Maastricht University

Stephan Smeekes

Detrending Bootstrap Unit Root Tests

METEOR

Detrending Bootstrap Unit Root Tests

Stephan Smeekes[*]

Department of Quantitative Economics
Maastricht University
November 6, 2009

Abstract

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 points should be treated separately. Asymptotic validity of sieve bootstrap ADF unit root tests is shown for test statistics based on full sample and recursive OLS and 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 detrending should be preferred 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.

JEL Classification: C15, C22.

Keywords: unit root test, deterministic trends, sieve bootstrap.

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 a nonstationary setting. Especially in the field of unit root testing, where finite sample size distortions are known to occur frequently, a large literature has arisen. The literature has focused mainly on how to deal with serial

correlation, but it stays relatively silent on an important aspect of unit root testing, that is how to deal with deterministic trends. Our aim in this paper is to investigate how the method of detrending impacts the performance of bootstrap unit root tests in univariate time series.

It is very important in practical applications to allow for deterministic trends. Many economic series such as real 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 Φ-tests of Dickey and Fuller (1981). The alternative way, which has become the most popular in recent years, is to perform an initial step of detrending, with the goal of eliminating the deterministic components, and then performing 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, Rothenberg, and Stock (1996) showed that GLS (or quasi-difference) detrending is optimal in terms of local asymptotic power if the initial condition is equal to zero. Simulations also show that the finite sample power of GLS detrended tests, in particular the DF-GLS test, is higher than that of their OLS detrended counterparts for a zero initial condition. Another method that has been proposed is recursive (OLS) detrending, originally proposed by Shin and So (2001) and Taylor (2002).[1] Shin and So (2001) show that with recursive demeaning 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.

The work on bootstrap unit root testing has become quite extensive. The foundations were laid by Basawa, Mallik, McCormick, Reeves, and Taylor (1991a,b) and Ferretti and Romo (1996), who considered settings with simple correlation structures. Their work was later extended to fairly general settings by Park (2002), Chang and Park (2003), Paparoditis and Politis (2003), Swensen (2003a) and Parker, Paparoditis, and Politis (2006) among others. The tests that have been proposed in these papers differ in three respects. First, the bootstrap method. Some tests use the sieve bootstrap, others the moving-blocks or stationary bootstrap. Second, the test statistic. All these tests are based on Dickey-Fuller (DF)

---

[1]The method proposed by Shin and So (2001) can only be used for recursive demeaning; when applied to linear trends their method is not invariant to the trend parameter (Rodrigues, 2006). An extension of their method that is invariant to the trend parameter was proposed by Sul (2009).

type of test statistics, some methods use the augmented DF (ADF) test while others use the non-augmented test. Finally, the methods differ in whether estimation is done under the null or under the alternative.

Given the large array of options, the question becomes how to deal with them in this paper. We choose to focus on one single bootstrap test, based on the following. In Palm, Smeekes, and Urbain (2008) these tests are compared and it is found that the ADF test is clearly preferable to the DF test.[2] Furthermore, the sieve bootstrap usually outperforms the block bootstrap, especially for linear models. Regarding the use of differences and residuals, it is strongly argued in Paparoditis and Politis (2005) to use residuals as using differences leads to a misspecified model if the alternative is true. For these reasons, we focus here on the residual-based ADF sieve bootstrap t-test, a test that performed well in the simulation study of Palm et al. (2008) and was advocated by Paparoditis and Politis (2005).

The framework covered by Palm et al. (2008) is obviously not complete by any means. Cavaliere and Taylor (2009) propose bootstrap versions of the M unit root tests of Ng and Perron (2001) based on GLS detrending. Richard (2007) proposes an ARMA sieve bootstrap unit root test, instead of the regular AR sieve method. Simulations indicate that the method has quite some potential. Another interesting extension is to allow for nonstationary volatility, and apply the tests of Cavaliere and Taylor (2008). However, we restrict ourselves to one specific bootstrap unit root test, in order to analyze the effects of detrending only without having to consider differences in bootstrap tests.

In this paper we extend the proof of asymptotic validity given in Palm et al. (2008) to a setting with deterministic components in the DGP, and allowing for a range of detrending methods that includes full sample and recursive OLS and GLS detrending. Most of the bootstrap unit root tests considered in the literature that take deterministic trends into account are based on full sample OLS detrending or including the deterministic components in the test regression (which closely resembles OLS detrending). The exceptions are Swensen (2003b) who also considers a DF test based on GLS detrending, although in a setting without serial correlation, and Cavaliere and Taylor (2009). To our knowledge no bootstrap version of a test based on recursive detrending has yet been proposed. As a side-product we obtain a single framework that nests full sample and recursive OLS and GLS detrending, and a rigorous derivation of the asymptotic distribution of the ADF t-statistics for these detrending methods. While the limiting distributions are well known and accepted, to our knowledge no such rigorous derivations as in Chang and Park (2002) can be found in the literature for ADF t-tests with deterministic components.

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

---

[2]This conclusion is not surprising given that ADF tests are asymptotically pivotal and therefore may provide asymptotic refinements (Park, 2003) whereas DF tests are not.

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, Leybourne, and Taylor (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. This is therefore a very interesting and empirically relevant issue. However, 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 outline of the paper is as follows. Section 2 will describe the model used for the theoretical analysis. The tests will be explained and their limit distributions derived in Section 3. The bootstrap tests are the topic of Section 4. In Section 5 a simulation study will be undertaken. Section 6 concludes. All proofs are contained in the appendix.

A word on notation. $\lfloor x \rfloor$ is the largest integer smaller than or equal to x. We define $\Delta[a_t b_t]$ such that $\Delta[a_t b_t] = a_t b_t - a_{t-1} b_{t-1}$ for any sequences a_t and b_t. 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. $W(r)$ denotes a univariate standard Brownian motion.

2 The model with deterministic trends

We consider the following Data Generating Process (DGP), where y_t is a scalar variable.

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

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.

We need the following assumption on the linear process $\psi(z)$.

Assumption 1.

(i) Let ε_t be i.i.d. with $\mathrm{E}\,\varepsilon_t = 0$, $\mathrm{E}\,\varepsilon_t^2 = \sigma^2$ and $\mathrm{E}\,\varepsilon_t^4 < \infty$.

(ii) $\psi(z) \neq 0$ for all $|z| \leq 1$, and $\sum_{j=0}^{\infty} j |\psi_j| < \infty$.

4

These assumptions, which are comparable to those found in the literature (cf. Phillips and Solo, 1992; Chang and Park, 2002, 2003), are sufficient for the derivation of the asymptotic distribution of the test statistic and its bootstrap counterpart.

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 conditions on $\psi(z)$, the process is integrated of order zero.

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 Elliott (1999), 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 5.

3   Detrended ADF statistics and their asymptotic properties

We consider ADF statistics with different methods of detrending. We can describe the method in a general framework, of which OLS, GLS and recursive detrending are special cases.[3]

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 test statistics

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

$$y_{t,\gamma,\lambda}^d = y_t - \hat{\beta}_{t,\gamma,\lambda}' z_t. \tag{2}$$

We suggest to detrend the data using $\hat{\beta}_{t,\gamma,\lambda}$ which can be described as

$$\hat{\beta}_{t,\gamma,\lambda} = \left(\sum_{t=1}^{\lambda^*} z_{c_T,\gamma,t} z_{c_T,\gamma,t}' \right)^{-1} \left(\sum_{t=1}^{\lambda^*} z_{c_T,\gamma,t} y_{c_T,\gamma,t} \right). \tag{3}$$

Several parameters need to be explained here. The first, $\lambda^* = \max(t, \lfloor T\lambda \rfloor)$, has the same meaning as in Taylor (2002). It indicates if and how recursive detrending is used, as only

---

[3]A general framework that nests all these options was presented by Broda, Carstensen, and Paolella (2009). Ours slightly deviates from theirs as our objectives are different.

observations up to λ^* 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 λ is between 0 and 1, recursive detrending is used but a minimum proportion of the sample is always used in estimating β.

As mentioned above, we consider the variant of recursive detrending of Taylor (2002). It is easier to apply and does not require the adjustment of Sul (2009), which is necessary as the Shin and So (2001) method is not invariant to the trend parameter. 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. This is for example the case 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 $z_{c_{T,1},1} = z_t - (1 - \bar{c}T^{-1})z_{t-1}$. $y_{c_{T,\gamma}}$ is defined accordingly. Elliott et al. (1996) consider the construction of unit root tests that are point optimal against a local alternative $\rho = 1 - \bar{c}T^{-1}$. Local alternatives are the relevant framework if one is interested in alternatives that are close to the null hypothesis. 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 γ and λ when no confusion can arise. Hence, we usually write y_t^d and $\hat{\beta}_t$ when the context is clear. Also we will usually just write $z_{c,t}$ and $y_{c,t}$ instead of $z_{c_{T,\gamma}}$ and $y_{c_{T,\gamma}}$.

Let $\delta = \rho - 1$. The estimate for δ is then obtained from the OLS regression

$$\Delta y_t^d = \delta y_{t-1}^d + \sum_{j=1}^{p} \phi_j \Delta y_{t-j}^d + \varepsilon_{p,t}^d. \tag{4}$$

Letting $w_{p,t}^d = (\Delta y_{t-1}^d, \ldots, \Delta y_{t-p}^d)'$, we can define

$$\hat{\delta} = A_T B_T^{-1}, \tag{5}$$

where

$$A_T = \sum_{t=1}^{T} y_{t-1}^d \Delta y_t^d - \left(\sum_{t=1}^{T} y_{t-1}^d w_{p,t}^{d\prime} \right) \left(\sum_{t=1}^{T} w_{p,t}^d w_{p,t}^{d\prime} \right)^{-1} \left(\sum_{t=1}^{T} w_{p,t}^d \Delta y_t^d \right)$$

$$B_T = \sum_{t=1}^{T} y_{t-1}^{d2} - \left(\sum_{t=1}^{T} y_{t-1}^d w_{p,t}^{d\prime} \right) \left(\sum_{t=1}^{T} w_{p,t}^d w_{p,t}^{d\prime} \right)^{-1} \left(\sum_{t=1}^{T} w_{p,t}^d y_{t-1}^d \right). \qquad (6)$$

We can then define the ADF t-statistic as

$$ADF_{\gamma,\lambda} = \hat{\delta} \left[\hat{\sigma}^2 \widehat{\text{Var}}(\hat{\delta}) \right]^{-1/2} = \hat{\sigma}^{-1} A_T B_T^{-1/2}, \qquad (7)$$

where $\hat{\sigma}^2$ is the OLS residual variance estimator in (4) (defined explicitly in Lemma 2).

3.2 Asymptotic properties

In this section we derive the limiting distributions of the test statistics under the null hypothesis. Our first goal is to obtain an autoregressive approximation for the detrended series, on which the ADF test is based.

As Assumption 1 implies that $\psi(z)$ is invertible, we can define $\phi(z)$ as $\phi(z) = \psi(z)^{-1} = 1 - \sum_{j=1}^{\infty} \phi_j z^j$ and write

$$u_t = \sum_{j=1}^{\infty} \phi_j u_{t-j} + \varepsilon_t. \qquad (8)$$

Now define $\varepsilon_{p,t}$ such that

$$u_t = \sum_{j=1}^{p} \phi_j u_{t-j} + \varepsilon_{p,t}. \qquad (9)$$

Combining (8) and (9) we obtain

$$\varepsilon_{p,t} = \varepsilon_t + \sum_{j=p+1}^{\infty} \phi_j u_{t-j}. \qquad (10)$$

As $y_t^d = y_t - \hat{\beta}_t' z_t$ and $y_t = x_t + \beta' z_t$, we have that

$$y_t^d = x_t + \beta' z_t - \hat{\beta}_t' z_t = x_t - (\hat{\beta}_t - \beta)' z_t. \qquad (11)$$

Then $u_t = \Delta x_t = \Delta y_t^d + \Delta[(\hat{\beta}_t - \beta)'z_t]$. Now we can write

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

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

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

$$\Delta y_t^d = \sum_{j=1}^{p} \phi_j \Delta y_{t-j}^d + \varepsilon_{p,t}^d. \tag{12}$$

Similarly we can define ε_t^d such that

$$\varepsilon_t^d = \Delta y_t^d - \sum_{j=1}^{\infty} \phi_j \Delta y_{t-j}^d = \varepsilon_t - \phi(L)\Delta[(\hat{\beta}_t - \beta)'z_t].$$

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

$$\Delta Y^d = M_p^d \Phi_p + \varepsilon_p^d$$

and

$$A_T = Y_{-1}^{d\prime} \Delta Y^d - Y_{-1}^{d\prime} M_p^d \left(M_p^{d\prime} M_p^d \right)^{-1} M_p^{d\prime} \Delta Y^d = Y_{-1}^{d\prime} \varepsilon_p^d - Y_{-1}^{d\prime} M_p^d \left(M_p^{d\prime} M_p^d \right)^{-1} M_p^{d\prime} \varepsilon_p^d$$
$$B_T = Y_{-1}^{d\prime} Y_{-1}^d - Y_{-1}^{d\prime} M_p^d \left(M_p^{d\prime} M_p^d \right)^{-1} M_p^{d\prime} Y_{-1}^d, \tag{13}$$

or equivalently

$$A_T = \sum_{t=1}^{T} y_{t-1}^d \varepsilon_{p,t}^d - \left(\sum_{t=1}^{T} y_{t-1}^d w_{p,t}^{d\prime} \right) \left(\sum_{t=1}^{T} w_{p,t}^d w_{p,t}^{d\prime} \right)^{-1} \left(\sum_{t=1}^{T} w_{p,t}^d \varepsilon_{p,t}^d \right)$$
$$B_T = \sum_{t=1}^{T} y_{t-1}^{d2} - \left(\sum_{t=1}^{T} y_{t-1}^d w_{p,t}^{d\prime} \right) \left(\sum_{t=1}^{T} w_{p,t}^d w_{p,t}^{d\prime} \right)^{-1} \left(\sum_{t=1}^{T} w_{p,t}^d y_{t-1}^d \right). \tag{14}$$

We need the following assumption on the lag length p.

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

Using the expressions developed above, one can derive the asymptotic distribution of the test statistics.

The first step in deriving the limiting distribution is to consider the limiting behavior of

the elements of A_T and B_T, as in the following lemma.

Lemma 1. *Let Assumptions 1 and 2 hold. Let $y_t^d = y_{t,\gamma,\lambda}^d$ be defined as in equation (2) with $\gamma = 0, 1$ and $\lambda \in [0, 1]$. Then*

(a) $T^{-2} \sum_{t=1}^{T} y_{t-1}^{d2} \xrightarrow{d} \psi(1)^2 \sigma^2 \int_0^1 W_\gamma(r, \lambda)^2 dr$,

(b) $T^{-1} \sum_{t=1}^{T} y_{t-1}^d \varepsilon_{p,t}^d \xrightarrow{d} \frac{1}{2}\psi(1)\sigma^2 (W_\gamma(1, \lambda)^2 - W_\gamma(0, \lambda)^2 - 1)$,

(c) $\left\| \left(T^{-1} \sum_{t=1}^{T} w_{p,t}^d w_{p,t}^{d\prime} \right)^{-1} \right\| = O_p(1)$,

(d) $\left| T^{-1} \sum_{t=1}^{T} y_{t-1}^d w_{p,t}^{d\prime} \right| = O_p(p^{1/2})$,

(e) $\left| T^{-1} \sum_{t=1}^{T} w_{p,t}^d \varepsilon_{p,t}^d \right| = o_p(p^{-1/2})$,

where

$$W_0(r, \lambda) = W(r) - 2\bar{r}^{-2}(2 - 3r) \int_0^{\bar{r}} W(s)ds - 6\bar{r}^{-3}(2r - 1) \int_0^{\bar{r}} sW(s)ds,$$

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

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

The next step is to show the consistency of the residual variance estimator, as done in the following lemma.

Lemma 2. *Let Assumptions 1 and 2 hold. Let $\hat{\sigma}^2$ be defined as*

$$\hat{\sigma}^2 = T^{-1}(\Delta Y^d - Y_{-1}^d \hat{\delta})'(I - M_p^d(M_p^{d\prime} M_p^d)^{-1} M_p^{d\prime})(\Delta Y^d - Y_{-1}^d \hat{\delta}).$$

Then $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$.

We can then straightforwardly derive the limiting distribution of the ADF t-statistic as given below.

Theorem 1. *Let Assumptions 1 and 2 hold. Let $ADF_{\gamma,\lambda}$ be defined as in (7) with $\gamma = 0, 1$ and $\lambda \in [0, 1]$. Then, as $T \to \infty$, we have that*

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

9

Remark 1. Under the local alternative $\rho = 1 - cT^{-1}$ the limit distribution will remain the same as in Theorem 1, but with $W_\gamma(r, \lambda)$ replaced by $W_{c,\gamma}(r, \lambda)$, where

$$W_{c,0}(r, \lambda) = W_c(r) - 2\bar{r}^{-2}(2 - 3r) \int_0^{\bar{r}} W_c(s)ds - 6\bar{r}^{-3}(2r - 1) \int_0^{\bar{r}} sW_c(s)ds,$$

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

and $W_c(r) = \int_0^r e^{-(r-s)c}dW(s)$.

