Unit Root Inference for Non-Stationary Linear Processes driven by Infinite Variance Innovations*

Giuseppe Cavaliere[a], Iliyan Georgiev[a] and A.M.Robert Taylor[b]
[a]Department of Statistical Sciences, University of Bologna
[b]Essex Business School, University of Essex

February 15, 2016

Abstract

The contribution of this paper is two-fold. First, we derive the asymptotic null distribution of the familiar augmented Dickey-Fuller [ADF] statistics in the case where the shocks follow a linear process driven by infinite variance innovations. We show that these distributions are free of serial correlation nuisance parameters but depend on the tail index of the infinite variance process. These distributions are shown to coincide with the corresponding results for the case where the shocks follow a finite autoregression, provided the lag length in the ADF regression satisfies the same $o(T^{1/3})$ rate condition as is required in the finite variance case. In addition, we establish the rates of consistency and (where they exist) the asymptotic distributions of the ordinary least squares sieve estimates from the ADF regression. Given the dependence of their null distributions on the unknown tail index, our second contribution is to explore sieve wild bootstrap implementations of the ADF tests. Under the assumption of symmetry, we demonstrate the asymptotic validity (bootstrap consistency) of the wild bootstrap ADF tests. This is done by establishing that (conditional on the data) the wild bootstrap ADF statistics attain the same limiting distribution as that of the original ADF statistics taken conditional on the magnitude of the innovations.

Keywords: Unit root; infinite variance; linear process; sieve estimator; wild bootstrap.

J.E.L. Classifications: C12, C14, C22.

1 Introduction

Practitioners routinely report the outcomes of unit root tests applied to macroeconomic and financial data. These tests are invariably constructed with reference to published asymptotic critical values which have been obtained under the assumption that the innovations driving the time series process display finite variance. However, it is well known that many macroeconomic and financial series appear to violate this assumption having heavy-tailed distributions; see, *inter alia*, Adler *et al.* (1998) and Embrechts *et al.* (1997).

*Cavaliere and Georgiev gratefully acknowledge financial support provided by the Fundação para a Ciência e a Tecnologia, Portugal, through grant PTDC/EGE-ECO/108620/2008. Cavaliere thanks the Italian Ministry of Education, University and Research (MIUR) for financial support (PRIN project: "Multivariate statistical models for risk assessment"). Cavaliere and Taylor would also like to thank the Danish Council for Independent Research, Sapere Aude | DFF Advanced Grant (Grant nr: 12-124980) for financial support. Correspondence to: Robert Taylor, Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK. Email: rtaylor@essex.ac.uk

The asymptotic null distributions of conventional regression-based unit root statistics, such as the ordinary least squares (OLS) based statistics of Dickey and Fuller (1979), Said and Dickey (1987) and the semi-parametric statistic of Phillips (1987), differ from the finite variance case when the innovations lie in the domain of attraction of a stable law, such that they have infinite variance [InfV]; see, in particular, Chan and Tran (1989), Phillips (1990), Samarakoon and Knight (2009), Rachev *et al.* (1998) and Caner (1998). Indeed, in such cases these limiting null distributions are no longer pivotal, depending on the so-called tail index of the stable law and on the relative weights of the left and right tails of the stable distribution. Comparing these statistics to critical values tabulated for the finite variance case therefore leads to incorrectly sized unit root tests. Rachev and Mittnik (2000) tabulate critical values from these limiting distributions for various values of the tail index and these could be used for the case where the innovations are symmetric as approximate critical values based on estimates of the tail index, as might be obtained using the method of McCulloch (1986), as is done for inflation data in Charemza *et al.* (2005). However, it is known that it is very difficult to estimate the tail index well; see, for example, Resnick (1997).

Where the innovations display InfV the properties of unit root tests are strongly affected by the way in which inference procedures treat extreme observations. In particular, Knight (1989,1991) and Samarakoon and Knight (2009) show that robust estimation methods based on the down-weighting of large errors through appropriate M-estimators can result in order of magnitude gains in efficiency compared to standard OLS estimation in cases where the innovations display InfV. The same phenomenon occurs if large outliers are dummied out in an iterated OLS regression, as is not uncommon in applied work; see Cavaliere and Georgiev (2013). Knight (1989,1991) and Samarakoon and Knight (2009) develop unit root tests based on M estimation (some of these papers additionally propose unit root tests based on least absolute deviation estimation) and demonstrate that these have Gaussian limiting null distributions (which do not depend on the tail index).

The use of these robust estimators remains uncommon in the empirical analysis of economic and financial time series, however, possibly because applied researchers typically do not want to commit to the assumption of infinite variance innovations, but rather view this as a possibility they would ideally like their results to be robust against. Moreover, the M estimators are dominated by OLS in the finite variance case; see, for example, Maronna *et al.* (2006,p.269) for a comparison of M and OLS estimators. Related to this, while simulations presented in Moreno and Romo (2012) suggest that under InfV the M-based unit root tests can display significant finite sample power gains over the OLS-based tests, these gains remain relatively small when the tail index is close to two, whereas under finite variance the situation tends to be reversed with OLS-based tests displaying greater power. In practice, therefore, OLS remains an attractive estimation method for applied workers in economics and finance and so for the purposes of this paper we restrict our attention to OLS-based unit root tests with the goal to develop implementations of these which are robust to InfV.

To that end, the contribution of this paper is two-fold. Our first contribution is to extend upon the work of Chan and Tran (1989) and Samarakoon and Knight (2009) to derive the limiting null distribution of the usual OLS-based augmented Dickey-Fuller [ADF] test in the case where the shocks follow a linear process which is driven by InfV innovations. We show that, provided the lag length used in the ADF regression satisfies the standard $o(T^{1/3})$ rate condition, these distributions are free of serial correlation nuisance parameters and coincide with those given in Chan and Tran (1989) and Samarakoon and Knight (2009) who assume a first-order and finite-order autoregressive process respectively. The required rate condition therefore coincides with that required for analogous results to hold in the case where the innovations have finite variance; cf. Chang and Park (2002). We also establish the rates of consistency and (where they exist) the asymptotic distributions of the OLS sieve estimates from the ADF regression, in each case again establishing the rate conditions required on the lag length for these results to hold. These rates are shown to depend on the tail index.

The second contribution of this paper is to explore sieve wild bootstrap implementations of the ADF tests. While the formulation of the OLS-based ADF statistics gives no special treatment to large innovations, these could still be taken into account through the choice of reference population with respect to which the statistics are compared in forming a test and bootstrap methods can be used to achieve this. It should be stressed that extant bootstrap methods proposed in this literature do not do this. Zarepour and Knight (1999) and Moreno and Romo (2012) establish the asymptotic validity of the m-out-of-n bootstrap with respect to the *unconditional* asymptotic null distribution of the M-based Dickey-Fuller statistic for a first-order autoregression driven by InfV errors. These authors also demonstrate the invalidity of i.i.d. bootstrap unit root tests when applied to InfV data. For OLS-based Dickey-Fuller statistics, again in the context of a first-order autoregression driven by InfV innovations, Horvath and Kokoszka (2003) and Jach and Kokoszka (2004) do the same for the m-out-of-n bootstrap and for subsampling inference, respectively. These approaches are also valid under finite variance and do not require estimation of the tail index. However, the simulation evidence provided by their authors suggests that they can be rather over-sized in finite samples.

The wild bootstrap approach we examine is one based on approximating the limiting null distributions of the ADF statistics conditional on the absolute sizes of the innovations. Doing so restricts the reference population and hence should be expected to gain precision. The rationale behind this is that the restricted reference population corresponds to outcomes whose information content is compatible with that of the sample. Lepage and Podgórski (1996) propose this idea and implement it by means of a permutations bootstrap in the context of fixed-design regressions, though their formal analysis is incomplete; Cavaliere *et al.* (2013) choose a wild bootstrap implementation and complete the analysis in the representative case of inference on the location parameter of an i.i.d. sample. A similar approach is applied by Aue *et al.* (2008) to the distribution of the CUSUM statistic. As with the bootstrap approximations to the unconditional limiting distribution of the ADF statistics mentioned above, the wild bootstrap ADF tests we discuss are known to be valid in the finite-variance case, though without the conditioning interpretation of their null limits; see, for example, Cavaliere and Taylor (2008,2009).

Under the assumption that the innovations are symmetrically distributed, we demonstrate the asymptotic validity (bootstrap consistency) of the sieve wild bootstrap ADF tests by showing that the wild bootstrap ADF statistics attain the same limiting distribution as that of the original ADF statistics when taken conditional on the magnitude of the innovations. We show that for this result to hold a Rademacher distribution (*i.e.*, a two point, symmetric distribution) needs to be employed in the wild bootstrap re-sampling scheme. Importantly, our results are established under a general linear process framework rather than the first-order autoregressive model considered in the previous bootstrap approaches outlined above. We establish that valid wild bootstrap ADF tests can be formed using either the unrestricted sieve estimates from the ADF regression or the sieve estimates obtained from a sieve regression which imposes the unit root null hypothesis, as is done in the finite variance case in the context of i.i.d. bootstrap ADF tests in Chang and Park (2003).

It should be stressed that our conditioning proposal does not require the practitioner to do anything different from what is done in the finite variance case; that is, the wild bootstrap

ADF tests we discuss are constructed in exactly the same way as outlined for the finite variance case in Cavaliere and Taylor (2008,2009). Conditioning is simply a useful concept in understanding the meaning of an otherwise standard wild bootstrap inference procedure when the variance happens to be infinite. The bootstrap procedure can therefore be successfully applied in ignorance of whether the innovations display finite or infinite variance, and lies fully in the regression framework.

The plan of the paper is as follows. In section 2 we detail our reference InfV linear process data generating process (DGP) together with the usual ADF tests obtained from a sieve approximation to this DGP. In section 3 we derive the asymptotic null distribution of the ADF test statistics, thereby extending the results of Said and Dickey (1987) and Chang and Park (2002) to the InfV case. We do this assuming that the order of the sieve approximation increases with the sample size at the same rate as is required under finite variance. The asymptotic properties of the (appropriately normalised) sieve estimators from the ADF regression are also established and related back to the extant results in the literature. In section 4 we outline how the sieve wild bootstrap principle can be applied to the ADF testing problem. Here we also develop the necessary asymptotic theory to demonstrate the validity of the sieve wild bootstrap ADF tests. The finite sample properties of these tests are explored in section 5 and compared with those of standard ADF tests together with m-out-of-n and sub-sampling implementations of the ADF tests. Extensions to allow for deterministic components are briefly discussed in section 6. Some conclusions and directions for further research are offered in section 7. Proofs are contained in the Appendix.

In the following we use P^*, E^* and Var^* respectively to denote probability, expectation and variance, conditional on the original sample. We denote weak convergence and convergence in probability by $\overset{w}{\to}$ and $\overset{P}{\to}$, respectively. The Euclidean norm of the vector x is $\|x\| := (x'x)^{1/2}$, where $x := y$ indicates that x is defined by y. Also, $\mathbb{I}(\cdot)$ denotes the indicator function; $\lfloor \cdot \rfloor$ denotes the integer part of its argument; $\mathbf{I}_k$ denotes the $k \times k$ identity matrix, $\mathbf{I}$ is the infinite identity matrix and $\mathbf{0}_{j\times k}$ the $j \times k$ matrix of zeroes; $\mathbf{i}_k \in \mathbb{R}^k$ denote a vector of ones and $\mathbf{i} \in \mathbb{R}^\infty$ an infinite column-sequence of ones. The space of càdlàg functions from the unit interval $[0, 1]$ to $\mathbb{R}^k$ is denoted by $D_k[0, 1]$ (or $D[0, 1]$ for $k = 1$) and is endowed with the Skorokhod topology.

2 The Model, Assumptions and ADF Tests

We consider the time series process $\{y_t\}$ generated according to the recursion

$$y_t = \rho y_{t-1} + u_t, \ (t \in \mathbb{N}) \tag{1}$$

where the initial value y_0 is assumed to be $O_P(1)$ and available to the practitioner, and where $\{u_t\}$ is a stationary and invertible linear process of potentially infinite order with innovations following an InfV distribution.

Our primary interest in this paper is on testing the usual unit root null hypothesis $\mathsf{H}_0 : \rho = 1$ in (1) against the stable root (stationary) alternative $\mathsf{H}_1 : |\rho| < 1$. The key feature of our setup is that the innovations driving the linear process u_t display *infinite variance.* More precisely, u_t is taken to satisfy the $MA(\infty)$ scheme

$$u_t = \sum_{i=0}^{\infty} \gamma_i \varepsilon_{t-i}, \quad (t \in \mathbb{Z}), \tag{2}$$

with the following set of assumptions, collectively labelled Assumption 1 in what follows, taken to hold.

Assumption 1

(i) The random variables ε_t *(*$t \in \mathbb{Z}$*) form an i.i.d. sequence which is in the domain of attraction of an* α*-stable law,* $\alpha \in (0,2)$*; that is, the tails of the distribution of* ε_t *exhibit the power law decay:*

$$P(|\varepsilon_t| > x) = x^{-\alpha} L(x) \;\; \text{for } x > 0$$

with $L(\cdot)$ *a slowly varying function at infinity, and* $\lim_{x\to\infty} P(\varepsilon_t > x)/P(|\varepsilon_t| > x) =: p \in [0,1]$*,* $\lim_{x\to\infty} P(\varepsilon_t < -x)/P(|\varepsilon_t| > x) = 1-p$*. Where* $\mathrm{E}|\varepsilon_1| < \infty$*, it is assumed that* $\mathrm{E}\,\varepsilon_1 = 0$*; that is, where the mean of* ε_1 *exists, it is assumed to be zero. Moreover, we assume that there exists a normalising sequence* a_T *such that* $a_T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor}\varepsilon_t \xrightarrow{w} \mathcal{S}(\cdot)$ *as random elements of* $D[0,1]$*, where* $\mathcal{S}$ *is an* α*-stable process (or* α*-stable motion).*

(ii) There exists a $\delta \in (0,\alpha) \cap [0,1]$ *such that* $\sum_{i=0}^{\infty} i|\gamma_i|^{\delta/2} < \infty$*.*

(iii) The power series $\gamma(z) := \sum_{i=0}^{\infty}\gamma_i z^i$*, where we set* $\gamma_0 = 1$ *with no loss of generality in what follows, has no roots on the closed complex unit disk.*

(iv) Its reciprocal $\sum_{i=0}^{\infty}\beta_i z^i := (\sum_{i=0}^{\infty}\gamma_i z^i)^{-1}$ *satisfies* $\sum_{i=0}^{\infty}|\beta_i|^{\delta} < \infty$ *where* δ *is as defined in part (ii).*

Some remarks are in order.

Remark 2.1. The parameter α in part (i) of Assumption 1, which will be treated as unknown in this paper, controls the thickness of the tails of the distribution of ε_t, and, as such, is often referred to as the tail index, index of stability or characteristic exponent; for further details see Chapter XVII of Feller (1971). The parameter p, also defined in part (i), is a measure of the relative heaviness of the two tails of the distribution of ε_t. Where ε_t has a symmetric distribution, $p = 0.5$. Moments $\mathrm{E}|\varepsilon_t|^r$ are finite for $r < \alpha$ and infinite for $r > \alpha$; the moment $\mathrm{E}|\varepsilon_t|^{\alpha}$ can be either finite or infinite, discriminating between some results in section 3. The tail index, α, is inherited by the limiting process $\mathcal{S}$. Heavy tailed data are widely encountered in financial, macroeconomic, actuarial, telecommunication network traffic, and meteorological time series; see, *inter alia*, Embrechts *et al* (1997), Finkenstädt and Rootzén (2003) and Davis (2010) for examples and references. Reported estimates of α include 1.85 for stock returns (McCulloch, 1997), above 1.5 for income, about 1.5 for wealth and trading volumes, about 1 for firm and city sizes (all in Gabaix, 2009, and references therein) and even below 1 for returns from technological innovations (Silverberg and Verspagen, 2007).

Remark 2.2. In part (i) of Assumption 1, choosing $a_T := \inf\{x : P(|\varepsilon_1| > x) \leq T^{-1}\}$, which implies the existence of a slowly varying sequence l_T such that $a_T = T^{1/\alpha} l_T$, ensures that the stated invariance principle holds without extra conditions for all $\alpha \in (0,1)$. It also holds for this choice of a_T when $\alpha \in (1,2)$ and $\mathrm{E}\,\varepsilon_1 = 0$. For $\alpha = 1$ a sufficient condition for it to hold, again for this choice of a_T, is that the distribution of ε_t is symmetric around zero; see Resnick and Greenwood (1979). In all the three cases the limiting process $\mathcal{S}(\cdot)$ induces on $D[0,1]$ the same measure as the pure jump process $\sum_{i=1}^{\infty}\delta_i\Gamma_i^{-1/\alpha}\mathbb{I}(\{U_i \leq (\cdot)\})$, where $\{\Gamma_i\}$ is the sequence of arrival times of a Poisson process of unit intensity, $\{\delta_i\}$ is an i.i.d. binomial sequence with $P(\delta_i = 1) = p = 1 - P(\delta_i = -1)$, $\{U_i\}$ is an i.i.d. sequence of uniform $[0,1]$ random variables, and the sequences $\{\Gamma_i\}$, $\{\delta_i\}$ and $\{U_i\}$ are jointly independent of one another; see LePage *et al.* (1997). The sequence of jump magnitudes $\{\Gamma_i^{-1/\alpha}\}$ is the weak limit in $\mathbb{R}^{\infty}$ of the order statistics of $\{|\varepsilon_t|_{t=1}^{T}\}$, $\{\delta_i\}$ is the weak limit of their respective signs, and $\{U_i\}$, of their relative location in the sample. Where the invariance principle holds with $a_T = aT^{1/\alpha}$, ε_t is said to be in the *normal* domain of attraction of a stable law.

Remark 2.3. Most of the unit root literature deals with cases where ε_t has finite variance. Under finite variance, the invariance principle in part (i) of Assumption 1 is satisfied with $a_T = aT^{1/2}$ and with $\mathcal{S}$ replaced by a standard Brownian motion (or 2-stable motion). In this case ε_t therefore belongs to the normal domain of attraction of the Gaussian distribution.

Remark 2.4. Part (ii) of Assumption 1 guarantees the almost sure convergence of the series $\sum_{i=0}^{\infty} \gamma_i \varepsilon_{t-i}$ (as well as certain series in ε_t^2) and underlies the asymptotics for sample correlations (see Davis and Resnick 1985b, p.270, and 1986, p.547). It also implies that $\sum_{i=1}^{\infty} i^{2/\delta}|\gamma_i| < \infty$ holds on the γ_i coefficients in (2). It can therefore be seen that part (ii) of Assumption 1 would also impose weak stationarity on u_t in the case where the mean and variance of ε_t were both finite and constant. Part (iii) ensures that the MA polynomial, $\gamma(z)$, is invertible, while part (iv) implies, among other things, that the infinite autoregressive series in (3) below converges absolutely with probability one. □

Under Assumption 1 the process u_t is strictly stationary and invertible and, equivalently, solves the (potentially) infinite order difference equation

$$u_t = \sum_{i=1}^{\infty} \beta_i u_{t-i} + \varepsilon_t,\ t \in \mathbb{Z}. \tag{3}$$

Notice that the coefficients in (3) satisfy $\sum_{i=1}^{\infty} i^{2/\delta}|\beta_i| < \infty$ due to the analogous property imposed on the $\{\gamma_i\}$ coefficients in part (ii) of Assumption 1; see Brillinger (2001, pp.76-77). Assumption 1 therefore allows us to re-write (1)-(2) as

$$\Delta y_t = \phi y_{t-1} + \sum_{i=1}^{\infty} \beta_i u_{t-i} + \varepsilon_t \tag{4}$$

where $\phi := \rho - 1$.

In order to obtain an operational unit root test, the ADF sieve-based regression of Said and Dickey (1984) is then formed by truncating the autoregression in (4) at a given order $k = k(T)$ and substituting u_{t-j} by Δy_{t-j}, $j = 1, ...k$, yielding

$$\Delta y_t = \phi y_{t-1} + \sum_{i=1}^{k} \beta_i \Delta y_{t-i} + \varepsilon_{t,k} \tag{5}$$

where $\varepsilon_{t,k} := \varepsilon_t + \rho_{t,k}$ with $\rho_{t,k} := \sum_{i=k+1}^{\infty} \beta_i u_{t-i}$. As will be formally established in section 3, in order to ensure that the error associated with the sieve approximation is asymptotically negligible, it is sufficient that the lag truncation k used in (5) increases with the sample size, though at a sufficiently slow rate such that consistent estimation of the sieve coefficients, β_i, $i = 1, ..., k$, remains feasible.

Dickey and Fuller (1979) and Said and Dickey (1984) proposed the use of the so-called normalised bias and t-ratio ADF unit root statistics, each based on OLS estimation of (5), for testing $\mathsf{H}_0 : \phi = 0$ against $\mathsf{H}_1 : \phi < 0$; see also Chang and Park (2002). To define the ADF statistics, first let $\hat{\phi}_k$ and $\hat{\boldsymbol{\beta}}_k$ denote the OLS estimators of ϕ and $\boldsymbol{\beta}_k := (\beta_1, ..., \beta_k)'$ respectively from (5), based on the sample data $y_0, ..., y_T$; that is estimating (5) over $t = k+1, ..., T$. The normalised bias ADF statistic is then given by

$$R_T := \frac{T_k \hat{\phi}_k}{\hat{\beta}(1)} \tag{6}$$

where $T_k := T - k$ and $\hat{\beta}(z) := 1 - \sum_{i=1}^{k} \hat{\beta}_i z^i$. The corresponding ADF regression t-statistic from (5) is given by

$$Q_T := \frac{\hat{\phi}_k}{s(\hat{\phi}_k)} \tag{7}$$

where $s(\hat{\phi}_k)$ denotes the usual OLS standard error of $\hat{\phi}_k$ from (5).

In each case, H_0 is rejected for large negative values of these statistics. Critical values from the limiting null distributions of these statistics, obtained for the case where ε_t is a finite variance process with finite fourth moments, are provided in Fuller (1996). Bootstrap implementations of these sieve-based ADF tests have been provided based on standard i.i.d. re-sampling in Chang and Park (2003) and on wild bootstrap re-sampling in Cavaliere and Taylor (2009). In both cases the validity of the bootstrap tests is demonstrated under the assumption that the variance of ε_t is finite (although in the case of the wild bootstrap ADF tests it can be time-varying) with finite fourth moments.

In what follows, we will show that the sieve wild bootstrap ADF unit root tests, implemented using a Rademacher distribution in the wild bootstrap re-sampling scheme, developed in Cavaliere and Taylor (2009) are also valid for symmetrically distributed innovations satisfying Assumption 1, so that neither the fourth order moment nor the variance are finite. As a result, the wild bootstrap ADF test can be validly applied regardless of whether the innovations have finite or infinite variance and, in the case of the latter, without knowledge of the tail index, α. The same is not true of the standard ADF tests based on the critical values from Fuller (1996) since, as we show in the next section, the null distribution of these under Assumption 1 differs from the finite variance case and, moreover, depends on the unknown tail index, α. The i.i.d. sieve bootstrap ADF tests of Chang and Park (2003) are invalid under InfV innovations; see Zarepour and Knight (1999), to which the reader is referred for further discussion on this point.

