Essex Finance Centre

Working Paper Series

Working Paper No16: 01-2017

"Unit Root Tests and Heavy-Tailed Innovations"

"Iliyan Georgiev, Paulo M. M. Rodrigues and A.M.Robert Taylor"

Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ
Web site: http://www.essex.ac.uk/ebs/

Unit Root Tests and Heavy-Tailed Innovations*

Iliyan Georgiev[a], Paulo M. M. Rodrigues[b] and A.M.Robert Taylor[c]

[a] Department of Statistical Sciences, University of Bologna
[b] Banco de Portugal and Nova School of Business and Economics, Universidade Nova de Lisboa
[c] Essex Business School, University of Essex

September 22, 2016

Abstract

We evaluate the impact of heavy-tailed innovations on some popular unit root tests. In the context of a near-integrated series driven by linear-process shocks, we demonstrate that their limiting distributions are altered under infinite variance *vis-à-vis* finite variance. Reassuringly, however, simulation results suggest that the impact of heavy-tailed innovations on these tests are relatively small. We use the framework of Amsler and Schmidt (2012) whereby the innovations have local-to-finite variances being generated as a linear combination of draws from a thin-tailed distribution (in the domain of attraction of the Gaussian distribution) and a heavy-tailed distribution (in the normal domain of attraction of a stable law). We also explore the properties of ADF tests which employ Eicker-White standard errors, demonstrating that these can yield significant power improvements over conventional tests.

JEL: C12, C22, C58.
Keywords: Infinite variance, α-stable distribution, Eicker-White standard errors, asymptotic local power functions, weak dependence.

1 Introduction

Gaussianity is a commonly employed assumption in time series analysis and in econometrics in general. Characteristically, tests for stochastic trends or persistence have been based on ordinary least squares (OLS) estimation, achieving optimal power properties under normality. For example, the asymptotic local power envelopes frequently discussed in the unit root literature (see, among others, Elliott, Rothenberg and Stock, 1996) are derived under the assumption that the latent innovations driving the processes are Gaussian. In practice, however, many variables in macroeconomics and finance appear to be driven by heavy-tailed shocks, often exhibiting large and sudden movements, similar to additive outliers. To give just two examples, Falk and Wang (2003) test the long-run purchasing power parity hypothesis considering exchange rate returns and inflation rates to be heavy-tailed stochastic processes, using residual and likelihood-ratio-based co-integration tests which explicitly allow for infinite-variance innovations, while Charemza, Hristova and Burridge (2005) apply unit root tests to inflation data allowing for infinite variance innovations.

As discussed in Rothenberg and Stock (1997), if the data are non-Gaussian but remain finite variance, then OLS-based inference typically remains valid in large samples, but is inefficient compared to methods which exploit the correct form of the likelihood. Rothenberg and Stock (1997)

*The authors would like to thank participants of the Econometrics and Finance seminar at the Leibniz Universität Hannover, Germany for useful feedback on an earlier version of the paper. Correspondence to: Paulo Rodrigues, Banco de Portugal, Economics and Research Department, Av. Almirante Reis, 71-6th floor, 1150-012 Lisbon, Portugal.

and Jansson (2008) discuss the development of power envelopes for non-Gaussian data; see also Haldrup and Jansson (2006, pp.260-270) for a useful summary of this material. As Haldrup and Jansson (2006) discuss, minimal sufficient statistics which are derived under the assumption of Gaussianity are invariant with respect to the distribution of the disturbances as long as these have an expected value of zero and an unconditional variance of one. The Gaussian asymptotic local power envelope can therefore be seen as the lower bound of the maximal attainable local asymptotic power. Rothenberg and Stock (1997) derive an upper bound on the magnitude of the power gains available when the innovations are non-Gaussian under the assumption that the underlying innovation distribution is known. Jansson (2008) derives further results for the case where the innovation distribution is not assumed known. Where the innovations are symmetrically distributed he demonstrates that the result in Rothenberg and Stock (1997) can be obtained without knowledge of the innovation distribution.

Since the seminal work of Mandelbrot (1967) and Fama (1965) heavy tails have received considerable attention in the statistics, econometrics and finance literatures; see Resnick (1997) for a comprehensive review. Given that unusually large movements in economic and financial time series seem to occur more often than is implied by Gaussianity and that in practice the exact distribution of the innovations cannot be assumed known, information on the performance of unit root tests and an evaluation of their resilience with respect to deviations from Gaussianity is of considerable empirical relevance. Moreover, a growing literature on unit root testing in processes with infinite-variance innovations, drawn from the α-stable class of distributions, has developed in recent years; see Ahn, Fotopoulos and He (2001) for a partial review. Chan and Tran (1989) and Rachev, Mittnik and Kim (1998) detail large sample results for OLS-based unit root tests where the underlying data generation process [DGP] is an AR(1) driven by independent and identically distributed [IID] heavy-tailed innovations. Knight (1989) develops M-estimation and least absolute deviation based unit root tests in the context of the same DGP, while Knight (1991) extends these results to the case where the driving shocks display infinite-order moving average dynamics. Phillips (1990) generalises the Phillips and Perron (1988) unit root tests to the context of processes driven by weakly dependent shocks whose innovations display infinite variance. More recently, Samarakoon and Knight (2009) consider M-based testing for unit roots in finite-order autoregressive processes driven by infinite variance innovations. Bootstrap based unit root tests are proposed by Horváth and Kokoszka (2003), Moreno and Romo (2012), and Cavaliere, Georgiev and Taylor (2016b), the latter focusing on the popular augmented Dickey Fuller [ADF] tests of Dickey and Fuller (1979). Chan, Peng and Qi (2006) derive results for inference in a near-integrated first-order autoregressive process with IID innovations and infinite variance based on quantile regression methods.

Cappuccio and Lubian (2007) investigate the behaviour of a variety of stationarity and non-stationarity tests under a class of local-to-finite variance IID errors, originally introduced in Amsler and Schmidt (1999) and subsequently published as Amsler and Schmidt (2012). The local-to-finite variance formulation is intended to deliver an improvement over standard Wiener asymptotics for data that, on the one hand, possess finite variance, while on the other hand exhibit heavy-tailed behaviour that can render the Wiener approximation inadequate in finite samples. Using this local-to-finite variance set up, in this paper we evaluate both analytically and through Monte Carlo experiments the impact that heavy tailed innovations have on the size and power performance of popular unit root tests. In particular, we generalise existing results in three directions. First, we work within a local-to-unity framework, thereby allowing an evaluation of the impact of heavy-tailed innovations on the asymptotic local power functions of the tests. Second, we allow for the case where the driving shocks follow a linear process in heavy-tailed innovations rather than simply follow an IID process. Finally, we explore how variants of the familiar ADF unit root tests which employ Eicker-White (rather than OLS) standard errors behave when applied to heavy-tailed data. In conventional stationary regression settings, these standard errors are designed to yield regression t- and F-tests which are robust to certain forms of heteroskedasticity in the latent error process. It therefore seems worth exploring how they behave with heavy-tailed data. Our results suggest that

a very simple implementation of Eicker-White can deliver ADF-type tests which display significant finite sample power gains relative to other standard tests.

The remainder of the paper is structured as follows. In section 2 we present and discuss our reference time series model, a local-to-unity autoregressive process driven by shocks which follow a stable and invertible linear process driven by heavy-tailed innovations. Section 3 briefly outlines some of the most widely used unit root tests in the literature, including the variants of the conventional ADF tests implemented with Eicker-White standard errors. Section 4 details the large sample behaviour of these unit root statistics when the innovations are heavy-tailed. Section 5 reports results from a Monte Carlo simulation study into the finite sample size and power properties of the tests in such cases. Section 6 concludes. All proofs are contained in a mathematical appendix.

2 Near-Integrated Local-to-Finite Variance Processes

Consider the data $\{x_t\}_{t=1}^T$ generated from the near-integrated process,

$$x_t = \phi_T x_{t-1} + u_t \tag{1}$$

where $x_0 = O_p(1)$ and $\phi_T := 1 - c/T$, with c a fixed finite constant, which characterises the local-to-unity properties of x_t. Our interest lies in testing the unit root null hypothesis $\mathsf{H_0} : \phi_\mathsf{T} = 1$ $(c = 0)$ against the locally stable alternative $\mathsf{H_{1,c}} : |\phi_\mathsf{T}| < 1$ $(c > 0)$, although the large sample results which we provide do not in fact restrict c to be non-negative and so could also be used to evaluate the performance of these tests against locally explosive alternatives of the form $\mathsf{H_{1,c^-}} : |\phi_\mathsf{T}| > 1$ $(c < 0)$.

In the context of (1) we specify u_t to follow a linear process; that is,

$$u_t = \psi(L)\varepsilon_t, \;\; \varepsilon_t = \varepsilon_{1t} + \gamma a_T^{-1} T^{1/2} \varepsilon_{2t} \tag{2}$$

where $\psi(L)$ is a potentially infinite-ordered moving average polynomial, conditions on which will be specified below. The innovations $\{\varepsilon_{1t}\}_{t\in\mathbb{Z}}$ and $\{\varepsilon_{2t}\}_{t\in\mathbb{Z}}$ are independent sequences with $\varepsilon_{1t} \sim IID(0, \sigma_1^2)$ and where ε_{2t} is IID and symmetrically distributed in the normal domain of attraction of a stable law with characteristic exponent (or tail index) $\alpha \in [1, 2)$, denoted $\varepsilon_{2t} \in \mathcal{ND}(\alpha)$; see, for example, Ibragimov and Linnik (1971, pp.92–93). The normalisation of ε_{2t}, by $\gamma > 0$ and $a_T := aT^{1/\alpha}$, is such that the partial sum process of $T^{-1/2}\varepsilon_t$ converges weakly to a Lévy process with a Wiener component and an α-stable component whose relative weights depend on σ_1 and γ.

The process in (2) can be equivalently written as,

$$\begin{aligned} u_t &= \psi(L)\varepsilon_{1t} + \gamma a_T^{-1} T^{1/2} \psi(L)\varepsilon_{2t} \\ &= \varepsilon_{1t}^* + \gamma a_T^{-1} T^{1/2} \varepsilon_{2t}^* \end{aligned} \tag{3}$$

where $\varepsilon_{it}^* := \psi(L)\varepsilon_{it}$, $i = 1, 2$. The specification in (2) was originally proposed by Amsler and Schmidt (1999, 2012) for the particular case where $\psi(L) = 1$, so that u_t in (2) is also an IID process. In their case, (3) therefore simplifies to

$$u_t = \varepsilon_{1t} + \gamma a_T^{-1} T^{1/2} \varepsilon_{2t}. \tag{4}$$

The error process u_t in (3) maintains the infinite variance property in finite samples, while collapsing to the standard finite variance assumption asymptotically (for $T \to \infty$ and t fixed); see also Callegari *et al.* (2003), Capuccio and Lubain (2007) and Samarakoon and Knight (2009, pp. 330-331). It is important to recognise, however, that the passage to the limit in (4), and in (3), for a fixed t is distinct from the passage to the limit which occurs for statistics computed by aggregation over t; in the latter case, the contribution of ε_{2t} rather than vanishing as $T \to \infty$ gives rise to a Lévy α-stable component. Anticipating this functional convergence, we recall its one-dimensional counterpart. It follows from the definition of the $\mathcal{ND}(\alpha)$ domain that the non-degenerate limit of

a (scaled and centred) sum of IID $\mathcal{ND}(\alpha)$ variables is an α-stable random variable. It is usual to denote an α-stable distribution by $S_\alpha(\beta,\sigma,\mu)$, indicating that the distribution is defined by four parameters: (i) the characteristic exponent α, $\alpha \in (0,2]$; when $\alpha = 2$ we obtain the Gaussian distribution with variance $2\sigma^2$, and when $\alpha < 2$ the resulting distribution is heavy-tailed with the key property that the moments $\mathrm{E}\,|S_\alpha(\beta,\sigma,\mu)|^r$ are finite for $r < \alpha$ and infinite for $r \geq \alpha$, (ii) the skewness parameter β which determines the shape of the distribution (for $\beta > 0$ (< 0) the distribution is positively (negatively) skewed and for $\beta = 0$ it is symmetric), (iii) the scale parameter σ, $\sigma \in \mathbb{R}^+$, and (iv) the location parameter $\mu \in \mathbb{R}$. Because we have assumed symmetry here, we have that $a_T^{-1}\sum_{t=1}^{T}\varepsilon_{2t} \Rightarrow S_\alpha(0,1,0)$, with no need for centring, where $\Rightarrow$ denotes weak convergence and noting that the scale parameter in the limit distribution is made unity by an appropriate choice of the constant a in the definition of a_T. For more details on stable distributions see, for example, Feller (1971), and for stable processes, Samorodnitsky and Taqqu (2000).

In what follows, we place the following set of conditions, collectively labelled Assumption $\mathcal{A}$, on u_t of (3).

Assumption $\mathcal{A}$:

$\mathcal{A}.1$: $\{\varepsilon_{it}^*\}_{t\in\mathbb{Z}}$, $i = 1,2$, are generated by a linear process such that

$$\varepsilon_{it}^* = \psi(L)\varepsilon_{it} = \sum_{j=0}^{\infty}\psi_j\varepsilon_{ij}$$

with $\psi_0 := 1$ and $\psi(z) := \sum_{j=0}^{\infty}\psi_j z^j$ has no roots on the closed unit complex disk.

$\mathcal{A}.2$: $\{\varepsilon_{1t}\}_{t\in\mathbb{Z}}$ is an IID sequence with $E[\varepsilon_{1t}] = 0$, $E[\varepsilon_{1t}^2] = \sigma_1^2 > 0$ and $E[\varepsilon_{1t}^4] < \infty$, $t \in \mathbb{Z}$.

$\mathcal{A}.3$: $\{\varepsilon_{2t}\}_{t\in\mathbb{Z}}$ is an IID sequence of symmetrically distributed random variables in the normal domain of attraction of a stable law with characteristic exponent $\alpha \in [1,2)$, denoted as $\varepsilon_{2t} \in \mathcal{ND}(\alpha)$. In particular, $E[\varepsilon_{2t}] = 0$, $t \in \mathbb{Z}$, whenever the expectation exists.

$\mathcal{A}.4$: $\{\varepsilon_{1t}\}_{t\in\mathbb{Z}}$ and $\{\varepsilon_{2t}\}_{t\in\mathbb{Z}}$ are independent.

$\mathcal{A}.5$: $\sum_{j=0}^{\infty} j\left|\psi_j\right|^{\delta/2} < \infty$ for some $\delta \in (0,\alpha)\cap(0,1]$ and the inverse $1+\sum_{j=1}^{\infty}\beta_j z^j := (\sum_{j=0}^{\infty}\psi_j z^j)^{-1}$ satisfies $\sum_{j=1}^{\infty}|\beta_j|^\delta < \infty$.

Remark 2.1: Although the more general assumption $\alpha \in (0,2)$ could be considered instead of $\alpha \in [1,2)$, as specified in $\mathcal{A}.3$, values for the tail index in the range $\alpha \in (0,1)$ are arguably of limited empirical relevance for economic and financial variables in view of the finiteness of moments property of α-stable random variables noted above; hence, in what follows we restrict our attention to $\alpha \in [1,2)$. Regarding summability, Assumption $\mathcal{A}.1$ and the condition that $\sum_{j=0}^{\infty} j\left|\psi_j\right|^{\delta/2} < \infty$ from Assumption $\mathcal{A}.5$ are sufficient for $(\sum_{j=0}^{\infty}\psi_j z^j)^{-1}$ to be bounded and bounded away from zero for $|z| \leq 1$. The conditions placed on $\psi(z)$ in Assumptions $\mathcal{A}.1$ and $\mathcal{A}.5$ imply Assumption 2 of Chang and Park (2002) and coincide with Assumption A(b,d) of Cavaliere *et al.* (2016a), allowing us to use certain results from these papers. □

Remark 2.2: Assumptions $\mathcal{A}.2$ and $\mathcal{A}.3$ impose an IID condition on $\{\varepsilon_{1t}\}$ and $\{\varepsilon_{2t}\}$, respectively. In the case of $\{\varepsilon_{1t}\}$ this can be relaxed to the considerably weaker globally stationary martingale difference sequence [MDS] assumption of, *inter alia*, Davidson (1994, p.450), without altering the limiting results given in this paper. In the case of $\{\varepsilon_{2t}\}$, however, we cannot weaken this assumption because of the absence in the literature of any counterparts for the IID convergence results given in Lemma 4.1 below for the case where $\{\varepsilon_{2t}\}$ is a MDS. □

3 Unit Root Tests

A large number of procedures has been proposed to test for the presence of an autoregressive unit root. In this section we briefly review the most widely used of these tests, summarising known facts about their behaviour in the pure finite variance case (*i.e.*, $u_t = \varepsilon_{1t}^*$) and pure infinite variance case (*i.e.*, $u_t = \varepsilon_{2t}^*$). We will subsequently evaluate the behaviour of these tests, both analytically and through Monte Carlo simulations, in the context of heavy-tailed data generated from (1) and (3).

3.1 Augmented Dickey-Fuller Tests

The well-known ADF tests are computed from the sieve regression

$$\Delta x_t = \rho x_{t-1} + \sum_{j=1}^{p_T} \beta_j \Delta x_{t-j} + \varepsilon_{p_T,t}, \qquad t = p_T + 1, ..., T. \tag{5}$$

In the context of (5) we can test the unit root null hypothesis, $\mathsf{H_0}$, against the locally stable root alternative, $\mathsf{H_{1,c}}$, using either the ADF t-ratio, $t_{\widehat{\rho}} := \widehat{\rho}/se\,(\widehat{\rho})$, or the associated normalised bias statistic $Z_{\widehat{\rho}} := T\widehat{\rho}/(1 - \sum_{i=1}^{p_T} \widehat{\beta}_i)$, where $\widehat{\rho}$ and $\widehat{\beta}_j$, $j = 1, ..., p_T$ are the OLS slope estimates of ρ and β_j, $j = 1, ..., p_T$, respectively, and $se\,(\widehat{\rho})$ is the OLS standard error of $\widehat{\rho}$.

In the pure finite variance case ($u_t = \varepsilon_{1t}^*$), Chang and Park (2002) show that provided the lag truncation parameter, p_T in (5), satisfies the rate condition $1/p_T + p_T^3/T \to 0$ as $T \to \infty$, and that standard summability and moment conditions hold, that $t_{\widehat{\rho}}$ and $Z_{\widehat{\rho}}$ have the usual pivotal Dickey-Fuller limiting null distributions (which are functionals of a standard Brownian motion process) regardless of any weak dependence present in u_t. Cavaliere *et al.* (2016b) demonstrate that the same rate condition is sufficient in the pure infinite variance case, $u_t = \varepsilon_{2t}^*$. In particular, they show that under the summability conditions of Assumption $\mathcal{A}.5$ this rate condition on p_T ensures that $t_{\widehat{\rho}}$ and $Z_{\widehat{\rho}}$ have the same limiting null distributions when weak dependence is present as in the case where u_t is IID. In the pure infinite variance case, however, these limiting null distributions are functionals of an α-stable motion (with $\alpha < 2$). The limiting null distributions, in both the pure finite variance and pure infinite variance cases, are special cases of the limiting distributions we shall report in Theorem 4.1, below.