This can be shown straightforwardly, though tediously, using standard results for the invariance principle (cf. Phillips and Perron, 1988) and our proofs in the Appendix. Note that under local alternatives $u_t = \Delta x_t + cT^{-1}x_{t-1}$ and, analogously to the derivation of (12), we can then derive that $\Delta y_t^d = \sum_{j=1}^p \phi_j \Delta y_{t-j}^d + \varepsilon_{p,t}^{c,d}$ where $\varepsilon_{p,t}^{c,d} = \varepsilon_{p,t}^d - cT^{-1}\phi_p(L)x_{t-1}$. As we can further derive that $\varepsilon_t^{c,d} = \varepsilon_t^d - cT^{-1}\phi(L)x_{t-1}$, we can then plug these quantities into the proofs to obtain the results given above.

4 Bootstrap ADF statistics and their asymptotic properties

4.1 Bootstrap algorithm

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.

Bootstrap Algorithm 1.

1. Calculate

$$\tilde{y}_t^d = y_t - \tilde{\beta}_t' z_t, \tag{15}$$

    where $\tilde{\beta}_t = \hat{\beta}_{t,\tilde{\gamma},\tilde{\lambda}}$ is defined in (3) 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

$$\hat{\varepsilon}_{q,t}^d = \Delta\tilde{y}_t^d - \hat{\delta}\tilde{y}_{t-1}^d - \sum_{j=1}^q \hat{\phi}_j \Delta\tilde{y}_{t-j}^d. \tag{16}$$

3. Resample with replacement from the recentered residuals $(\hat{\varepsilon}_{q,t}^d - \bar{\hat{\varepsilon}}_{q,t}^d)$ to obtain bootstrap errors ε_t^*.

4. Build u_t^* recursively as

$$u_t^* = \sum_{j=1}^q \hat{\phi}_j u_{t-j}^* + \varepsilon_t^*, \tag{17}$$

10

using the estimated parameters $\hat{\phi}_j$ from Step 2, and build x_t^* as

$$x_t^* = x_{t-1}^* + u_t^*. \tag{18}$$

Finally let

$$y_t^* = x_t^* + \beta^{*\prime} z_t. \tag{19}$$

See Remark 2 for the choice of β^*.

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 $y_t^{*d} = y_{t,\gamma,\lambda}^{*d}$, where

$$y_{t,\gamma,\lambda}^{*d} = y_t^* - \hat{\beta}_{t,\gamma,\lambda}^{*\prime} z_t, \tag{20a}$$

$$\hat{\beta}_{t,\gamma,\lambda}^* = \left(\sum_{t=1}^{\lambda^*} z_{cT,\gamma,t} z_{cT,\gamma,t}' \right)^{-1} \left(\sum_{t=1}^{\lambda^*} z_{cT,\gamma,t} y_{cT,\gamma,t}^* \right). \tag{20b}$$

Estimate by OLS the ADF regression of order p^*,

$$\Delta y_t^{*d} = \delta^* y_{t-1}^{*d} + \sum_{j=1}^{p^*} \phi_j^* \Delta y_{t-j}^{*d} + \varepsilon_{p^*,t}^{*d}, \tag{21}$$

and calculate the ADF test statistic as

$$ADF_{\gamma,\lambda}^* = \hat{\delta}^* \left[\hat{\sigma}^{*2} \widehat{\mathrm{Var}}(\hat{\delta}^*) \right]^{-1/2} = \hat{\sigma}^{*-1} A_T^* B_T^{*-1/2}, \tag{22}$$

where

$$\begin{aligned}
A_T^* &= \sum_{t=1}^T y_{t-1}^{*d} \Delta y_t^{*d} - \left(\sum_{t=1}^T y_{t-1}^{*d} w_{p^*,t}^{*d\prime} \right) \left(\sum_{t=1}^T w_{p^*,t}^{*d} w_{p^*,t}^{*d\prime} \right)^{-1} \left(\sum_{t=1}^T w_{p^*,t}^{*d} \Delta y_t^{*d} \right) \\
B_T^* &= \sum_{t=1}^T y_{t-1}^{*d2} - \left(\sum_{t=1}^T y_{t-1}^{*d} w_{p^*,t}^{*d\prime} \right) \left(\sum_{t=1}^T w_{p^*,t}^{*d} w_{p^*,t}^{*d\prime} \right)^{-1} \left(\sum_{t=1}^T w_{p^*,t}^{*d} y_{t-1}^{*d} \right).
\end{aligned} \tag{23}$$

where $w_{p^*,t}^{*d} = (\Delta y_{t-1}^{*d}, \dots, \Delta y_{t-p^*}^{*d})'$.

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_α^* as $c_\alpha^* = \max\{c : \sum_{b=1}^B I(ADF_{\gamma,\lambda}^{*b} < c) \le \alpha\}$, or equivalently as the α-quantile of the ordered $ADF_{\gamma,\lambda}^{*b}$ statistics. Reject the null of a unit root if $ADF_{\gamma,\lambda}$ is smaller than c_α^*, where α 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^*). In general it will be a logical

11

choice to set $q = p$, as both are based on an ADF regression. However we do not wish to impose this a priori in order to be as general as possible. For example, if the methods of detrending differ, in finite samples one might obtain a different p and q if the choice is based on an information criterion.

What is more important however is 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 length q.

Assumption 3. Let $q \to \infty$ and $q = o((n/\ln n)^{1/3})$ as $n \to \infty$.

We also need the following assumption to relate q to p (p^*).

Assumption 4. Let $p/q \to \kappa > 1$ as $T \to \infty$, where κ may be infinite.

This 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. Note that the arguments still hold for different values of β^*. It would however not be valid to set $y_t^* = x_t^* + \beta_t^{*\prime} z_t$, with β_t^* varying over t (for example $\beta_t^* = \hat{\beta}_t$), as this would mean that the parameters of the deterministic trends are time-varying, which is not the case in the original sample.

4.2 Detrending within the bootstrap

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 $\hat{\beta}_{t,\gamma,\lambda}$. 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 3. *Define $\tilde{\phi}_j$, $j = 1, \ldots, q$ as the OLS estimators in a regression of u_t on $u_{t-1}, \ldots, u_{t-q}$ and $\tilde{\varepsilon}_{q,t}$ as the corresponding residuals. Let $\hat{\phi}_j$ and $\hat{\varepsilon}_{q,t}^d$ be defined as in (16). Let $\tilde{\beta}_t = \hat{\beta}_{t,\tilde{\gamma},\tilde{\lambda}}$ be defined as in (3) 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 $j = 1, \ldots, q$. Moreover,

$$\max_{1 \le t \le T} |\hat{\varepsilon}^d_{q,t} - \tilde{\varepsilon}_{q,t}| = O_p(T^{-1/2}).$$

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 (1995, 1997), used in a unit root setting by Park (2002) and Chang and Park (2003) (also see Remark 4). Given Lemma 3 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 $\tilde{\beta}$ is that it satisfies the conditions

$$\tilde{\beta}_1 - \beta_1 = O_p(T^{1/2}) \quad \text{and} \quad \tilde{\beta}_2 - \beta_2 = O_p(T^{-1/2}), \tag{24}$$

thus allowing for trend estimators beyond the OLS and GLS framework (see Smeekes, 2009, 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.[4] In fact, the results of Hannan and Kavalieris (1986) and Bühlmann (1995, 1997) are derived for Yule-Walker estimators. However, Theorem 1 of Poskitt (1994) implies that these results are valid for OLS estimation as well.

4.3 Asymptotic bootstrap validity

In this section we show that the bootstrap tests are asymptotically valid. In order to establish asymptotic validity we need to show that the bootstrap t-statistic converges to the same distribution as its asymptotic counterpart if the null hypothesis is true.

The first step in the derivation of the bootstrap limit distribution is the construction of an invariance principle for y^{*d}_t. The several steps that are needed for the construction are detailed in the Appendix. Here we give the final invariance principle.

Lemma 4. *Let Assumptions 1 and 3 hold. Let $y^{*d}_{t,\gamma,\lambda}$ be defined as in (20) with $\gamma = 0, 1$ and $\lambda \in [0,1]$. Then*

$$T^{-1/2} y^{*d}_{\lfloor Tr \rfloor, \gamma, \lambda} \xrightarrow{d^*} \sigma \psi(1) W_\gamma(r, \lambda) \text{ in probability.}$$

Next we must derive the autoregressive approximation on which the ADF regression is

---

[4]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).

based, as for the asymptotic test. For the bootstrap error process u_t^*, we can write

$$u_t^* = \sum_{j=1}^{q} \hat{\phi}_j u_{t-j}^* + \varepsilon_t^*. \tag{25}$$

In analogy with the original sample, define $\varepsilon_{p,t}^*$ such that

$$u_t^* = \sum_{j=1}^{p} \hat{\phi}_j u_{t-j}^* + \varepsilon_{p,t}^*. \tag{26}$$

Combining (25) and (26) we obtain

$$\varepsilon_{p,t}^* = \varepsilon_t^* + \sum_{j=p+1}^{q} \hat{\phi}_j u_{t-j}^*. \tag{27}$$

However, it is clear from our Assumption 4 that for large T one obtains $\varepsilon_{p,t}^* = \varepsilon_t^*$. Therefore our proofs can proceed as if we set $p = q$.

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

$$\varepsilon_t^{*d} = \varepsilon_t^* - \hat{\phi}_q(L)\Delta[(\hat{\beta}_t^* - \beta^*)'z_t] = \varepsilon_t^* - \hat{\phi}(L)\Delta[(\hat{\beta}_t^* - \beta^*)'z_t].$$

It will then also be clear that for large T we have $\varepsilon_{p,t}^{*d} = \varepsilon_t^{*d}$.

Now let ΔY^{*d}, Y_{-1}^{*d}, M_p^{*d}, ε_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

$$\begin{aligned}
A_T^* &= Y_{-1}^{*d\prime}\varepsilon_p^{*d} - Y_{-1}^{*d\prime}M_p^{*d}\left(M_p^{*d\prime}M_p^{*d}\right)^{-1}M_p^{*d\prime}\varepsilon_p^{*d} \\
B_T^* &= Y_{-1}^{*d\prime}Y_{-1}^{*d} - Y_{-1}^{*d\prime}M_p^{*d}\left(M_p^{*d\prime}M_p^{*d}\right)^{-1}M_p^{*d\prime}Y_{-1}^{*d} \\
\hat{\sigma}^{*2} &= T^{-1}(\Delta Y^{*d} - Y_{-1}^{*d}\hat{\alpha}^*)'(I - M_p^{*d}(M_p^{*d\prime}M_p^{*d})^{-1}M_p^{*d\prime})(\Delta Y^{*d} - Y_{-1}^{*d}\hat{\alpha}^*).
\end{aligned} \tag{28}$$

Next we can establish the bootstrap counterparts of Lemma 1 and 2.

Lemma 5. *Let Assumptions 1, 2, 3 and 4 hold. Let $y_t^{*d} = y_{t,\gamma,\lambda}^{*d}$ be defined as in (20) with $\gamma = 0, 1$ and $\lambda \in [0, 1]$. Then*

(a) $T^{-2} \sum_{t=1}^T y_{t-1}^{*d2} \xrightarrow{d^*} \psi(1)^2 \sigma^2 \int_0^1 W_\gamma(r, \lambda)^2 dr$ in probability.

(b) $T^{-1} \sum_{t=1}^T y_{t-1}^{*d} \varepsilon_{p,t}^{*d} \xrightarrow{d^*} \frac{1}{2}\psi(1)\sigma^2(W_\gamma(1, \lambda)^2 - W_\gamma(0, \lambda)^2 - 1)$ in probability.

(c) $\left\| \left(T^{-1} \sum_{t=1}^T w_{p,t}^{*d} w_{p,t}^{*d\prime} \right)^{-1} \right\| = O_p^*(1),$

(d) $\left| T^{-1} \sum_{t=1}^T y_{t-1}^{*d} w_{p,t}^{*d\prime} \right| = O_p^*(p^{1/2}),$

(e) $\left| T^{-1} \sum_{t=1}^T w_{p,t}^{*d} \varepsilon_{p,t}^{*d} \right| = O_p^*(T^{-1/2} p^{1/2}).$

Lemma 6. *Let Assumptions 1, 2, 3 and 4 hold. Let $\hat{\sigma}^{*2}$ be defined as in (28). Then $\hat{\sigma}^{*2} \xrightarrow{p^*} \sigma^2$.*

This leads to 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 $ADF_{\gamma,\lambda}^*$ be defined as in (22) with $\gamma = 0, 1$ and $\lambda \in [0, 1]$. Let Assumptions 1, 2, 3 and 4 hold. Then, as $T \to \infty$, we have that*

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

4.4 Bootstrap tests under the alternative hypothesis

The asymptotic validity of the bootstrap tests that we established in the previous section is purely a property of the bootstrap tests under the null hypothesis. In this section we investigate how the bootstrap performs under the alternative hypothesis. This is what we investigate in this section. We discern two different alternative hypotheses, local and fixed alternatives.

Under local alternatives the bootstrap tests should have the same asymptotic distribution as under the null hypothesis. It is only then that the bootstrap tests will have the same asymptotic local power function as the asymptotic tests. Swensen (2003b) shows that this is the case for full sample OLS and GLS tests when there is no correlation in the residuals and the i.i.d. bootstrap can be used.

Under fixed alternatives the bootstrap test should converge to some limiting distribution (i.e. it should not diverge) in order to achieve consistency. However, to have the highest power possible one wants again that the limit distribution is the same as under the null.

We will not go into the technical details in this paper but we try to show intuitively why the bootstrap tests considered here satisfy these requirements. It is not difficult to see that under local alternatives the bootstrap tests will have the same asymptotic distribution as under the null hypothesis. Under local alternatives 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 3, remain valid.[5] It then follows from this Lemma that the bootstrap tests will have the same distributions as under the null.

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,[6] which means that this will not cause any problems. For these reasons the bootstrap tests will have the same distributions under fixed alternatives as under the null hypothesis.

5 Finite-sample performance

5.1 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 two sets of simulations, the first with a small (zero) initial condition, the second 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

---

[5]In fact, one just needs to modify the proof of 3 using the expressions for $\varepsilon_{q,t}^{c,d}$ and $\varepsilon_t^{c,d}$ obtained in Remark 1.

[6]See for example Hamilton (1994, Chapter 16) for the OLS estimator in a model with intercept and trend.

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 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 in the tables are denoted by $ADF_{\gamma,\lambda}$ where γ and λ 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 $\bar{c} = 13.5$ as Elliott et al. (1996) suggest.[7]

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.

$$y_t = x_t + \beta' z_t$$
$$x_t = \rho x_{t-1} + u_t$$
$$u_t = \phi u_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$$

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 β.

For the first set of simulations we set the initial condition equal to zero, i.e. $x_0 = 0$. For the second set of simulations, we follow Harvey et al. (2009) and set

$$x_0 = a\sqrt{\omega_u/(1-\rho^2)},$$

where $\omega_u = \lim_{T\to\infty} T^{-1} \mathrm{E}(\sum_{t=1}^{T} u_t)^2$. We set $a = 2.5$, 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 MAIC proposed by Ng and Perron (2001). All results are obtained using 5000 simulations and the Warp-speed bootstrap method of Giacomini, Politis, and White (2007).

---

[7]There is 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.

5.2 Simulation results

Tables 1 and 2 present results for size ($c = 0$). It can be seen that all bootstrap tests perform better than the asymptotic tests. The size of the asymptotic tests is quite sensitive to the values in the simulation DGP of both the AR and MA parameter. There is undersize for most parameter combinations, although generally not too severe, while there is the familiar oversize for negative MA parameters. Among the asymptotic tests the OLS tests seem to be more sensitive to the parameters of the dynamics than the GLS tests.

INSERT TABLE 1 ABOUT HERE

INSERT TABLE 2 ABOUT HERE

The bootstrap tests are far less sensitive than the asymptotic tests to the values of the AR and MA 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 considerably less than for the asymptotic tests. What is also noticeable is that the bootstrap not only corrects oversize of the asymptotic tests, but also undersize.

If we consider the method of detrending in the first step of the bootstrap, we see that the recursively detrended tests have a tendency to reject less often than the full sample detrended tests. This is positive for the DGPs with negative MA parameters, but leads to undersize elsewhere, which could affect power negatively. The effects of the detrending method used for the calculation of the test statistic follows that of the corresponding asymyptotic tests (although less pronounced), with the largest size distortions for the OLS full sample detrended test.

Tables 3 and 4 give size-adjusted powers for $c = 10$ for a model with zero initial condition. Regarding the asymptotic tests, we see that the size-adjusted power of the full sample OLS detrended test is clearly lower than that of the other tests (which are fairly comparable). This is in line with results from the literature on unit root testing (Elliott et al., 1996; Shin and So, 2001).

INSERT TABLE 3 ABOUT HERE

INSERT TABLE 4 ABOUT HERE

It can also be seen that the size-adjusted powers of the bootstrap tests are somewhat lower than but still quite close to those of the asymptotic tests. Considering the method of detrending used in the first step of the bootstrap, there is again a slight tendency for the recursive detrending methods to have lower power. For the detrending method used for the test statistic, it is very clear that the bootstrap tests follow their asymptotic counterparts as the full sample OLS tests have lowest power.

Based on these results, it seems that the power properties of the bootstrap tests are determined by the power properties of their asymptotic counterparts. We will try to confirm this conclusion by next looking at models with large initial conditions.

