A Sieve Bootstrap Test for Cointegration in a Conditional Error Correction Model

Franz C. Palm* Stephan Smeekes Jean-Pierre Urbain

Department of Quantitative Economics
Maastricht University
December 16, 2008

Abstract

In this paper we propose a bootstrap version of the Wald test for cointegration in a single-equation conditional error correction model. The multivariate sieve bootstrap is used to deal with dependence in the series. We show that the introduced bootstrap test is asymptotically valid.

We also analyze the small sample properties of our test by simulation and compare it with the asymptotic test and several alternative bootstrap tests. The bootstrap test offers significant improvements in terms of size properties over the asymptotic test, while having similar power properties.

The sensitivity of the bootstrap test to the allowance for deterministic components is also investigated. Simulation results show that the tests with sufficient deterministic components included are insensitive to the true value of the trends in the model, and retain correct size.

JEL Classification: C15, C32.

Keywords: sieve bootstrap, cointegration, error correction model.

*Corresponding author: Department of Quantitative Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands. Email: F.Palm@ke.unimaas.nl. Previous versions of this paper have been presented at the Econometric Society European Meeting in Milan, August 2008, at the International Workshop on Recent Advances in Time Series Analysis in Cyprus, June 2008, at the Workshop on Bootstrap and Time Series in Kaiserslautern, June 2008, at the conference entitled Inference and Tests in Econometrics, A Tribute to Russell Davidson in Marseille, April 2008, at an Ente Luigi Einaudi Seminar in Econometrics in Rome, November 2007, and at the first workshop of the Methods in International Finance Network in Maastricht, September 2007. We gratefully acknowledge the comments by participants at these seminars and the comments by Anders Swensen, Pentti Saikkonen and three anonymous referees. The usual disclaimer applies.

1 Introduction

In this paper we present a bootstrap version of the single-equation error correction model (ECM) Wald test for cointegration originally proposed by Boswijk (1994).

Broadly speaking, tests proposed in the literature to test for the absence of cointegration can be classified in two groups. Tests that allow for more than one cointegrating vector under the alternative using for example a VAR framework, see e.g. Johansen (1995), and tests that consider single-equation models and assuming at most a single cointegrating vector under the alternative. Among the latter ones, we can further distinguish between approaches based on the triangular representation of a cointegration system that naturally leads to residual-based tests for cointegration (e.g. Phillips and Ouliaris, 1990) that make use of semi-parametric correction for endogeneity and serial correlation; and those based on fully specified parametric data generating processes that naturally lead to single equation dynamic models. The ECM test considered in this paper falls in this category. As already discussed in the literature, ECM tests are an attractive option for cointegration testing, as, contrary to the more popular residual-based tests, ECM tests do not suffer from imposing potentially invalid common factor restrictions (Kremers, Ericsson, and Dolado, 1992; Banerjee, Dolado, and Mestre, 1998; Zivot, 2000). Moreover, Pesavento (2004) analyzes several tests which have as null hypothesis no cointegration, including the residual ADF test by Engle and Granger (1987) and the maximum eigenvalue test by Johansen and Juselius (1990), and finds that among these the ECM tests perform best in terms of power both in small and large samples, while performing similarly as the other tests in terms of size. ECM tests thus appear to be an appealing tool of testing for cointegration.

The ECM Wald test has as main advantage over the ECM t-test (Banerjee et al., 1998) that it is more intuitive and one does not have to add a redundant regressor if no particular cointegrating vector is specified. Although the Wald ECM test performs well in general, especially in terms of power, it still suffers from size distortions in finite samples (see for example Boswijk and Franses, 1992). It is well known that the bootstrap's ability to provide asymptotic refinements often leads to a reduction of size distortions for hypothesis tests. Even under "non-favorable" conditions for the bootstrap, under which it is unclear whether it provides asymptotic refinements, such as when dealing with nonstationary time series, the bootstrap has been shown to reduce size distortions in finite samples (see for example the tests for unit roots considered in Chang and Park, 2003, Palm, Smeekes, and Urbain, 2008 or Paparoditis and Politis, 2003).[1]

Little is known so far about the application of the bootstrap to cointegration testing in error correction models. Swensen (2006) and Trenkler (2006) provide theoretical and

[1]The notable exception to the lack of theoretical results is Park (2003), who shows that bootstrap ADF tests offer asymptotic refinements under the assumption that the errors are a finite AR process with known order.

simulation results on bootstrap versions of the trace test for cointegration rank by Johansen (1995). Their setting differs from ours in that we a priori assume that the cointegrating rank is at most one. Seo (2006) provides analytical and simulation results for a residual-based bootstrap test in a threshold vector error correction model. Closer to our setting, Mantalos and Shukur (1998) and Ahlgren (2000) consider a bootstrap version of the test with known cointegrating vector by Kremers et al. (1992), however they only provide simulation results for a simple model. In this paper we will allow for more general dependence over time in our model, and we provide analytical as well as simulation results.

Our paper relies on the sieve bootstrap introduced by Bühlmann (1997), a method that can handle time series dependence in the form of a general linear process that is approximated by an autoregressive process. The sieve bootstrap method is easy to use and performs well relative to other time series bootstrap methods, especially the block bootstrap (for a comparison between methods in the unit root setting, see Palm et al., 2008). The condition of linearity is fulfilled by a large class of processes, and is needed to validate the use of the Wald test without the bootstrap as well.

The contribution of the paper is threefold. First, we prove that the sieve bootstrap version of the single-equation Wald test of no cointegration is asymptotically valid. The proofs are given in detail for the multivariate setting, such that proofs of other types of tests could be done along the same lines as presented here. Second, we provide simulation results showing that the bootstrap version of the Wald test has better properties in finite samples than the asymptotic test. Third, we investigate the sensitivity of the bootstrap to various specifications of deterministic components and alternative distributional assumptions.

The structure of the paper is as follows. Section 2 explains the model and assumptions. The construction of the bootstrap test and the establishment of its asymptotic validity are discussed in Section 3. Our simulation study is presented in Section 4. The inclusion of deterministic components is discussed in Section 5. Section 6 concludes. All proofs are contained in Appendix A.

Finally, a word on notation. We use $|\cdot|$ to denote the Euclidean norm for vectors and matrices, i.e. $|v| = (v'v)^{1/2}$ for a vector v and $|M| = (\operatorname{tr} M'M)^{1/2}$ for a matrix M. For matrices we also use the operator norm $||M|| = \max_v |Mv|/|v|$. $W(r) = (W_1(r), W_2(r)')'$ denotes a multivariate standard Brownian motion of dimension $(1 + l)$. $[x]$ is the largest integer smaller than or equal to x. Convergence in distribution (probability) is denoted by $\xrightarrow{d}$ ($\xrightarrow{p}$). Bootstrap quantities (conditional on the original sample) are indicated by appending a superscript $*$ to the standard notation. Subscripts p (or q) are used to indicate quantities depending on approximations of infinite order models by models of order p (or q). For simplicity we suppress these subscripts whenever clarity allows it.

2 The model

Our Data Generating Process (DGP) is closely related to that of Pesavento (2004). We let our $(1+l)$-dimensional time series $z_t = (y_t, x_t')'$ be described by the process

$$z_t = \mu + \tau t + \zeta_t. \tag{1}$$

The stochastic component ζ_t is given by

$$\Delta\zeta_t = (\rho - 1)\alpha\beta'\zeta_{t-1} + u_t, \tag{2}$$

where

$$u_t = \Psi(L)\varepsilon_t \tag{3}$$

with $\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j$. Furthermore we assume that $\zeta_0 = 0$.[2] The null hypothesis is $H_0 : \rho = 1$, there is no cointegration. Under the alternative $H_1 : \rho < 1$ there is cointegration with a single cointegrating vector β and the error correction term must be present in the equation for y_t. Also, we impose $\alpha_1 = 1$ and $\alpha_2 = 0$, which follows from the triangular representation of the model as in Pesavento (2004) and is needed for identification purposes.[3] These points are formalized in Assumption 1.

Assumption 1. We assume

(i) $\alpha\beta'$ is of rank 1, i.e. there is a single $(1+l)$-dimensional cointegrating vector β,

(ii) β is normalized on the coefficient of y_t, i.e. $\beta = (1, -\gamma')'$,

(iii) $\alpha = (1, 0')'$.

It is important to remark that Assumption 1 is of no importance for the derivation of the null distribution of the tests as it only concerns the situation where cointegration is present in the system. It is however important to derive the equivalence between the triangular representation and the ECM form. Assumption 1 is also important to enable us to focus on a single equation ECM and to rule out cases where the ECM tests would trivially have low power. This would for example occur under the alternative if the cointegration vector only appears in the equation for the conditioning variables x_t.

Equation (3) shows that we take u_t to be a linear process (Phillips and Solo, 1992). Assumption 2 ensures the invertibility of u_t and the existence of moments of ε_t. These assumptions are not very stringent and encompass many assumptions (including all finite VARMA models) that are often used in cointegration analysis.

[2] This assumption is made for expositional simplicity only and can be extended to $\zeta_0 = O_p(1)$.

[3] Pesavento (2004) shows that this restriction corresponds to the assumption that x_t are not mutually cointegrated, as required under Assumption 1(i), and are known a priori to be $I(1)$.

Assumption 2. We assume

(i) ε_t are i.i.d. with $\mathrm{E}(\varepsilon_t)=0$, $\mathrm{E}(\varepsilon_t\varepsilon_t')=\Sigma$ and $\mathrm{E}\,|\varepsilon_t|^4<\infty$.

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

By Assumption 2 we may write $\Phi(L)=\sum_{j=0}^{\infty}\Phi_j L^j=\Psi(L)^{-1}$. We may then substitute equation (1) into (2) and apply the Beveridge-Nelson decomposition to show as in Pesavento (2004) that this model can be rewritten in VECM form

$$\Delta z_t=(\rho-1)\Phi(1)\alpha\beta'(z_{t-1}-\mu-\tau(t-1))+\tilde{\tau}+\Phi^*(L)\Delta z_{t-1}+\varepsilon_t \tag{4}$$

where,

$$\Phi^*(L)=\sum_{j=0}^{\infty}\left((1-\rho)\left(\sum_{i=j+1}^{\infty}\Phi_i\right)\alpha\beta'-\Phi_{j+1}\right)L^j$$

and

$$\tilde{\tau}=\left(\sum_{j=0}^{\infty}\Phi_j+(\rho-1)\left(\sum_{j=0}^{\infty}\sum_{i=j+1}^{\infty}\Phi_i\right)\alpha\beta'\right)\tau.$$

It can be seen from the above representation that z_t has a drift if $\tau\neq 0$, and this drift leads to a linear trend in the cointegrating relation if $\beta'\tau\neq 0$. The constant μ only appears in the cointegrating relation; note that the cointegrating relation has mean zero if $\beta'\mu=0$.

Pesavento shows that the model can be written in triangular form as well, which makes it a very flexible model. As we do not need that representation here, we continue with the VECM representation (4) and condition on x_t to obtain

$$\Delta y_t=(\rho-1)\theta\beta'(z_{t-1}-\mu-\tau(t-1))+\tilde{\tau}_1+\pi_0'\Delta x_t+\sum_{j=1}^{\infty}\pi_j'\Delta z_{t-j}+\xi_t, \tag{5}$$

where $\xi_t=\varepsilon_{1,t}-\Sigma_{12}\Sigma_{22}^{-1}\varepsilon_{2,t}\sim$ i.i.d. $(0,\omega^2)$ and $\theta=\Phi_1(1)\alpha-\Sigma_{12}\Sigma_{22}^{-1}\Phi_2(1)\alpha$ with $\Phi(1)=(\Phi_1(1)',\Phi_2(1)')'$.[4]

The advantage of this framework is that its assumptions are weaker than what is usually assumed for tests based on a conditional ECM, as it does not impose that x_t are weakly exogenous for β under the alternative of cointegration. Under the null however, the error correction term does not appear in the marginal equations, which makes a test on the error correction term in the conditional model a valid test for cointegration (Boswijk, 1994).

[4]Note that $\omega^2=\sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. Σ and ε_t have been partitioned conformably with y_t and x_t, i.e. $\varepsilon_{1,t}$ is a scalar and $\varepsilon_{2,t}$ is an l-dimensional vector.

3 The bootstrap test and asymptotics

3.1 Test statistic

The Wald test proposed by Boswijk (1994) is based on the conditional model (5). Consider the regression

$$\Delta y_t = \delta' \tilde{z}_{t-1} + \lambda' D_t + \pi_0' \Delta x_t + \sum_{j=1}^{p} \pi_j' \Delta z_{t-j} + \xi_{p,t}, \tag{6}$$

where D_t are the (unrestricted) deterministic components included in the regression, $\tilde{z}_{t-1} = (z_{t-1}', D_{t-1}^{r})'$ where D_t^r are the deterministic components that are restricted to be equal to zero under the null (see Section 5) and $\xi_{p,t} = \sum_{j=p+1}^{\infty} \pi_j' \Delta z_{t-j} + \xi_t$. If $\rho = 1$, $\delta' = (\rho - 1)\theta\beta' = 0$, which leads to the test statistic

$$T_{\text{wald}} = \hat{\delta}' \widehat{\text{Var}(\hat{\delta})}^{-1} \hat{\delta}, \tag{7}$$

where $\hat{\delta}$ is the OLS estimator of δ in (6) and $\widehat{\text{Var}(\hat{\delta})}$ is its estimated covariance matrix. The null hypothesis of no cointegration is then rejected for large values of T_{wald}.

We let the lag length p in regression (6) grow to infinity at a controlled rate.

Assumption 3. Let $p \to \infty$ and $p = o(n^{1/2})$ as $n \to \infty$.

The limiting distribution of T_{wald} can be found in Boswijk (1994) for the ECM with finite autoregressive dependence and in Pesavento (2004) for the infinite-order model. The asymptotic distribution of the test without the inclusion of any deterministic components (and with $\mu = \tau = 0$) is given for completeness in Lemma 1 without proof.

Lemma 1. *Under Assumptions 2 and 3 we have that*

$$T_{\text{wald}} \xrightarrow{d} \int_0^1 dW_1(r) W(r)' \left[\int_0^1 W(r) W(r)' dr \right]^{-1} \int_0^1 W(r) dW_1(r)$$

where T_{wald} *is defined in equation* (7).

3.2 Bootstrap method

The multivariate sieve bootstrap method we employ here is similar to the one employed by Chang, Park, and Song (2006). It is important to note that they study bootstrap inference on the cointegrating regressions and they do not consider bootstrap tests for no cointegration. The full algorithm is given below.

Bootstrap Algorithm.

Step 1: Fit a VAR(q) process to Δz_t by OLS and save the residuals

$$\hat{\varepsilon}_{q,t} = \Delta z_t - \hat{\lambda}_s D_t^s - \sum_{j=1}^{q} \hat{\Phi}_j \Delta z_{t-j}, \tag{8}$$

where D_t^s are the deterministic components included in this sieve estimation (see Section 5 for details). Recenter the residuals $\hat{\varepsilon}_{q,t}$ in the case where no constant is included to eliminate any drifts in the resampled series and save the recentered residuals $\tilde{\varepsilon}_{q,t} = \hat{\varepsilon}_{q,t} - (n-q-1)^{-1} \sum_t \hat{\varepsilon}_{q,t}$.[5]

Step 2: Resample with replacement from $\tilde{\varepsilon}_{q,t}$ to obtain bootstrap errors ε_t^*.

Step 3: Build u_t^* recursively as

$$u_t^* = \sum_{j=1}^{q} \hat{\Phi}_j u_{t-j}^* + \varepsilon_t^*, \tag{9}$$

using the estimated parameters $\hat{\Phi}_j$ from Step 1, and build z_t^* as

$$z_t^* = z_{t-1}^* + u_t^*. \tag{10}$$

Note that it is unnecessary to include deterministic components in this step, as the tests we consider are asymptotically similar (see Remark 8 in Section 5).

Step 4: Using the bootstrap sample z_t^*, obtain $\hat{\delta}^*$ from the regression

$$\Delta y_t^* = \delta^{*\prime} \tilde{z}_{t-1}^* + \lambda^{*\prime} D_t^* + \pi_0^{*\prime} \Delta x_t^* + \sum_{j=1}^{p^*} \pi_j^{*\prime} \Delta z_{t-j}^* + \xi_{p^*,t}^*, \tag{11}$$

where p^* is the lag length selected in the bootstrap regression (see Remark 6) and $\tilde{z}_{t-1}^* = (z_{t-1}^{*\prime}, D_{t-1}^{r*})'$, and calculate the bootstrap test statistic

$$T_{\text{wald}}^* = \hat{\delta}^{*\prime} \widehat{\text{Var}^*(\hat{\delta}^*)}^{-1} \hat{\delta}^*. \tag{12}$$

D_t^* and D_t^{r*} are the bootstrap counterparts of D_t and D_t^r. In order to get the correct asymptotic bootstrap distribution, one should always take $D_t^* = D_t$ and $D_t^{r*} = D_t^r$.

