadvantage of Yule–Walker is that it may have substantial finite sample bias (Poskitt, 1994). Another option if one is worried about the noninvertibility of the OLS estimates is to impose a root bound as in Burridge and Taylor (2004).

all rates of convergence remain the same as under the null hypothesis, including those of the trend estimators, which will ensure that all results, including Lemma 1, remain valid. It then follows directly from Lemma 1 that Theorem 2 will continue to hold.

Under fixed alternatives, the bootstrap test needs to converge to some limiting distribution in order to achieve consistency. However, to have the highest power possible one needs that Theorem 2 continues to hold under fixed alternatives. For fixed alternatives we may write  $x_t = (1 - \rho L)^{-1} \psi(L) \varepsilon_t = \psi^+(L) \varepsilon_t$ , where  $\psi^+(L)$  is an invertible polynomial. Therefore, one may approximate  $x_t$  with a finite order autoregressive model, or in other words, directly apply the sieve bootstrap of Bühlmann (1997) to it. Our ADF regression is equivalent to the direct autoregressive approximation and therefore valid as well. As such, the estimates  $\hat{\phi}_j$  will converge to their population counterparts with rates as in Hannan and Kavalieris (1986). The only complication arising is the detrending, as the trend estimators have different properties in the stationary setting. However, the trend estimators will converge at higher rates,<sup>5</sup> which means that this will not cause any problems. For these reasons, Theorem 2 will continue to hold, and the bootstrap tests will have the same distributions under fixed alternatives as under the null hypothesis.

## 4. FINITE SAMPLE PERFORMANCE

### 4.1. Simulation Setup

In this section, a Monte Carlo study is performed to investigate the performance of the methods in finite samples. Our goal is twofold. First, we wish to investigate whether the power properties of the asymptotic tests carry over to the bootstrap setting. For example, it is well known that the GLS detrended test is more powerful than the OLS detrended test if the initial condition, the deviation of the initial observation from the deterministic components, is small, while it is the other way around if the initial condition is large (cf. Müller and Elliott, 2003). Therefore, we will perform simulations both with a small (zero) initial condition and with a large initial condition. Our goal is certainly not to give a complete analysis of the power properties of the tests, but simply to get an idea of whether power properties carry over to the bootstrap.

The second goal is to investigate whether the method of detrending in the first step of the bootstrap procedure has an impact on the performance of the test (both size and power). As discussed in the previous section, the

<sup>5</sup>See for example (Hamilton, 1994, Chapter 16) for the OLS estimator in a model with intercept and trend.

method of detrending in the bootstrap does not have to be the same as the method performed for the construction of the test statistic.

In order to investigate this we will consider all combinations of OLS ( $\gamma = 0$ ), GLS ( $\gamma = 1$ ), full sample ( $\lambda = 1$ ) and full recursive ( $\lambda = 0$ ) detrending for use in the bootstrap and the construction of the test statistic, including their asymptotic variants. The asymptotic tests are denoted by  $ADF_{\gamma,\lambda}$  where  $\gamma$  and  $\lambda$  indicate the method of detrending for the calculation of the test statistic as before. The bootstrap tests are denoted by  $ADF_{\gamma,\lambda}^{*,\tilde{\gamma},\tilde{\lambda}}$ , where  $\tilde{\gamma}, \tilde{\lambda}$  indicate the method of detrending used in the first step of the bootstrap. For GLS detrending we use  $\bar{c} = 13.5$ .<sup>6</sup>

The DGP we use in our simulations is almost identical to the one given in (1), except that we restrict  $u_t$  to be a (stationary and invertible) ARMA(1,1) process, and we generalize the initial condition. The DGP is given below:

$$\begin{aligned} y_t &= x_t + \beta' z_t, & x_t &= \rho x_{t-1} + u_t \\ u_t &= \phi u_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}, \end{aligned} \tag{4}$$

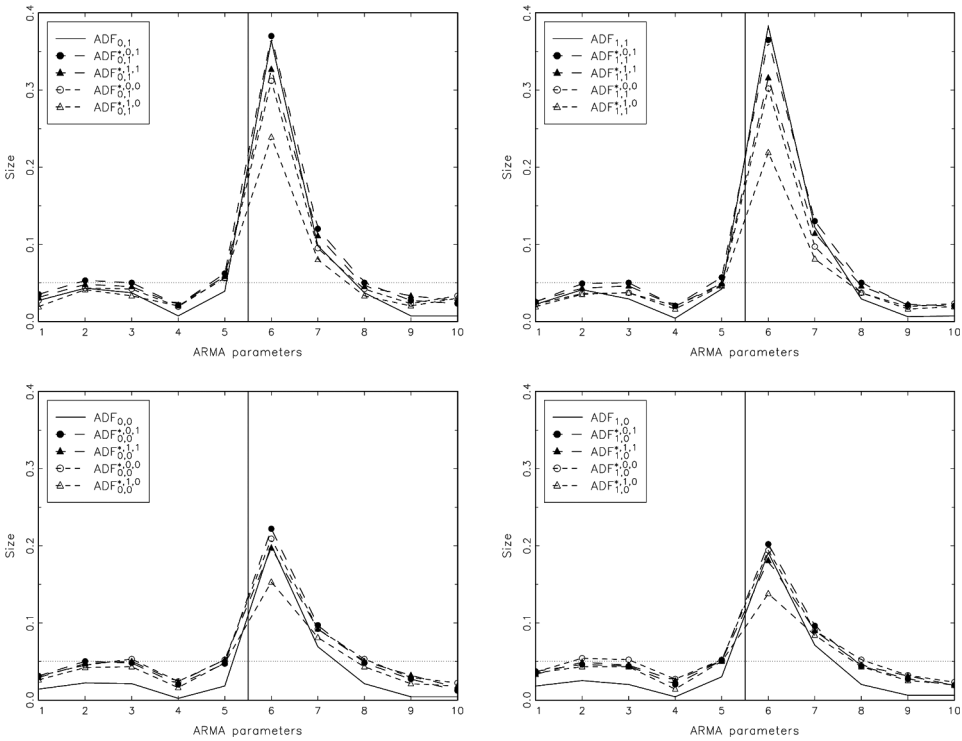
where  $\varepsilon_t \sim N(0,1)$  and  $\rho = 1 - cT^{-1}$ . We set the true deterministic components equal to zero (take  $\beta = (0,0)'$ ); as we perform all tests under the assumption that  $z_t = (1, t)$ , all tests are invariant to the true value of  $\beta$ . For the size analysis we set  $x_0 = 0$  without loss of generality. For the power analysis we consider both a small and large initial condition; that is, we follow Harvey et al. (2009) and set  $x_0 = a\sqrt{\omega_u/(1 - \rho^2)}$ , where  $\omega_u = \lim_{T \rightarrow \infty} T^{-1}E(\sum_{t=1}^T u_t)^2$ . We set  $a = 0$  which gives a zero initial condition, and  $a = 2.5$ , a value that gives a clear power advantage to the OLS test in Harvey et al. (2009).