We will also consider implementations of the ADF t-statistic which, rather than using the OLS standard error of $\widehat{\rho}$, use Eicker-White standard errors. Such implementations seem worth exploring in the present context, given that these standard errors are used to control for heteroskedasticity in the latent error process in the case of standard stationary regression models. While they will clearly not control for heavy-tailed behaviour in the present context they might mitigate against the impact of the type of large innovations associated with heavy-tails. To that end, let $\Delta X_{t-1} := (\Delta x_{t-1}, ..., \Delta x_{t-p_T})'$. Then the Eicker-White analogues of the ADF $t_{\widehat{\rho}}$ statistic can be defined in potentially several ways, including $t_{\widehat{\rho},i}^W = \widehat{\rho}/se_{W,i}\,(\widehat{\rho})$, $i = 1, 2$, where

$$se_{W,1}\,(\widehat{\rho}) := \left(\sum_{t=p_T+1}^{T} x_{t-1}^2 \right)^{-1} \left(\sum_{t=p_T+1}^{T} x_{t-1}^2 \widehat{\varepsilon}_{p_T,t}^2 \right)^{1/2} \tag{6}$$

with $\widehat{\varepsilon}_{p_T,t}$ denoting the OLS residual from estimating (5), and

$$se_{W,2}\,(\widehat{\rho}) := \left(\sum_{t=p_T+1}^{T} x_{t-1}^2 - S_{1\Delta}(S_{\Delta\Delta})^{-1} S_{1\Delta}' \right)^{-1} \left(\sum_{t=p_T+1}^{T} x_{t-1}^2 \widehat{\varepsilon}_{p_T,t}^2 - S_{1\Delta,\varepsilon}(S_{\Delta\Delta,\varepsilon})^{-1} S_{1\Delta,\varepsilon}' \right)^{1/2} \tag{7}$$

with $S_{1\Delta} = \sum_{t=p_T+1}^{T} x_{t-1} \Delta X_{t-1}'$, $S_{\Delta\Delta} = \sum_{t=p_T+1}^{T} \Delta X_{t-1} \Delta X_{t-1}'$, $S_{1\Delta,\varepsilon} = \sum_{t=p_T+1}^{T} x_{t-1} \Delta X_{t-1}' \widehat{\varepsilon}_{p_T,t}^2$ and $S_{\Delta\Delta,\varepsilon} = \sum_{t=p_T+1}^{T} \Delta X_{t-1} \Delta X_{t-1}' \widehat{\varepsilon}_{p_T,t}^2$. In what follows we will focus attention on the $t_{\widehat{\rho},1}^W$ which

provides a simple theoretical benchmark, but we will also provide some results for $t^W_{\hat{\rho},2}$ in order to make comparison with Demetrescu (2010) who derived the asymptotic null distribution of $t^W_{\hat{\rho},2}$ in the pure finite-variance case, $u_t = \varepsilon^*_{1t}$, for the case where $\psi(L)$ is the inverse of a finite $AR(p)$ polynomial. Specifically, from the results in Demetrescu (2010) it follows that under finite variance, homoskedasticity (and some additional technical conditions), the standard ADF statistic $t_{\hat{\rho}}$ and its robust versions $t^W_{\hat{\rho},i}$, $i = 1, 2$, are all asymptotically equivalent. One of our findings in this paper, which will be discussed in section 4, is that this is no longer the case when $\gamma \neq 0$ in (3). Not only does the limit distribution of $t_{\hat{\rho}}$ differ from those of $t^W_{\hat{\rho},i}$, $i = 1, 2$, but $t^W_{\hat{\rho},1}$ and $t^W_{\hat{\rho},2}$ also converge to different limiting distributions under the heavy-tailed error specification in (3).

3.2 Phillips-Perron Tests

The so-called Phillips-Perron [PP] unit root tests of Phillips (1987) and Phillips and Perron (1988) control for weak dependence in u_t through non-parametric correction factors, based on consistent estimates of both the long-run and short-run variances of u_t. The PP test statistics are constructed as

$$\mathcal{Z}_\phi := \quad T\left(\widehat{\phi} - 1\right) - \frac{1}{2}\left(\widehat{\omega}^2_u - \widehat{\sigma}^2_u\right)\left(T^{-2}\sum_{t=1}^{T} x^2_{t-1}\right)^{-1} \tag{8}$$

$$\mathcal{Z}_t := \quad \frac{\widehat{\sigma}_u}{\widehat{\omega}_u} t_{\widehat{\phi}=1} - \frac{1}{2}\left(\widehat{\omega}^2_u - \widehat{\sigma}^2_u\right)\left(T^{-2}\widehat{\omega}^2_u\sum_{t=1}^{T} x^2_{t-1}\right)^{-1/2} \tag{9}$$

where $\widehat{\phi}$ is the OLS estimate of the slope parameter obtained from regressing x_t on x_{t-1} ($t = 1, ..., T$), and $t_{\widehat{\phi}=1} := \widehat{\sigma}_u^{-1}(\widehat{\phi}-1)(\sum_{t=1}^{T} x^2_{t-1})^{1/2}$, with $\widehat{\sigma}^2_u := T^{-1}\sum_{t=1}^{T}\widehat{u}^2_t$ and $\widehat{u}_t := x_t - \widehat{\phi}x_{t-1}$. In the pure finite variance case, $\widehat{\sigma}^2_u$ is a consistent estimator of the short run variance $\sigma^2_u := E(u^2_t) = \Psi^2\sigma^2_1$, $\Psi^2 := \sum_{j=0}^{\infty}\psi^2_j$. Finally, $\widehat{\omega}^2_u$, is defined such that, at least in the pure finite variance case, it is a consistent estimator of the long-run variance $\omega^2_u := 2\pi f_u(0) = \{\psi(1)\}^2\sigma^2_1$, where $f_u(\cdot)$ denotes the spectral density of u_t.

In the pure finite variance context, $u_t = \varepsilon^*_{1t}$, following Perron and Ng (1996), two alternative estimators for the long-run variance are usually considered. First, a non-parametric kernel estimator based on the sample autocovariances, $\widehat{\omega}^2_u = s^2_{WA}$, with $s^2_{WA} := \sum_{h=-T+1}^{T-1} K(h/m)\hat{\gamma}_h$, where $\hat{\gamma}_h := T^{-1}\sum_{t=1}^{T-|h|}\widehat{u}_t\widehat{u}_{t+|h|}$, $K(\cdot)$ is a kernel function satisfying the general conditions reported in Jansson (2002, Assumption A3), and the bandwidth parameter $m > 0$ satisfies the rate condition $1/m + m^2/T \to 0$ as $T \to \infty$ (which corresponds to Assumption A4 of Jansson, 2002). Second, a parametric autoregressive spectral density [ASD] estimator, $\widehat{\omega}^2_u = s^2_{AR}$, of the form suggested by Berk (1974), with $s^2_{AR} := s^2_{p_T}/(1 - \sum_{i=1}^{k}\widehat{\beta}_i)^2$ and $s^2_{p_T} := T^{-1}\sum_{t=p_T+1}^{T}\widehat{\varepsilon}^2_{p_T,t}$, where $\widehat{\varepsilon}_{p_T,t}$ are the OLS residuals from estimating (5). The conditions for the asymptotic validity of the finite order autoregressive approximation in (5) have been discussed in Section 3.1. The consistency of the short-run variance, $s^2_{p_T}$, follows from Chang and Park (2002) and Berk (1974, p.492) who showed that, $s^2_{p_T} := T^{-1}\sum_{t=p_T+1}^{T}\widehat{\varepsilon}^2_{p_T,1t} = T^{-1}\sum_{t=1}^{T}\varepsilon^2_{1t} + o_p(1) \xrightarrow{p} \sigma^2_1$.

Although originally proposed for the finite variance case, the PP test statistics defined in (8) and (9) can also be meaningfully implemented in the pure infinite variance case. In particular, the correction factors used in (8) and (9) attain the same purpose of cleansing the limiting null distributions of the pseudo AR(1) regression statistics $\widehat{\phi}$ and $t_{\widehat{\phi}=1}$ from nuisance parameters arising from weak dependence in u_t. The only difference is that while these correction quantities were previously related to σ^2_1 (which can be interpreted as the quadratic variation at the point 1 of a Brownian motion) they are now related to the quadratic variation at 1 of an infinite-variance random process. We give further details on this in Remark 3.1.

Remark 3.1: In the pure α-stable context ($\alpha < 2$, $u_t = \varepsilon^*_{2t}$), it holds that

$$a_T^{-2} \sum_{t=1}^{T} \varepsilon^{*2}_{2t} \Rightarrow \Psi^2 [\mathcal{U}_\alpha]_1, \tag{10}$$

where $[\mathcal{U}_\alpha]_1$ is the quadratic variation at the point 1 of a Lévy α-stable process $\mathcal{U}_\alpha(r)$ (see Lemma 4.1 below), $\Psi^2 := \sum_{j=0}^{\infty} \psi_j^2$ and $a_T := aT^{1/\alpha}$. From Phillips (1990), when $\widehat{\phi}$ is a T-consistent estimator of ϕ_T, the same result as in (10) holds if ε^*_{2t} is replaced by the residual $\widehat{\varepsilon}^*_{2t} := x_t - \widehat{\phi} x_{t-1}$, and hence, $T a_T^{-2} \widehat{\sigma}_u^2 \Rightarrow \Psi^2 [\mathcal{U}_\alpha]_1$. Furthermore, if $\widehat{\omega}_u^2$ is defined by a kernel procedure (as detailed above), then it is argued in Phillips (1990, p.53) that $T a_T^{-2} \widehat{\omega}_u^2 \Rightarrow \{\psi(1)\}^2 [\mathcal{U}_\alpha]_1$. The same convergence holds for the ASD estimator, $\widehat{\omega}_u^2 = s^2_{AR}$, based on (5). Specifically, in the pure infinite variance context the consistency of the estimators $\widehat{\beta}_j$, $j = 1, ..., p_T$, and the convergence $T a_T^{-2} s^2_{p_T} \Rightarrow [\mathcal{U}_\alpha]_1$ are established in Cavaliere *et al.* (2016b). Although $\widehat{\sigma}_u^2$ and $\widehat{\omega}_u^2$ need to be re-normalised to achieve non-trivial convergence, no re-normalisations are needed in (8) and (9) because the contributions of a_T cancel out, as does $[\mathcal{U}_\alpha]_1$ in the limits of the statistics. □

In both the pure finite variance case and the pure infinite variance case, under suitable regularity conditions, the PP $\mathcal{Z}_\phi$ statistic is known to share the same limiting null distribution as the ADF normalised bias statistic, $Z_{\widehat{\rho}}$ from (5), while $\mathcal{Z}_t$ shares the same limiting null distribution as $t_{\widehat{\rho}}$; see, among others, Phillips (1987), Phillips and Perron (1988), and Phillips (1990).

3.3 $\mathcal{M}$ Unit Root Tests

We also consider the popular trinity of modified or so-called $\mathcal{M}$ unit root tests due to Stock (1999) and Perron and Ng (1996). These are defined as follows:

$$\mathcal{MSB} := \left(\widehat{\omega}_u^{-2} T^{-2} \sum_{t=1}^{T} x_{t-1}^2 \right)^{1/2} \tag{11}$$

$$\mathcal{MZ}_\phi := \left(2T^{-2} \sum_{t=1}^{T} x_{t-1}^2 \right)^{-1} \left(T^{-1} x_T^2 - \widehat{\omega}_u^2 \right) \tag{12}$$

$$\mathcal{MZ}_t := \mathcal{MSB} \times \mathcal{MZ}_\phi \tag{13}$$

where $\widehat{\omega}_u^2$ is as defined previously in the context of the PP statistics in (8) and (9). Haldrup and Jansson (2006) argue that the $\mathcal{M}$ tests, when coupled with the modified AIC lag selection method of Ng and Perron (2001), are preferable to the standard ADF tests in the pure finite variance case because of their superior finite sample properties when weak dependence is present in $\{u_t\}$. Representations for the limiting null distributions of the $\mathcal{M}$ statistics in the pure finite variance case, under suitable regularity conditions, are provided in, among others, Haldrup and Jansson (2006). In particular, $\mathcal{MZ}_\phi$ and $\mathcal{MZ}_t$ are asymptotically equivalent to $Z_{\widehat{\rho}}$ and $\mathcal{Z}_t$, respectively. As with the PP tests discussed above, in the pure infinite-variance case, $u_t = \varepsilon^*_{2t}$, the $\mathcal{M}$ unit root tests can again be meaningfully implemented with no change in (11)-(13). To the best of our knowledge the large sample behaviour of the $\mathcal{M}$ statistics has not been established in the pure infinite variance case. These can be obtained as a special case of the large sample results we will subsequently provide in section 4.2.

3.4 Breitung's Variance Ratio Test

Finally, we will also consider the variance ratio test ($\mathcal{VRT}$) proposed by Breitung (2002),

$$\mathcal{VRT} := T^{-2} \left(\sum_{t=1}^{T} x_t^2 \right)^{-1} \sum_{t=1}^{T} \left(\sum_{j=1}^{t} x_j \right)^2 . \tag{14}$$

The variance ratio test has some appealing properties in the pure finite variance case. First, it requires no correction, either parametric or non-parametric, to account for any weak dependence in u_t. Second, by virtue of this, it has been advocated by some authors (see, for example, Müller, 2008) as a unit root test which avoids the criticisms of Faust (1996) regarding the (theoretical) uncontrollability of the size of unit root tests based around (parametric or non-parametric) corrections for general weak dependence in u_t. Representations for the limiting null distributions of the $\mathcal{VRT}$ statistic in the pure finite variance case are given, under suitable regularity conditions, in Breitung (2002) and Breitung and Taylor (2003). In the pure infinite-variance case, $u_t = \varepsilon_{2t}^*$, the $\mathcal{VRT}$ test can again be meaningfully implemented with no change needed in the formulation of (14). Again to the best of our knowledge, the large sample behaviour of the $\mathcal{VRT}$ statistic has not been established in the pure infinite variance case, and again can be obtained as a special case of the large sample results we give in section 4.2.

4 Large Sample Results

As we have seen, in the pure infinite variance case, $u_t = \varepsilon_{2t}^*$, commonly used unit root tests designed for the pure finite variance case, $u_t = \varepsilon_{1t}^*$, may still be implemented without modifications to the functional form of the test statistics. Their limiting null distributions differ between the two cases, however. In the pure finite variance case these distributions are particular functionals of a Brownian motion process (which is 2-stable), while in the pure infinite variance case they take the form of the corresponding functionals of an α-stable motion (with $\alpha < 2$). In this section we will show that under the Amsler and Schmidt (2012) specification in (4), weak convergence is again obtained (again without modifications to the test statistics) and, typically, the resulting limit distributions are expressible as functionals of a Lévy process with a Wiener and an α-stable ($\alpha < 2$) component, corresponding to the two components in (4). The results we give in this section can therefore be used to evaluate the asymptotic null distributions of the statistics and the asymptotic local power functions of the associated tests. A by-product of our results is that they also enable us to detail the limiting null distributions of the $\mathcal{M}$ and $\mathcal{VRT}$ statistics in the pure infinite variance case.

4.1 Preliminary Results

Here we present some functional and product moment convergence facts needed for our main results. For a semi-martingale $\mathcal{S}$, recall that the associated quadratic variation process is given by $[\mathcal{S}]_r := \mathcal{S}^2(r) - 2\int_0^r \mathcal{S}(u)d\mathcal{S}(u)$, $r \in [0,1]$. In particular, for a standard Brownian motion W the quadratic variation is deterministic: $[\sigma_1 W]_r = \sigma_1^2 r$, whereas for a pure jump Lévy α-stable process ($\alpha < 2$) it is the sum (infinite series) of squared jumps until r, which are countably many. Throughout, the space $D_n[0,1]$ of cadlag functions on $[0,1]$ with values in $\mathbb{R}^n$ is equipped with Skorokhod's (1956) J_1 topology.

We first state Lemma 4.1 which details the joint weak convergence properties of the (scaled) partial sum processes relevant for establishing the large sample behaviour of the unit root statistics from section 3. The proof of Lemma 4.1 is discussed in Remark 4.1.

Lemma 4.1 *Let Assumptions $\mathcal{A}.1 - \mathcal{A}.5$ hold. Then, as $T \to \infty$, it holds that*

$$\left\{ T^{-1/2}\sum_{t=1}^{[rT]}\varepsilon_{1t}^*,\ a_T^{-1}\sum_{t=1}^{[rT]}\varepsilon_{2t},\ T^{-1}\sum_{t=1}^{[rT]}\varepsilon_{1t}^{*2},\ a_T^{-2}\sum_{t=1}^{[rT]}\varepsilon_{2t}^2 \right\} \Rightarrow \left\{\psi(1)\sigma_1 W(r), \mathcal{U}_\alpha(r), \Psi^2\sigma_1^2 r, [\mathcal{U}_\alpha]_r\right\} \tag{15}$$

in $D_4[0,1]$*, and*

$$\left\{ a_T^{-1} \sum_{t=1}^{[rT]} \varepsilon_{2t}^*,\ a_T^{-2} \sum_{t=1}^{[rT]} \varepsilon_{2t}^{*2} \right\} \overset{f.d.d.}{\to} \left\{ \psi(1)\mathcal{U}_\alpha(r), \Psi^2 \left[\mathcal{U}_\alpha\right]_r \right\}, \tag{16}$$

where W *and* $\mathcal{U}_\alpha$ *are, respectively, independent standard Wiener and Lévy* α*-stable processes,* $\overset{f.d.d.}{\to}$ *denotes convergence of the finite dimensional distributions (f.d.d.) jointly with (15),* $\psi(1) := \sum_{j=0}^{\infty} \psi_j$, $\Psi^2 := \sum_{j=0}^{\infty} \psi_j^2$ *and the norming sequence* a_T *is given by* $a_T := aT^{1/\alpha}$.

Remark 4.1: For a proof of the marginal convergence involving the Brownian motion process W alone, see Phillips (1987) and Hamilton (1994, Proposition 17.3, pp.505-506), and for that involving $\mathcal{U}_\alpha$ alone, see Resnick (1986, pp. 94-95), Phillips (1990), Avram and Taqqu (1992), and Phillips and Solo (1992, p.975). As $\{\varepsilon_{1t}\}_{t\in\mathbb{Z}}$ and $\{\varepsilon_{2t}\}_{t\in\mathbb{Z}}$ are independent, then so are W and $\mathcal{U}_\alpha$, and because W is also continuous a.s., convergence is in $D_4[0,1]$ as asserted in Lemma 4.1. □

Remark 4.2: As shown by Avram and Taqqu (1992) - see also Phillips (1990) and Phillips and Solo (1992) - the finite dimensional convergence in (16) cannot be replaced by weak convergence in $D_2[0,1]$, because the linear structure of the errors under heavy-tails induces collisions of large jumps in the partial sum process. As discussed by, among others, Phillips and Solo (1992) and Chan and Zhang (2009), in considering possible dependent noise sequences with infinite variances a coarser topology is required to ensure weak convergence of partial sums. For instance, Chan and Zhang (2009) develop, in a topology known as M_1, interesting results for near-integrated processes under strong dependence and infinite variances. □

Following Phillips (1987) and given the local-to-unity behaviour of our DGP, we define the (Gaussian) Ornstein-Uhlenbeck [OU] process $J_c(r) := \int_0^r e^{(s-r)c} dW(s) = W(r) - c\int_0^r e^{(s-r)c} W(s)ds$, $r \in [0,1]$, generated by the stochastic differential equation $dJ_c(r) = -cJ_c(r)dr + dW(r)$, with $J_c(0) = 0$. Given the specification of u_t considered in (4), we will also require the analogue of J_c in the α-stable context introduced by Chan (1990); that is, a Lévy driven OU process, defined as $\mathfrak{J}_{c,\alpha}(r) := \mathcal{U}_\alpha(r) - c\int_0^r e^{(s-r)c}\mathcal{U}_\alpha(s)ds$, $r \in [0,1]$, and which results from $d\mathfrak{J}_{c,\alpha}(r) = -c\mathfrak{J}_{c,\alpha}(r)dr + d\mathcal{U}_\alpha(r)$, with $\mathfrak{J}_{c,\alpha}(0) = 0$.

In Lemma 4.2 we now provide intermediate results which will enable us subsequently to provide characterisations of the limit distributions of the unit root test statistics outlined in Section 2. Lemma 4.2 is formulated in terms of the process $H_{c,\nu} := J_c + \nu\mathfrak{J}_{c,\alpha}$, where $\nu := \gamma/\sigma_1$. In particular, relating to the null hypothesis, $\mathsf{H_0}$ (where $c = 0$), $H_{0,\nu} = W + \nu\mathcal{U}_\alpha$ and its quadratic variation $[H_{0,\nu}]_r$ satisfies $[H_{0,\nu}]_r = [W]_r + [\nu\mathcal{U}_\alpha]_r = r + \nu^2[\mathcal{U}_\alpha]_r$, $r \in [0,1]$. Although under the conditions of Lemma 4.1 the weak convergence of $T^{-1/2}\sum_{t=1}^{[rT]} u_t$ to $\psi(1)\sigma_1 H_{0,\nu}(r)$ holds only in f.d.d. sense, it is still true that $T^{-1/2}\sum_{t=1}^{[rT]} \varepsilon_t \Rightarrow \sigma_1 H_{0,\nu}(r)$ in $D[0,1]$, which is sufficient for Lemma 4.2.