Step 5: Repeat Steps 2 to 4 B times, obtaining bootstrap test statistics $T_{\text{wald}}^{*b}, b = 1, \ldots, B$, and select the bootstrap critical value c_α^* as $c_\alpha^* = \min\{c : \sum_{b=1}^{B} I(T_{\text{wald}}^{*b} > c) \leq \alpha\}$, or equivalently as the $(1-\alpha)$-quantile of the ordered T_{wald}^{*b} statistics. Reject the null

[5] In the cases where we do not include a constant in this regression the residuals may have a sample mean unequal to zero, even though their theoretical mean is zero. As the sample mean of the residuals becomes the population mean of the bootstrap errors, this may lead to (unwanted) drifts in the bootstrap sample.

of no cointegration if T_{wald}, calculated from equations (6) and (7), is larger than c^*_α, where α is the nominal level of the test.

We need to allow the lag length q in the sieve bootstrap to go to infinity at a controlled rate. We will use two assumptions.

Assumption 4. Let $q \to \infty$ and $q = o((n/\ln n)^{1/2})$ as $n \to \infty$.

Assumption 4′. Let $q \to \infty$ and $q = o((n/\ln n)^{1/3})$ as $n \to \infty$.

We also need an assumption on the relative speed of the lag lengths p and q.

Assumption 5. Let $p/q \to \kappa > 1$ as $n \to \infty$, where κ may be infinite.

Note that by allowing κ to be infinite, we do not impose the same rate on p and q. Assumption 5 imposes a lower bound but not an upper bound on the rate of p (or equivalently an upper bound but not a lower bound on the rate of q).

Remark 1. In Step 3 we need to initialize u^*_t in (9) and z^*_t in (10). We propose to generate a large number of values of u^*_t and delete the first generated values. This will ensure that u^*_t is a stationary process. The initial values in (9) will then become unimportant as the realization of u^*_t will not depend on them; hence they may be set equal to zero. An alternative is to take the first q values of u^*_t equal to the first q values of u_t; this however does not ensure stationarity of u^*_t.

As asymptotically the effect of z^*_0 disappears, we simply set $z^*_0 = 0$. The logical alternative here would be to set $z^*_0 = z_0$, especially in applications.

Remark 2. Instead of estimating the sieve under the null of no cointegration (which we impose by fitting the VAR model to Δz_t in Step 1), we may also estimate it under the alternative of cointegration. In this case we would estimate the residuals as

$$\hat{\varepsilon}_{q,t} = \Delta z_t - \hat{\lambda}_b D^{s,a}_t - \hat{\Phi}_0 z_{t-1} - \sum_{j=1}^{q} \hat{\Phi}_j \Delta z_{t-j}, \tag{13}$$

where $\hat{\Phi}_0$ denotes the unrestricted OLS estimator and $D^{s,a}_t$ are the deterministic components included in this alternative-based sieve estimation. Note that even for the same deterministic setting, $D^{s,a}_t$ is not necessarily the same as D^s_t in (8), as is explained in Section 5 (Remark 10).

In the context of unit root testing, Paparoditis and Politis (2005) advocate the use of such a "residual-based" estimation as opposed to the "difference-based" estimation in (8), claiming that the residual-based tests have better power properties. We will return to this point in our simulations in Section 4.

Remark 3. A second alternative bootstrap strategy would be to base the sieve bootstrap on the conditional/marginal ECM model instead of the VECM/VAR model. In this case we would need two separate equations to estimate residuals in Step 1. We would estimate the residuals from the conditional model as

$$\hat{\varepsilon}_{1,q,t} = \Delta y_t - \hat{\lambda}_{s,1} D_{t,1}^s - \hat{\pi}_0' \Delta x_t - \sum_{j=1}^{q} \hat{\pi}_j \Delta z_{t-j}$$

and the residuals from the marginal model as

$$\hat{\varepsilon}_{2,q,t} = \Delta x_t - \hat{\lambda}_{s,2} D_{t,2}^s - \sum_{j=1}^{q} \hat{\Phi}_{2,j} \Delta z_{t-j}$$

for the difference-based alternative. We can of course also construct a residual-based version of this test. In the simulations in Section 4 we will look at these alternatives as well.

Although such an approach is closer in spirit to the single-equation Wald test statistic, it is basically just a reparametrization of the VECM approach, as the model on which the bootstrap is based is still completely specified. An alternative approach, which would be "truly conditional" on x_t, is to take x_t as fixed and only resample y_t. To justify such an approach we would have to assume strong exogeneity, see Van Giersbergen and Kiviet (1996) for a discussion. This last approach will not be investigated in this paper.

Remark 4. Although estimation under the alternative is an option in Step 1, it is not possible to build the bootstrap sample z_t^* in Step 3 based on the alternative hypothesis, i.e. using

$$z_t^* = (I + \hat{\Phi}_0) z_{t-1}^* + u_t^*. \tag{14}$$

Basawa, Mallik, McCormick, Reeves, and Taylor (1991) show that if such an alternative-based recursion is used in the unit root setting, the limiting distribution of the bootstrap test statistic is random due to the discontinuity of the limiting distribution at the unit root. The same logic applies here, therefore the null hypothesis of no cointegration must be imposed in Step 3.

Remark 5. To obtain the theoretical results in the next subsection, we set all deterministic components equal to zero, both in the model (μ and τ) and in the test (all variants of D_t). In Section 5 we will go into more detail about the inclusion of deterministic components, and present some simulation results. We conjecture that asymptotic validity still holds in the presence of deterministic components.

Remark 6. In Step 4 we specify the lag length in the bootstrap test regression (11) as p^*, in order to emphasize that this lag length does not have to be the same as the lag length in the original test regression (6). In finite samples the performance of the bootstrap test will be better if the lag length is allowed to be different. Just as for the original test regression (and the sieve bootstrap), the lag length can be chosen in practice using information criteria like AIC and BIC.

Obviously, p^* has to fulfil the same conditions as p. Therefore, we can write p^* as p in the theoretical results, which is done for notational simplicity.

Remark 7. As we will see in the next subsection, Assumption 4 is sufficient to prove Theorem 1. However, to prove the second result needed for Theorem 2, we require the stronger assumption 4′. The result in the proof of Theorem 3.3 of Park (2002, p. 487, line 12), where it is stated (in Park's notation) that

$$\sum_{k=1}^{p} k|\hat{\alpha}_{p,k}| = \sum_{k=1}^{p} k|\alpha_{p,k}| + o(1) \text{ a.s.},$$

with $\hat{\alpha}_{p,k}$ being the OLS estimators of the p-th order autoregressive approximation of the univariate general linear process considered by Park (2002) with coefficients $\alpha_{p,k}$, does not go through with $p = o((n/\ln n)^{1/2})$. One needs a stronger restriction on p to make the second part $o(1)$.[6] With our stronger Assumption 4′ one can show that Theorem 3.3 of Park (2002) (and consequently our Theorem 2) holds.

3.3 Asymptotic results

In this section we will give the main theoretical results needed to show the asymptotic validity of the bootstrap test. As stated in Remark 5, we derive these results for the tests (and DGP) without deterministic components. The proofs of all the results here plus additional lemmas can be found in Appendix A. Most of the proofs are based on the proofs in Chang et al. (2006), and the papers they refer to.

As we present all our proofs for vector processes, the theory employed in the paper can be used to prove validity of other multivariate bootstrap procedures as well. Note that all our bootstrap weak convergence results hold in probability as we derive all underlying results in probability.

The first step in proving the asymptotic validity is the development of an invariance principle for the bootstrap errors ε_t^*.

Theorem 1. *Under Assumptions 2 and 4, we have that*

$$W_n^*(r) = n^{-1/2} \sum_{t=1}^{[nr]} \varepsilon_t^* \xrightarrow{d^*} LW(r) \quad \textit{in probability}$$

where L is a $(1+l) \times (1+l)$-dimensional lower triangular matrix such that the Cholesky decomposition of Σ is equal to LL'.

[6]We thank Anders Swensen for bringing this point to our attention in a personal communication.

We can show this result by first showing that $\mathrm{E}^* |\varepsilon_t^*|^a = O_p(1)$ for some $a > 2$, and then referring to Einmahl (1987), who shows that an invariance principle holds if this condition is met.

From this result, with the help of the Beveridge-Nelson decomposition, we can construct an invariance principle for u_t^*.

Theorem 2. *Under Assumptions 2 and 4′ we have that*

$$B_n^*(r) = n^{-1/2} \sum_{t=1}^{[nr]} u_t^* \xrightarrow{d^*} B(r) \quad \textit{in probability}$$

where $B(r)$ is a $(1 + l)$-dimensional Brownian motion such that $B(r) = \Psi(1)LW(r)$.

Then, using Theorem 2, we can derive the limiting distributions of the elements of the test statistic, and finally show the consistency of the bootstrap variance estimator. With these results, we can then present Theorem 3 which establishes the asymptotic distribution of the bootstrap test statistic.

Theorem 3. *Under Assumptions 2, 3, 4′ and 5 we have that*

$$T_{\text{wald}}^* \xrightarrow{d^*} \int_0^1 dW_1(r)W(r)' \left[\int_0^1 W(r)W(r)'dr\right]^{-1} \int_0^1 W(r)dW_1(r) \quad \textit{in probability}$$

where T_{wald}^* is defined in equation (12).

Note that Theorem 3 shows that the bootstrap test statistic has the same asymptotic distribution as the original test statistic, which shows that the bootstrap test is asymptotically valid. Also note that the test statistic is asymptotically pivotal, which means that the bootstrap may offer asymptotic refinements, although this does not have to be so.

4 Simulations

We wish to study the small sample properties of our test by simulation. We compare our test with the test based on asymptotic critical values (provided by Boswijk, 1994) and with the three alternative bootstrap tests mentioned in Remarks 2 and 3. Our bootstrap test is denoted by $T_{v,n}^*$, where the subscript v stands for estimation based on the VAR/VECM model, and the n for estimation of the sieve bootstrap under the null. The alternative test discussed in Remark 2 is denoted by $T_{v,a}^*$, with the subscript a indicating estimation under the alternative. Similarly, the two alternatives discussed in Remark 3 are given as $T_{c,n}^*$ and $T_{c,a}^*$, where the subscript c indicates that these are based on the conditional/marginal model. Finally, the asymptotic test is denoted as T_{as}.

For the simulation study we use the same setup as Pesavento (2004). We let the bivariate series $(y_t, x_t)'$ be generated by the triangular system

$$\begin{aligned} y_t &= \gamma x_t + w_t, \\ w_t &= \rho w_{t-1} + v_{1t}, \\ \Delta x_t &= v_{2t}. \end{aligned} \tag{15}$$

We take $\rho = 1$ to analyze the size of tests, and $\rho < 1$ for the power. For the local power analysis, $\rho = 1 + c/n$, where n, the sample size, is either 50 or 100. The tests are invariant to the true value of γ as long as it is non-zero, therefore we set $\gamma = 1$. Furthermore we set $w_0 = x_0 = 0$.

The errors $v_t = (v_{1t}, v_{2t})'$ are generated as

$$(1 - \Phi L)v_t = (1 + \Theta L)\varepsilon_t,$$

where ε_t is generated as an i.i.d. sequence from a bivariate normal distribution with covariance matrix

$$\Sigma = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}.$$

The exact parameter combinations considered are summarized in Table 1.

Insert Table 1 about here

We can rewrite the above DGP in terms of the model in (2) by setting $\alpha = (1, 0)'$, $\beta = (1, -\gamma)'$ and $u_t = \left[\begin{smallmatrix} 1 & \gamma \\ 0 & 1 \end{smallmatrix}\right] v_t$ in equation (2).

The lag lengths in (6), (8) and (11) are selected by BIC, with maximum lag lengths of 8 for $n = 50$ and 11 for $n = 100$. Each generated sample is used to perform all the tests, such that the lag length p in (6) is always the same for all tests. Our results are based on 2000 simulations, with 999 bootstrap replications per simulation.

The results for the DGPs with white noise errors ($\Phi = \Theta = 0$) are given in Table 2. For this case, the asymptotic test has a reasonably good size, but the bootstrap tests clearly have sizes even closer to the nominal size, especially for $n = 50$. The rejection frequencies of the bootstrap tests are somewhat smaller than those of the asymptotic test under the alternatives considered, but it is difficult to compare powers as sizes are not equal. We therefore also report size-corrected powers for the asymptotic test (in the Table as T_{sc}).[7] The size-corrected power of the asymptotic test is close to the power of the bootstrap tests, which shows that the higher raw power of the asymptotic test is mainly due to the higher size distortions. All bootstrap

[7]There is no need to correct the power of the bootstrap tests, as they have virtually no size distortions; their size-corrected powers would be almost the same as their raw powers.

tests perform similarly both in terms of size and power, indicating that there is no evidence of reduced power for the difference-based tests in this setting.

Insert Table 2 about here

Table 3 gives the results for the size of the tests for DGPs with autoregressive and moving-average errors. For all DGPs considered here, there is a clear advantage of using the bootstrap, which virtually eliminates all size distortions except for the negative moving-average coefficients. Again note that the difference between the bootstrap and asymptotic test is the largest for $n = 50$. The bootstrap tests perform fairly similarly, with a minor advantage for the difference-based tests. This is especially noticeable for the DGP with negative moving-average coefficients.

Insert Table 3 about here

To illustrate the power properties for DGPs allowing for some dependence in the errors, we selected one DGP with autoregressive and one with moving-average coefficients from the set considered above. The results are given in Table 4. We again have to be cautious when comparing raw powers as the sizes vary across the tests. We see that the asymptotic test has somewhat higher rejection frequencies than the bootstrap tests, but as in Table 2 the differences are due to high size distortions of the asymptotic test. This is confirmed by the size-corrected power of the asymptotic test, which is not better, and in some cases considerably worse, than the power of the bootstrap tests. The difference-based tests appear to have higher power than the residual-based tests (especially for $n = 50$ and for alternatives close to the null). This is quite surprising, as it is exactly the opposite of what Paparoditis and Politis (2005) found for unit root tests. This may possibly be a small sample phenomenon reflecting the fact that very often imposing invalid restrictions may lead to improved finite sample statistical inference by reducing the effect of sampling errors.

Insert Table 4 about here

These results show that the bootstrap tests all offer significant size improvements over the asymptotic test, while retaining quite good power properties. Note that the four bootstrap tests perform similarly, with a small advantage for the difference-based tests, both in terms of size and power. The bootstrap tests based on the conditional-marginal representation perform as their counterparts based on the vector representation, thus giving no reason to prefer the conditional representation over the more straightforward vector representation.

As suggested by a referee, the similar performances of the bootstrap tests based on the vector representation and the conditional-marginal representation may be due to the normality of the innovations in our DGP. In order address this issue, we also performed simulations where the ε_t's are generated from non-normal distributions, in particular central χ^2- and

t-distributions. The simulation results not reported here show that the two representations also lead to very similar results if the variables are not normal.

We also investigated the sensitivity to the form of v_t. In the first analysis we generate the innovations ε_t as multivariate GARCH errors, which fall outside the class of processes defined by Assumption 2. In the second analysis we consider a Markov-switching model in which the parameters of the short-run dynamics are generated by a Markov process. The results show that the bootstrap tests are robust against both types of processes.

Finally, we also performed simulations with the original DGP using AIC instead of BIC to select lag lengths. The results show that the bootstrap tests are somewhat undersized. The only notable improvement of the size of the bootstrap tests with respect to lag length selection by BIC occurs in the case of the large negative MA parameters. The power of the bootstrap tests is adversely affected by the use of AIC. Surprisingly, the asymptotic test has larger size distortions using AIC than BIC.[8]

5 Deterministic components

In this section we will discuss how to include deterministic components in the tests. Deterministic components have to be included both in the test regression (D_t and D_t^r in equation (6) and their bootstrap counterparts in equation (11)) and in Step 1 of the bootstrap procedure (D_t^s in equation (8)). We consider the five different options proposed by Boswijk (1994).

The first option is to simply leave out all deterministic components, which is the case we analyzed before in the paper. Obviously this is only valid if both μ and τ in equation (1) are equal to zero.

The second and third options (Boswijk's ξ_μ^* and ξ_μ) arise if there is no drift in z_t ($\tau = 0$). In this case we include an intercept in regression (6) and its bootstrap equivalent (11). The intercept can but need not be restricted to zero under the null of no cointegration. In the first case $D_t = 0$ and $D_t^r = 1$, in the second case $D_t = 1$ and $D_t^r = 0$. As in both cases z_t does not have a drift, there is no need to include any deterministic components in Step 1 of the bootstrap procedure, hence $D_t^s = 0$.

If the variables are generated by a process with drift, we have to include a linear trend as well as an intercept in equations (6) and (11) (Boswijk's ξ_τ^* and ξ_τ). Again we can either restrict the trend to be equal to zero under the null, in which case $D_t = 1$ and $D_t^r = t$, or we leave it unrestricted, in which case $D_t = (1, t)'$ and $D_t^r = 0$. As Δz_t now has a nonzero mean, we include a constant term in equation (8) in Step 1 of the bootstrap procedure, i.e. we set $D_t^s = 1$ in both cases.