Tables 5 and 6 give the size-adjusted powers for $c = 10$ for the model with a large initial condition. The initial condition used is based on Harvey et al. (2009), where this value led to a clear power advantage of OLS over GLS detrending. It is not clear yet from the literature how the recursive methods perform for such a large initial condition.

INSERT TABLE 5 ABOUT HERE

INSERT TABLE 6 ABOUT HERE

Considering the asymptotic tests first, we see that the test based on full sample OLS detrending is clearly the most powerful now. The power advantage of OLS detrending over GLS detrending is in line with the results in Harvey et al. (2009). The power advantage of OLS over GLS is also noticeable for recursive detrending, although the power of the recursively detrended tests is smaller than that of their full sample counterparts.

Next we turn to the bootstrap tests, and in particular the detrending method used for the calculation of the test statistic. It can again be seen that the bootstrap tests follow their asymptotic counterparts closely. Bootstrap tests based on OLS detrending are more powerful than bootstrap tests based on GLS detrending and full sample detrending is more powerful than recursive detrending. The impact of the method of detrending in the first step of the bootstrap algorithm is similar as before: recursive detrending leads to somewhat lower power. These results confirm our conclusion that power properties of the bootstrap tests are mainly determined by the power properties of their asymptotic counterparts.

6   Conclusion

We have investigated the role of detrending in bootstrap unit root tests. We have shown 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 the method of detrending in the first step of the bootstrap algorithm does not have a huge impact on the size and the power of the test, although there are differences between full sample and recursive detrending. 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, while having a minor impact on the size. Moreover, the power properties of the bootstrap tests are determined by the power properties of their asymptotic counterparts.

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. In our analysis we found that full sample detrending outperformed recursive detrending both in terms of size and power. The difference between OLS and GLS detrending was fairly minor. Second, the choice of the detrending 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. For example, one could simply adapt the arguments used in Harvey et al. (2009) when there is uncertainty over the initial condition to the bootstrap setting.

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. Third, we could explicitly use the bootstrap to tackle the problem of uncertainty about deterministic trends and/or the initial condition. Instead of simply adapting the ideas of Harvey et al. (2009) to the bootstrap test, one could explicitly use the bootstrap to control size exactly when the rejection strategy is based on the union of rejections of individual tests as in Harvey et al. (2009). To do so however would not be trivial. Finally, one could view detrending in a broader perspective and analyze more general trends, such as polynomial trends of higher order or broken trends.

References

Basawa, I. V., A. K. Mallik, W. P. McCormick, J. H. Reeves, and R. L. Taylor (1991a). Bootstrap test of significance and sequential bootstrap estimation for unstable first order autoregressive processes. *Communications in Statistics - Theory and Methods 20*, 1015–1026.

Basawa, I. V., A. K. Mallik, W. P. McCormick, J. H. Reeves, and R. L. Taylor (1991b). Bootstrapping unstable first-order autoregressive processes. *Annals of Statistics 19*, 1098–1101.

Berk, K. N. (1974). Consistent autoregressive spectral estimates. *Annals of Statistics 2*, 489–502.

Broda, S., K. Carstensen, and M. S. Paolella (2009). Assessing and improving the performance of nearly efficient unit root tests in small samples. *Econometric Reviews 28*, 468–494.

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.

Burridge, P. and A. M. R. Taylor (2004). Bootstrapping the HEGY seasonal unit root tests. *Journal of Econometrics 123*, 67–87.

Cavaliere, G. and A. M. R. Taylor (2008). Bootstrap unit root tests for time series with nonstationary volatility. *Econometric Theory 24*, 43–71.

Cavaliere, G. and A. M. R. Taylor (2009). Bootstrap M unit root tests. *Econometric Reviews 28*, 393–421.

Chang, Y. and J. Y. Park (2002). On the asymptotics of ADF tests for unit roots. *Econometric Reviews 21*, 431–447.

Chang, Y. and J. Y. Park (2003). A sieve bootstrap for the test of a unit root. *Journal of Time Series Analysis 24*, 379–400.

Dickey, D. A. and W. A. Fuller (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica 49*, 1057–1072.

Elliott, G. (1999). Efficient tests for a unit root when the initial observation is drawn from its unconditional distribution. *International Economic Review 40*, 767–783.

Elliott, G. and U. K. Müller (2006). Minimizing the impact of the initial condition on testing for unit roots. *Journal of Econometrics 135*, 285–310.

Elliott, G., T. J. Rothenberg, and J. H. Stock (1996). Efficient tests for an autoregressive unit root. *Econometrica 64*, 813–836.

Ferretti, N. and J. Romo (1996). Unit root bootstrap tests for AR(1) models. *Biometrika 83*, 849–860.

Giacomini, R., D. N. Politis, and H. White (2007). A Warp-Speed method for conducting Monte Carlo experiments involving bootstrap estimators. Working paper.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.

Hannan, E. J. and L. Kavalieris (1986). Regression, autoregression models. *Journal of Time Series Analysis 7*, 27–49.

Harvey, D. I., S. J. Leybourne, and A. M. R. Taylor (2009). Unit root testing in practice: dealing with uncertainty over the trend and initial condition. *Econometric Theory 25*, 587–636.

Müller, U. K. and G. Elliott (2003). Tests for unit roots and the initial condition. *Econometrica 71*, 1269–1286.

Ng, S. and P. Perron (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica 69*, 1519–1554.

Palm, F. C., S. Smeekes, and J.-P. Urbain (2008). Bootstrap unit root tests: comparison and extensions. *Journal of Time Series Analysis 29*, 371–401.

Palm, F. C., S. Smeekes, and J.-P. Urbain (2010). A sieve bootstrap test for cointegration in a conditional error correction model. *Econometric Theory*, forthcoming.

Paparoditis, E. and D. N. Politis (2003). Residual-based block bootstrap for unit root testing. *Econometrica 71*, 813–855.

Paparoditis, E. and D. N. Politis (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.

Park, J. Y. (2003). Bootstrap unit root tests. *Econometrica 71*, 1845–1895.

Parker, C., E. Paparoditis, and D. N. Politis (2006). Unit root testing via the stationary bootstrap. *Journal of Econometrics 133*, 601–638.

Phillips, P. C. B. and P. Perron (1988). Testing for a unit root in time series regression. *Biometrika 75*, 335–346.

Phillips, P. C. B. and V. Solo (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. and B. S. So (2001). Recursive mean adjustment for unit root tests. *Journal of Time Series Analysis 22*, 595–612.

Smeekes, S. (2009). *Bootstrapping Nonstationary Time Series.* Ph. D. thesis, Maastricht University.

Stock, J. H. (1994). Unit roots, structural breaks and trends. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, Chapter 46, pp. 2739–2841. Amsterdam: North Holland Publishing.

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. and P. C. B. Phillips (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.

A   Appendix: Proofs

A.1   Proofs for Section 3

For completeness, we start with two results that are well known in the literature (Phillips and Solo, 1992). We let $W(r)$ denote a standard Brownian motion.

Lemma A.1. *Let Assumption 1 hold. Then*

$$T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \varepsilon_t \xrightarrow{d} \sigma W(r).$$

Lemma A.2. *Let Assumption 1 hold. Then*

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

The first step is to derive the distribution of the estimator of β. This is done in Lemma A.3. The results in Lemma A.3 hold for OLS and GLS (possibly recursive) detrending. Note that these are fairly standard results (cf. Stock, 1994; Elliott et al., 1996; Taylor, 2002).

Lemma A.3. *Let $(\hat{\beta}_{1,t}, \hat{\beta}_{2,t})' = \hat{\beta}_t = \hat{\beta}_{t,\gamma,\lambda}$ be defined as in (3), with $\gamma = 0, 1$ and $\lambda \in [0,1]$. Let Assumption 1 hold. Then*

$$\begin{pmatrix} T^{-1/2}(\hat{\beta}_{1,\lfloor Tr \rfloor} - \beta_1) \\ T^{1/2}(\hat{\beta}_{2,\lfloor Tr \rfloor} - \beta_2) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \psi(1)\sigma V_{1,\gamma}(r,\lambda) \\ \psi(1)\sigma V_{2,\gamma}(r,\lambda) \end{pmatrix},$$

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

$$V_{1,0}(r,\lambda) = 4\bar{r}^{-1} \int_0^{\bar{r}} W(s)ds - 6\bar{r}^{-2} \int_0^{\bar{r}} sW(s)ds,$$

$$V_{2,0}(r,\lambda) = -6\bar{r}^{-2} \int_0^{\bar{r}} W(s)ds + 12\bar{r}^{-3} \int_0^{\bar{r}} sW(s)ds,$$

$$V_{1,1}(r,\lambda) = 0,$$

$$V_{2,1}(r,\lambda) = \bar{r}^{-1}(1 + \bar{c}\bar{r} + \tfrac{1}{3}\bar{c}^2\bar{r}^2)^{-1}\left[(1 + \bar{c}\bar{r})W(\bar{r}) + \bar{c}^2 \int_0^{\bar{r}} sW(s)ds\right]. \tag{A.1}$$

Proof of Lemma A.3. We have that

$$\hat{\beta}_{\lfloor Tr \rfloor} - \beta = \left(\sum_{s=1}^{\lambda^*} z_{c_T,s} z'_{c_T,s}\right)^{-1} \left(\sum_{s=1}^{\lambda^*} z_{c_T,s} x_{c_T,s}\right), \tag{A.2}$$

where $\lambda^* = \max(\lfloor Tr \rfloor, \lfloor T\lambda \rfloor)$, $x_{c_T,1} = x_1$ and $x_{c_T,t} = \Delta x_t + c_T x_{t-1}$ for $t = 2, \ldots, T$. Then,

$$\begin{pmatrix} T^{-1/2}(\hat{\beta}_{1,\lfloor Tr \rfloor} - \beta_1) \\ T^{1/2}(\hat{\beta}_{2,\lfloor Tr \rfloor} - \beta_2) \end{pmatrix} = \begin{pmatrix} T^{\gamma-1}\sum_{s=1}^{\lambda^*} z_{c_T,1s}^2 & T^{3\gamma/2-2}\sum_{s=1}^{\lambda^*} z_{c_T,1s}z_{c_T,2s} \\ T^{3\gamma/2-2}\sum_{s=1}^{\lambda^*} z_{c_T,2s}z_{c_T,1s} & T^{2\gamma-3}\sum_{s=1}^{\lambda^*} z_{c_T,2s}^2 \end{pmatrix}^{-1}$$

$$\times \begin{pmatrix} T^{\gamma-3/2}\sum_{s=1}^{\lambda^*} z_{c_T,1s}x_{c_T,s} \\ T^{2\gamma-5/2}\sum_{s=1}^{\lambda^*} z_{c_T,2s}x_{c_T,s} \end{pmatrix} \tag{A.3}$$

$$= M_{\lambda^*}^{-1} N_{\lambda^*}.$$

We start with the denominator. Note that $z_{c_T,1} = (1,1)'$. Also we have that $z_{c_T,1s} = \Delta z_{1s} + c_T z_{1,s-1} = c_T$ and $z_{c_T,2s} = \Delta z_{2s} + c_T z_{2,s-1} = 1 + c_T(s-1)$. Then

$$\sum_{s=1}^{\lambda^*} z_{c_T,1s}^2 = 1 + \sum_{s=2}^{\lambda^*} c_T^2 = 1 + c_T^2(\lambda^* - 1).$$

Furthermore

$$\sum_{s=1}^{\lambda^*} z_{c_T,1s}z_{c_T,2s} = 1 + c_T \sum_{s=2}^{\lambda^*}(1 + c_T(s-1)) = 1 + c_T(\lambda^* - 1) + \frac{1}{2}c_T^2\lambda^*(\lambda^* - 1).$$

Finally,

$$\sum_{s=1}^{\lambda^*} z_{c_T,2s}^2 = 1 + \sum_{s=2}^{\lambda^*}\left(1 + 2c_T(s-1) + c_T^2(s-1)^2\right)$$

$$= \lambda^* + c_T\lambda^*(\lambda^* - 1) + \frac{1}{6}c_T^2\lambda^*(\lambda^* - 1)(2\lambda^* - 1).$$

Therefore we have that, using that $c_T = \bar{c}^\gamma T^{-\gamma}$

$$M_{\lambda^*} = \begin{pmatrix} T^{\gamma-1} + \bar{c}^{2\gamma}T^{-\gamma-1}(\lambda^* - 1) \\ T^{3\gamma/2-2} + \bar{c}^\gamma T^{\gamma/2-2}(\lambda^* - 1) + \frac{1}{2}\bar{c}^{2\gamma}T^{-\gamma/2-2}\lambda^*(\lambda^* - 1) \end{pmatrix}$$

$$\begin{pmatrix} T^{3\gamma/2-2} + \bar{c}^\gamma T^{\gamma/2-2}(\lambda^* - 1) + \frac{1}{2}\bar{c}^{2\gamma}T^{-\gamma/2-2}\lambda^*(\lambda^* - 1) \\ T^{2\gamma-3}\lambda^* + \bar{c}^\gamma T^{\gamma-3}\lambda^*(\lambda^* - 1) + \frac{1}{6}\bar{c}^{2\gamma}T^{-3}\lambda^*(\lambda^* - 1)(2\lambda^* - 1) \end{pmatrix}.$$

If $\gamma = 0$, then

$$M_{\lambda^*} \to \begin{pmatrix} \bar{r} & \frac{1}{2}\bar{r}^2 \\ \frac{1}{2}\bar{r}^2 & \frac{1}{3}\bar{r}^3 \end{pmatrix}. \tag{A.4}$$

If $\gamma = 1$, then

$$M_{\lambda^*} \to \begin{pmatrix} 1 & 0 \\ 0 & \bar{r}(1 + \bar{c}\bar{r} + \frac{1}{3}\bar{c}^2\bar{r}^2) \end{pmatrix}. \tag{A.5}$$

Next we consider N_{λ^*}. Note that

$$\sum_{s=1}^{\lambda^*} z_{c_T,1s}x_{c_T,s} = x_1 + c_T\sum_{s=2}^{\lambda^*}(\Delta x_s + c_T x_{s-1}) = x_1 + c_T(x_{\lambda^*} - x_1) + c_T^2\sum_{s=2}^{\lambda^*}x_{s-1}$$

and

$$\sum_{s=1}^{\lambda^*} z_{c_T,2s}x_{c_T,s} = x_1 + \sum_{s=2}^{\lambda^*}(\Delta x_s + c_T x_{s-1}) + c_T\sum_{s=2}^{\lambda^*}(s-1)(\Delta x_s + c_T x_{s-1})$$

$$= x_{\lambda^*} + c_T\sum_{s=2}^{\lambda^*}(x_{s-1} + (s-1)\Delta x_s) + c_T^2\sum_{s=2}^{\lambda^*}(s-1)x_{s-1}$$

$$= x_{\lambda^*} + c_T(\lambda^* - 1)x_{\lambda^*} + c_T^2\sum_{s=2}^{\lambda^*}(s-1)x_{s-1},$$

as

$$\sum_{s=2}^{\lambda^*}(x_{s-1} + (s-1)\Delta x_s) = \sum_{s=2}^{\lambda^*}((s-1)x_s - (s-2)x_{s-1})$$

$$= \sum_{s=2}^{\lambda^*}(s-1)x_s - \sum_{s=2}^{\lambda^*-1}(s-1)x_s = (\lambda^* - 1)x_{\lambda^*}.$$

Hence, using that $c_T = \bar{c}^\gamma T^{-\gamma}$

$$N_{\lambda^*} = \begin{pmatrix} T^{\gamma-3/2}x_1 + \bar{c}^\gamma T^{-3/2}(x_{\lambda^*} - x_1) + \bar{c}^{2\gamma}T^{-\gamma-3/2}\sum_{s=2}^{\lambda^*}x_{s-1} \\ T^{2\gamma-5/2}x_{\lambda^*} + \bar{c}^\gamma T^{\gamma-5/2}(\lambda^* - 1)x_{\lambda^*} + \bar{c}^{2\gamma}T^{-5/2}\sum_{s=2}^{\lambda^*}(s-1)x_{s-1} \end{pmatrix}.$$

Then, if $\gamma = 0$,

$$N_{\lambda^*} \xrightarrow{d} \begin{pmatrix} \psi(1)\sigma \int_0^{\bar{r}} W(s)ds \\ \psi(1)\sigma \int_0^{\bar{r}} sW(s)ds \end{pmatrix}. \tag{A.6}$$

If $\gamma = 1$,

$$N_{\lambda^*} \xrightarrow{d} \begin{pmatrix} 0 \\ \psi(1)\sigma(1 + \bar{c}\bar{r})W(\bar{r}) + \psi(1)\sigma\bar{c}^2 \int_0^{\bar{r}} sW(s)ds \end{pmatrix}. \tag{A.7}$$

Putting everything together we get for $\gamma = 0$

$$\begin{pmatrix} T^{1/2}(\hat{\beta}_{1,\lfloor Tr \rfloor} - \beta_1) \\ T^{-1/2}(\hat{\beta}_{2,\lfloor Tr \rfloor} - \beta_2) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \psi(1)\sigma \left(4\bar{r}^{-1} \int_0^{\bar{r}} W(s)ds - 6\bar{r}^{-2} \int_0^{\bar{r}} sW(s)ds \right) \\ \psi(1)\sigma \left(-6\bar{r}^{-2} \int_0^{\bar{r}} W(s)ds + 12\bar{r}^{-3} \int_0^{\bar{r}} sW(s)ds \right) \end{pmatrix}, \tag{A.8}$$

and for $\gamma = 1$

$$\begin{pmatrix} T^{1/2}(\hat{\beta}_{1,\lfloor Tr \rfloor} - \beta_1) \\ T^{-1/2}(\hat{\beta}_{2,\lfloor Tr \rfloor} - \beta_2) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} 0 \\ \psi(1)\sigma\bar{r}^{-1}(1 + \bar{c}\bar{r} + \frac{1}{3}\bar{c}^2\bar{r}^2)^{-1} \left[(1 + \bar{c}\bar{r})W(\bar{r}) + \bar{c}^2 \int_0^{\bar{r}} sW(s)ds \right] \end{pmatrix}. \tag{A.9}$$

This completes the proof. □

Lemma A.4 provides the invariance principle for y_t^d, which follows straightforwardly from the previous lemma.