3 Asymptotic Results

In this section we establish the large sample properties of the OLS estimators $\hat{\phi}_k$ and $\hat{\boldsymbol{\beta}}_k$ from (5). We first establish the result that these estimators are consistent. We then derive the asymptotic null distributions of the sieve ADF unit root statistics R_T and Q_T from section 2, along with those (where they exist) for the sieve estimates, $\hat{\beta}_1, ..., \hat{\beta}_k$, thereby establishing the consistency rates of the estimators from (5). The results for the latter are shown to coincide with the corresponding results which obtain in the stationary case provided $\alpha > 1$ but to differ otherwise.

In Theorem 1 we first establish the consistency of the OLS estimators from (5), in the sense that under H_0, both $\hat{\phi}_k$ and $\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k$ become arbitrarily small (in probability) as the sample size increases.

Theorem 1 *Let y_t be generated according to* (1)-(2) *under Assumption 1. Then, under $\mathsf{H}_0 : \phi = 0$, the OLS estimates $\hat{\phi}_k$ and $\hat{\boldsymbol{\beta}}_k$ from the ADF regression (5) are such that:*

(i) if $1/k + k^2/T \to 0$ and $(a_T/T) \sum_{i=k+1}^{\infty} |\beta_i| \to 0$, in each case as $T \to \infty$, then $\hat{\phi}_k = o_P(1)$;

(ii) if, in addition to the conditions given in part (i), it also holds that $k^{1/2}(a_T/T) \sum_{i=k+1}^{\infty} |\beta_i| \to 0$, as $T \to \infty$, then $\|\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\| = o_P(1)$.

Remark 3.1. The conditions $(a_T/T)\sum_{i=k+1}^{\infty}|\beta_i| \to 0$ and $k^{1/2}(a_T/T)\sum_{i=k+1}^{\infty}|\beta_i| \to 0$ required in parts (i) and (ii) of Theorem 1, respectively, are redundant for $\alpha \in (1,2)$, since they are then implied by the summability condition $\sum_{i=1}^{\infty} i^{2/\delta}|\beta_i| < \infty$ ensured by Assumption 1(ii). It is stronger than its standard finite-variance counterpart $\sum_{i=k+1}^{\infty} i^{1/2}|\beta_i| < \infty$; see, for example, Lütkepohl (2005, Proposition 15.1). □

In Theorem 2 we present the asymptotic null distributions of the ADF statistics R_T and Q_T from section 2.

Theorem 2 *Let y_t be generated according to* (1)-(2) *under Assumption 1 and let* $\mathsf{H}_0 : \phi = 0$ *hold. Then, provided* $1/k + k^3/T \to 0$ *and* $a_T \sum_{i=k+1}^{\infty}|\beta_i| \to 0$ *in each case as* $T \to \infty$*:*

$$R_T \overset{w}{\to} \left(\int_0^1 \mathcal{S}^2\right)^{-1} \int_0^1 \mathcal{S} d\mathcal{S} \tag{8}$$

$$Q_T \overset{w}{\to} \left([\mathcal{S}]_1 \int_0^1 \mathcal{S}^2\right)^{-1/2} \int_0^1 \mathcal{S} d\mathcal{S} \tag{9}$$

where $\mathcal{S}$ is the α-stable process defined in part (i) of Assumption 1 and $[\mathcal{S}]_1$ denotes the quadratic variation of the semimartingale $\mathcal{S}$ at unity; that is, $[\mathcal{S}]_1 := \mathcal{S}(1)^2 - 2\int_0^1 \mathcal{S}d\mathcal{S}$.[1]

Remark 3.2. The asymptotic distributions in (8) and (9) first appeared in Chan and Tran (1989). They are expressed as the same functionals of the α-stable motion $\mathcal{S}$ (with $\alpha < 2$) as the so-called Dickey-Fuller distributions are of a standard Brownian motion (an α-stable motion with $\alpha = 2$). A graphical comparison of the distribution in (8) for $\alpha = 1$ with the corresponding Dickey-Fuller distribution is given in Figure 1 in Chan and Tran (1989, p.361). Although scale-invariant, the distributions in (8) and (9) depend on the distribution of ε_t through the two scalar quantities α and p defined in part (i) of Assumption 1. Where the distribution of ε_t is symmetric (so that $p = 1/2$), this dependence is characterised by the tail index, α, alone.

Remark 3.3. For the case where the lag length k is set to zero in (5), Phillips (1990) shows that the limiting null distribution of $\hat{\phi}_0 := (\sum_t y_{t-1}^2)^{-1}\sum_t y_{t-1}\Delta y_t$ is given by (under technical conditions similar to our Assumption 1)

$$T_k\hat{\phi}_0 \overset{w}{\to} \left(\int_0^1 \mathcal{S}^2\right)^{-1}\left(\int_0^1 \mathcal{S}d\mathcal{S} + \tfrac{1}{2}(1-\kappa_u^2)[\mathcal{S}]_1\right)$$

where $\kappa_u^2 := (\sum_{i=0}^{\infty}\gamma_i^2)/(\sum_{i=0}^{\infty}\gamma_i)^2$. He further demonstrates that the semi-parametric analogues of R_T and Q_T proposed in Phillips (1987) and Phillips and Perron (1988), which use a non-parametric correction for the weak dependence in u_t rather than an autoregressive sieve device, achieve the same limiting null distributions as given for those statistics in Theorem 2 above.

Remark 3.4. The analogue of our condition that $a_T\sum_{i=k+1}^{\infty}|\beta_i| \to 0$ in the finite-variance case is that $T^{1/2}\sum_{i=k+1}^{\infty}|\beta_i| \to 0$; see Berk (1974) and Lewis and Reinsel (1985). Both conditions involve the order of magnitude of the (possibly centred) error sums $\sum_{t=1}^{T}\varepsilon_t$, respectively a_T and $T^{1/2}$ for infinite and finite variance. Our condition entails that k is, in

[1]Notice that because $\mathcal{S}$ is a pure jump process for $\alpha \in (0,2)$, it holds that $[\mathcal{S}]_1 = \sum_{u\in(0,1]}[\Delta\mathcal{S}(u)]^2$, where $\Delta\mathcal{S}(u) := \mathcal{S}(u) - \mathcal{S}(u-)$ is non-zero when a jump occurs in $\mathcal{S}$ at time u (see also Remark 2.2).

general, required to grow at a faster rate the smaller is α. Nevertheless, in the important special case of a finite-order autoregression, it would suffice to choose k at least as large as the true autoregressive order, while in the case where the β_i, $i = 1, 2, \ldots$ exhibit exponential decay (as happens for finite-order ARMA processes), any power rate of the form $k = T^r$ ($r \in (0,1)$) would be sufficient uniformly in α.

Remark 3.5. In common with most of the literature on unit root testing when the innovations are infinite variance, we have assumed that no deterministic component is present in either the DGP (1)-(2) or in the ADF regression (5); this amounts to assuming that the distribution of the innovations is (known to be) centred on zero. More generally, one could consider replacing (1) with $(y_t - d_t) = \rho(y_{t-1} - d_{t-1}) + u_t$, where d_t is a purely deterministic component. In the case where d_t was a slowly evolving deterministic component (which includes a constant as a special case) of the form considered in Elliott *et al.* (1996,p.816), the results given in this paper should not alter, at least for $\alpha > 1$, if the ADF regression in (5) was constructed using the appropriate quasi-difference de-trended data in place of y_t; see Elliott *et al.* (1996,p.824) for details. Where d_t is either a constant or a linear trend, inference should be based on the appropriately de-trended data; as in Phillips (1990, p.55), we will discuss this extension for the case of OLS de-trending further in section 6 below. It is also worth noting that where $\alpha < 1$, the usual constant and linear trend cases of d_t both cease to be an issue so far as the results in Theorem 2 are concerned because here the scaling term a_T^{-1} is of $o(T^{-1})$. □

We now move to establishing the asymptotic distributions of the OLS sieve estimators, $\hat{\beta}_1, \ldots, \hat{\beta}_k$, from (5) under H_0. A by-product of this is to provide us with the rate at which $\|\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\|$ shrinks to zero under the unit root null, information which will be subsequently used for establishing the validity of the bootstrap ADF tests discussed in section 4. There we shall also discuss the sieve estimator of $\boldsymbol{\beta}_k$ obtained under the restriction of H_0. This estimator, employed in the bootstrap tests of Chang and Park (2003), will be denoted $\check{\boldsymbol{\beta}}_k := (\check{\beta}_1, \ldots, \check{\beta}_k)'$and is computed by OLS regression of Δy_t on $\mathbf{X}_{t-1}^k := (\Delta y_{t-1}, \ldots, \Delta y_{t-k})'$, $t = k+1, \ldots, T$. Therefore $\check{\boldsymbol{\beta}}_k$ under H_0 coincides with the sieve estimator analysed for the case of stationary linear processes driven by InfV innovations in Cavaliere *et al.* (2016a). The results therein show that three kinds of asymptotic behaviour are possible under H_0 for $\check{\boldsymbol{\beta}}_k$, depending on the existence of the α-order moment $\mathrm{E}\,|\varepsilon_t|^\alpha$ and, moreover, on the behaviour of the ratio $P(|\varepsilon_1\varepsilon_2| > n)/P(|\varepsilon_1| > n)$ in the tails of the distribution, i.e. for $n \to \infty$. This feature was first observed by Davis and Resnick (1985b, 1986) for the case of stationary finite-order autoregressions driven by InfV innovations. Theorem 3, below, shows that the same trichotomy occurs with the unrestricted sieve estimator $\hat{\boldsymbol{\beta}}_k$ for $\alpha > 1$, albeit not for $\alpha < 1$ where also the consistency rate is affected.

Before we present Theorem 3 we need to define some additional notation. First, for $\mathrm{E}\,|\varepsilon_1|^\alpha < \infty$ and $\lim_{n\to\infty} P(|\varepsilon_1\varepsilon_2| > n)/P(|\varepsilon_1| > n) = 2\,\mathrm{E}\,|\varepsilon_1|^\alpha$ define $\tilde{a}_T := a_T$, whereas otherwise define $\tilde{a}_T := \inf\{x : P\left(|\varepsilon_1\varepsilon_2| > x\right) \leq T^{-1}\}$. In the latter case, $\tilde{a}_T = a_T\tilde{l}_T$ for some $\tilde{l}_T$, slowly varying at infinity, such that $\tilde{l}_T \to \infty$ as $T \to \infty$ (see Davis and Resnick,1985b, p.263, 1986, p.542). Second, define the cross-product matrices $S_{00}^k := \sum_{t=k+1}^T \mathbf{X}_{t-1}^k(\mathbf{X}_{t-1}^k)'$ and $S_{01}^k := \sum_{t=k+1}^T \mathbf{X}_{t-1}^k y_{t-1}$. Third, define the infinite Toeplitz matrix $\Sigma := (r_{|i-j|})_{i,j=0}^\infty$ formed from the *scale-free autocovariances*, $r_{|i-j|} := \sum_{s=0}^\infty \gamma_s\gamma_{s+|i-j|}$. Finally, denote by L a generic $m \times \infty$ selection matrix of constants, with (i,j)th element l_{ij}, and let $L_k := (L_{\cdot 1}, \ldots, L_{\cdot k})$ denote the matrix formed from the first k columns of L.

Theorem 3 *Let the conditions of Theorem 2 hold, including the rate conditions on k imposed therein, with the additional condition that k is not a slowly varying function of T for the particular value $\alpha = 1$. Also assume that there exists some $\delta' \in (\delta, \frac{2\alpha}{2+\alpha})$, where δ is as defined in part (i) of Assumption 1, such that the selection matrix L has δ'-summable rows under linear weighting; that is, such that $\sum_{j=1}^{\infty} j|l_{ij}|^{\delta'} < \infty$, $i = 1, ..., m$. Then:*

Case *(i): If* $\mathrm{E}\,|\varepsilon_t|^\alpha = \infty$, *then*

$$a_T^2 \tilde{a}_T^{-1} L_k \left\{ (\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k) - g_T \right\} \xrightarrow{w} S^{-1} \sum_{j=1}^{\infty} A_j S_j, \tag{10}$$

where $g_T := d_T - \hat{\phi}_k (S_{00}^k)^{-1} S_{01}^k$ *with* $d_T := T_k \gamma(1) \mu_T (S_{00}^k)^{-1} \mathbf{i}_k$, $\mu_T := \mathrm{E}(\varepsilon_1 \varepsilon_2 \mathbb{I}_{\{|\varepsilon_1 \varepsilon_2| \le \tilde{a}_T\}})$, $A_j \in \mathbb{R}^m$ *and are given by* $A_j := \sum_{i=1}^{j} L \Sigma_{\cdot i}^{-1} \gamma_{j-i}$ $(j \in \mathbb{N})$, $\{S_j\}_{j=1}^{\infty}$ *is an i.i.d. sequence of α-stable random variables and S is an almost surely positive $\alpha/2$-stable random variable independent of* $\{S_j\}_{j=1}^{\infty}$. *For $\alpha \in (1, 2)$, the term $\hat{\phi}_k (S_{00}^k)^{-1} S_{01}^k$ in the definition of g_T above can be omitted.*

Case *(ii): If* $\mathrm{E}\,|\varepsilon_t|^\alpha < \infty$ *and* $\lim_{n\to\infty} P(|\varepsilon_1 \varepsilon_2| > n)/P(|\varepsilon_1| > n) = \mathrm{E}\,|\varepsilon_1|^\alpha$, *then the convergence result in (10) holds with* $a_T^2 \tilde{a}_T^{-1} = a_T$, *and where $\{S_j\}_{j=1}^{\infty}$ and S are as described in Case (i) except that they are now dependent with joint distribution as given in Theorem 3.5 of Davis and Resnick (1985b).*

Case *(iii): If* $\mathrm{E}\,|\varepsilon_t|^\alpha < \infty$ *and* $P(|\varepsilon_1 \varepsilon_2| > n)/P(|\varepsilon_1| > n)$ *does not converge to* $\mathrm{E}\,|\varepsilon_1|^\alpha$ *as* $n \to \infty$, *then*

$$\|L_k \{ (\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k) - g_T \} - \sigma_T^{-2} \sum_{j=1}^{\infty} A_j \sum_{t=k+1}^{T} (\varepsilon_{t-j} \varepsilon_t - \mu_T) \| = o_P(a_T^{-2} \tilde{a}_T), \tag{11}$$

where $\sigma_T^2 := \sum_{t=k+1}^{T} \varepsilon_t^2$, *although no limiting distribution for* $L_k \{ (\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k) - g_T \}$ *needs to exist.*

Remark 3.6. The need to place a summability condition on the rows of L in order to obtain the consistency rates and, where appropriate, distributional results (rather than simply establishing consistency) given in Theorem 3 is standard in the sieve literature. A similar condition is, for example, imposed on L in the finite-variance case; see Theorem 2(iv) of Lewis and Reinsel (1985).

Remark 3.7. In the finite variance case the condition $T^{1/2} \sum_{i=k+1}^{\infty} |\beta_i| \to 0$, which as discussed in Remark 3.4 is the analogue of our condition that $a_T \sum_{i=k+1}^{\infty} |\beta_i| \to 0$, is required; see Berk (1974) and Lewis and Reinsel (1985).

Remark 3.8. Given part (i) of Assumption 1 and the assumption of δ'-row-summability of L under linear weighting, the A_j $(j \in \mathbb{N})$ are also row-wise δ'-summable under linear weighting; that is, $\sum_{j=1}^{\infty} j|A_{ij}|^{\delta'} < \infty$, $i = 1, ..., m$. This property allows us to employ Davis and Resnick's (1985b and 1986) asymptotic theory for sample autocovariances. The upper bound on δ' is used to control the convergence rate of the quantity $L_k \Sigma_k^{-1}$ to $L\Sigma$.

Remark 3.9. The results stated in Theorem 3 also hold for the restricted sieve estimator, $\breve{\boldsymbol{\beta}}_k$, which imposes the unit root null hypothesis, on deleting the term $\hat{\phi}_k (S_{00}^k)^{-1} S_{01}^k$ from the definition of g_T given in Theorem 3. These results therefore coincide with those given for sieve estimators in the case of stationary linear processes driven by InfV innovations in Cavaliere *et al.* (2016a). As a consequence, it would be possible to develop bootstrap specification tests on the elements of $\boldsymbol{\beta}_k$, based on either $\hat{\boldsymbol{\beta}}_k$ or $\check{\boldsymbol{\beta}}_k$ as is done in Cavaliere *et al.* (2016a), which could then be used for data-based selection of the lag truncation k to use in (5). It should also be possible to select k via standard information criteria; Burridge and Hristova (2008) show that k can be consistently estimated in (5) by standard information criteria in cases where the true autoregressive order in (4) is finite. We conjecture that their results should carry over to the setting considered in this paper.

Remark 3.10. Using standard least squares algebra, the restricted sieve estimator $\check{\boldsymbol{\beta}}_k$ can be written in terms of the unrestricted estimator as $\check{\boldsymbol{\beta}}_k = \hat{\boldsymbol{\beta}}_k + \hat{\phi}_k(S_{00}^k)^{-1}S_{01}^k$. Hence,

$$L_k(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k - d_T) = L_k(\check{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k - d_T) - \hat{\phi}_k L_k(S_{00}^k)^{-1}S_{01}^k. \tag{12}$$

For $\alpha \in (0,1)$, the first term on the right hand side of (12), $L_k(\check{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k - d_T)$, is asymptotically dominated by the second term, $\hat{\phi}_k L_k(S_{00}^k)^{-1}S_{01}^k$, under H_0. It is shown in the Appendix that in this case, without correcting for the latter term,

$$TL_k(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k - d_T) \xrightarrow{w} -\frac{\int \mathcal{S}d\mathcal{S}}{\int \mathcal{S}^2}\left(\gamma(1)\frac{\int \mathcal{S}d\mathcal{S}}{[\mathcal{S}]_1}L\Sigma^{-1}\mathbf{i} + \gamma(1)^{-1}L\mathbf{i}\right), \tag{13}$$

where $d_T = o(T^{-1})$ can be omitted. For $\alpha = 1$, (13) holds if $a_T^2/(T\tilde{a}_T) \to \infty$, whereas its right-hand side is replaced by $A + B\lim_{T\to\infty} a_T^2/(T\tilde{a}_T)$ if $0 \le \lim_{T\to\infty} a_T^2/(T\tilde{a}_T) < \infty$ and cases (i) or (ii) of Theorem 3 apply; here A and B are the right hand sides of (10) and (13), respectively. As a consequence, when $\alpha \in (0,1)$ the rate of consistency of the unrestricted sieve estimator $\hat{\boldsymbol{\beta}}_k$ under H_0 is reduced to T, slower than that of $\check{\boldsymbol{\beta}}_k$, while they attain the same rate of consistency when $\alpha \in (1,2)$; at $\alpha = 1$ the consistency rates differ by a slowly varying factor. For the wild bootstrap to be asymptotically valid a sufficiently fast consistency rate is required on the sieve estimates, such that the corresponding sieve residuals are sufficiently close to the true innovations. As we shall see in section 4.2, the consistency rates established here for both the restricted and unrestricted sieve estimates are sufficient for these purposes, regardless of the value of α.

Remark 3.11. The results given in this section are also sufficient to determine corresponding results for the class of modified unit root tests originally proposed in Stock (1999) and developed further in Perron and Ng (1996) and Ng and Perron (2001); see also Haldrup and Jansson (2006). Where these are based on semi-parameteric estimation of the serial correlation nuisance parameters, the limiting null distributions of these statistics are an immediate corollary of the results given in Phillips (1990). Results for the analogous tests based on an autoregressive spectral density (ASD) estimator of the type proposed in Berk (1974) follow immediately from the results given here. Moreover, the sieve wild bootstrap principle outlined in section 4 can also be applied to these modified unit root tests using the ASD estimator from Cavaliere and Taylor (2009). The results given in this paper are sufficient to show that these would also constitute asymptotically valid bootstrap tests in the presence of infinite variance innovations, under the same conditions as given for the sieve wild bootstrap ADF tests here. □

4 Wild Bootstrap ADF Tests

It is well known that in the case where the innovations have InfV, commonly used bootstrap re-sampling methods such as those based on either i.i.d. re-sampling or the wild bootstrap are unable to replicate the (first-order) asymptotic null distribution of test statistics; see Athreya (1987) for an early general reference and Zarepour and Knight (1999) for the specific case of unit root statistics. Valid bootstrap solutions can, however, be obtained based on either subsampling techniques (see, *inter alia*, Romano and Wolf, 1999) or the m-out-of-n bootstrap (see, *inter alia*, Arcones and Giné, 1989, 1991). The validity of the latter is shown by Zarepour and Knight (1999), Horvath and Kokoszka (2003) and Moreno and Romo (2012), and for the former by Jach and Kokoszka (2004). However, in many testing problems these methods, which entail the use of a bootstrap sample which has a smaller sample size than the original sample, have been shown to result in tests with somewhat unreliable finite sample size properties; see, in particular, Cornea and Davidson (2015). Moreover, the validity of these methods is only demonstrated by the aforementioned authors for an $AR(1)$ process.

In this section we establish the result that a particular version of the sieve wild bootstrap ADF unit root test, originally proposed in Cavaliere and Taylor (2009) for time series driven by unconditionally heteroskedastic shocks with finite variance, can provide p-values which are (first order) asymptotically valid (such that they are uniformly distributed on $[0, 1]$ under the unit root null hypothesis) also in the presence of symmetric InfV innovations. We show that a sufficient requirement for this to hold is that the wild bootstrap innovations are generated by multiplying the sieve residuals obtained in section 3 by a sequence of i.i.d. symmetric two point (Rademacher) distributions. In the finite variance case, Cavaliere and Taylor (2009) show that the wild bootstrap ADF tests are valid for any i.i.d. sequence with mean zero, unit variance and bounded fourth moment. As a result, the wild bootstrap ADF tests discussed here are asymptotically valid, when implemented using the Rademacher form, for both finite and infinite variance innovations.

The key feature of our approach is that we do not attempt to deliver a bootstrap algorithm which is able to approximate the unconditional asymptotic distribution of the unit root test statistics given in (8) and (9). Instead, in contrast to the finite variance framework, we show that the sieve wild bootstrap tests we propose replicate particular *conditional* asymptotic distributions, where the conditioning is upon the absolute values of the original innovations. Because the sieve wild bootstrap replicates the asymptotic distribution of the unit root statistics *conditional* on the sample extremes, it has several potential advantages; see Cavaliere *et al.* (2013), who applied this idea to a simple location model. Specifically: (i) the unit root statistics are evaluated with respect to a more concentrated distribution than the unconditional distribution so power gains might be expected; (ii) the sample size of the bootstrap sample coincides with the original sample size; (iii) preliminary knowledge or estimation of the tail index α is not required. The downside of conditioning on the absolute values of the innovations is that these will need to be assumed symmetric. While this assumption is commonly made in the InfV literature it is still important to stress that our proposed unit root tests would not be valid if this assumption did not hold. A property close to symmetry is also imposed by Zarepour and Knight (1999) in establishing the asymptotic validity of their m-out-of-n unit root tests.