Remark 8. While it is possible to account for the presence of deterministic components in Step 3 of the bootstrap algorithm as well, it is not necessary. By specifying the tests as above,

[8] The results of all the additional simulations discussed above can be found on the website `http://www.personeel.unimaas.nl/s.smeekes/research.htm`.

the tests are similar, i.e. their asymptotic distributions do not depend on the true value of the deterministic components. Therefore, building the bootstrap process with or without deterministic components will both lead to the correct limiting distribution, as long as the deterministic specification in the bootstrap test regression (11) is the same as the specification in the original regression (6), i.e. $D_t^* = D_t$ and $D_t^{r*} = D_t^r$.[9]

Remark 9. One might want to use the test with the unrestricted constant term to deal with the situation where the variables have a drift, but the drift does not lead to a time trend in the cointegrating relation ($\beta'\tau = 0$). However, Boswijk (1994) stresses that in this case the asymptotic distribution of the test will not be similar and depend on whether the drift is zero or not. Therefore we do not consider this to be a viable option.

Remark 10. One can also adapt the bootstrap procedure mentioned in Remark 2 to the inclusion of deterministic components. As estimation in Step 1 is done under the alternative hypothesis, the inclusion of deterministic components is slightly different. If we only include a constant term in the regression, then a constant term must be included in equation (13) as well, hence $D_t^{s,a} = 1$. If the variables are generated by a drift, and a trend is added to the regression, $D_t^{s,a} = (1, t)'$.

To illustrate the tests with deterministic components, we perform a small simulation study. The DGP used for the simulations corresponds to the DGP used in Section 4, except that we now add deterministic components to the triangular system as follows.

$$\begin{aligned} y_t &= \mu_1 + \tau_1 t + \gamma x_t + w_t, \\ w_t &= \rho w_{t-1} + v_{1t}, \\ \Delta(x_t - \mu_2 - \tau_2 t) &= v_{2t}. \end{aligned} \tag{16}$$

Note that μ_1 and τ_1 correspond to $\beta'\mu$ and $\beta'\tau$ respectively in equation (4). To reduce the size of the experiment we only report simulations for $n = 50$, and for $c = 0$ and $c = -10$. Also, we only consider three combinations of Φ and Θ: $\Phi = \Theta = 0$; $\Phi = \left[\begin{smallmatrix} 0.2 & 0.5 \\ 0.5 & 0.2 \end{smallmatrix}\right]$ and $\Theta = 0$; and $\Phi = 0$ and $\Theta = \left[\begin{smallmatrix} 0.2 & 0.5 \\ 0.5 & 0.2 \end{smallmatrix}\right]$. We restrict our attention to the two bootstrap variants $T_{v,n}^*$ and $T_{v,a}^*$ and the asymptotic test T_{as}.

We consider two models without a drift, and two where a drift is present. For the models without drift, a DGP with no deterministic components and one with just a constant term are chosen. For the models with drift, we select one DGP where the drift cancels out in the direction of the cointegrating vector (i.e. $\tau_1 = 0$), and one where it does not. For each model we perform the tests with every deterministic specification that is appropriate for that specific model. The specific values used and the corresponding empirical rejection frequencies can be found in Table 5.

[9] Unreported simulation results, which can also be found on the website mentioned above, show that in finite samples the tests perform the same whether or not deterministic components are included in Step 3.

Insert Table 5 about here

It can be seen from the table that the size of the bootstrap test is satisfactory for all settings considered. As in Section 4, the null-based test has slightly better size than the alternative-based test in the presence of serial correlation. The asymptotic test has again large size distortions almost everywhere. In terms of power the conclusions are similar to those drawn in Section 4 as well. Also, both in terms of size and power, the rejection frequencies for a particular deterministic specification of the tests ($D_t^{(r)}$), are comparable across different specifications for the trends in the DGP (μ and τ), confirming the similarity of the tests.

Noticeable is that the bootstrap tests lose power if deterministic components are included unnecessarily. This is very much a small sample effect, unreported simulations for $n = 100$ show that this effect, although still present, is less pronounced there. The asymptotic test does not seem to lose as much power. This can be explained by the fact that (contrary to the bootstrap tests) the size distortions of the asymptotic test increase when deterministic components are added unnecessarily. It also appears that the tests with unrestricted deterministic components are slightly more powerful than their restricted counterparts.

6 Conclusion

In this paper we present a bootstrap version of the Wald test for cointegration in a conditional single-equation ECM originally proposed by Boswijk (1994) and also considered by Pesavento (2004). A multivariate sieve bootstrap method is used to deal with dependence in the data, and shown to be asymptotically valid. We also consider several alternative bootstrap tests, for which the asymptotic validity can be established in a similar fashion, and show how deterministic components can be included in the test.

The small sample properties of our bootstrap tests are studied by simulation, and compared to those of the asymptotic test and several alternative bootstrap tests. All bootstrap tests clearly outperform the asymptotic test in terms of size, while retaining good power. Our bootstrap test based on the null hypothesis performs slightly better in terms of size and power than the bootstrap test based on the alternative, while the performance of the tests based on the vector representation is very similar to that of the tests based on the conditional representation. The bootstrap tests with deterministic components retain excellent size properties and are insensitive to the true value of the trends in the model as long as sufficient deterministic components are included.

The results show that our bootstrap version of the Wald ECM test is worth being considered in empirical research, as our test can be seen to improve upon the original Wald test considered by Boswijk (1994) and Pesavento (2004). The Wald ECM test easily allows for other bootstrap variants as well, such as those considered in the simulation study, or block bootstrap methods, which account for somewhat more general DGPs. Such tests could easily be placed in the framework presented here.

A Proofs

All bootstrap weak convergence results that we present in the following are in probability. We do not add this explicitly to every result in order to simplify the notation.

Also note that we define bootstrap stochastic order symbols $O_p^*(\cdot)$ and $o_p^*(\cdot)$ in the same way as $O_p(\cdot)$ and $o_p(\cdot)$ for the original sample (see Chang and Park, 2003, Remark 1).

In order to prove Theorem 1, we first need the following lemma.

Lemma 2. *Under Assumptions 2 and 4 we have for any* $2 < a \leq 4$

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

Proof of Lemma 2. Our proof follows Park (2002, Proof of Lemma 3.2). Note that[10]

$$\begin{aligned}
\mathrm{E}^* \left|\varepsilon_t^*\right|^a &= \frac{1}{n}\sum_{t=1}^n \left|\hat{\varepsilon}_{q,t} - \frac{1}{n}\sum_{t=1}^n \hat{\varepsilon}_{q,t}\right|^a \\
&= \frac{1}{n}\sum_{t=1}^n \left|\hat{\varepsilon}_{q,t} - \varepsilon_{q,t} + \varepsilon_{q,t} - \varepsilon_t + \varepsilon_t - \frac{1}{n}\sum_{t=1}^n \hat{\varepsilon}_{q,t}\right|^a \\
&\leq 4^{a-1}\frac{1}{n}\sum_{t=1}^n \left\{ \left|\hat{\varepsilon}_{q,t} - \varepsilon_{q,t}\right|^a + \left|\varepsilon_{q,t} - \varepsilon_t\right|^a + \left|\varepsilon_t\right|^a + \left|\frac{1}{n}\sum_{t=1}^n \hat{\varepsilon}_{q,t}\right|^a \right\} \\
&= c(A_n + B_n + C_n + D_n)
\end{aligned}$$

where

$$A_n = \frac{1}{n}\sum_{t=1}^n \left|\varepsilon_t\right|^a \qquad B_n = \frac{1}{n}\sum_{t=1}^n \left|\varepsilon_{q,t} - \varepsilon_t\right|^a$$

$$C_n = \frac{1}{n}\sum_{t=1}^n \left|\hat{\varepsilon}_{q,t} - \varepsilon_{q,t}\right|^a \qquad D_n = \left|\frac{1}{n}\sum_{t=1}^n \hat{\varepsilon}_{q,t}\right|^a$$

and $c = 4^{a-1}$ is a constant not depending on n. Note that $\varepsilon_{q,t}$ is defined as

$$\varepsilon_{q,t} = u_t - \sum_{j=1}^q \Phi_j u_{t-j} = \varepsilon_t + \sum_{j=q+1}^\infty \Phi_j u_{t-j}. \tag{17}$$

[10] Every convex function $f(x)$ has the property that $f(\sum_{i=1}^k x_i/k) \leq \sum_{i=1}^k f(x_i)/k$. Applying this to the function $f(x) = |x|^a$, we have

$$\left|\sum_{i=1}^k x_i\right|^a = k^a \left|\sum_{i=1}^k x_i/k\right|^a \leq k^a \sum_{i=1}^k \left|x_i\right|^a/k = k^{a-1}\sum_{i=1}^k \left|x_i\right|^a.$$

We first look at A_n. By the weak law of large numbers, $\frac{1}{n}\sum_{t=1}^n |\varepsilon_t|^a \xrightarrow{p} \mathrm{E}\,|\varepsilon_t|^a$. As by Assumption 2 $\mathrm{E}\,|\varepsilon_t|^a = O(1)$, we have that $A_n = O_p(1)$.

For B_n, we wish to show that $\mathrm{E}\,|\varepsilon_{q,t} - \varepsilon_t|^a = o(q^{-a})$ as $\frac{1}{n}\sum_{t=1}^n |\varepsilon_{q,t} - \varepsilon_t|^a \xrightarrow{p} \mathrm{E}\,|\varepsilon_{q,t} - \varepsilon_t|^a$. Using Minkowski's inequality we have

$$
\begin{aligned}
\mathrm{E}\,|\varepsilon_{q,t} - \varepsilon_t|^a &= \mathrm{E}\left|\sum_{j=q+1}^{\infty} \Phi_j u_{t-j}\right|^a \leq \left(\sum_{j=q+1}^{\infty} \left(\mathrm{E}\,|\Phi_j u_{t-j}|^a\right)^{1/a}\right)^a \\
&\leq \left(\sum_{j=q+1}^{\infty} |\Phi_j| \left(\mathrm{E}\,|u_{t-j}|^a\right)^{1/a}\right)^a = \left(\left(\mathrm{E}\,|u_t|^a\right)^{1/a} \sum_{j=q+1}^{\infty} |\Phi_j|\right)^a \\
&= \mathrm{E}\,|u_t|^a \left(\sum_{j=q+1}^{\infty} |\Phi_j|\right)^a = o(q^{-a}).
\end{aligned}
$$

The final step comes from Bühlmann (1995), where it is shown in Lemma 2.1 that Assumption 2 implies that $\sum_{j=0}^{\infty} j|\Phi_j| < \infty$. It is also shown (in the proof of Theorem 3.1) that $\sum_{j=q+1}^{\infty} j|\Phi_j| = o(1)$ if $\sum_{j=0}^{\infty} j|\Phi_j| < \infty$. Consequently $\sum_{j=q+1}^{\infty} |\Phi_j| = o(q^{-1})$ as $q\sum_{j=q+1}^{\infty} |\Phi_j| \leq \sum_{j=q+1}^{\infty} j|\Phi_j|$.

Next we turn to C_n. We can write

$$
\begin{aligned}
\hat{\varepsilon}_{q,t} &= u_t - \sum_{j=1}^{q} \hat{\Phi}_j u_{t-j} = \varepsilon_{q,t} + \sum_{j=1}^{q} \Phi_j u_{t-j} - \sum_{j=1}^{q} \hat{\Phi}_j u_{t-j} \\
&= \varepsilon_{q,t} - \sum_{j=1}^{q} (\hat{\Phi}_j - \Phi_{q,j}) u_{t-j} - \sum_{j=1}^{q} (\Phi_{q,j} - \Phi_j) u_{t-j}
\end{aligned}
\tag{18}
$$

where $\Phi_{q,j}$ is defined as the coefficient of y_{t-j} in the best linear predictor of y_t in terms of $y_{t-1}, \ldots, y_{t-q}$. Then

$$
|\hat{\varepsilon}_{q,t} - \varepsilon_{q,t}|^a \leq 2^{a-1}\left(\left|\sum_{j=1}^{q} (\hat{\Phi}_j - \Phi_{q,j}) u_{t-j}\right|^a + \left|\sum_{j=1}^{q} (\Phi_{q,j} - \Phi_j) u_{t-j}\right|^a\right).
$$

We define

$$
C_{1n} = \frac{1}{n}\sum_{t=1}^{n} \left|\sum_{j=1}^{q} (\hat{\Phi}_j - \Phi_{q,j}) u_{t-j}\right|^a, \qquad C_{2n} = \frac{1}{n}\sum_{t=1}^{n} \left|\sum_{j=1}^{q} (\Phi_{q,j} - \Phi_j) u_{t-j}\right|^a
$$

and show that $C_{1n}, C_{2n} = o_p(1)$. Then we have that

$$\begin{aligned} C_{1n} &= \frac{1}{n}\sum_{t=1}^{n}\left|\sum_{j=1}^{q}(\hat{\Phi}_j - \Phi_{q,j})u_{t-j}\right|^a \leq q^{a-1}\frac{1}{n}\sum_{t=1}^{n}\sum_{j=1}^{q}\left|(\hat{\Phi}_j - \Phi_{q,j})\right|^a |u_{t-j}|^a \\ &\leq q^{a-1}\left(\max_{1\leq j\leq q}\left|\hat{\Phi}_j - \Phi_{q,j}\right|^a\right)\frac{1}{n}\sum_{t=1}^{n}\sum_{j=1}^{q}|u_{t-j}|^a, \end{aligned}$$

As every value of $|u_{t-j}|^a$ for $j = 1-q, \dots, n-1$ occurs at most q times in the double sum $\sum_{t=1}^{n}\sum_{j=1}^{q}|u_{t-j}|^a$, we have that

$$\begin{aligned} C_{1n} &\leq q^{a-1}\left(\max_{1\leq j\leq q}\left|\hat{\Phi}_j - \Phi_{q,j}\right|^a\right)\frac{1}{n}\sum_{t=1}^{n}\sum_{j=1}^{q}|u_{t-j}|^a \\ &\leq q^{a}\left(\max_{1\leq j\leq q}\left|\hat{\Phi}_j - \Phi_{q,j}\right|^a\right)\frac{1}{n}\left(\sum_{t=0}^{n-1}|u_t|^a + \sum_{t=-1}^{1-q}|u_t|^a\right) \\ &= O_p((\ln n/n)^{a/2})(q^a/n)O_p(n) = O_p(q^a(\ln n/n)^{a/2}), \end{aligned}$$

where we use that

$$\max_{1\leq j\leq q}\left|\hat{\Phi}_j - \Phi_{q,j}\right| = O_p((\ln n/n)^{1/2}) \tag{19}$$

uniformly in $q < Q_n$, where $Q_n = o((n/\ln n)^{1/2})$, from Hannan and Kavalieris (1986). Note that while Hannan and Kavalieris (1986) show their result for the Yule-Walker estimator, (19) is valid for OLS as well by Theorem 1 of Poskitt (1994). To conclude this part of the proof, note that $C_{1n} = o_p(1)$ as $q = o((n/\ln n)^{1/2})$.

For C_{2n}, note that by Markov's inequality for any $\epsilon > 0$

$$\mathrm{P}(|C_{2n}| > \epsilon) \leq \epsilon^{-1}\,\mathrm{E}\left|\frac{1}{n}\sum_{t=1}^{n}\left|\sum_{j=1}^{q}(\Phi_{q,j} - \Phi_j)u_{t-j}\right|^a\right|$$

Then, using Minkowski's inequality and the stationarity of u_t, we have

$$\begin{aligned} \mathrm{E}\left|\sum_{j=1}^{q}(\Phi_{q,j} - \Phi_j)u_{t-j}\right|^a &\leq \left(\sum_{j=1}^{q}(\mathrm{E}\left|(\Phi_{q,j} - \Phi_j)u_{t-j}\right|^a)^{1/a}\right)^a \\ &\leq \left((\mathrm{E}\,|u_t|^a)^{1/a}\sum_{j=1}^{q}|\Phi_{q,j} - \Phi_j|\right)^a \\ &= \mathrm{E}\,|u_t|^a\left(\sum_{j=1}^{q}|\Phi_{q,j} - \Phi_j|\right)^a. \end{aligned}$$

Again from Bühlmann (1995, p. 337), we have that

$$\sum_{j=1}^{q} |\Phi_{q,j} - \Phi_j| \leq c \sum_{j=q+1}^{\infty} |\Phi_j| = o(q^{-1}) \tag{20}$$

with c some constant. Hence, $C_{2n} = o(q^{-a})$ which completes the proof for C_n.