Lemma 4.2 *Let* $\{x_t\}$ *be generated according to the DGP in (1) and (3), and let Assumptions*

$\mathcal{A}.1 - \mathcal{A}.5$ *hold. Then, as* $T \to \infty$*, the following convergence results hold jointly:*

$$
\begin{aligned}
i) \quad & T^{-1}\sum_{t=1}^{T} u_t^2 \;\Rightarrow\; \Psi^2\sigma_1^2[H_{0,\nu}]_1 \\
ii) \quad & T^{-3/2}\sum_{t=1}^{T} x_t \;\Rightarrow\; \psi(1)\sigma_1\int_0^1 H_{c,\nu}(r)dr \\
iii) \quad & T^{-2}\sum_{t=1}^{T} x_{t-1}^2 \;\Rightarrow\; \{\psi(1)\}^2\sigma_1^2\int_0^1 H_{c,\nu}^2(r)dr \\
iv) \quad & T^{-1}\sum_{t=1}^{T} x_{t-1}u_t \;\Rightarrow\; \{\psi(1)\}^2\sigma_1^2\left\{\int_0^1 H_{c,\nu}(r)dH_{0,\nu}(r) + \frac{1}{2}\left(1 - \frac{\Psi^2}{\{\psi(1)\}^2}\right)[H_{0,\nu}]_1\right\} \\
v) \quad & T^{-1}\sum_{t=1}^{T} x_{t-1}\varepsilon_t \;\Rightarrow\; \psi(1)\sigma_1^2\int_0^1 H_{c,\nu}(r)dH_{0,\nu}(r) \\
vi) \quad & T^{-2}\sum_{t=1}^{T} x_{t-1}^2\varepsilon_t^2 \;\Rightarrow\; \{\psi(1)\}^2\sigma_1^4\int_0^1 H_{c,\nu}^2(r)d[H_{0,\nu}]_r
\end{aligned}
$$

where $H_{c,\nu} := J_c + \nu\mathfrak{J}_{c,\alpha}$*, with* $\nu := \gamma/\sigma_1$*, and where* $[H_{0,\nu}]_1 = 1 + \nu^2[\mathcal{U}_\alpha]_1$*,* $\Psi^2 := \sum_{j=0}^{\infty}\psi_j^2$ *and* $\psi(1) := \sum_{j=0}^{\infty}\psi_j$*.*

Remark 4.3: In Lemma 4.2(vi), the presence of a stochastic integral with respect to the quadratic variation $[H_{0,\nu}]_r$ is natural on noting that $T^{-2}\sum_{t=1}^{T} x_{t-1}^2\varepsilon_t^2$ is the quadratic variation at 1 of the process $T^{-1}\sum_{t=1}^{[Tr]} x_{t-1}\varepsilon_t$, $r \in [0,1]$. As the latter process has $\mathcal{I}(r) := \psi(1)\sigma_1^2\int_0^r H_{c,\nu}(s)dH_{0,\nu}(s)$, $r \in [0,1]$, as its weak limit, it is then not surprising that $T^{-2}\sum_{t=1}^{T} x_{t-1}^2\varepsilon_t^2$ converges weakly to the quadratic variation at 1 of $\mathcal{I}$, which is given by $[\mathcal{I}]_1 := \{\psi(1)\}^2\sigma_1^4\int_0^1 H_{c,\nu}^2(r)d[H_{0,\nu}]_r$. It could also be useful to note that

$$\int_0^1 H_{c,\nu}^2(r)d[H_{0,\nu}]_r = \int_0^1 H_{c,\nu}^2(r)dr + \nu^2\sum_{r\in(0,1]} H_{c,\nu}^2(r-)\{\Delta\mathcal{U}_\alpha(r)\}^2,$$

where $\int_0^1 H_{c,\nu}^2(r)dr = \left[\int_0^r H_{c,\nu}(s)dW(s)\right]_1$ and $\sum_{r\in(0,1]} H_{c,\nu}^2(r-)\{\Delta\mathcal{U}_\alpha(r)\}^2 = \left[\int_0^r H_{c,\nu}(s)d\mathcal{U}_\alpha(s)\right]_1$. Moreover, the non-zero jumps $\Delta\mathcal{U}_\alpha(r)$, $r \in (0,1]$, are countably many, and $\sum_{r\in(0,1]} H_{c,\nu}^2(r-)\{\Delta\mathcal{U}_\alpha(r)\}^2$ converges a.s. because $H_{c,\nu}$ is bounded a.s. and $\sum_{r\in(0,1]}\{\Delta\mathcal{U}_\alpha(r)\}^2 = [\mathcal{U}_\alpha]_1 < \infty$, a.s. □

Recall from section 3.2 that the PP unit root tests are based on modifications of the $T(\widehat{\phi} - 1)$ and $t_{\widehat{\phi}=1}$ statistics, obtained from the regression of x_t on x_{t-1}. In Proposition 4.1 we detail the large sample properties of these statistics when x_t is generated by the DGP (1) and (3). We will subsequently use these results in section 4.2 to derive the large sample distributions of the corresponding PP tests, $\mathcal{Z}_\phi$ and $\mathcal{Z}_t$.

Proposition 4.1 *Let the conditions of Lemma 4.2 hold. Then, as* $T \to \infty$*, the following weak convergence results hold:*

$$T\left(\widehat{\phi} - 1\right) \Rightarrow -c + \frac{\int_0^1 H_{c,\nu}(r)dH_{0,\nu}(r) + \frac{[H_{0,\nu}]_1}{2}\left(1 - \frac{\Psi^2}{\{\psi(1)\}^2}\right)}{\int_0^1 H_{c,\nu}^2(r)dr}$$

and

$$t_{\widehat{\phi}=1} \Rightarrow \frac{\psi(1)}{\Psi[H_{0,\nu}]_1^{1/2}} \left\{ -c \left(\int_0^1 H_{c,\nu}^2(r)dr \right)^{1/2} + \frac{\int_0^1 H_{c,\nu}(r)dH_{0,\nu}(r) + \frac{[H_{0,\nu}]_1}{2}\left(1 - \frac{\Psi^2}{\{\psi(1)\}^2}\right)}{\left(\int_0^1 H_{c,\nu}^2(r)dr\right)^{1/2}} \right\}$$

where $H_{c,\nu}$, $[H_{0,\nu}]_1$, $\psi(1)$ *and* Ψ *are as defined in Lemma 4.2.*

Remark 4.4: The proof of Proposition 4.1 follows straightforwardly from the results given in Lemma 4.2. Hence, as is the case for both the pure finite variance and pure infinite variance cases (see, *inter alia*, Phillips, 1987, and Phillips, 1990, respectively), the limit distributions of $T(\widehat{\phi}-1)$ and $t_{\widehat{\phi}=1}$ are not nuisance parameter free, and therefore any procedure based on these statistics needs to employ corrections for any weak dependence present in u_t, and in particular those discussed in section 3.2. □

4.2 Limiting Distributions of Unit Root Statistics

In Theorem 4.1 we now collect together representations for the limiting distributions of the unit root statistics detailed in section 3, with the exception of the second heteroskedasticity-robust ADF statistics $t_{\hat{\rho},2}^W$, from section 3.1. This turns out to behave rather differently from the other statistics and we will discuss aspects of this statistic in section 4.3.

Theorem 4.1 *Let the conditions of Lemma 4.2 hold. Then, as* $T \to \infty$*, the following weak convergence results hold:*

i) Provided the lag length p_T *in (5) satisfies the rate condition that* $1/p_T + p_T^3/T \to 0$*, as* $T \to \infty$*, then the ADF test statistics from section 3.1 satisfy,*

$$\begin{aligned}
Z_{\widehat{\rho}} &\Rightarrow -c + \frac{\int_0^1 H_{c,\nu}(r)dH_{0,\nu}(r)}{\int_0^1 H_{c,\nu}^2(r)dr} =: \mathcal{Z}_\infty, \\
t_{\widehat{\rho}} &\Rightarrow -c\frac{\left(\int_0^1 H_{c,\nu}^2(r)dr\right)^{1/2}}{[H_{0,\nu}]_1^{1/2}} + \frac{\int_0^1 H_{c,\nu}(r)dH_{0,\nu}(r)}{[H_{0,\nu}]_1^{1/2}\left(\int_0^1 H_{c,\nu}^2(r)dr\right)^{1/2}} =: t_\infty, \\
t_{\hat{\rho},1}^W &\Rightarrow -c\frac{\int_0^1 H_{c,\nu}^2(r)dr}{\left(\int_0^1 H_{c,\nu}^2(r)d[H_{0,\nu}]_r\right)^{1/2}} + \frac{\int_0^1 H_{c,\nu}(r)dH_{0,\nu}(r)}{\left(\int_0^1 H_{c,\nu}^2(r)d[H_{0,\nu}]_r\right)^{1/2}};
\end{aligned}$$

ii) Provided the conditions stated on $\hat{\omega}_u^2$ *in section 3.2 hold, then the PP statistics are such that,*

$$\mathcal{Z}_\phi \Rightarrow \mathcal{Z}_\infty \text{ and } \mathcal{Z}_t \Rightarrow t_\infty;$$

iii) Again provided the conditions stated on $\hat{\omega}_u^2$ *in section 3.2 hold, then the* $\mathcal{M}$ *statistics from*

section 3.3 are such that,

$$\mathcal{MSB} \Rightarrow \left(\frac{\int_0^1 H_{c,\nu}^2(r)dr}{[H_{0,\nu}]_1} \right)^{1/2} \quad , \quad \mathcal{MZ}_\phi \Rightarrow \mathcal{Z}_\infty \ , \ \textit{and} \ \mathcal{MZ}_t \Rightarrow t_\infty;$$

iv) Breitung's variance ratio statistic from section 3.4 is such that,

$$\mathcal{VRT} \Rightarrow \frac{\int_0^1 \left(\int_0^r H_{c,\nu}(s)ds \right)^2}{\int_0^1 H_{c,\nu}^2(r)dr}.$$

Remark 4.5: All of the limiting distributions given in Theorem 4.1 can be seen to depend, regardless of whether $\mathsf{H_0}$ holds or not, on two key parameters: the maximal moment exponent, α, characterising the Lévy process $\mathcal{U}_\alpha$, and the relative weight $\nu := \gamma/\sigma_1$ of the infinite-variance component. It is, however, important to observe that these limiting representations are all invariant to any weak dependence present in u_t. For all of the statistics given, representations for their limiting null distributions obtain setting $c = 0$, while expressions for their asymptotic local power functions can be obtained from the representations with $c \neq 0$. □

Remark 4.6: The results given in Theorem 4.1 include as special cases the limiting distributions of unit root test statistics under the pure finite variance and pure infinite variance specifications. The pure finite variance ($u_t = \varepsilon_{1t}^*$) limits obtain by setting $\gamma = \nu = 0$ and $H_{c,0} = J_c$ ($H_{0,0} = W$) in the limiting representations, whereas the limiting distributions in the pure infinite variance case ($u_t = \varepsilon_{2t}^*$) obtain by letting $\nu \to \infty$ and replacing $H_{c,\nu}$ by $\mathfrak{J}_{c,\alpha}$ ($H_{0,\nu}$ by $\mathcal{U}_\alpha$) in the limits from Theorem 4.1. For example, in the pure finite variance case we then have (noting that the stochastic differential equation satisfied by $J_c(r)$ is used to write the numerator of (17) more succinctly) that

$$t_{\widehat{\rho},1}^W \Rightarrow \frac{\int_0^1 J_c(r)dJ_c(r)}{\left(\int_0^1 J_c^2(r)dr \right)^{1/2}}, \tag{17}$$

which is the same as the limiting distribution for the standard ADF statistic $t_{\widehat{\rho}}$ in the pure finite variance case, whereas in the pure infinite variance case the limit distributions of the two statistics are distinct, *viz*:

$$t_{\widehat{\rho}} \Rightarrow \frac{\int_0^1 \mathfrak{J}_{c,\alpha}(r)d\mathfrak{J}_{c,\alpha}(r)}{([\mathcal{U}_\alpha]_1)^{1/2} \left(\int_0^1 \mathfrak{J}_{c,\alpha}^2(r)dr \right)^{1/2}}, \quad \text{and} \quad t_{\widehat{\rho},1}^W \Rightarrow \frac{\int_0^1 \mathfrak{J}_{c,\alpha}(r)d\mathfrak{J}_{c,\alpha}(r)}{\left(\int_0^1 \mathfrak{J}_{c,\alpha}^2(r)d[\mathcal{U}_\alpha]_r \right)^{1/2}}. \tag{18}$$

Notice, moreover, that for $\alpha = 2$ the two limiting distributions given in (18) both reduce to the right member of (17). □

Remark 4.7: The large sample results given in Theorem 4.1 relate to the case where no deterministic component is allowed for in computing the unit root statistic in question. These results can be generalised to the case where the statistics are computed using de-meaned or de-trended data, as is typically done in the unit root testing literature. This involves replacing x_t in the computation of the statistics from section 3 by either: the residual $\widehat{x}_t := x_t - z_t'\widehat{\theta}$ obtained from the OLS regression of x_t on $z_t := 1$, where $\theta := \mu$ or on $z_t := (1, t)'$, where $\theta := (\mu, \tau)$; or, by setting $\overline{\phi}_T = 1 - \overline{c}/T$ (where

the value of $\overline{c}$ is pre-specified according to the deterministic specification considered and the desired significance level), the residual $\widetilde{x}_t := x_t - z_t'\widetilde{\theta}$, with $\widetilde{\theta}$ obtained from the quasi-differenced (GLS) regression of $X_{\overline{c}} := (x_0,\, x_1 - \overline{\phi}_T x_0,..,\, x_T - \overline{\phi}_T x_{T-1})'$ on $Z_{\overline{c}} := (z_0,\, z_1 - \overline{\phi}_T z_0,..,\, z_T - \overline{\phi}_T z_{T-1})'$, as in Elliott *et al.* (1996), where again $z_t := 1$ when de-meaned data is required and $z_t := (1,\, t)'$ when de-trended data is required. In terms of the limiting distributions given in Proposition 4.1 and Theorem 4.1, these remain valid in such cases provided the limiting processes $H_{0,\nu}$ and $H_{c,\nu}$, and the functionals thereof, are replaced by their appropriate de-meaned ($H_{0,\nu}^{k,\mu}$ and $H_{c,\nu}^{k,\mu}$, $k = OLS$ or GLS) or de-trended ($H_{0,\nu}^{k,\tau}$ and $H_{c,\nu}^{k,\tau}$, $k = OLS$ or GLS) counterparts. In particular, $H_{i,\nu}^{OLS,\mu}(r) := H_{i,\nu}(r) - \int_0^1 H_{i,\nu}(s)ds$, $H_{i,\nu}^{OLS,\tau}(r) := H_{i,\nu}^{OLS,\mu}(r) - 12(r-1/2)\int_0^1 (s-1/2)H_{i,\nu}(s)ds$, $H_{i,\nu}^{GLS,\tau}(r) := H_{i,\nu}(r) - r\{\overline{c}^* H_{i,\nu}(1) + 3(1-\overline{c}^*)\int_0^1 sH_{i,\nu}(s)ds\}$, for $i = 0, c$, and where $\overline{c}^* := (1+\overline{c})/(1+\overline{c}+\overline{c}^2/3)$. As in Müller and Elliott (2003), in the case where OLS de-meaning or de-trending is employed, the term $-T\widehat{x}_0^2$ (which weakly converges to $H_{c,\nu}^{OLS,\mu}(0)^2$ under de-meaning and to $H_{c,\nu}^{OLS,\tau}(0)^2$ under de-trending) needs to be added to the numerator of $\mathcal{MZ}_\phi$ from section 3.3 in order that $\mathcal{MZ}_\phi$ and $\mathcal{MZ}_t$ remain asymptotically equivalent to $Z_{\widehat{\rho}}$ and $t_{\widehat{\rho}}$, respectively. □

4.3 Further Results Relating to the Eicker-White ADF Tests

We conclude this section with some further considerations regarding the limit distributions of the statistics $t_{\hat{\rho},1}^W$ and $t_{\hat{\rho},2}^W$ from section 3.1, which modify the corresponding standard ADF test statistics through the use of Eicker-White rather than OLS standard errors. In section 4.3.1 we first explore the relationship between the limiting distributions of $t_{\hat{\rho}}$ and $t_{\hat{\rho},1}^W$ in order to try and shed some light on some of the findings we subsequently make in our Monte Carlo study in section 5. Secondly, in section 4.3.2 we present some material relating to the limiting distribution of $t_{\hat{\rho},2}^W$.

4.3.1 Further Discussion of the Large Sample Behaviour of $t_{\hat{\rho},1}^W$ Relative to $t_{\hat{\rho}}$

Under the conditions of Theorem 4.1, the limiting distribution of the ratio of $t_{\hat{\rho},1}^W$ to $t_{\hat{\rho}}$ can be shown to be given by

$$\frac{t_{\hat{\rho},1}^W}{t_{\hat{\rho}}} \Rightarrow \left(\frac{[H_{0,\nu}]_1 \int_0^1 H_{c,\nu}^2(r)dr}{\int_0^1 H_{c,\nu}^2(r)d[H_{0,\nu}]_r} \right)^{1/2}. \tag{19}$$

In the pure finite variance case ($u_t = \varepsilon_{1t}^*$), it is known from Demetrescu (2010) that $t_{\hat{\rho}}$ and $t_{\hat{\rho},1}^W$ are asymptotically equivalent, such that the limit in (19) equal to 1 (with the same result holding for the ratio of $t_{\hat{\rho},2}^W$ to $t_{\hat{\rho}}$); see also Remark 4.6. In particular, $T^{-2}\sum_{t=1}^T x_{t-1}^2\varepsilon_t^2 = \sigma_1^2 T^{-2}\sum_{t=1}^T x_{t-1}^2 + o_p(1)$ and $s_{p_T}^2 T^{-2}\sum_{t=1}^T x_{t-1}^2 = \sigma_1^2 T^{-2}\sum_{t=1}^T x_{t-1}^2 + o_p(1)$ jointly converge to the same weak limit, so that the impact of the Eicker-White correction is asymptotically negligible.