In section 4.1 we now detail our sieve wild bootstrap algorithm. Proof of its asymptotic validity is then given in section 4.2.

4.1 The Sieve Wild Bootstrap Algorithm

Our bootstrap algorithm involves the OLS estimators of ϕ and $\boldsymbol{\beta}_k = (\beta_1, ..., \beta_k)'$ from (5). As discussed in section 3 these can either be obtained unrestrictedly, denoted by $(\hat{\phi}_k, \hat{\boldsymbol{\beta}}_k')'$, or under the restriction of the null hypothesis $\phi = 0$, denoted by $(\check{\phi}_k, \check{\boldsymbol{\beta}}_k')' = (0, \check{\boldsymbol{\beta}}_k')'$. Additionally, let $(\tilde{\phi}_k, \tilde{\boldsymbol{\beta}}_k')'$ generically denote either the unrestricted or the restricted estimator.

Algorithm 1

(i) *Compute the OLS residuals,* $\hat{\varepsilon}_t := \Delta y_t - \tilde{\phi}_k y_{t-1} - \sum_{i=1}^k \tilde{\beta}_i \Delta y_{t-i}$, $t = k+1, ..., T$.

(ii) *Using the OLS residuals,* $\hat{\varepsilon}_t$*, from step (i), generate the bootstrap innovations* $\varepsilon_t^* := \hat{\varepsilon}_t w_t$, $t = k+1, ..., T$, *where the* w_t *are i.i.d. Rademacher random variables, i.e. such that* $w_t \in \{-1, +1\}$*, each outcome occurring with probability* $\frac{1}{2}$*, and are independent of the original data.*

(iii) *Construct the bootstrap shocks* u_t^* *using the recursion*

$$u_t^* = \sum_{i=1}^k \tilde{\beta}_i u_{t-i}^* + \varepsilon_t^*, \; t = k+1, ..., T \tag{14}$$

initialized at $(u_1^*, ..., u_k^*) = (0, ..., 0)$*; accordingly, the bootstrap data are generated as*

$$y_t^* = y_0^* + \sum_{i=1}^t u_i^*, \; t = 1, ..., T \tag{15}$$

initialized at $y_0^* = 0$.

(iv) *Estimate the ADF regression*

$$\Delta y_t^* = \phi y_{t-1}^* + \sum_{i=1}^k \beta_i \Delta y_{t-i}^* + \varepsilon_{t,k}^*, \quad t = k+1, ..., T$$

on the bootstrap sample, yielding the corresponding OLS estimators $\hat{\phi}_k^*$ *and* $\hat{\boldsymbol{\beta}}_k^* := (\hat{\beta}_1^*, ..., \hat{\beta}_k^*)'$*, together with* $s(\hat{\phi}_k^*)$*, the OLS estimate of the standard error of* $\hat{\phi}_k^*$.

(v) *Define the bootstrap statistics*

$$R_T^* := \frac{T_k \hat{\phi}_k^*}{1 - \sum_{i=1}^k \hat{\beta}_i^*} \qquad \text{and} \qquad Q_T^* := \frac{\hat{\phi}_k^*}{s(\hat{\phi}_k^*)}.$$

(vi) *The associated bootstrap p-values are defined as* $p_{R,T}^* := G_{R,T}^*(R_T)$ *and* $p_{Q,T}^* := G_{Q,T}^*(Q_T)$*, where* $G_{R,T}^*$ *and* $G_{Q,T}^*$ *denote the cumulative distribution functions (cdfs) of* R_T^* *and* Q_T^**, in each case conditional on the original data.*

Some remarks are in order.

Remark 4.1. In step (iii) of Algorithm 1 the bootstrap process u_t^* is initialized at zero. This is not strictly necessary; for instance, one may alternatively set $u_t^* = \Delta y_t$, $t = 1, ..., k$ (i.e., the initial values are set to the corresponding value that u_t takes under the null hypothesis), or $u_t^* = \Delta y_t - \hat{\phi}_0 y_{t-1}$, where $\hat{\phi}_0$ is the OLS estimator from the regression of Δy_t on y_{t-1}. These alternative initialisations for u_t^* would not alter the asymptotic results given in section 4.2. Other initialisations for y_t^* could also be considered; Chang and Park (2003, p.390), for example, set $y_0^* = y_0$.

Remark 4.2. Algorithm 1 uses the OLS estimates of $\beta_1, ..., \beta_k$ to recolour the bootstrap shocks ε_t^* through the recursion in (14) of step (iii). Although it seems likely this will improve the finite sample properties of the bootstrap test, this step is not strictly required. One could in fact simply set $u_t^* = \varepsilon_t^*$ in place of (14) so that the bootstrap sample obtained in step (iv) would be the random walk with (conditionally on the data) i.i.d. increments, $y_t^* = y_0^* + \sum_{i=1}^t \varepsilon_i^*$, $t = 1, ..., T$. The asymptotic results given in section 4.2 relate to the use of the recoloured bootstrap shocks but apply equally where no recolouring is used. Some finite sample comparisons between these two versions of Algorithm 1 are reported in section 5.

Remark 4.3. The sieve estimates used in step (i) of Algorithm 1 to generate the residuals $\hat{\varepsilon}_t$ can be either the unrestricted OLS estimates $(\hat{\phi}_k, \hat{\boldsymbol{\beta}}_k')'$ or the restricted ones $(0, \check{\boldsymbol{\beta}}_k')'$. In order to prove bootstrap validity under the null hypothesis (which is based on a conditional argument, where conditioning is upon the absolute values of the innovations, $|\varepsilon_t|$), it is crucial that the estimator used in step (i) is sufficiently precise to guarantee that the residuals $\hat{\varepsilon}_t$ and the true innovations ε_t are sufficiently close. In section 4.2 we shall show that both $(\hat{\phi}_k, \hat{\boldsymbol{\beta}}_k')'$ and $(0, \check{\boldsymbol{\beta}}_k')'$ satisfy this requirement for any $\alpha \in (0, 2)$. However, it should be noted that under the alternative hypothesis $\check{\boldsymbol{\beta}}_k$ will not necessarily be close to the true $\boldsymbol{\beta}_k$ and this may affect the finite sample power of the corresponding test. These issues will be explored further by Monte Carlo simulation in Section 5.

Remark 4.4. Computation of the bootstrap p-values in step (vi) of Algorithm 1 requires the (conditional) cdfs to be known. Taking R_T to illustrate, these can be approximated numerically in practice by generating B (conditionally) independent bootstrap statistics, $R_{T:b}^*$, $b = 1, ..., B$, computed as in Algorithm 1, steps (i)–(v). Then, $p_{R,T}^*$ is approximated by $\hat{p}_{R,T,B}^* := B^{-1} \sum_{b=1}^B \mathbb{I}(R_{T:b}^* \leq R_T)$, and is such that $\hat{p}_{R,T,B}^* \overset{a.s.}{\to} p_{R,T}^*$ as $B \to \infty$; cf. Hansen (1996), Andrews and Buchinsky (2000) and Davidson and MacKinnon (2000). □

4.2 Bootstrap Asymptotic Theory

In this section we establish that the sieve wild bootstrap tests from Algorithm 1 deliver asymptotically valid inference *conditional* on the (unobservable) vector of magnitudes of the innovations, $|\boldsymbol{\varepsilon}_T| := (|\varepsilon_1|, ..., |\varepsilon_T|)$, under the assumption that the distribution of the innovations is symmetric. In order to simplify exposition and save on space we will only present the results relating to the normalised bias test. Analogous results for the t-ratio test follow similarly as does the fundamental result that the wild bootstrap t-test is asymptotically valid in exactly the same sense as the normalised bias test.

Specifically, in the remainder of this section we will show that under the unit root null hypothesis the distribution of R_T, conditional on $|\boldsymbol{\varepsilon}_T|$, is consistently estimated by the distribution of R_T^*, conditional on the data. More formally, we will establish that the distributions of R_T conditional on $|\boldsymbol{\varepsilon}_T|$ and of R_T^* conditional on the data converge jointly to the same random distribution. We obtain this result as a consequence of a consistency property of the wild bootstrap that we formulate in Theorem 4 below in a general setup, abstracting both from the particular unit root inference problem and the InfV assumption.

Theorem 4 is based on a few ingredients which we now define:

(a) Let $\mathbf{Z}_T := (Z_0, ..., Z_T)$ denote a possible sample of observables which may depend on a vector $\boldsymbol{\varepsilon}_T := (\varepsilon_1, ..., \varepsilon_T)$ of (potentially unobservable) i.i.d. shocks, defined on the same probability space as the data $\mathbf{Z}_T$. In applications of Theorem 4, the data $\mathbf{Z}_T$ will typically be a function of some random 'shocks', including ε_T, and possibly other random variables. For instance, in our unit root testing problem $\mathbf{Z}_T$ corresponds to the data $\{y_t\}_{t=0}^T$, which depend on the shocks $\boldsymbol{\varepsilon}_T := (\varepsilon_1, ..., \varepsilon_T)$ as well as on the pre-sample innovations $\{\varepsilon_t\}_{t=-\infty}^0$ (which are not included in $\boldsymbol{\varepsilon}_T$). It is important to stress that, for the purposes of Theorem 4, we do not need to assume that $\boldsymbol{\varepsilon}_T$ satisfies Assumption 1; only symmetry is further imposed on $\boldsymbol{\varepsilon}_T$.

(b) The shocks $\boldsymbol{\varepsilon}_T$ are re-sampled using the wild bootstrap, based on Rademacher bootstrap signs. In particular, we let $\mathbf{w}_T := (w_1, ..., w_T)$ denote a vector of i.i.d. Rademacher random variables (possibly defined upon expansion of the probability space on which $\mathbf{Z}_T$ and $\boldsymbol{\varepsilon}_T$ are defined), independent of $\mathbf{Z}_T$ and $\boldsymbol{\varepsilon}_T$. The corresponding wild bootstrap process is then denoted as $\boldsymbol{\varepsilon}_T^\dagger = (\varepsilon_1^\dagger, ..., \varepsilon_T^\dagger)$, where $\varepsilon_t^\dagger := \varepsilon_t w_t$, $t = 1, ..., T$. Notice that this bootstrap is in general a merely theoretical device, except in the special case where $\boldsymbol{\varepsilon}_T$ is observable.

(c) We consider two statistics: the first, denoted by g_T, which depends on the original data $\mathbf{Z}_T$ only; the second, denoted by g_T^*, which is a classical wild bootstrap statistic, depending on both $\mathbf{Z}_T$ and the wild bootstrap Rademacher shocks $\mathbf{w}_T$.

The results in Theorem 4 will allow us to detail the proximity in distribution of g_T and g_T^*. Specifically, it provides sufficient conditions such that, conditionally on the original data, the bootstrap quantiles of g_T^* can be consistently used in order to evaluate g_T.

Theorem 4 *Let $\mathbf{Z}_T$, $\boldsymbol{\varepsilon}_T$ and $\boldsymbol{\varepsilon}_T^\dagger$ be as defined above, with the additional requirement that $\boldsymbol{\varepsilon}_T$ has a symmetric distribution. Moreover, let there be given:*

(a) a statistic $g_T = g_T(\mathbf{Z}_T)$ and a wild bootstrap statistic $g_T^* = g_T^*(\mathbf{Z}_T, \mathbf{w}_T)$;

(b) a measurable function $\gamma_T : \mathbb{R}^T \to \mathbb{R}$, $T \in \mathbb{N}$.

As $T \to \infty$, if the following two conditions hold:

(i) for every $\eta > 0$,

$$P_{|\varepsilon|}\left(|g_T - \gamma_T(\boldsymbol{\varepsilon}_T)| > \eta\right) \xrightarrow{P} 0 \qquad \text{and} \qquad P_{Z,\varepsilon}(|g_T^* - \gamma_T(\boldsymbol{\varepsilon}_T^\dagger)| > \eta) \xrightarrow{P} 0,$$

where $P_{|\varepsilon|}$ and $P_{Z,\varepsilon}$ denote probability conditional on $|\varepsilon_T| := (|\varepsilon_1|, ..., |\varepsilon_T|)$ and $\{\mathbf{Z}_T, \varepsilon_T\}$, respectively;

(ii) the distribution of $\gamma_T(\boldsymbol{\varepsilon}_T)$ conditional on $|\varepsilon_T|$ converges weakly to a random measure φ with cumulative distribution process $\varphi((-\infty, \cdot])$ which is a.s. continuous at every point in $\mathbb{R}$,[2]

then,

$$P_Z\left(g_T^* \le g_T\right) \xrightarrow{w} U[0,1], \tag{16}$$

where P_Z denotes probability conditional on the data $\mathbf{Z}_T$ and $U[0,1]$ denotes the uniform distribution on $[0,1]$.

[2]For the relation between weak convergence to random measures and the associated cumulative processes, see e.g. Daley and Vere-Jones (2008, pp.143-144).

Remark 4.5. The consistency of the wild bootstrap established in Theorem 4 is based on the following triangular argument. First, under condition (i), the statistic of interest g_T is well approximated, in $P_{|\varepsilon|}$-probability, by a function of the shocks $\boldsymbol{\varepsilon}_T$, namely $\gamma_T(\boldsymbol{\varepsilon}_T)$. Second, the bootstrap statistic g_T^* is well approximated, in $P_{Z,\varepsilon}$-probability, by the same function, but evaluated at the bootstrap shocks $\boldsymbol{\varepsilon}_T^\dagger$. Third, due to the assumed symmetry of $\boldsymbol{\varepsilon}_T$, the (random) distributions of the approximating quantities $\gamma_T(\boldsymbol{\varepsilon}_T)$ under $P_{|\varepsilon|}$ and $\gamma_T(\boldsymbol{\varepsilon}_T^\dagger)$ under $P_{Z,\varepsilon}$ (equivalently, under P_ε) coincide a.s. Notice that the specification of w_t to be Rademacher random variables is crucial for this distributional equality to hold; see also Remark 4.8 below. By combining these three facts we can conclude that the cumulative distribution process of g_T under $P_{|\varepsilon|}$ is close to that of g_T^* under $P_{Z,\varepsilon}$ (equivalently, under P_Z). Moreover, as both processes converge weakly to the same a.s. continuous limiting process under condition (ii), their proximity is uniform:

$$\sup_{x\in\mathbb{R}} \left| P_{|\varepsilon|}(g_T \le x) - P_Z(g_T^* \le x) \right| \xrightarrow{P} 0$$

as $T \to \infty$. It is in this sense that the wild bootstrap approximates the distribution of g_T conditional on the shock magnitudes, $|\boldsymbol{\varepsilon}_T|$. Again using a.s. continuity of the limiting cumulative distribution process, (16) follows.

Remark 4.6. The triangular argument used to prove the result in (16), relies on the distributional equality of $\boldsymbol{\varepsilon}_T$ under $P_{|\varepsilon|}$ and $\boldsymbol{\varepsilon}_T^\dagger$ under P_ε. Given symmetry of $\boldsymbol{\varepsilon}_T$, this property is ensured by choosing the Rademacher distribution for the bootstrap signs in $\mathbf{w}_T$. Other choices, such as the $N(0,1)$ distribution proposed in Cavaliere and Taylor (2008), will, in general, compromise the required distributional equality. □

As anticipated earlier in this section, our approach to formally establishing bootstrap validity of sieve bootstrap ADF tests under InfV is to verify that the conditions of Theorem 4 are satisfied by the unit root statistic $g_T = R_T$ of (8) and its sieve wild bootstrap counterpart, $g_T^* = R_T^*$ from Algorithm 1, with $\mathbf{Z}_T = \{y_t\}_{t=0}^T$. In order to do so, we choose the function $\gamma_T(\cdot)$ appearing in Theorem 4 as follows. For any vector $\mathbf{x}_T := (x_1, ..., x_T)$ in $\mathbb{R}^T$, $T \in \mathbb{N}$, we let

$$\gamma_T(\mathbf{x}_T) := T\frac{\sum_{t=1}^T x_t \sum_{s=1}^{t-1} x_s}{\sum_{t=1}^T (\sum_{s=1}^{t-1} x_s)^2}\,. \tag{17}$$

Notice that $\gamma_T(\mathbf{x}_T)$ is the normalised OLS estimator of ϕ for a random walk starting at 0 and steps collected in the vector $\mathbf{x}_T$.

Our interest therefore centres on the two quantities

$$\gamma_T(\boldsymbol{\varepsilon}_T) = T\frac{\sum_{t=1}^T \varepsilon_t \sum_{s=1}^{t-1} \varepsilon_s}{\sum_{t=1}^T (\sum_{s=1}^{t-1} \varepsilon_s)^2} \qquad \text{and} \qquad \gamma_T(\boldsymbol{\varepsilon}_T^\dagger) = T\frac{\sum_{t=1}^T \varepsilon_t^\dagger \sum_{s=1}^{t-1} \varepsilon_s^\dagger}{\sum_{t=1}^T (\sum_{s=1}^{t-1} \varepsilon_s^\dagger)^2}.$$

Again, these two quantities correspond to the normalised OLS estimator of ϕ for a random walk with steps collected in the vectors $\boldsymbol{\varepsilon}_T$ and $\boldsymbol{\varepsilon}_T^\dagger$, respectively.

Verification that condition (i) of Theorem 4 holds in our framework is given in Lemmas A.2 and A.3 in the Appendix. In particular, in these lemmas we show that, as $T \to \infty$,

$$P_{|\varepsilon|}(|R_T - \gamma_T(\boldsymbol{\varepsilon}_T)| > \eta) \xrightarrow{P} 0 \qquad \text{and} \qquad P^*(|R_T^* - \gamma_T(\boldsymbol{\varepsilon}_T^\dagger)| > \eta) \xrightarrow{P} 0,$$

as required in (i).

That condition (ii) also holds is established in the following lemma, which provides a weak convergence result for the distribution of $\gamma_T(\boldsymbol{\varepsilon}_T)$ conditionally on the magnitude of the shocks, $|\boldsymbol{\varepsilon}_T|$. The lemma provides a representation for the relevant (random) limiting measure and states the required result that this measure has an a.s. continuous cumulative distribution process. To establish the latter, we assume that the unconditional limiting distribution in (8) has a continuous distribution function; although this property appears to be treated as known in the profession (e.g., Jach and Kokoszka, 2004, p.78), we could not find reference to a formal proof. Before introducing the lemma, recall from Remark 2.2 the distributional equality $\mathcal{S}(\cdot) \stackrel{d}{=} \sum_{i=1}^{\infty} \Gamma_i^{-1/\alpha} \delta_i \mathbb{I}\{U_i \leq (\cdot)\}$ in $D[0,1]$.

Lemma 1 *Let $\Gamma := \{\Gamma_i\}_{i\in\mathbb{N}}$ denote the sequence of arrival times of a Poisson process with unit intensity, $U := \{U_i\}_{i\in\mathbb{N}}$ denote an i.i.d. sequence of uniform random variables on $[0,1]$, and let $\delta = \{\delta_i\}$ denote a sequence of i.i.d. Rademacher random variables, such that Γ, U and δ are independent. Finally define $J(\cdot) := \sum_{i=1}^{\infty} \Gamma_i^{-1/\alpha} \delta_i \mathbb{I}\{U_i \leq (\cdot)\}$. Under Assumption 1, if the distribution of ε_t is symmetric, it holds that*

$$\mathcal{L}\left(\gamma_T(\boldsymbol{\varepsilon}_T) \middle| |\boldsymbol{\varepsilon}_T|\right) \xrightarrow{w} \mathcal{L}\left(\frac{1}{2}\frac{J(1)^2 - \sum_{i=1}^{\infty}\Gamma_i^{-2/\alpha}}{\int_0^1 J^2} \middle| \Gamma, U\right)$$

$$\stackrel{d}{=} \mathcal{L}\left(\frac{\int_0^1 \mathcal{S} d\mathcal{S}}{\int_0^1 \mathcal{S}^2} \middle| \{|\Delta\mathcal{S}(u)|\}_{u\in(0,1]}\right) \tag{18}$$

in the sense of weak convergence of random measures on $\mathbb{R}$, with conditioning on the magnitude and the location of the jumps of J and $\mathcal{S}$ respectively. Moreover, the cumulative process of the limiting random measure is a.s. continuous at every point in $\mathbb{R}$.

We are now in a position to use Theorem 4 and show our main theoretical result; that is, the consistency of the sieve wild bootstrap ADF test under the unit root null hypothesis. This is provided in Theorem 5 below. The theorem also allows us to conclude that the bootstrap unit root statistic R_T^* approximates the same limit distribution as that found in Lemma 1.

Theorem 5 *Under the conditions of Theorem 2 and if the distribution of ε_t is symmetric, it holds under the unit root hypothesis $\mathsf{H}_0 : \phi = 0$ that*

$$\mathcal{L}\left(R_T^* \middle| y_0, ..., y_T\right) \xrightarrow{w} \mathcal{L}\left(\frac{\int_0^1 \mathcal{S} d\mathcal{S}}{\int_0^1 \mathcal{S}^2} \middle| \{|\Delta\mathcal{S}(u)|\}_{u\in(0,1]}\right) \tag{19}$$

in the sense of weak convergence of random measures on $\mathbb{R}$, provided the shocks w_t used in step (ii) of Algorithm 1 form an i.i.d. sequence of Rademacher random variables, i.e. such that $P(w_t = 1) = P(w_t = -1) = 0.5$, and are independent of $\{y_t\}_{t=0}^T$. Under this condition, it also holds that

$$p_{R,T}^* = P^*\left(R_T^* \leq R_T\right) \xrightarrow{w} U[0,1]. \tag{20}$$

Remark 4.7. An immediate implication of the result in (20) is that the wild bootstrap implementation of the R_T tests (the same holds for the wild bootstrap Q_T tests) detailed in Algorithm 1 of section 4 will have correct asymptotic size in the presence of InfV innovations, regardless of the value of tail index, α, provided the Rademacher distribution is used in step (ii) of the algorithm. Combined with the results in Cavaliere and Taylor (2008,2009), this therefore establishes the result that the Rademacher-based wild bootstrap R_T and Q_T

ADF tests are asymptotically valid regardless of whether the innovations display finite or infinite variance. From a practical perspective this is a powerful result as it implies that using these wild bootstrap implementations of the ADF tests allows the practitioner to take an ambivalent stance on whether the innovations have finite or infinite variance.