Lag lengths  $p$ ,  $q$  and  $p^*$  are selected separately using the modified Akaike information criterion (MAIC) (Ng and Perron, 2001). As suggested by Perron and Qu (2007), lag lengths  $p$  and  $p^*$  are always determined from OLS detrended data. All results are obtained using 1000 simulations and 499 bootstrap replications. For the asymptotic tests, we use small sample critical values. The nominal level is taken to be 0.05 everywhere.

## 4.2. Simulation Results

Figures 1 and 2 present results for size ( $c = 0$ ) for  $T = 50$  and  $T = 100$ , respectively. The figures are split according to the test statistic used, e.g., the top left cell consists of all tests that use  $ADF_{0,1}$ , the ADF

<sup>6</sup>Elliott et al. (1996) suggest and investigate the optimality of this value for full sample GLS detrending. There is however no reason why this value should be optimal for recursive GLS detrending. However, we will use it as it is a well accepted value in the literature. Moreover, a study into the optimal value for  $\bar{c}$  is outside the scope of this paper. Broda et al. (2009) go into more detail.

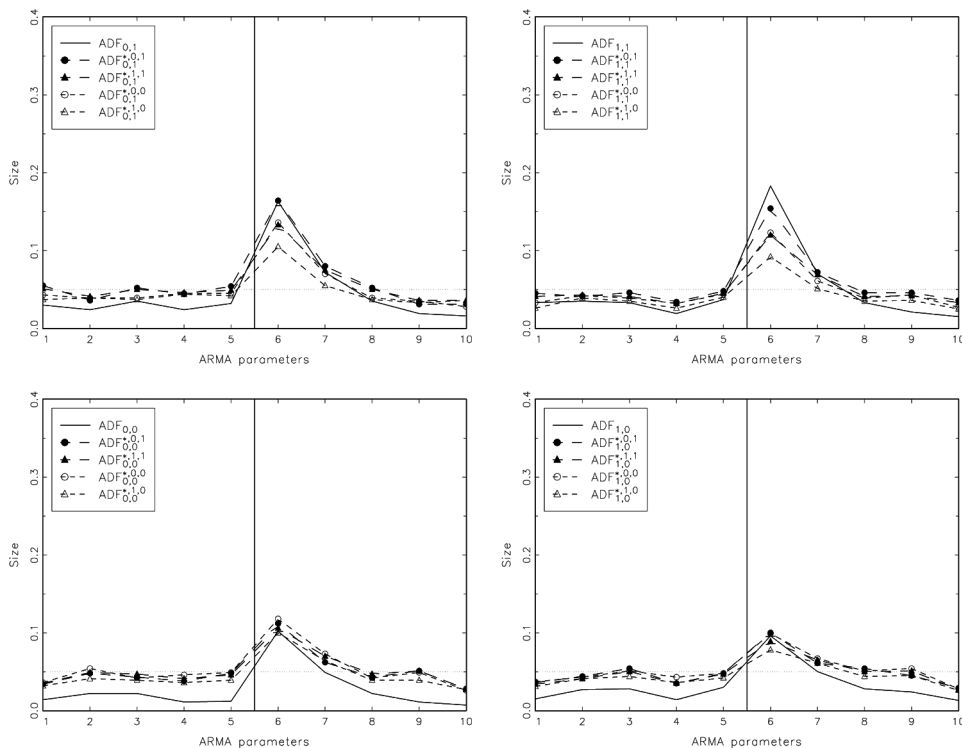


<sup>a</sup> ARMA parameters 1-5:  $\theta = 0$  and  $\phi = -0.8, -0.4, 0, 0.4, 0.8$ ; 6-10:  $\phi = 0$  and  $\theta = -0.8, -0.4, 0, 0.4, 0.8$ . Test statistics:  $ADF_{\gamma, \tilde{\gamma}, \lambda}^{*}$ , where  $\gamma$  (and  $\tilde{\gamma}$ ) and  $\lambda$  (and  $\tilde{\lambda}$ ) denote the type of detrending used to obtain the test statistic (and in the first step of the bootstrap):  $\gamma = 0$  OLS,  $\gamma = 1$  GLS,  $\lambda = 0$  recursive,  $\lambda = 1$  full sample detrending.

FIGURE 1 Size,  $T = 50$ .

test based on full sample OLS detrending, as test statistic, and they differ in the method that is applied in the first step of the bootstrap. The numbers 1, 2, 3, 4, 5 on the x-axis correspond to setting the autoregressive moving average (ARMA) parameters in (4) as  $\theta = 0$  and  $\phi = -0.8, -0.4, 0, 0.4, 0.8$ , respectively, while 6, 7, 8, 9, 10 correspond to  $\phi = 0$  and  $\theta = -0.8, -0.4, 0, 0.4, 0.8$ , respectively (the value for  $\phi = \theta = 0$  is given in both parts of the graph).

The asymptotic tests are undersized for most parameter combinations, while there is the familiar oversize for negative moving average (MA) parameters. Note that the recursively detrended tests have smaller size distortions in this case, although for other parameter combinations they have more undersize. The bootstrap tests appear to be less sensitive than the asymptotic tests to the values of the ARMA parameters, and have size close to the nominal level in general. The exception is the DGP with the large negative MA parameter, where there is still oversize, although generally less than for the asymptotic tests. What is quite noticeable is that the bootstrap corrects most of the undersize of the asymptotic tests.