In contradistinction, in the pure infinite variance case ($u_t = \varepsilon_{2t}^*$), the analogue of the result in Lemma 4.2(vi) is $T^{-2}\sum_{t=1}^T x_{t-1}^2\varepsilon_t^2 \Rightarrow \gamma^4\{\psi(1)\}^2 \int_0^1 \mathfrak{J}_{c,\alpha}^2(r)d\,[\mathcal{U}_\alpha]_r$, whereas $s_{p_T}^2 T^{-2}\sum_{t=1}^T x_{t-1}^2 \Rightarrow \gamma^4\{\psi(1)\}^2\,[\mathcal{U}_\alpha]_1 \int_0^1 \mathfrak{J}_{c,\alpha}^2(r)dr$, with the convergence joint and the two weak limits distinct. As a result, in this case it holds that

$$\frac{t_{\hat{\rho},1}^W}{t_{\hat{\rho}}} \Rightarrow \left(\frac{[\mathcal{U}_\alpha]_1 \int_0^1 \mathfrak{J}_{c,\alpha}^2(r)dr}{\int_0^1 \mathfrak{J}_{c,\alpha}^2(r)d[\mathcal{U}_\alpha]_r} \right)^{1/2} =: \tau_{c,\alpha}. \tag{20}$$

In order to gain some insight into how the non-trivial limit in (19) is likely to manifest itself in the behaviour of the test statistics, we will focus attention on the pure infinite variance case where

$\sigma_1 = 0$. We might expect that the qualitative conclusions from this exercise would be maintained, at least for relatively small values of σ_1, when $\sigma_1 \neq 0$. This is indeed confirmed by the Monte Carlo simulation results later reported in section 5. To that end, consider decomposing the weak limit of $t_{\hat{\rho}}$ in the pure infinite variance case into $t_\infty = t_\infty^{(1)} + t_\infty^{(2)}$ with

$$t_\infty^{(1)} := -c\frac{\left(\int_0^1 \mathfrak{J}_{c,\alpha}^2(r)dr\right)^{1/2}}{[\mathcal{U}_\alpha]_1^{1/2}} \qquad \text{and} \qquad t_\infty^{(2)} := \frac{\int_0^1 \mathfrak{J}_{c,\alpha}^2 d\mathcal{U}_\alpha(r)}{[\mathcal{U}_\alpha]_1^{1/2}\left(\int_0^1 \mathfrak{J}_{c,\alpha}^2(r)dr\right)^{1/2}}.$$

Under $\mathsf{H_0}$, $t_\infty^{(1)} = 0$ and so the limiting null distributions of $t_{\hat{\rho}}$ and $t_{\hat{\rho},1}^W$ are given by $t_\infty^{(2)}$ and $\tau_{0,\alpha}t_\infty^{(2)}$, respectively. Under $\mathsf{H_c}$ ($c > 0$), however, a shift leftwards by, respectively, $t_\infty^{(1)}$ and $\tau_{c,\alpha}t_\infty^{(1)}$ occurs.

To quantify these effects, we investigate two special cases: first, the null hypothesis, $\mathsf{H_0}$, where $c = 0$, and second one particular local alternative, $\mathsf{H_c}$ for $c = 15$. For that purpose, in Table 1 we report simulations[1] of the cumulative distribution function [cdf] of $\tau_{c,\alpha}$ for $\alpha \in \{1,\ 1.5,\ 1.75\}$, both unconditionally and conditionally on the event that $t_\infty^{(2)} < -1.95$.[2] Specifically, Table 1 reports values for $F_{c,\alpha}(x) = P[\tau_{c,\alpha} \leq x]$, which is the unconditional cdf of $\tau_{c,\alpha}$, and $F_{c,\alpha}(x|t_\infty^{(2)} < -1.95) = P[\tau_{c,\alpha} \leq x|t_\infty^{(2)} < -1.95]$, which corresponds to a cdf of $\tau_{c,\alpha}$ conditional on $t_\infty^{(2)}$ taking a (relatively) large negative value. The purpose is to see how frequently and by what magnitude the multiplication of $t_\infty^{(1)}$ and $t_\infty^{(2)}$ by $\tau_{c,\alpha}$, resulting from the Eicker-White correction, inflates or deflates the statistics, both in general and for outcomes in the left tail of $t_\infty^{(2)}$ which are conducive to rejections of $\mathsf{H_0}$. Although the value of -1.95 is chosen as the fifth percentile of the Dickey-Fuller distribution in the finite variance case, qualitative conclusions are the same if the critical value is taken from the limit null distribution for the true α.

[Please insert Table 1 about here]

In Table 1, the simulated probabilities for $x = 1$ correspond to the event that $|t_{1,\infty}^W| \leq |t_\infty|$, where $t_{1,\infty}^W$ is the weak limit of $t_{\hat{\rho},1}^W$, and are highlighted in bold. Starting from the null hypothesis $c = 0$, the results suggest that the unconditional distribution of $\tau_{0,\alpha} = t_{1,\infty}^W/t_\infty$ displays positive asymmetry for all values of α considered, yielding a predominance of cases where $|t_{1,\infty}^W| > |t_\infty|$. In other words, and with probabilities ranging between 0.8 and 0.6 (decreasing in α), it is seen that in large samples the Eicker-White correction gives rise to a larger (in magnitude) value of the t-statistic, and that this effect is more pronounced the smaller is α. Nevertheless, because $\tau_{c,\alpha}$ and t_∞ are not independent, this does not imply, *per se*, that the Eicker-White corrected ADF test will reject $\mathsf{H_0}$ more frequently than the standard ADF test based on $t_{\hat{\rho}}$, even if the same critical value is used in both tests. In fact, when $\mathsf{H_0}$ is true, the results for the cdf of $\tau_{0,\alpha}$ conditional on the occurrence of a large negative value of $t_\infty = t_\infty^{(2)}$ (conducive to rejections of $\mathsf{H_0}$), are the exact opposite of the general case, with predominance of outcomes where $|t_{1,\infty}^W| \leq |t_\infty|$ and the respective probabilities ranging from 0.6 to 0.8 (decreasing in α). Thus, although under $\mathsf{H_0}$ the Eicker-White correction more frequently inflates $|t_\infty|$ than deflates it, in the proximity of the conventional critical regions for $\mathsf{H_0}$ the situation is reversed and $|t_\infty|$ is more frequently deflated. In the Monte Carlo results reported in section 5 it will be seen that this leads to undersizing in the test based on $t_{\hat{\rho},1}^W$.

Turning to $c = 15$, the unconditional cdf of $\tau_{15,\alpha}$ is essentially unaltered compared to that of $\tau_{0,\alpha}$, while variations in the simulated conditional cdf are minor.[3] The conclusions drawn under

[1] The random processes are discretised over a grid of 10000 points and 10000 replications are performed.

[2] For both $c = 0$ and $c = 15$, the event $\{t_\infty^{(2)} < -1.95\}$ occurred with simulated frequencies of 0.028, 0.041 and 0.048 for $\alpha = 1$, $\alpha = 1.5$ and $\alpha = 1.75$, respectively.

[3] It should of course be noted that the conditional cdf is less precisely simulated because the proportion of replications where the conditioning event occurs is relatively small by construction.

H_0 for t_∞, therefore, remain valid under $c = 15$ for $t_\infty^{(2)}$. However, t_∞ and $t_{1,\infty}^W$ are now shifted downwards by, respectively, $t_\infty^{(1)}$ and $\tau_{c,\alpha} t_\infty^{(1)}$ which are proportional to $-c$ which, obviously, is independent of $t_\infty^{(2)}$. Therefore, through the downward shift, the unconditionally more frequent inflation than deflation of $-c$ by $\tau_{c,\alpha}$ becomes relevant and, at least for large c, will dominate the contribution of $t_\infty^{(2)}$ and $\tau_{c,\alpha} t_\infty^{(2)}$ to t_∞ and $t_{1,\infty}^W$. In the Monte Carlo results reported in section 5 it will be seen that this leads to power gains in the test based on $t_{\hat{\rho},1}^W$ relative to that based on $t_{\hat{\rho}}$.

4.3.2 Some Limiting Results for $t_{\hat{\rho},2}^W$

We conclude this section by reporting some asymptotic results for the Eicker-White corrected ADF $t_{\hat{\rho},2}^W$ statistic. Our purpose here is to show that under the Amsler and Schmidt (2012) infinite variance specification in (4), the statistics $t_{\hat{\rho},1}^W$ and $t_{\hat{\rho},2}^W$ are not asymptotically equivalent. Specifically, in large samples $t_{\hat{\rho},1}^W$ tends to be smaller in magnitude than $t_{\hat{\rho},2}^W$ and the ratio $t_{\hat{\rho},1}^W / t_{\hat{\rho},2}^W$ is bounded away from one, in probability, potentially leading to asymptotically more frequent rejections using $t_{\hat{\rho},2}^W$ than $t_{\hat{\rho},1}^W$, for the same critical value. As this turns out to be a consequence of the interaction between the two additive terms, ε_{1t} and ε_{2t} in (4), which is an artefact of the Amsler-Schmidt specification, we restrict ourselves to a rigorous yet indirect argument that avoids any explicit representation of the limiting distribution of $t_{\hat{\rho},2}^W$. Moreover, we focus on the case where x_t is a random walk because the lack of asymptotic equivalence in this special case implies that asymptotic equivalence does not hold in general either. We stress that the random walk case is the simplest, yet by no means exceptional, counterexample to asymptotic equivalence.

Our reasoning is as follows. Recall definitions (6) and (7). On the one hand, it holds that

$$\sum_{t=p_T+1}^{T} x_{t-1}^2 \widehat{\varepsilon}_{p_T,t}^2 - S_{1\Delta,\varepsilon} (S_{\Delta\Delta,\varepsilon})^{-1} S_{1\Delta,\varepsilon}' \leq \sum_{t=p_T+1}^{T} x_{t-1}^2 \widehat{\varepsilon}_{p_T,t}^2$$

by considerations of positive definiteness, and we also showed in the proof of Theorem 4.1 that, under its assumptions,

$$T^{-2} \sum_{t=p_T+1}^{T} x_{t-1}^2 - T^{-2} S_{1\Delta} (S_{\Delta\Delta})^{-1} S_{1\Delta}' = T^{-2} \sum_{t=p_T+1}^{T} x_{t-1}^2 + o_p(1), \tag{21}$$

where $T^{-2} \sum_{t=p_T+1}^{T} x_{t-1}^2$ is bounded away from 0 in probability. Therefore, for every $\delta > 0$,

$$P\left(t_{\hat{\rho},1}^W / t_{\hat{\rho},2}^W \leq 1 + \delta\right) = P\left(se_{W,2}\left(\widehat{\rho}\right) / se_{W,1}\left(\widehat{\rho}\right) \leq 1 + \delta\right) \to 1$$

as $T \to \infty$, under the same conditions as in Theorem 4.1. This could be restated as $\max\left\{t_{\hat{\rho},1}^W / t_{\hat{\rho},2}^W, 1\right\} \xrightarrow{P} 1$ and interpreted as meaning that in large samples $t_{\hat{\rho},1}^W$ tends to be no larger in magnitude than $t_{\hat{\rho},2}^W$. The same result also holds in the pure finite variance case and does not, by itself, preclude the asymptotic equivalence of $t_{\hat{\rho},1}^W$ and $t_{\hat{\rho},2}^W$.

On the other hand, however, let x_t be a pure random walk with $x_0 = 0$, so that $\Delta x_t = \varepsilon_t$. In this case we establish, under Assumption $\mathcal{A}$, the existence of a random variable θ with $P(\theta \in (0,1)) = 1$ such that, for every $z \in \mathbb{R}$ where the cdf of θ is continuous, it holds that

$$\liminf_{T\to\infty} P\left(t_{\hat{\rho},1}^W / t_{\hat{\rho},2}^W \leq z\right) \geq P\left(\theta \leq z\right). \tag{22}$$

In particular, it can be concluded that $P(t_{\hat{\rho},1}^W / t_{\hat{\rho},2}^W < 1) \to 1$ and, moreover, that $t_{\hat{\rho},1}^W / t_{\hat{\rho},2}^W$ is bounded away from one, in probability.

Details about the random variable θ are given in Proposition 4.2, which relies on the weak convergence of the process $\sum_{t=1}^{[Tr]} (\varepsilon_t, \varepsilon_{2t}^2 \boldsymbol{\varepsilon}_{1t}')'$ upon appropriate normalisation, where we define $\boldsymbol{\varepsilon}_{1t} := (1, \varepsilon_{1,t-1}, \varepsilon_{1,t-1}^2, \varepsilon_{1,t+1}^2)'$.

Proposition 4.2 *Under Assumption $\mathcal{A}$ it holds that*

$$\sum_{t=1}^{[Tr]}(T^{-1/2}\varepsilon_t, a_T^{-2}\varepsilon_{2t}^2\boldsymbol{\varepsilon}_{1t}')' \Rightarrow (\sigma_1 H_{\nu,0}(r), [\mathcal{U}_\alpha]_r, S(r), Q_-(r), Q_+(r))' \tag{23}$$

in $D_5[0,1]$, where S and $Q_\pm$ are $\frac{\alpha}{2}$-stable processes with $Q_\pm \geq 0$ a.s. Furthermore, as a consequence, for the random walk $x_t = \sum_{s=1}^t \varepsilon_s$ inequality (22) holds with

$$\theta = \left(1 - \frac{\nu^4\left[\int_0^1 H_{0,\nu}(r)dS(r)\right]^2}{\{\sigma_1^2 + \nu^2\left[Q_-(1) + Q_+(1)\right]\}\int_0^1 H_{c,\nu}^2(r)d[H_{0,\nu}]_r}\right)^{1/2}. \tag{24}$$

Remark 4.8: The dependence structure of the limit random process in (23) can be inferred from the proof of the proposition given in the Appendix. However, this fact is secondary with respect to our main point which follows from the functional form of the right-hand side in (24) alone. Specifically, $\theta < 1$ a.s. because $\int_0^1 H_{0,\nu}dS \neq 0$ a.s. □

5 Finite Sample Simulations

In this section we report results from an in-depth Monte Carlo simulation study into the finite sample size (section 5.2) and local power (section 5.2) properties of the unit root tests considered in section 3 under the Amsler and Schmidt (2012) local-to-finite variance specification in (4). Before doing so we first briefly explore in section 5.1 estimation of the tail index of a best stable approximation to local-to-finite variance processes of the form given for u_t in (2).

5.1 Tail Index Estimates

Consider u_t from (2). By design, the tail index of u_t for fixed T coincides with the tail index α of ε_{2t} (whenever $\gamma > 0$), and therefore, is independent of T. Nevertheless, given the decreasing weight of ε_{2t} in the definition of u_t, we may also conceive of pseudo tail indices for u_t that behave continuously and converge to 2 as $T \to \infty$, in agreement with the localisation idea. For instance, we might expect the tail indices of the sequence of stable distributions that, for each T, best approximate the distribution of u_t with respect to some meaningful criterion, to behave in this way. We adopt the approach of Koutrouvelis (1980,1981), where the approximation criterion is defined in terms of proximity of the characteristic functions, and the corresponding tail indices (denoted by α_u) of the best stable approximation can be estimated by Monte Carlo simulation, by applying the estimator of Koutrouvelis to samples of generated u_t's.[4]

To that end, we generate samples of u_t, as $u_t = \varepsilon_{1t} + \gamma T^{-1/\alpha_2+1/2}\varepsilon_{2t}$, for $T = 100$, 1000, 10000 and 100000 observations, with $\varepsilon_{1t} \sim IID\, N(0,1)$ and independent of $\varepsilon_{2t} \sim IID\, S_{\alpha_2}(0,1,0)$. For the infinite variance component, ε_{2t}, we consider values of the tail index $\alpha_2 \in \{1,\ 1.25,\ 1.5,\ 1.75,\ 2\}$. To generate data from a stable distribution with index α_2 we use the method of Samorodnitsky and Taqqu (1994, Proposition 1.7.1). As regards the relative weighting parameter between the finite and infinite variance components, γ, we follow Amsler and Schmidt (2012) and consider $\gamma \in \{0.1, 0.316, 1, 3.16, 10, 31.6\}$.

[Please insert Table 2 about here]

[4] We also investigated the quantile-based estimator of McCulloch (1986) and found the results to be very similar (although it should be stressed that the pseudo parameters estimated by this approach need not coincide with those obtained using Koutrouvelis's method).

Table 2 provides the average value (taken across 1000 Monte Carlo replications) of the tail index estimate, denoted $\widehat{\alpha}_u$, of u_t. The results illustrate the local-to-finite behaviour of the underlying distribution of u_t. In particular, we observe that, other things being equal, and as might be expected, $\widehat{\alpha}_u$ will lie further from 2 the bigger is γ, the smaller is α_2 and the smaller is T. For small values of γ ($\gamma = 0.1$ and $\gamma = 0.316$) with small samples ($T = 100$) and as α_2 approaches 1, the empirical average of the empirical stability index $\widehat{\alpha}_u$ can be well below 2, although it quickly reverts towards 2 as the sample size increases, other things equal. However, it can also be seen from the results in Table 2 that for $\gamma > 0.316$, substantially larger sample sizes are required for the average of the estimated tail index to lie close to 2; for example, when $\gamma = 3.16$ and $\alpha_2 < 2$, even $T = 100000$ is not a sufficiently large sample size to yield a value of the average of $\widehat{\alpha}_u$ close to 2.

5.2 Empirical Size

In this subsection (empirical size) and the next (empirical power) we compare the empirical rejection frequencies [ERFs] of the unit root tests discussed in this paper to data generated according to

$$x_t = \phi_T x_{t-1} + u_t \tag{25}$$

where $x_0 = 0$, $\phi_T := 1 - c/T$ with $c \geq 0$ and fixed. The error process $\{u_t\}$ is generated as

$$(1 - \varphi L)u_t = (1 - \theta L)\varepsilon_t \tag{26}$$

where $\varepsilon_t = \varepsilon_{1t} + \gamma T^{-1/\alpha_2+1/2}\varepsilon_{2t}$, with ε_{1t} and ε_{2t} generated as detailed in section 5.1. Results are reported for $\alpha_2 \in \{1,\ 1.25,\ 1.5,\ 1.75,\ 2\}$ and $\gamma \in \{0.1, 1, 10\}$. The finite sample size and power of the tests is evaluated for samples of size $T = 200$ and $T = 500$, based on 5000 Monte Carlo replications. All simulations were performed in MATLAB 7.8.0 (R2009a). The size and power results reported pertain to the empirically most relevant case where the unit root tests are based on (either OLS or local GLS) de-meaned data. Corresponding results for tests based on de-trended data were qualitatively similar and can be obtained from the authors on request.

For, the finite sample size analysis reported in this section we set $\phi_T = 1$ ($c = 0$), such that the unit root null hypothesis holds, and allow for autoregressive and moving average dynamics in the error term in (26) by allowing $\varphi \in \{-0.5, 0, 0.5\}$ and $\theta \in \{-0.5, 0, 0.5\}$, respectively. Results are reported for tests based on the ADF t-statistic, $t_{\widehat{\rho}}$, Breitung's variance ratio statistic, $\mathcal{VRT}$, the trinity of $\mathcal{M}$ statistics, $\mathcal{MSB}$, $\mathcal{MZ}_\phi$ and $\mathcal{MZ}_t$, and the Eicker-White corrected ADF $t^W_{\widehat{\rho},1}$ statistic. Results are not reported here for the PP tests from section 3.2 or for the normalised bias ADF test $Z_{\widehat{\rho}}$ from section 3.1, because these displayed very unreliable finite sample size properties in the presence of serial correlation, a phenomenon also well documented for the former in the pure finite variance case. In each case the statistics were compared to the standard (Gaussian) asymptotic critical values that would be relevant in the pure finite variance case. We also consider a further test based on the $t^W_{\widehat{\rho},1}$ statistic which, rather than using the standard finite variance critical value, uses a critical value simulated from the limiting null distribution of this statistic based on the pseudo tail index estimate, $\widehat{\alpha}_u$, calculated as outlined in section 5.1. In order to distinguish between these two tests, we will denote the latter by $t^{W,\alpha_u}_{\widehat{\rho},1}$ in what follows. All tests were run at the nominal asymptotic 5% level.

For the implementation of the semi-parametric unit root tests, the reported results relate to the use of parametric ASDEs of the long run variances as we found these to deliver significantly better finite sample performance than the corresponding tests based on sums-of-covariances estimators using either the Bartlett or Quadratic Spectral kernels. The autoregressive lag order used in constructing the ASDEs was determined using modified MAIC criteria with Schwert's rule applied to obtain the maximum lag length allowed; that is, $k_{\max K} := [K\left(\frac{T}{100}\right)^{1/4}]$ with $K = 12$. As in Perron and Qu (2007) the MAIC criteria is computed based on OLS de-trended data. The same lag length selection method was used in the context of the standard ADF test, $t_{\widehat{\rho}}$ and for the

corresponding Eicker-White corrected ADF tests. The empirical size results for these tests are reported in Tables 3-8.