Remark 4.8 It could be shown that if the Rademacher assumption on w_t was dropped and a different zero-mean finite-variance symmetric distribution was used instead, then the limit in (19) would involve, in place of $\mathcal{S}$, a process $\mathcal{S}^w(\cdot) \stackrel{d}{=} \sum_{i=1}^{\infty} \Gamma_i^{-1/\alpha} w_i \mathbb{I}\{U_i \leq (\cdot)\}$. Such a limit would differ from that of $\gamma_T(\boldsymbol{\varepsilon}_T)$ in (18) and, more importantly, it is not obvious that such a limit could be attained by $\gamma_T(\boldsymbol{\varepsilon}_T)$ upon any change of the conditioning σ-algebra in (18). In this sense the requirement of a Rademacher distribution for w_t is essential for showing bootstrap validity. □

5 Finite Sample Simulations

In this section we use Monte Carlo simulation methods to investigate the finite-sample performance of the wild bootstrap implementations of the R_T and Q_T ADF unit root tests outlined in the previous section. We will report results for both recoloured and non recoloured versions of the wild bootstrap tests from Algorithm 1 (see Remark 4.2). We only report results for the wild bootstrap tests based on restricted residuals (see Remark 4.3). Corresponding results for (recoloured and non-recoloured) wild bootstrap tests based on unrestricted residuals were inferior and, hence, are not reported here, but can be found in the accompanying working paper, Cavaliere *et al.* (2016b). Comparison will also be made with analogous tests derived using the sub-sampling procedure of Jach and Kokoszka (2004) and the m-out-of-n bootstrap of Zarepour and Knight (1999), in each case extended to include a sieve element and associated recolouring. We will also include results for tests based on the R_T and Q_T statistics using either the asymptotic critical values appropriate to the finite variance case from Fuller (1996) or critical values simulated from the asymptotic null distributions in Theorem 2. The latter are of course infeasible in practice as they are based on knowledge of α; they do, however provide a useful benchmark to compare the performance of the bootstrap tests against and, moreover, the power results for these tests essentially quantify the size-adjusted power of the tests based on the asymptotic critical values from Fuller (1996).

As our reference DGP we consider the $MA(1)$

$$\begin{aligned} \Delta y_t &= \phi y_{t-1} + u_t, \quad t = 1, ..., T, \\ u_t &= \varepsilon_t + \theta \varepsilon_{t-1}, \quad t = 1, ..., T, \end{aligned}$$

with $\varepsilon_{-1} = y_0 = 0$, and where ε_t was generated as an i.i.d. sequence of α-stable random variables; that is, $\varepsilon_t \sim i.i.d.$ stable(α). Results are reported for the following values of the tail index, $\alpha \in \{2, 1.5, 1\}$. The Gaussian case $\alpha = 2$ is included to illustrate the uniform behaviour of the wild bootstrap across finite and infinite variance. Corresponding results for $\alpha = 1.7$ are reported in the accompanying working paper, Cavaliere *et al.* (2016b); these were little different to the results reported here for $\alpha = 2$ and, hence, have been omitted. We consider the following values of the MA parameter, $\theta \in \{0, \pm 0.5, \pm 0.8\}$, allowing both positive and negative and moderate and large MA behaviour, as well as the i.i.d. case. We report results both for the null hypothesis, $\phi = 0$, and for alternatives of the form $\phi = -c/T$ where we set $c = 7$.

Following Schwert (1989), the lag length used in the ADF regression, (5), for the purposes of computing the ADF statistics, R_T and Q_T, was set to $k = \lfloor \kappa (T/100)^{1/4} \rfloor$, and as is common in the literature, we report results for $\kappa = 4$ and $\kappa = 12$. For all of the procedures considered, this lag length was also used in the corresponding bootstrap analogue of (5) (for the wild bootstrap procedures, see step (iv) of Algorithm 1), regardless of whether recolouring was used or not. Where recolouring was employed this was also based on this same choice of k.

In Table 1 we report results for empirical size while Table 2 reports results for empirical power (non size-adjusted). Results are reported for the following tests: (i) the (non-bootstrap) R_T and Q_T tests of section 2 based on either the critical values from Fuller (1996), denoted R_T^2 and Q_T^2 or using critical values where α is assumed known, denoted R_T^α and Q_T^α; (ii) wild bootstrap implementations of the R_T and Q_T tests constructed as in Algorithm 1 using recolouring and restricted residuals, denoted $R_T^{rc,r}$ and $Q_T^{rc,r}$; (iii) tests as in (ii) but with no recolouring used, denoted $R_T^{n,r}$ and $Q_T^{n,r}$.

Tables 1-2 also report results relating to the m-out-of-n bootstrap testing approach of Zarepour and Knight (1999), extended to allow for the same recolouring device as we use with the wild bootstrap tests discussed in this paper. This procedure parallels Algorithm 1 except that the bootstrap errors in step (ii) are generated using the m-out-of-n re-sampling scheme with $m = \lfloor T/\ln\ln T \rfloor$, and with the lag length again set to k in both the original and bootstrap ADF regressions.[3] We denote these tests R_T^m and Q_T^m. The reported results relate to the use of the unrestricted sieve residuals in step (i) of the algorithm; we also considered versions of the m-out-of-n bootstrap ADF tests based on restricted residuals and without the use of recolouring - these alternatives delivered no discernible improvements over the results reported here. Finally, we also report corresponding results for implementations of the R_T and Q_T tests based on the sub-sampling approach outlined in Jach and Kokoszka (2004). The procedure we adopted is as outlined on page 76 of their paper but applied to the sieve residuals in order to construct the bootstrap sample. In particular, the sieve residuals are centred and all possible blocks of consecutive residuals of length $b = \lfloor 0.125 \times T \rfloor$ are then sub-sampled from the centred residuals and then cumulated with no recolouring employed (since the residuals are re-sampled in blocks). A regression of order k is then fitted to the bootstrap sample. These tests will be denoted R_T^{jk} and Q_T^{jk}.

All simulations were programmed in Ox 7.01 using 10,000 Monte Carlo replications and $B = 399$ bootstrap replications (see Remark 4.4) using an implementation of the algorithm of Chambers *et al.* (1976) to generate the α-stable random variables. Results are reported here for $T = 100$. Results for $T = 500$ are available on request.

Table 1 about here

Consider first the results reported in Table 1 relating to the empirical size properties of the various tests considered. There are two striking features in these results. First, the naïve ADF Q_T^2 t-type test, based on the standard critical values from Fuller (1996), is surprisingly robust to InfV. Abstracting away from the large negative MA case ($\theta = -0.8$) which causes oversize in most of the tests, size is reasonably well controlled regardless of the value of α, although there is a tendency to undersizing for the smaller values of α considered. The same cannot be said for the corresponding normalised bias test, R_T^2, which is often significantly over-sized, most notably when $\kappa = 12$, suggesting that the standard (Dickey-Fuller) asymptotic critical value may be an increasingly poor approximation to the true finite sample critical value as k is increased in the ADF regression (5). The same pattern is seen with the infeasible R_T^α

[3]While one might consider setting the lag length in the m-out-of-n bootstrap ADF regression to $\lfloor \kappa(m/100)^{1/4} \rfloor$ (reflecting the smaller bootstrap sample size under this method of re-sampling) this would not be appropriate because of the k-th order recolouring polynomial employed in the analogue of step (iii) of Algorithm 1.

test which uses the simulated asymptotic critical value based on the knowledge of α. These patterns are not replicated in either the corresponding wild or m-out-of-n bootstrap tests, further supporting this conjecture. The second striking feature is the failure of the tests based on sub-sampling, Q_T^{jk} and R_T^{jk}. These tend to be over-sized when θ is large and negative with $\kappa = 4$ but are often very heavily under-sized when $\kappa = 12$ regardless of the value of θ. As we will see when we discuss the results in Table 2 this has dramatic implications for their finite sample power properties.

Among the wild bootstrap tests considered some variation is seen between the results based on recoloured data *vis-à-vis* non-recoloured data. These differences are relatively small, although the tests based on non-recoloured data do perhaps perform slightly better overall. While it is clearly seen from the results in Table 1 that the wild bootstrap does tend to deliver significant improvements on the naïve normalised bias R_T^2 test, it is also fair to say from the results in Table 1 that any improvements seen in the finite sample size of the corresponding Q_T^2 test are far from spectacular. This is perhaps not too surprising given the apparent robustness to InfV shown by this test, noted above. However, excepting some cases where $\theta = -0.8$, the wild bootstrap procedure does consistently deliver, albeit relatively small, improvements on the size of the Q_T^2 test.

The m-out-of-n bootstrap, despite being calculated from a bootstrap sample which has a smaller sample size than the original sample, also performs well in our experiments. It is much less affected than the wild bootstrap tests by large negative values of θ when α is relatively small ($\alpha = 1.5$ and $\alpha = 1.0$), but on the other hand tends to be more affected by such MA behaviour when $\alpha = 2$ (and also when $\alpha = 1.7$; see Cavaliere *et al.*, 2016b). These cases aside, the m-out-of-n bootstrap performs very well overall with broadly comparable size control to the wild bootstrap tests.

Table 2 about here

Turning to the results in Table 2 relating to empirical power, perhaps the most obvious feature is the tendency to very low power in the tests based on sub-sampling, consistent with the heavy under-sizing in these cases observed in Table 1. Based on the results in Tables 1 and 2, the Q_T^{jk} and R_T^{jk} tests cannot be recommended for use in practice.

In terms of the naïve Q_T^2 and R_T^2 tests, a comparison with those for the corresponding tests based on known α critical values, Q_T^α and R_T^α, shows that the latter can be significantly more powerful than the former, although this may in part be attributable to size differences between these tests; cf. Table 1. Focusing on the t-type tests (given the lack of size control seen in the R_T^2 and R_T^α tests), the best performing (in terms of size control) among the wild bootstrap tests display similar power to the corresponding Q_T^α and Q_T^2 tests in the Gaussian case, $\alpha = 2.0$ (the same is also seen for $\alpha = 1.7$; see Cavaliere *et al.*, 2016b), but outperform the Q_T^α test (and often even more so the Q_T^2 test) for $\alpha = 1.5$ and smaller. This suggests that the wild bootstrap tests simultaneously improve on the finite sample size properties of the naïve ADF tests, and deliver gains in finite sample power over both these tests and on the benchmark given by the infeasible ADF tests based on knowledge of α. The m-out-of-n t-type bootstrap test, Q_T^m, also performs well although its power is more comparable with the infeasible ADF tests than the wild bootstrap tests. It is also apparent from the results in Table 2 that the m-out-of-n implementation of the normalised bias test, R_T^m, can display very low power relative to the other tests when $\kappa = 12$.

To conclude this section it is worth commenting that the usual trade-off between empirical size and power relating to the choice of the lag length k (see, in particular, Schwert, 1989) is apparent for all of the tests (except for the ill-behaved R_T^2 and sub-sample unit root tests discussed above) in the results in Tables 1 and 2. In those cases where significant oversize is seen, empirical size is often considerably improved by increasing κ from 4 to 12; these improvements are most obvious for those values of θ where the size distortions are the most pronounced and in particular in the near cancellation region where $\theta = -0.8$. The usual flip side of the coin is that empirical power is decreased between $\kappa = 4$ and $\kappa = 12$, other things equal. That this usual trade-off exists in the InfV case, just as it does in the finite variance case, is not surprising and suggests that the use of data-based lag selection methods, see Remark 3.10, is likely to be a useful tool here too and the development and evaluation of such methods in the InfV environment constitutes a useful topic for further research.

6 Deterministic Terms

We now provide some brief discussion of how deterministic terms, in particular a constant or a linear trend, can be introduced into the model.

To that end, let y_t satisfy (1)-(2), but now suppose that we have observations on $z_t^\mu = y_t + \mu$ or $z_t^\tau = y_t + \mu + \tau t$, where μ, τ are unknown parameters.[4] Then suppose z_t^μ is demeaned, yielding $\hat{y}_t^\mu := z_t^\mu - T^{-1}\sum_{s=1}^T z_s^\mu$, or that z_t^τ is detrended by the OLS regression of z_t^τ on $(1,t)'$, yielding residuals $\hat{y}_t^\tau$. The AR sieve (5) may then be applied to either $\hat{y}_t^\mu$ or $\hat{y}_t^\tau$, as appropriate. We may then define, analogously to R_T and Q_T from Section 2, the corresponding unit root test statistics based on the de-meaned and de-trended data as R_T^μ, Q_T^μ and R_T^τ, Q_T^τ, respectively, using an obvious notation. Under the assumption that the distribution of ε_t is symmetric and using the notation of Lemma 1, it could be shown that, for $\kappa \in \{\mu, \tau\}$,

$$\mathcal{L}\left(R_T^\kappa \middle| |\varepsilon_T|\right) \xrightarrow{w} \mathcal{L}\left(\left.\frac{\int_0^1 J^\kappa dJ}{\int_0^1 (J^\kappa)^2}\right| \Gamma, U\right) \tag{21}$$

in the sense of weak convergence of random measures on $\mathbb{R}$, where, for $u \in [0,1]$,

$$J^\mu(u) \quad : \quad = J(u) - \int_0^1 J(r)dr = \sum_{i=1}^\infty \Gamma_i^{-1/\alpha}\delta_i[\mathbb{I}\{U_i \le u\} + U_i - 1], \tag{22}$$

$$\begin{aligned} J^\tau(u) \quad : \quad &= J(u) - \int_0^1 (4-6r)J(r)dr + 6u\int_0^1 (1-2r)J(r)dr \\ &= \sum_{i=1}^\infty \Gamma_i^{-1/\alpha}\delta_i[\mathbb{I}\{U_i \le u\} - 1 + (4-6u)U_i - (3-6u)U_i^2]. \end{aligned} \tag{23}$$

The corresponding representations for Q_T^μ and Q_T^τ are similarly obtained. Importantly, from the expressions for J^μ and J^τ it can be seen that randomness in the conditional limit distributions (21) is fully due to the Rademacher signs δ_i, which are independent of Γ, U, as in Lemma 1. Consequently, the wild bootstrap based on Rademacher signs w_t could be successfully used for inference with the reference distribution given by the limit in (21). To this end, after estimating (5) for the demeaned data $\hat{y}_t^\mu$ (or the detrended data, $\hat{y}_t^\tau$), bootstrap data could be generated exactly as in steps (i)-(iii) of Algorithm 1. Prior to conducting step (iv), the same transformation, either demeaning or detrending, should then be applied to the

[4]More generally we could allow for the case where $z_t^\mu = y_t + d_t$ with d_t a pth order, $0 \le p \le P < \infty$, polynomial in t, but the leading constant and linear trend cases would seem sufficient for the purposes of exposition.

bootstrap data as was applied to the original data, z_t^κ. Step (iv) of Algorithm 1 should then be performed on the transformed bootstrap data.

Remark 6.1. It is worth noting that for inference based on any of the R_T^μ, Q_T^μ, R_T^τ and Q_T^τ statistics, the assumption made in section 2 that y_0 is of $O_P(1)$ is made without loss of generality because all these statistics are exact similar with respect to y_0; cf. Müller and Elliott (2003).

Remark 6.2. The foregoing conclusions would continue to hold if rather than OLS de-trended data, the appropriate quasi-difference de-trended data of Elliott *et al.* (1996, p.824) were used. In such a case, replace the de-meaned and de-trended processes J^μ and J^τ in (22) and (23), respectively, by the corresponding quasi-difference de-trended processes; see Elliott *et al.* (1996, p.817) for further details. □

7 Conclusions

We have extended the corpus of available asymptotic distribution theory for the ADF test to cover the case where the shocks follow a linear process driven by i.i.d. infinite variance [InfV] innovations. We have demonstrated that the limiting null distributions of these statistics coincide with those previously derived in the literature for the case where the shocks follow a finite-order autoregression driven by InfV innovations, provided the lag length in the ADF regression satisfies the usual rate condition given in Said and Dickey (1987) for the case where the innovations have finite variance. Because these distributions depend on the tail index of the innovation distribution they cannot be used directly for inference on the unit root hypothesis. We have also established the large sample properties, including consistency rates, of the associated sieve estimates from the ADF regression and of the corresponding sieve estimates which are obtained under the restriction of the unit root null hypothesis. Using these results we then established that asymptotically valid wild bootstrap ADF tests can be formed using the residuals from either of these sieve regressions combined with a Rademacher-based implementation of the wild bootstrap re-sampling scheme. Although these results rest on the assumption that the innovations are symmetrically distributed they do not require knowledge of the tail index. Monte Carlo simulation results were reported which suggested that the sieve wild bootstrap ADF tests perform well in practice.

It would be interesting to investigate whether the symmetry assumption could be dropped. In order to do so, we would need to find an alternative method of re-sampling to the wild device currently used in step (ii) of Algorithm 1 such that the resulting sieve bootstrap tests were still asymptotically valid. This would likely entail convergence to a different conditional limiting null distribution from the one shown to hold for the wild bootstrap statistics here. Unfortunately, we have been unable to find such a conditional distribution and so we leave this as a suggestion for further research. An alternative approach, along the lines suggested by Cornea-Madeira and Davidson (2015) for tests in a simple location model, would be to base inference on direct simulation of the asymptotic limiting distribution from Theorem 2 using consistent estimates of the tail index α and the asymmetry parameter p. It should be possible, given the results obtained in this paper, to show the asymptotic validity of such an approach. However, it is important to note that this method will approximate the unconditional, rather than a conditional, distribution of the ADF test statistics. The power gains observed for the wild bootstrap tests discussed here could not then be expected. Again, we leave a detailed exploration of this alternative approach for further research.

The Monte Carlo results reported in this paper also suggested that a sieve m-out-of-n

bootstrap procedure also performed well in finite samples. Formally establishing the asymptotic validity of these tests would also constitute an interesting topic for further research, although it should be noted that symmetry is also assumed by Zarepour and Knight (1999) for establishing the asymptotic validity of m-out-of-n bootstrap unit root tests they consider in the context of a first-order autoregressive model.

Finally, our focus in this paper has been on the case where the innovations driving the linear process in (2) are drawn from a class of infinite *unconditional* variance processes. In many applications in finance an equally plausible model is one where the innovations are drawn from a class of infinite variance *conditional* variance processes, a leading example being the well-known integrated GARCH (IGARCH) and explosive GARCH models considered in Zhang, Sin and Ling (2015). An analysis of such cases is clearly beyond the scope of this paper, but would constitute an interesting area for further research and the work of Zhang, Sin and Ling (2015) should provide a useful basis for doing so. It is important to note, however, that (under certain regularity conditions) the tail index of such processes is $\alpha = 2$, and so they are in the Gaussian domain of attraction. As a consequence, the wild bootstrap approach we have outlined here would not have the conditional interpretation that is central to the present paper and the arguments and technical results involved will be quite different.

References

Andrews, D.W.K. and Buchinsky, M. (2000). A Three-Step Method for Choosing the Number of Bootstrap Repetitions, *Econometrica* 68, 23–51.

Adler, R.J., Feldman, R.E. and Taqqu, M.S. (1998) *A Practical Guide to Heavy Tails.* Boston, MA Birkhauser.

Arcones, M. and Giné, E. (1989). The bootstrap of the mean with arbitrary bootstrap sample size, *Annales de l'I.H.P. Probabilités et statistiques, Section B* 25, 457-481.

Arcones, M. and Giné, E. (1991). Additions and correction to "The bootstrap of the mean with arbitrary bootstrap sample size", *Annales de l'I.H.P. Probabilités et statistiques, Section B* 27, 583-595.

Athreya, K.B. (1987). Bootstrap of the mean in the infinite variance case, *The Annals of Statistics* 15, 724-731.

Aue, A., Berkes, I. and Horváth, L. (2008). Selection from a stable box. *Bernoulli* 14, 125-139.

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

Brillinger, D. (2001). *Time Series. Data Analysis and Theory.* SIAM, Philadelphia.

Burridge, P. and Hristova, D. (2008). Consistent estimation and order selection for non-stationary autoregressive processes with stable innovations. *Journal of Time Series Analysis* 29, 695–718.

Caner, M. (1998). Tests for cointegration with infinite variance errors. *Journal of Econometrics* 86, 155—175.

Cavaliere, G. and Georgiev, I. (2013). Exploiting infinite variance through dummy variables in non-stationary autoregressions, *Econometric Theory* 29, 1162-1195.

Cavaliere, G., Georgiev, I. and Taylor, A.M.R. (2013). Wild bootstrap of the sample mean in the infinite variance case, *Econometric Reviews* 32, 204-219.

Cavaliere, G., Georgiev, I. and Taylor, A.M.R. (2016a). Sieve inference for infinite-variance stationary linear processes. *The Annals of Statistics*, forthcoming, with on-line supplement.

Cavaliere, G., Georgiev, I. and Taylor, A.M.R. (2016b). Unit root inference for non-stationary linear processes driven by infinite variance innovations. *Quaderni di Dipartimento 1, Department of Statistics, University of Bologna.* Downloadable from `https://ideas.repec.org/p/bot/quadip/wpaper130.html`

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

Cavaliere, G. and Taylor, A.M.R. (2009). Bootstrap M Unit Root Tests, *Econometric Reviews* 28, 393—421.

Chan, N.H. and Tran, L.T. (1989). On the first order autoregressive process with infinite variance. *Econometric Theory* 5, 354-362.

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

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

Charemza, W.W., Hristova, D. and Burridge, P. (2005). Is inflation stationary? *Applied Economics* 37, 901–3.

Cornea-Madeira, A. and Davidson, R. (2015), A Parametric Bootstrap for Heavy-Tailed Distributions, *Econometric Theory* 31, 449–470.

Daley, D. and Vere-Jones, D. (2008), *An Introduction to the Theory of Point Processes*, Volume II: General Theory and Structure, Springer-Verlag.

Davidson, R. and MacKinnon, J. (2000). Bootstrap Tests: How Many Bootstraps?, *Econometric Reviews* 19, 55–68.

Davis, R. A. (2010). Heavy tails in financial time series, In *Encyclopedia Quantitative Finance* (ed. R. Cont), John Wiley and Sons.

Davis, R. and Resnick, S. (1985a). Limit theory for moving averages of random variables with regularly varying tail probabilities. *The Annals of Probability* 13, 179-195.

Davis, R. and Resnick, S. (1985b). More limit theory for the sample correlation function of moving averages. *Stochastic Processes and their Applications* 20, 257-279.

Davis, R. and Resnick, S. (1986). Limit theory for the sample covariance and correlation functions of moving averages. *The Annals of Statistics* 14, 533-558.

Dickey, D. A. and Fuller, W. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427–431.

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

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, Springer-Verlag.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, Vol. 2, John Wiley & Sons.

Finkenstädt, B. and Rootzén, H. (2003). *Extreme values in finance, telecommunication and the environment*, Chapman & Hall.

Fuller, W.A. (1996). *Introduction to Statistical Time Series* (2nd Edition). Wiley, New York.

Gabaix, X. (2009). Power laws in economics and finance, *Annual Review of Economics* 1, 255–293.

Haldrup, N. and Jansson, M. (2006). Improving power and size in unit root testing. In T.C. Mills and K. Patterson (eds.), *Palgrave Handbooks of Econometrics, volume 1*, chap. 7. Palgrave MacMillan, pp. 252–277.

Hansen, B.E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica* 64, 413–430.

Horvath, L. and Kokoszka, P. (2003). A bootstrap approximation to a unit root test statistic for heavy-tailed observations. *Statistics & Probability Letters* 62, 163-173.

Jach, A. and Kokozska, P. (2004). Subsampling unit root tests for heavy-tailed observations. *Methodology & Computing in Applied Probability* 6, 73-97.

Knight, K. (1989). Limit theory for autoregressive parameter estimates in an infinite-variance random walk. *Canadian Journal of Statistics* 17, 261-278.