Finally, we look at D_n. We want to show that

$$\frac{1}{n}\sum_{t=1}^{n} \hat{\varepsilon}_{q,t} = \frac{1}{n}\sum_{t=1}^{n} \varepsilon_{q,t} + o_p(1) = \frac{1}{n}\sum_{t=1}^{n} \varepsilon_t + o_p(1).$$

Using equations (17) and (18) we can write

$$\hat{\varepsilon}_{q,t} = \varepsilon_t + \sum_{j=q+1}^{\infty} \Phi_j u_{t-j} - \sum_{j=1}^{q} (\hat{\Phi}_j - \Phi_{q,j}) u_{t-j} - \sum_{j=1}^{q} (\Phi_{q,j} - \Phi_j) u_{t-j}.$$

Hence, what we need to show is that

$$\frac{1}{n}\sum_{t=1}^{n} \sum_{j=q+1}^{\infty} \Phi_j u_{t-j} \xrightarrow{p} 0 \tag{21}$$

$$\frac{1}{n}\sum_{t=1}^{n} \sum_{j=1}^{q} (\Phi_{q,j} - \Phi_j) u_{t-j} \xrightarrow{p} 0 \tag{22}$$

$$\frac{1}{n}\sum_{t=1}^{n} \sum_{j=1}^{q} (\hat{\Phi}_j - \Phi_{q,j}) u_{t-j} \xrightarrow{p} 0. \tag{23}$$

Note that, using Markov's inequality, we have for any $\varepsilon > 0$

$$\mathrm{P}\left(\left| \frac{1}{n}\sum_{t=1}^{n} \sum_{j=q+1}^{\infty} \Phi_j u_{t-j} \right| > \epsilon \right) \leq \epsilon^{-a}\, \mathrm{E} \left| \frac{1}{n}\sum_{t=1}^{n} \sum_{j=q+1}^{\infty} \Phi_j u_{t-j} \right|^a = o(q^{-a})$$

for any $2 < a \leq 4$ which follows from the proof for B_n. This shows (21). To show (22), we can use the proof of C_{2n} to show that

$$\mathrm{P}\left(\left| \frac{1}{n}\sum_{t=1}^{n} \sum_{j=1}^{q} (\Phi_{q,j} - \Phi_j) u_{t-j} \right| > \epsilon \right) \leq \epsilon^{-a}\, \mathrm{E} \left| \frac{1}{n}\sum_{t=1}^{n} \sum_{j=1}^{q} (\Phi_{q,j} - \Phi_j) u_{t-j} \right|^a = o(q^{-a}).$$

Finally, to prove (23), note that

$$\left|\frac{1}{n}\sum_{t=1}^{n}\sum_{j=1}^{q}(\hat{\Phi}_j-\Phi_{q,j})u_{t-j}\right| \leq q\left(\max_{1\leq j\leq q}\left|\hat{\Phi}_j-\Phi_{q,j}\right|\right)\frac{1}{n}\left(\sum_{t=0}^{n-1}|u_t|+\sum_{t=-1}^{1-q}|u_t|\right)$$
$$= O_p(q(\ln n/n)^{1/2})$$

which follows exactly as in the proof of C_{1n}. This shows that $D_n = o_p(1)$, and the proof is complete. □

Before proceeding with the proof of Theorem 1, we need one additional lemma to ensure that the covariance matrix of the bootstrap errors correctly mimics that of the original errors.

Lemma 3. *Under Assumptions 2 and 4 we have that*

$$\Sigma^* = \mathrm{E}^*(\varepsilon_t^*\varepsilon_t^{*\prime}) = \Sigma + o_p(1).$$

Proof of Lemma 3. This proof follows Paparoditis (1996, Proof of Theorem 2.5, p. 288). First note that $\mathrm{E}^*(\varepsilon_t^*\varepsilon_t^{*\prime}) = n^{-1}\sum_{t=1}^{n}\tilde{\varepsilon}_{q,t}\tilde{\varepsilon}_{q,t}'$. Then

$$\begin{aligned}
|\Sigma^*-\Sigma| &= \left|n^{-1}\sum_{t=1}^{n}\tilde{\varepsilon}_{q,t}\tilde{\varepsilon}_{q,t}'-\Sigma\right| = \left|n^{-1}\sum_{t=1}^{n}(\tilde{\varepsilon}_{q,t}\tilde{\varepsilon}_{q,t}'-\varepsilon_t\varepsilon_t')\right| + \left|n^{-1}\sum_{t=1}^{n}\varepsilon_t\varepsilon_t'-\Sigma\right| \\
&= n^{-1}\sum_{t=1}^{n}\left|\tilde{\varepsilon}_{q,t}\tilde{\varepsilon}_{q,t}'-\varepsilon_t\varepsilon_t'\right| + o_p(1) \\
&= n^{-1}\sum_{t=1}^{n}\left|(\tilde{\varepsilon}_{q,t}-\varepsilon_t)\tilde{\varepsilon}_{q,t}'+\varepsilon_t(\tilde{\varepsilon}_{q,t}-\varepsilon_t)'\right| + o_p(1) \\
&\leq n^{-1}\sum_{t=1}^{n}|\tilde{\varepsilon}_{q,t}-\varepsilon_t|\,|\tilde{\varepsilon}_{q,t}| + n^{-1}\sum_{t=1}^{n}|\varepsilon_t|\,|\tilde{\varepsilon}_{q,t}-\varepsilon_t| + o_p(1) \\
&\leq \max_{1\leq t\leq n}|\tilde{\varepsilon}_{q,t}|\,n^{-1}\sum_{t=1}^{n}|\tilde{\varepsilon}_{q,t}-\varepsilon_t| + \max_{1\leq t\leq n}|\varepsilon_t|\,n^{-1}\sum_{t=1}^{n}|\tilde{\varepsilon}_{q,t}-\varepsilon_t| + o_p(1),
\end{aligned}$$

as $n^{-1}\sum_{t=1}^{n}\varepsilon_t\varepsilon_t'-\Sigma \xrightarrow{p} 0$ by the weak law of large numbers.

Note that

$$|\tilde{\varepsilon}_{q,t}| \leq |\hat{\varepsilon}_{q,t}| + \left|n^{-1}\sum_{t=1}^{n}\hat{\varepsilon}_t\right| \leq |\hat{\varepsilon}_{q,t}-\varepsilon_{q,t}| + |\varepsilon_{q,t}-\varepsilon_t| + |\varepsilon_t| + \left|n^{-1}\sum_{t=1}^{n}\hat{\varepsilon}_{q,t}\right|$$

and

$$|\tilde{\varepsilon}_{q,t}-\varepsilon_t| = \left|\hat{\varepsilon}_{q,t}-n^{-1}\sum_{t=1}^{n}\hat{\varepsilon}_{q,t}-\varepsilon_t\right| \leq |\hat{\varepsilon}_{tq,}-\varepsilon_{q,t}| + |\varepsilon_{q,t}-\varepsilon_t| + \left|n^{-1}\sum_{t=1}^{n}\hat{\varepsilon}_t\right|.$$

It now follows that $|\Sigma^* - \Sigma| = o_p(1)$ by applying the methods from the proof of Lemma 2 (with $a = 1$). $\square$

Proof of Theorem 1. In this proof we draw heavily on results by Einmahl (1987), as in Chang et al. (2006). Therefore, we first need to introduce notation used by Einmahl (1987). Let $(\mathbb{R}^d, |\cdot|)$ denote the d-dimensional Euclidean space. Let $C_d[0,1]$ be the space of all continuous $\mathbb{R}^d$-valued functions on $[0,1]$ endowed with the sup-norm $||\cdot||$.

Let $\lambda(Q_1, Q_2, \delta)$ denote the δ-*distance* of two measures Q_1 and Q_2, that is

$$\lambda(Q_1, Q_2, \delta) = \sup\{Q_1(A) - Q_2(A^\delta) : A \subseteq C_d[0,1] \text{ closed}\}$$

where $A^\delta = \{g \in C_d[0,1] : \exists f \in A, ||f - g|| < \delta\}$. Then Einmahl (1987) shows that

$$\lambda(W_n^*, W, \delta) \leq c\delta^{-a} K_{an}^*, \tag{24}$$

holds for all δ if $2 < a < 4$, and $\delta > K_{an}^{*\gamma}$ if $a \geq 4$; where $K_{an}^* = \sum_{t=1}^n \mathrm{E}^* |\varepsilon_t^*|^a$, c is a positive constant depending only on a, l and γ, and $0 < \gamma < 1/(2a-4)$.

By noting that $K_{an}^* = \sum_{t=1}^n \mathrm{E}^* |\varepsilon_t^*|^a = n\,\mathrm{E}^* |\varepsilon_t^*|^a$, we can, as in Chang et al. (2006), transform this into the following condition:

$$\mathrm{P}^* \left\{ \sup_{0 \leq r \leq 1} |W_n^*(r) - W(r)| > n^{-1/2} c_n \right\} \leq K n c_n^{-a}\, \mathrm{E}^* |\varepsilon_t^*|^a \tag{25}$$

for any sequence $\{c_n\}$, $c_n = n^{1/a + \delta_2}$ for any $\delta_2 > 0$, where K is an absolute constant depending only on a and l.

Once we have the result in (25), we can take $0 < \delta_2 < 1/2 - 1/a$, or alternatively, $\delta_2 = 1/2 - 1/a - \epsilon$, where $0 < \epsilon (< 1/2 - 1/a)$. Then on the left-hand side we have $n^{-1/2} c_n = n^{-1/2 + 1/a + \delta_2} = n^{-\epsilon}$. On the right-hand side we have $c_n^{-a} = (n^{1/a + \delta_2})^{-a} = n^{-1 - a\delta_2}$, to show that

$$\mathrm{P}^* \left\{ \sup_{0 \leq r \leq 1} |W_n^*(r) - W(r)| > n^{-\epsilon} \right\} \leq K n^{-(1 + a\delta_2)}\, \mathrm{E}^* |\varepsilon_t^*|^a,$$

from which we can deduce that, as $n \to \infty$,

$$\sup_{0 \leq r \leq 1} |W_n^*(r) - W(r)| = o_p^*(1).$$

$\square$

Proof of Theorem 2. Using the Beveridge-Nelson decomposition, we can write

$$\begin{aligned}\varepsilon_t^* &= u_t^* - \sum_{j=1}^{q} \hat{\Phi}_j u_{t-j}^* \\ &= (I - \sum_{j=1}^{q} \hat{\Phi}_j) u_t^* + \sum_{i=1}^{q} \sum_{j=i}^{q} \hat{\Phi}_j (u_{t-i+1}^* - u_{t-i}^*) \\ &= \hat{\Phi}(1) u_t^* - \sum_{i=1}^{q} \sum_{j=i}^{q} \hat{\Phi}_j (u_{t-i}^* - u_{t-i+1}^*)\end{aligned}$$

and hence

$$u_t^* = \hat{\Psi}(1)\varepsilon_t^* + \hat{\Psi}(1) \sum_{i=1}^{q} \left(\sum_{j=i}^{q} \hat{\Phi}_j \right) (u_{t-i}^* - u_{t-i+1}^*) = \hat{\Psi}(1)\varepsilon_t^* + (\bar{u}_{t-1}^* - \bar{u}_t^*),$$

where $\bar{u}_{t-1}^* = \hat{\Psi}(1) \sum_{i=1}^{q} (\sum_{j=i}^{q} \hat{\Phi}_j) u_{t-i}^*$ and $\hat{\Psi}(1) = \hat{\Phi}(1)^{-1}$. Then

$$\begin{aligned}B_n^*(r) &= n^{-1/2} \sum_{t=1}^{[nr]} u_t^* = n^{-1/2} \sum_{t=1}^{[nr]} \hat{\Psi}(1)\varepsilon_t^* + n^{-1/2} \sum_{t=1}^{[nr]} (\bar{u}_{t-1}^* - \bar{u}_t^*) \\ &= \hat{\Psi}(1) W_n^*(r) + n^{-1/2} (\bar{u}_0^* - \bar{u}_{[nr]}^*)\end{aligned}$$

Hence, we need to show that

$$\hat{\Phi}(1) \xrightarrow{p} \Phi(1) \tag{26}$$

$$\mathrm{P}^* \left\{ \max_{1 \le t \le n} |n^{-1/2} \bar{u}_t^*| > \epsilon \right\} = o_p(1) \tag{27}$$

We can follow Chang et al. (2006, Proof of Theorem 3.3) for the proofs of these result.

We first show (26). Using equations (19) and (20) we have that

$$\left| \hat{\Phi}(1) - \Phi(1) \right| \le \sum_{j=1}^{q} \left| \hat{\Phi}_j - \Phi_{q,j} \right| + \sum_{j=1}^{q} |\Phi_{q,j} - \Phi_j| + \sum_{j=q+1}^{\infty} |\Phi_j| = O_p(q(\ln n/n)^{1/2}) + o(q^{-1}).$$

Hence $\hat{\Phi}(1) = \Phi(1) + o_p(1)$. This proves (26).

To prove (27), we have as in Park (2002)

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

The second inequality follows from Markov's inequality. Hence, we have to show that

$$n^{1-a/2}\, \mathrm{E}^* |\bar{u}_t^*|^a = o_p(1), \tag{28}$$

which for $a > 2$ implies that we have to show that $\mathrm{E}^* |\bar{u}_t^*|^a = O_p(1)$. If the Yule-Walker method is used to estimate (8), the estimated autoregression is always invertible. Although invertibility of the estimated autoregression is not guaranteed for finite samples using OLS, the asymptotic equivalence of OLS to Yule-Walker (Poskitt, 1994, Theorem 1) implies that for large n we can write $u_t^* = \sum_{j=0}^{\infty} \hat{\Psi}_j \varepsilon_{t-j}^*$ and furthermore $\bar{u}_t^* = \sum_{j=0}^{\infty} \bar{\Psi}_j \varepsilon_{t-j}^*$, where $\bar{\Psi}_j = \sum_{i=j+1}^{\infty} \hat{\Psi}_j$. Let $\bar{u}_{(k),t}^*$ be the k-th element of $\bar{u}_t^*$ and let $\bar{\Psi}_{(k),j}$ be the k-th row of $\bar{\Psi}_j$. By successive application of the Marcinkiewicz-Zygmund inequality (Berger, 1991) and Minkowski's inequality element by element, we have

$$
\begin{aligned}
\mathrm{E}^* |\bar{u}_t^*|^a = \mathrm{E}^* \left(\sum_{k=1}^{l+1} \bar{u}_{(k),t}^{*2} \right)^{a/2} & \leq (l+1)^{a/2-1} \sum_{k=1}^{l+1} \mathrm{E}^* |\bar{u}_{(k),t}^*|^a \\
& \leq c_a (l+1)^{a/2-1} \sum_{k=1}^{l+1} \mathrm{E}^* \left(\sum_{j=0}^{\infty} |\bar{\Psi}_{(k),j} \varepsilon_{t-j}^*|^2 \right)^{a/2} \\
& \leq c_a (l+1)^{a/2-1} \sum_{k=1}^{l+1} \left(\sum_{j=0}^{\infty} \left(\mathrm{E}^* |\bar{\Psi}_{(k),j} \varepsilon_{t-j}^*|^a \right)^{2/a} \right)^{a/2} \\
& \leq c_a \sum_{k=1}^{l+1} \left(\sum_{j=0}^{\infty} |\bar{\Psi}_{(k),j}|^2 \right)^{a/2} \mathrm{E}^* |\varepsilon_t^*|^a \\
& \leq c_a (l+1)^{a/2-1} \left(\sum_{j=0}^{\infty} |\bar{\Psi}_j|^2 \right)^{a/2} \mathrm{E}^* |\varepsilon_t^*|^a
\end{aligned} \tag{29}
$$

for some constant c_a not depending on n. Phillips and Solo (1992, p. 973) show that a sufficient condition for $\sum_{j=1}^{\infty} |\bar{\Psi}_j|^2 = O_p(1)$ is

$$
\sum_{j=1}^{\infty} j^{1/2} |\hat{\Psi}_j| = O_p(1). \tag{30}
$$

This will in turn hold if (Hannan and Kavalieris, 1986)

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

We have

$$\begin{aligned}\sum_{j=1}^{q} j^{1/2}|\hat{\Phi}_j| &= \sum_{j=1}^{q} j^{1/2}|\hat{\Phi}_j - \Phi_{q,j} + \Phi_{q,j} - \Phi_j + \Phi_j| \\ &\leq \sum_{j=1}^{q} j^{1/2}|\hat{\Phi}_j - \Phi_{q,j}| + \sum_{j=1}^{q} j^{1/2}|\Phi_{q,j} - \Phi_j| + \sum_{j=1}^{q} j^{1/2}|\Phi_j| \\ &\leq q^{1/2}\sum_{j=1}^{q} |\hat{\Phi}_j - \Phi_{q,j}| + q^{1/2}\sum_{j=1}^{q} |\Phi_{q,j} - \Phi_j| + \sum_{j=1}^{\infty} j^{1/2}|\Phi_j| \\ &\leq q^{3/2}\max_{1\leq j\leq q} |\hat{\Phi}_j - \Phi_{q,j}| + q^{1/2}\sum_{j=1}^{q} |\Phi_{q,j} - \Phi_j| + \sum_{j=1}^{\infty} j^{1/2}|\Phi_j| \\ &= O_p(q^{3/2}(\ln n/n)^{1/2}) + o(q^{-1/2}) + O(1) = O_p(1),\end{aligned}$$

by (19), (20) and Assumption $4'$. Together with Lemma 2 this shows that

$$\mathrm{E}^* |\bar{u}_t^*|^a = O_p(1). \tag{31}$$

This concludes the proof of this theorem. □

Next we need several lemmas in order to show the limiting distribution of the bootstrap test statistic.