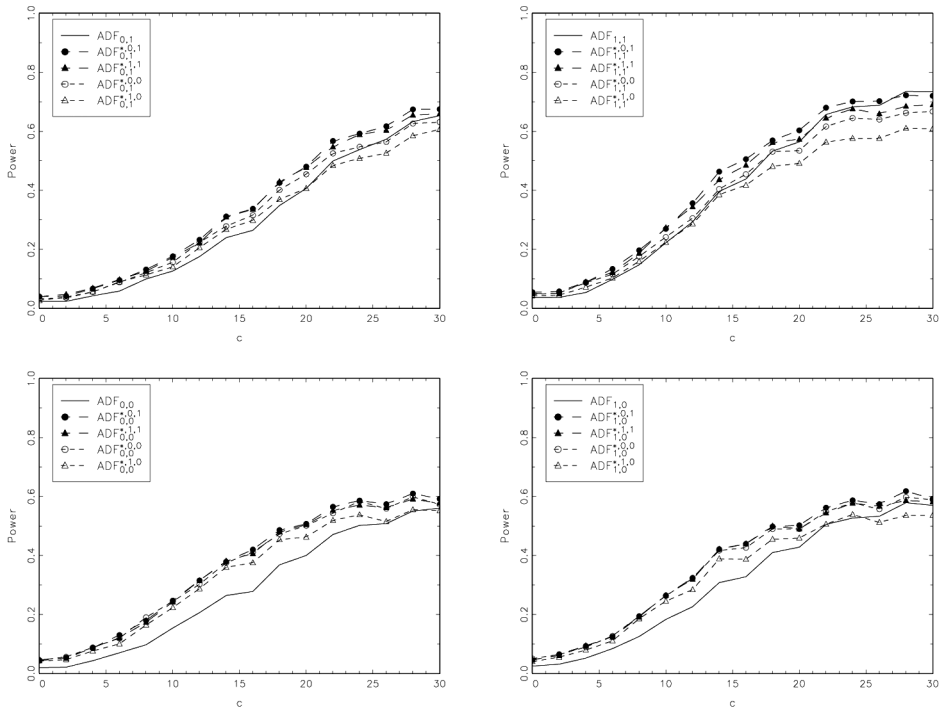
<sup>a</sup> ARMA parameters 1-5:  $\theta = 0$  and  $\phi = -0.8, -0.4, 0, 0.4, 0.8$ ; 6-10:  $\phi = 0$  and  $\theta = -0.8, -0.4, 0, 0.4, 0.8$ . Test statistics:  $ADF_{\gamma, \lambda}^{*, \tilde{\gamma}, \tilde{\lambda}}$ , where  $\gamma$  (and  $\tilde{\gamma}$ ) and  $\lambda$  (and  $\tilde{\lambda}$ ) denote the type of detrending used to obtain the test statistic (and in the first step of the bootstrap):  $\gamma = 0$  OLS,  $\gamma = 1$  GLS,  $\lambda = 0$  recursive,  $\lambda = 1$  full sample detrending.

**FIGURE 2** Size,  $T = 100$ .

The effects of the detrending method used for the calculation of the test statistic follow that of the corresponding asymptotic tests, although less pronounced. Considering the method of detrending in the first step of the bootstrap, we see that the recursively detrended tests, in particular the GLS version, have a tendency to reject less often than the full sample detrended tests. With the exception of the models with a negative MA parameter, where recursive GLS detrending leads to size closest to nominal, this leads to undersize, which could affect power negatively. Full sample detrending, in particular using OLS, in the first step of the bootstrap leads to size closest to nominal although it has larger size distortions if there is a large negative MA parameter.

We next consider the power properties of the tests. For this purpose we report (unadjusted) power curves for the DGP with i.i.d. errors.<sup>7</sup> We investigate the power properties of the tests for both a small (zero) and

<sup>7</sup>We consider i.i.d. errors to avoid size distortions caused by the ARMA parameters. These curves are representative though for the other models.



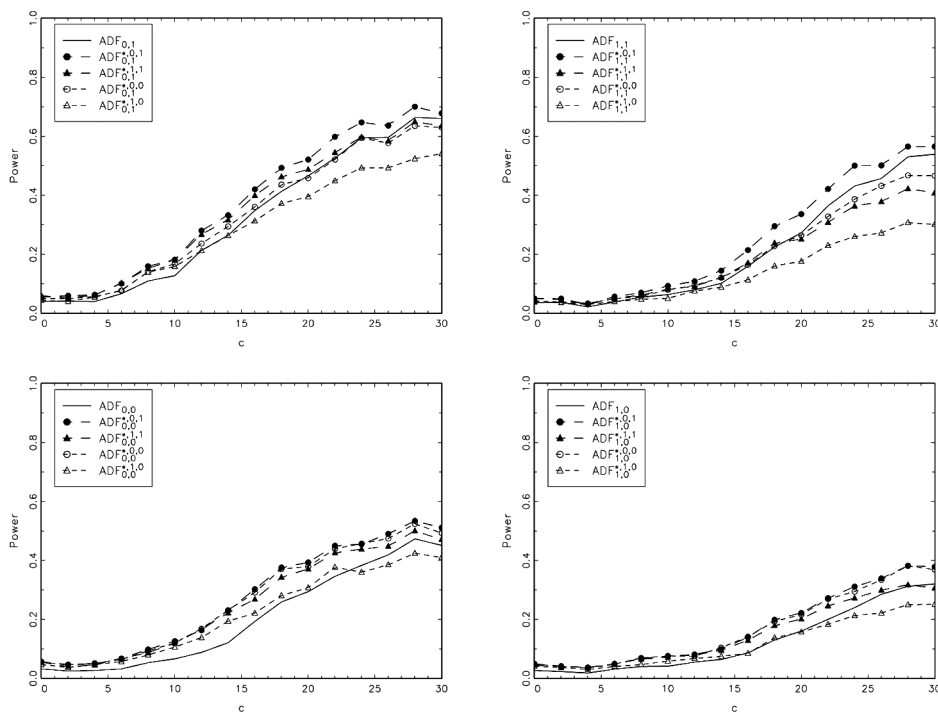
<sup>a</sup> Test statistics:  $ADF_{\gamma, \lambda}^{*, \tilde{\gamma}, \tilde{\lambda}}$ , where  $\gamma$  (and  $\tilde{\gamma}$ ) and  $\lambda$  (and  $\tilde{\lambda}$ ) denote the type of detrending used to obtain the test statistic (and in the first step of the bootstrap):  $\gamma = 0$  OLS,  $\gamma = 1$  GLS,  $\lambda = 0$  recursive,  $\lambda = 1$  full sample detrending.