Knight, K. (1991). Limit theory for M-estimates in an integrated infinite variance process. *Econometric Theory* 7, 200-212.

LePage, R. (1992). Bootstrapping signs, in Lepage R. and Billard L. (eds.) *Exploring the Limits of the Bootstrap*, John Wiley & Sons.

LePage, R.Podgórsky, K. (1996). Resampling permutations in regression without second moments, *Journal of Multivariate Analysis* 57, 119–141.

LePage, R., Podgórski K. and Ryznar, M. (1997). Strong and conditional invariance principles for samples attracted to stable laws. *Probability Theory and Related Fields* 108, 281–298.

Lewis, R. and Reinsel, G. (1985). Prediction of multivariate time series by autoreressive model fitting, *Journal of Multivariate Analysis* 16, 393-411.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis.* Springer: Berlin.

Moreno, M. and Romo, J. (2012). Unit root bootstrap tests under infinite variance. *Journal of Time Series Analysis* 33, 32–47.

Maronna, R., Martin, D. and Yohai, V. (2006). *Robust Statistics: Theory and Methods*, John Wiley & Sons.

McCulloch, J. C. (1986). Simple consistent estimators of stable distribution parameters. *Communications in Statistics, Simulation and Computation* 15, 1109–1136.

McCulloch, J. (1997). Measuring Tail Thickness to Estimate the Stable Index α: A Critique, *Journal of Business and Economic Statistics* 15, 74-81.

Müller, U. and Elliott, G. (2003). Tests for Unit Roots and the Initial Condition, *Econometrica* 71, 1269–1286.

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

Perron, P. and Ng, S. (1996). Useful modifications to some unit root tests with dependent errors and their local asymptotic properties, *Review of Economic Studies* 63, 435—463.

Phillips, P.C.B. (1987). Time Series Regression with a Unit Root. *Econometrica* 55, 277–301.

Phillips, P.C.B. (1990). Time series regression with a unit root and infinite-variance errors. *Econometric Theory* 6, 44–62.

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

Rachev, S. T. and Mittnik, S. (2000). *Stable Paretian Models in Finance.* New York: Wiley.

Rachev, S.T., Mittnik, S. and Kim, J.R. (1998) Time series with unit roots and infinite variance disturbances. *Applied Mathematics Letters* 11, 69—74.

Resnick, S. (1986). Point processes, regular variation and weak convergence. *Advances in Applied Probability* 18, 66-138.

Resnick, S. (1997). Heavy tailed modeling and teletraffic data. *Annals of Statistics* 25, 1805–1869.

Resnick, S. and Greenwood, P. (1979), A bivariate stable characterization and domains of attraction, *Journal of Multivariate Analysis* 9, 206–221.

Romano, J.P. and Wolf, M. (1999). Inference for the mean in the heavy-tailed case. *Metrika* 50, 55–69.

Said, S.E. and Dickey, D.A. (1984). Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order. *Biometrika* 71, 599–608.

Samarakoon, M. and Knight, K. (2009). A note on unit root tests with infinite variance noise. *Econometric Reviews* 28, 314–334.

Schwert, G.W. (1989). Tests for Unit Roots: A Monte Carlo Investigation, *Journal of Business and Economic Statistics* 7, 147–59.

Silverberg, G. and Verspagen, B. (2007). The size distribution of innovations revisited: an application of extreme value statistics to citation and value measures of patent significance, *Journal of Econometrics* 139, 318–339.

Stock, J.H. (1999). A class of tests for integration and cointegration. In: Engle, R.F. and White, H. (eds.) *Cointegration, Causality and Forecasting. A Festschrift in Honour of Clive W.J. Granger.* Oxford: Oxford University Press, pp. 137—167.

Zarepour, M. and Knight, K. (1999). Bootstrapping unstable first order autoregressive process with errors in the domain of attraction of stable law. *Communications in Statistics. Stochastic Models* 15, 11–27.

Zhang, R.-M., Sin, C.Y. and Ling, S. (2015). On functional limits of short- and long-memory linear processes with GARCH(1,1) noises. *Stochastic Processes and their Applications* 125, 482–512.

A Appendix

This appendix is organised as follows. In section A.1 we introduce a preliminary lemma and subsequently we give proofs of theorems 1 and 2 from Section 3. Bootstrap asymptotics are considered in section A.2. In section A.2.1 we provide the proof of Theorem 4 while, in section A.2.2 we verify that condition (i) of Theorem 4 is indeed satisfied under the assumptions of our main bootstrap theorem (Theorem 5). Finally, Section A.2.3 contains the proof of Lemma 1 and of Theorem 5.

A.1 Sieve Asymptotics

The proof of Theorem 1 relies on the following lemma where, for square matrices, $\|(\cdot)\|_2 := \sup_{\|x\|=1} \|(\cdot)x\|$ is used to denote the linear-space norm induced by the Euclidean vector norm, with x and the matrix having matching dimensions. In particular, for positive semi-definite matrices, $\|(\cdot)\|_2 = \lambda_{\max}(\cdot)$, the largest eigenvalue.

Lemma A.1 *Under Assumption 1 and if $k^2/T + 1/k \to 0$, for every $\epsilon > 0$,*

a. $S_{00}^k := \sum_{t=k+1}^T \mathbf{X}_{t-1}^k (\mathbf{X}_{t-1}^k)'$ satisfies $\|S_{00}^k - \Sigma_k \sigma_T^2\|_2 = O_P(l_T \tilde{a}_T) \max\{k^\epsilon a_k, k\}$, where $\sigma_T^2 := \sum_{t=1}^T \varepsilon_t^2$, $l_T = 1$ for $\alpha \neq 1$ and l_T is slowly varying for $\alpha = 1$; Moreover, $(S_{00}^k)^{-1}$ exists with probability approaching one and $\|(S_{00}^k)^{-1} - \Sigma_k^{-1}\sigma_T^{-2}\|_2 = O_P(l_T \tilde{a}_T a_T^{-4}) \max\{k^\epsilon a_k, k\}$.

b. $S_{0\varepsilon}^k := \sum_{t=k+1}^T \mathbf{X}_{t-1}^k \varepsilon_{t,k}$ satisfies $\|S_{0\varepsilon}^k - \sum_{t=k+1}^T \mathbf{X}_{t-1}^k \varepsilon_t\| = o_P(a_T^{1-\zeta}) + O_P(a_T^2) \sum_{j=k+1}^\infty |\beta_j|$ and $\|S_{0\varepsilon}^k\| = o_P(k^\epsilon a_k l_T \tilde{a}_T) + O_P(a_T^2) \sum_{j=k+1}^\infty |\beta_j|$ with $\zeta > 0$ sufficiently small and l_T as in (a).

c. $S_{1\varepsilon}^k := \sum_{t=k+1}^T y_{t-1}\varepsilon_{t,k} = \sum_{t=k+1}^T y_{t-1}\varepsilon_t + o_P(a_T^2) + o_P(a_T^3) \sum_{i=k+1}^\infty |\beta_i|$ if $k^3/T \to 0$.

d. $S_{01}^k := \sum_{t=k+1}^T \mathbf{X}_{t-1}^k y_{t-1}$ has $\|S_{01}^k\| = O_P(k^{1/2} a_T^2)$.

e. $S_{11}^k := \sum_{t=k+1}^T y_{t-1}^2 = \gamma(1)^2 \sum_{t=k+1}^T (\sum_{s=1}^{t-1} \varepsilon_s)^2 + O_P(a_T \max\{T, a_T T^\epsilon\})$.

Proof. Parts (a) and (b) reproduce Lemma 2(a,b) of Cavaliere *et al.* (2016a). For part (c), write

$$y_t = \gamma(1) \sum_{s=1}^t \varepsilon_s + v_t - v_0 + y_0,\ v_t := \sum_{i=0}^\infty \gamma_i^* \varepsilon_{t-i}, \gamma_i^* := -\sum_{j=i+1}^\infty \gamma_j; \tag{A.1}$$

the series ν_t is a.s. well-defined since $\sum_{i=1}^{\infty} i|\gamma_i|^\delta < \infty$. Then $S_{1\varepsilon}^k - \sum_{t=k+1}^T y_{t-1}\varepsilon_t = \sum_{t=k+1}^T y_{t-1}\rho_{t,k} = \gamma(1)(R_1 + R_2) + R_3 + (y_0 - v_0)R_4$ with

$$\begin{aligned} R_1 &= \textstyle\sum_{t=k+1}^T \sum_{i=k+1}^{t-1} \beta_i \sum_{j=0}^{t-i-1} \gamma_j \varepsilon_{t-i-j}^2, \\ R_2 &= \textstyle\sum_{i=k+1}^{\infty} \beta_i \sum_{t=k+1}^T \sum_{s=1}^{t-1} \sum_{j=0}^{\infty} \mathbb{I}_{\{s\neq t-i-j\}} \gamma_j \varepsilon_s \varepsilon_{t-i-j}, \\ R_3 &= \textstyle\sum_{i=k+1}^{\infty} \beta_i \sum_{t=k+1}^T v_{t-1} \sum_{j=0}^{\infty} \gamma_j \varepsilon_{t-i-j} \end{aligned}$$

and $R_4 = \sum_{t=k+1}^T \sum_{i=k+1}^{\infty} \beta_i u_{t-i}$ which are evaluated next. Define $D_k^\eta := \sum_{i=k+1}^{\infty} |\beta_i|^\eta \sum_{j=0}^{\infty} |\gamma_j|^\eta = O(1) \sum_{i=k+1}^{\infty} |\beta_i|^\eta$.

First, for every $M > 0$ with $I_{M,T} := \prod_{t=1}^T \mathbb{I}_{\{|\varepsilon_t| < Ma_T\}}$,

$$\mathrm{E}\,|a_T^{-2} I_{M,T} R_1| \leq a_T^{-2} T\, \mathrm{E}(\varepsilon_1^2 \mathbb{I}_{|\varepsilon_1| \leq Ma_T}) D_k^1 = O(1) \sum_{i=k+1}^{\infty} |\beta_i|$$

by Karamata's theorem [KT], so $I_{M,T} R_1 = O_P(a_T^2) \sum_{i=k+1}^{\infty} |\beta_i|$. Given any $\omega \in (0,1)$, the weak convergence of $\max_{t=1,\ldots,T} |a_T^{-1}\varepsilon_t|$ to an a.s. finite random variable implies the existence of M such that $P(I_{M,T} \neq 1) < \omega/2$ for all T, so also $R_1 = O_P(a_T^2) \sum_{i=k+1}^{\infty} |\beta_i|$.

Second, for $\alpha \in (0,1)$ and $\eta \in [\delta, \alpha)$, it holds that $\mathrm{E}\,|R_2|^\eta \leq T^2 (\mathrm{E}\,|\varepsilon_1|^\eta)^2 D_k^\eta$, so

$$R_2 = O_P(T^{2/\eta}) (\sum_{i=k+1}^{\infty} |\beta_i|^\eta)^{1/\eta} = o_P(a_T^2) + o_P(a_T^3) \sum_{i=k+1}^{\infty} |\beta_i|$$

by Lemma S.1 in the supplement to Cavaliere *et al.* (2016a) with $k^3/T \to 0$ and η sufficiently close to α. For $\alpha = 1$, $\mathrm{E}(|a_T^{-2} R_2|) \leq \{a_T^{-1} T\, \mathrm{E}(|\varepsilon_1| \mathbb{I}_{|\varepsilon_1| \leq Ma_T})\}^2 D_k = O(T^\epsilon) \sum_{i=k+1}^{\infty} |\beta_i|$ for every $\epsilon > 0$ since $a_T^{-1} T\, \mathrm{E}(|\varepsilon_1| \mathbb{I}_{|\varepsilon_1| \leq Ma_T})$ is slowly varying. Finally, for $\alpha \in (1,2)$, $R_2 = \sum_{i=1}^5 R_{2i}$ with the following summands:
(i) $R_{21} := \sum_{i=k+1}^{\infty} \beta_i \sum_{t=k+1}^T \sum_{s=1}^{t-1} \sum_{j=0}^{t-i-1} \mathbb{I}_{\{s\neq t-i-j\}} \gamma_j \varepsilon_s \varepsilon_{t-i-j} \mathbb{I}_{\{|\varepsilon_s| > a_T \text{ or } |\varepsilon_{t-i-j}| > a_T\}}$ having

$$\mathrm{E}(|a_T^{-2} R_{21}|) \leq a_T^{-1} T\, \mathrm{E}\,|\varepsilon_1| \{a_T^{-1} T\, \mathrm{E}(|\varepsilon_1| |\mathbb{I}_{|\varepsilon_1| > a_T})\} D_k^1 = O(a_T^{-1} T D_k^1) = o(a_T) \sum_{i=k+1}^{\infty} |\beta_i|$$

by KT, (ii) $R_{22} := \sum_{i=k+1}^{\infty} \beta_i \sum_{t=k+1}^T \sum_{s=1}^{t-1} \sum_{j=0}^{t-i-1} \mathbb{I}_{\{s\neq t-i-j\}} \gamma_j (\varepsilon_s \mathbb{I}_{|\varepsilon_s| \leq a_T} - \mu_T)(\varepsilon_{t-i-j} \mathbb{I}_{|\varepsilon_{t-i-j}| \leq a_T} - \mu_T)$ (with $\mu_T := \mathrm{E}(\varepsilon_s \mathbb{I}_{|\varepsilon_s| \leq a_T})$) having

$$\mathrm{E}(a_T^{-4} R_{22}^2) \leq 2\{a_T^{-2} T\, \mathrm{E}(\varepsilon_1^2 \mathbb{I}_{|\varepsilon_1| \leq a_T})\}^2 (D_k^1)^2 = O(1) (\sum_{i=k+1}^{\infty} |\beta_i|)^2$$

by independence and KT, (iii) $R_{23} := \mu_T \sum_{i=k+1}^{\infty} \beta_i \sum_{t=k+1}^T \sum_{s=1}^{t-1} (\varepsilon_s \mathbb{I}_{|\varepsilon_s| \leq a_T} - \mu_T) \sum_{j=0}^{t-i-1} \mathbb{I}_{\{s\neq t-i-j\}} \gamma_j$ having

$$\mathrm{E}\,|a_T^{-4} R_{23}^2| \leq (a_T^{-2} T^2 \mu_T^2) \{a_T^{-2} T\, \mathrm{E}(\varepsilon_1^2 \mathbb{I}_{|\varepsilon_1| \leq a_T})\} (D_k^1)^2 = O(1) (\sum_{i=k+1}^{\infty} |\beta_i|)^2$$

as, with $\mathrm{E}\,\varepsilon_s = 0$ it holds that $|\mu_T| \leq \mathrm{E}(|\varepsilon_s| \mathbb{I}_{|\varepsilon_s| > a_T}) = O(a_T/T)$ by KT, (iv) $R_{24} := \mu_T \sum_{i=k+1}^{\infty} \beta_i \sum_{t=k+1}^T \sum_{s=1}^{t-1} \sum_{j=0}^{t-i-1} \mathbb{I}_{\{s\neq t-i-j\}} \gamma_j (\varepsilon_{t-i-j} \mathbb{I}_{|\varepsilon_{t-i-j}| \leq a_T} - \mu_T)$ having the same upper bound as $\mathrm{E}\,|a_T^{-4} R_{24}^2|$, and (v) $R_{25} := \mu_T^2 \sum_{i=k+1}^{\infty} \beta_i \sum_{t=k+1}^T \sum_{s=1}^{t-1} \sum_{j=0}^{t-i-1} \mathbb{I}_{\{s\neq t-i-j\}} \gamma_j$

having $|R_{25}| \le T^2\nu_T^2 D_k^1 = O(a_T^2)\sum_{i=k+1}^{\infty}|\beta_i|$. Summarizing, $R_2 = o_P(a_T^3)\sum_{i=k+1}^{\infty}|\beta_i|$ for all $\alpha \in (0,2)$.

Third, $R_3 = R_{31} + R_{32}$ with $R_{31} := \sum_{i=k+1}^{\infty}\beta_i\sum_{t=k+1}^{T}(\sum_{j=0}^{\infty}\gamma_j\gamma_{j+i-1}^*)\varepsilon_{t-i-j}^2$ satisfying

$$|R_{31}| \le \sum_{i=k+1}^{\infty}|\beta_i|\sum_{i=k+1}^{\infty}|\sum_{t=k+1}^{T}\sum_{j=0}^{\infty}\gamma_j\gamma_{j+i-1}^*|\varepsilon_{t-i-j}^2 = o_P(a_T^3)\sum_{i=k+1}^{\infty}|\beta_i|$$

since, for $\eta \in (\delta,\alpha)$ under Assumption 1,

$$\mathrm{E}(\sum_{i=k+1}^{\infty}|\sum_{t=k+1}^{T}\sum_{j=0}^{\infty}\gamma_j\gamma_{j+i-1}^*|\varepsilon_{t-i-j}^2)^{\eta/2} \le T\,\mathrm{E}\,|\varepsilon_1|^{\eta}(\sum_{j=0}^{\infty}j|\gamma_j|^{\eta/2})^2 = O(T),$$

and $R_{32} := \sum_{i=k+1}^{\infty}\beta_i\sum_{t=k+1}^{T}\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\mathbb{I}_{l\neq j-i+1}\gamma_j^*\gamma_l\varepsilon_{t-1-j}\varepsilon_{t-i-l}$ satisfying, for $\eta \in [\delta,\alpha)\cap(0,1]$ under Assumption 1,

$$\mathrm{E}\,|R_{32}|^{\eta} \le T(\mathrm{E}\,|\varepsilon_1|^{\eta})^2\sum_{i=k+1}^{\infty}|\beta_i|^{\eta}(\sum_{j=0}^{\infty}j|\gamma_j|^{\eta})^2,$$

so $R_{32} = o_P(a_T^2)\sum_{i=k+1}^{\infty}|\beta_i|$ for $\alpha \in (1,2)$ (using $\eta = 1$) and $R_{32} = o_P(1)+o_P(a_T^3)\sum_{i=k+1}^{\infty}|\beta_i|$ by Lemma S.1 of the supplement to Cavaliere *et al.* (2016a) for $\alpha \in (0,1]$. Similarly, for the same η, $\mathrm{E}\,|R_4|^{\eta} \le T(\mathrm{E}\,|u_1|^{\eta})\sum_{i=k+1}^{\infty}|\beta_i|^{\eta}$, so $R_4 = o_P(a_T^2)\sum_{i=k+1}^{\infty}|\beta_i|$ for $\alpha \in (1,2)$, whereas $R_4 = o_P(1) + o_P\left(a_T^3\right)\sum_{i=k+1}^{\infty}|\beta_i|$ for $\alpha \in (0,1]$. By collecting the magnitude orders of R_i $(i = 1,...,4)$, $S_{1\varepsilon}^k - \sum_{t=k+1}^{T}y_{t-1}\varepsilon_t = o_P(1) + o_P(a_T^3)\sum_{i=k+1}^{\infty}|\beta_i|$ follows.

In part (d), using partial summation and the inequality $(\sum_{i=1}^{4}a_i)^2 \le 3\sum_{i=1}^{4}a_i^2$, we find

$$\begin{aligned}\|S_{01}^k\|^2 &= \sum_{i=1}^{k}(\sum_{t=k+1}^{T}u_{t-i}y_{t-1})^2 = \frac{1}{4}\sum_{i=1}^{k}(y_{T-i}^2 - y_{k-i}^2 + \sum_{t=k+1}^{T}u_{t-i}^2 + 2\sum_{j=1}^{i-1}\sum_{t=k+1}^{T}u_{t-j}u_{t-i})^2\\ &\le \Phi_1 + 3\Phi_2,\end{aligned}$$

with $\Phi_1 := \sum_{i=1}^{k}y_{T-i}^4 + \sum_{i=1}^{k}y_{k-i}^4$ and

$$\Phi_2 := \sum_{i=1}^{k}(\sum_{j=1}^{i}\sum_{t=k+1}^{T}u_{t-j}u_{t-i})^2 = \sum_{i=1}^{k}\{\sum_{j=1}^{i}(S_{00}^k)_{ij}\}^2.$$

These satisfy $\Phi_1 \le 2k\max_{1\le t\le T}|y_t|^4 \le 8k(\gamma^4(1)\max_{1\le t\le T}|\sum_{s=1}^{T}\varepsilon_s|^4 + \max_{1\le t\le T}|v_t|^4 + O_P(1)) = O_P(ka_T^4)$ as, under Assumption 1(i), $\max_{1\le t\le T}|a_T^{-1}\sum_{s=1}^{T}\varepsilon_s| \xrightarrow{w} \sup_{[0,1]}|\mathcal{S}| < \infty$ a.s. by the Continuous Mapping Theorem [CMT] and $\max_{1\le t\le T}|v_t| = O_P(a_T)$ by Theorem 3.2 of Davis and Resnick (1985a), and

$$\begin{aligned}\frac{1}{4}\Phi_2 &\le \frac{1}{2}\sigma_T^4\sum_{i=1}^{k}(\sum_{j=1}^{i}|(\Sigma_k)_{ij}|)^2 + \frac{1}{2}\sum_{i=1}^{k}(\sum_{j=1}^{i}|(S_{00}^k)_{ij} - (\Sigma_k)_{ij}\sigma_T^2|)^2\\ &\le \sigma_T^4\sum_{i=1}^{k}(\sum_{j=1}^{i}|r_{ij}|)^2 + k^3\|S_{00}^k - \Sigma_k\sigma_T^2\|_2^2,\end{aligned}$$

where $a_T^{-2}\sigma_T^2 \xrightarrow{w} [\mathcal{S}]_1 < \infty$ a.s. by a result in section 4.4 of Resnick (1986), $\sum_{j=1}^{i}|r_{ij}| \le (\sum_{j=0}^{\infty}|\gamma_j|)^2 < \infty$ and $k^3\|S_{00}^k - \Sigma_k\sigma_T^2\|_2^2 = O_P(ka_T^4)$ when $k^2/T \to 0$, from where $\Phi_2 = O_P(ka_T^4)$. Thus, $S_{01}^k = O_P(k^{1/2}a_T^2)$.