[Please insert Tables 3 - 8 about here]

Consider first the results for $T = 200$ reported in Tables 3-5. In the case where no serial correlation is present in u_t (i.e., $\varphi = \theta = 0$) the ERFs conform well in general to the nominal level. With the exception of the $\mathcal{VRT}$ test based on local GLS de-meaned data, the empirical sizes of the various tests when $\gamma = 0.1$, 1 and 10, lie between [0.035, 0.067], [0.023, 0.060], and [0.003, 0.068], respectively. In general, again with the exception of the local GLS de-meaned $\mathcal{VRT}$ test, the tests appear to become more conservative the smaller is α_2. Again excluding the $\mathcal{VRT}$ test, little differences are seen between the tests based on OLS de-meaning and the corresponding tests based on local GLS de-meaning. The behaviour of the $\mathcal{VRT}$ test is, however, quite different depending on which method of de-meaning is used; while the empirical sizes of $\mathcal{VRT}$ test based on OLS de-meaning show only small deviations from the nominal 5% level, those based on local GLS de-meaning show a tendency to significant over-sizing as α_2 decreases and γ increases. For example, for $\alpha_2 = 1$ and $\gamma = 10$ the empirical size of the local GLS de-meaned $\mathcal{VRT}$ test is close to 11%. In contrast the empirical size of the OLS de-meaned version of the test is 4.4%. It is worth noting that even in the pure finite variance case, the results reported in Breitung and Taylor (2003) show that the $\mathcal{VRT}$ based on OLS de-meaned data is preferable to the version based on local GLS de-meaned data.

Where serial correlation is present in u_t (i.e., $\varphi = \pm 0.5$ or $\theta = \pm 0.5$) the results in Tables 3-5 suggest that in general the empirical sizes of the tests remain quite robust to the heavy tailed behaviour in u_t, particularly bearing in mind the sample size is only $T = 200$ here. These results can essentially be summarised as follows: the size of the $t_{\widehat{\rho}}$, $\mathcal{VRT}$, $\mathcal{MSB}$, $\mathcal{MZ}_\phi$, $\mathcal{MZ}_t$, $t^W_{\widehat{\rho},1}$ and $t^{W,\alpha_u}_{\widehat{\rho},1}$ tests when $\varphi = \pm 0.5$ or $\theta = 0.5$ for $\gamma = 0.1$, 1 and 10, when OLS de-meaning is considered, lie between [0.029, 0.069], [0.023, 0.059], and [0.004, 0.070], respectively; whereas when local GLS de-meaning is used, excluding $\mathcal{VRT}$, they lie between [0.039, 0.076], [0.030, 0.068], and [0.019, 0.074], respectively. As is also well known to occur in the pure finite variance case, empirical sizes deteriorate in the negative moving average case, where $\theta = -0.5$, for some of the tests; in particular, we observe that for $\gamma = 0.1$, 1 and 10, when OLS de-meaning is used the empirical sizes of the tests lies between [0.038, 0.109], [0.042, 0.101] and [0.033, 0.126], respectively, whereas when local GLS de-meaning is used they lie between [0.054, 0.114], [0.043, 0.123], and [0.035, 0.173], respectively. The largest distortions are observed for the Eicker-White corrected statistics, whereas $t_{\widehat{\rho}}$, $\mathcal{MSB}$, $\mathcal{MZ}_\phi$ and $\mathcal{MZ}_t$ are close to the 5% nominal significance level.

Excluding the negative moving average case, the Eicker-White corrected ADF tests, $t^W_{\widehat{\rho},1}$ and $t^{W,\alpha_u}_{\widehat{\rho},1}$, become increasingly conservative as γ increases, other things equal, particularly for the OLS de-meaned versions. For example, the OLS de-meaned $t^W_{\widehat{\rho},1}$ and $t^{W,\alpha_u}_{\widehat{\rho},1}$ tests display ERFs of 0.3% and 1.1%, respectively, when $\gamma = 10$, $\alpha_2 = 1$ and where no serial correlation is present in u_t, and display similar behaviour for $\varphi = \pm 0.5$ or $\theta = 0.5$). The local GLS de-meaned $t^W_{\widehat{\rho},1}$ and $t^{W,\alpha_u}_{\widehat{\rho},1}$ tests also become more conservative as γ increases when $\varphi = 0$, $\varphi = \pm 0.5$ or $\theta = 0.5$ and $\alpha_2 = 1$, but, although the tests are conservative, their ERFs are in most cases considerably higher than those of the corresponding OLS de-meaned tests. Notice that this observed under-sizing seen in the simulation results accords with the discussion in Section 4.3 concerning the asymptotic distributions of the Eicker-White corrected ADF statistics in the benchmark case of no deterministics.

Consider now Tables 6 - 8 which display the corresponding results for $T = 500$. Overall, the relative performance of the tests remains qualitatively similar to what was seen in the results for $T = 200$, although in general the size distortions of the tests improve somewhat relative to $T = 200$. This is particularly so in the serially correlated cases considered, as is of course anticipated by the asymptotic distribution theory presented in section 4.

5.3 Empirical Power

We now turn to a comparison of the finite sample local power properties of the unit root tests. For the local power analysis the Monte Carlo data are generated by (25)-(26) for the local alternatives generated by $c \in \{1, ..., 50\}$. In order to eliminate finite sample differences between the tests which are attributable solely to weak dependence, we set $\varphi = \theta = 0$. Results are reported for the same set of tests as were considered in the finite sample size simulations in section 5.2 except that among the three $\mathcal{M}$ tests we only report results for the $\mathcal{MSB}$ test, as its finite sample local power properties were essentially indistinguishable from those of $\mathcal{MZ}_\phi$ and $\mathcal{MZ}_t$. All comments which follow regarding $\mathcal{MSB}$ therefore apply equally to the corresponding $\mathcal{MZ}_\phi$ and $\mathcal{MZ}_t$ tests.

[Please insert Figures 1 - 8 about here]

Consider first Figures 1-4 which graph the finite sample local power properties of both the OLS and local GLS de-meaned tests for $T = 200$. A comparison of the results in each of these Figures suggests that, for both OLS and local GLS de-meaned data, the finite sample power properties of the $t_{\widehat{\rho}}$, $\mathcal{VRT}$, and $\mathcal{MSB}$ statistics vary relatively little with γ and α_2, while in contrast the two Eicker-White corrected tests, $t^{W}_{\widehat{\rho},1}$ and $t^{W,\alpha_u}_{\widehat{\rho},1}$, display considerable improvements in power as α_2 decreases and as γ increases. As is also known to happen in the pure finite variance case, finite sample local power for a given test is higher when local GLS de-meaning is used relative to OLS de-meaning, although the converse holds for the $\mathcal{VRT}$ test; see Elliott *et al.* (1996) and Breitung and Taylor (2003), among others. Where OLS de-meaning is used the $\mathcal{MSB}$ test is more powerful than the Eicker-White corrected tests when $\gamma = 0.1$ for $c < 25$ with the reverse holding otherwise. For the larger values of γ considered the Eicker-White corrected tests tend to be more powerful than $\mathcal{MSB}$. These three tests clearly dominate the other tests on power. Where local GLS de-meaned data is used the Eicker-White corrected tests display superior power properties to the other tests for all the values of γ considered. The $\mathcal{MSB}$ and $t_{\widehat{\rho}}$ are, however, only marginally less powerful here than the Eicker-White corrected tests. In the case of local GLS de-meaning the power of the $\mathcal{VRT}$ test is very much lower than that of all the other tests. In the case of OLS de-meaning it is more competitive.

Consider next the results in Figures 5-8 which give the corresponding results for $T = 500$. The most striking feature of these results is how close they are to the results in Figures 1-4 for $T = 200$. Indeed there are almost no discernable differences between the results for the two sample sizes. This is very encouraging because it suggests that the local asymptotic distribution theory given in section 4 provides a very close approximation to the finite sample local power functions of the tests.

We conclude this section by considering some extra graphs designed to provide further insight into the changes in the finite sample power properties of the tests that are seen as α_2 decreases and γ increases. To that end, Figures 9 and 10 for $T = 200$ and $T = 500$, respectively, depict the changes that occur, for a given value of α_2, in the finite sample local power properties of the tests between the two extreme cases considered for the weight parameter, γ, namely between $\gamma = 0.1$ and $\gamma = 10$. For each test what is graphed is the difference between the power for $\gamma = 10$ and $\gamma = 0.1$ scaled by the power for $\gamma = 0.1$. Essentially then, these graphs show the rate of change in power of the tests between $\gamma = 0.1$ and $\gamma = 10$.

[Please insert Figures 9 and 10 about here]

A number of interesting observations can be drawn from these graphs. First, we again see little differences between the results for $T = 200$ and $T = 500$ confirming what was seen in Figures 1-8. Second, it is seen that as α_2 decreases then so the use of the Eicker-White correction can prove useful in increasing power (regardless of whether OLS or local GLS de-meaning is used) and that this improvement becomes larger the smaller is α_2. Moreover, this is not a small sample artifact, as is seen by comparing Figures 9 and 10. An explanation of this behaviour based on asymptotic

considerations was provided in Section 4.3 for the tests not involving deterministics, although the same logic seems likely to apply. Third, as c increases then so, other things equal, the power differences between the two values of γ becomes negligible for all of the tests, indicating that for large c the most relevant determinant of local power for any given test (for a given method of de-meaning) is the local-to-unity parameter, c. Fourth, among all the tests, the test whose local power shows least dependence on γ is the $\mathcal{VRT}$ test. Fifth, the largest power gains among the tests are seen with the $t_{\widehat{\rho},1}^{W,\alpha_u}$ test, closely followed by $t_{\widehat{\rho},1}^{W}$, and these gains are generally maximised for values of c of about 5, exactly the region of the alternative parameter space where maximising local power is most useful.

6 Conclusions

In this paper, we have provided representations for the large sample distributions of a number of the most commonly used unit root test statistics in the scenario where the data generation process is a near-integrated process driven by linear process shocks whose innovations are heavy-tailed, following a local-to-finite variance process. The form of the local-to-finite variance process we use is based on the recent framework of Amsler and Schmidt (2012) which models the innovations as a linear combination of a pure finite variance component and a pure infinite variance component, the latter scaled such that it does not dominate the former in the limit. This has enabled us to provide analytical expressions for both the limiting null distributions and asymptotic local power of these tests within this framework, thereby facilitating a comparison of these results with those which obtain under the standard pure finite variance assumption and also under a pure infinite variance assumption. We have also reported results from a finite sample Monte Carlo study which suggest that, although the limiting distributions of the unit root statistics differ under local-to-finite variance innovations relative to the pure finite variance case, the impact of such innovations on standard unit root tests which compare these statistics with the asymptotic critical values which obtain in the pure finite variance case can be relatively modest with some of the tests displaying almost no deviations from the nominal level. Our results suggest that a variant of the usual ADF tests based on the use of Eicker-White standard errors can deliver significant gains in local power relative to other standard unit root tests. From an empirical perspective our results are important in that they help validate the use of standard unit root tests on heavy-tailed data, as is frequently done in empirical finance research.

References

Ahn, S.K., S.B. Fotopoulos and L. He (2001) Unit root tests with infinite variance errors, Econometric Reviews 20(4), 461-483.

Amsler, C. and P. Schmidt (1999) Tests of short memory with thick tailed errors, Cowles Foundation Econometrics Conference "New Developments in Time Series Econometrics.

Amsler, C. and P. Schmidt (2012) Tests of short-memory with thick-tailed errors, Journal of Business and Economic Statistics 30(3), 381-390.

Avram, F. and M.S. Taqqu (1992) Weak convergence of sums of moving averages in the α-stable domain of attraction, The Annals of Probability 20(1), 483-503.

Basrak B., R.A. Davis and T. Mikosch (2002) A characterization of multivariate regular variation, The Annals of Applied Probability 12(3), 908-920.

Berk, K.N. (1974) Consistent autoregressive spectral estimates, The Annals of Statistics 2, 389-502.

Billingsley, P. (1968) Convergence of probability measures, New york: Wiley.

Breitung, J. (2002) Nonparametric tests for unit roots and cointegration, Journal of Econometrics 108, 343-363.

Breitung, J. and A.M.R. Taylor (2003) Corrigendum to. Nonparametric tests for unit roots and cointegration, Journal of Econometrics 117, 401-404 (J. Econom. 108 (2002) 343–363).

Charemza, W.W., D. Hristova and P. Burridge (2005) Is inflation stationary?, Applied Economics 37(8), 901-903.

Callegari, F., N. Cappuccio and D. Lubian (2003) Asymptotic inference in time series regressions with a unit root and infinite variance errors, Journal of Statistical Planning and Inference 116, 277-303.

Cappuccio, N. and D. Lubian (2007) Asymptotic null distributions of stationarity and nonstationarity tests under local-to-finite variance errors, Annals of the Institute of Statistical Mathematics 59(3), 403-423.

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. Econometric Theory, forthcoming. Downloadable from `https://ideas.repec.org/p/bot/quadip/wpaper130.html`.

Chan, N.H. (1990) Inference for near-integrated time series with infinite variance, Journal of the American Statistical Association 85(412), 1069-1074.

Chan, N.H., L. Peng and Y. Qi (2006) Quantile inference for near-integrated autoregressive time series with infinite variance, Statistica Sinica 16, 15-28.

Chan, N.H. and R.M. Zhang (2009) Inference for nearly nonstationary processes under strong dependence with infinite variance, Statistica Sinica 19, 925-947.

Chan, N.H. and L.T. Tran (1989) On the first-order autoregressive process with infinite variance, Econometric Theory 5(3), 354-362.

Chang, Y. and J.Y. Park (2002) On the asymptotics of ADF tests for unit roots, Econometric Reviews 21(4), 431-447.

Davidson, J. (1994) Stochastic Limit Theory, Oxford University Press.

Davis, R.A. (1983) Stable limits for partial sums of dependent random variables, The Annals of Probability 11(2), 262-269.

Davis, R.A. and S. Resnick (1985) Limit theory for moving averages of random variables with regularly varying tail probabilities, The Annals of Probability 13(1), 179-195.

Demetrescu M. (2010) On the Dickey-Fuller test with White standard errors, Statistical Papers 51, 11-25.

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

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

Fama, E.F. (1965) Portfolio analysis in a stable Paretian market, Management Science 11(3), 404-419.

Falk, B. and C.-H. Wang (2003) Testing long-run PPP with infinite-variance returns, Journal of Applied Econometrics 18(4), 471-484.

Faust, J. (1996) Near observational equivalence and theoretical size problems with unit root tests, Econometric Theory 12, 724-731.

Feller, W. (1971) An introduction to probability theory and its applications. Volume II. 2nd Ed. Wiley, New York.

Haldrup, N., and M. Jansson, 2006, Improving power and size in unit root testing. Palgrave Handbooks of Econometrics: Vol. 1 Econometric Theory, Chapter 7. T. C. Mills and K. Patterson (eds.). Palgrave MacMillan, Basingstoke.

Hamilton, J.D. (1994). Time series analysis, Princeton University Press.

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

Ibragimov U. and Y. Linnik (1971) Independent and stationary sequences of random variables, Wolters-Noordhoff Groningen.

Jacod, J. and A.N. Shiryaev (2003) Skorokhod topology and convergence of processes. "Limit theorems for stochastic processes." Springer Berlin Heidelberg, 324-388.

Jansson, M. (2002) Consistent covariance matrix estimation for linear processes, Econometric Theory 18, 1449-1459.

Jansson, M. (2008) Semiparametric power envelopes for tests of the unit root hypothesis, Econometrica 76, 1103–1142.

Knight, K. (1989) Limit theory for autoregressive-parameter estimates in an infinite-variance random walk, The Canadian Journal of Statistics 17(3), 261-278.

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

Koutrouvelis, I.A. (1980) Regression-type estimation of the parameters of stable laws, Journal of the American Statistical Association 75(372).

Koutrouvelis, I.A. (1981) An iterative procedure for the estimation of the parameters of stable law, Communications in Statistics - Simulations and Computations 10(1), 17-28.

Kurtz T. and P. Protter (1991) Weak limit theorems for stochastic integrals and stochastic differential equations, The Annals of Probability 19(3), 1035-1070.

LePage, R, M. Woodroofe and J. Zinn (1981). Convergence to a stable distribution via order statistics, The Annals of Probability 9(4), 624-632.

Mandelbrot, B. (1967) The variation of some other speculative prices, The Journal of Business 40(4), 393-413.

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

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

Müller, U.K. (2008) The impossibility of consistent discrimination between I(0) and I(1) processes, Econometric Theory 24, 616 630.

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

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

Paulauskas, V., and S. Rachev (1998) Cointegrated processes with infinite variance innovations, The Annals of Applied Probability 8(3), 775-792.

Perron, P. and S. Ng (1996) Useful modifications to some unit root tests with dependent errors and their local asymptotic properties, The Review of Economic Studies 63(3), 435-463.

Perron, P. and Z. Qu (2007) A simple modification to improve the finite sample properties of Ng and Perron's unit root tests, Economics Letters 94(1), 12-19.

Phillips, P.C.B. (1987) Time series regression with a unit root, Econometrica 55, 277-301.

Phillips, P.C.B. (1988) Towards a unified asymptotic theory for autoregression, Biometrica 74, 535-547.

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

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

Phillips, P.C.B. and V. Solo (1992) Asymptotics for linear precesses, The Annals of Statistics 20(2), 971-1001.

Rachev, S.T., S. Mittnik and J., -R., Kim (1998) Time series with unit roots and infinite-variance disturbances, Applied Mathematics Letters 11(5), 69-74.

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

Resnick, S.I. (1997) Heavy tail modelling and teletraffic data (with discussion), The Annals of Statistics 25, 1805-1869.

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

Rothenberg, T.J., and J.H. Stock (1997) Inference in a nearly integrated autoregressive model with nonnormal innovations, Journal of Econometrics 80(2), 269-286.

Samarakoon, D.M.M. and K. Knight (2009) A note on unit root tests with infinite variance noise, Econometric Reviews 28(4), 314-334.

Samorodnitsky G. and M. Taqqu (2000) Stable Non-Gaussian Random Processes, Chapman &Hall Boca Raton.

Skorokhod A. (1956) Limit theorems for stochastic processes, Theory of Probability and Its Applications 1, 261-290.

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

Tsihrintzis, G.A., M. Shao and C.L. Nikias (1996) Recent results in applications and processing of α-stable-distributed time series, Journal of the Franklin Institute 333(B) No. 4, 467-497.

Tyran-Kamińska M. (2010) Convergence to Lévy stable processes under some weak dependence conditions, Stochastic Processes and Their Applications 120, 1629-1650.

A Appendix

For later reference, define $\mathcal{P}_t := \sum_{j=0}^{t-1} \phi_T^{t-j} \varepsilon_{t-j}$ and $\mathcal{P}_{i,t} := \sum_{j=0}^{t-1} \phi_T^{t-j} \varepsilon_{i,t-j}$ $(i = 1, 2)$, so that $\mathcal{P}_t = \mathcal{P}_{1,t} + \gamma a_T^{-1} T^{1/2} \mathcal{P}_{2,t}$. In view of Lemma 4.1, summation by parts can be used to justify the standard joint convergence $(T^{-1/2}\mathcal{P}_{1,[Tr]}, a_T^{-1}\mathcal{P}_{2,[Tr]})' \Rightarrow (\sigma_1 J_c(r), \mathfrak{J}_{c,\alpha}(r))'$ in $D_2[0,1]$, which by the continuity of $J_c(r)$ implies that $T^{-1/2}\mathcal{P}_{[Tr]} \Rightarrow \sigma_1 H_{c,\nu}(r)$ on $D[0,1]$. As a direct result, for $\xi_t := \psi(1)\mathcal{P}_t$ it holds on $D[0,1]$ that

$$T^{-1/2}\xi_{[Tr]} \Rightarrow \sigma_1 \psi(1) H_{c,\nu}(r). \tag{A.1}$$

Recall further that,

$$u_t = \psi(L)\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

with $\psi(L) := \sum_{j=0}^{\infty} \psi_j L^j$ and $\varepsilon_t := \varepsilon_{1t} + \gamma a_T^{-1} T^{1/2} \varepsilon_{2t}$. Thus, considering a Beveridge-Nelson decomposition of u_t we obtain,

$$u_t = \psi(1)\varepsilon_t + \widetilde{\varepsilon}_{t-1} - \widetilde{\varepsilon}_t \tag{A.2}$$

where $\widetilde{\varepsilon}_t := \widetilde{\psi}(L)\varepsilon_t = \sum_{j=0}^{\infty} \widetilde{\psi}_j \varepsilon_{t-j}$, with $\widetilde{\psi}_j := \sum_{k=j+1}^{\infty} \psi_k$. Alternatively, considering u_t as given in (3), we can write (A.2) as,

$$u_t = \left[\psi(1)\varepsilon_{1t} + \widetilde{\varepsilon}_{1,t-1} - \widetilde{\varepsilon}_{1t}\right] + \gamma a_T^{-1} T^{1/2} \left[\psi(1)\varepsilon_{2t} + \widetilde{\varepsilon}_{2,t-1} - \widetilde{\varepsilon}_{2t}\right]. \tag{A.3}$$

Here the series for $\widetilde{\varepsilon}_{it}$, $i = 1, 2$, are well-defined *a.s.* given that $\sum_{j=0}^{\infty} |\widetilde{\psi}_j|^\delta < \infty$ for δ of Assumption $\mathcal{A}.5$ ($\sum_{j=0}^{\infty} |\widetilde{\psi}_j|^\delta < \sum_{k=0}^{\infty} k|\psi_k|^\delta < \infty$; cf. Phillips and Solo (1992, pp.976,984)), and $\widetilde{\varepsilon}_{2t}$ belongs to the normal domain of attraction of a stable law with characteristic exponent α.