**FIGURE 3** Power curves for i.i.d. errors,  $a = 0$ ,  $T = 100$ .

large initial condition, as it is known from the literature that the initial condition has a major impact on the relative power of tests using OLS and GLS detrending (cf. Harvey et al., 2009). Fig. 3 gives the power curves for a zero initial condition, while Fig. 4 gives the curves for a large initial condition ( $a = 2.5$ ). For both cases we take  $T = 100$ .

As expected for the asymptotic tests, the power curves show that for a small initial condition, full sample GLS detrending is the most powerful, while for a large initial condition, full sample OLS detrending is the most powerful. The power of the recursively detrended tests is smaller than that of their full sample counterparts. The bootstrap tests follow their asymptotic counterparts closely regarding the detrending method used for the calculation of the test statistic, but the curves also show that in general the bootstrap tests tend to be more powerful than the asymptotic tests, which is likely caused by the fact that the asymptotic tests are undersized. The power of the bootstrap tests also clearly depends on the method of detrending used within the bootstrap. The tests that use full sample OLS detrending in the first step of the bootstrap have the highest power,





<sup>a</sup> Test statistics:  $ADF_{\gamma,\lambda}^{*,\tilde{\gamma},\tilde{\lambda}}$ , where  $\gamma$  (and  $\tilde{\gamma}$ ) and  $\lambda$  (and  $\tilde{\lambda}$ ) denote the type of detrending used to obtain the test statistic (and in the first step of the bootstrap):  $\gamma = 0$  OLS,  $\gamma = 1$  GLS,  $\lambda = 0$  recursive,  $\lambda = 1$  full sample detrending.

**FIGURE 4** Power curves for i.i.d. errors,  $a = 2.5$ ,  $T = 100$ .

while the power of recursive GLS detrending is lowest. Full sample GLS detrending is very similar to OLS for a small initial condition, but leads to lower power for a large initial condition.

These results show that in the first step of the bootstrap full sample OLS detrending should be preferred from a power perspective. The only downside to full sample OLS detrending is its size for negative MA parameters, but given the power properties this does not seem to justify using another detrending method in the first step of the bootstrap. The major determinant of this result appears to be the selection of the lag length  $q$ . Under recursive detrending, structurally a higher lag length is selected than under full sample detrending. Also, a higher lag length is selected under full sample GLS detrending than under full sample OLS detrending.<sup>8</sup> Unreported simulations show that if the lag length is selected in the same way for all methods, the differences disappear. Therefore,

<sup>8</sup>Average selected lags are available on request.

it mainly seems to be important to let the selection of  $q$  be based on full sample OLS detrended data.

While the detrending in the first step of the bootstrap clearly matters, overall it only has a minor impact on the power properties of the bootstrap tests, as these are mainly determined by the properties of the corresponding asymptotic test. Therefore, just as for the asymptotic tests, we can conclude that the GLS full sample detrended test is the preferred test if the initial condition is small, while the OLS full sample detrended test is to be preferred if the initial condition is large. Recursive detrending does not outperform full sample detrending. Of course, the initial condition will usually be unknown and therefore this conclusion may not be very helpful. However, the same conclusion holds if asymptotic tests are applied, and as the properties of the asymptotic tests carry over to the bootstrap setting, any solution to this problem of the initial condition for asymptotic tests (such as considered in Harvey et al., 2009, and Harvey et al., 2012) will remain valid in a bootstrap context (as in Smeekes and Taylor, 2012).

## 5. CONCLUSION

We have investigated the role of detrending in bootstrap unit root tests. We have pointed out that the method of detrending used for the construction of the test statistic does not have to be the same as the method of detrending performed in the first step of the bootstrap algorithm. The bootstrap has been shown to be valid for a wide range of possible detrending methods, irrespective of the method used in the construction of the test statistic.

A simulation study has been conducted to investigate the impact of detrending on the size and power properties of the bootstrap unit root tests. The first important conclusion is that in the first step of the bootstrap algorithm, full sample OLS detrending outperforms the other methods in terms of power, and this is mainly caused by the lag selection in the bootstrap. The second important conclusion is that the method of detrending used for the construction of the test statistic has a major impact on the power of the test. Moreover, the power properties of the bootstrap tests are determined by the power properties of their asymptotic counterparts, although the bootstrap tests are in general more powerful.

These two conclusions have the following implications. First, the choice of detrending used in the first step of the bootstrap algorithm should be seen separately from the choice of the detrending method for the test statistic. Our simulation study indicates that full sample OLS detrending should be preferred here, or at least that the lags used in the ADF regression in the second step of the bootstrap should be selected from full sample OLS detrended data. Second, the choice of the detrending method used in the construction of the test statistic should be based on

power considerations. As the power properties of the asymptotic tests carry over to the bootstrap setting, the choice of the detrending method for the bootstrap tests should be based on the same considerations as for the asymptotic tests, that is, GLS detrending should be preferred if the initial condition is small while OLS detrending should be preferred if the initial condition is large.

There are several extensions possible to this paper. First, one could consider alternative methods of detrending. We have limited our analysis to OLS and GLS detrending, but one can easily imagine other methods. Second, one could extend the analysis to other types of unit root tests. Examples of these points are given in Leybourne et al. (2005). Finally, one could view detrending in a broader perspective and analyze more general trends, such as polynomial trends of higher order or broken trends.