Finally, using (A.1), we have $S_{11}^k - \gamma(1)^2 \sum_{t=k+1}^T (\sum_{s=1}^{t-1} \varepsilon_s)^2 = 2\gamma(1)\Upsilon_1 + \Upsilon_2$, with, first,

$$|\Upsilon_1| = |\sum_{t=k+1}^T (\sum_{s=1}^{t-1} \varepsilon_s)(v_t - v_0 + y_0)| \le \max_{t=1,\ldots,T} |\sum_{s=1}^{t-1} \varepsilon_s| \{ \sum_{t=k+1}^T |v_t| + T(|v_0| + |y_0|)\}.$$

Here $\max_{t=1,\ldots,T} |a_T^{-1} \sum_{s=1}^{t-1} \varepsilon_s| \xrightarrow{w} \sup_{[0,1]} |\mathcal{S}| < \infty$ a.s., $\sum_{t=k+1}^T |v_t| = O_P(T)$ for $\alpha > 1$ by Markov's inequality (as $\mathrm{E} \sum_{t=k+1}^T |v_t| \le T \, \mathrm{E} |v_1|$ with $\mathrm{E} |v_1| \le \mathrm{E} |\varepsilon_1| \sum_{i=0}^\infty i |\gamma_i| < \infty$), and $\sum_{t=k+1}^T |v_t| = O_P(a_T)$ for $\alpha \in (0,1]$ by truncation, Markov's inequality and KT (as $\max_{1 \le t \le T} |v_t| = O_P(a_T)$, see above) and $\mathrm{E}(\sum_{t=k+1}^T |v_t| \mathbb{I}_{\{|v_t| < M a_T\}}) \le T \, \mathrm{E}(|v_1| \mathbb{I}_{\{|v_1| < M a_T\}}) = O(a_T l_T)$ for every $M \in \mathbb{R}$), so $\Upsilon_1 = O_P(a_T \max\{T, a_T l_T\})$ with l_T as in (a). Second,

$$|\Upsilon_2| = \sum_{t=k+1}^T (v_t - v_0 + y_0)^2 \le 2 \sum_{t=k+1}^T v_t^2 + 2T(v_0 - y_0)^2 = O_P(a_T^2)$$

since $\sum_{t=k+1}^T v_t^2 = O_P(a_T^2)$ by truncation, Markov's inequality and KT:

$$\mathrm{E}(\sum_{t=k+1}^T v_t^2 \mathbb{I}_{\{|v_t| < M a_T\}}) \le T \, \mathrm{E}(v_1^2 I_{\{|v_1| < M a_T\}}) = O(a_T^2).$$

Combining the magnitude orders of Υ_1 and Υ_2 completes the proof. ■

We now provide a proof of Theorem 1.

Proof of Theorem 1. Let $\check{\boldsymbol{\beta}}_k$ be the OLS estimator of $\boldsymbol{\beta}_k$ from the regression of Δy_t on $\mathbf{X}_{t-1}^k$. It holds that

$$\hat{\phi}_k = \{S_{11}^k - S_{10}^k (S_{00}^k)^{-1} S_{01}^k\}^{-1} \{S_{1\varepsilon}^k - S_{10}^k (\check{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)\},$$

where, for all $\epsilon > 0$, first,

$$\|\check{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\| = O_P(a_k a_T^{\epsilon - 1}) + O_P(1) \sum_{j=k+1}^\infty |\beta_j| = o_P(1) \tag{A.2}$$

by (7.1) of Cavaliere *et al.* (2016a); second,

$$\begin{aligned} S_{10}^k (\check{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k) &= O_P(T^\epsilon k^{1/2} a_k a_T) + O_P(k^{1/2} a_T^2) \sum_{j=k+1}^\infty |\beta_j| \\ &= o_P(a_T^2) + o_P(a_T^3) \sum_{i=k+1}^\infty |\beta_i| \end{aligned}$$

using also Lemma A.1(d) and the condition $k^2/T \to 0$; third,

$$\begin{aligned} S_{11}^k - S_{10}^k (S_{00}^k)^{-1} S_{01}^k &= \gamma(1)^2 \sum_{t=k+1}^T (\sum_{s=1}^{t-1} \varepsilon_s)^2 + o_P(T a_T^2) + O_P(\|S_{10}^k\|^2 \lambda_{\max}\{(S_{00}^k)^{-1}\}) \\ &= \gamma(1)^2 \sum_{t=k+1}^T (\sum_{s=1}^{t-1} \varepsilon_s)^2 + o_P(T a_T^2) \end{aligned}$$

using Lemma A.1(a,d,e) and the fact that $\lambda_{\max}\left((S_{00}^k)^{-1}\right) = (\sum_{t=1}^T \varepsilon_t^2)^{-1}\{\lambda_{\max}(\Sigma_k^{-1}) + o_P(1)\} = O_P(a_T^{-2})$; here $T^{-1}a_T^{-2}\sum_{t=k+1}^T(\sum_{s=1}^{t-1}\varepsilon_s)^2 = T^{-1}a_T^{-2}\sum_{t=1}^T(\sum_{s=1}^{t-1}\varepsilon_s)^2 + o_P(T^{-1}a_T^{-2}ka_k^2) \xrightarrow{w} \int \mathcal{S}^2$ by the CMT with $\int \mathcal{S}^2 > 0$ a.s. (see p.359 of Chan and Tran, 1989). By combining these results with Lemma A.1(c), it follows that

$$\hat{\phi}_k = \gamma(1)^{-2}\{\sum_{t=1}^T(\sum_{s=1}^{t-1}\varepsilon_s)^2\}^{-1}\sum_{t=k+1}^T y_{t-1}\varepsilon_t + o_P(T^{-1}) + o_P(T^{-1}a_T)\sum_{i=k+1}^\infty|\beta_i|. \tag{A.3}$$

Using (A.1), $\sum_{t=k+1}^T y_{t-1}\varepsilon_t = \gamma(1)\sum_{t=1}^T\sum_{s=1}^{t-1}\varepsilon_s\varepsilon_t - \gamma(1)\sum_{t=1}^k\sum_{s=1}^{t-1}\varepsilon_s\varepsilon_t + \sum_{t=k+1}^T v_{t-1}\varepsilon_t + (y_0-\nu_0)\sum_{t=k+1}^T\varepsilon_t$, where (i), $a_k^{-2}\sum_{t=1}^k\sum_{s=1}^{t-1}\varepsilon_s\varepsilon_t \xrightarrow{w} \frac{1}{2}(\mathcal{S}(1)^2 - [\mathcal{S}]_1) = \int \mathcal{S}d\mathcal{S}$ by partial and integration summation, the joint convergence $(a_k^{-1}\sum_{t=1}^{\lfloor k\cdot\rfloor}\varepsilon_t, a_k^{-2}\sum_{t=1}^{\lfloor k\cdot\rfloor}\varepsilon_t^2) \xrightarrow{w} (\mathcal{S}(\cdot), [\mathcal{S}]_{(\cdot)})$ in $D_2[0,1]$ from section 4.4. of Resnick (1986), and the CMT, (ii), $a_T^{-2}\sum_{t=k+1}^T v_{t-1}\varepsilon_t \xrightarrow{P} 0$ by Theorem 4.2 of Davis and Resnick (1985a), and (iii), $a_T^{-1}\sum_{t=k+1}^T\varepsilon_t = a_T^{-1}\sum_{t=1}^T\varepsilon_t + o_P(a_T^{-1}a_k) \xrightarrow{w} \mathcal{S}(1)$, so $a_T^{-2}\sum_{t=k+1}^T y_{t-1}\varepsilon_t = a_T^{-2}\gamma(1)\sum_{t=1}^T\sum_{s=1}^{t-1}\varepsilon_s\varepsilon_t + o_P(1) \xrightarrow{w} \gamma(1)\int \mathcal{S}d\mathcal{S}$ and

$$\hat{\phi}_k = O_P(T^{-1}) + o_P(T^{-1}a_T)\sum_{i=k+1}^\infty|\beta_i| = o_P(1).$$

Finally,

$$\begin{aligned}\|\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\| &= \|\check{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k + \hat{\phi}_k(S_{00}^k)^{-1}S_{01}^k\| \le \|\check{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\| + |\hat{\phi}_k|\|S_{10}^k\|\lambda_{\max}\{(S_{00}^k)^{-1}\} \quad (A.4)\\ &= \|\check{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\| + |\hat{\phi}_k|O_P(k^{1/2})\\ &= \|\check{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\| + O_P(k^{1/2}/T) + o_P((k^{1/2}/T)a_T)\sum_{i=k+1}^\infty|\beta_i| = o_P(a_T^{-\epsilon})\end{aligned}$$

for sufficiently small $\epsilon > 0$, by (A.2) and since $a_T\sum_{i=k+1}^\infty|\beta_i| \to 0$. ■

Proof of Theorem 2. From (A.3) and its discussion, under $a_T\sum_{i=k+1}^\infty|\beta_i| \to 0$,

$$T\hat{\phi}_k = \gamma(1)^{-1}\{\sum_{t=1}^T(\sum_{s=1}^{t-1}\varepsilon_s)^2\}^{-1}\sum_{t=1}^T\sum_{s=1}^{t-1}\varepsilon_s\varepsilon_t + o_P(1) \xrightarrow{w} \frac{1}{\gamma(1)}\frac{\int \mathcal{S}d\mathcal{S}}{\int \mathcal{S}^2},$$

where $T^{-2}\sum_{t=k+1}^T(\sum_{s=1}^{t-1}\varepsilon_s)^2 \xrightarrow{w} \int \mathcal{S}^2$ and $a_T^{-2}\sum_{t=1}^T\sum_{s=1}^{t-1}\varepsilon_s\varepsilon_t \xrightarrow{w} \gamma(1)\int \mathcal{S}d\mathcal{S}$ and . Further,

$$\begin{aligned}|1 - \sum_{i=1}^k\hat{\beta}_i - \frac{1}{\gamma(1)}| &= |\sum_{i=1}^k\hat{\beta}_i - \sum_{i=1}^\infty\beta_i| \le \sum_{i=1}^k|\hat{\beta}_i - \beta_i| + \sum_{i=k+1}^\infty|\beta_i|\\ &\le \sqrt{k}\|\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\| + \sum_{i=k+1}^\infty|\beta_i| \to 0\end{aligned}$$

using (A.4) with $k/T \to 0$ and $a_T\sum_{i=k+1}^\infty|\beta_i| \to 0$. The convergence of $R_T = T_k\hat{\phi}_k/(1-\sum_{i=1}^k\hat{\beta}_i)$ obtains from the two previous displays. The convergence of Q_T uses also the fact that $a_T^{-2}\sigma_T^2 \xrightarrow{w} [\mathcal{S}]_1$ jointly with R_T; see the proof of Theorem 1. ■

We now turn to convergence (13). It holds that $\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k = \check{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k - \hat{\phi}_k(S_{00}^k)^{-1}S_{01}^k$, where under the unit root null,

$$S_{01}^k = \mathbf{i}_k\sum_{t=k}^{T-1}y_{t-1}u_t + U_kS_{00}^k\mathbf{e}_k + o_P(a_T^2),$$

U_k is lower triangular with ones on and below the main diagonal, and $\mathbf{e}_k$ is the first canonical base vector in $\mathbb{R}^k$. Using also Lemma A.1(a), upon substitution,

$$L_k(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k - d_T) = L_k(\check{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k - d_T) - \hat{\phi}_k L_k\{\Sigma_k^{-1}\mathbf{i}_k(\sum_{t=1}^{T}\varepsilon_t^2)^{-1}\sum_{t=k}^{T-1} y_{t-1}u_t + \Sigma_k^{-1}U_k\Sigma_k\mathbf{e}_k\} + o_P(T^{-1}),$$

where the first term after the equality sign has magnitude order $\tilde{a}_T a_T^{-2}$ with asymptotics given in Lemma 1 and Theorem 2 of Cavaliere *et al.* (2016a), whereas the term involving $\hat{\phi}_k$ has magnitude order T^{-1}:

$$T\hat{\phi}_k L_k\{\Sigma_k^{-1}\mathbf{i}_k(a_T^{-2}\sum_{t=1}^{T}\varepsilon_t^2)^{-1}(a_T^{-2}\sum_{t=k}^{T-1} y_{t-1}u_t) + \Sigma_k^{-1}U_k\Sigma_k\mathbf{e}_k\}$$
$$\xrightarrow{w} \frac{\int \mathcal{S}d\mathcal{S}}{\gamma(1)\int \mathcal{S}^2} L\Big(\Sigma^{-1}\mathbf{i}\gamma(1)^2\frac{\int \mathcal{S}d\mathcal{S}}{[\mathcal{S}]_1} - \Sigma^{-1}\mathbf{i}\sum_{j=0}^{\infty}\gamma_j^*\gamma_j + \Sigma^{-1}U\Sigma\mathbf{e}\Big).$$

As $U\Sigma\mathbf{e} = \Sigma\mathbf{i} - (\sum_{i=1}^{\infty} r_{i0})\mathbf{i} = \Sigma\mathbf{i} + (\sum_{j=0}^{\infty}\gamma_j^*\gamma_j)\mathbf{i}$ and $\gamma_0 = 1$**,** the previous limit is the negative of the limit in (13).

A.2 Bootstrap Asymptotics

A.2.1 Proof of Theorem 4

Proof of Theorem 4. We follow the triangular scheme outlined in Remark 4.2.5.

Denote the finite-sample cumulative process of g_T under $P_{|\varepsilon|}$ by $\Phi_T(\cdot) := P_{|\varepsilon|}(g_T \le \cdot)$. Hypothesis (i) implies that

$$\rho\left(\Phi_T(\cdot), P_{|\varepsilon|}(\gamma_T(\boldsymbol{\varepsilon}_T) \le \cdot)\right) = o_P(1) \text{ and } \rho(P_{Z,\varepsilon}(g_T^* \le \cdot), P_{Z,\varepsilon}(\gamma_T(\boldsymbol{\varepsilon}_T^\dagger) \le \cdot)) = o_P(1),$$

where ρ is Lévy distance. Moreover, under symmetry of $\boldsymbol{\varepsilon}_T$, and the Rademacher and independence assumption about $\mathbf{w}_T$, it holds that $\mathcal{L}_{|\varepsilon|}(\boldsymbol{\varepsilon}_T) = \mathcal{L}_{\varepsilon}(\boldsymbol{\varepsilon}_T^\dagger) = \mathcal{L}_{Z,\varepsilon}(\boldsymbol{\varepsilon}_T^\dagger)$ a.s., with $\mathcal{L}_{(\cdot)}$ for law conditional on $(\cdot)$, the last equality because $\boldsymbol{\varepsilon}_T^\dagger$ is independent of $\mathbf{Z}_T$ given $\boldsymbol{\varepsilon}_T$. By the triangle inequality, $\rho\left(\Phi_T(\cdot), P_{Z,\varepsilon}(g_T^* \le \cdot)\right) = o_P(1)$. As g_T^* is independent of $\boldsymbol{\varepsilon}_T$ given $\mathbf{Z}_T$, also $\rho\left(\Phi_T(\cdot), P_Z(g_T^* \le \cdot)\right)$. The same holds for the Skorokhod distance in $D[0,1]$, say d_S, see Daley and Vere-Jones (2008, pp.143-144).

We now pass to uniform distance. Under hypothesis (ii), let the limiting cumulative process of $\gamma_T(\boldsymbol{\varepsilon}_T)$ under $P_{|\varepsilon|}$ be $\Phi(\cdot) = \phi((-\infty,\cdot))$ such that $P_{|\varepsilon|}(\gamma_T(\boldsymbol{\varepsilon}_T) \le \cdot) \xrightarrow{w} \Phi(\cdot)$ in $D[0,1]$; under hypothesis (i), Φ is also the limiting cumulative process of g_T under $P_{|\varepsilon|}$, such that $\Phi_T \xrightarrow{w} \Phi$ in $D[0,1]$. As Φ is a.s. continuous, having $d_S(\Phi_T(\cdot), P_Z(g_T^* \le \cdot)) = o_P(1)$ implies that $\sup_{x\in\mathbb{R}}|\Phi_T(x) - P_Z(g_T^* \le x)| = o_P(1)$. Therefore, $P_Z(g_T^* \le g_T) = \Phi_T(g_T) + o_P(1)$.

Further, define the quantile transformation using the right-continuous version of the generalized inverse. As the quantile transformation is continuous in the Lévy, and hence, in the Skorokhod metric, it holds that $(\Phi_T, \Phi_T^{-1}) \xrightarrow{w} (\Phi, \Phi^{-1})$ in $D_2[0,1]$. For $\theta \in [0,1]$,

$$\begin{aligned} P_{|\varepsilon|}\left(\Phi_T(g_T) \ge \theta\right) &= P_{|\varepsilon|}(g_T \ge \Phi_T^{-1}(\theta)) = 1 - P_{|\varepsilon|}(g_T < \Phi_T^{-1}(\theta)) \\ &= 1 - \Phi_T(\Phi_T^{-1}(\theta)-) \xrightarrow{w} 1 - \Phi(\Phi^{-1}(\theta)) = \theta \end{aligned}$$

using the continuity of Φ, and the same holds in probability as the limit is a constant. By the Bounded convergence theorem, integration over $|\varepsilon_T|$ yields $P\left(\Phi_T(g_T) \ge \theta\right) \to \theta$ for $\theta \in [0,1]$. Therefore, $\Phi_T(g_T) \xrightarrow{w} U[0,1]$. Since $P_Z(g_T^* \le g_T) = \Phi_T(g_T) + o_P(1)$, equation (16) follows. ■

A.2.2 Condition (i) of Theorem 4

Our plan in what follows is to check that the conditions of Theorem 4 are satisfied by the statistics R_T and R_T^* in place of g_T and g_T^*. We start from a conditional version of the large-sample expansion of $\hat{\phi}_k$.

Lemma A.2 *Under the conditions of Theorem 2, it holds in P-probability that*

$$T\hat{\phi}_k = T\gamma(1)^{-1}\{\sum_{t=1}^{T}(\sum_{s=1}^{t-1}\varepsilon_s)^2\}^{-1}\sum_{t=1}^{T}(\sum_{s=1}^{t-1}\varepsilon_s)\varepsilon_t + o_{P_{|\varepsilon|}}(1), \quad \text{(A.5)}$$

i.e., conditionally on $|\boldsymbol{\varepsilon}_T| = (|\varepsilon_1|, ..., |\varepsilon_T|)$.

PROOF. As in the proof of Theorem 2, expansion (A.5) holds with remainder term, say r_T, such that $r_T = o_P(1)$. But then $r_T = o_{P_{|\varepsilon|}}(1)$ as well. Indeed, as $\mathbb{I}\{|r_T| \geq \eta\} \xrightarrow{P} 0$ for any $\eta > 0$, it follows that $P_{|\varepsilon|}(|r_T| \geq \eta) \xrightarrow{P} 0$ by the Bounded convergence theorem for conditional expectations. This proves the lemma. ■

We now turn to the derivation of a similar expansion of the bootstrap estimator

$$\phi^* = \{S_{11}^* - S_{10}^* S_{00}^{*-1} S_{01}^*\}^{-1}\{S_{1\varepsilon}^* - S_{10}^* S_{00}^{*-1} S_{0\varepsilon}^*\},$$

where $S_{1\varepsilon}^* := \sum_{t=k+1}^{T} y_{t-1}^* \varepsilon_t^*$, $S_{0\varepsilon}^{*k} := \sum_{t=k+1}^{T} \mathbf{X}_{t-1}^{*k} \varepsilon_t^*$, $S_{01}^{*k} := \sum_{t=k+1}^{T} \mathbf{X}_{t-1}^{*k} y_{t-1}^* =: (S_{10}^*)'$, $S_{11}^* := \sum_{t=k+1}^{T} (y_{t-1}^*)^2$ and $S_{00}^{*k} := \sum_{t=k+1}^{T} \mathbf{X}_{t-1}^{*k}(\mathbf{X}_{t-1}^{*k})'$. To this aim, jointly with the actual bootstrap errors $\varepsilon_t^* = w_t \hat{\varepsilon}_t$, we consider the benchmark errors $\varepsilon_t^\dagger = w_t \varepsilon_t$ with associated (infeasible) bootstrap data $y_t^\dagger = 0$ $(t = 0, ..., k)$, $\Delta y_t^\dagger = \sum_{i=0}^{t-k-1} \gamma_i \varepsilon_{t-i}^\dagger$ $(t = k+1, ..., T)$ and $y_t^\dagger = \sum_{s=k+1}^{t} \Delta y_t^\dagger$ $(t = k+1, ..., T)$. With $\mathbf{\Delta Y}_{t-1}^{k\dagger} := (\Delta y_{t-1}^\dagger, ..., \Delta y_{t-k}^\dagger)' = \sum_{i=0}^{t-k-2} \boldsymbol{\gamma}_{i:k} \varepsilon_{t-i-1}^\dagger$, we define $S_{1\varepsilon}^\dagger := \sum_{t=k+1}^{T} y_{t-1}^\dagger \varepsilon_t^\dagger$, $S_{0\varepsilon}^\dagger := \sum_{t=k+1}^{T} \mathbf{\Delta Y}_{t-1}^{k\dagger} \varepsilon_t^\dagger =: (S_{\varepsilon 0}^\dagger)'$, $S_{01}^\dagger := \sum_{t=k+1}^{T} \mathbf{\Delta Y}_{t-1}^{k\dagger} y_{t-1}^\dagger =: (S_{10}^\dagger)'$, $S_{11}^\dagger := \sum_{t=k+1}^{T} y_{t-1}^{\dagger 2}$ and $S_{00}^\dagger := \sum_{t=k+1}^{T} \mathbf{\Delta Y}_{t-1}^{k\dagger}(\mathbf{\Delta Y}_{t-1}^{k\dagger})'$. Let $P_\dagger$ denote probability conditional on the data and $\boldsymbol{\varepsilon}_T$, and $(\cdot)_T = o_{P_\dagger}(1)$ in P-probability mean that $P_\dagger(|(\cdot)_T| > \omega) \xrightarrow{P} 0$ as $T \to \infty$ for every $\omega > 0$.

Lemma A.3 *Let* $k^3/T + 1/k \to 0$ *and* $a_T \sum_{i=k+1}^{\infty} |\beta_i| \to 0$ *as* $T \to \infty$*. Then the following r.v.'s are* $o_{P_\dagger}(1)$ *in P-probability: (a)* $a_T^{-2}|S_{1\varepsilon}^* - S_{1\varepsilon}^\dagger|$*, (b)* $a_T^{-2}|S_{1\varepsilon}^\dagger - \gamma(1)\sum_{t=k+1}^{T}(\sum_{s=k+1}^{t-1}\varepsilon_s^\dagger)\varepsilon_t^\dagger|$*, (c)* $a_T^{-2}T^{-1}|S_{11}^* - S_{11}^\dagger|$*, (d)* $a_T^{-2}T^{-1}|S_{11}^\dagger - \gamma(1)^2\sum_{t=k+1}^{T}(\sum_{s=k+1}^{t-1}\varepsilon_s^\dagger)^2|$*, (e)* $a_T^{-2}|S_{10}^\dagger(S_{00}^\dagger)^{-1}S_{0\varepsilon}^\dagger|$*, (f)* $a_T^{-2}|S_{10}^* S_{00}^{*-1} S_{0\varepsilon}^*|$*, (g)* $a_T^{-2}T^{-1}|S_{10}^\dagger(S_{00}^\dagger)^{-1}S_{01}^\dagger|$*, (h)* $a_T^{-2}T^{-1}|S_{10}^* S_{00}^{*-1} S_{01}^*|$*. As a consequence, it holds in P-probability that*

$$T\phi_k^* = T\gamma(1)^{-1}\{\sum_{t=1}^{T}(\sum_{s=1}^{t-1}\varepsilon_s^\dagger)^2\}^{-1}\sum_{t=1}^{T}(\sum_{s=1}^{t-1}\varepsilon_s^\dagger)\varepsilon_t^\dagger + o_{P_\dagger}(1).$$

PROOF. We give the proof for $\tilde{\phi} = \hat{\phi}$ and $\hat{\boldsymbol{\beta}}_k = \tilde{\boldsymbol{\beta}}_k$; the restricted case is analogous. The proof uses some known magnitude orders like $\|\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\| = o_P(a_k a_T^{\epsilon-1} + k^{1/2}T^{-1})$ for $\epsilon > 0$ (see eqs. (A.2)-(A.4)) and $\hat{\phi}_k = O_P(T^{-1})$ that hold under the assumptions of the lemma.