Lemma A.4. *Let Assumption 1 hold and let $y_{t,\gamma,\lambda}^d$ be defined as in (2) with $\gamma = 0, 1$ and $\lambda \in [0, 1]$. We have that*

$$T^{-1/2}y_{\lfloor Tr \rfloor}^d \xrightarrow{d} \psi(1)\sigma W_\gamma(r, \lambda),$$

where

$$W_0(r, \lambda) = W(r) - 2\bar{r}^{-2}(2 - 3r) \int_0^{\bar{r}} W(s)ds - 6\bar{r}^{-3}(2r - 1) \int_0^{\bar{r}} sW(s)ds,$$

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

Proof of Lemma A.4. We have that

$$T^{-1/2}y_{\lfloor Tr \rfloor}^d = T^{-1/2}y_{\lfloor Tr \rfloor} - T^{-1/2}\hat{\beta}_{\lfloor Tr \rfloor}' z_{\lfloor Tr \rfloor}$$

$$= T^{-1/2}x_{\lfloor Tr \rfloor} - T^{-1/2}(\hat{\beta}_{1,\lfloor Tr \rfloor} - \beta_1) - \frac{\lfloor Tr \rfloor}{T}T^{1/2}(\hat{\beta}_{2,\lfloor Tr \rfloor} - \beta_2)$$

By Lemma A.2 we have that

$$T^{-1/2}x_{\lfloor Tr \rfloor} \xrightarrow{d} \psi(1)\sigma W(r).$$

The result then follows straightforwardly from Lemma A.3. □

Proof of Lemma 1. Part (a) follows directly from Lemma A.4 using the continuous mapping theorem. For part (b), we write

$$T^{-1}\sum_{t=1}^{T} y_{t-1}^d \varepsilon_{p,t}^d = T^{-1}\sum_{t=1}^{T} y_{t-1}^d \varepsilon_t^d + T^{-1}\sum_{t=1}^{T} y_{t-1}^d (\varepsilon_{p,t}^d - \varepsilon_t^d).$$

We want to show that $T^{-1}\sum_{t=1}^{T} y_{t-1}^d (\varepsilon_{p,t}^d - \varepsilon_t^d) = o_p(1)$. Note that

$$\varepsilon_{p,t}^d - \varepsilon_t^d = \sum_{j=p+1}^{\infty} \phi_j u_{t-j} + \sum_{j=p+1}^{\infty} \phi_j \Delta[(\hat{\beta}_{t-j} - \beta)' z_{t-j}].$$

As $y_t^d = x_t - (\hat{\beta}_t - \beta)' z_t$, we have

$$T^{-1}\sum_{t=1}^{T} y_{t-1}^d (\varepsilon_{p,t}^d - \varepsilon_t^d) = T^{-1}\sum_{t=1}^{T} \left[x_{t-1} \sum_{j=p+1}^{\infty} \phi_j u_{t-j} + x_{t-1} \sum_{j=p+1}^{\infty} \phi_j \Delta[(\hat{\beta}_{t-j} - \beta)' z_{t-j}] \right.$$

$$- (\hat{\beta}_{t-1} - \beta)' z_{t-1} \sum_{j=p+1}^{\infty} \phi_j u_{t-j}$$

$$\left. - (\hat{\beta}_{t-1} - \beta)' z_{t-1} \sum_{j=p+1}^{\infty} \phi_j \Delta[(\hat{\beta}_{t-j} - \beta)' z_{t-j}] \right]$$

$$= A_T^b + B_T^b - C_T^b - D_T^b.$$

It follows from Chang and Park (2002, Proof of Lemma 3.1a) that $A_T^b = o_p(1)$. Then,

$$|B_T^b| = \left| T^{-1}\sum_{t=1}^{T} x_{t-1} \sum_{j=p+1}^{\infty} \phi_j \Delta[(\hat{\beta}_{t-j} - \beta)' z_{t-j}] \right|$$

$$\leq \left(T^{-1}\sum_{t=1}^{T} x_{t-1}^2 \right)^{1/2} \left(T^{-1}\sum_{t=1}^{T} \left[\sum_{j=p+1}^{\infty} \phi_j \Delta[(\hat{\beta}_{t-j} - \beta)' z_{t-j}] \right]^2 \right)^{1/2}$$

$$\leq \left(T^{-1}\sum_{t=1}^{T} x_{t-1}^2 \right)^{1/2} \left(\sum_{j=p+1}^{\infty} |\phi_j| \left[T^{-1}\sum_{t=1}^{T} \left\{ \Delta[(\hat{\beta}_{t-j} - \beta)' z_{t-j}] \right\}^2 \right]^{1/2} \right)$$

$$= \left(T^{-1}\sum_{t=1}^{T} x_{t-1}^2 \right)^{1/2} \left(\sum_{j=p+1}^{\infty} |\phi_j| \left[B_T^{b'} \right]^{1/2} \right)$$

$$= O_p(T^{1/2}) o(p^{-1}) O_p(T^{-1/2}) = o_p(p^{-1}),$$

where we use Cauchy's inequality followed by Minkowski's inequality. The result follows from the fact that $B_T^{b'} = O_p(T^{-1})$. To see this note that

$$B_T^{b'} = T^{-1}\sum_{t=1}^{T} \{ \Delta(\hat{\beta}_{1,t-j} - \beta_1) + \Delta[(\hat{\beta}_{2,t-j} - \beta_2)(t - j)] \}^2$$

For ease of exposition suppose that $\lambda = 0$ and so $\lambda^* = t$. The general case follows by splitting the

sample according to where $t \le \lambda^*$ and $t > \lambda^*$ and combining the proof for $\lambda = 0$ with the (trivial) proof for $\lambda = 1$. Let $M_t^{i,j}$ denote the (i,j)-th element of M_t^{-1} and let N_{t-j}^i be the i-th element of N_t, where M_t and N_t are defined in equation (A.3). Then

$$
\begin{aligned}
\Delta(\hat{\beta}_{1,t-j} - \beta_1) &= T^{1/2} M_{t-j}^{1,1} N_{t-j}^1 + T^{1/2} M_{t-j}^{1,2} N_{t-j}^2 - T^{1/2} M_{t-j-1}^{1,1} N_{t-j-1}^1 - T^{1/2} M_{t-j-1}^{1,2} N_{t-j-1}^2 \\
&= T^{1/2}(M_{t-j}^{1,1} N_{t-j}^1 - M_{t-j-1}^{1,1} N_{t-j-1}^1) + T^{1/2}(M_{t-j}^{1,2} N_{t-j}^2 - M_{t-j-1}^{1,2} N_{t-j-1}^2) \\
&= T^{1/2} M N_{t-j}^{1,1} + T^{1/2} M N_{t-j}^{1,2},
\end{aligned}
$$

and

$$
\begin{aligned}
\Delta[(\hat{\beta}_{2,t-j} - \beta_2)(t-j)] &= T^{-1/2} M_{t-j}^{2,1} N_{t-j}^1 (t-j) + T^{-1/2} M_{t-j}^{2,2} N_{t-j}^2 (t-j) \\
&\quad - T^{-1/2} M_{t-j-1}^{2,1} N_{t-j-1}^1 (t-j-1) - T^{-1/2} M_{t-j-1}^{2,2} N_{t-j-1}^2 (t-j-1) \\
&= T^{-1/2}(M_{t-j}^{2,1} N_{t-j}^1 (t-j) - M_{t-j-1}^{2,1} N_{t-j-1}^1 (t-j-1)) \\
&\quad + T^{-1/2}(M_{t-j}^{2,2} N_{t-j}^2 (t-j) - M_{t-j-1}^{1,2} N_{t-j-1}^2 (t-j-1)) \\
&= T^{-1/2} M N_{t-j}^{2,1}(t-j-1) + T^{-1/2} M_{t-j}^{2,1} N_{t-j}^1 \\
&\quad + T^{-1/2} M N_{t-j}^{2,2}(t-j+1) + T^{-1/2} M_{t-j}^{2,2} N_{t-j}^2 \\
&= T^{-1/2} M N 2_{t-j}^{2,1} + T^{-1/2} M N 2_{t-j}^{2,2}.
\end{aligned}
$$

Then

$$
\begin{aligned}
B_T^{b'} &= T^{-1} \sum_{t=1}^{T} \{ T^{1/2} M N_{t-j}^{1,1} + T^{1/2} M N_{t-j}^{1,2} + T^{-1/2} M N 2_{t-j}^{2,1} + T^{-1/2} M N 2_{t-j}^{2,2} \}^2 \\
&\le 4 T^{-1} \sum_{t=1}^{T} \{ T (M N_{t-j}^{1,1})^2 + T(M N_{t-j}^{1,2})^2 + T^{-1}(M N 2_{t-j}^{2,1})^2 + T^{-1}(M N 2_{t-j}^{2,2})^2 \}.
\end{aligned}
$$

Now

$$
\begin{aligned}
M N_{t-j}^{i,1} &= M_{t-j}^{i,1} \left[T^{\gamma-3/2} x_1 + \bar{c}^{\gamma} T^{-3/2}(x_{t-j} - x_1) + \bar{c}^{2\gamma} T^{-\gamma-3/2} \sum_{s=2}^{t-j} x_{s-1} \right] \\
&\quad - M_{t-j-1}^{i,1} \left[T^{\gamma-3/2} x_1 + \bar{c}^{\gamma} T^{-3/2}(x_{t-j-1} - x_1) + \bar{c}^{2\gamma} T^{-\gamma-3/2} \sum_{s=2}^{t-j-1} x_{s-1} \right] \\
&= \left(M_{t-j}^{i,1} - M_{t-j-1}^{i,1} \right) \left[T^{\gamma-3/2} x_1 + \bar{c}^{\gamma} T^{-3/2}(x_{t-j-1} - x_1) + \bar{c}^{2\gamma} T^{-\gamma-3/2} \sum_{s=2}^{t-j-1} x_{s-1} \right] \\
&\quad + M_{t-j}^{i,1} \left[\bar{c}^{\gamma} T^{-3/2} u_{t-j} + \bar{c}^{2\gamma} T^{-\gamma-3/2} x_{t-j-1} \right],
\end{aligned}
$$

and

$$MN_{t-j}^{i,2} = M_{t-j}^{i,2} \left[T^{2\gamma - 5/2} x_{t-j} + \bar{c}^\gamma T^{\gamma - 5/2}(t-j-1)x_{t-j} + \bar{c}^{2\gamma} T^{-5/2} \sum_{s=2}^{t-j} (s-1)x_{s-1} \right]$$

$$- M_{t-j-1}^{i,2} \left[T^{2\gamma - 5/2} x_{t-j-1} + \bar{c}^\gamma T^{\gamma - 5/2}(t-j-2)x_{t-j-1} + \bar{c}^{2\gamma} T^{-5/2} \sum_{s=2}^{t-j-1} (s-1)x_{s-1} \right]$$

$$= \left(M_{t-j}^{i,1} - M_{t-j-1}^{i,2} \right) \left[T^{2\gamma - 5/2} x_{t-j-1} + \bar{c}^\gamma T^{\gamma - 5/2}(t-j-2)x_{t-j-1} + \bar{c}^{2\gamma} T^{-5/2} \sum_{s=2}^{t-j-1} (s-1)x_{s-1} \right]$$

$$+ M_{t-j}^{i,2} \left[T^{2\gamma - 5/2} u_{t-j} + \bar{c}^\gamma T^{\gamma - 5/2}(x_{t-j-1} + (t-j-1)u_{t-j}) + \bar{c}^{2\gamma} T^{-5/2}(t-j-1)x_{t-j-1} \right].$$

It follows from the proof of Lemma A.3 that

$$M_{t-j} = \begin{bmatrix} 4(t/T)^{-1} & -6(t/T)^{-2} \\ -6(t/T)^{-2} & 12(t/T)^{-3} \end{bmatrix} + O(T^{-1})$$

for $\gamma = 0$ and

$$M_{t-j} = \begin{bmatrix} 1 & 0 \\ 0 & (t/T)^{-1}[1 + \bar{c}(t/T) + \frac{1}{3}\bar{c}^2(t/T)^2]^{-1} \end{bmatrix} + O(T^{-1})$$

for $\gamma = 1$. Then for $\gamma = 0$,

$$\left| M_{t-j}^{1,1} - M_{t-j-1}^{1,1} \right| = 4 \left| \frac{T^{-1}}{T^{-2}t(t-1)} \right| + O(T^{-1}) = O(T^{-1})$$

$$\left| M_{t-j}^{1,2} - M_{t-j-1}^{1,2} \right| = 6 \left| \frac{T^{-2}(2t-1)}{T^{-4}t^2(t-1)^2} \right| + O(T^{-1}) = O(T^{-1})$$

$$\left| M_{t-j}^{2,2} - M_{t-j-1}^{1,2} \right| = 12 \left| \frac{T^{-3}(3t^2 + 3t - 1)}{T^{-6}t^3(t-1)^3} \right| + O(T^{-1}) = O(T^{-1}),$$

while for $\gamma = 1$,

$$|M_{t-j}^{1,1} - M_{t-j-1}^{1,1}| = |1 - 1| + O(T^{-1}) = O(T^{-1})$$
$$|M_{t-j}^{1,2} - M_{t-j-1}^{1,2}| = 0 + O(T^{-1})$$
$$|M_{t-j}^{2,2} - M_{t-j-1}^{1,2}| = \left| \frac{T^{-1}(1 + \bar{c}(2t-1)/T + \bar{c}^2(3t^2 + 3t - 1)/T^2)}{[1 + \bar{c}(t/T) + \frac{1}{3}\bar{c}^2(t/T)^2][1 + \bar{c}((t-1)/T) + \frac{1}{3}\bar{c}^2((t-1)/T)^2]t(t-1)/T^2} \right| + O(T^{-1})$$
$$= O(T^{-1}).$$

Then

$$\left| MN_{t-j}^{i,1} \right| \leq \left| M_{t-j}^{i,1} - M_{t-j-1}^{i,1} \right| \left| T^{\gamma - 3/2} x_1 + \bar{c}^\gamma T^{-3/2}(x_{t-j-1} - x_1) + \bar{c}^{2\gamma} T^{-\gamma - 3/2} \sum_{s=2}^{t-j-1} x_{s-1} \right|$$

$$+ \left| M_{t-j}^{i,1} \right| \left| \bar{c}^\gamma T^{-3/2} u_{t-j} + \bar{c}^{2\gamma} T^{-\gamma - 3/2} x_{t-j-1} \right|$$

$$= O(T^{-1})O_p(1) + O(1)O_p(T^{-1}) = O_p(T^{-1}),$$

29

and

$$\left| MN_{t-j}^{i,2} \right| \le \left| M_{t-j}^{i,1} - M_{t-j-1}^{i,2} \right| \left| T^{2\gamma-5/2} x_{t-j-1} + \bar{c}^{\gamma} T^{\gamma-5/2} (t-j-2) x_{t-j-1} + \bar{c}^{2\gamma} T^{-5/2} \sum_{s=2}^{t-j-1} (s-1) x_{s-1} \right|$$

$$+ \left| M_{t-j}^{i,2} \right| \left| T^{2\gamma-5/2} u_{t-j} + \bar{c}^{\gamma} T^{\gamma-5/2} (x_{t-j-1} + (t-j+1) u_{t-j}) + \bar{c}^{2\gamma} T^{-5/2} (t-j-1) x_{t-j-1} \right|$$

$$= O(T^{-1}) O_p(1) + O(1) O_p(T^{-1}) = O_p(T^{-1}).$$

It then follows that

$$MN2_{t-j}^{2,i} = MN_{t-j}^{2,i}(t-j-1) + M_{t-j}^{2,i} N_{t-j}^{i} = O_p(1),$$

and consequently that

$$B_T^{b'} \le 4T^{-1} \sum_{t=1}^{T} \{ T(MN_{t-j}^{1,1})^2 + T(MN_{t-j}^{1,2})^2 + T^{-1}(MN2_{t-j}^{2,1})^2 + T^{-1}(MN2_{t-j}^{2,2})^2 \} = O_p(T^{-1}).$$