**APPENDIX: PROOFS**

*Proof of Theorem 1.* It follows directly from results in Elliott et al. (1996) and Taylor (2002) that

$$T^{-1/2}y_{[T\tau]}^d \xrightarrow{d} \psi(1)\sigma W_\gamma(r, \lambda).$$

By Assumption 1, we may define  $\phi(z) = \psi(z)^{-1} = 1 - \sum_{j=1}^\infty \phi_j z^j$ . Now define  $\varepsilon_{p,t} = u_t - \sum_{j=1}^p \phi_j u_{t-j} = \varepsilon_t + \sum_{j=p+1}^\infty \phi_j u_{t-j}$ . As  $y_t^d = x_t - (\hat{\beta}_t - \beta)'z_t$ , we can write

$$\varepsilon_{p,t} = \Delta y_t^d - \sum_{j=1}^p \phi_j \Delta y_{t-j}^d + \Delta[(\hat{\beta}_t - \beta)'z_t] - \sum_{j=1}^p \phi_j \Delta[(\hat{\beta}_{t-j} - \beta)'z_{t-j}],$$

where  $\Delta[a_t b_t] = a_t b_t - a_{t-1} b_{t-1}$  for any sequences  $a_t$  and  $b_t$ . Then, letting  $\phi_p(z) = 1 - \sum_{j=1}^p \phi_j z^j$ , we can write  $\varepsilon_{p,t}^d = \varepsilon_{p,t} - \phi_p(L)\Delta[(\hat{\beta}_t - \beta)'z_t]$  such that  $\Delta y_t^d = \sum_{j=1}^p \phi_j \Delta y_{t-j}^d + \varepsilon_{p,t}^d$ . Similarly, we can define  $\varepsilon_t^d$  such that  $\varepsilon_t^d = \Delta y_t^d - \sum_{j=1}^\infty \phi_j \Delta y_{t-j}^d = \varepsilon_t - \phi(L)\Delta[(\hat{\beta}_t - \beta)'z_t]$ .

Letting  $\Delta Y^d = (\Delta y_1^d, \dots, \Delta y_T^d)'$ ,  $Y_{-1}^d = (y_0^d, \dots, y_{T-1}^d)'$ ,  $w_{p,t}^d = (\Delta y_{t-1}^d, \dots, \Delta y_{t-p}^d)'$ ,  $M_p^d = (w_{p,1}^d, \dots, w_{p,T}^d)'$ ,  $\Phi_p = (\phi_1, \dots, \phi_p)'$ , and  $\varepsilon_p^d = (\varepsilon_{p,1}^d, \dots, \varepsilon_{p,T}^d)'$ , we have  $\Delta Y^d = M_p^d \Phi_p + \varepsilon_p^d$  and

$$\begin{aligned} A_T &= Y_{-1}^{d'} \Delta Y^d - Y_{-1}^{d'} M_p^d \left( M_p^{d'} M_p^d \right)^{-1} M_p^{d'} \Delta Y^d \\ &= Y_{-1}^{d'} \varepsilon_p^d - Y_{-1}^{d'} M_p^d \left( M_p^{d'} M_p^d \right)^{-1} M_p^{d'} \varepsilon_p^d, \\ B_T &= Y_{-1}^d Y_{-1}^d - Y_{-1}^{d'} M_p^d \left( M_p^{d'} M_p^d \right)^{-1} M_p^{d'} Y_{-1}^d, \\ \hat{\sigma}^2 &= T^{-1} (\Delta Y^d - Y_{-1}^d \hat{\delta})' (I - M_p^d (M_p^{d'} M_p^d)^{-1} M_p^{d'}) (\Delta Y^d - Y_{-1}^d \hat{\delta}), \end{aligned}$$

from which we can now construct the ADF statistic  $ADF_{\gamma,\lambda} = \hat{\delta} \left[ \hat{\sigma}^2 \widehat{\text{Var}}(\hat{\delta}) \right]^{-1/2} = \hat{\sigma}^{-1} A_T B_T^{-1/2}$ .

It can next be shown, in similar spirit as in Chang and Park (2002), that

$$\begin{aligned} T^{-2} Y_{-1}^{d'} Y_{-1}^d &\xrightarrow{d} \psi(1)^2 \sigma^2 \int_0^1 W_\gamma(r, \lambda)^2 dr, \\ T^{-1} Y_{-1}^{d'} \varepsilon_p^d &\xrightarrow{d} \frac{1}{2} \psi(1) \sigma^2 (W_\gamma(1, \lambda)^2 - W_\gamma(0, \lambda)^2 - 1), \\ \left\| (T^{-1} M_p^{d'} M_p^d)^{-1} \mathbf{1} \right\| &= O_p(1), \quad \left| T^{-1} Y_{-1}^{d'} M_p^d \right| = O_p(p^{1/2}), \\ \left| T^{-1} M_p^{d'} \varepsilon_p^d \right| &= O_p(p^{-1/2}), \end{aligned} \tag{A.1}$$

and  $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$ . The proofs of Lemma 1 and 2 in Smeekes (2009b) present details. With these results the limit distributions follow straightforwardly.  $\square$

**Proof of Lemma 1.** Let  $\bar{\Phi}_q$  denote the vector of OLS estimators  $\bar{\phi}_1, \dots, \bar{\phi}_q$  in a regression of  $\Delta y_t^d$  on  $\Delta y_{t-1}^d, \dots, \Delta y_{t-q}^d$  (hence imposing the null hypothesis of a unit root). Then

$$\hat{\Phi}_q = \bar{\Phi}_q + (M_q^{d'} M_q^d)^{-1} M_q^{d'} Y_{-1}^d \hat{\delta} = \bar{\Phi}_q + O_p(T^{-1} q^{1/2}).$$

The next step is to show that  $\bar{\Phi}_q = \tilde{\Phi}_q + O_p(T^{-1} q^{1/2})$ . Let  $\Delta \tilde{B}z = (\Delta[(\tilde{\beta}_1 - \beta)' z_1], \dots, \Delta[(\tilde{\beta}_T - \beta)' z_T])'$ , such that  $\varepsilon_q^d = \varepsilon_q - \phi_q(L) \Delta \tilde{B}z$ . Also define  $w_{q,t} = (u_{t-1}, \dots, u_{t-q})'$  and  $M_q = (w_{q,1}, \dots, w_{q,T})'$ . Now note that

$$\begin{aligned} \bar{\Phi}_q - \tilde{\Phi}_q &= (M_q^{d'} M_q^d)^{-1} M_q^{d'} \varepsilon_q^d - (M_q' M_q)^{-1} M_q' \varepsilon_q \\ &= [(M_q^{d'} M_q^d)^{-1} - (M_q' M_q)^{-1}] M_q' \varepsilon_q - (M_q^{d'} M_q^d)^{-1} \left[ M_q' \phi_q(L) \Delta \tilde{B}z \right. \\ &\quad \left. + \Delta \tilde{B}z' \varepsilon_q - \Delta \tilde{B}z' \phi_q(L) \Delta \tilde{B}z \right] \\ &= A_T + B_T + C_T + D_T. \end{aligned}$$