Lemma 4. *Let ξ_t^* be the bootstrap equivalent of ξ defined in equation (5), i.e.*

$$y_t^* = \pi_0^{*\prime}\Delta x_t^* + \sum_{j=1}^{\infty}\pi_j^{*\prime}\Delta z_{t-j}^* + \xi_t^*. \tag{32}$$

Then, if Assumptions 2 and $4'$ hold,

$$n^{-1/2}\sum_{t=1}^{[nr]}\xi_t^* \xrightarrow{d^*} B_\xi(r),$$

where $B_\xi(r)$ is a scalar Brownian motion with variance ω^2, i.e $B_\xi(r) = \omega W_1(r)$, where $W_1(r)$ is the first element of the standard Brownian motion $W(r)$.

Proof of Lemma 4. Follows immediately from Theorem 1. □

Lemma 5. *Let f^* denote the spectral density and $\Gamma^*(k)$ the autocovariance function of u_t^*. Under Assumptions 2 and $4'$, we have*

$$\sup_\lambda |f^*(\lambda) - f(\lambda)| = o_p^*(1) \tag{33}$$

and

$$\sum_{k=-\infty}^{\infty} \Gamma^*(k) = \sum_{k=-\infty}^{\infty} \Gamma(k) + o_p^*(1). \tag{34}$$

Proof of Lemma 5. The spectral density $f^*(\lambda)$ of u_t^* is

$$f^*(\lambda) = \frac{1}{2\pi}(I - \sum_{j=1}^{q} \hat{\Phi}_j e^{-\mathrm{i}j\lambda})^{-1}\Sigma^*(I - \sum_{j=1}^{q} \hat{\Phi}_j' e^{\mathrm{i}j\lambda})^{-1}.$$

Note that by Lemma 3 $\Sigma^* \xrightarrow{p} \Sigma$. Furthermore,

$$\begin{aligned}\left|\sum_{j=1}^{q}\left(\hat{\Phi}_j - \Phi_j\right)e^{-\mathrm{i}j\lambda}\right| &\leq \sum_{j=1}^{q}\left|\hat{\Phi}_j - \Phi_{q,j}\right|\left|e^{-\mathrm{i}jv}\right| + \sum_{j=1}^{q}\left|\Phi_{q,j} - \Phi_j\right|\left|e^{-\mathrm{i}jv}\right| \\ &\leq q \max_{1\leq j\leq q}\left|\hat{\Phi}_j - \Phi_{q,j}\right| + \sum_{j=1}^{q}\left|\Phi_{q,j} - \Phi_j\right| = o_p(1)\end{aligned}$$

by (19) and (20). Now the result in (33) follows straightforwardly.

The result in (34) follows trivially by noting that $\sum_{k=-\infty}^{\infty}\Gamma(k) = 2\pi f(0)$ and correspondingly $\sum_{k=-\infty}^{\infty}\Gamma^*(k) = 2\pi f^*(0)$. □

Now we can derive the limiting distributions of the different elements of the test statistic. First define $w_{p,t} = (\Delta x_t', \Delta z_{t-1}', \ldots, \Delta z_{t-p})'$, and let $W_p = (w_{p,1}, \ldots, w_{p,n})'$, $Z_{-1} = (z_0, \ldots, z_{n-1})'$, $\Xi_p = (\xi_{p,1}, \ldots, \xi_{p,n})'$ and $\Delta Y = (\Delta y_1, \ldots, \Delta y_n)'$, and define their bootstrap versions accordingly.

Lemma 6. *Under Assumptions 2, 3, 4′ and 5 we have*

a) $$n^{-2}Z_{-1}^{*\prime}Z_{-1}^* = n^{-2}\sum_{t=1}^{n} z_{t-1}^* z_{t-1}^{*\prime} \xrightarrow{d^*} \int_0^1 B(r)B(r)'dr \tag{35}$$

b) $$n^{-1}Z_{-1}^{*\prime}\Xi_p^* = n^{-1}\sum_{t=1}^{n} z_{t-1}^*\xi_{p,t}^* \xrightarrow{d^*} \int_0^1 B(r)dB_\xi(r) \tag{36}$$

c) $$\left|\left|\left(n^{-1}W_p^{*\prime}W_p^*\right)^{-1}\right|\right| = \left|\left|\left(\frac{1}{n}\sum_{t=1}^{n} w_{p,t}^* w_{p,t}^{*\prime}\right)^{-1}\right|\right| = O_p^*(1) \tag{37}$$

d) $$\left|Z_{-1}^{*\prime}W_p^*\right| = \left|\sum_{t=1}^{n} z_{t-1}^* w_{p,t}^{*\prime}\right| = O_p^*(np^{1/2}) \tag{38}$$

e) $$\left|W_p^{*\prime}\Xi_p^*\right| = \left|\sum_{t=1}^{n} w_{p,t}^*\xi_{p,t}^*\right| = O_p^*(n^{1/2}p^{1/2}). \tag{39}$$

Proof of Lemma 6. First we look at a). As we set $z_0^* = 0$, we have

$$z_t^* = \sum_{i=1}^{t} u_t^*$$

and therefore

$$B_n^*(r) = n^{-1/2} z_{[nr]}.$$

Then by Theorem 2 and the continuous mapping theorem we have

$$n^{-2} \sum_{t=1}^{n} z_{t-1}^* z_{t-1}^{*\prime} = n^{-1} \sum_{t=1}^{n} \int_{(t-1)/n}^{t/n} z_{[nr]}^* z_{[nr]}^{*\prime} dr = \sum_{t=1}^{n} \int_{(t-1)/n}^{t/n} B_n^*(r) B_n^{*\prime}(r) dr$$
$$= \int_0^1 B_n^*(r) B_n^{*\prime}(r) dr \xrightarrow{d^*} \int_0^1 B(r) B(r)' dr$$

as in Chang et al. (2006, Proof of Lemma 3.4).

Next we look at b). We have

$$|n^{-1} \sum_{t=1}^{n} z_{t-1}^* \xi_{p,t}^*| \leq |n^{-1} \sum_{t=1}^{n} z_{t-1}^* \xi_t^*| + |n^{-1} \sum_{t=1}^{n} z_{t-1}^* (\xi_{p,t}^* - \xi_t^*)|.$$

Hence, we first have to show that $n^{-1} \sum_{t=1}^{n} z_{t-1}^* (\xi_{p,t}^* - \xi_t^*) = o_p(1)$. We can follow Chang et al. (2006, Proof of Lemma A.6) for the proof.

Note that $\xi_{p,t}^* = \sum_{k=p+1}^{\infty} \pi_k^{*\prime} u_{t-k}^* + \xi_t^*$, where

$$\pi_k^{*\prime} = \hat{\Phi}_{1,k} - \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1} \hat{\Phi}_{2,k} \tag{40}$$

and $\hat{\Phi}_k = (\hat{\Phi}_{1,k}', \hat{\Phi}_{2,k}')'$. As $\hat{\Phi}_k = 0$ for $k > q$ and using Assumption 5, we have that

$$\sum_{j=p+1}^{\infty} |\pi_j^*| = \sum_{k=p+1}^{q} |\pi_j^*| = o_p(1).$$

Then define $\hat{\Psi}_{p,j}$ such that

$$\xi_t^* - \xi_{p,t}^* = \sum_{k=p+1}^{\infty} \pi_k^{*\prime} u_{t-k}^* = \sum_{j=p+1}^{\infty} \sum_{k=p+1}^{j} \pi_k^{*\prime} \hat{\Psi}_{j-k} \varepsilon_{t-j}^* = \sum_{j=p+1}^{\infty} \hat{\Psi}_{p,j} \varepsilon_{t-j}^*, \tag{41}$$

We then have that

$$\sum_{j=p+1}^{\infty} \left|\hat{\Psi}_{p,j}\right| \leq \sum_{k=p+1}^{j} |\pi_k^{*\prime}| \sum_{j=p+1}^{\infty} |\hat{\Psi}_{j-k}| \leq \left(\sum_{j=p+1}^{\infty} |\pi_j^*| \right) \left(\sum_{i=0}^{\infty} |\hat{\Psi}_i| \right) = \left(\sum_{j=p+1}^{\infty} |\pi_j^*| \right) O_p(1).$$

Define $\eta_t^* = \sum_{i=1}^t \varepsilon_i^*$, such that we can write

$$z_t^* = \hat{\Psi}(1)\eta_t^* + (\bar{u}_0^* - \bar{u}_t^*).$$

Then we have

$$\begin{aligned}\sum_{t=1}^n z_{t-1}^*(\xi_{p,t}^* - \xi_t^*)' &= \sum_{t=1}^n \hat{\Psi}(1)\eta_{t-1}^*(\xi_{p,t}^* - \xi_t^*)' + \sum_{t=1}^n \bar{u}_0^*(\xi_{p,t}^* - \xi_t^*)' - \sum_{t=1}^n \bar{u}_{t-1}^*(\xi_{p,t}^* - \xi_t^*)' \\ &= R_{1n}^* + R_{2n}^* + R_{3n}^*.\end{aligned}$$

We first want to show that $R_{1n}^* = o_p^*(n)$. Let δ_{ij} be the Kronecker delta. We have

$$\begin{aligned}\left|\sum_{t=1}^n \eta_{t-1}^*(\xi_{p,t}^* - \xi_t^*)\right| &= \left|\sum_{t=1}^n \eta_{t-1}^* \left(\sum_{j=p+1}^\infty \hat{\Psi}_{p,j}\varepsilon_{t-j}^*\right)\right| = \left|\sum_{j=p+1}^\infty \sum_{t=1}^n \sum_{i=1}^{t-1} \varepsilon_{t-i}^*\varepsilon_{t-j}^{*\prime}\hat{\Psi}_{p,j}'\right| \\ &= \left|\sum_{j=p+1}^{n-1}(n-j)\Sigma\hat{\Psi}_{p,j}' + \sum_{j=p+1}^\infty \sum_{t=1}^n \sum_{i=1}^{t-1}\left(\varepsilon_{t-i}^*\varepsilon_{t-j}^{*\prime} - \delta_{ij}\Sigma\right)\hat{\Psi}_{p,j}'\right| \\ &= \sum_{j=p+1}^\infty |\hat{\Psi}_{p,j}|O_p^*(n) + \sum_{j=p+1}^\infty |\hat{\Psi}_{p,j}|O_p^*(n^{1/2}) \\ &= \left(\sum_{j=p+1}^\infty |\pi_j^*|\right) O_p^*(n) = o_p^*(n).\end{aligned}$$

Next we turn to R_{2n}^*. We have

$$\begin{aligned}\sum_{t=1}^n(\xi_{p,t}^* - \xi_t^*) &= \sum_{t=1}^n \sum_{j=p+1}^\infty \hat{\Psi}_{p,j}\varepsilon_{t-j}^* = \sum_{j=p+1}^\infty \hat{\Psi}_{p,j}\sum_{t=1}^n \varepsilon_{t-j}^* \\ &= \left(\sum_{j=p+1}^\infty |\hat{\Psi}_{p,j}|\right) O_p^*(n^{1/2}) = \left(\sum_{j=p+1}^\infty |\pi_j^*|\right) O_p^*(n^{1/2}),\end{aligned}$$

from which we can easily see that $R_{2n}^* = o_p^*(n^{1/2})$.

Finally we look at R_{3n}^*. First note that by applying the Beveridge-Nelson decomposition in a slightly different way than before, we can derive that $\bar{u}_t^* = \sum_{j=0}^\infty \sum_{i=j+1}^\infty \hat{\Psi}_i\varepsilon_{t-j-i}^*$. Then, using the same approach as for R^*_{1n}, we have for $|R^*_{3n}|$ that

$$\begin{aligned}
\left|\sum_{t=1}^{n} \bar{u}^*_{t-1}(\xi^*_{p,t} - \xi^*_t)\right| &= \left|\sum_{t=1}^{n} \bar{u}^*_{t-1} \sum_{j=p+1}^{\infty} \hat{\Psi}_{p,j}\varepsilon^*_{t-j}\right| = \left|\sum_{t=1}^{n}\sum_{i=0}^{\infty}\sum_{k=i+1}^{\infty} \hat{\Psi}_k \varepsilon^*_{t-i-1} \sum_{j=p+1}^{\infty} \hat{\Psi}_{p,j}\varepsilon^*_{t-j}\right| \\
&= \left|\sum_{i=0}^{\infty}\sum_{j=p+1}^{\infty}\sum_{k=i}^{\infty} \hat{\Psi}_k \sum_{t=1}^{n} \left(\varepsilon^*_{t-i-1}\varepsilon^{*\prime}_{t-j}\right) \hat{\Psi}'_{p,j}\right| \\
&= \left|n \sum_{j=p+1}^{\infty}\sum_{k=j}^{\infty} \hat{\Psi}_k \Sigma \hat{\Psi}'_{p,j} + \sum_{i=0}^{\infty}\sum_{k=i}^{\infty} \hat{\Psi}_k \sum_{j=p+1}^{\infty}\sum_{t=1}^{n} \left(\varepsilon^*_{t-i-1}\varepsilon^*_{t-j} - \delta_{i+1,j}\Sigma\right) \hat{\Psi}'_{p,j}\right| \\
&= \sum_{j=p+1}^{\infty}\sum_{k=j}^{\infty} |\hat{\Psi}_k||\hat{\Psi}_{p,j}| O^*_p(n) + \sum_{i=0}^{\infty}\sum_{k=i}^{\infty} |\hat{\Psi}_k| \sum_{j=p+1}^{\infty} |\hat{\Psi}_{p,j}| O^*_p(n^{1/2}) \\
&= \sum_{j=p+1}^{\infty} |\hat{\Psi}_{p,j}| O^*_p(n) + \sum_{j=p+1}^{\infty} |\hat{\Psi}_{p,j}| O^*_p(n^{1/2}) \\
&= \left(\sum_{j=p+1}^{\infty} |\pi^*_j|\right) O^*_p(n).
\end{aligned}$$

Therefore, $R^*_{3n} = o^*_p(n)$, and hence

$$n^{-1}\sum_{t=1}^{n} z^*_{t-1}(\xi^*_{p,t} - \xi^*_t) = n^{-1}\left(R^*_{1n} + R^*_{2n} + R^*_{3n}\right) = o^*_p(1).$$

Then

$$n^{-1}\sum_{t=1}^{n} z^*_{t-1}\xi^*_{p,t} = n^{-1}\sum_{t=1}^{n} z^*_{t-1}\xi^*_t + o^*_p(1),$$

while by Park and Phillips (1989, Lemma 2.1), Theorem 2 and Lemma 4, we have that

$$n^{-1}\sum_{t=1}^{n} z^*_{t-1}\xi^*_t \xrightarrow{d^*} \int_0^1 B(r)dB_\xi(r).$$

This completes the proof of part b).

For c), we want to show that

$$\mathrm{E}^*\left\|\left(\frac{1}{n}\sum_{t=1}^{n} w^*_{p,t}w^{*\prime}_{p,t}\right)^{-1}\right\| = O_p(1). \tag{42}$$

Following Chang and Park (2003, Proof of Lemma 3) we first want to show that

$$\mathrm{E}^* \left| \sum_{t=1}^{n} \left[u^*_{t-i} u^{*\prime}_{t-j} - \Gamma^*(i-j) \right] \right|^2 = O_p(n). \tag{43}$$

For this to hold it is sufficient to show that

$$\mathrm{E}^* \left(\sum_{t=1}^{n} \left[u^*_{(a),t-i} u^{*\prime}_{(b),t-j} - \Gamma^*_{(ab)}(i-j) \right] \right)^2 = O_p(n), \tag{44}$$

for all $1 \leq a, b \leq 1+l$, where $u^*_{(a),t}$ is the a-th element of u^*_t, and similarly $\Gamma^*_{(ab)}(i-j)$ is the (a,b)-th element of $\Gamma^*(i-j)$.

Analogous to the case for univariate time series models discussed in Berk (1974, eqs (2.10) and (2.11), p. 491), we have that

$$\mathrm{E}^* \left(\sum_{t=1}^{n} \left[u^*_{(a),t-i} u^*_{(b),t-i} - \Gamma^*_{(ab)}(i-j) \right] \right)^2 \leq 2n \sum_{k=-\infty}^{\infty} \Gamma^*_{(ab)}(k)^2$$
$$+ n \sum_{c,d,e,f} |\tilde{\kappa}^*_{cdef}| \left(\sum_{k=0}^{\infty} \Psi^{*2}_{(ac),k} \right)^{1/2} \left(\sum_{k=0}^{\infty} \Psi^{*2}_{(bd),k} \right)^{1/2} \left(\sum_{k=0}^{\infty} \Psi^{*2}_{(ae),k} \right)^{1/2} \left(\sum_{k=0}^{\infty} \Psi^{*2}_{(bf),k} \right)^{1/2},$$

where $\tilde{\kappa}^*_{cdef} = \mathrm{E}^*(\varepsilon^*_{(c),t}\varepsilon^*_{(d),t}\varepsilon^*_{(e),t}\varepsilon^*_{(f),t}) - \sigma_{cd}\sigma_{ef} - \sigma_{ce}\sigma_{df} - \sigma_{cf}\sigma_{de}$ and $\sigma_{cd} = \mathrm{E}^*(\varepsilon^*_{(c),t}\varepsilon^*_{(d),t})$. Note that $|\tilde{\kappa}^*_{cdef}| = O_p(1)$ as $\mathrm{E}^*|\varepsilon^*_t|^4 = O_p(1)$ (take $a = 4$ in Lemma 2). Furthermore, $\sum_{k=-\infty}^{\infty} \Gamma^*_{(ab)}(k)^2 = O_p(1)$ through Lemma 5 and $\left(\sum_{k=0}^{\infty} \Psi^{*2}_{(ac),k}\right)^{1/2} = O_p(1)$ as $\sum_{k=0}^{\infty} k^{1/2}|\hat{\Psi}_k| = O_p(1)$, which we demonstrated in the proof of Theorem 2, equation (30). Now equation (44) follows straightforwardly.