Finally, for x_t of (1) we find that

$$\begin{aligned} x_t &= \sum_{j=0}^{t} \phi_T^j u_{t-j} + \phi_T^t x_0 = \psi(1)\mathcal{P}_t - \widetilde{\varepsilon}_t + (1 - \phi_T) \sum_{j=1}^{t-1} \phi_T^{j-1} \widetilde{\varepsilon}_{t-j} + \widetilde{\varepsilon}_0 + \phi_T^t x_0 \\ &= \xi_t - \widetilde{\varepsilon}_t + \zeta_t, \end{aligned} \tag{A.4}$$

where the equality defines ζ_t. From $\max_{t=1,\ldots,T} |\sum_{j=1}^{t-1} \phi_T^{j-1} \widetilde{\varepsilon}_{t-j}| \leq \sum_{t=1}^{T} |\widetilde{\varepsilon}_{1t}| + \gamma a_T^{-1} T^{1/2} \sum_{t=1}^{T} |\widetilde{\varepsilon}_{2t}|$, Markov's inequality and, for $\alpha = 1$, Karamata's theorem, it follows that $\max_{t=1,\ldots,T} |\zeta_t| = O_p(1)$.

Proof of Lemma 4.2

Without loss of generality under our assumption that $x_0 = O_p(1)$, we may set $x_0 = 0$ in what follows.

i) Let $\Psi^2 := \sum_{j=0}^{\infty} \psi_j^2$. Consider first the sample variance of u_t, that is,

$$T^{-1}\sum_{t=1}^{T} u_t^2 = T^{-1}\sum_{t=1}^{T} [\psi(L)\varepsilon_{1t}]^2 + \frac{\gamma^2}{a_T^2}\sum_{t=1}^{T} [\psi(L)\varepsilon_{2t}]^2 + \frac{2\gamma}{T^{1/2}a_T}\sum_{t=1}^{T} [\psi(L)\varepsilon_{1t}]\,[\psi(L)\varepsilon_{2t}]\,.$$

Here $T^{-1}\sum_{t=1}^{T} [\psi(L)\varepsilon_{1t}]^2 \xrightarrow{p} Var(\psi(L)\varepsilon_{11}) = \Psi^2\sigma_1^2$ by a law of large numbers [LLN], $a_T^{-2}\sum_{t=1}^{T} [\psi(L)\varepsilon_{2t}]^2 \Rightarrow \Psi^2[\mathcal{U}_\alpha]_1$ by Theorem 4.2 of Davis and Resnick (1985), and $\sum_{t=1}^{T} [\psi(L)\varepsilon_{1t}]\,[\psi(L)\varepsilon_{2t}] = o_p(T^{1/2}a_T)$ by Markov's inequality. In fact,

$$E\left|\sum_{t=1}^{T} [\psi(L)\varepsilon_{1t}]\,[\psi(L)\varepsilon_{2t}]\right|^{\eta} \leq TE|\varepsilon_{1t}|^{\eta}E|\varepsilon_{2t}|^{\eta}(\sum_{j=1}^{\infty}|\psi_j|^{\eta})^2 = O(T),$$

where $\eta = 1$ if $\alpha > 1$ and $\eta \in [\delta/2, 1)$ is arbitrary if $\alpha = 1$, so $\sum_{t=1}^{T} [\psi(L)\varepsilon_{1t}]\,[\psi(L)\varepsilon_{2t}] = O_p(T) = o_p(T^{1/2}a_T)$ if $\alpha > 1$ and $\sum_{t=1}^{T} [\psi(L)\varepsilon_{1t}]\,[\psi(L)\varepsilon_{2t}] = O(T^{1+\epsilon})$ for all $\epsilon > 0$ if $\alpha = 1$, with $O(T^{1+\epsilon}) = o_p(T^{1/2}a_T)$ for $\epsilon \in (0, \frac{1}{2})$ in the latter case. By collecting these facts, we establish that,

$$T^{-1}\sum_{t=1}^{T} u_t^2 \Rightarrow \Psi^2\sigma_1^2 + \gamma^2\Psi^2[\mathcal{U}_\alpha]_1 = \Psi^2\sigma_1^2\left(1 + \nu^2[\mathcal{U}_\alpha]_1\right).$$

ii) Using (A.4) and the uniform evaluation of ξ_t there, we find that

$$T^{-3/2}\sum_{t=1}^{T} x_t = T^{-3/2}\sum_{t=1}^{T} \xi_t - T^{-3/2}\sum_{t=1}^{T} \widetilde{\varepsilon}_t + o_p(1),$$

where further $\sum_{t=1}^{T} \widetilde{\varepsilon}_t = O_p(T)$ by the same argument as for the remainder in (A.4). Hence, by (A.1) and the Continuous mapping theorem [CMT], $T^{-3/2}\sum_{t=1}^{T} x_t \Rightarrow \psi(1)\sigma_1 \int_0^1 H_{c,\nu}dr$.

iii) Again by (A.4) with a uniformly $O_p(1)$ remainder ζ_t,

$$\begin{aligned}
\frac{1}{4}\left|T^{-1}\sum_{t=1}^{T}(x_{t-1}^2 - \xi_{t-1}^2)\right| &\leq T^{-1}\sum_{t=1}^{T}|\xi_{t-1}(\widetilde{\varepsilon}_{t-1} - \zeta_{t-1})| + \frac{T^{-1}}{4}\sum_{t=1}^{T}(\widetilde{\varepsilon}_{t-1} - \zeta_{t-1})^2 \\
&\leq \max_{t=1,\ldots,T}|\xi_{t-1}|\left(\frac{1}{T}\sum_{t=1}^{T}|\widetilde{\varepsilon}_{2,t-1}| + \frac{1}{T^{1/2}a_T}\sum_{t=1}^{T}|\widetilde{\varepsilon}_{2,t-1}| + O_p(1)\right) \\
&\quad + T^{-1}\sum_{t=1}^{T}\widetilde{\varepsilon}_{1,t-1}^2 + \frac{1}{a_T^2}\sum_{t=1}^{T}\widetilde{\varepsilon}_{2,t-1}^2 + o_p(T) = o_p(T)
\end{aligned}$$

because $\max_{t=1,\ldots,T}|T^{-1/2}\xi_{t-1}| \Rightarrow \sigma_1|\psi(1)|\sup_{[0,1]}|H_{c,\nu}| < \infty$ a.s. by (A.1) and the CMT, $\sum_{t=1}^{T}|\widetilde{\varepsilon}_{2,t-1}^i| = O_p(T)$, $i = 1, 2$, by an LLN, $\sum_{t=1}^{T}|\widetilde{\varepsilon}_{2,t-1}| = O_p(T)$ for $\alpha > 1$ by an LLN, $\sum_{t=1}^{T}|\widetilde{\varepsilon}_{2,t-1}| = O_p(Tl_T)$ with a slowly varying l_T for $\alpha = 1$ by Markov's inequality, and $\sum_{t=1}^{T}\widetilde{\varepsilon}_{2,t-1}^2 = O_p(a_T^2)$ by Theorem 4.2 of Davis and Resnick (1985). Therefore, $T^{-2}\sum_{t=1}^{T} x_{t-1}^2 = \sum_{t=1}^{T}\xi_{t-1}^2 + o_p(1) \Rightarrow \{\psi(1)\}^2\sigma_1^2\int_0^1 H_{c,\nu}^2 dr$ by (A.1) and the CMT.

iv) Regarding $T^{-1}\sum_{t=1}^{T} x_{t-1}u_t$, following Phillips (1988, 1990) we observe that:

$$\sum_{t=1}^{T} x_t^2 = \sum_{t=1}^{T}(\phi_T x_{t-1} + u_t)^2 = \sum_{t=1}^{T}(\phi_T^2 x_{t-1}^2 + 2\phi_T x_{t-1}u_t + u_t^2).$$

Since $\phi_T^2 = (1-c/T)^2 = 1 - 2c/T + c^2/T^2$, it follows that

$$\sum_{t=1}^{T} x_t^2 = \sum_{t=1}^{T} x_{t-1}^2 - \frac{2c}{T}\sum_{t=1}^{T} x_{t-1}^2 + \frac{c^2}{T^2}\sum_{t=1}^{T} x_{t-1}^2 + 2\sum_{t=1}^{T} x_{t-1}u_t - \frac{2c}{T}\sum_{t=1}^{T} x_{t-1}u_t + \sum_{t=1}^{T} u_t^2.$$

Hence,

$$x_T^2 = \sum_{t=1}^{T}\left(x_t^2 - x_{t-1}^2\right) = -\frac{2c}{T}\sum_{t=1}^{T} x_{t-1}^2 + 2\sum_{t=1}^{T} x_{t-1}u_t + \sum_{t=1}^{T} u_t^2 + o_p(T).$$

Thus, we establish that

$$T^{-1}\sum_{t=1}^{T} x_{t-1}u_t = \frac{1}{2}\left(T^{-1}x_T^2 + T^{-2}2c\sum_{t=1}^{T} x_{t-1}^2 - T^{-1}\sum_{t=1}^{T} u_t^2\right) + o_p(1) \tag{A.5}$$

From (A.4), $x_T^2 = \xi_T^2 - 2\xi_T(\widetilde{\varepsilon}_T - \zeta_T) + (\widetilde{\varepsilon}_T - \zeta_T)^2$, where $T^{-1/2}\xi_T \Rightarrow \sigma_1\psi(1)H_{c,\nu}(1)$ by (A.1) and the CMT, and $\widetilde{\varepsilon}_T = \widetilde{\varepsilon}_{1T} + T^{1/2}a_T^{-1}\widetilde{\varepsilon}_{2T} = \widetilde{\varepsilon}_{1T} + o_p(1) = O_p(1)$ because $\{\widetilde{\varepsilon}_{1t}\}$ and $\{\widetilde{\varepsilon}_{2t}\}$ are stationary with a.s. finite terms. Thus, $x_T^2 \Rightarrow \{\psi(1)\}^2\sigma_1^2 H_{c,\nu}^2(1)$. Considering also Lemma 4.2(i, iii), we establish that,

$$T^{-1}\sum_{t=1}^{T} x_{t-1}u_t \Rightarrow \frac{1}{2}\left\{\{\psi(1)\}^2\sigma_1^2 H_{c,\nu}^2(1) + 2c\sigma_1^2\{\psi(1)\}^2\int_0^1 H_{c,\nu}^2(r)dr - \Psi^2\sigma_1^2[H_{0,\nu}]_1\right\}.$$

Finally, we obtain the limit in Lemma 4.2(iv) by straightforward manipulations and using the identity

$$H_{c,\nu}^2(1) \equiv [H_{0,\nu}]_1 - 2c\int_0^1 H_{c,\nu}^2(r)dr + 2\int_0^1 H_{c,\nu}(r)dH_{0,\nu}(r). \tag{A.6}$$

v) The convergence of $T^{-1}\sum_{t=1}^{T} x_{t-1}\varepsilon_t$ can be deduced from part (iv) and the identities

$$\begin{aligned}
\sum_{t=1}^{T} x_{t-1}\varepsilon_t &= \{\psi(1)\}^{-1}\sum_{t=1}^{T} x_{t-1}u_t + \sum_{t=1}^{T} x_{t-1}\Delta\widetilde{\varepsilon}_t \\
&= \{\psi(1)\}^{-1}\sum_{t=1}^{T} x_{t-1}u_t - \sum_{t=1}^{T}\Delta x_t\widetilde{\varepsilon}_t + x_T\widetilde{\varepsilon}_T \\
&= \{\psi(1)\}^{-1}\sum_{t=1}^{T} x_{t-1}u_t - \sum_{t=1}^{T} u_t\widetilde{\varepsilon}_t + T^{-1}c\sum_{t=1}^{T} x_{t-1}\widetilde{\varepsilon}_t + x_T\widetilde{\varepsilon}_T.
\end{aligned}$$

Handling mixed products as in the proof of part (i), we find that

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} u_t\widetilde{\varepsilon}_t &= \frac{1}{T}\sum_{t=1}^{T} u_{1t}\widetilde{\varepsilon}_{1t} + \frac{\gamma^2}{a_T^2}\sum_{t=1}^{T} u_{2t}\widetilde{\varepsilon}_{2t} + o_p(1) \\
&\Rightarrow Cov(u_{11},\widetilde{\varepsilon}_{11}) + \gamma^2[\mathcal{U}_\alpha]_1\sum_{i=0}^{\infty}\psi_i\widetilde{\psi}_i = \sigma_1^2[H_{0,\nu}]_1\sum_{i=0}^{\infty}\psi_i\widetilde{\psi}_i
\end{aligned}$$

by an LLN and Theorem 4.2 of Davis and Resnick (1985). As $|\sum_{t=1}^{T} x_{t-1}\widetilde{\varepsilon}_t| \le \max_{t=1,\ldots,T}|x_t|\sum_{t=1}^{T}|\widetilde{\varepsilon}_t| = O_p(T^{3/2})$, see the derivation of (A.4), and $x_T\widetilde{\varepsilon}_T = O_p(T^{1/2})$, it remains to apply part (iv) to $\sum_{t=1}^{T} x_{t-1}u_t$ and to observe that $\sum_{i=0}^{\infty}\psi_i\widetilde{\psi}_i = \frac{1}{2}(\{\psi(1)\}^2 - \Psi^2)$.

vi) First, $T^{-2}\sum_{t=1}^{T} x_{t-1}^2\varepsilon_t^2 = T^{-2}\sum_{t=1}^{T}\xi_{t-1}^2\varepsilon_t^2 + o_p(1)$ since, using (A.4),

$$\begin{aligned}\sum_{t=1}^{T}\left|x_{t-1}^2-\xi_{t-1}^2\right|\varepsilon_t^2 &= \sum_{t=1}^{T}\left|2\xi_{t-1}(\widetilde{\varepsilon}_t-\zeta_t)+(\widetilde{\varepsilon}_t-\zeta_t)^2\right|\varepsilon_t^2 \\ &\leq 2\max_{t=1,\ldots,T}|\xi_t|\,(\sum_{t=1}^{T}|\widetilde{\varepsilon}_{t-1}|\varepsilon_t^2+\max_{t=1,\ldots,T}|\zeta_t|\sum_{t=1}^{T}\varepsilon_t^2) \\ &\quad +2\sum_{t=1}^{T}\widetilde{\varepsilon}_{t-1}^2\varepsilon_t^2+2\max_{t=1,\ldots,T}\zeta_t^2\sum_{t=1}^{T}\varepsilon_t^2\end{aligned}$$

with (i) $\max|T^{-1/2}\xi_t| = O_p(1)$, as a consequence of the fact that it converges weakly, (ii),

$$\sum_{t=1}^{T}|\widetilde{\varepsilon}_{t-1}|\varepsilon_t^2 \leq 2\sum_{i,j=1}^{2}(\gamma a_T^{-1}T^{1/2})^{i+2j-3}\sum_{t=1}^{T}|\widetilde{\varepsilon}_{i,t-1}|\varepsilon_{jt}^2 = o_p(T^{3/2})$$

by LLN for $j=1$ and by Markov's inequality for $j=2$:

$$E(\sum_{t=1}^{T}|\widetilde{\varepsilon}_{i,t-1}|\varepsilon_{2t}^2)^{\eta/2} \leq \sum_{t=1}^{T}E|\widetilde{\varepsilon}_{i,t-1}|^{\eta/2}E|\varepsilon_{2t}|^{\eta} = O\left(T\right),$$

so $\sum|\widetilde{\varepsilon}_{i,t-1}|\varepsilon_{2t}^2 = O\left(T^{2/\eta}\right)$ for all $\eta\in(0,\alpha)$, (iii), $T^{-1}\sum\varepsilon_t^2 = O_p(1)$, again because it converges weakly, and (iv),

$$\sum_{t=1}^{T}\widetilde{\varepsilon}_{t-1}^2\varepsilon_t^2 \leq 4\sum_{i,j=1}^{2}(\gamma^2 a_T^{-2}T)^{i+j-2}\sum_{t=1}^{T}\widetilde{\varepsilon}_{i,t-1}^2\varepsilon_{jt}^2 = o_p(T^2)$$

by LLN for $i=j=1$ and by Markov's inequality applied to the $\eta/2$ powers otherwise.

Second, we turn to $T^{-2}\sum\xi_{t-1}^2\varepsilon_t^2$. It holds that $(T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}\varepsilon_t, T^{-1}\sum_{t=1}^{\lfloor Tr\rfloor}\varepsilon_t^2) \Rightarrow (\sigma_1 H_{\nu,0}(r), \sigma_1^2[H_{\nu,0}]_r)$ in $D_2[0,1]$ because, (i), $(T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}\varepsilon_{1t}, a_T^{-1}\sum_{t=1}^{\lfloor Tr\rfloor}\varepsilon_{2t}, a_T^{-2}\sum_{t=1}^{\lfloor Tr\rfloor}\varepsilon_{2t}^2) \Rightarrow (\sigma_1 W(r), \mathcal{U}_\alpha(r), [\mathcal{U}_\alpha]_r)$ in $D_3[0,1]$ by Theorem 4 of Resnick and Greenwood (1979) and the independence of $\{\varepsilon_{1t}\}$ and $\{\varepsilon_{2t}\}$, and (ii),

$$\frac{1}{T}\sum_{t=1}^{\lfloor Tr\rfloor}\varepsilon_t^2 = \frac{1}{T}\sum_{t=1}^{\lfloor Tr\rfloor}\varepsilon_{1t}^2+\frac{\gamma^2}{a_T^2}\sum_{t=1}^{\lfloor Tr\rfloor}\varepsilon_{2t}^2+\frac{2\gamma}{T^{1/2}a_T}\sum_{t=1}^{\lfloor Tr\rfloor}\varepsilon_{1t}\varepsilon_{2t} \Rightarrow \sigma_1^2 r+\gamma^2[\mathcal{U}_\alpha]_r = \sigma_1^2[H_{\nu,0}]_r \qquad (A.7)$$

because $\{\varepsilon_{1t}\varepsilon_{2t}\}$ is IID with tail index α, so $\max_{r\in[0,1]}|T^{-1/2}a_T^{-1}\sum_{t=1}^{\lfloor Tr\rfloor}\varepsilon_{1t}\varepsilon_{2t}| \xrightarrow{p} 0$. By Theorem 2.7 of Kurtz and Protter (1991), it follows that $T^{-2}\sum_{t=1}^{T}\xi_{t-1}^2\varepsilon_t^2 = [\psi(1)]^2T^{-2}\sum_{t=1}^{T}(\sum_{s=1}^{t-1}\varepsilon_s)^2\varepsilon_t^2 \Rightarrow [\psi(1)]^2\sigma_1^4\int H_{\nu,0}^2 d[H_{\nu,0}]$, where condition C2.7 of the theorem can be checked as on pp.784-786 of Paulauskas and Rachev (1998). Recalling the previous paragraph, we conclude that $T^{-2}\sum_{t=1}^{T}x_{t-1}^2\varepsilon_t^2$ converges weakly to the same limit as that of $T^{-2}\sum_{t=1}^{T}\xi_{t-1}^2\varepsilon_t^2$. ■

Proof of Theorem 4.1. To discuss the ADF statistics, we need a precise expression for the error term in the AR sieve. Thus, if $(\sum_{i=0}^{\infty}\psi_i z^i)^{-1} = 1+\sum_{i=1}^{\infty}\beta_i z^i$, then $\Delta x_t = -(c/T)x_{t-1} + \sum_{i=1}^{p_T}\beta_i\Delta x_{t-i}+\varepsilon_{t,p_T}$ with $\varepsilon_{t,p_T} = \varepsilon_t+(c/T)\sum_{i=1}^{p_T}\beta_i x_{t-i-1}+\sum_{i=p_T+1}^{\infty}\beta_i u_{t-i}$. Without loss of generality under our stated rate condition on p_T, we proceed with an effective sample size of T (instead of $T-p_T$), assuming the availability of pre-sample values related to x_0 via equation (1).