Now define  $\Delta \tilde{B}z_q$  as a  $T \times q$  matrix with element  $(i, j)$  as  $\Delta \tilde{B}z_q^{(i,j)} = \Delta[(\tilde{\beta}_{i-j} - \beta)' z_{i-j}]$ . Then

$$\begin{aligned} \|A_T\| &\leq 2 \left\| T^{-1} M_q' \varepsilon_q \right\| \left\| (T^{-1} M_q^{d'} M_q^d)^{-1} \right\| \left\| (T^{-1} M_q' M_q)^{-1} \right\| \\ &\quad \times \left( \left\| T^{-1} M_q' \Delta \tilde{B}z_q \right\| + \left\| \Delta \tilde{B}z_q' \Delta \tilde{B}z_q \right\| \right) = o_p(T^{-1} q^{1/2}), \end{aligned}$$

$$\begin{aligned} \|B_T\| &\leq \left\| (T^{-1}M_q^{d'}M_q^d)^{-1} \right\| \left| T^{-1}M_q'\phi_q(L)\Delta\tilde{B}z \right| = O_p(T^{-1}q^{1/2}), \\ \|C_T\| &\leq \left\| (T^{-1}M_q^{d'}M_q^d)^{-1} \right\| \left| T^{-1}\Delta\tilde{B}z'\varepsilon_q \right| = O_p(T^{-1}q^{1/2}), \\ \|D_T\| &\leq \left\| (T^{-1}M_q^{d'}M_q^d)^{-1} \right\| \left| T^{-1}\Delta\tilde{B}z'\phi_q(L)\Delta\tilde{B}z \right| = O_p(T^{-1}q^{1/2}), \end{aligned}$$

which follows directly from the proof of (A.1) as shown in the proof of Lemma 3 in Smeekes (2009b). Hence we may conclude that  $\hat{\phi}_j = \tilde{\phi}_j + O_p(T^{-1}q^{1/2})$  uniformly in  $j$ ,  $1 \leq j \leq q$ .  $\square$

**Proof of Theorem 2.** We must first show that the following invariance principle for  $u_t^*$  holds:

$$T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} u_t^* \xrightarrow{d^*} \sigma\psi(1)W(r) \text{ in probability.} \tag{A.2}$$

The first step towards proving (A.2) is to show that  $E^*|\varepsilon_t^*|^{2+\epsilon} < \infty$  for some  $\epsilon > 0$ . Define  $\tilde{\varepsilon}_{q,t}$  as the residuals of a regression of  $u_t$  on  $u_{t-1}, \dots, u_{t-q}$ . We then have for any  $2 < a \leq 4$

$$\begin{aligned} E^*|\varepsilon_t^*|^a &\leq 2^{a-1}T^{-1} \sum_{t=1}^T \left| \hat{\varepsilon}_{q,t}^d - \tilde{\varepsilon}_{q,t} - T^{-1} \sum_{\tau=1}^T (\hat{\varepsilon}_{q,\tau}^d - \tilde{\varepsilon}_{q,\tau}) \right|^a \\ &\quad + 2^{a-1}T^{-1} \sum_{t=1}^T \left| \tilde{\varepsilon}_{q,t} - T^{-1} \sum_{\tau=1}^T \tilde{\varepsilon}_{q,\tau} \right|^a \end{aligned}$$

so that we can write  $E^*|\varepsilon_t^*|^a \leq 2^{a-1}(R_{1,T} + R_{2,T})$ . First note that

$$\begin{aligned} R_{1,T} &\leq 2^{a-1}T^{-1} \sum_{t=1}^T \left| \hat{\varepsilon}_{q,t}^d - \tilde{\varepsilon}_{q,t} \right|^a + 2^{a-1} \left| T^{-1} \sum_{t=1}^T (\hat{\varepsilon}_{q,t}^d - \tilde{\varepsilon}_{q,t}) \right|^a \\ &= 2^{a-1}(R_{11,T} + R_{12,T}). \end{aligned}$$

As  $\hat{\varepsilon}_{q,t}^d - \tilde{\varepsilon}_{q,t} = -\hat{\delta}y_{t-1}^d + \sum_{j=1}^q(\hat{\phi}_j - \tilde{\phi}_j)u_{t-j} - \hat{\phi}_q(L)\Delta[(\tilde{\beta}_t - \beta)'z_t]$ , we have that

$$\begin{aligned} R_{11,T} &\leq 3^{a-1} \left( |\hat{\delta}|^a T^{-1} \sum_{t=1}^T |y_{t-1}^d|^a + T^{-1} \sum_{t=1}^T \left| \sum_{j=1}^q (\hat{\phi}_j - \tilde{\phi}_j)u_{t-j} \right|^a \right. \\ &\quad \left. + T^{-1} \sum_{t=1}^T |\hat{\phi}_q(L)\Delta[(\tilde{\beta}_t - \beta)'z_t]|^a \right). \end{aligned}$$

It follows that  $R_{11,T} = o_p(1)$  as  $|\hat{\delta}|^a T^{-1} \sum_{t=1}^T |y_{t-1}^d|^a = o_p(1)$ ,

$$\begin{aligned} T^{-1} \sum_{t=1}^T \left| \sum_{j=1}^q (\hat{\phi}_j - \tilde{\phi}_j) u_{t-j} \right|^a &\leq T^{-1} \max_{1 \leq j \leq q} |\hat{\phi}_j - \tilde{\phi}_j|^a \sum_{t=1}^T \sum_{j=1}^q |u_{t-j}|^a \\ &= O_p(T^{-a} q^{a/2+1}), \end{aligned}$$

and finally, defining  $\hat{\phi}_0 = -1$  and realizing that  $\Delta[(\tilde{\beta}_{t-j} - \beta)' z_{t-j}] = (\tilde{\beta}_{2,t-j} - \beta_2) + o_p(1)$ ,

$$\begin{aligned} T^{-1} \sum_{t=1}^T \left| \sum_{j=0}^q \hat{\phi}_j \Delta[(\tilde{\beta}_{t-j} - \beta)' z_{t-j}] \right|^a &\leq \max_{1 \leq t \leq T} |\Delta[(\tilde{\beta}_{t-j} - \beta)' z_{t-j}]|^a \left| \sum_{j=0}^q \hat{\phi}_j \right|^a \\ &= O_p(T^{-a/2}). \end{aligned}$$

It follows similarly that  $R_{12,T} = o_p(1)$ . It then follows from Lemma 3.2 in Park (2002) that  $R_{2,T} = O_p(1)$ , and consequently that  $E^*|\varepsilon_t^*|^a = O_p(1)$  and  $T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \varepsilon_t^* \xrightarrow{d^*} \sigma W(r)$  in probability.