We can now continue with C_T^b. We have

$$C_T^b = T^{-1} \sum_{t=1}^{T} (\hat{\beta}_{1,t-1} - \beta_1) \left(\sum_{j=p+1}^{\infty} \phi_j u_{t-j} \right) + T^{-1} \sum_{t=1}^{T} (\hat{\beta}_{2,t-1} - \beta_2)(t-1) \left(\sum_{j=p+1}^{\infty} \phi_j u_{t-j} \right)$$

$$= C_{1T}^b + C_{2T}^b.$$

Define $\psi_{p,j}$ such that

$$\sum_{j=p+1}^{\infty} \phi_j u_{t-j} = \sum_{j=p+1}^{\infty} \psi_{p,j} \varepsilon_{t-j}$$

and note that $\sum_{j=p+1}^{\infty} |\psi_{p,j}| = o(p^{-1})$ (Chang and Park, 2002, Proof of Lemma 3.1a). Then

$$C_{1T}^b = \sum_{j=p+1}^{\infty} \psi_{p,j} T^{-1} \sum_{t=1}^{T} (\hat{\beta}_{1,t-1} - \beta_1) \varepsilon_{t-j}$$

and

$$C_{2T}^b = \sum_{j=p+1}^{\infty} \psi_{p,j} T^{-1} \sum_{t=1}^{T} (t-1)(\hat{\beta}_{2,t-1} - \beta_2) \varepsilon_{t-j}$$

Let $M_{t-j}^{i,j}$ be defined as before. Again we prove the result for $\lambda = 0$, as the general case follows

straightforwardly. Then

$$\sum_{t=1}^{T}(\hat{\beta}_{i,t-1}-\beta_i)\varepsilon_t = T^{3/2-i}\sum_{t=1}^{T}T^{i-3/2}(\hat{\beta}_{i,t-1}-\beta_i)\varepsilon_t$$

$$= T^{3/2-i}\sum_{t=1}^{T}\left[M_{t-1}^{i,1}\left(T^{\gamma-3/2}x_1 + \bar{c}^{\gamma}T^{-3/2}(x_{t-1}-x_1) + \bar{c}^{2\gamma}T^{-\gamma-3/2}\sum_{s=2}^{t-1}x_{s-1}\right)\varepsilon_t\right.$$

$$\left. + M_{t-1}^{i,2}\left(T^{2\gamma-5/2}x_{t-1} + \bar{c}^{\gamma}T^{\gamma-5/2}(t-2)x_{t-1} + \bar{c}^{2\gamma}T^{-5/2}\sum_{s=2}^{t-1}(s-1)x_{s-1}\right)\varepsilon_t\right]$$

$$= O_p(T^{2-i}),$$

as $\sum_{t=1}^{T}M_{t-1}^{i,j}\sum_{s=2}^{t-1}x_{s-1}\varepsilon_t = \sum_{t=2}^{T}x_{t-1}\sum_{s=t+1}^{T}M_{s-1}^{i,j}\varepsilon_s = O_p(T^2)$. It then follows that $C_{1T}^b, C_{2T}^b = o_p(p^{-1})$.

Finally,

$$D_T^b = T^{-1}\sum_{t=1}^{T}(\hat{\beta}_{1,t-1}-\beta_1)\sum_{j=p+1}^{\infty}\phi_j\Delta[(\hat{\beta}_{t-j}-\beta)'z_{t-j}]$$

$$+ T^{-1}\sum_{t=1}^{T}(\hat{\beta}_{2,t-1}-\beta_2)(t-1)\sum_{j=p+1}^{\infty}\phi_j\Delta[(\hat{\beta}_{t-j}-\beta)'z_{t-j}]$$

$$= D_{1T}^b + D_{2T}^b.$$

Then

$$|D_{1T}^b| \leq \left(T^{-1}\sum_{t=1}^{T}(\hat{\beta}_{1,t-1}-\beta_1)^2\right)^{1/2}\left(\sum_{j=p+1}^{\infty}|\phi_j|\left[T^{-1}\sum_{t=1}^{T}(\Delta[(\hat{\beta}_{t-j}-\beta)'z_{t-j}])^2\right]^{1/2}\right)$$

$$= O_p(T^{1/2})o(p^{-1})O_p(T^{-1/2}) = o_p(p^{-1}),$$

$$|D_{2T}^b| \leq \left(T^{-1}\sum_{t=1}^{T}(t-1)^2(\hat{\beta}_{2,t-1}-\beta_2)^2\right)^{1/2}\left(\sum_{j=p+1}^{\infty}|\phi_j|\left[(\Delta[(\hat{\beta}_{t-j}-\beta)'z_{t-j}])^2\right]^{1/2}\right)$$

$$= O_p(T^{1/2})o(p^{-1})O_p(T^{-1/2}) = o_p(p^{-1}).$$

Hence,

$$T^{-1}\sum_{t=1}^{T}y_{t-1}^d\varepsilon_{p,t}^d = T^{-1}\sum_{t=1}^{T}y_{t-1}^d\varepsilon_t^d + o_p(1).$$

Now define

$$\eta_t^d = \sum_{s=1}^{t}\varepsilon_s^d = \sum_{s=1}^{t}\varepsilon_t - \sum_{s=1}^{t}\phi(L)\Delta[(\hat{\beta}_t-\beta)'z_t] = \eta_t - \phi(L)(\hat{\beta}_t-\beta)'z_t,$$

where $\eta_t = \sum_{s=1}^{t}\varepsilon_t$. We can then straightforwardly show that

$$T^{-1/2}\eta_{\lfloor Tr\rfloor}^d = T^{-1/2}\eta_{\lfloor Tr\rfloor} - \phi(1)T^{-1/2}(\hat{\beta}_t-\beta)'z_t + o_p(1) \xrightarrow{d} \sigma W_{\gamma}(r,\lambda).$$

By the Beveridge-Nelson decomposition we can write

$$\Delta y_t^d = \psi(1)\varepsilon_t^d + (\tilde{u}_{t-1} - \tilde{u}_t),$$

with $\tilde{u}_t = \sum_{j=0}^{\infty} \tilde{\psi}_j \varepsilon_{t-j}^d$ and $\tilde{\psi}_j = \sum_{k=j+1}^{\infty} \psi_k$. Then

$$y_t^d = \psi(1)\eta_t^d + \tilde{u}_0 - \tilde{u}_t.$$

Then we may write

$$T^{-1}\sum_{t=1}^{T} y_{t-1}^d \varepsilon_t^d = \psi(1)T^{-1}\sum_{t=1}^{T} \eta_{t-1}^d \varepsilon_t^d + \tilde{u}_0 T^{-1}\sum_{t=1}^{T} \varepsilon_t^d - T^{-1}\sum_{t=1}^{T} \tilde{u}_{t-1}\varepsilon_t^d.$$

Now

$$\tilde{u}_0 T^{-1}\sum_{t=1}^{T} \varepsilon_t^d = T^{-1}\tilde{u}_0 \eta_T^d - T^{-1}\tilde{u}_0(\hat{\beta}_T - \beta)' z_T = O_p(T^{-1/2})$$

and

$$
\begin{aligned}
T^{-1}\sum_{t=1}^{T} \tilde{u}_{t-1}\varepsilon_t^d &= T^{-1}\sum_{t=1}^{T}\sum_{j=0}^{\infty} \tilde{\psi}_j \varepsilon_{t-j-1}^d \varepsilon_t^d \\
&= T^{-1}\sum_{t=1}^{T}\sum_{j=0}^{\infty} \tilde{\psi}_j \varepsilon_{t-j-1}\varepsilon_t - T^{-1}\sum_{t=1}^{T}\sum_{j=0}^{\infty} \tilde{\psi}_j \varepsilon_{t-j-1}\phi(L)\Delta[(\hat{\beta}_t - \beta)' z_t] \\
&\quad - T^{-1}\sum_{t=1}^{T}\sum_{j=0}^{\infty} \tilde{\psi}_j \phi(L)\Delta[(\hat{\beta}_{t-j-1} - \beta)' z_{t-j-1}]\varepsilon_t \\
&\quad + T^{-1}\sum_{t=1}^{T}\sum_{j=0}^{\infty} \tilde{\psi}_j \phi(L)\Delta[(\hat{\beta}_{t-j-1} - \beta)' z_{t-j-1}]\phi(L)\Delta[(\hat{\beta}_t - \beta)' z_t] \\
&= O_p(T^{-1/2}),
\end{aligned}
$$

by the results for B_T^b, C_T^b and Assumption 1 (ii). Hence,

$$T^{-1}\sum_{t=1}^{T} y_{t-1}^d \varepsilon_t^d = \psi(1)T^{-1}\sum_{t=1}^{T} \eta_{t-1}^d \varepsilon_t^d + o_p(1).$$

As

$$T^{-1}\sum_{t=1}^{T} \eta_{t-1}^d \varepsilon_t^d = \frac{1}{2}\left(\eta_T^{d2} - \eta_0^{d2} - T^{-1}\sum_{t=1}^{T} \varepsilon_t^{d2}\right) \xrightarrow{d} \frac{\sigma}{2}\left(W_\gamma(1)^2 - W_\gamma(0,\lambda)^2 - 1\right),$$

the result follows.

We continue with (c). Let Ω_{pp} be defined as in Chang and Park (2002, Proof of Lemma 3.2), i.e. $\Omega_{pp} = (\Gamma_{i-j})_{i,j=1}^{p}$ where $\Gamma_k = \mathrm{E}(u_t u_{t-k})$.

Let $z_{p,t} = (z'_{t-1}, \ldots, z'_{t-p})'$ and $B_{p,t}$ an $p \times 2p$-matrix with

$$
B_{p,t} = \begin{bmatrix} (\hat{\beta}_{1,t-1} - \beta) & (\hat{\beta}_{2,t-1} - \beta) & 0 & \cdots & & 0 \\ & & \ddots & & & \\ 0 & & \cdots & 0 & (\hat{\beta}_{1,t-p} - \beta) & (\hat{\beta}_{1,t-p} - \beta) \end{bmatrix},
$$

such that $w^d_{p,t} = w_{p,t} - \Delta[B_{p,t} z_{p,t}]$. Then

$$
\left\| T^{-1} \sum_{t=1}^{T} w^d_{p,t} w^{d\prime}_{p,t} - \Omega_{pp} \right\| \leq \left\| T^{-1} \sum_{t=1}^{T} w_{p,t} w'_{p,t} - \Omega_{pp} \right\| + 2 \left\| T^{-1} \sum_{t=1}^{T} w_{p,t} \Delta[z'_{p,t} B'_{p,t}] \right\|
$$

$$
+ \left\| T^{-1} \sum_{t=1}^{T} \Delta[B_{p,t} z_{p,t}] \Delta[z'_{p,t} B'_{p,t}] \right\|
$$

$$
= A^c_T + 2 B^c_T + C^c_T.
$$

By Berk (1974, Proof of Lemma 3) and Chang and Park (2002, Proof of Lemma 3.2a) $A^c_T = O_p(T^{-1/2} p)$. Next consider

$$
T^{-1} \sum_{t=1}^{T} u_{t-i} \Delta[(\hat{\beta}_{t-j} - \beta)' z_{t-j}] = O_p(T^{-1}),
$$

which follows from combining the arguments used for $B^{b\prime}_T$ and C^b_T.

It also follows from $B^{b\prime}_T$ that

$$
\left| T^{-1} \sum_{t=1}^{T} \Delta[(\hat{\beta}_{t-i} - \beta)' z_{t-i}] \Delta[(\hat{\beta}_{t-j} - \beta)' z_{t-j}] \right|
$$

$$
\leq \left(T^{-1} \sum_{t=1}^{T} \Delta[(\hat{\beta}_{t-i} - \beta)' z_{t-i}] \right)^{1/2} \left(T^{-1} \sum_{t=1}^{T} \left[\Delta[(\hat{\beta}_{t-j} - \beta)' z_{t-j}] \right]^2 \right)^{1/2} = O_p(T^{-1/2}) O_p(T^{-1/2}).
$$

As this holds uniformly in $i, j = 1, \ldots, p$, we can conclude that $B^c_T, C^c_T = O_p(T^{-1} p)$.

The proof now follows as in Chang and Park (2002, Proof of Lemma 3.2a).

Next we look at (d). We have that

$$
\left| T^{-1} \sum_{t=1}^{T} y^d_{t-1} w^d_{p,t} \right| \leq \left| T^{-1} \sum_{t=1}^{T} x_{t-1} w_{p,t} \right| + \left| T^{-1} \sum_{t=1}^{T} x_{t-1} \Delta[B_{p,t} z_{p,t}] \right|
$$

$$
+ \left| T^{-1} \sum_{t=1}^{T} (\hat{\beta}_{t-1} - \beta)' z_{t-1} w_{p,t} \right| + \left| T^{-1} \sum_{t=1}^{T} (\hat{\beta}_{t-1} - \beta)' z_{t-1} \Delta[B_{p,t} z_{p,t}] \right|
$$

$$
= A^d_T + B^d_T + C^d_T + D^d_T.
$$

Now $A_T^d = O_p(p^{1/2})$ by Chang and Park (2002, Proof of Lemma 3.2b). Furthermore

$$\left| T^{-1} \sum_{t=1}^{T} x_{t-1} \Delta(\hat{\beta}_{t-j} - \beta)' z_{t-j} \right| \leq \left(T^{-1} \sum_{t=1}^{T} x_{t-1}^2 \right)^{1/2} \left(T^{-1} \sum_{t=1}^{T} \left[\Delta(\hat{\beta}_{t-j} - \beta)' z_{t-j} \right]^2 \right)^{1/2}$$
$$= O_p(T^{1/2}) O_p(T^{-1/2}),$$

from which we can conclude that $B_T^d = O_p(p^{1/2})$. It also follows from the arguments used for C_T^b that

$$T^{-1} \sum_{t=1}^{T} (\hat{\beta}_{t-1} - \beta)' z_{t-1} u_{t-j} = T^{-1} \sum_{t=1}^{T} (\hat{\beta}_{1,t-1} - \beta_1) u_{t-j} + T^{-1} \sum_{t=1}^{T} (\hat{\beta}_{2,t-1} - \beta_2)(t-1) u_{t-j}$$
$$= O_p(1) + O_p(1),$$

by which $C_T^d = O_p(p^{1/2})$. Finally, it follows again from $B_T^{b'}$ that

$$\left| T^{-1} \sum_{t=1}^{T} (\hat{\beta}_{t-1} - \beta)' z_{t-1} \Delta[(\hat{\beta}_{t-j} - \beta)' z_{t-j}] \right|$$
$$= \left| T^{-1} \sum_{t=1}^{T} (\hat{\beta}_{1,t-1} - \beta_1) \Delta[(\hat{\beta}_{t-j} - \beta)' z_{t-j}] + T^{-1} \sum_{t=1}^{T} (\hat{\beta}_{2,t-1} - \beta_2)(t-1) \Delta[(\hat{\beta}_{t-j} - \beta)' z_{t-j}] \right|$$
$$\leq \left(T^{-1} \sum_{t=1}^{T} (\hat{\beta}_{1,t-1} - \beta_1)^2 \right)^{1/2} \left(T^{-1} \sum_{t=1}^{T} \left[\Delta[(\hat{\beta}_{t-j} - \beta)' z_{t-j}] \right]^2 \right)^{1/2}$$
$$+ \left(T^{-1} \sum_{t=1}^{T} (t-1)^2 (\hat{\beta}_{2,t-1} - \beta_2)^2 \right)^{1/2} \left(T^{-1} \sum_{t=1}^{T} (t-1)^2 \left[\Delta[(\hat{\beta}_{t-j} - \beta)' z_{t-j}] \right]^2 \right)^{1/2}$$
$$= O_p(T^{1/2}) O_p(T^{-1/2}) + O_p(T^{1/2}) O_p(T^{-1/2}),$$

by which $D_T^d = O_p(p^{1/2})$. This concludes the proof for part (d).

For part (e) we can write

$$\left| T^{-1} \sum_{t=1}^{T} w_{p,t}^d \varepsilon_{p,t}^d \right| \leq \left| T^{-1} \sum_{t=1}^{T} w_{p,t} \varepsilon_{p,t} \right| + \left| T^{-1} \sum_{t=1}^{T} \Delta[B_{p,t} z_{p,t}] \varepsilon_{p,t} \right|$$
$$+ \left| T^{-1} \sum_{t=1}^{T} w_{p,t} \phi_p(L) \Delta[(\hat{\beta}_t - \beta)' z_t] \right|$$
$$+ \left| T^{-1} \sum_{t=1}^{T} \Delta[B_{p,t} z_{p,t}] \phi_p(L) \Delta[(\hat{\beta}_t - \beta)' z_t] \right|$$
$$= A_T^e + B_T^e + C_T^e + D_T^e.$$

By Chang and Park (2002, Proof of Lemma 3.2c) we have that $A_T^e = o_p(p^{-1/2})$. Appealing again to $B_T^{b'}$ and C_T^b, we have that

$$T^{-1} \sum_{t=1}^{T} \Delta[(\hat{\beta}_{t-j} - \beta)' z_{t-j}] \varepsilon_{p,t} = \sum_{j=p+1}^{\infty} \psi_{p,j} T^{-1} \sum_{t=1}^{T} \Delta[(\hat{\beta}_{t-j} - \beta)' z_{t-j}] \varepsilon_{t-j} + T^{-1} \sum_{t=1}^{T} \Delta[(\hat{\beta}_{t-j} - \beta)' z_{t-j}] \varepsilon_t$$
$$= O_p(T^{-1}) o_p(p^{-1}) + O_p(T^{-1}).$$

34

From the same reasoning it follows that

$$T^{-1}\sum_{t=1}^{T}u_{t-j}\phi_p(L)\Delta[(\hat{\beta}_t-\beta)'z_t]=O_p(T^{-1}),$$

and

$$\left|T^{-1}\sum_{t=1}^{T}\Delta[(\hat{\beta}_{t-i}-\beta)'z_{t-i}]\phi_p(L)\Delta[(\hat{\beta}_t-\beta)'z_t]\right|$$
$$\leq\left(T^{-1}\sum_{t=1}^{T}\left[\Delta[(\hat{\beta}_{t-i}-\beta)'z_{t-i}]\right]^2\right)^{1/2}\left(\sum_{j=0}^{p}|\phi_j|\left[T^{-1}\sum_{t=1}^{T}\left[\Delta[(\hat{\beta}_{t-j}-\beta)'z_{t-j}]\right]^2\right]^{1/2}\right)$$
$$=O_p(T^{-1/2})O_p(T^{-1/2}).$$

Therefore $B_T^e, C_T^e, D_T^e=O_p(T^{-1}p^{1/2})$, which concludes the proof. $\square$

Corollary A.1. *Let Assumptions 1 and 2 hold. Let A_T and B_T be defined as in (14). Let $\gamma=0,1$ and $\lambda\in[0,1]$. Then*

1. $T^{-1}A_T\xrightarrow{d}\frac{1}{2}\psi(1)\sigma^2[W_\gamma(1,\lambda)^2-W_\gamma(0,\lambda)^2-1]$,

2. $T^{-2}B_T\xrightarrow{d}\psi(1)^2\sigma^2\int_0^1 W_\gamma(r,\lambda)^2dr$.