Let $\varepsilon_{t,k} = \varepsilon_t + \sum_{i=k+1}^{\infty} \beta_i u_{t-i}$ so that $\Delta y_t = \boldsymbol{\beta}_k' \mathbf{X}_{t-1}^k + \varepsilon_{t,k}$ under the null hypothesis. From $(\hat{\varepsilon}_t - \varepsilon_{t,k})^2 \leq 2\hat{\phi}_k^2 y_{t-1}^2 + 2\{(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)' \mathbf{X}_{t-1}^k\}^2$ and Lemma A.1 2(a,e) we find that

$$\sum_{t=k+1}^{T} (\hat{\varepsilon}_t - \varepsilon_{t,k})^2 \leq 2\hat{\phi}_k^2 S_{11}^k + 2(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)' S_{00}^k (\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k) = O_P(a_T^2\{T^{-1} + \|\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\|^2\}),$$

$$\sum_{t=k+1}^{T} (\hat{\varepsilon}_t - \varepsilon_t)^2 \leq 2\sum_{t=k+1}^{T} (\hat{\varepsilon}_t - \varepsilon_{t,k})^2 + 2S_{\rho\rho}^k = o_P(a_T^2), \tag{A.6}$$

$$\hat{\sigma}_{Tk}^2 := \sum_{t=k+1}^{T} \hat{\varepsilon}_t^2 \leq 2\sigma_T^2 + 2\sum_{t=1}^{T} (\hat{\varepsilon}_t - \varepsilon_t)^2 = O_P(a_T^2), \tag{A.7}$$

where $S_{\rho\rho}^k := \sum_{t=k+1}^{T} (\sum_{i=k+1}^{\infty} \beta_i u_{t-i})^2 = o_P(a_T^2)$ by the proof of Lemma 3 of Cavaliere *et al.* (2016a) provided in the supplement thereof. Further, let $\hat{\gamma}_i$ be defined by $(1 - \sum_{i=1}^{k} \hat{\beta}_i z^i)(1 + \sum_{i=1}^{\infty} \hat{\gamma}_i z^i) = 1$, where well-definition is guaranteed with P-probability approaching one. For $t = k+2, ..., T$ and $i = 1, ..., t-k-1$, we obtain

$$\begin{aligned}(\hat{\varepsilon}_{t-i} \sum_{j=0}^{i-1} \hat{\gamma}_j - \varepsilon_{t-i} \sum_{j=0}^{i-1} \gamma_j)^2 &\leq 2(\hat{\varepsilon}_{t-i} - \varepsilon_{t-i})^2 (\sum_{j=0}^{i-1} \hat{\gamma}_j)^2 + 2\varepsilon_{t-i}^2 \{\sum_{j=0}^{i-1} (\hat{\gamma}_j - \gamma_j)\}^2 \\ &\leq 2(\hat{\varepsilon}_{t-i} - \varepsilon_{t-i})^2 (\sum_{j=0}^{T-k-2} |\hat{\gamma}_j|)^2 + 2\varepsilon_{t-i}^2 (\sum_{j=1}^{T-k-2} |\hat{\gamma}_j - \gamma_j|)^2\end{aligned} \tag{A.8}$$

where $\sum_{j=0}^{T-k-2} |\hat{\gamma}_j| \leq \sum_{j=0}^{\infty} |\gamma_j| + \sum_{j=1}^{T-k-2} |\hat{\gamma}_j - \gamma_j|$ and $\sum_{j=1}^{T-k-2} |\hat{\gamma}_j - \gamma_j| = o_P(1)$ as in (S.4.14) of the supplement to Cavaliere *et al.* (2016a), so using (A.6),

$$\max_{k+1 \leq t \leq T} \sum_{i=1}^{t-k-1} (\hat{\varepsilon}_{t-i} \sum_{j=0}^{i-1} \hat{\gamma}_j - \varepsilon_{t-i} \sum_{j=0}^{i-1} \gamma_j)^2 \leq O_P(1) \sum_{t=k+1}^{T} (\hat{\varepsilon}_t - \varepsilon_t)^2 + o_P(1)\, \sigma_T^2 = o_P(a_T^2). \tag{A.9}$$

Regarding part (a), $S_{1\varepsilon}^* - S_{1\varepsilon}^\dagger = \sigma_1 + \sigma_2$ with $\sigma_1 = \sum_{t=k+2}^{T} (\varepsilon_t^* - \varepsilon_t^\dagger) y_{t-1}^*$ and $\sigma_2 = \sum_{t=k+2}^{T} \varepsilon_t^\dagger (y_{t-1}^* - y_{t-1}^\dagger)$. For expressions of the form $\sigma = \sum_{t=k+2}^{T} w_t e_t \sum_{i=1}^{t-k-1} w_{t-i} g_{ti}$, where (e_t, g_{ti}) are measurable transformations of $|\boldsymbol{\varepsilon}_T|$ and the data, it holds that

$$\mathrm{E}_\dagger\, \sigma = 0,\ \mathrm{Var}_\dagger\, \sigma = \sum_{t=k+2}^{T} e_t^2 \sum_{i=1}^{t-k-1} g_{ti}^2. \tag{A.10}$$

To apply (A.10) to $\sigma_{1,2}$, we write $y_{t-1}^* = \sum_{s=k+1}^{t-1} \sum_{i=0}^{s-k-1} \hat{\gamma}_i \varepsilon_{s-i}^* = \sum_{i=1}^{t-k-1} \varepsilon_{t-i}^* \sum_{j=0}^{i-1} \hat{\gamma}_j$ and $y_{t-1}^* - y_{t-1}^\dagger = \sum_{i=1}^{t-k-1} w_{t-i} (\hat{\varepsilon}_{t-i} \sum_{j=0}^{i-1} \hat{\gamma}_j - \varepsilon_{t-i} \sum_{j=0}^{i-1} \gamma_j)$. Then $\mathrm{E}_\dagger\, \sigma_1 = \mathrm{E}_\dagger\, \sigma_2 = 0$,

$$\begin{aligned}\mathrm{Var}_\dagger\, \sigma_1 &= \sum_{t=k+2}^{T} (\hat{\varepsilon}_t - \varepsilon_t)^2 \sum_{i=1}^{t-k-1} \hat{\varepsilon}_{t-i}^2 (\sum_{j=0}^{i-1} \hat{\gamma}_j)^2 \leq \hat{\sigma}_{Tk}^2 (\sum_{j=0}^{T-k-2} |\hat{\gamma}_j|)^2 \sum_{t=k+2}^{T} (\hat{\varepsilon}_t - \varepsilon_t)^2 = o_P(a_T^4), \\ \mathrm{Var}_\dagger\, \sigma_2 &= \sum_{t=k+2}^{T} \varepsilon_t^2 \sum_{i=1}^{t-k-1} (\hat{\varepsilon}_{t-i} \sum_{j=0}^{i-1} \hat{\gamma}_j - \varepsilon_{t-i} \sum_{j=0}^{i-1} \gamma_j)^2 \leq \sigma_T^2 o_P(a_T^2) = o_P(a_T^4)\end{aligned}$$

using (A.6), (A.7) and (A.9). This, and Chebyshev's inequality, proves item (a).

For part (b), let $\sigma_3 := S^\dagger_{1\varepsilon} - \gamma(1)\sum_{t=k+1}^T(\sum_{s=k+1}^{t-1}\varepsilon^\dagger_s)\varepsilon^\dagger_t = \sum_{t=k+1}^T(\sum_{i=0}^{t-k-2}\gamma^*_i\varepsilon^\dagger_{t-i-1})\varepsilon^\dagger_t$. It holds that $\mathrm{E}_\dagger\,\sigma_3^2 = \sum_{t=k+1}^T\varepsilon_t^2\sum_{i=0}^{t-k-2}(\gamma^*_i)^2\varepsilon^2_{t-i-1} = o_P(a_T^4)$ by Markov's inequality, since for all $\eta\in[\delta,\alpha)$,

$$\mathrm{E}\,|\,\mathrm{E}_\dagger\,\sigma_3^2|^{\eta/2} \le T(\mathrm{E}\,|\varepsilon_1|^\eta)^2\sum_{i=0}^\infty i|\gamma_i|^\eta = O\left(T\right).$$

Therefore, by Chebyshev's inequality, $a_T^{-2}\,|\sigma_3| = o_{P_\dagger}(1)$ in P-probability.

Next, $S^*_{11}-S^\dagger_{11} = \rho_1+\rho_2$ with $\rho_1 = \sum_{t=k+2}^T y^*_{t-1}(y^*_{t-1}-y^\dagger_{t-1})$ and $\rho_2 = \sum_{t=k+2}^T y^\dagger_{t-1}(y^*_{t-1}-y^\dagger_{t-1})$. For an expression of the form $\rho = \sum_{t=k+2}^T\{\sum_{i=1}^{t-k-1}w_{t-i}e_{t-i}a_i\sum_{i=1}^{t-k-1}w_{t-i}g_{t-i,i}\} = \sum_{s,t=k+1}^{T-1}w_sw_te_s\sum_{i=\max(s,t)+1}^T a_{i-s}g_{t,i-t}$, where $(e_t,a_i,g_{t,i})$ are measurable functions of $\boldsymbol{\varepsilon}_T$ and the data, it holds that

$$\mathrm{E}_\dagger\,\rho = \sum_{t=k+2}^T\sum_{i=1}^{t-k-1}e_{t-i}a_ig_{t-i,i},\ \mathrm{Var}_\dagger\,\rho \le 2\sum_{s,t=k+1}^{T-1}e_s^2\left(\sum_{i=\max(s,t)+1}^T a_{i-s}g_{t,i-t}\right)^2. \qquad \text{(A.11)}$$

We find

$$\begin{aligned}
|\,\mathrm{E}_\dagger\,\rho_1| &= \left|\sum_{t=k+2}^T\sum_{i=1}^{t-k-1}\{\hat\varepsilon_{t-i}(\sum_{j=0}^{i-1}\hat\gamma_j)(\hat\varepsilon_{t-i}\sum_{j=0}^{i-1}\hat\gamma_j - \varepsilon_{t-i}\sum_{j=0}^{i-1}\gamma_j)\}\right| \\
&\le (\sum_{j=0}^{T-k-2}|\hat\gamma_j|)\sum_{t=k+2}^T\sum_{i=1}^{t-k-1}\left|\hat\varepsilon_{t-i}(\hat\varepsilon_{t-i}\sum_{j=0}^{i-1}\hat\gamma_j - \varepsilon_{t-i}\sum_{j=0}^{i-1}\gamma_j)\right| \\
\text{(Cauchy-Schwartz)}\quad &\le (\sum_{j=0}^{T-k-2}|\hat\gamma_j|)\hat\sigma_{Tk}\sum_{t=k+2}^T\{\sum_{i=1}^{t-k-1}(\hat\varepsilon_{t-i}\sum_{j=0}^{i-1}\hat\gamma_j - \varepsilon_{t-i}\sum_{j=0}^{i-1}\gamma_j)^2\}^{1/2} \\
&\le (\sum_{j=0}^{T-k-2}|\hat\gamma_j|)\hat\sigma_{Tk}o_P(Ta_T) = o_P(Ta_T^2)
\end{aligned}$$

using (A.9), and similarly, $|\,\mathrm{E}_\dagger\,\rho_2| \le \sum_{j=0}^\infty|\gamma_j|)\sigma_{Tk}o_P(Ta_T) = o_P(Ta_T^2)$. Regarding variances,

$$\begin{aligned}
\mathrm{Var}_\dagger\,\rho_1 &\le 2\sum_{s,t=k+1}^{T-1}\hat\varepsilon_s^2\left(\sum_{i=\max(s,t)+1}^T(\sum_{j=0}^{i-s-1}\hat\gamma_j)(\hat\varepsilon_t\sum_{j=0}^{i-t-1}\hat\gamma_j - \varepsilon_t\sum_{j=0}^{i-t-1}\gamma_j)\right)^2 \\
\text{(Cauchy-Schwartz)}\quad &\le 2T(\sum_{j=0}^{T-k-2}|\hat\gamma_j|)^2\sum_{s,t=k+1}^{T-1}\hat\varepsilon_s^2\sum_{i=t+1}^T(\hat\varepsilon_t\sum_{j=0}^{i-t-1}\hat\gamma_j - \varepsilon_t\sum_{j=0}^{i-t-1}\gamma_j)^2 \\
\text{cf. eq. (A.8)}\quad &\le 4T^2(\sum_{j=0}^{T-k-2}|\hat\gamma_j|)^2\sum_{s,t=k+1}^{T-1}\hat\varepsilon_s^2\left\{(\hat\varepsilon_t-\varepsilon_t)^2(\sum_{j=0}^{T-k-2}|\hat\gamma_j|)^2 + \varepsilon_t^2(\sum_{j=0}^{T-k-2}|\hat\gamma_j-\gamma_j|)^2\right\} \\
&= 4T^2(\sum_{j=0}^{T-k-2}|\hat\gamma_j|)^2\hat\sigma_{Tk}^2\left\{\sum_{t=k+1}^T(\hat\varepsilon_t-\varepsilon_t)^2(\sum_{j=0}^{T-k-2}|\hat\gamma_j|)^2 + \sigma_T^2(\sum_{j=1}^{T-k-2}|\hat\gamma_j-\gamma_j|)^2\right\} \\
&= o_P(T^2a_T^4)
\end{aligned}$$

using (A.6) and (A.7), and analogously,

$$\mathrm{Var}_\dagger\,\rho_2 \le 4T^2(\sum_{j=0}^{T-k-2}|\gamma_j|)^2\sigma_T^2\left\{\sum_{t=k+1}^T(\hat\varepsilon_t-\varepsilon_t)^2(\sum_{j=0}^{T-k-2}|\hat\gamma_j|)^2 + \sigma_T^2(\sum_{j=1}^{T-k-2}|\hat\gamma_j-\gamma_j|)^2\right\} = o(T^2a_T^4).$$

In part (d), $\rho_3 := S_{11}^{\dagger} - \gamma(1)^2 \sum_{t=k+1}^{T} (\sum_{s=k+1}^{t-1} \varepsilon_s^{\dagger})^2$ can be written as

$$\rho_3 = \sum_{t=k+2}^{T} \{ \sum_{i=0}^{t-k-2} \gamma_i^* \varepsilon_{t-i-1}^{\dagger} \sum_{i=0}^{t-k-2} [\gamma_i^* + 2\gamma(1)] \varepsilon_{t-i-1}^{\dagger} \}.$$

As the sequence $\{\gamma_i^* + 2\gamma(1)\}_i$ is bounded, there is an $M < \infty$ such that

$$\begin{aligned} |\mathrm{E}_{\dagger} \rho_3| &= \left| \sum_{t=k+2}^{T} \sum_{i=0}^{t-k-2} \gamma_i^* [\gamma_i^* + 2\gamma(1)] \varepsilon_{t-i-1}^2 \right| \leq M \sum_{t=k+2}^{T} \sum_{i=0}^{t-k-2} \varepsilon_{t-i-1}^2 |\gamma_i^*| \\ &\leq M \sigma_T^2 \sum_{i=1}^{\infty} i |\gamma_i| = O_P(a_T^2), \\ \mathrm{Var}_{\dagger} \rho_3 &\leq 2 \sum_{s,t=k+1}^{T-1} \varepsilon_s^2 \varepsilon_t^2 \left(\sum_{i=\max(s,t)+1}^{T} \gamma_{i-s}^* [\gamma_{i-t-1}^* + 2\gamma(1)] \right)^2 \\ &\leq 2M^2 \sum_{s,t=1}^{T-1} \varepsilon_s^2 \varepsilon_t^2 (\sum_{i=0}^{\infty} |\gamma_i^*|)^2 \leq 2M \sigma_T^4 (\sum_{i=1}^{\infty} i |\gamma_i|)^2 = O_P(a_T^4), \end{aligned}$$

so $a_T^{-2} |\rho_3| = O_{P_{\dagger}}(1)$ in P-probability.

We proceed to $S_{00}^{\dagger}$,$S_{0\varepsilon}^{\dagger}$, $S_{10}^{\dagger}$ and the respective $S_{\cdot\cdot}^{*}$'s. Under the assumptions $k^3/T + 1/k \to 0$ and $a_T \sum_{i=k+1}^{\infty} |\beta_i| \to 0$, using the fact that $\|\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\| = o_P(a_k a_T^{\epsilon-1} + k^{1/2} T^{-1})$ for $\epsilon > 0$, it follows as in Lemma 3(a) of Cavaliere *et al.* (2016a) that, in P-probability, $(S_{00}^{\dagger})^{-1}$ exists with $P_{\dagger}$-probability approaching one, $\lambda_{\max}((S_{00}^{\dagger})^{-1}) = O_{P_{\dagger}}(a_T^{-2})$ and likewise for S_{00}^*. Additionally, $\lambda_{\max}(S_{00}^{\dagger}) = O_{P_{\dagger}}(a_T^{2+\epsilon})$ in P-probability for all $\epsilon > 0$ since

$$S_{00}^{\dagger} = \sum_{t=k+1}^{T} \sum_{i=1}^{t-k-1} \boldsymbol{\gamma}_{i-1:k} \boldsymbol{\gamma}_{i-1:k}' \varepsilon_{t-i}^2 + \sum_{t=k+1}^{T} \sum_{i \neq j=1}^{t-k-1} \boldsymbol{\gamma}_{i-1:k} \boldsymbol{\gamma}_{i-1:k}' \varepsilon_{t-i}^{\dagger} \varepsilon_{t-j}^{\dagger} =: \Lambda_1 + \Lambda_2,$$

where $\|\Lambda_1 - \sigma_T^2 \Sigma_k\|_2 = o_P(a_T^2)$ as in the proof of Lemma 1(a) of Cavaliere *et al.* (2016a), so $\lambda_{\max}(\Lambda_1) = \lambda_{\max}(\sigma_T^2 \Sigma_k) + o_P(a_T^2) = O_P(a_T^2)$ because $\lambda_{\max}(\Sigma_k)$ is bounded, whereas $\|\Lambda_2\|_2 = o_{P_{\dagger}}(a_T^{2+\epsilon})$ in P-probability for all $\epsilon > 0$ by Markov's inequality, since

$$\begin{aligned} \mathrm{E}(\mathrm{E}_{\dagger} \|\Lambda_2\|^2)^{\eta/2} &\leq (2 \sum_{s,t=k+1}^{T} \sum_{i,j=1+\max\{s-t,0\}}^{t-k-1} \mathbb{I}_{i \neq j} \|\boldsymbol{\gamma}_{i-1:k}\| \|\boldsymbol{\gamma}_{j-1:k}\| \|\boldsymbol{\gamma}_{s-t+i-1:k}\| \|\boldsymbol{\gamma}_{s-t+j-1:k}\| \varepsilon_{t-i}^2 \varepsilon_{t-j}^2)^{\eta/2} \\ &\leq 2^{\eta/2} (\mathrm{E} |\varepsilon_1|^{\eta})^2 \sum_{t=k+1}^{T} \sum_{i,j=0}^{t-k-2} \|\boldsymbol{\gamma}_{i:k}\|^{\eta/2} \|\boldsymbol{\gamma}_{j:k}\|^{\eta/2} \sum_{s=\max\{-i,-j\}}^{T-k-2-\max\{i,j\}} \|\boldsymbol{\gamma}_{s+i:k}\|^{\eta/2} \|\boldsymbol{\gamma}_{s+j:k}\|^{\eta/2} \\ &\leq 2^{\eta/2} (\mathrm{E} |\varepsilon_1|^{\eta})^2 \sum_{t=k+1}^{T} \sum_{i,j=0}^{t-k-2} \|\boldsymbol{\gamma}_{i:k}\|^{\eta/2} \|\boldsymbol{\gamma}_{j:k}\|^{\eta/2} \sum_{s=0}^{T-k-2} \|\boldsymbol{\gamma}_{s:k}\|^{\eta} \\ &= O(T) (\sum_{i=0}^{T-k-2} \|\boldsymbol{\gamma}_{i:k}\|^{\eta/2})^2 (\sum_{s=0}^{T-k-2} \|\boldsymbol{\gamma}_{s:k}\|^{\eta}) = O\left(T k^3\right) = o(T^2) \end{aligned}$$

for $\eta \in [\delta, \alpha)$ under $k^3/T \to 0$.

Using (A.10), for $S^{\dagger}_{0\varepsilon} = \sum_{t=k+1}^{T}(\sum_{i=1}^{t-k-1}\boldsymbol{\gamma}_{i-1:k}\varepsilon^{\dagger}_{t-i})\varepsilon^{\dagger}_t$ we find that $\mathrm{E}_{\dagger} S^{\dagger}_{0\varepsilon} = 0$ and $\mathrm{E}_{\dagger}\|S^{\dagger}_{0\varepsilon}\|^2 = \sum_{t=k+1}^{T}\sum_{i=1}^{t-k-1}\varepsilon^2_{t-i}\varepsilon^2_t\|\boldsymbol{\gamma}_{i-1:k}\|^2 = o_P(a_k^2 a_T^{2+\epsilon})$ for all $\epsilon > 0$, since, for $\eta \in [\delta, \alpha)$,

$$\mathrm{E}(\mathrm{E}_{\dagger}\|S^{\dagger}_{0\varepsilon}\|^2)^{\eta/2} \leq T(\mathrm{E}\,|\varepsilon_1|^{\eta})^2\sum_{i=0}^{\infty}\|\boldsymbol{\gamma}_{i:k}\|^{\eta} \leq Tk(\mathrm{E}\,|\varepsilon_1|^{\eta})^2\sum_{i=0}^{\infty}|\gamma_i|^{\eta}.$$

Thus, $P_{\dagger}(a_k^{-1}a_T^{-1-\epsilon}\|S^{\dagger}_{0\varepsilon}\| > \omega) = o_P(1)$ for all $\epsilon, \omega > 0$. Further, by the same argument as given in the proof of Lemma 3(b) in the supplement to Cavaliere *et al.* (2016a), under $k^3/T \to 0$ and $\|\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\| = o_P(a_k a_T^{\epsilon-1} + k^{1/2}T^{-1})$ for $\epsilon > 0$ it follows that

$$\|S^{*}_{0\varepsilon}\| = \|S^{\dagger}_{0\varepsilon}\| + \|\sum_{t=k}^{T-1}\Delta\mathbf{Y}^{k\dagger}_t(\mathbf{X}^k_t)'w_t\|_2\|\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\| + o_{P_{\dagger}}(k^{-1}a_T^2)$$

in P-probability. Under these conditions, given the previous evaluation of $\|S^{\dagger}_{0\varepsilon}\|$ it follows that

$$\begin{aligned}\|S^{*}_{0\varepsilon}\| &= \|\sum_{t=k}^{T-1}\Delta\mathbf{Y}^{k\dagger}_t(\mathbf{X}^k_t)'w_t\|_2\|\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\| + o_{P_{\dagger}}(k^{-1}a_T^2)\\ &\leq \lambda^{1/2}_{\max}(S^{\dagger}_{00})\lambda^{1/2}_{\max}(S^{k}_{00})\|\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\| + o_{P_{\dagger}}(k^{-1}a_T^2) = o_{P_{\dagger}}(k^{-1}a_T^2)\end{aligned}$$

in P-probability, using the inequality $\|\sum_{t=k}^{T-1}a_t b_t'\|_2^2 \leq \lambda_{\max}(\sum_{t=k}^{T-1}a_t a_t')\lambda_{\max}(\sum_{t=k}^{T-1}b_t b_t')$ and the magnitude orders $\lambda_{\max}(S^{\dagger}_{00}) = O_{P_{\dagger}}(a_T^{2+\epsilon})$ in P-probability established previously and $\lambda_{\max}(S^{k}_{00}) = O_P(a_T^2)$ implied by Lemma A.1(a).