To prove the invariance principle for  $u_t^*$  we can apply the Beveridge–Nelson decomposition as in (Park, 2002, p. 478). We then need to show that

$$\hat{\phi}(1) \xrightarrow{p} \phi(1), \quad \mathbf{P}^* \left\{ \max_{1 \leq t \leq T} |T^{-1/2} \bar{u}_t^*| > \epsilon \right\} = o_p(1), \quad (\text{A.3})$$

where  $\bar{u}_t^* = \hat{\phi}(1)^{-1} \sum_{i=1}^q (\sum_{j=i}^q \hat{\phi}_j) u_{t-i+1}^*$ . The first result in (A.3) follows as

$$\left| \hat{\phi}(1) - \phi(1) \right| \leq \left| \hat{\phi}(1) - \tilde{\phi}(1) \right| + \left| \tilde{\phi}(1) - \phi(1) \right| = O_p(T^{-1} q^{3/2}) + o_p(1),$$

where the first part follows from Lemma 1 and the second part from Lemma 3.1 in Park (2002). To prove the second result in (A.3) we need to show that  $\sum_{j=1}^q j^{1/2} |\hat{\phi}_j| = O_p(1)$ , (see Palm et al., 2010, Proof of Theorem 2, p. 671). We can write

$$\sum_{j=1}^q j^{1/2} |\hat{\phi}_j| \leq \sum_{j=1}^q j^{1/2} |\hat{\phi}_j - \tilde{\phi}_j| + \sum_{j=1}^q j^{1/2} |\tilde{\phi}_j| = O_p(T^{-1} q^2) + O_p(1) = O_p(1),$$

where the first part follows from Lemma 1 and the second part follows from the proof of Theorem 2 in Palm et al. (2010). This completes the proof of (A.2). It then follows straightforwardly from (A.2) that  $T^{-1/2} y_{\lfloor Tr \rfloor, \gamma, \lambda}^{*d} \xrightarrow{d^*} \sigma \psi(1) W_\gamma(r, \lambda)$  in probability.

In analogy with the original sample, define  $\varepsilon_{p,t}^* = u_t^* - \sum_{j=1}^p \hat{\phi}_j u_{t-j}^* = \varepsilon_t^* + \sum_{j=p+1}^q \hat{\phi}_j u_{t-j}^*$ . However, it is clear from our Assumption 3(ii) that there is some  $T_0$  such that for all  $T > T_0$  one obtains  $\varepsilon_{p,t}^* = \varepsilon_t^*$ .<sup>9</sup> Therefore our proofs can proceed as if we set  $\varepsilon_{p,t}^* = \varepsilon_t^*$ . In analogy with the original sample we can derive that  $\Delta y_t^{*d} = \sum_{j=1}^p \hat{\phi}_j \Delta y_{t-j}^{*d} + \varepsilon_{p,t}^{*d}$ , where, letting  $\hat{\phi}_p(z) = 1 - \sum_{j=1}^p \hat{\phi}_j z^j$ ,  $\varepsilon_{p,t}^{*d} = \varepsilon_{p,t}^* - \hat{\phi}_p(L) \Delta[(\hat{\beta}_t^* - \beta^*)' z_t]$ . Similarly, we can define  $\varepsilon_t^{*d} = \varepsilon_t^* - \hat{\phi}(L) \Delta[(\hat{\beta}_t^* - \beta^*)' z_t]$ . It will then also be clear that for large  $T$  we have  $\varepsilon_{p,t}^{*d} = \varepsilon_t^{*d}$ . Now let  $\Delta Y^{*d}$ ,  $Y_{-1}^{*d}$ ,  $M_p^{*d}$ ,  $\varepsilon_p^{*d}$ , and  $\hat{\Phi}_p$  be defined analogously as their original sample counterparts. Then  $\Delta Y^{*d} = M_p^{*d} \hat{\Phi}_p + \varepsilon_p^{*d}$  and

$$\begin{aligned} A_T^* &= Y_{-1}^{*d'} \varepsilon_p^{*d} - Y_{-1}^{*d'} M_p^{*d} \left( M_p^{*d'} M_p^{*d} \right)^{-1} M_p^{*d'} \varepsilon_p^{*d}, \\ B_T^* &= Y_{-1}^{*d'} Y_{-1}^{*d} - Y_{-1}^{*d'} M_p^{*d} \left( M_p^{*d'} M_p^{*d} \right)^{-1} M_p^{*d'} Y_{-1}^{*d}, \\ \hat{\sigma}^{*2} &= T^{-1} (\Delta Y^{*d} - Y_{-1}^{*d} \hat{\delta}^*)' (I - M_p^{*d} (M_p^{*d'} M_p^{*d})^{-1} M_p^{*d'}) (\Delta Y^{*d} - Y_{-1}^{*d} \hat{\delta}^*), \end{aligned}$$

such that  $ADF_{\gamma,\lambda}^* = \hat{\delta}^* \left[ \hat{\sigma}^{*2} \widehat{\text{Var}}(\hat{\delta}^*) \right]^{-1/2} = \hat{\sigma}^{*-1} A_T^* B_T^{*-1/2}$ .