Proof of Corollary A.1. Given the expressions for A_T and B_T in (14), it follows immediately from Lemma 1 that

$$T^{-1}A_T=T^{-1}Y_{-1}^{d\prime}\varepsilon_p^d-T^{-1}Y_{-1}^{d\prime}M_p^d\left(T^{-1}M_p^{d\prime}M_p^d\right)^{-1}T^{-1}M_p^{d\prime}\varepsilon_p^d$$
$$=T^{-1}Y_{-1}^{d\prime}\varepsilon_p^d+O_p(p^{1/2})O_p(1)o_p(p^{-1/2})=T^{-1}Y_{-1}^{d\prime}\varepsilon_p^d+o_p(1)$$
$$\xrightarrow{d}\frac{1}{2}\psi(1)\sigma^2[W_\gamma(1,\lambda)^2-W_\gamma(0,\lambda)^2-1],$$

and

$$T^{-2}B_T=T^{-2}Y_{-1}^{d\prime}Y_{-1}^d-T^{-1}(T^{-1}Y_{-1}^{d\prime}M_p^d)\left(T^{-1}M_p^{d\prime}M_p^d\right)^{-1}T^{-1}M_p^{d\prime}Y_{-1}^d$$
$$=T^{-2}Y_{-1}^{d\prime}Y_{-1}^d+T^{-1}O_p(p^{1/2})O_p(1)O_p(p^{1/2})=T^{-1}Y_{-1}^{d\prime}Y_{-1}^d+o_p(1)$$
$$\xrightarrow{d}\psi(1)^2\sigma^2\int_0^1 W_\gamma(r,\lambda)^2dr.$$

This completes the proof. $\square$

Proof of Lemma 2. Note that

$$T\hat{\sigma}^2=(\Delta Y^d-Y_{-1}^d\hat{\delta})'(I-M_p^d(M_p^{d\prime}M_p^d)^{-1}M_p^{d\prime})(\Delta Y^d-Y_{-1}^d\hat{\delta})$$
$$=\Delta Y^{d\prime}(I-M_p^d(M_p^{d\prime}M_p^d)^{-1}M_p^{d\prime})\Delta Y^d-\Delta Y^{d\prime}(I-M_p^d(M_p^{d\prime}M_p^d)^{-1}M_p^{d\prime})Y_{-1}^d\hat{\delta}$$
$$-\hat{\delta}Y_{-1}^{d\prime}(I-M_p^d(M_p^{d\prime}M_p^d)^{-1}M_p^{d\prime})\Delta Y^d+\hat{\delta}Y_{-1}^{d\prime}(I-M_p^d(M_p^{d\prime}M_p^d)^{-1}M_p^{d\prime})Y_{-1}^d\hat{\delta}$$
$$=\varepsilon_p^{d\prime}(I-M_p^d(M_p^{d\prime}M_p^d)^{-1}M_p^{d\prime})\varepsilon_p^d-\varepsilon_p^{d\prime}(I-M_p^d(M_p^{d\prime}M_p^d)^{-1}M_p^{d\prime})Y_{-1}^d\hat{\delta}$$
$$-\hat{\delta}Y_{-1}^{d\prime}(I-M_p^d(M_p^{d\prime}M_p^d)^{-1}M_p^{d\prime})\varepsilon_p^d+\hat{\delta}Y_{-1}^{d\prime}(I-M_p^d(M_p^{d\prime}M_p^d)^{-1}M_p^{d\prime})Y_{-1}^d\hat{\delta}.$$

which we can write as

$$\hat{\sigma}^2 = C_T - 2D_T + E_T$$

We first consider C_T. Write

$$C_T = T^{-1}\varepsilon_p^{d\prime}\varepsilon_p^d - T^{-1}\varepsilon_p^{d\prime}M_p^d(M_p^{d\prime}M_p^d)^{-1}M_p^{d\prime}\varepsilon_p^d.$$

Given the results from Lemma 1, we have that

$$T^{-1}\left|\varepsilon_p^{d\prime}M_p^d(M_p^{d\prime}M_p^d)^{-1}M_p^{d\prime}\varepsilon_p^d\right| \leq \left|T^{-1}\varepsilon_p^{d\prime}M_p^d\right|\left|\left|(T^{-1}M_p^{d\prime}M_p^d)^{-1}\right|\right|\left|T^{-1}M_p^{d\prime}\varepsilon_p^d\right|$$
$$= o_p(p^{-1/2})O_p(1)o_p(p^{-1/2}) = o_p(p^{-1}).$$

Hence,

$$C_T = T^{-1}\varepsilon_p^{d\prime}\varepsilon_p^d + o_p(1).$$

Next we turn to D_T. We can write D_T as

$$D_T = T^{-1}\varepsilon_p^{d\prime}Y_{-1}^d\hat{\delta} - T^{-1}\varepsilon_p^{d\prime}M_p^d(M_p^{d\prime}M_p^d)^{-1}M_p^{d\prime}Y_{-1}^d\hat{\delta}.$$

Again using Lemma 1 and $\hat{\delta} = T^{-1}(T^{-1}A_T)(T^{-2}B_T)^{-1} = O_p(T^{-1})$, we have

$$|D_T| \leq \left|T^{-1}\varepsilon_p^{d\prime}Y_{-1}^d\right||\hat{\delta}| + \left|T^{-1}\varepsilon_p^{d\prime}M_p^d\right|\left|\left|T^{-1}(M_p^{d\prime}M_p^d)^{-1}\right|\right|\left|T^{-1}M_p^{d\prime}Y_{-1}^d\right||\hat{\delta}|$$
$$= O_p(1)O_p(T^{-1}) + o_p(p^{-1/2})O_p(1)O_p(T^{-1}) = O_p(T^{-1}).$$

Finally we look at E_T:

$$E_T = T^{-1}\hat{\delta}Y_{-1}^{d\prime}Y_{-1}^d\hat{\delta} - T^{-1}\hat{\delta}Y_{-1}^{d\prime}M_p^d(M_p^{d\prime}M_p^d)^{-1}M_p^{d\prime}Y_{-1}^d\hat{\delta}.$$

As before, we use the results from Lemma 1 and $\hat{\delta} = O_p(T^{-1})$ to obtain

$$|E_T| \leq T|\hat{\delta}|^2\left|\left|T^{-2}Y_{-1}^{d\prime}Y_{-1}^d\right|\right| + |\hat{\delta}|^2\left|T^{-1}Y_{-1}^{d\prime}M_p^d\right|\left|\left|(T^{-1}M_p^{d\prime}M_p^d)^{-1}\right|\right|\left|T^{-1}M_p^{d\prime}Y_{-1}^d\right|$$
$$= TO_p(T^{-2})O_p(1) + O_p(T^{-2})O_p(p^{1/2})O_p(1)O_p(p^{1/2})$$
$$= O_p(T^{-1}).$$

Therefore, we have that

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{p,t}^{d2} + o_p(1).$$

Now

$$\left|\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{p,t}^{d2}\right)^{1/2} - \left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{p,t}^2\right)^{1/2}\right| \leq \left[\frac{1}{T}\sum_{t=1}^{T}(\varepsilon_{p,t}^d - \varepsilon_{p,t})^2\right]^{1/2}.$$

Then we have that

$$
\frac{1}{T}\sum_{t=1}^{T}(\varepsilon_{p,t}^{d}-\varepsilon_{p,t})^2 = \frac{1}{T}\sum_{t=1}^{T}\left[\Delta[(\hat{\beta}_t-\beta)'z_t] - \sum_{j=1}^{p}\phi_j\Delta[(\hat{\beta}-\beta)'z_{t-j}]\right]^2
$$

$$
= \frac{1}{T}\sum_{t=1}^{T}\left[\Delta[(\hat{\beta}_t-\beta)'z_t] - \sum_{j=1}^{p}\phi_j\Delta[(\hat{\beta}_{t-j}-\beta)'z_{t-j}]\right]^2
$$

$$
\leq \left(\sum_{j=0}^{p}\phi_j^2\right)\left(T^{-1}\sum_{j=0}^{p}\sum_{t=1}^{T}\left(\Delta[(\hat{\beta}_{t-j}-\beta)'z_{t-j}]\right)^2\right)
$$

$$
= O_p(T^{-1}).
$$

Hence,

$$
\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{p,t}^{d2} = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{p,t}^2 + o_p(1).
$$

Then, by Chang and Park (2002, Proof of Lemma 3.1c), we have that

$$
\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{p,t}^2 = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 + o_p(1),
$$

and by the law of large numbers $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 \overset{p}{\to} \sigma^2$. $\qquad\square$

Proof of Theorem 1. We have that

$$
ADF_{\gamma,\lambda} = \frac{T^{-1}A_T}{(T^{-2}B_T\hat{\sigma}^2)^{1/2}} \overset{d}{\to} \frac{\frac{1}{2}\psi(1)\sigma^2(W_\gamma(1,\lambda)^2 - W_\gamma(0,\lambda)^2 - 1)}{\left(\psi(1)^2\sigma^4\int_0^1 W_\gamma(r,\lambda)^2 dr\right)^{1/2}} = \frac{W_\gamma(1,\lambda)^2 - W_\gamma(0,\lambda)^2 - 1}{2\left(\int_0^1 W_\gamma(r,\lambda)^2 dr\right)^{1/2}},
$$

which follows straightforwardly from Corollary A.1 and Lemma 2. $\qquad\square$

A.2   Proofs for Section 4

We start with the proof of the lemma that demonstrates the equivalence of the different detrending techniques in the first step in the bootstrap.

Proof of Lemma 3. For the first part, we first make the step from ADF estimation to estimation under the null of a unit root (denote the vector of autoregressive estimators as $\bar{\Phi}_q$). Then

$$
\hat{\Phi}_q = \bar{\Phi}_q + (M_q^{d\prime}M_q^d)^{-1}M_q^{d\prime}Y_{-1}^d\hat{\delta} = \bar{\Phi}_q + O_p(T^{-1}q^{1/2}).
$$

The next step is to show that $\bar{\Phi}_q = \tilde{\Phi}_q + O_p(T^{-1}q^{1/2})$. Let $\Delta\tilde{B}z = (\Delta[(\tilde{\beta}_1-\beta)'z_1], \ldots, \Delta[(\tilde{\beta}_T-\beta)'z_T])'$.

Then $\varepsilon_q^d = \varepsilon_q - \phi_q(L)\Delta\tilde{B}z$. Now note that

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

Then

$$
|\bar{\phi}_j - \tilde{\phi}_j| = |e_j'(\bar{\Phi}_q - \tilde{\Phi}_q)| \leq |e_j|(\|A_T\| + \|B_T\| + \|C_T\| + \|D_T\|).
$$

First we look at A_T. Note that

$$
\begin{aligned}
(T^{-1}M_q^{d\prime}M_q^d)^{-1} - (T^{-1}M_q'M_q)^{-1} &= (T^{-1}M_q^{d\prime}M_q^d)^{-1} \\
&\quad \times (T^{-1}M_q'M_q - T^{-1}M_q^{d\prime}M_q^d)(T^{-1}M_q'M_q)^{-1}.
\end{aligned}
$$

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

$$
\begin{aligned}
T^{-1}M_q'M_q - T^{-1}M_q^{d\prime}M_q^d &= -T^{-1}M_q'\Delta\tilde{B}z_q \\
&\quad - T^{-1}\Delta\tilde{B}z_q'M_q + \Delta\tilde{B}z_q'\Delta\tilde{B}z_q.
\end{aligned}
$$

Therefore

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

which follows directly from the proof of Lemma 1(c) and (e). It also follows directly from the proof of Lemma 1(c) and (e) that

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

Therefore we may conclude that $\bar{\phi}_j = \tilde{\phi}_{o,j} + O_p(T^{-1}q^{1/2})$ and consequently that $\hat{\phi}_j = \tilde{\phi}_j + O_p(T^{-1}q^{1/2})$ uniformly in j, $1 \leq j \leq q$.

For the second part we have that

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

from which we can conclude that

$$\max_{1 \le t \le T} |\hat{\varepsilon}^d_{q,t} - \tilde{\varepsilon}_{q,t}| \le |\hat{\alpha}| \max_{1 \le t \le T} |y_{t-1}| + \max_{1 \le t \le T} |u_t| \sum_{j=1}^{q} |\hat{\phi}_j - \tilde{\phi}_j| + \max_{1 \le t \le T} |\Delta[(\tilde{\beta}_t - \beta)' z_t]||\hat{\phi}(1)|$$

$$= O_p(T^{-1/2}) + O_p(T^{-1} q^{3/2}) + O_p(T^{-1/2}).$$

This completes the proof. $\qquad\square$

The first step towards an invariance principle for u_t^* is to show that higher than second order moments exist for ε_t^*.

Lemma A.5. *Let Assumptions 1 and 3 hold. Then we have for any $2 < a \le 4$*

$$\mathrm{E}^* |\varepsilon_t^*|^a = O_p(1).$$

Proof of Lemma A.5. We have that

$$\mathrm{E}^* |\varepsilon_t^*|^a = T^{-1} \sum_{t=1}^{T} |\hat{\varepsilon}^d_{q,t} - T^{-1} \sum_{\tau=1}^{T} \hat{\varepsilon}^d_{q,\tau}|^a \le 2^{a-1} T^{-1} \sum_{t=1}^{T} |\hat{\varepsilon}^d_{q,t} - \tilde{\varepsilon}_{q,t}$$

$$- T^{-1} \sum_{\tau=1}^{T} (\hat{\varepsilon}^d_{q,\tau} - \tilde{\varepsilon}_{q,\tau})|^a + 2^{a-1} T^{-1} \sum_{t=1}^{T} |\tilde{\varepsilon}_{q,t} - T^{-1} \sum_{\tau=1}^{T} \tilde{\varepsilon}_{q,\tau}|^a,$$

where the first part is $o_p(1)$ by Lemma 3; the second part is $O_p(1)$ by Park (2002, Lemma 3.2). $\quad\square$

Lemma A.6. *Let Assumptions 1 and 3 hold. Then $\sigma^* \xrightarrow{p} \sigma$.*

Proof of Lemma A.6. Follows directly from Lemma 2. $\qquad\square$

Lemma A.7. *Let Assumptions 1 and 3 hold. Then*

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

Proof of Lemma A.7. Follows directly from Lemma A.5 and A.6 as in Park (2002, Theorem 2.2). $\quad\square$

Lemma A.8. *Let Assumptions 1 and 3 hold. Then*

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

Proof of Lemma A.8. As in Park (2002, p. 478) we need to show that

$$\hat{\phi}(1) \xrightarrow{p} \phi(1), \tag{A.10}$$

$$\mathrm{P}^* \left\{ \max_{1 \le t \le T} |T^{-1/2} \bar{u}_t^*| > \epsilon \right\} = o_p(1), \tag{A.11}$$

where $\bar{u}_t^* = \hat{\phi}(1)^{-1} \sum_{i=1}^{q} (\sum_{j=i}^{q} \hat{\phi}_j) u_{t-i+1}^*$.

We first show (A.10). Park (2002, Lemma 3.1) shows that $\left|\tilde{\phi}(1) - \phi(1)\right| = o_p(1)$; therefore we have that

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

where the first part follows from Lemma 3. Hence $\hat{\phi}(1) = \phi(1) + o_p(1)$. This proves (A.10).