Next, partition $\Gamma^*(k)$ conformably with y_t and x_t as

$$\Gamma^*(k) = \begin{bmatrix} \Gamma^*_{11}(k) & \Gamma^*_{12}(k) \\ \Gamma^*_{21}(k) & \Gamma^*_{22}(k) \end{bmatrix}$$

and define $\Gamma^*_{2\cdot}(k) = [\Gamma^*_{21}(k), \Gamma^*_{22}(k)]$ and $\Gamma^*_{\cdot 2}(k) = [\Gamma^*_{12}(k)', \Gamma^*_{22}(k)']'$. Then define Ω^*_{pp} as

$$\Omega^*_{pp} = \begin{bmatrix} \Gamma^*_{22}(0) & \Gamma^*_{2\cdot}(-1) & \dots & \Gamma^*_{2\cdot}(-p) \\ \Gamma^*_{\cdot 2}(1) & \Gamma^*(0) & \dots & \Gamma^*(1-p) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma^*_{\cdot 2}(p) & \Gamma^*(p-1) & \cdots & \Gamma^*(0) \end{bmatrix}.$$

As for any matrix M, $||M||^2 \leq \sum_{i,j} ||M_{ij}||^2$,[11] we can write

$$\mathrm{E}^* \left|\left| \frac{1}{n} \sum_{t=1}^{n} w_{p,t}^* w_{p,t}^{*\prime} - \Omega_{pp}^* \right|\right|^2 \leq \mathrm{E}^* \left(\left|\left| \frac{1}{n} \sum_{t=1}^{n} u_{2,t}^* u_{2,t}^{*\prime} - \Gamma_{22}^*(0) \right|\right|^2 + \sum_{j=1}^{p} \left|\left| \frac{1}{n} \sum_{t=1}^{n} u_{2,t}^* u_{t-j}^{*\prime} - \Gamma_{2\cdot}^*(-j) \right|\right|^2 \right.$$
$$\left. + \sum_{i=1}^{p} \left|\left| \frac{1}{n} \sum_{t=1}^{n} u_{t-i}^* u_{2,t}^{*\prime} - \Gamma_{\cdot 2}^*(i) \right|\right|^2 + \sum_{(i=1,j=1)}^{(p,p)} \left|\left| \frac{1}{n} \sum_{t=1}^{n} u_{t-i}^* u_{t-j}^{*\prime} - \Gamma^*(j-i) \right|\right|^2 \right)$$
$$= O_p(n^{-1}) + O_p(n^{-1}p) + O_p(n^{-1}p) + O_p(n^{-1}p^2) = O_p(n^{-1}p^2).$$

Next we need to show that

$$||\Omega_{pp}^{*-1}|| \leq \left[2\pi \left(\inf_{\lambda} f^*(\lambda) \right) \right]^{-1} = O_p(1).$$

Let us consider an "extended" Ω_{pp}^* matrix, i.e.

$$\tilde{\Omega}_{pp}^* = \begin{bmatrix} \Gamma_{11}^*(0) & \Gamma_{12}^*(0) & \Gamma_{1\cdot}^*(-1) & \dots & \Gamma_{1\cdot}^*(-p) \\ \Gamma_{21}^*(0) & & & & \\ \Gamma_{\cdot 1}^*(1) & & \Omega_{pp}^* & & \\ \vdots & & & & \\ \Gamma_{\cdot 1}^*(p) & & & & \end{bmatrix} = \begin{bmatrix} \Gamma^*(0) & \Gamma^*(-1) & \dots & \Gamma^*(-p) \\ \Gamma^*(1) & \Gamma^*(0) & \dots & \Gamma^*(1-p) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma^*(p) & \Gamma^*(p-1) & \cdots & \Gamma^*(0) \end{bmatrix}.$$

Let $\lambda^* = (\lambda_1^*, \dots, \lambda_{(l+1)p}^*)'$ be the eigenvalues of $\tilde{\Omega}_{pp}^*$ and define $0 < F_1^* = \inf_\lambda ||f^*(\lambda)||$. Then as a direct consequence of Lemma A.2 of Chang et al. (2006) we have that

$$||\Omega_{pp}^{*-1}|| \leq (2\pi F_1^*)^{-1} = O_p(1).$$

[11] We let M_{ij} denote submatrices into which one can partition M.

As $\left|\left|\Omega_{pp}^{*-1}\right|\right| \leq \left|\left|\tilde{\Omega}_{pp}^{*-1}\right|\right|$,[12] we know that $\left|\left|\Omega_{pp}^{*-1}\right|\right| = O_p(1)$ as well. Then

$$\begin{aligned}
\left|\left|\left(\frac{1}{n}\sum_{t=1}^{n} w_{p,t}^{*} w_{p,t}^{*\prime}\right)^{-1}\right|\right| &\leq \left|\left|\Omega_{pp}^{*-1}\right|\right| + \left|\left|\left(\frac{1}{n}\sum_{t=1}^{n} w_{p,t}^{*} w_{p,t}^{*\prime}\right)^{-1} - \Omega_{pp}^{*-1}\right|\right| \\
&= \left|\left|\Omega_{pp}^{*-1}\right|\right| + \left|\left|\Omega_{pp}^{*-1}\left(\Omega_{pp}^{*} - \frac{1}{n}\sum_{t=1}^{n} w_{p,t}^{*} w_{p,t}^{*\prime}\right)\left(\frac{1}{n}\sum_{t=1}^{n} w_{p,t}^{*} w_{p,t}^{*\prime}\right)^{-1}\right|\right| \\
&\leq \left|\left|\Omega_{pp}^{*-1}\right|\right| + \left|\left|\Omega_{pp}^{*-1}\right|\right| \left|\left|\frac{1}{n}\sum_{t=1}^{n} w_{p,t}^{*} w_{p,t}^{*\prime} - \Omega_{pp}^{*}\right|\right| \left|\left|\left(\frac{1}{n}\sum_{t=1}^{n} w_{p,t}^{*} w_{p,t}^{*\prime}\right)^{-1}\right|\right|
\end{aligned}$$

which implies that

$$\left|\left|\left(\frac{1}{n}\sum_{t=1}^{n} w_{p,t}^{*} w_{p,t}^{*\prime}\right)^{-1}\right|\right| \leq \frac{\left|\left|\Omega_{pp}^{*-1}\right|\right|}{1 - \left|\left|\Omega_{pp}^{*-1}\right|\right| \left|\left|\frac{1}{n}\sum_{t=1}^{n} w_{p,t}^{*} w_{p,t}^{*\prime} - \Omega_{pp}^{*}\right|\right|}$$

holds for large n with probability 1 as $\left|\left|\frac{1}{n}\sum_{t=1}^{n} w_{p,t}^{*} w_{p,t}^{*\prime} - \Omega_{pp}^{*}\right|\right| = O_p^{*}(n^{-1/2}p)$. As $\left|\left|\Omega_{pp}^{*-1}\right|\right| = O_p(1)$, the result in (42) follows.

For d) we want to show that

$$\mathrm{E}^{*}\left|\sum_{t=1}^{n} z_{t-1}^{*} w_{p,t}^{*\prime}\right| = O_p(np^{1/2}).$$

Following Chang and Park (2002, Proof of Lemma 3.2), we write

$$\sum_{t=1}^{n} z_{t-1}^{*} u_{t-j}^{*\prime} = \sum_{t=1}^{n} z_{t-1}^{*} u_{t}^{*\prime} + \sum_{t=1}^{n} z_{t-1}^{*} u_{t-j}^{*\prime} - \sum_{t=1}^{n} z_{t-1}^{*} u_{t}^{*\prime} = \sum_{t=1}^{n} z_{t-1}^{*} u_{t}^{*\prime} + R_n^{*}$$

and we want to show that $R_n^{*} = O_p^{*}(n)$ uniformly in $j = 1, \ldots, p$.

[12]Suppose we have a matrix M and a vector v that we can write as

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \qquad v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.$$

Then we have that

$$\begin{aligned}
||M||^2 &= \max_{v} |Mv|^2 / |v|^2 = \max_{v_1,v_2} \left|\begin{matrix} M_{11}v_1 + M_{12}v_2 \\ M_{21}v_1 + M_{22}v_2 \end{matrix}\right|^2 \Bigg/ \left|\begin{matrix} v_1 \\ v_2 \end{matrix}\right|^2 \\
&= \max_{v_1,v_2} \frac{|M_{11}v_1 + M_{12}v_2|^2 + |M_{21}v_1 + M_{22}v_2|^2}{|v_1|^2 + |v_2|^2} \geq \max_{v_1,v_2} \frac{|M_{11}v_1|^2 + |M_{12}v_2|^2 + |M_{21}v_1|^2 + |M_{22}v_2|^2}{|v_1|^2 + |v_2|^2} \\
&\geq \max_{v_2(v_1=0)} \frac{|M_{12}v_2|^2 + |M_{22}v_2|^2}{|v_2|^2} \geq \max_{v_2} \frac{|M_{22}|^2}{|v_2|^2} = ||M_{22}||^2.
\end{aligned}$$

First, write

$$\sum_{t=1}^{n} z_{t-1}^* u_t^{*\prime} = \sum_{t=1}^{n-j} z_{t-1}^* u_t^{*\prime} + \sum_{t=n-j+1}^{n} z_{t-1}^* u_t^{*\prime} = \sum_{t=1}^{n} z_{t-j-1}^* u_{t-j}^{*\prime} + \sum_{t=n-j+1}^{n} z_{t-1}^* u_t^{*\prime}$$

(as $u_t^* = 0$ for $t = 0, \ldots, -p$) and rewrite R_n^* as

$$\sum_{t=1}^{n} (z_{t-1}^* - z_{t-j-1}^*) u_{t-j}^{*\prime} - \sum_{t=n-j+1}^{n} z_{t-1}^* u_t' = R_{1n}^* - R_{2n}^*.$$

Then

$$\begin{aligned} R_{1n}^* &= \sum_{t=1}^{n} (z_{t-1}^* - z_{t-j-1}^*) u_{t-j}^{*\prime} = \sum_{t=1}^{n} \left(\sum_{i=1}^{j} u_{t-i}^* \right) u_{t-j}^{*\prime} = \sum_{i=1}^{j} \sum_{t=1}^{n} u_{t-i}^* u_{t-j}^{*\prime} \\ &= n \sum_{i=1}^{j} \Gamma^*(i-j) + \sum_{i=1}^{j} \left[\sum_{t=1}^{n} (u_{t-i}^* u_{t-j}^{*\prime} - \Gamma^*(i-j)) \right] \\ &= O_p(n) + O_p^*(n^{1/2} p) \end{aligned}$$

as $1 \leq j \leq p$, the result in equation (43), and the fact that $\sum_{k=-\infty}^{\infty} \Gamma^*(k) = O_p(1)$ by Assumption 2 and Lemma 5. We can write R_{2n}^* as

$$\begin{aligned} R_{2n}^* &= \sum_{t=n-j+1}^{n} \sum_{i=1}^{t-1} u_{t-i}^* u_t^{*\prime} = \sum_{t=n-j+1}^{n} \left(\sum_{i=1}^{n-j} u_{t-i}^* + \sum_{i=n-j+1}^{t-1} u_{t-i}^* \right) u_t^{*\prime} \\ &= \sum_{t=n-j+1}^{n} \sum_{i=1}^{n-j} u_{t-i}^* u_t^{*\prime} + \sum_{t=n-j+1}^{n} \sum_{i=n-j+1}^{t-1} u_{t-i}^* u_t^{*\prime} = R_{2n}^{a*} + R_{2n}^{b*}. \end{aligned}$$

Then we have

$$R_{2n}^{a*} = j \sum_{i=1}^{n-j} \Gamma^*(i) + \sum_{t=n-j+1}^{n} \left[\sum_{i=1}^{n-j} (u_{t-i}^* u_t^{*\prime} - \Gamma(i)^*) \right] = O_p(p) + O_p^*(n^{1/2} p)$$

and

$$R_{2n}^{b*} = j \sum_{i=n-j+1}^{t-j} \Gamma^*(i) + \sum_{t=n-j+1}^{n} \left[\sum_{i=n-j+1}^{t-1} (u_{t-i}^* u_t^{*\prime} - \Gamma(i)^*) \right] = O_p(p) + O_p^*(p^{3/2}),$$

as

$$\sum_{t=n-j+1}^{n}\left[\sum_{i=n-j+1}^{t-1}(u_{t-i}^{*}u_{t}^{*\prime}-\Gamma(i)^{*})\right] = \sum_{t=n-j+1}^{n}\left[\sum_{i=1}^{t-(n-j)-1}(u_{t-n+j-i}^{*}u_{t}^{*\prime}-\Gamma(i-n+j)^{*})\right]$$
$$= \sum_{t=n-j+1}^{n} O_p^{*}((t-n+j)^{1/2}) = \sum_{t=n-j+1}^{n} O_p^{*}(j^{1/2})$$
$$= \sum_{t=n-j+1}^{n} O_p^{*}(p^{1/2}) = jO_p^{*}(p^{1/2}) = O_p^{*}(p^{3/2}).$$

Hence,

$$\sum_{t=1}^{n} z_{t-1}^{*}u_{t-j}^{*\prime} = \sum_{t=1}^{n} z_{t-1}^{*}u_{t}^{*\prime} + R_n^{*} = \sum_{t=1}^{n} z_{t-1}^{*}u_{t}^{*\prime} + O_p(n) + O_p^{*}(n^{1/2}p).$$

Note that

$$\left|\sum_{t=1}^{n} z_{t-1}^{*}u_{t}^{*\prime}\right| = O_p^{*}(n),$$

by Phillips (1988), and

$$\left|\sum_{t=1}^{n} z_{t-1}^{*}u_{2,t}^{*\prime}\right| = \left|\sum_{t=1}^{n} z_{t-1}^{*}u_{t}^{*\prime}\right| + O_p^{*}(n).$$

Then

$$\mathrm{E}^{*}\left|\sum_{t=1}^{n} z_{t-1}^{*}w_{p,t}^{*\prime}\right| = \mathrm{E}^{*}\left|\sum_{t=1}^{n}\left[z_{t-1}^{*}u_{2,t}^{*\prime} \quad z_{t-1}^{*}u_{t-1}^{*\prime} \quad \dots \quad z_{t-1}^{*}u_{t-p}^{*\prime}\right]\right|$$
$$= \mathrm{E}^{*}\left(\left|\sum_{t=1}^{n} z_{t-1}^{*}u_{2,t}^{*\prime}\right|^{2} + \sum_{j=1}^{p}\left|\sum_{t=1}^{n} z_{t-1}^{*}u_{t-j}^{*\prime}\right|^{2}\right)^{1/2}$$
$$= \mathrm{E}^{*}\left(\left|\sum_{t=1}^{n} z_{t-1}^{*}u_{2,t}^{*\prime}\right|^{2} + \sum_{j=1}^{p}\left(\left|\sum_{t=1}^{n} z_{t-1}^{*}u_{t}^{*\prime}\right| + O_p(n) + O_p^{*}(n^{1/2}p)\right)^{2}\right)^{1/2}$$
$$= \mathrm{E}^{*}\left(\left(\left|\sum_{t=1}^{n} z_{t-1}^{*}u_{t}^{*\prime}\right| + O_p^{*}(n)\right)^{2} + p\left(\left|\sum_{t=1}^{n} z_{t-1}^{*}u_{t}^{*\prime}\right| + O_p(n) + O_p^{*}(n^{1/2}p)\right)^{2}\right)^{1/2}$$
$$= O_p(np^{1/2}).$$

Finally, we look at e). We want to show that

$$\left|\sum_{t=1}^{n} w_{p,t}^{*}\xi_{p,t}^{*\prime}\right| = O_p^*(n^{1/2}p^{1/2}).$$

Write

$$\sum_{t=1}^{n} w_{p,t}^{*}\xi_{p,t}^{*} = \sum_{t=1}^{n} w_{p,t}^{*}\xi_{t}^{*} + \sum_{t=1}^{n} w_{p,t}^{*}\left(\xi_{p,t}^{*} - \xi_{t}^{*}\right).$$