As $S^{\dagger}_{01} = \sum_{t=k+1}^{T}(\sum_{i=1}^{t-k-1}\varepsilon^{\dagger}_{t-i}\boldsymbol{\gamma}_{i-1:k})(\sum_{i=1}^{t-k-1}\varepsilon^{\dagger}_{t-i}\sum_{j=0}^{i-1}\gamma_j)$, from (A.11) in k dimensions,

$$\begin{aligned}\|\mathrm{E}_{\dagger} S^{\dagger}_{01}\| &= \|\sum_{t=k+2}^{T}\sum_{i=1}^{t-k-1}\varepsilon^2_{t-i}\boldsymbol{\gamma}_{i-1:k}\sum_{j=0}^{i-1}\gamma_j\| \leq (\sum_{t=k+2}^{T}\sum_{i=1}^{t-k-1}\varepsilon^2_{t-i}\|\boldsymbol{\gamma}_{i-1:k}\|)(\sum_{j=0}^{\infty}|\gamma_j|)\\ &\leq \sigma^2_T(\sum_{i=0}^{\infty}\|\boldsymbol{\gamma}_{i:k}\|)(\sum_{j=0}^{\infty}|\gamma_j|) \leq k\sigma^2_T(\sum_{j=0}^{\infty}|\gamma_j|)^2 = O_P(ka_T^2),\end{aligned}$$

$$\begin{aligned}\mathrm{E}_{\dagger}\|S^{\dagger}_{01}\|^2 - \|\mathrm{E}_{\dagger} S^{\dagger}_{01}\|^2 &\leq 2\sum_{s,t=k+1}^{T-1}\varepsilon_s^2\varepsilon_t^2\|\sum_{i=\max(s,t)+1}^{T}\boldsymbol{\gamma}_{i-s-1:k}\sum_{j=0}^{i-t-1}\gamma_j\|^2 \leq 2k^2\sigma^4_T(\sum_{j=0}^{\infty}|\gamma_j|)^4\\ &= O_P(k^2a_T^4).\end{aligned}$$

Similar relations hold for S^{*}_{01}, with $\hat{\varepsilon}_t, \hat{\boldsymbol{\gamma}}_{i-1:k}, \hat{\gamma}_j, \hat{\sigma}^2_{Tk}$ in place of, respectively, $\varepsilon_t, \boldsymbol{\gamma}_{i-1:k}, \gamma_j, \sigma^2_T$. Hence, for every $\eta \in (0,1)$ there exist $K^{\dagger}_{\eta}, \tilde{K}_{\eta} < \infty$ such that $P(P_{\dagger}(k^{-1}a_T^{-2}\|S^{\dagger}_{01}\| > K^{\dagger}_{\eta}) < \eta) > 1-\eta$ and $P(P_{\dagger}(k^{-1}a_T^{-2}\|S^{*}_{01}\| > \tilde{K}_{\eta}) < \eta) > 1-\eta$.

By combining the conclusions about $S^{\dagger}_{00}$, $S^{\dagger}_{0\varepsilon}$ and $S^{\dagger}_{01}$, parts (e) and (g) follow, and similarly, parts (f) and (h) from the conclusions about S^{*}_{00}, $S^{*}_{0\varepsilon}$ and S^{*}_{01}. Next, parts (a)-(h) jointly with the weak convergence of the distribution of $a_T^{-2}\sum_{t=k+1}^{T}(\sum_{s=k+1}^{t-1}\varepsilon^{\dagger}_s)^2$ conditional on $\boldsymbol{\varepsilon}_T$ to the random distribution of $\int\mathcal{S}^2$ conditional on $\{|\Delta\mathcal{S}(u)|\}_{u\in(0,1]}$ (proved similarly to Lemma 1), it follows that

$$P_{\dagger}\left(T\left|\phi^{*}_k - \gamma(1)^{-1}\{\sum_{t=k+1}^{T}(\sum_{s=k+1}^{t-1}\varepsilon^{\dagger}_s)^2\}^{-1}\sum_{t=k+1}^{T}(\sum_{s=k+1}^{t-1}\varepsilon^{\dagger}_s)\varepsilon^{\dagger}_t\right| > \eta\right) \xrightarrow{P} 0$$

for all $\eta > 0$. As $\max_{t=1,\ldots,k}|\varepsilon_t| = O_{P_{\dagger}}(a_k)$ in P-probability and $k^3/T \to 0$, in the previous display summations can start at 1, which completes the proof. ■

A.2.3 Concluding Results

Proof of Lemma 1. By Corollary 1 of LePage *et al.* (1997), we can consider a probability space where Γ, U and δ are defined together with $\{\tilde{\varepsilon}_t\}_{t\in\mathbb{N}}$ distributed like $\{\varepsilon_t\}_{t\in\mathbb{N}}$, and such that $a_T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor}\tilde{\varepsilon}_t \to \sum_{i=1}^{\infty}\delta_i\Gamma_i^{-1/\alpha}\mathbb{I}\{U_i \le (\cdot)\}$ a.s. in $D[0,1]$. Without loss of generality, we proceed as if $\{\varepsilon_t\}_{t\in\mathbb{N}}$ have this property themselves:

$$a_T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor}\varepsilon_t \overset{a.s}{\to} \mathcal{S}(\cdot) := \sum_{i=1}^{\infty}\delta_i\Gamma_i^{-1/\alpha}\mathbb{I}\{U_i \le (\cdot)\} \tag{A.12}$$

in $D[0,1]$, and argue that on such a probability space it holds that

$$\mathrm{E}_{|\varepsilon|}\, f(2\gamma_T(\boldsymbol{\varepsilon}_T)) \overset{P}{\to} \mathrm{E}\left[f(\frac{\mathcal{S}(1)^2 - \sum_{i=1}^{\infty}\Gamma_i^{-2/\alpha}}{\int \mathcal{S}^2(s)\,ds})|\Gamma,U\right] \tag{A.13}$$

for bounded and continuous real f, where $\mathrm{E}_{|\varepsilon|}$ denotes expectation under $P_{|\varepsilon|}$. Then it follows that on a general probability space convergence in (A.13) holds in the weak sense instead of in probability.

For every $K = 1, ..., T$, let e_K be the K-th order statistic of $\{|\varepsilon_t|\}_{t=1}^T$, and $e_0 := \infty$. By the continuity on $D[0,1]$ of ordered jumps and their locations, and since $\{\Gamma_i\}$ is a.s. strictly increasing in i, (A.12) implies that

$$e_{\lfloor T\cdot\rfloor,K} := a_T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor}\varepsilon_t\mathbb{I}_{\{e_K\le|\varepsilon_t|<e_{K-1}\}} \overset{a.s.}{\to} \delta_K\Gamma_K^{-1/\alpha}\mathbb{I}_{\{U_K\le(\cdot)\}} \tag{A.14}$$

in $D[0,1]$. For every $K\in\mathbb{N}$ and $T\ge K$, define $\varepsilon_{t,K} := \varepsilon_t\mathbb{I}_{\{|\varepsilon_t|\ge e_K\}}$ $(t = 1, ..., T)$ and $\boldsymbol{\varepsilon}_{T,K} := (\varepsilon_{1,K}, ..., \varepsilon_{T,K})$. As there are no ties in the order statistics of $\{|\varepsilon_t|\}_{t=1}^T$, with $P_{|\varepsilon|}$-probability converging to one in P-probability, we find by direct calculation that $\mathrm{E}_{|\varepsilon|}\, f(2\gamma_T(\boldsymbol{\varepsilon}_{T,K}))$ equals

$$\begin{aligned}
\mathrm{E}_{|\varepsilon|}\, f(\frac{(\sum_{t=1}^T\varepsilon_{t,K})^2 - \sum_{t=1}^T\varepsilon_{t,K}^2}{\int(\sum_{t=1}^{\lfloor Ts\rfloor}\varepsilon_{t,K})^2ds}) &= 2^{-K}\sum_{b\in\{-1,1\}^K} f(\frac{(\sum_{i=1}^K b_i|e_{T,i}|)^2 - \sum_{i=1}^K e_{T,i}^2}{\int(\sum_{i=1}^K b_i|e_{\lfloor Ts\rfloor,K}|)^2ds}) + o_P(1)\\
&\overset{P}{\to} 2^{-K}\sum_{b\in\{-1,1\}^K} f(\frac{(\sum_{i=1}^K b_i\Gamma_i^{-1/\alpha})^2 - \sum_{i=1}^K\Gamma_i^{-2/\alpha}}{\int(\sum_{i=1}^K b_i\Gamma_i^{-1/\alpha}\mathbb{I}_{\{U_i\le s\}})^2ds})\\
&= \mathrm{E}\left[f(\frac{(\sum_{i=1}^K\delta_i\Gamma_i^{-1/\alpha})^2 - \sum_{i=1}^K\Gamma_i^{-2/\alpha}}{\int(\sum_{i=1}^K\delta_i\Gamma_i^{-1/\alpha}\mathbb{I}_{\{U_i\le s\}})^2ds})\middle|\Gamma,U\right]
\end{aligned}$$

for bounded and continuous real f, the convergence using (A.14) and the 'in probability' CMT. It can be further established that

$$\lim_{K\to\infty}\limsup_{T\to\infty}P\left\{P_{|\varepsilon|}\left(|\gamma_T(\boldsymbol{\varepsilon}_T) - \gamma_T(\boldsymbol{\varepsilon}_{T,K})| \ge \eta\right) \ge \omega\right\} \to 0$$

and

$$\lim_{K\to\infty}P\left\{P\left(\left|\frac{(\sum_{i=1}^K\delta_i\Gamma_i^{-1/\alpha})^2 - \sum_{i=1}^K\Gamma_i^{-2/\alpha}}{\int(\sum_{i=1}^K\delta_i\Gamma_i^{-1/\alpha}\mathbb{I}_{\{U_i\le s\}})^2ds} - \frac{(\sum_{i=1}^{\infty}\delta_i\Gamma_i^{-1/\alpha})^2 - \sum_{i=1}^{\infty}\Gamma_i^{-2/\alpha}}{\int(\sum_{i=1}^{\infty}\delta_i\Gamma_i^{-1/\alpha}\mathbb{I}_{\{U_i\le s\}})^2ds}\right| \ge \eta\middle|\Gamma,U\right) \ge \omega\right\} \to 0,$$

for every $\eta, \omega > 0$, implying that

$$\mathrm{E}_{|\varepsilon|}\, f(2\gamma_T(\boldsymbol{\varepsilon}_T)) \xrightarrow{P} \mathrm{E}\left[f\{\frac{(\sum_{i=1}^{\infty} \delta_i \Gamma_i^{-1/\alpha})^2 - \sum_{i=1}^{\infty} \Gamma_i^{-2/\alpha}}{\int (\sum_{i=1}^{\infty} \delta_i \Gamma_i^{-1/\alpha} \mathbb{I}_{\{U_i \le s\}})^2 ds}\} | \Gamma, U \right].$$

To establish (A.13), it remains to notice that $(\sum_{i=1}^{\infty} \delta_i \Gamma_i^{-1/\alpha})^2 - \sum_{i=1}^{\infty} \Gamma_i^{-2/\alpha} = \mathcal{S}(1)^2 - \sum_{i=1}^{\infty} \Gamma_i^{-2/\alpha} = 2 \int \mathcal{S} d\mathcal{S}$ and $\int (\sum_{i=1}^{\infty} \delta_i \Gamma_i^{-1/\alpha} \mathbb{I}_{\{U_i \le s\}})^2 ds = \int \mathcal{S}^2$.

Now we argue that for every $c \in \mathbb{R}$,

$$P\left(\left(\int \mathcal{S}^2 \right)^{-1} \int \mathcal{S} d\mathcal{S} = c \,\middle|\, \Gamma, U \right) = 0 \ P\text{-a.s.} \tag{A.15}$$

It is known in the literature (see Jach and Kokoszka, 2004, p.78) that the distribution function of the limit in (8) is continuous. Thus,

$$\mathrm{E}\, P\left(\left(\int \mathcal{S}^2 \right)^{-1} \int \mathcal{S} d\mathcal{S} = c \,\middle|\, \Gamma, U \right) = P\left(\left(\int \mathcal{S}^2 \right)^{-1} \int \mathcal{S} d\mathcal{S} = c \right) = 0.$$

Since the conditional probability is non-negative, (A.15) follows. ■

We show next that Theorem 4 is applicable to $g_T = R_T$ and $g_T^* = R_T^*$.

Proof of Theorem 5. With $\mathbf{Z}_T = \{y_t\}_{t=0}^T$, $g_T = R_T$ and $g_T^* = R_T^*$ and γ_T defined in (17), condition (i) of Theorem 4 is satisfied in view of Lemmas A.2 and A.3, and condition (ii), in view of convergence (18), with φ equal to the limiting measure in (18). The convergence of R_T in conditional distribution follows by combining Lemma A.2 and (18), whereas the weak convergence of $P^*(R_T^* \le R_T)$ follows from (16) in Theorem 4. ■

Table 1: Empirical Size of the Unit Root Tests

	wild bootstrap				m-out of-n		sub-sample		ADF			
	recolored		non-recolored		bootstrap							
θ	$Q_T^{rc,r}$	$R_T^{rc,r}$	$Q_T^{n,r}$	$R_T^{n,r}$	Q_T^m	R_T^m	Q_T^{jk}	R_T^{jk}	Q_T^{α}	R_T^{α}	Q_T^2	R_T^2
	$\alpha = 2$	$\kappa = 4$										
−0.8	28.2	27.4	28.4	27.0	35.0	31.8	18.3	5.8	28.6	32.2	28.4	31.6
−0.5	5.2	5.0	4.6	5.0	11.7	6.1	4.2	0.7	7.4	8.1	7.1	7.9
0	5.3	4.9	5.3	4.9	6.8	4.8	3.4	0.5	5.3	6.8	5.2	6.6
0.5	5.0	4.1	4.6	4.1	5.8	4.0	3.6	0.4	4.6	6.0	4.5	5.8
0.8	4.0	3.4	3.9	3.4	4.9	3.2	3.7	0.3	3.9	5.2	3.8	5.0
	$\alpha = 2$	$\kappa = 12$										
−0.8	7.8	5.5	7.7	4.5	8.7	3.4	0.0	8.5	8.0	13.9	7.8	13.7
−0.5	5.8	3.7	5.7	3.6	5.7	3.2	0.0	6.7	6.2	11.7	6.0	11.5
0	5.8	3.7	5.7	3.8	5.1	3.7	0.0	7.3	6.0	12.3	5.9	12.0
0.5	5.7	3.9	5.9	4.0	5.1	4.0	0.0	7.6	6.1	12.4	6.0	12.2
0.8	5.5	3.7	5.6	3.8	4.9	3.7	0.0	7.3	5.9	12.0	5.8	11.8
	$\alpha = 1.5$	$\kappa = 4$										
−0.8	28.4	28.1	30.4	28.9	32.0	29.4	14.0	7.7	25.3	28.6	23.4	27.0
−0.5	7.3	5.8	5.6	5.8	10.2	4.0	2.8	1.2	5.4	6.9	6.4	6.2
0	5.1	4.6	5.1	4.6	5.7	4.0	2.2	0.8	4.4	5.7	3.9	5.1
0.5	4.6	3.7	4.1	3.6	5.0	3.4	2.3	0.8	4.2	5.0	3.1	4.4
0.8	3.1	2.7	3.1	2.6	4.0	2.7	2.3	0.7	3.2	4.2	2.8	3.7
	$\alpha = 1.5$	$\kappa = 12$										
−0.8	8.5	6.6	8.3	5.5	7.2	3.3	0.1	6.8	7.9	13.0	5.9	11.1
−0.5	6.1	4.3	5.9	3.9	4.7	3.0	0.0	5.6	6.2	11.2	4.8	9.6
0	5.8	4.4	5.9	4.5	4.2	3.5	0.0	6.1	6.4	11.4	4.6	9.8
0.5	5.7	4.4	5.9	4.6	3.9	3.9	0.0	6.2	6.3	11.5	4.4	9.9
0.8	5.3	4.0	5.5	4.3	3.6	3.8	0.0	6.0	6.0	11.1	4.4	9.6
	$\alpha = 1.0$	$\kappa = 4$										
−0.8	36.4	35.9	40.1	39.0	29.6	27.7	11.3	14.3	27.5	30.7	19.7	23.8
−0.5	8.4	6.7	7.7	7.2	9.2	38.5	2.2	3.3	6.9	8.2	3.9	5.6
0	6.1	5.6	6.0	5.5	5.4	4.3	1.8	2.7	5.8	7.1	3.1	4.7
0.5	5.8	4.9	4.4	3.9	5.2	3.6	1.7	2.5	5.0	6.2	3.0	4.1
0.8	2.9	2.4	2.7	2.3	3.8	2.8	1.7	2.3	4.3	5.4	2.2	3.5
	$\alpha = 1.0$	$\kappa = 12$										
−0.8	12.9	11.5	12.8	10.2	6.2	3.9	2.4	6.1	7.4	11.9	4.4	9.4
−0.5	7.8	6.4	7.1	5.6	4.1	3.8	1.2	5.2	6.2	10.4	3.6	8.3
0	7.4	6.2	7.3	5.9	3.8	4.3	1.4	5.5	6.4	11.2	3.8	8.7
0.5	7.3	6.0	7.4	6.0	3.7	4.6	1.2	5.8	6.6	11.2	3.7	9.0
0.8	6.7	5.5	6.7	5.3	3.5	4.5	0.7	5.6	6.3	10.8	3.6	8.7

Table 2: Empirical Power of the Unit Root Tests

	wild bootstrap				m-out of-n		sub-sample		ADF			
	recoloured		non-recoloured		bootstrap							
θ	$Q_T^{rc,r}$	$R_T^{rc,r}$	$Q_T^{n,r}$	$R_T^{n,r}$	Q_T^m	R_T^m	Q_T^{jk}	R_T^{jk}	Q_T^α	R_T^α	Q_T^2	R_T^2
	$\alpha = 2$	$\kappa = 4$										
-0.8	95.8	95.4	95.9	95.0	97.4	96.1	82.5	43.4	96.3	97.6	96.1	97.5
-0.5	48.1	45.4	47.2	41.8	58.6	40.7	32.0	9.4	47.6	55.3	46.9	54.3
0	41.8	38.0	41.7	38.0	48.3	37.8	26.4	4.8	42.5	48.3	41.8	47.1
0.5	37.2	35.1	37.0	37.4	42.2	31.9	26.6	4.4	39.8	43.5	38.6	42.3
0.8	31.0	27.9	30.9	27.9	36.5	26.1	26.7	3.7	31.4	37.3	31.1	36.2
	$\alpha = 2$	$\kappa = 12$										
-0.8	46.1	37.3	45.4	31.1	49.6	25.1	0.0	49.1	46.6	66.2	46.1	65.5
-0.5	32.6	23.2	31.9	20.4	33.4	16.4	0.0	35.2	32.7	51.4	32.2	50.8
0	30.7	21.1	30.4	19.7	29.0	16.7	0.0	33.4	31.2	49.1	30.8	48.3
0.5	29.8	20.7	29.8	19.6	27.7	16.9	0.0	32.8	30.7	48.2	30.3	47.4
0.8	28.7	19.6	28.9	18.7	26.4	16.0	0.0	31.6	29.5	47.0	29.1	46.2
	$\alpha = 1.5$	$\kappa = 4$										
-0.8	95.4	95.4	95.6	95.3	97.9	97.6	78.7	55.4	97.1	98.6	96.4	98.1
-0.5	58.0	55.2	58.0	54.7	61.0	38.2	28.8	15.7	48.6	55.0	44.0	49.7
0	51.0	47.3	50.8	46.8	53.1	37.0	23.3	10.1	41.4	47.2	37.3	42.8
0.5	45.6	42.4	45.3	44.8	44.9	30.3	23.0	9.8	37.4	43.1	32.6	37.8
0.8	39.0	34.8	38.9	34.5	37.9	23.6	22.8	7.9	30.2	34.7	26.5	31.2
	$\alpha = 1.5$	$\kappa = 12$										
-0.8	55.8	45.8	54.6	39.9	51.9	24.3	0.4	46.0	54.3	73.0	43.8	67.0
-0.5	42.5	31.9	42.1	29.1	31.9	14.2	0.2	30.3	38.3	54.5	29.8	47.7
0	40.1	29.1	39.7	27.1	27.8	14.0	0.3	28.1	35.9	51.3	28.0	44.9
0.5	39.1	28.4	38.8	26.9	26.1	14.1	0.2	27.8	35.2	50.2	27.3	43.7
0.8	37.9	27.1	37.7	25.8	24.7	13.3	0.3	26.5	33.8	48.3	26.3	42.1
	$\alpha = 1.0$	$\kappa = 4$										
-0.8	94.6	95.0	94.9	94.3	97.9	98.4	74.2	72.8	97.7	99.3	96.4	98.4
-0.5	70.8	67.5	69.8	65.0	66.6	41.1	26.1	33.7	63.3	72.6	38.2	40.7
0	64.7	61.4	64.5	60.8	61.2	38.6	20.8	26.8	59.4	63.9	30.1	35.1
0.5	61.4	58.1	60.7	58.3	50.2	29.2	20.4	23.9	52.5	54.7	28.1	32.2
0.8	52.4	48.3	51.8	47.3	39.1	19.7	20.0	21.0	39.1	40.6	20.4	23.4
	$\alpha = 1.0$	$\kappa = 12$										
-0.8	67.0	57.7	64.7	52.0	55.1	24.6	8.6	40.0	63.3	81.3	36.7	70.2
-0.5	56.2	45.6	54.5	41.7	29.5	11.8	6.5	24.8	38.3	59.0	25.2	40.4
0	53.5	42.8	52.3	40.2	24.8	11.3	7.5	22.1	35.6	51.9	23.5	37.2
0.5	53.1	42.3	51.9	40.0	23.5	11.7	6.2	23.1	34.9	50.1	23.3	36.5
0.8	51.8	41.0	51.0	39.1	22.1	11.1	6.7	20.1	33.4	47.4	22.4	34.6