To prove (A.11) we have, as in Park (2002, Proof of Theorem 3.3),

$$\mathrm{P}^* \left\{ \max_{1 \leq t \leq T} |T^{-1/2}\bar{u}_t^*| > \epsilon \right\} \leq T\, \mathrm{P}^* \left\{ |T^{-1/2}\bar{u}_t^*| > \epsilon \right\} \leq (1/\epsilon^a)T^{1-a/2}\, \mathrm{E}^*\, |\bar{u}_t^*|^a.$$

Hence, we have to show that

$$T^{1-a/2}\, \mathrm{E}^*\, |\bar{u}_t^*|^a = o_p(1). \tag{A.12}$$

As in Palm, Smeekes, and Urban (2010, Proof of Theorem 2) this amounts to showing that

$$\sum_{j=1}^{q} j^{1/2}|\hat{\phi}_j| = O_p(1).$$

We can write

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

where the first part follows from Lemma 3 and the second part follows from Palm et al. (2010, Proof of Theorem 2). This concludes the proof of this theorem. $\qquad\square$

Lemma A.9. *Let* $(\hat{\beta}_{1,t}^*, \hat{\beta}_{2,t}^*)' = \hat{\beta}_t^* = \hat{\beta}_{t,\gamma,\lambda}^*$ *be defined as in* (20b), *with* $\gamma = 0, 1$ *and* $\lambda \in [0,1]$. *Let Assumptions 1 and 3 hold. Then*

$$\begin{pmatrix} T^{-1/2}(\hat{\beta}_{1,\lfloor Tr \rfloor}^* - \beta_1) \\ T^{1/2}(\hat{\beta}_{2,\lfloor Tr \rfloor}^* - \beta_2) \end{pmatrix} \xrightarrow{d^*} \begin{pmatrix} V_{1,\gamma}(r,\lambda) \\ V_{2,\gamma}(r,\lambda) \end{pmatrix} \text{ in probability,}$$

where $\bar{r}$ *and* $V_{i,\gamma}(r,\lambda)$ *for* $i = 1, 2$ *are defined in Lemma A.3.*

Proof of Lemma A.9. The proof follows trivially from the proof of Lemma A.3 using Lemma A.8. $\qquad\square$

Proof of Lemma 4. The result follows directly from Lemma A.8 and A.9 along the lines of the proof of Lemma A.4. $\qquad\square$

Proof of Lemma 5. Part (a) follows from Lemma 4 using the continuous mapping theorem.

For part (b), we write

$$T^{-1} \sum_{t=1}^{T} y_{t-1}^{*d} \varepsilon_{p,t}^{*d} = T^{-1} \sum_{t=1}^{T} y_{t-1}^{*d} \varepsilon_t^{*d} + T^{-1} \sum_{t=1}^{T} y_{t-1}^{*d} (\varepsilon_{p,t}^{*d} - \varepsilon_t^{*d}).$$

40

Note that

$$\varepsilon_{p,t}^{*d} - \varepsilon_t^{*d} = \sum_{j=p+1}^{q} \hat{\phi}_j u_{t-j}^* + \sum_{j=p+1}^{q} \hat{\phi}_j \Delta[(\hat{\beta}_{t-j}^* - \beta^*)' z_{t-j}].$$

By Assumption 4 there is some $\tilde{T}$ such that for all $T > \tilde{T}$ we have that $\varepsilon_{p,t}^{*d} - \varepsilon_t^{*d} = 0$. It then follows as in the proof of Lemma 1(b) that $T^{-1} \sum_{t=1}^{T} y_{t-1}^{*d} \varepsilon_t^{*d} \xrightarrow{d^*} \frac{\sigma}{2} \left[W_\gamma(1, \lambda)^2 - W_\gamma(0, \lambda)^2 - 1 \right]$ in probability.

For (c) we define Ω_{pp}^* and $B_{p,t}^*$ analogously to Ω_{pp} and $B_{p,t}$, and write

$$\left\| T^{-1} \sum_{t=1}^{T} w_{p,t}^{*d} w_{p,t}^{*d\prime} - \Omega_{pp}^* \right\| \leq \left\| T^{-1} \sum_{t=1}^{T} w_{p,t}^* w_{p,t}^{*\prime} - \Omega_{pp}^* \right\| + 2 \left\| T^{-1} \sum_{t=1}^{T} w_{p,t}^* \Delta[z_{p,t}' B_{p,t}^{*\prime}] \right\|$$

$$+ \left\| T^{-1} \sum_{t=1}^{T} \Delta[B_{p,t}^* z_{p,t}] \Delta[z_{p,t}' B_{p,t}^{*\prime}] \right\|$$

$$= A_T^{c*} + 2 B_T^{c*} + C_T^{c*}.$$

By Chang and Park (2003, Proof of Lemma 3a) and Lemma 3 $A_T^{c*} = O_p^*(T^{-1/2} p)$. We can then show in the same way as in the proof of Lemma 1(c) that $B_T^{c*}, C_T^{c*} = O_p^*(T^{-1} p)$. Therefore we have that $\left\| T^{-1} \sum_{t=1}^{T} w_{p,t}^{*(r)d} w_{p,t}^{*(r)d\prime} - \Omega_{pp}^* \right\| = O_p^*(T^{-1/2} p)$ and we can conclude the proof as in Chang and Park (2003, Proof of Lemma 3a).

Next we look at (d). As in the proof of lemma 1 we can write

$$\left| T^{-1} \sum_{t=1}^{T} y_{t-1}^{*d} w_{p,t}^{*d} \right| \leq \left| T^{-1} \sum_{t=1}^{T} x_{t-1}^* w_{p,t}^* \right| + \left| T^{-1} \sum_{t=1}^{T} x_{t-1}^* \Delta[B_{p,t}^* z_{p,t}] \right|$$

$$+ \left| T^{-1} \sum_{t=1}^{T} (\hat{\beta}_{t-1}^* - \beta^*)' z_{t-1} w_{p,t}^* \right| + \left| T^{-1} \sum_{t=1}^{T} (\hat{\beta}_{t-1}^* - \beta^*)' z_{t-1} \Delta[B_{p,t}^* z_{p,t}] \right|$$

$$= A_T^{d*} + B_T^{d*} + C_T^{d*} + D_T^{d*}.$$

We can show that $A_T^{d*} = O_p^*(p^{1/2})$ along the same lines as Chang and Park (2003, Proof of Lemma 3b) using Lemma 3. Furthermore we can show in the same way as in Lemma 1 that $B_T^{d*}, C_T^{d*}, D_T^{d*} = O_p^*(p^{1/2})$.

For part (e) we can write

$$\left| T^{-1} \sum_{t=1}^{T} w_{p,t}^{*d} \varepsilon_{p,t}^{*d} \right| \leq \left| T^{-1} \sum_{t=1}^{T} w_{p,t}^* \varepsilon_{p,t}^* \right| + \left| T^{-1} \sum_{t=1}^{T} \Delta[B_{p,t}^* z_{p,t}] \varepsilon_{p,t}^* \right|$$

$$+ \left| T^{-1} \sum_{t=1}^{T} w_{p,t}^* \hat{\phi}_p(L) \Delta[(\hat{\beta}_t^* - \beta^*)' z_t] \right|$$

$$+ \left| T^{-1} \sum_{t=1}^{T} \Delta[B_{p,t}^* z_{p,t}] \hat{\phi}_p(L) \Delta[(\hat{\beta}_t^* - \beta)' z_t] \right|$$

$$= A_T^{e*} + B_T^{e*} + C_T^{e*} + D_T^{e*}.$$

Now by Assumption 4 we have that for large T we may write

$$A_T^{e*} = \left| T^{-1} \sum_{t=1}^{T} w_{p,t}^* \varepsilon_t^* \right|.$$

Then for any $1 \le j \le p$

$$E^* \left(\sum_{t=1}^{T} u_{t-j}^* \varepsilon_t^* \right)^2 = T\sigma^{*2} \Gamma_0^*,$$

by which it follows from Chang and Park (2003, Proof of Lemma 3c) that $A_T^{e*} = O_p^*(T^{-1/2} p^{1/2})$. It then again follows from the proof of Lemma 1, together with (A.10), that $B_T^{e*}, C_T^{e*}, D_T^{e*} = O_p^*(T^{-1} p^{1/2})$. This completes the proof. $\qquad\square$

Corollary A.2. *Let Assumptions 1, 2, 3 and 4 hold. Let A_T^* and B_T^* be defined as in (28) and let $\gamma = 0, 1$ and $\lambda \in [0,1]$. Then*

1. $T^{-1} A_T^* \xrightarrow{d^*} \frac{1}{2}\psi(1)\sigma^2 [W_\gamma(1,\lambda)^2 - W_\gamma(0,\lambda)^2 - 1]$ in probability,

2. $T^{-2} B_T^* \xrightarrow{d^*} \psi(1)^2 \sigma^2 \int_0^1 W_\gamma(r,\lambda)^2 dr$ in probability.

Proof of Corollary A.2. The results follow immediately from Lemma 5, given the expressions for A_T^* and B_T^*. $\qquad\square$

Proof of Lemma 6. As in the proof of Lemma 2, we have that

$$T\hat{\sigma}^{*2} = \varepsilon_p^{*d\prime}(I - M_p^{*d}(M_p^{*d\prime} M_p^{*d})^{-1} M_p^{*d\prime})\varepsilon_p^{*d} - \varepsilon_p^{*d\prime}(I - M_p^{*d}(M_p^{*d\prime} M_p^{*d})^{-1} M_p^{*d\prime})Y_{-1}^{*d}\hat{\delta}^*$$
$$- \hat{\delta}^* Y_{-1}^{*d\prime}(I - M_p^{*d}(M_p^{*d\prime} M_p^{*d})^{-1} M_p^{*d\prime})\varepsilon_p^{*d} + \hat{\delta}^* Y_{-1}^{*d\prime}(I - M_p^{*d}(M_p^{*d\prime} M_p^{*d})^{-1} M_p^{*d\prime})Y_{-1}^{*d}\hat{\delta}^*,$$

which we can write as

$$\hat{\sigma}^{*2} = C_T^* - 2D_T^* + E_T^*.$$

Using Lemma 5 and Corollary A.2 we can show in the same way as in the proof of Lemma 2 that

$$C_T^* = T^{-1} \varepsilon_p^{*d\prime} \varepsilon_p^{*d} + o_p^*(1),$$
$$D_T^* = o_p^*(1),$$
$$E_T^* = o_p^*(1).$$

Therefore, we have that

$$\hat{\sigma}^{*2} = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{p,t}^{*d2} + o_p^*(1),$$

which, by Assumption 4, we can write as

$$\hat{\sigma}^{*2} = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^{*d2} + o_p^*(1).$$

Now

$$\left| \left(\frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^{*d2} \right)^{1/2} - \left(\frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^{*2} \right)^{1/2} \right| \leq \left[\frac{1}{T} \sum_{t=1}^{T} (\varepsilon_t^{*d} - \varepsilon_t^{*})^2 \right]^{1/2}.$$

Then

$$\begin{aligned}
\frac{1}{T} \sum_{t=1}^{T} (\varepsilon_t^{*d} - \varepsilon_t^{*})^2 &= \frac{1}{T} \sum_{t=1}^{T} \left[\Delta[(\hat{\beta}_t^* - \beta^*)'z_t] - \sum_{j=1}^{p} \hat{\phi}_j \Delta[(\hat{\beta}_{t-j}^* - \beta^*)'z_{t-j}] \right]^2 \\
&\leq \left(\sum_{j=0}^{p} \hat{\phi}_j^2 \right) \left[T^{-1} \sum_{t=1}^{T} \left(\Delta[(\hat{\beta}_{t-j}^* - \beta^*)'z_{t-j}] \right)^2 \right] \\
&= O_p^*(T^{-1}).
\end{aligned}$$

Hence,

$$\frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^{*d2} = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^{*2} + o_p^*(1).$$

Next we show that

$$\left| T^{-1} \sum_{t=1}^{T} \varepsilon_t^{*2} - \sigma^2 \right| \leq \left| T^{-1} \sum_{t=1}^{T} \varepsilon_t^{*2} - \sigma^{*2} \right| + \left| \sigma^{*2} - \sigma^2 \right| = o_p^*(1).$$

To show that $\left| \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^{*2} - \sigma^{*2} \right| = o_p^*(1)$, note that we have

$$\begin{aligned}
\mathrm{P}^* \left(\left| T^{-1} \sum_{t=1}^{T} \varepsilon_t^{*2} - \sigma^{*2} \right| > \epsilon \right) &\leq \epsilon^{-2} \, \mathrm{E}^* \left(T^{-1} \sum_{t=1}^{T} \varepsilon_t^{*2} - \sigma^{*2} \right)^2 \\
&= \epsilon^{-2} T^{-2} \sum_{t=1}^{T} \left[\mathrm{E}^*(\varepsilon_t^{*4}) - \left(\mathrm{E}^*(\varepsilon_t^{*2}) \right)^2 \right] = O_p(T^{-1}).
\end{aligned}$$

It follows from Lemma A.6 that $\sigma^{*2} \xrightarrow{p} \sigma^2$. This completes the proof. $\qquad \square$

Proof of Theorem 2. The result follows straightforwardly from Corollary A.2 and Lemma 6. $\qquad \square$

T	ϕ	θ	$ADF_{0,1}$	$ADF^{*,0,1}_{0,1}$	$ADF^{*,1,1}_{0,1}$	$ADF^{*,1,0}_{0,1}$	$ADF^{*,0,0}_{0,1}$	$ADF_{1,1}$	$ADF^{*,0,1}_{1,1}$	$ADF^{*,1,1}_{1,1}$	$ADF^{*,1,0}_{1,1}$	$ADF^{*,0,0}_{1,1}$
50	0	0	0.031	0.046	0.042	0.032	0.027	0.028	0.043	0.045	0.031	0.034
	-0.8	0	0.025	0.035	0.033	0.021	0.020	0.020	0.031	0.031	0.022	0.031
	-0.4	0	0.042	0.037	0.040	0.026	0.024	0.037	0.034	0.036	0.031	0.032
	0.4	0	0.010	0.026	0.025	0.026	0.026	0.005	0.021	0.023	0.021	0.021
	0.8	0	0.030	0.055	0.054	0.051	0.052	0.032	0.048	0.046	0.043	0.045
	0	-0.8	0.373	0.269	0.269	0.168	0.168	0.259	0.144	0.141	0.096	0.104
	0	-0.4	0.092	0.073	0.077	0.037	0.040	0.080	0.061	0.061	0.042	0.042
	0	0.4	0.008	0.023	0.027	0.020	0.018	0.003	0.015	0.014	0.012	0.013
	0	0.8	0.006	0.029	0.027	0.026	0.026	0.008	0.032	0.024	0.020	0.025
	-0.4	-0.4	0.077	0.041	0.044	0.019	0.015	0.066	0.037	0.041	0.032	0.035
	0.4	-0.4	0.035	0.045	0.047	0.031	0.032	0.029	0.043	0.045	0.033	0.032
	-0.4	0.4	0.034	0.048	0.050	0.029	0.033	0.030	0.053	0.045	0.032	0.037
	0.4	0.4	0.022	0.043	0.046	0.045	0.048	0.024	0.045	0.041	0.047	0.046
100	0	0	0.035	0.049	0.050	0.045	0.041	0.038	0.052	0.060	0.054	0.059
	-0.8	0	0.030	0.044	0.048	0.038	0.034	0.027	0.041	0.042	0.040	0.042
	-0.4	0	0.029	0.040	0.041	0.029	0.029	0.031	0.046	0.039	0.036	0.042
	0.4	0	0.022	0.041	0.045	0.043	0.036	0.024	0.045	0.050	0.041	0.044
	0.8	0	0.031	0.048	0.049	0.044	0.041	0.037	0.055	0.048	0.045	0.043
	0	-0.8	0.195	0.078	0.078	0.050	0.047	0.117	0.057	0.051	0.054	0.053
	0	-0.4	0.049	0.054	0.052	0.036	0.036	0.050	0.054	0.054	0.056	0.049
	0	0.4	0.019	0.043	0.036	0.032	0.036	0.022	0.037	0.039	0.039	0.035
	0	0.8	0.011	0.028	0.029	0.026	0.024	0.013	0.026	0.029	0.026	0.028
	-0.4	-0.4	0.045	0.048	0.050	0.035	0.039	0.041	0.043	0.044	0.043	0.053
	0.4	-0.4	0.033	0.047	0.043	0.040	0.040	0.029	0.047	0.042	0.040	0.049
	-0.4	0.4	0.033	0.052	0.051	0.035	0.038	0.029	0.042	0.044	0.045	0.045
	0.4	0.4	0.026	0.039	0.039	0.035	0.032	0.028	0.039	0.041	0.041	0.036

Table 1: Size ($c = 0$) of full sample detrended tests ($\lambda = 1$)