We first show that

$$\sum_{t=1}^{n} u_{t-j}^{*}\left(\xi_{p,t}^{*} - \xi_{t}^{*}\right) = o_p^*(n^{1/2})$$

uniformly in $1 \leq j \leq p$. We have that

$$\begin{aligned}
\left|\sum_{t=1}^{n} u_{t-j}^{*}\left(\xi_{p,t}^{*} - \xi_{t}^{*}\right)\right| &= \left|\sum_{t=1}^{n} u_{t-j}^{*}\left(\sum_{k=p+1}^{\infty} \hat{\Psi}_{p,k}\varepsilon_{t-k}^{*}\right)\right| = \left|\sum_{t=1}^{n}\left(\sum_{m=0}^{\infty} \hat{\Psi}_{m}\varepsilon_{t-j-m}\right)\left(\sum_{k=p+1}^{\infty} \hat{\Psi}_{p,k}\varepsilon_{t-k}^{*}\right)\right| \\
&= \left|\sum_{k=p+1}^{\infty}\sum_{t=1}^{n}\sum_{m=0}^{\infty} \hat{\Psi}_{m}\varepsilon_{t-j-m}^{*}\varepsilon_{t-k}^{*\prime}\hat{\Psi}_{p,k}^{\prime}\right| \\
&= \left|n\sum_{k=p+1}^{n-1} \hat{\Psi}_{k}\Sigma\hat{\Psi}_{p,k}^{\prime} + \sum_{k=p+1}^{\infty}\sum_{m=0}^{\infty} \hat{\Psi}_{m}\sum_{t=1}^{n}\left(\varepsilon_{t-j-m}^{*}\varepsilon_{t-k}^{*\prime} - \delta_{(j+m)k}\Sigma\right)\hat{\Psi}_{p,k}^{\prime}\right| \\
&= \sum_{k=p+1}^{\infty} |\hat{\Psi}_{k}||\hat{\Psi}_{p,k}|O_p^*(n) + \sum_{m=0}^{\infty}\sum_{k=p+1}^{\infty} |\hat{\Psi}_{m}||\hat{\Psi}_{p,k}|O_p^*(n^{1/2}) = o_p^*(n^{1/2}),
\end{aligned}$$

as in Chang et al. (2006, Proof of Lemma A.6), such that

$$\sum_{t=1}^{n} u_{t-j}^{*}\xi_{p,t}^{*} = \sum_{t=1}^{n} u_{t-j}^{*}\xi_{t}^{*} + o_p^*(n^{1/2}).$$

Furthermore,

$$\begin{aligned}
\mathrm{E}^*\left|\sum_{t=1}^{n} u_{t-j}^{*}\xi_{t}^{*}\right|^2 &= \mathrm{E}^*\left(\sum_{s=1}^{n} u_{s-j}^{*}\xi_{s}^{*}\right)^{\prime}\left(\sum_{t=1}^{n} u_{t-j}^{*}\xi_{t}^{*}\right) \\
&= \sum_{s=1}^{n}\sum_{t=1}^{n} \mathrm{E}^*\, u_{s-j}^{*\prime}u_{t-j}^{*}\xi_{s}^{*}\xi_{t}^{*} = \sum_{s=1}^{n}\sum_{t=1}^{n} \mathrm{E}^*\, u_{s-j}^{*\prime}u_{t-j}^{*}\,\mathrm{E}^*\,\xi_{s}^{*}\xi_{t}^{*} \\
&= \sum_{t=1}^{n} \mathrm{E}^*\, u_{t-j}^{*\prime}u_{t-j}^{*}\,\mathrm{E}^*\,\xi_{t}^{*2} = O_p(n).
\end{aligned}$$

Then

$$\left|\sum_{t=1}^{n} w_{p,t}^{*}\xi_{p,t}^{*}\right| = \left(\sum_{j=1}^{p}\left|\sum_{t=1}^{n} u_{t-j}^{*}\xi_{p,t}^{*}\right|^{2}\right)^{1/2} = O_p^{*}(n^{1/2}p^{1/2}),$$

which concludes the proof. □

The following lemma shows the consistency of the bootstrap variance estimator.

Lemma 7. *Let $\hat{\omega}^{*2}$ be the estimator of the variance of the bootstrap errors $\xi_{p,t}^{*}$ in regression* (11)*, i.e.*

$$\hat{\omega}^{*2} = \frac{1}{n}(\Delta y^{*} - Z_{-1}^{*}\hat{\delta}^{*})'(I - W_p^{*}(W_p^{*\prime}W_p^{*})^{-1}W_p^{*\prime})(\Delta y^{*} - Z_{-1}^{*}\hat{\delta}^{*}).$$

Then $\hat{\omega}^{*2} \xrightarrow{p^{*}} \omega^{2}$ under Assumptions 2, 3, 4′ and 5.

Proof of Lemma 7. Note that

$$\begin{aligned}
n\hat{\omega}^{*2} &= (\Delta y^{*} - Z_{-1}^{*}\hat{\delta}^{*})'(I - W_p^{*}(W_p^{*\prime}W_p^{*})^{-1}W_p^{*\prime})(\Delta y^{*} - Z_{-1}^{*}\hat{\delta}^{*}) \\
&= \Delta y^{*\prime}(I - W_p^{*}(W_p^{*\prime}W_p^{*})^{-1}W_p^{*\prime})\Delta y^{*} - \Delta y^{*\prime}(I - W_p^{*}(W_p^{*\prime}W_p^{*})^{-1}W_p^{*\prime})Z_{-1}^{*}\hat{\delta}^{*} \\
&\quad - \hat{\delta}^{*\prime}Z_{-1}^{*\prime}(I - W_p^{*}(W_p^{*\prime}W_p^{*})^{-1}W_p^{*\prime})\Delta y^{*} + \hat{\delta}^{*\prime}Z_{-1}^{*\prime}(I - W_p^{*}(W_p^{*\prime}W_p^{*})^{-1}W_p^{*\prime})Z_{-1}^{*}\hat{\delta}^{*} \\
&= \Xi_p^{*\prime}(I - W_p^{*}(W_p^{*\prime}W_p^{*})^{-1}W_p^{*\prime})\Xi_p^{*} - \Xi_p^{*\prime}(I - W_p^{*}(W_p^{*\prime}W_p^{*})^{-1}W_p^{*\prime})Z_{-1}^{*}\hat{\delta}^{*} \\
&\quad - \hat{\delta}^{*\prime}Z_{-1}^{*\prime}(I - W_p^{*}(W_p^{*\prime}W_p^{*})^{-1}W_p^{*\prime})\Xi_p^{*} + \hat{\delta}^{*\prime}Z_{-1}^{*\prime}(I - W_p^{*}(W_p^{*\prime}W_p^{*})^{-1}W_p^{*\prime})Z_{-1}^{*}\hat{\delta}^{*}.
\end{aligned}$$

which we write as

$$\hat{\omega}^{*2} = C_n^{*} - 2D_n^{*} + E_n^{*}$$

We first look at C_n^{*}. Write

$$C_n^{*} = n^{-1}\Xi_p^{*\prime}\Xi_p^{*} - n^{-1}\Xi_p^{*\prime}W_p^{*}(W_p^{*\prime}W_p^{*})^{-1}W_p^{*\prime}\Xi_p^{*}.$$

Using that $\hat{\delta}^{*} = O_p(n^{-1})$ and the results from Lemma 6, we have that

$$\begin{aligned}
n^{-1}\left|\Xi_p^{*\prime}W_p^{*}(W_p^{*\prime}W_p^{*})^{-1}W_p^{*\prime}\Xi_p^{*}\right| &\leq n^{-1}\left|n^{-1}\Xi_p^{*\prime}W_p^{*}\right| \left|\left|(W_p^{*\prime}W_p^{*})^{-1}\right|\right| \left|W_p^{*\prime}\Xi_p^{*}\right| \\
&= n^{-1}O_p^{*}(n^{1/2}p^{1/2})O_p^{*}(n^{-1})O_p^{*}(n^{1/2}p^{1/2}) = o_p^{*}(n^{-1/2}).
\end{aligned}$$

Hence,

$$C_n^{*} = n^{-1}\Xi_p^{*\prime}\Xi_p^{*} + o_p^{*}(1).$$

Next we turn to D_n^*. We can write D_n^* as

$$D_n^* = n^{-1}\Xi_p^{*\prime} Z_{-1}^* \hat{\delta}^* - n^{-1}\Xi_p^{*\prime} W_p^* (W_p^{*\prime} W_p^*)^{-1} W_p^{*\prime} Z_{-1}^* \hat{\delta}^*.$$

Again using Lemma 6 and $\hat{\delta}^* = O_p^*(n^{-1})$, we have

$$\begin{aligned} |D_n^*| &\le \left|\Xi_p^{*\prime} Z_{-1}^*\right| \left|\hat{\delta}^*\right| + \left|\Xi_p^{*\prime} W_p^*\right| \left|\left|(W_p^{*\prime} W_p^*)^{-1}\right|\right| \left|W_p^{*\prime} Z_{-1}^*\right| \left|\hat{\delta}^*\right| \\ &= n^{-1} O_p^*(n) O_p^*(n^{-1}) + n^{-1} O_p^*(n^{1/2} p^{1/2}) O_p^*(n^{-1}) O_p^*(n p^{1/2}) O_p^*(n^{-1}) = O_p^*(n^{-1}). \end{aligned}$$

Finally we look at E_n^*:

$$E_n^* = \hat{\delta}^{*\prime} Z_{-1}^{*\prime} Z_{-1}^* \hat{\delta}^* - \hat{\delta}^{*\prime} Z_{-1}^{*\prime} W_p^* (W_p^{*\prime} W_p^*)^{-1} W_p^{*\prime} Z_{-1}^* \hat{\delta}^*.$$

As before, we use the results from Lemma 6 and $\hat{\delta}^* = O_p^*(n^{-1})$ to obtain

$$\begin{aligned} |E_n^*| &\le n^{-1} \left|\hat{\delta}^{*\prime}\right| \left|\left|Z_{-1}^{*\prime} Z_{-1}^*\right|\right| \left|\hat{\delta}^*\right| + n^{-1} \left|\hat{\delta}^{*\prime}\right| \left|Z_{-1}^{*\prime} W_p^*\right| \left|\left|(W_p^{*\prime} W_p^*)^{-1}\right|\right| \left|W_p^{*\prime} Z_{-1}^*\right| \left|\hat{\delta}^*\right| \\ &= n^{-1} O_p^*(n^{-1}) O_p^*(n^2) O_p^*(n^{-1}) + n^{-1} O_p^*(n^{-1}) O_p^*(n p^{1/2}) O_p^*(n^{-1}) O_p^*(n p^{1/2}) O_p^*(n^{-1}) \\ &= O_p^*(n^{-1}). \end{aligned}$$

Therefore, we have that

$$\hat{\omega}^{*2} = \frac{1}{n} \sum_{t=1}^{n} \xi_{p,t}^{*2} + o_p^*(1).$$

Next we wish to show that $\frac{1}{n}\sum_{t=1}^n \xi_{p,t}^{*2} = \frac{1}{n}\sum_{t=1}^n \xi_t^{*2} + o_p^*(1)$, for which our proof is similar as Chang and Park (2002, Proof of Lemma 3.1(c)). Note that

$$\begin{aligned} \frac{1}{n}\sum_{t=1}^{n} \left(\xi_{p,t}^* - \xi_t^*\right)^2 &= \frac{1}{n}\sum_{t=1}^{n} \left(\sum_{j=p+1}^{\infty} \hat{\Psi}_{p,j} \varepsilon_{t-j}^*\right)^2 = \sum_{j=p+1}^{\infty} \sum_{i=p+1}^{\infty} \hat{\Psi}_{p,j} \left(\frac{1}{n}\sum_{t=1}^{n} \varepsilon_{t-j}^* \varepsilon_{t-i}^{*\prime}\right) \hat{\Psi}_{p,i}^{\prime} \\ &= \sum_{j=p+1}^{\infty} \hat{\Psi}_{p,j} \Sigma^* \hat{\Psi}_j^{\prime} + \sum_{j=p+1}^{\infty} \sum_{i=p+1}^{\infty} \left(\hat{\Psi}_{p,j} \left(\frac{1}{n}\sum_{t=1}^{n} \varepsilon_{t-j}^* \varepsilon_{t-i}^{*\prime}\right) \hat{\Psi}_{p,i}^{\prime} - \delta_{ij}\Sigma\right) \\ &= \sum_{j=p+1}^{\infty} \left|\hat{\Psi}_{p,j}\right|^2 O_p^*(1) + \sum_{j=p+1}^{\infty} \sum_{i=p+1}^{\infty} \left|\hat{\Psi}_{p,j}\right| \left|\hat{\Psi}_{p,i}\right| O_p^*(n^{-1/2}) = o_p^*(1). \end{aligned}$$

Then, as

$$\left| \left(\frac{1}{n}\sum_{t=1}^{n} \xi_{p,t}^{*2}\right)^{1/2} - \left(\frac{1}{n}\sum_{t=1}^{n} \xi_t^{*2}\right)^{1/2} \right| \le \left(\frac{1}{n}\sum_{t=1}^{n} \left(\xi_{p,t}^* - \xi_t^*\right)^2\right)^{1/2}$$

as a consequence from the triangle inequality it follows that

$$\frac{1}{n}\sum_{t=1}^{n}\xi_{p,t}^{*2} = \frac{1}{n}\sum_{t=1}^{n}\xi_t^{*2} + o_p^*(1)$$

which concludes this step.

For the final step we show that

$$\left|\frac{1}{n}\sum_{t=1}^{n}\xi_t^{*2} - \omega^2\right| \leq \left|\frac{1}{n}\sum_{t=1}^{n}\xi_t^{*2} - \omega^{*2}\right| + \left|\omega^{*2} - \omega^2\right| = o_p^*(1),$$

where $\omega^{*2} = \mathrm{E}^*(\xi_t^{*2})$. First, we show that $\left|\frac{1}{n}\sum_{t=1}^{n}\xi_t^{*2} - \omega^{*2}\right| = o_p(1)$. Note that

$$\begin{aligned}
\mathrm{P}^*\left(\left|\frac{1}{n}\sum_{t=1}^{n}\xi_t^{*2} - \omega^{*2}\right| > \epsilon\right) &\leq \epsilon^{-2}\,\mathrm{E}^*\left(\frac{1}{n}\sum_{t=1}^{n}\xi_t^{*2} - \omega^{*2}\right)^2 \\
&= \epsilon^{-2}\,\mathrm{E}^*\left(\frac{1}{n}\sum_{t=1}^{n}\left(\xi_t^{*2} - \mathrm{E}^*(\xi_t^{*2})\right)\right)^2 \\
&= \epsilon^{-2}\frac{1}{n^2}\sum_{s=1}^{n}\sum_{t=1}^{n}\mathrm{E}^*\left(\xi_s^{*2} - \mathrm{E}^*(\xi_s^{*2})\right)\left(\xi_t^{*2} - \mathrm{E}^*(\xi_t^{*2})\right) \\
&= \epsilon^{-2}\frac{1}{n^2}\sum_{s=1}^{n}\sum_{t=1}^{n}\left(\mathrm{E}^*\left(\xi_s^{*2}\xi_t^{*2}\right) - \mathrm{E}^*(\xi_s^{*2}) - \mathrm{E}^*(\xi_t^{*2})\right) \\
&= \epsilon^{-2}\frac{1}{n^2}\sum_{t=1}^{n}\left(\mathrm{E}^*(\xi_t^{*4}) - \left(\mathrm{E}^*(\xi_t^{*2})\right)^2\right) = O_p(n^{-1}).
\end{aligned}$$

Next, we next show that $\omega^{*2} \xrightarrow{p} \omega^2$. As

$$\omega^{*2} = \sigma_{11}^* - \Sigma_{12}^*\Sigma_{22}^{*-1}\Sigma_{21}^* \quad \text{and} \quad \omega^2 = \sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21},$$

and $\Sigma^* \xrightarrow{p} \Sigma$ by Lemma 3, the result follows. This completes the proof. □

Proof of Theorem 3. Note that

$$T_{\text{wald}}^* = n\hat{\delta}^{*\prime}\left(n^2\widehat{\mathrm{Var}^*(\hat{\delta}^*)}\right)^{-1} n\hat{\delta}^*.$$

We first look at $\hat{\delta}^*$. We can write $n\hat{\delta}^*$ as

$$\begin{aligned}
n\hat{\delta}^* &= n[Z_{-1}^{*\prime}(I - W_p^*(W_p^{*\prime}W_p^*)^{-1}W_p^{*\prime})Z_{-1}^*]^{-1}Z_{-1}^{*\prime}(I - W_p^*(W_p^{*\prime}W_p^*)^{-1}W_p^{*\prime})\Xi_p^* \\
&= [n^{-2}Z_{-1}^{*\prime}Z_{-1}^* - A_n^*]^{-1}(n^{-1}Z_{-1}^{*\prime}\Xi_p^* - B_n^*),
\end{aligned}$$

where

$$A_n^* = n^{-2} Z_{-1}^* W_p^* (W_p^{*\prime} W_p^*)^{-1} W_p^{*\prime} Z_{-1}^* \quad \text{and} \quad B_n^* = n^{-1} Z_{-1}^* W_p^* (W_p^{*\prime} W_p^*)^{-1} W_p^{*\prime} \Xi_p^*.$$