The validity of the AR sieve will require that $\widehat{\rho}$ computed from (5) should satisfy

$$T\widehat{\rho} = -c+\left(T^{-2}\sum_{t=1}^{T}x_{t-1}^2\right)^{-1}T^{-1}\sum_{t=1}^{T}x_{t-1}\varepsilon_t+o_p(1). \qquad (A.8)$$

By standard OLS algebra,

$$\widehat{\rho} = -\frac{c}{T} + \left(\sum_{t=1}^{T} x_{t-1}^2 - S_{10} S_{00}^{-1} S_{01} \right)^{-1} \left(\sum_{t=1}^{T} x_{t-1} \varepsilon_{t,p_T} - S_{10} S_{00}^{-1} S_{0\varepsilon} \right), \tag{A.9}$$

where $S_{00} := \sum_{t=1}^{T} \Delta X_{t-1} \Delta X_{t-1}'$, $S_{0\varepsilon} := \sum_{t=1}^{T} \Delta X_{t-1} \varepsilon_{t,p_T}$ and $S_{01} := \sum_{t=1}^{T} \Delta X_{t-1} x_{t-1} = S_{10}'$. We evaluate S_{00} first. Upon splitting the observations and the product moments into the contributions of the finite and the infinite variance components, with notation corresponding to decomposition in (4), we argue in steps that $\| S_{00} - S_{00}^{(1)} - S_{00}^{(2)} \|_* = o_p(T)$, where $\| \cdot \|_*$ denotes the spectral matrix norm, $S_{00}^{(i)} := \sum_{t=1}^{T} \Delta X_{i,t-1} \Delta X_{i,t-1}'$ $(i = 1, 2)$ and the norming sequence $\gamma a_T^{-1} T^{1/2}$ is incorporated into ΔX_{2t}. Thus, defining $C_T := \{c_{ij}\}_{i,j=1}^{p_T}$ with $c_{ij} := \sum_{t=0}^{T-1} \Delta x_{1t} \Delta x_{2,t-|i-j|}$, we find that

$$\begin{aligned} \frac{1}{2} \| S_{00} - S_{00}^{(1)} - S_{00}^{(2)} \|_* &\leq \| \sum_{t=1}^{T} \Delta X_{1,t-1} \Delta X_{2,t-1}' \|_* \leq \| C_T \|_* + p_T^2 \max_{i=1,\ldots,p_T} \{ (\Delta x_{-i})^2 + (\Delta x_{T-i})^2 \} \\ &= \| C_T \|_* + p_T^2 \{ o_p(p_T) + a_T^{-2} T O_p(a_{p_T}^2) \} = \| C_T \|_* + o_p(T) \end{aligned} \tag{A.10}$$

under $p_T^3 / T \to 0$. Further, given the Toeplitz structure of C_T,

$$\frac{1}{2} \| C_T \|_* \leq \sum_{i=1}^{p_T} |c_{1i}| \leq \sum_{i=1}^{p_T} | \sum_{t=0}^{T-1} u_{1t} u_{2,t-i+1} | + O_p(p_T T^{1/2+\epsilon}), \tag{A.11}$$

where $O_p(p_T T^{1/2+\epsilon})$, with $\epsilon > 0$ arbitrary, stands for

$$\sum_{i=1}^{p_T} \{ (\frac{c}{T})^2 \sum_{t=0}^{T-1} x_{1,t-1} x_{2,t-i} + \frac{c}{T} \max_{t=1,\ldots,T} (|x_{1,t-1}| + |x_{2,t-i}|) \sum_{t=0}^{T-1} (|u_{1t}| + |u_{2,t-i+1}|) \},$$

given that $\max_{t=-p_T,\ldots,T} |x_{1t}| = O_p(T^{1/2})$ and $\sum_{i=1}^{p_T} \sum_{t=0}^{T-1} (|u_{1,t}| + |u_{2,t-i+1}|) \leq p_T \sum_{t=0}^{T-1} |u_{1,t}| + p_T \sum_{t=-p_T}^{T-1} |u_{2,t}| = O_p(p_T l_T T)$ with a slowly varying l_T (constant except for $\alpha = 1$). Regarding $\sum_{t=0}^{T-1} u_{1,t} u_{2,t-i+1} = a_T^{-1} T^{1/2} \gamma (\chi_i^{\leq} + \chi_i^{>})$, with

$$\chi_i^R := \sum_{t=0}^{T-1} \sum_{u,v=0}^{\infty} \psi_u \psi_v \varepsilon_{1,t-u} \varepsilon_{2,t-v-i+1} \mathbb{I}_{|\varepsilon_{2,t-v-i+1}| R a_T}, \ R \in \{\leq, >\},$$

it holds that (i), $E \sum_{i=1}^{p_T} |\chi_i^{\leq}| \leq \sum_{i=1}^{p_T} \{ E(\chi_i^{\leq})^2 \}^{1/2}$ by Jensen's inequality, where, using Karamata's theorem, we find that

$$E(\chi_i^{\leq})^2 \leq T E \varepsilon_{11}^2 E(\varepsilon_{21}^2 \mathbb{I}_{|\varepsilon_{21}| \leq a_T}) (\sum_{u=0}^{\infty} |\psi_u|)^4 = O(a_T^2),$$

where $\mathbb{I}$ denotes the usual indicator function, because $\{\varepsilon_{1t}\}$ and $\{\varepsilon_{2t}\}$ are independent, $E\varepsilon_{1,t-u} = 0$, $E(\varepsilon_{2,t-v-i+1} \mathbb{I}_{|\varepsilon_{2,t-v-i+1}| \leq a_T}) = 0$ by symmetry, and $E(\varepsilon_{21}^2 \mathbb{I}_{|\varepsilon_{21}| \leq a_T}) = E(\varepsilon_{2,t-v-i+1}^2 \mathbb{I}_{|\varepsilon_{2,t-v-i+1}| \leq a_T}) = O(T^{-1} a_T^2)$, and (ii), $E(\sum_{i=1}^{p_T} |\chi_i^{>}|)^\eta \leq \sum_{i=1}^{p_T} E|\chi_i^{>}|^\eta$, where $\eta = 1$ for $\alpha > 1$, $\eta \in [\delta, 1)$ is arbitrary for $\alpha = 1$, and

$$E|\chi_i^{>}|^\eta \leq \sum_{t=0}^{T-1} \sum_{u,v=0}^{\infty} |\psi_u|^\eta |\psi_v|^\eta E|\varepsilon_{11}|^\eta E(|\varepsilon_{21}|^\eta \mathbb{I}_{|\varepsilon_{21}| > a_T}) = O(a_T^\eta) (\sum_{u=0}^{\infty} |\psi_u|^\eta)^2 = O(a_T^\eta)$$

using Karamata's theorem again, so eventually, by Markov's inequality, $\| S_{00} - S_{00}^{(1)} - S_{00}^{(2)} \|_* \leq O_p(p_T T^{1/2+\epsilon}) + o_p(T) = o_p(T)$, because $p_T^3 / T \to 0$ as $T \to \infty$, where $\epsilon > 0$ is arbitrary. Let $\Sigma_p :=$

$\{r_{|i-j|}\}_{i,j=1}^{p_T}$ with $r_i := \sum_{j=0}^{\infty} \psi_i \psi_{j+i}$; then the eigenvalues of Σ_p are bounded and bounded away from zero under Assumptions $\mathcal{A}.1$ and $\mathcal{A}.5$. As additionally, under $p_T^3/T \to 0$, $\|S_{00}^{(1)} - T\Sigma_p \sigma_1^2\|_* = o_p(T)$ by Lemma 3 of Berk (1974) and $\|S_{00}^{(2)} - a_T^{-2} T\gamma^2 \Sigma_p \sum_{t=1}^{T} \varepsilon_{2t}^2\|_* = o_p(T)$ by Lemma 2 of Cavaliere *et al.* (2016a), by combining the previous results it follows that

$$\|T^{-1} S_{00} - \Sigma_p(\sigma_1^2 + \gamma^2 a_T^{-2} \sum_{t=1}^{T} \varepsilon_{2t}^2)\|_* = o_p(1), \tag{A.12}$$

and using inequality (2.15) of Berk (1974), it also holds that

$$\|(T^{-1} S_{00})^{-1} - \Sigma_p^{-1}(\sigma_1^2 + \gamma^2 a_T^{-2} \sum_{t=1}^{T} \varepsilon_{2t}^2)^{-1}\|_* = o_p(1). \tag{A.13}$$

Second, $S_{00}^{-1} S_{0\varepsilon}$ can be decomposed as

$$(S_{00}^{-1} S_{00}^{(1)})(S_{00}^{(1)})^{-1} S_{0\varepsilon}^{(1)} + (S_{00}^{-1} S_{00}^{(2)})(S_{00}^{(2)})^{-1} S_{0\varepsilon}^{(2)} + S_{00}^{-1} \sum_{t=1}^{T} (\Delta X_{1,t-1} \varepsilon_{2t,p_T} + \Delta X_{2,t-1} \varepsilon_{1t,p_T}),$$

where $\|(S_{00}^{(1)})^{-1} S_{0\varepsilon}^{(1)}\| = o_p(p_T^{-2})$ as in Lemma 3.2 of Chang and Park (2002), $(S_{00}^{(2)})^{-1} S_{0\varepsilon}^{(2)} = O_p(a_{p_T} a_T^{\epsilon-1} + \sum_{i=p_T+1}^{\infty} |\beta_i|)$ for all $\epsilon > 0$ as in Equation (7.1) of Cavaliere *et al.* (2016a), both under the condition that $p_T^2/T + 1/p_T \to 0$ as $T \to \infty$, and $\|\sum_{t=1}^{T} (\Delta X_{1,t-1} \varepsilon_{2t,p_T} + \Delta X_{2,t-1} \varepsilon_{1t,p_T})\| = O_p(p_T T^{1/2})$ by Markov's inequality and Karamata's theorem as, e.g., for the first kind of summands,

$$\begin{aligned}
\|\sum_{t=1}^{T} \Delta X_{1,t-1} \varepsilon_{2t,p_T}\|^2 &= \sum_{i=1}^{p_T} (\sum_{t=1}^{T} \Delta x_{1,t-i} \varepsilon_{2t,p_T})^2 \\
&\leq 2 \sum_{i=1}^{p_T} \{\sum_{t=1}^{T} \Delta x_{1,t-i} (a_T^{-1} T^{1/2} \varepsilon_{2t} + \sum_{i=p_T+1}^{\infty} \beta_i u_{2,t-i})\}^2 \\
&\quad + 2c^2 p_T T^{-2} \max_{t=-p_T,\ldots,T} x_{2t}^2 (\sum_{t=1-p_T}^{T-1} |\Delta x_{1t}|)^2 (\sum_{i=1}^{\infty} |\beta_i|)^2
\end{aligned}$$

with $\max_{t=-p_T,\ldots,T} |x_{2t}| = O_p(T^{1/2})$, $\sum_{t=1-p_T}^{T-1} |\Delta x_{1t}| = O_p(T)$,

$$\begin{aligned}
E \sum_{i=1}^{p_T} \{\sum_{t=1}^{T} \Delta x_{1,t-i} \varepsilon_{2t} \mathbb{I}_{|\varepsilon_{2t}| \leq a_T}\}^2 &= p_T T E(\varepsilon_{21}^2 \mathbb{I}_{|\varepsilon_{21}| \leq a_T})[E(\Delta x_{11})^2] = O(p_T a_T^2), \\
E[\sum_{i=1}^{p_T} \{\sum_{t=1}^{T} \Delta x_{1,t-i} \varepsilon_{2t} \mathbb{I}_{|\varepsilon_{2t}| > a_T}\}^2]^{\eta/2} &\leq p_T T E(|\varepsilon_{21}|^\eta \mathbb{I}_{|\varepsilon_{21}| > a_T}) E|\Delta x_{11}|^\eta = O(p_T a_T^\eta)
\end{aligned}$$

for $\eta = 1$ if $\alpha > 1$ and $\eta \in [\delta, 1)$ arbitrary if $\alpha = 1$, and similarly for the terms involving u_{2t}:

$$\begin{aligned}
&E \sum_{m=1}^{p_T} \{\sum_{t=1}^{T} \Delta x_{1,t-m} \sum_{i=p_T+1}^{\infty} \beta_i \sum_{j=0}^{\infty} \psi_j \varepsilon_{2,t-i-j} \mathbb{I}_{|\varepsilon_{2,t-i-j}| \leq a_T}\}^2 \\
\leq\ & p_T T E(\varepsilon_{21}^2 \mathbb{I}_{|\varepsilon_{21}| \leq a_T})[E(\Delta x_{11})^2] (\sum_{i=p_T+1}^{\infty} |\beta_i| \sum_{j=0}^{\infty} |\psi_j|)^2 = o(a_T^2),
\end{aligned}$$

$$\begin{aligned}
&E[\sum_{m=1}^{p_T} \{\sum_{t=1}^{T} \Delta x_{1,t-m} \sum_{i=p_T+1}^{\infty} \beta_i \sum_{j=0}^{\infty} \psi_j \varepsilon_{2,t-i-j} \mathbb{I}_{|\varepsilon_{2,t-i-j}| > a_T}\}^2]^{\eta/2} \\
\leq\ & p_T T E(|\varepsilon_{21}|^\eta \mathbb{I}_{|\varepsilon_{21}| > a_T}) E|\Delta x_{11}|^\eta \sum_{i=p_T+1}^{\infty} |\beta_i|^\eta \sum_{j=0}^{\infty} |\psi_j|^\eta = o(p_T a_T^\eta)
\end{aligned}$$

since $p_T \sum_{i=p_T+1}^{\infty} |\beta_i| \to 0$, $\sum_{i=p_T+1}^{\infty} |\beta_i|^\eta \to 0$ as $p_T \to \infty$. Accounting also for (A.13), which implies that $\|TS_{00}^{-1}\|_* = O_p(1)$, it follows that for all $\epsilon > 0$,

$$\|S_{00}^{-1} S_{0\varepsilon}\| = O_p(p_T^{-2} + T^{\epsilon - 1/2} p_T + \sum_{i=p_T+1}^{\infty} |\beta_i|). \tag{A.14}$$

Third, it holds that

$$\|S_{10} - \mathbf{1}'_{p_T} \sum_{t=1}^{T} x_{t-1}\Delta x_t - Tr\{(S_{00}^{(1)} + S_{00}^{(2)} + 2C_T)\Upsilon_{p_T}\}\| = o_p(T),$$

where $\mathbf{1}_{p_T}$ is a p_T-vector of ones, Υ_{p_T} is an upper triangular matrix with ones on and above the main diagonal, and the difference is due to presample and end-of-sample contributions as in (A.10). Thus, further,

$$\begin{aligned} \|S_{10}\| &\leq p_T^{1/2} |\sum_{t=1}^{T} x_{t-1}\Delta x_t| + \|Tr(S_{00}^{(1)}\Upsilon_{p_T})\| + \|Tr(S_{00}^{(2)}\Upsilon_{p_T})\| + 2\|Tr(C_T\Upsilon_{p_T})\| + o_p(T) \\ &\leq O_p(p_T^{1/2} T) + p_T^{1/2} \sum_{i=1}^{p_T} |c_{1i}| = O_p(p_T^{1/2} T) \end{aligned} \tag{A.15}$$

since $\sum_{t=1}^{T} x_{t-1}\Delta x_t = \sum_{t=1}^{T} x_{t-1}u_t - (c/T)\sum_{t=1}^{T} x_{t-1}^2 = O_p(T)$ by Lemma 4.2(iii,iv), $\|Tr(S_{00}^{(i)}\Upsilon_{p_T})\| = O_p(p_T^{1/2} T)$ $(i = 1, 2)$ is shown in the proof of Lemma 3.2(b) of Chang and Park (2002) and Lemma A.1(d) of Cavaliere *et al.* (2016b), and $\sum_{i=1}^{p_T} |c_{1i}| = O_p(p_T T^{1/2+\epsilon})$ for all $\epsilon > 0$ by the argument following (A.11) and $p_T^3/T \to 0$.

From (A.9), (A.13), (A.14) and (A.15), using the boundedness away from zero in probability of $T^{-2}\sum_{t=1}^{T} x_{t-1}^2$ and the finiteness of $\sum_{i=1}^{\infty} i|\beta_i|$, it follows that

$$T\widehat{\rho} = -c + \left(\sum_{t=1}^{T} x_{t-1}^2\right)^{-1} \sum_{t=1}^{T} x_{t-1}\varepsilon_{t,p_T} + o_p(1)$$

under the rate condition that $p_T^3/T + 1/p_T \to 0$ as $T \to \infty$. To obtain (A.8), it remains to observe that

$$\begin{aligned} \sum_{t=1}^{T} x_{t-1}\varepsilon_{t,p_T} &= \sum_{t=1}^{T} x_{t-1}\varepsilon_t + \frac{c}{T}\sum_{i=1}^{p_T}\beta_i \sum_{t=1}^{T} x_{t-1}^2 \\ &\quad + \frac{c}{T}\sum_{i=1}^{p_T}\beta_i \sum_{t=1}^{T} x_{t-1}(x_{t-i-1} - x_{t-1}) + \sum_{t=1}^{T} x_{t-1}(\sum_{i=p_T+1}^{\infty}\beta_i u_{t-i}), \end{aligned}$$

where $\sum_{t=1}^{T} x_{t-1}(x_{t-i-1} - x_{t-1}) = o_p(T^2)$ uniformly in $i = 1, ..., p_T$ by partial summation and the evaluation of S_{00}, and $\sum_{t=1}^{T} x_{t-1}(\sum_{i=p_T+1}^{\infty}\beta_i u_{t-i}) = o_p(T)$ by relating it to $(\sum_{t=1}^{T} x_{t-1}u_t)(\sum_{i=p_T+1}^{\infty}\beta_i) = o_p(T)$. Then (A.8) and Lemma 4.2(iii,v) imply that

$$T\widehat{\rho} \Rightarrow \frac{-c + \int_0^1 H_{c,\nu}(r)dH_{0,\nu}(r)}{\psi(1)\int_0^1 H_{c,\nu}^2(r)dr} \tag{A.16}$$

under the rate condition that $p_T^3/T + 1/p_T \to 0$, as $T \to \infty$, where $\psi(1) = (1 - \sum_{i=1}^{\infty}\beta_i)^{-1}$.