T	ϕ	θ	$ADF_{0,0}$	$ADF^{*,0,1}_{0,0}$	$ADF^{*,1,1}_{0,0}$	$ADF^{*,1,0}_{0,0}$	$ADF^{*,0,0}_{0,0}$	$ADF_{1,0}$	$ADF^{*,0,1}_{1,0}$	$ADF^{*,1,1}_{1,0}$	$ADF^{*,1,0}_{1,0}$	$ADF^{*,0,0}_{1,0}$
50	0	0	0.020	0.047	0.038	0.030	0.031	0.018	0.045	0.048	0.040	0.046
	-0.8	0	0.016	0.040	0.036	0.025	0.026	0.013	0.033	0.034	0.030	0.030
	-0.4	0	0.027	0.043	0.044	0.031	0.033	0.024	0.039	0.040	0.043	0.039
	0.4	0	0.003	0.021	0.024	0.018	0.020	0.006	0.027	0.026	0.028	0.030
	0.8	0	0.009	0.048	0.049	0.036	0.046	0.021	0.047	0.053	0.049	0.053
	0	-0.8	0.199	0.170	0.163	0.093	0.095	0.128	0.093	0.091	0.058	0.061
	0	-0.4	0.051	0.069	0.064	0.037	0.040	0.044	0.050	0.050	0.040	0.045
	0	0.4	0.001	0.018	0.017	0.013	0.013	0.008	0.023	0.028	0.021	0.024
	0	0.8	0.003	0.024	0.025	0.021	0.023	0.007	0.028	0.030	0.028	0.028
	-0.4	-0.4	0.043	0.046	0.045	0.023	0.022	0.033	0.042	0.047	0.044	0.036
	0.4	-0.4	0.021	0.045	0.047	0.036	0.031	0.021	0.045	0.044	0.039	0.039
	-0.4	0.4	0.021	0.044	0.044	0.034	0.029	0.022	0.041	0.042	0.036	0.042
	0.4	0.4	0.010	0.041	0.044	0.039	0.037	0.015	0.044	0.043	0.045	0.039
100	0	0	0.031	0.054	0.060	0.053	0.048	0.021	0.060	0.058	0.051	0.053
	-0.8	0	0.019	0.042	0.040	0.031	0.037	0.017	0.044	0.045	0.041	0.039
	-0.4	0	0.022	0.047	0.050	0.048	0.043	0.021	0.060	0.050	0.048	0.047
	0.4	0	0.013	0.052	0.049	0.048	0.044	0.016	0.049	0.048	0.049	0.052
	0.8	0	0.012	0.045	0.051	0.043	0.037	0.028	0.044	0.050	0.046	0.048
	0	-0.8	0.119	0.060	0.065	0.041	0.040	0.064	0.067	0.064	0.073	0.069
	0	-0.4	0.036	0.054	0.052	0.044	0.046	0.034	0.053	0.054	0.054	0.048
	0	0.4	0.013	0.037	0.039	0.034	0.038	0.019	0.045	0.043	0.042	0.042
	0	0.8	0.005	0.030	0.028	0.022	0.024	0.008	0.032	0.033	0.037	0.031
	-0.4	-0.4	0.027	0.046	0.046	0.039	0.043	0.025	0.052	0.051	0.052	0.043
	0.4	-0.4	0.023	0.046	0.047	0.038	0.049	0.022	0.049	0.048	0.051	0.051
	-0.4	0.4	0.024	0.045	0.047	0.043	0.044	0.022	0.047	0.044	0.045	0.045
	0.4	0.4	0.013	0.043	0.039	0.037	0.042	0.016	0.043	0.047	0.043	0.044

Table 2: Size ($c = 0$) of recursively detrended tests ($\lambda = 0$)

T	ϕ	θ	$ADF_{0,1}$	$ADF^{*,0,1}_{0,1}$	$ADF^{*,1,1}_{0,1}$	$ADF^{*,1,0}_{0,1}$	$ADF^{*,0,0}_{0,1}$	$ADF_{1,1}$	$ADF^{*,0,1}_{1,1}$	$ADF^{*,1,1}_{1,1}$	$ADF^{*,1,0}_{1,1}$	$ADF^{*,0,0}_{1,1}$
50	0	0	0.191	0.157	0.171	0.118	0.143	0.259	0.215	0.219	0.209	0.191
	-0.8	0	0.196	0.155	0.170	0.111	0.113	0.160	0.111	0.121	0.094	0.079
	-0.4	0	0.167	0.137	0.143	0.126	0.124	0.171	0.152	0.150	0.106	0.104
	0.4	0	0.049	0.040	0.035	0.030	0.035	0.109	0.087	0.075	0.070	0.078
	0.8	0	0.069	0.076	0.080	0.079	0.069	0.096	0.109	0.113	0.112	0.099
	0	-0.8	0.154	0.276	0.206	0.167	0.182	0.135	0.115	0.113	0.090	0.106
	0	-0.4	0.177	0.169	0.158	0.152	0.124	0.178	0.135	0.149	0.123	0.114
	0	0.4	0.097	0.083	0.064	0.069	0.072	0.188	0.149	0.140	0.129	0.129
	0	0.8	0.075	0.060	0.072	0.062	0.050	0.125	0.099	0.119	0.111	0.097
	-0.4	-0.4	0.162	0.119	0.114	0.078	0.090	0.129	0.081	0.066	0.047	0.052
	0.4	-0.4	0.188	0.150	0.153	0.144	0.145	0.252	0.210	0.232	0.155	0.184
	-0.4	0.4	0.166	0.164	0.153	0.128	0.125	0.233	0.159	0.201	0.144	0.167
	0.4	0.4	0.068	0.071	0.074	0.066	0.060	0.104	0.104	0.110	0.098	0.090
100	0	0	0.162	0.162	0.143	0.110	0.143	0.231	0.215	0.192	0.174	0.171
	-0.8	0	0.166	0.144	0.136	0.125	0.135	0.183	0.145	0.139	0.124	0.129
	-0.4	0	0.165	0.146	0.133	0.113	0.126	0.235	0.194	0.223	0.203	0.198
	0.4	0	0.112	0.120	0.103	0.093	0.103	0.178	0.194	0.158	0.163	0.151
	0.8	0	0.106	0.093	0.094	0.098	0.100	0.150	0.137	0.146	0.154	0.133
	0	-0.8	0.161	0.169	0.175	0.146	0.146	0.122	0.077	0.091	0.078	0.086
	0	-0.4	0.174	0.148	0.139	0.139	0.135	0.198	0.176	0.155	0.133	0.163
	0	0.4	0.106	0.097	0.098	0.096	0.089	0.228	0.222	0.223	0.200	0.199
	0	0.8	0.119	0.108	0.118	0.093	0.104	0.193	0.181	0.161	0.149	0.163
	-0.4	-0.4	0.162	0.110	0.119	0.107	0.101	0.161	0.139	0.135	0.113	0.100
	0.4	-0.4	0.164	0.143	0.137	0.142	0.147	0.267	0.239	0.258	0.192	0.184
	-0.4	0.4	0.165	0.128	0.126	0.145	0.158	0.282	0.237	0.253	0.233	0.241
	0.4	0.4	0.130	0.127	0.114	0.117	0.120	0.192	0.174	0.185	0.151	0.202

Table 3: Size-adjusted power ($c = 10$) of full sample detrended tests ($\lambda = 1$) with small initial condition.

T	ϕ	θ	$ADF_{0,0}$	$ADF_{0,0}^{*,0,1}$	$ADF_{0,0}^{*,1,1}$	$ADF_{0,0}^{*,1,0}$	$ADF_{0,0}^{*,0,0}$	$ADF_{1,0}$	$ADF_{1,0}^{*,0,1}$	$ADF_{1,0}^{*,1,1}$	$ADF_{1,0}^{*,1,0}$	$ADF_{1,0}^{*,0,0}$
50	0	0	0.235	0.202	0.225	0.197	0.183	0.248	0.209	0.206	0.170	0.172
	-0.8	0	0.192	0.152	0.162	0.132	0.130	0.134	0.111	0.112	0.106	0.119
	-0.4	0	0.187	0.154	0.154	0.132	0.120	0.182	0.134	0.140	0.114	0.124
	0.4	0	0.146	0.088	0.111	0.091	0.091	0.129	0.125	0.122	0.096	0.103
	0.8	0	0.114	0.112	0.105	0.104	0.098	0.088	0.111	0.102	0.093	0.094
	0	-0.8	0.151	0.190	0.197	0.166	0.176	0.109	0.111	0.106	0.111	0.104
	0	-0.4	0.180	0.157	0.150	0.127	0.118	0.143	0.148	0.146	0.122	0.100
	0	0.4	0.193	0.167	0.161	0.157	0.153	0.172	0.189	0.149	0.168	0.172
	0	0.8	0.154	0.130	0.134	0.127	0.116	0.143	0.128	0.124	0.119	0.123
	-0.4	-0.4	0.141	0.117	0.112	0.076	0.091	0.117	0.081	0.072	0.057	0.071
	0.4	-0.4	0.230	0.210	0.224	0.165	0.195	0.213	0.242	0.219	0.182	0.206
	-0.4	0.4	0.241	0.231	0.220	0.194	0.195	0.225	0.221	0.223	0.183	0.177
	0.4	0.4	0.132	0.112	0.114	0.115	0.127	0.122	0.125	0.123	0.107	0.129
100	0	0	0.205	0.189	0.188	0.162	0.166	0.240	0.181	0.198	0.199	0.207
	-0.8	0	0.229	0.200	0.196	0.190	0.166	0.162	0.157	0.141	0.168	0.169
	-0.4	0	0.205	0.176	0.184	0.147	0.171	0.216	0.158	0.184	0.179	0.176
	0.4	0	0.167	0.139	0.142	0.134	0.141	0.214	0.182	0.192	0.181	0.167
	0.8	0	0.159	0.134	0.131	0.135	0.144	0.150	0.170	0.137	0.155	0.148
	0	-0.8	0.155	0.155	0.155	0.120	0.116	0.121	0.085	0.091	0.068	0.073
	0	-0.4	0.194	0.169	0.166	0.156	0.162	0.192	0.163	0.174	0.147	0.171
	0	0.4	0.204	0.183	0.175	0.199	0.160	0.203	0.204	0.189	0.194	0.170
	0	0.8	0.183	0.153	0.167	0.151	0.150	0.159	0.178	0.175	0.173	0.154
	-0.4	-0.4	0.185	0.148	0.155	0.126	0.120	0.148	0.135	0.127	0.139	0.145
	0.4	-0.4	0.235	0.219	0.233	0.199	0.185	0.231	0.233	0.238	0.217	0.213
	-0.4	0.4	0.244	0.241	0.244	0.193	0.185	0.253	0.242	0.258	0.225	0.240
	0.4	0.4	0.172	0.175	0.157	0.172	0.133	0.214	0.181	0.178	0.145	0.178

Table 4: Size-adjusted power ($c = 10$) of recursively detrended tests ($\lambda = 0$) with small initial condition.

T	ϕ	θ	$ADF_{0,1}$	$ADF^{*,0,1}_{0,1}$	$ADF^{*,1,1}_{0,1}$	$ADF^{*,1,0}_{0,1}$	$ADF^{*,0,0}_{0,1}$	$ADF_{1,1}$	$ADF^{*,0,1}_{1,1}$	$ADF^{*,1,1}_{1,1}$	$ADF^{*,1,0}_{1,1}$	$ADF^{*,0,0}_{1,1}$
50	0	0	0.211	0.191	0.190	0.123	0.142	0.130	0.096	0.089	0.084	0.079
	-0.8	0	0.185	0.136	0.177	0.105	0.094	0.111	0.078	0.076	0.071	0.059
	-0.4	0	0.161	0.144	0.128	0.086	0.099	0.104	0.084	0.073	0.059	0.053
	0.4	0	0.143	0.116	0.118	0.105	0.109	0.060	0.049	0.047	0.043	0.043
	0.8	0	0.089	0.105	0.104	0.096	0.086	0.032	0.041	0.040	0.043	0.043
	0	-0.8	0.144	0.177	0.088	0.110	0.160	0.113	0.072	0.103	0.071	0.058
	0	-0.4	0.146	0.144	0.131	0.100	0.090	0.105	0.070	0.075	0.059	0.065
	0	0.4	0.168	0.146	0.139	0.132	0.128	0.086	0.069	0.067	0.052	0.057
	0	0.8	0.126	0.107	0.117	0.110	0.093	0.054	0.046	0.047	0.052	0.043
	-0.4	-0.4	0.139	0.098	0.095	0.061	0.065	0.091	0.059	0.057	0.052	0.041
	0.4	-0.4	0.204	0.193	0.180	0.145	0.135	0.129	0.110	0.097	0.071	0.072
	-0.4	0.4	0.181	0.192	0.155	0.128	0.124	0.124	0.086	0.097	0.057	0.063
	0.4	0.4	0.096	0.104	0.102	0.096	0.090	0.047	0.048	0.058	0.048	0.046
100	0	0	0.186	0.183	0.159	0.125	0.142	0.065	0.068	0.048	0.047	0.049
	-0.8	0	0.196	0.165	0.156	0.135	0.141	0.106	0.085	0.094	0.082	0.085
	-0.4	0	0.194	0.172	0.169	0.159	0.148	0.083	0.064	0.078	0.070	0.063
	0.4	0	0.168	0.155	0.155	0.150	0.168	0.054	0.057	0.049	0.054	0.051
	0.8	0	0.114	0.119	0.112	0.108	0.093	0.033	0.026	0.033	0.030	0.033
	0	-0.8	0.147	0.157	0.157	0.123	0.124	0.082	0.061	0.064	0.056	0.065
	0	-0.4	0.168	0.145	0.139	0.112	0.109	0.075	0.067	0.061	0.051	0.056
	0	0.4	0.161	0.141	0.177	0.148	0.143	0.062	0.055	0.056	0.044	0.056
	0	0.8	0.166	0.168	0.154	0.137	0.146	0.047	0.039	0.036	0.034	0.035
	-0.4	-0.4	0.163	0.133	0.118	0.096	0.098	0.091	0.087	0.088	0.090	0.061
	0.4	-0.4	0.216	0.203	0.187	0.159	0.167	0.076	0.058	0.069	0.050	0.044
	-0.4	0.4	0.201	0.162	0.187	0.176	0.163	0.091	0.076	0.068	0.059	0.075
	0.4	0.4	0.152	0.136	0.158	0.126	0.132	0.049	0.046	0.049	0.036	0.049

Table 5: Size-adjusted power ($c = 10$) of full sample detrended tests ($\lambda = 1$) with large initial condition.

T	ϕ	θ	$ADF_{0,0}$	$ADF^{*,0,1}_{0,0}$	$ADF^{*,1,1}_{0,0}$	$ADF^{*,1,0}_{0,0}$	$ADF^{*,0,0}_{0,0}$	$ADF_{1,0}$	$ADF^{*,0,1}_{1,0}$	$ADF^{*,1,1}_{1,0}$	$ADF^{*,1,0}_{1,0}$	$ADF^{*,0,0}_{1,0}$
50	0	0	0.134	0.109	0.138	0.096	0.078	0.103	0.064	0.060	0.053	0.048
	-0.8	0	0.119	0.099	0.108	0.068	0.071	0.088	0.069	0.073	0.071	0.060
	-0.4	0	0.115	0.099	0.088	0.075	0.060	0.080	0.066	0.082	0.052	0.051
	0.4	0	0.087	0.071	0.050	0.056	0.050	0.042	0.039	0.044	0.031	0.032
	0.8	0	0.041	0.036	0.032	0.043	0.033	0.013	0.017	0.013	0.016	0.016
	0	-0.8	0.129	0.141	0.174	0.126	0.133	0.087	0.089	0.089	0.095	0.080
	0	-0.4	0.109	0.101	0.084	0.070	0.053	0.072	0.068	0.069	0.058	0.067
	0	0.4	0.101	0.084	0.083	0.076	0.067	0.053	0.050	0.044	0.058	0.060
	0	0.8	0.080	0.073	0.070	0.053	0.057	0.041	0.038	0.035	0.034	0.034
	-0.4	-0.4	0.098	0.067	0.068	0.049	0.051	0.077	0.046	0.039	0.034	0.046
	0.4	-0.4	0.133	0.130	0.123	0.079	0.095	0.091	0.079	0.076	0.058	0.063
	-0.4	0.4	0.144	0.144	0.117	0.096	0.097	0.077	0.085	0.089	0.074	0.059
	0.4	0.4	0.067	0.062	0.052	0.058	0.062	0.035	0.035	0.040	0.034	0.038
100	0	0	0.109	0.097	0.101	0.081	0.084	0.065	0.039	0.051	0.047	0.054
	-0.8	0	0.142	0.130	0.119	0.117	0.100	0.081	0.069	0.077	0.081	0.092
	-0.4	0	0.120	0.116	0.115	0.076	0.097	0.071	0.044	0.052	0.046	0.047
	0.4	0	0.095	0.078	0.080	0.089	0.084	0.051	0.042	0.044	0.048	0.042
	0.8	0	0.068	0.056	0.057	0.053	0.062	0.023	0.031	0.021	0.026	0.024
	0	-0.8	0.117	0.115	0.117	0.083	0.085	0.082	0.064	0.072	0.048	0.055
	0	-0.4	0.106	0.102	0.107	0.086	0.080	0.071	0.066	0.066	0.056	0.066
	0	0.4	0.105	0.102	0.086	0.094	0.092	0.042	0.046	0.046	0.052	0.044
	0	0.8	0.087	0.070	0.083	0.075	0.068	0.034	0.034	0.034	0.032	0.036
	-0.4	-0.4	0.112	0.098	0.097	0.078	0.066	0.080	0.056	0.070	0.068	0.077
	0.4	-0.4	0.130	0.129	0.124	0.115	0.090	0.053	0.053	0.048	0.046	0.045
	-0.4	0.4	0.144	0.150	0.135	0.123	0.107	0.064	0.051	0.056	0.061	0.066
	0.4	0.4	0.086	0.085	0.079	0.084	0.066	0.040	0.037	0.036	0.036	0.040

Table 6: Size-adjusted power ($c = 10$) of recursively detrended tests ($\lambda = 0$) with large initial condition.