Using Lemma 6 c), d) and e), we have

$$\begin{aligned} |A_n^*| &\leq n^{-3} \left| Z_{-1}^* W_p^* \right| \left|\left| (n^{-1} W_p^{*\prime} W_p^*)^{-1} \right|\right| \left| W_p^{*\prime} Z_{-1}^* \right| \\ &= n^{-3} O_p^*(np^{1/2}) O_p^*(1) O_p^*(np^{1/2}) = O_p^*(n^{-1}p) \end{aligned}$$

and

$$\begin{aligned} |B_n^*| &\leq n^{-2} \left| Z_{-1}^* W_p^* \right| \left|\left| (n^{-1} W_p^{*\prime} W_p^*)^{-1} \right|\right| \left| W_p^{*\prime} \Xi_p^* \right| \\ &= n^{-2} O_p^*(np^{1/2}) O_p^*(1) O_p^*(n^{1/2}p^{1/2}) = O_p^*(n^{-1/2}p). \end{aligned}$$

Hence, as $p = o(n^{1/2})$ (Assumption 3), we have that

$$A_n^* = o_p^*(1) \quad \text{and} \quad B_n^* = o_p^*(1).$$

Then by Lemma 6 a) and b), we have

$$\begin{aligned} n\hat{\delta}^* &= (n^{-2} Z_{-1}' Z_{-1})^{-1} n^{-1} Z_{-1}' \Xi_p + o_p^*(1) \\ &\xrightarrow{d^*} \left[\int_0^1 B(r) B(r)' dr \right]^{-1} \int_0^1 B(r) dB_\xi(r). \end{aligned} \tag{45}$$

The estimated variance of $\hat{\delta}^*$, is defined as

$$\widehat{\text{Var}^*(\hat{\delta}^*)} = \hat{\omega}^{*2} [Z_{-1}^{*\prime} (I - W_p^* (W_p^{*\prime} W_p^*)^{-1} W_p^{*\prime}) Z_{-1}^*]^{-1}.$$

Using Lemma 6 and Lemma 7, we have that

$$n^2 \widehat{\text{Var}^*(\hat{\delta}^*)} \xrightarrow{d^*} \omega^2 \left[\int_0^1 B(r) B(r)' dr \right]^{-1}. \tag{46}$$

Finally, using equations (45) and (46) we can derive that

$$
\begin{aligned}
T^*_{\text{wald}} \xrightarrow{d^*} & \left(\int_0^1 dB_\xi(r)B(r)' \left[\int_0^1 B(r)B(r)'dr \right]^{-1} \right) \left(\omega^2 \left[\int_0^1 B(r)B(r)'dr \right]^{-1} \right)^{-1} \\
& \times \left(\left[\int_0^1 B(r)B(r)'dr \right]^{-1} \int_0^1 B(r)dB_\xi(r) \right) \\
= & \, \omega^{-2} \int_0^1 dB_\xi(r)B(r)' \left[\int_0^1 B(r)B(r)'dr \right]^{-1} \int_0^1 B(r)dB_\xi(r) \\
= & \, \omega^{-2} \int_0^1 dW_1(r)\omega W(r)'L'\Psi(1)' \left[\int_0^1 \Psi(1)LW(r)W(r)'L'\Psi(1)'dr \right]^{-1} \\
& \times \int_0^1 \Psi(1)LW(r)dW_1(r)\omega \\
= & \int_0^1 dW_1(r)W(r)' \left[\int_0^1 W(r)W(r)'dr \right]^{-1} \int_0^1 W(r)dW_1(r)
\end{aligned}
$$

as $B(r) = \Psi(1)LW(r)$ and $B_\xi(r) = \omega W_1(r)$. This completes the proof. □

References

Ahlgren, N. J. C. (2000). Bootstrapping the error correction model cointegration test. Working paper 428, Swedish School of Economics and Business Administration.

Banerjee, A., J. J. Dolado, and R. Mestre (1998). Error-correction mechanism tests for cointegration in a single-equation framework. *Journal of Time Series Analysis 19*, 267–283.

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

Berger, E. (1991). Majorization, exponential inequalities and almost sure behavior of vector-valued random variables. *Annals of Probability 19*, 1206–1226.

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

Boswijk, H. P. (1994). Testing for an unstable root in conditional and structural error correction models. *Journal of Econometrics 63*, 37–60.

Boswijk, H. P. and P. H. Franses (1992). Dynamic specification and cointegration. *Oxford Bulletin of Economics and Statistics 54*, 369–381.

Bühlmann, P. (1995). Moving-average representations of autoregressive approximations. *Stochastic Processes and their Applications 60*, 331–342.

Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli 3*, 123–148.

Chang, Y. 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.

Chang, Y., J. Y. Park, and K. Song (2006). Bootstrapping cointegrating regressions. *Journal of Econometrics 133*, 703–739.

Einmahl, U. (1987). A useful estimate in the multidimensional invariance principle. *Probability Theory and Related Fields 76*, 81–101.

Engle, R. F. and C. W. J. Granger (1987). Co-integration and error correction: representation, estimation and testing. *Econometrica 55*, 251–276.

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

Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models.* Oxford: Oxford University Press.

Johansen, S. and K. Juselius (1990). Maximum likelihood estimation and inference on cointegration - with applications to the demand for money. *Oxford Bulletin of Economics and Statistics 54*, 169–210.

Kremers, J. J. M., N. R. Ericsson, and J. J. Dolado (1992). The power of cointegration tests. *Oxford Bulletin of Economics and Statistics 52*, 325–348.

Mantalos, P. and G. Shukur (1998). Size and power of the error correction model cointegration test. A bootstrap approach. *Oxford Bulletin of Economics and Statistics 60*, 249–255.

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.

Paparoditis, E. (1996). Bootstrapping autoregressive and moving average parameter estimates of infinite order vector autoregressive processes. *Journal of Multivariate Analysis 57*, 277–296.

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.

Park, J. Y. and P. C. B. Phillips (1989). Statistical inference in regressions with integrated processes: part 2. *Econometric Theory 5*, 95–131.

Pesavento, E. (2004). Analytical evaluation of the power of tests for the absence of cointegration. *Journal of Econometrics 122*, 349–384.

Phillips, P. C. B. (1988). Weak convergence to the matrix stochastic integral $\int_0^1 BdB'$. *Journal of Multivariate Analysis 24*, 252–264.

Phillips, P. C. B. and S. Ouliaris (1990). Asymptotic properties of residual based tests for cointegration. *Econometrica 58*, 165–193.

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.

Seo, M. (2006). Bootstrap testing for the null of no cointegration in a threshold vector error correction model. *Journal of Econometrics 134*, 129–150.

Swensen, A. R. (2006). Bootstrap algorithms for testing and determining the cointegration rank in VAR models. *Econometrica 74*, 1699–1714.

Trenkler, C. (2006). Bootstrapping systems cointegration tests with a prior adjustment for deterministic terms. SFB 649 Discussion Paper 2006-012, Humboldt Universität zu Berlin.

Van Giersbergen, N. P. A. and J. F. Kiviet (1996). Bootstrapping a stable AD model: weak vs strong exogeneity. *Oxford Bulletin of Economics and Statistics 58*, 631–656.

Zivot, E. (2000). The power of single equation tests for cointegration when the cointegrating vector is prespecified. *Econometric Theory 16*, 407–439.

Table 1: Parameter combinations used in the simulation DGP

Φ	Θ	r	c
$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	$0, \sqrt{0.3}, \sqrt{0.7}$	0,-5,-10,-20
$\begin{bmatrix} 0.2 & 0 \\ 0 & 0.2 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	$\sqrt{0.3}$	0
$\begin{bmatrix} 0.8 & 0 \\ 0 & 0.8 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	$\sqrt{0.3}$	0
$\begin{bmatrix} 0.2 & 0.5 \\ 0.5 & 0.2 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	$\sqrt{0.3}$	0,-5,-10,-20
$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.2 & 0 \\ 0 & 0.2 \end{bmatrix}$	$\sqrt{0.3}$	0
$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.8 & 0 \\ 0 & 0.8 \end{bmatrix}$	$\sqrt{0.3}$	0
$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.2 & 0.5 \\ 0.5 & 0.2 \end{bmatrix}$	$\sqrt{0.3}$	0,-5,-10,-20
$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} -0.8 & 0 \\ 0 & -0.8 \end{bmatrix}$	$\sqrt{0.3}$	0

Table 2: Size and power for white noise errors

r	c	$T^*_{v,n}$	$T^*_{v,a}$	$T^*_{c,n}$	$T^*_{c,a}$	T_{as}	T_{sc}
				$n=50$			
0	0	0.042	0.042	0.043	0.040	0.072	0.050
	-5	0.102	0.109	0.105	0.103	0.163	0.121
	-10	0.281	0.283	0.284	0.283	0.414	0.322
	-20	0.831	0.839	0.829	0.835	0.918	0.866
$\sqrt{0.3}$	0	0.052	0.050	0.052	0.051	0.085	0.050
	-5	0.163	0.159	0.160	0.162	0.243	0.163
	-10	0.521	0.523	0.522	0.522	0.661	0.524
	-20	0.949	0.961	0.950	0.960	0.984	0.964
$\sqrt{0.7}$	0	0.052	0.051	0.051	0.052	0.079	0.050
	-5	0.501	0.505	0.498	0.503	0.618	0.488
	-10	0.898	0.901	0.898	0.903	0.935	0.898
	-20	0.983	0.996	0.985	0.997	0.999	0.998
				$n=100$			
0	0	0.048	0.048	0.048	0.049	0.059	0.050
	-5	0.108	0.107	0.108	0.109	0.133	0.113
	-10	0.317	0.320	0.315	0.314	0.381	0.334
	-20	0.868	0.859	0.865	0.861	0.906	0.875
$\sqrt{0.3}$	0	0.061	0.057	0.061	0.059	0.072	0.050
	-5	0.180	0.187	0.181	0.185	0.225	0.160
	-10	0.541	0.545	0.535	0.541	0.601	0.501
	-20	0.960	0.961	0.963	0.963	0.978	0.952
$\sqrt{0.7}$	0	0.056	0.056	0.053	0.058	0.071	0.050
	-5	0.545	0.543	0.538	0.539	0.597	0.524
	-10	0.936	0.936	0.938	0.933	0.949	0.933
	-20	0.999	1.000	1.000	1.000	1.000	1.000

Table 3: Size for serially correlated errors

Φ	Θ	$T^*_{v,n}$	$T^*_{v,a}$	$T^*_{c,n}$	$T^*_{c,a}$	T_{as}
		$n = 50$				
$\begin{bmatrix} 0.2 & 0 \\ 0 & 0.2 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	0.060	0.057	0.059	0.058	0.095
$\begin{bmatrix} 0.8 & 0 \\ 0 & 0.8 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	0.049	0.045	0.045	0.046	0.158
$\begin{bmatrix} 0.2 & 0.5 \\ 0.5 & 0.2 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	0.059	0.088	0.055	0.088	0.214
$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.2 & 0 \\ 0 & 0.2 \end{bmatrix}$	0.063	0.061	0.065	0.058	0.103
$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.8 & 0 \\ 0 & 0.8 \end{bmatrix}$	0.050	0.055	0.050	0.055	0.186
$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.2 & 0.5 \\ 0.5 & 0.2 \end{bmatrix}$	0.075	0.092	0.075	0.095	0.188
$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} -0.8 & 0 \\ 0 & -0.8 \end{bmatrix}$	0.459	0.625	0.453	0.625	0.677
		$n = 100$				
$\begin{bmatrix} 0.2 & 0 \\ 0 & 0.2 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	0.063	0.065	0.059	0.064	0.084
$\begin{bmatrix} 0.8 & 0 \\ 0 & 0.8 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	0.051	0.047	0.050	0.049	0.091
$\begin{bmatrix} 0.2 & 0.5 \\ 0.5 & 0.2 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	0.056	0.056	0.056	0.052	0.107
$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.2 & 0 \\ 0 & 0.2 \end{bmatrix}$	0.059	0.059	0.059	0.058	0.080
$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.8 & 0 \\ 0 & 0.8 \end{bmatrix}$	0.057	0.067	0.058	0.066	0.120
$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.2 & 0.5 \\ 0.5 & 0.2 \end{bmatrix}$	0.070	0.075	0.067	0.078	0.140
$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} -0.8 & 0 \\ 0 & -0.8 \end{bmatrix}$	0.485	0.518	0.481	0.525	0.611

Table 4: Power for serially correlated errors

dynamics	c	$T^*_{v,n}$	$T^*_{v,a}$	$T^*_{c,n}$	$T^*_{c,a}$	T_{as}	T_{sc}
				$n = 50$			
	-5	0.624	0.538	0.625	0.539	0.819	0.485
$\Phi = \begin{bmatrix} 0.2 & 0.5 \\ 0.5 & 0.2 \end{bmatrix}$	-10	0.857	0.844	0.856	0.845	0.964	0.735
	-20	0.929	0.978	0.930	0.978	0.996	0.899
	-5	0.401	0.301	0.411	0.298	0.604	0.259
$\Theta = \begin{bmatrix} 0.2 & 0.5 \\ 0.5 & 0.2 \end{bmatrix}$	-10	0.810	0.749	0.818	0.753	0.939	0.650
	-20	0.940	0.968	0.942	0.969	0.994	0.957
				$n = 100$			
	-5	0.893	0.895	0.898	0.896	0.948	0.891
$\Phi = \begin{bmatrix} 0.2 & 0.5 \\ 0.5 & 0.2 \end{bmatrix}$	-10	0.993	0.989	0.993	0.988	0.996	0.987
	-20	0.999	1.000	0.999	1.000	1.000	1.000
	-5	0.493	0.470	0.491	0.473	0.681	0.416
$\Theta = \begin{bmatrix} 0.2 & 0.5 \\ 0.5 & 0.2 \end{bmatrix}$	-10	0.907	0.889	0.905	0.884	0.973	0.851
	-20	0.998	0.998	0.996	0.996	1.000	0.996

Table 5: Size and power for tests with deterministic trends

μ_1 τ_1 / μ_2 τ_2	$D_t^{(r)}$	$T^*_{v,n}$	$T^*_{v,a}$	T_{as}	$T^*_{v,n}$	$T^*_{v,a}$	T_{as}	$T^*_{v,n}$	$T^*_{v,a}$	T_{as}
		$\Phi = \Theta = 0$			$\Phi = \begin{bmatrix} 0.2 & 0.5 \\ 0.5 & 0.2 \end{bmatrix}$			$\Theta = \begin{bmatrix} 0.2 & 0.5 \\ 0.5 & 0.2 \end{bmatrix}$		
					$c = 0$					
	$D_t^r = 1$	0.051	0.051	0.106	0.050	0.104	0.291	0.059	0.095	0.253
0 0	$D_t = 1$	0.052	0.048	0.093	0.039	0.102	0.244	0.073	0.121	0.254
0 0	$D_t^r = t$	0.046	0.043	0.133	0.050	0.135	0.372	0.064	0.105	0.342
	$D_t' = 1, t$	0.046	0.046	0.132	0.036	0.126	0.327	0.055	0.117	0.331
	$D_t^r = 1$	0.046	0.049	0.102	0.047	0.108	0.282	0.061	0.093	0.239
1 0	$D_t = 1$	0.045	0.045	0.086	0.042	0.099	0.241	0.068	0.096	0.221
1 0	$D_t^r = t$	0.059	0.051	0.134	0.042	0.131	0.373	0.063	0.107	0.358
	$D_t' = 1, t$	0.035	0.034	0.114	0.031	0.114	0.346	0.075	0.130	0.344
1 0	$D_t^r = t$	0.051	0.053	0.154	0.034	0.116	0.358	0.057	0.101	0.339
1 1	$D_t' = 1, t$	0.050	0.053	0.125	0.029	0.121	0.333	0.065	0.126	0.334
1 1	$D_t^r = t$	0.048	0.046	0.128	0.043	0.122	0.379	0.063	0.108	0.350
1 1	$D_t' = 1, t$	0.044	0.041	0.131	0.037	0.117	0.320	0.065	0.118	0.313
					$c = -10$					
	$D_t^r = 1$	0.244	0.249	0.420	0.667	0.612	0.902	0.552	0.436	0.796
0 0	$D_t = 1$	0.281	0.290	0.450	0.672	0.626	0.889	0.569	0.470	0.801
0 0	$D_t^r = t$	0.166	0.159	0.387	0.514	0.356	0.859	0.363	0.191	0.719
	$D_t' = 1, t$	0.173	0.170	0.384	0.497	0.344	0.801	0.360	0.209	0.707
	$D_t^r = 1$	0.247	0.245	0.428	0.682	0.624	0.906	0.548	0.452	0.818
1 0	$D_t = 1$	0.289	0.288	0.464	0.679	0.638	0.886	0.563	0.458	0.800
1 0	$D_t^r = t$	0.131	0.132	0.343	0.499	0.352	0.846	0.367	0.211	0.725
	$D_t' = 1, t$	0.158	0.157	0.365	0.477	0.322	0.784	0.374	0.215	0.711
1 0	$D_t^r = t$	0.133	0.140	0.355	0.508	0.334	0.855	0.356	0.191	0.707
1 1	$D_t' = 1, t$	0.166	0.159	0.391	0.484	0.332	0.819	0.369	0.221	0.717
1 1	$D_t^r = t$	0.156	0.153	0.370	0.501	0.327	0.843	0.364	0.201	0.728
1 1	$D_t' = 1, t$	0.162	0.170	0.383	0.481	0.318	0.809	0.366	0.213	0.700