Using the results in Chang and Park (2003), we can then show the following results in similar fashion as their non-bootstrap counterparts (Lemma 5 of Smeekes, 2009b, provides details):

$$\begin{aligned} T^{-2} Y_{-1}^{*d'} Y_{-1}^{*d} &\xrightarrow{d^*} \psi(1)^2 \sigma^2 \int_0^1 W_\gamma(r, \lambda)^2 dr, \\ T^{-1} Y_{-1}^{*d'} \varepsilon_p^{*d} &\xrightarrow{d^*} \frac{1}{2} \psi(1) \sigma^2 (W_\gamma(1, \lambda)^2 - W_\gamma(0, \lambda)^2 - 1), \end{aligned}$$

$$\begin{aligned} \|(T^{-1} M_p^{*d'} M_p^{*d})^{-1}\| &= O_p^*(1), \quad |T^{-1} Y_{-1}^{*d'} M_p^{*d}| = O_p^*(p^{1/2}), \\ |T^{-1} M_p^{*d'} \varepsilon_p^{*d}| &= O_p^*(T^{-1/2} p^{1/2}). \end{aligned}$$

It also follows in the same way that  $\hat{\sigma}^{*2} \xrightarrow{p^*} \sigma^2$ . The bootstrap limit distributions can then easily be derived. □

### ACKNOWLEDGMENTS

This research was financially supported by the Netherlands Organisation for Scientific Research (NWO). I would like to thank the associate editor, two anonymous referees, Christoph Hanck, Franz

<sup>9</sup>Note that if  $q < p$ ,  $\hat{\phi}_j = 0$  for  $j = q + 1, \dots, p$  and therefore  $\varepsilon_{p,t}^* = \varepsilon_t^*$ .

Palm, Anders Swensen, Robert Taylor and Jean-Pierre Urbain for helpful comments and suggestions.

## REFERENCES

- Broda, S., Carstensen, K., Paoletta, M. S. (2009). Assessing and improving the performance of nearly efficient unit root tests in small samples. *Econometric Reviews* 28:468–494.
- Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli* 3:123–148.
- Burridge, P., Taylor, A. M. R. (2004). Bootstrapping the HEGY seasonal unit root tests. *Journal of Econometrics* 123:67–87.
- Cavaliere, G., Taylor, A. M. R. (2008). Bootstrap unit root tests for time series with nonstationary volatility. *Econometric Theory* 24:43–71.
- Cavaliere, G., Taylor, A. M. R. (2009). Bootstrap  $M$  unit root tests. *Econometric Reviews* 28:393–421.
- 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.
- Dickey, D. A., Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* 49:1057–1072.
- Elliott, G., Müller, U. K. (2006). Minimizing the impact of the initial condition on testing for unit roots. *Journal of Econometrics* 135:285–310.
- Elliott, G., Rothenberg, T. J., Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica* 64:813–836.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.
- Hannan, E. J., Kavalieris, L. (1986). Regression, autoregression models. *Journal of Time Series Analysis* 7:27–49.
- Harvey, D. I., Leybourne, S. J., Taylor, A. M. R. (2009). Unit root testing in practice: dealing with uncertainty over the trend and initial condition. *Econometric Theory* 25:587–636.
- Harvey, D. I., Leybourne, S. J., Taylor, A. M. R. (2012). Testing for unit roots in the presence of uncertainty over both the trend and initial condition. *Journal of Econometrics* 169:188–195.
- Leybourne, S., Kim, T.-H., Newbold, P. (2005). Examination of some more powerful modifications of the Dickey–Fuller test. *Journal of Time Series Analysis* 26:355–369.
- Müller, U. K., Elliott, G. (2003). Tests for unit roots and the initial condition. *Econometrica* 71:1269–1286.
- Ng, S., Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69:1519–1554.
- Palm, F. C., Smeekes, S., Urbain, J.-P. (2008). Bootstrap unit root tests: comparison and extensions. *Journal of Time Series Analysis* 29:371–401.
- Palm, F. C., Smeekes, S., Urbain, J.-P. (2010). A sieve bootstrap test for cointegration in a conditional error correction model. *Econometric Theory* 26:647–681.
- 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.
- Parker, C., Paparoditis, E., Politis, D. N. (2006). Unit root testing via the stationary bootstrap. *Journal of Econometrics* 133:601–638.
- Park, J. Y. (2002). An invariance principle for sieve bootstrap in time series. *Econometric Theory* 18:469–490.
- Perron, P., Qu, Z. (2007). A simple modification to improve the finite sample properties of Ng and Perron’s unit root tests. *Economics Letters* 94:12–19.
- Phillips, P. C. B., Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika* 75:335–346.
- Phillips, P. C. B., Solo, V. (1992). Asymptotics for linear processes. *Annals of Statistics* 20:971–1001.
- Poskitt, D. S. (1994). A note on autoregressive modeling. *Econometric Theory* 10:884–899.



- Richard, P. (2007). Sieve bootstrap unit root tests. GREDI Working Paper 07-05, Université de Sherbrooke.
- Richard, P. (2009). Modified fast double sieve bootstraps for ADF tests. *Computational Statistics and Data Analysis* 53:4490–4499.
- Rodrigues, P. M. M. (2006). Properties of recursive trend-adjusted unit root tests. *Economics Letters* 91:413–419.
- Shin, D. W., So, B. S. (2001). Recursive mean adjustment for unit root tests. *Journal of Time Series Analysis* 22:595–612.
- Smeekes, S. (2009a). *Bootstrapping Nonstationary Time Series*. PhD thesis, Maastricht University, Maastricht, The Netherlands.
- Smeekes, S. (2009b). Detrending bootstrap unit root tests, METEOR Research Memorandum 09/056, Maastricht University, Maastricht, The Netherlands.
- Smeekes, S., Taylor, A. M. R. (2012). Bootstrap union tests for unit roots in the presence of nonstationary volatility. *Econometric Theory* 28:422–456.
- Sul, D. (2009). Panel unit root tests under cross section dependence with recursive mean adjustment. *Economics Letters* 105:123–126.
- 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.
- Taylor, A. M. R. (2002). Regression-based unit root tests with recursive mean adjustment for seasonal and nonseasonal time series. *Journal of Business and Economic Statistics* 20:269–281.
- Xiao, Z., Phillips, P. C. B. (1998). An ADF coefficient test for a unit root in ARMA models of unknown order with empirical applications to the US economy. *Econometrics Journal* 1:27–43.