The limits of the ADF statistics then follow by examining various normalisations of $\widehat{\rho}$. First,

$$\begin{aligned} \sum_{i=1}^{p_T} |\widehat{\beta}_i - \beta_i| &= \|S_{00}^{-1}S_{0\varepsilon} - (\widehat{\rho} + c/T)S_{00}^{-1}S_{01}\|_1 \\ &\leq p_T^{1/2}\|S_{00}^{-1}S_{0\varepsilon}\| + O_p(T^{-1}p_T^{1/2})\|S_{00}^{-1}\|_*\|S_{0\varepsilon}\| = o_p(1) \end{aligned} \tag{A.17}$$

by (A.13), (A.14), (A.15) and the condition $p_T^3/T + 1/p_T \to 0$. Since $\sum_{i=p_T+1}^{\infty} |\beta_i| \to 0$, it follows that $\sum_{i=1}^{p_T} \widehat{\beta}_i \xrightarrow{p} \sum_{i=1}^{\infty} \beta_i$, which jointly with (A.16) yields the limit of $Z_{\widehat{\rho}}$. Second, $s_{p_T}^2 = T^{-1} \sum_{t=1}^{T} \widehat{\varepsilon}_{p_T,t}^2$, where $\widehat{\varepsilon}_{p_T,t}$ are the OLS residuals from (5), satisfies

$$\begin{aligned} |T s_{p_T}^2 - \sum_{t=1}^{T} \varepsilon_t^2| &\leq \sum_{t=1}^{T} (\widehat{\varepsilon}_{p_T,t} - \varepsilon_t)^2 + 2|\sum_{t=1}^{T} (\widehat{\varepsilon}_{p_T,t} - \varepsilon_t)\varepsilon_t| \\ &\leq \sum_{t=1}^{T} (\widehat{\varepsilon}_{p_T,t} - \varepsilon_t)^2 + 2\{\sum_{t=1}^{T} (\widehat{\varepsilon}_{p_T,t} - \varepsilon_t)^2\}^{1/2} \{\sum_{t=1}^{T} \varepsilon_t^2\}^{1/2} = o_p(T) \end{aligned}$$

since $\sum_{t=1}^{T} \varepsilon_t^2 = O_p(T)$ and, for $\boldsymbol{\beta}_{p_T} := (\beta_1, ..., \beta_{p_T})'$,

$$\begin{aligned} \sum_{t=1}^{T} (\widehat{\varepsilon}_{p_T,t} - \varepsilon_t)^2 &\leq 3(\widehat{\rho} + c/T)^2 \sum_{t=1}^{T} x_{t-1}^2 + 3\|\widehat{\boldsymbol{\beta}}_{p_T} - \boldsymbol{\beta}_{p_T}\|^2 \|S_{00}\|_* \\ &\quad + 3\sum_{t=1}^{T} (\varepsilon_{p_T,t} - \varepsilon_t)^2 = o_p(T) \end{aligned}$$

with $\widehat{\rho} + c/T = O_p(T^{-1})$ by (A.16), $\sum_{t=1}^{T} x_{t-1}^2 = O_p(T^2)$ by Lemma 4.2(iii), $\|\widehat{\boldsymbol{\beta}}_{p_T} - \boldsymbol{\beta}_{p_T}\| \leq \sum_{i=1}^{p_T} |\widehat{\beta}_i - \beta_i| = o_p(1)$ by (A.17), $\|S_{00}\|_* = O_p(T)$ by (A.12) with $\|\Sigma_p\|_* = O(1)$, and

$$\begin{aligned} \sum_{t=1}^{T} (\varepsilon_{p_T,t} - \varepsilon_t)^2 &\leq 2c^2 T^{-2} \sum_{t=1}^{T} (\sum_{i=1}^{p_T} \beta_i x_{t-i-1})^2 + 2\sum_{t=1}^{T} (\sum_{i=p_T+1}^{\infty} \beta_i u_{t-i})^2 \\ &\leq 2c^2 T^{-1} \max_{t=-p_T,...,T} x_t^2 (\sum_{i=1}^{\infty} |\beta_i|)^2 + 2\sum_{t=1}^{T} (\sum_{i=p_T+1}^{\infty} \beta_i u_{1,t-i})^2 \\ &\quad + 2\gamma^2 T a_T^{-2} \sum_{t=1}^{T} (\sum_{i=p_T+1}^{\infty} \beta_i u_{2,t-i})^2 \\ &\quad + 4\gamma T^{1/2} a_T^{-1} \sum_{t=1}^{T} (\sum_{i=p_T+1}^{\infty} \beta_i u_{1,t-i})(\sum_{i=p_T+1}^{\infty} \beta_i u_{2,t-i}) = o_p(T) \end{aligned}$$

as $\max_{t=-p_T,...,T} x_t^2 = O_p(T)$, $\sum_{t=1}^{T} (\sum_{i=p_T+1}^{\infty} \beta_i u_{1,t-i})^2 = o_p(T)$ for $p_T \to \infty$ by Markov's inequality, $\sum_{t=1}^{T} (\sum_{i=p_T+1}^{\infty} \beta_i u_{2,t-i})^2 = o_p(a_T^2)$ for $p_T \to \infty$ by the proof of Lemma 3 of Cavaliere *et al.* (2016a), and $\sum_{t=1}^{T} (\sum_{i=p_T+1}^{\infty} \beta_i u_{1,t-i})(\sum_{i=p_T+1}^{\infty} \beta_i u_{2,t-i}) = o_p(a_T)$ similarly to the terms involving $\sum_{i=p_T+1}^{\infty} \beta_i u_{2,t-i}$ in the derivation of (A.14). Therefore, $s_{p_T}^2 = T^{-1} \sum_{t=1}^{T} \varepsilon_t^2 + o_p(1) \Rightarrow \sigma_1^2 [H_{0,\nu}]_1$. As $t_{\widehat{\rho}} = (\sum_{t=1}^{T} x_{t-1}^2 - S_{10} S_{00}^{-1} S_{01})^{1/2} \widehat{\rho} s_{p_T}^{-1}$, its limit now follows from the previous discussion of $\widehat{\rho}$.

Third, along similar lines,

$$\begin{aligned} |\sum_{t=1}^{T} x_{t-1}^2 (\widehat{\varepsilon}_{p_T,t} - \varepsilon_t^2)| &\leq \sum_{t=1}^{T} x_{t-1}^2 (\widehat{\varepsilon}_{p_T,t} - \varepsilon_t)^2 + 2|\sum_{t=1}^{T} x_{t-1}^2 (\widehat{\varepsilon}_{p_T,t} - \varepsilon_t)\varepsilon_t| \\ &\leq \sum_{t=1}^{T} x_{t-1}^2 (\widehat{\varepsilon}_{p_T,t} - \varepsilon_t)^2 + 2\{\sum_{t=1}^{T} x_{t-1}^2 (\widehat{\varepsilon}_{p_T,t} - \varepsilon_t)^2\}^{1/2} \{\sum_{t=1}^{T} x_{t-1}^2 \varepsilon_t^2\}^{1/2} \end{aligned}$$

is $o_p(T^2)$ because $\sum_{t=1}^{T} x_{t-1}^2 \varepsilon_t^2 = O_p(T^2)$ by Lemma 4.2(vi) and

$$\sum_{t=1}^{T} x_{t-1}^2 (\widehat{\varepsilon}_{p_T,t} - \varepsilon_t)^2 \leq \max_{t=1,...,T} x_t^2 \sum_{t=1}^{T} (\widehat{\varepsilon}_{p_T,t} - \varepsilon_t)^2 = o_p(T^2)$$

as $\max_{t=-p_T,...,T} x_t^2 = O_p(T)$ and $\sum_{t=1}^T (\widehat{\varepsilon}_{p_T,t} - \varepsilon_t)^2 = o_p(T)$. Hence, $T^{-2}\sum_{t=1}^T x_{t-1}^2 \widehat{\varepsilon}_{p_T,t} = \sum_{t=1}^T x_{t-1}^2 \varepsilon_t^2 + o_p(1)$ and the limit of $t_{\hat{\rho},1}^W$ follows from (A.16) and Lemma 4.2(iii,vi).

The remaining limits in Theorem 4.1 follow from Proposition 4.1, Lemma 4.2, the convergence of $\widehat{\sigma}_u^2$:

$$\begin{aligned}\widehat{\sigma}_u^2 &= T^{-1}\sum_{t=1}^T \hat{u}_t^2 = T^{-1}\sum_{t=1}^T u_t^2 + T^{-1}(\widehat{\phi}-1)^2 \sum_{t=1}^T x_{t-1}^2 + 2T^{-1}(\widehat{\phi}-1)\sum_{t=1}^T x_{t-1}u_t \\ &= T^{-1}\sum_{t=1}^T u_t^2 + o_p(1) \Rightarrow \Psi^2 \sigma_1^2 [H_{0,\nu}]_1\end{aligned}$$

which follows from results established above, and the fact that, under the conditions stated in section 3.2, the weak convergence result $\widehat{\omega}_u^2 \Rightarrow \{\psi(1)\}^2\sigma_1^2[H_{0,\nu}]_1$ holds for both $\widehat{\omega}_u^2 = s_{WA}^2$ and $\widehat{\omega}_u^2 = s_{AR}^2$, which again follows from results established above. Indeed, the stated results will hold for any estimator $\widehat{\omega}_u^2$ whose weak limit is $\{\psi(1)\}^2\sigma_1^2[H_{0,\nu}]_1$. We also remark that the weak limits of $\mathcal{MZ}_\phi$ and $\mathcal{MZ}_t$ obtain readily in the form

$$\mathcal{MZ}_\phi \Rightarrow \frac{H_{c,\nu}^2(1) - [H_{0,\nu}]_1}{2\int_0^1 H_{c,\nu}^2(r)dr}, \qquad \mathcal{MZ}_t \Rightarrow \frac{H_{c,\nu}^2(1) - [H_{0,\nu}]_1}{2\left\{[H_{0,\nu}]_1\right\}^{1/2}\left\{\int_0^1 H_{c,\nu}^2(r)dr\right\}^{1/2}},$$

and are seen to be the same as for the $t_{\widehat{\rho}}$ and $\mathcal{Z}_t$ statistics given in parts (i) and (ii) of Theorem 4.1 by using the Îto-type equation (A.6). ■

To introduce the idea behing Proposition 4.2, for lag lengths k between 1 and p_T define

$$\hat{\Omega}_k := \hat{\Omega}_0 - \sum_{t=p_T+1}^T x_{t-1}\Delta X_{t-1}^{(k)\prime}\widehat{\varepsilon}_{p_T,t}^2 \Big(\sum_{t=p_T+1}^T \Delta X_{t-1}^{(k)}\Delta X_{t-1}^{(k)\prime}\widehat{\varepsilon}_{p_T,t}^2\Big)^{-1} \sum_{t=p_T+1}^T x_{t-1}\Delta X_{t-1}^{(k)\prime}\widehat{\varepsilon}_{p_T,t}^2,$$

where $\hat{\Omega}_0 := \sum_{t=p_T+1}^T x_{t-1}^2 \widehat{\varepsilon}_{p_T,t}^2$ and $\Delta X_{t-1}^{(k)} := (\Delta x_{t-1}, ..., \Delta x_{t-k})'$. Using the poisitive definiteness of $\sum_{t=p_T+1}^T (x_{t-1}, \Delta X_{t-1}')(x_{t-1}, \Delta X_{t-1}')'\widehat{\varepsilon}_{p_T,t}^2$, it follows that $\hat{\Omega}_{k_1} \geq \hat{\Omega}_{k_2}$ for any $k_1 < k_2$ between 0 and p_T. As a result, it holds that

$$se_{W,1}(\widehat{\rho}) = \Big(\sum_{t=p_T+1}^T x_{t-1}^2\Big)^{-1}\hat{\Omega}_0^{1/2} \geq \Big(\sum_{t=p_T+1}^T x_{t-1}^2\Big)^{-1}\hat{\Omega}_1^{1/2} \geq \Big(\sum_{t=p_T+1}^T x_{t-1}^2\Big)^{-1}\hat{\Omega}_{p_T}^{1/2}.$$

If we could show that $\hat{\Omega}_1^{1/2}/\hat{\Omega}_0^{1/2}$ converges weakly to some r.v. θ, then from the previous chain of inequalities it would follow that

$$\liminf_{T\to\infty} P\left(\Big(\sum_{t=p_T+1}^T x_{t-1}^2\Big)^{-1}\hat{\Omega}_{p_T}^{1/2}/se_{W,1}(\widehat{\rho}) \leq z\right) \geq P(\theta \leq z)$$

for all $z \in \mathbb{R}$ where the cdf of θ is continuous. Finally, in view of (21), also

$$\liminf_{T\to\infty} P\left(t_{\hat{\rho},1}^W/t_{\hat{\rho},2}^W \leq z\right) = \liminf_{T\to\infty} P\left(se_{W,2}(\widehat{\rho})/se_{W,1}(\widehat{\rho}) \leq z\right) \geq P(\theta \leq z)$$

would hold.

We turn to the weak convergence of $\hat{\Omega}_1^{1/2}/\hat{\Omega}_0^{1/2}$ in the case where x_t is a pure random walk and $x_0 = 0$. Under the rate condition of Theorem 4.1, we can replace p_T by 1 in the definition of $\hat{\Omega}_k$ without changing their limit behaviour. First, $T^{-2}\hat{\Omega}_0 = T^{-2}\sum_{t=1}^T x_{t-1}^2\widehat{\varepsilon}_{p_T,t} = T^{-2}\sum_{t=1}^T x_{t-1}^2\varepsilon_t^2 +$

$o_p(1) \Rightarrow \sigma_1^4 \int_0^1 H_{c,\nu}^2(r) d[H_{0,\nu}]_r$, with $\psi(1) = 1$ for the random walk, as established in the proof of Theorem 4.1. Second, we need to discuss

$$T^{-2}\hat{\Omega}_1 = T^{-2}\hat{\Omega}_0 - \{T^{-1}\sum_{t=2}^{T}(\Delta X_{t-1})^2\widehat{\varepsilon}_{1,t}^2\}^{-1}\{T^{-3/2}\sum_{t=2}^{T}\Delta X_{t-1}x_{t-1}\widehat{\varepsilon}_{1,t}^2\}^2$$

with $\Delta X_t = \Delta x_t = \varepsilon_t$, where, as a result of a standard argument,

$$\begin{aligned} T^{-3/2}\sum_{t=2}^{T} x_{t-1}\Delta X_{t-1}\widehat{\varepsilon}_{1,t}^2 &= \gamma^2\sum_{t=1}^{T}(T^{-1/2}x_{t-1})\left(a_T^{-2}\varepsilon_{1,t-1}\varepsilon_{2t}^2\right) + o_p(1), \\ T^{-1}\sum_{t=2}^{T}(\Delta X_{t-1})^2\widehat{\varepsilon}_{1,t}^2 &= T^{-1}\sum_{t=1}^{T}\varepsilon_{t-1}^2\varepsilon_t^2 + o_p(1) \\ &= T^{-1}\sum_{t=1}^{T}\varepsilon_{1,t-1}^2\varepsilon_{1,t}^2 + a_T^{-2}\gamma^2\sum_{t=1}^{T}(\varepsilon_{1,t-1}^2\varepsilon_{2,t}^2 + \varepsilon_{2,t-1}^2\varepsilon_{1,t}^2) + o_p(1) \end{aligned}$$

as $T \to \infty$. To obtain the weak limit θ of $\hat{\Omega}_1^{1/2}/\hat{\Omega}_0^{1/2}$ as in (24), it remains to show that, in terms of the limiting processes in eq. (23),

$$\sum_{t=1}^{T}(T^{-1/2}x_{t-1})\left(a_T^{-2}\varepsilon_{1,t-1}\varepsilon_{2t}^2\right) \Rightarrow \sigma_1\int_0^1 H_{\nu,0}(r)dS(r), \tag{A.18}$$

$$T^{-1}\sum_{t=1}^{T}\varepsilon_{1,t-1}^2\varepsilon_{1,t}^2 + a_T^{-2}\gamma^2\sum_{t=1}^{T}(\varepsilon_{1,t-1}^2\varepsilon_{2,t}^2 + \varepsilon_{2,t-1}^2\varepsilon_{1,t}^2) \Rightarrow \sigma_1^4 + \gamma^2(Q_-(1) + Q_+(1)) \tag{A.19}$$

jointly with the convergence of $\hat{\Omega}_0$, which we do next.

Proof of Proposition 4.2. First, we argue that

$$\sum_{t=1}^{[Tr]}\begin{pmatrix} a_T^{-1}\varepsilon_{2t} \\ a_T^{-2}\boldsymbol{\varepsilon}_t\varepsilon_{2t}^2 \\ T^{-1/2}\varepsilon_{1t} \end{pmatrix} \Rightarrow (\mathcal{U}_\alpha(r), [\mathcal{U}_\alpha]_r, S(r), Q_-(r), Q_+(r), \sigma_1 W(r))' \tag{A.20}$$

in $D_6[0,1]$, where W is independent of the remaining components of the limit process. Second, as W is continuous a.s., (23) obtains by an application of the CMT.

To start from a situation with tail-balanced components, we introduce $z_t := (\mathrm{sgn}(\varepsilon_{2t}), \boldsymbol{\varepsilon}_t')'$ and $Z_t := z_t\varepsilon_{2t}^2$, where $\mathrm{sgn}(\varepsilon_{2t}) := 1$ if $\varepsilon_{2t} \geq 0$ and $\mathrm{sgn}(\varepsilon_{2t}) := -1$ if $\varepsilon_{2t} < 0$; then z_t and ε_{2t}^2 are independent under symmetry of ε_{2t}. Since z_t have finite variance, Z_t are regulary varying with tail index $\alpha/2$ (the same as ε_{2t}^2), by the multivariate Breiman lemma (Proposition A.1 of Basrak et al., 2002). If we were given an IID sequence $\{z_t^*\} = \{(\mathrm{sgn}(\varepsilon_{2t}), \boldsymbol{\varepsilon}_t^{*\prime})'\}$ independent of $\{\varepsilon_{2t}^2\}$ and with each $\boldsymbol{\varepsilon}_t^*$ distributed like $\boldsymbol{\varepsilon}_1$, for $Z_t^* = z_t^*\varepsilon_{2t}^2$ it would hold that $a_T^{-2}\sum_{t=1}^{[Tr]} Z_t^* \Rightarrow L(r)$ in $D_5[0,1]$, where L is a $\alpha/2$-stable Lévy process with series representation $L(r) = \sum_{i=1}^{\infty}\mathbb{I}_{\{U_i \leq r\}}\Gamma_i^{-2/\alpha}\theta_i$, $\{\Gamma_i\}_i$ is the partial sum sequence of an IID standard exponential sequence, $\{U_i\}_i$ is an IID sequence of uniform random variables on $[0,1]$, and $\{\Gamma_i\}_i$, $\{U_i\}_i$ and $\{\theta_i\}_i \stackrel{d}{=} \{z_i^*\}_i$ are jointly independent. This conclusion could be drawn as in Theorem 1 of LePage *et al.* (1981).

Although $\{z_t\}$ is not IID, $a_T^{-2}\sum_{t=1}^{[Tr]} Z_t$ converges like $a_T^{-2}\sum_{t=1}^{[Tr]} Z_t^*$ by a result of Tyran-Kamińska (2010) for regularly varying processes with isolated extremes in the sense of Davis (1983). Specifically, the periods of extreme values are inherited from ε_{2t}, which is IID and trivially has isolated extremes. This fact allows also for a direct proof of convergence, by approximation, as follows. Possibly upon an expansion of the probability space, take $\{\boldsymbol{\varepsilon}_t^*\}$ distributed as before and independent of $\{\varepsilon_{1t}, \varepsilon_{2t}\}$. For a fixed $\delta > 0$, let I_t be the indicator of the event that $\{|a_T^{-1}\varepsilon_{2t}| > \delta$ and

$|a_T^{-1}\varepsilon_{2,t-2}| \leq \delta$ and $\{|a_T^{-1}\varepsilon_{2,t+2}| \leq \delta\}$. Then $\tilde{Z}_t = \iota_{t\delta} Z_t + (1-\iota_{t\delta}) Z_t^*$, $t \in \mathbb{N}$, defines an IID sequence independent of $\{\varepsilon_{2t}\}$, and hence, $a_T^{-2}\sum_{t=1}^{[Tr]} \tilde{Z}_t \Rightarrow L(r)$ in $D_5[0,1]$. On the other hand, for every $\lambda > 0$,

$$\lim_{\delta\to 0} \limsup_{T\to\infty} P\left(\max_{s=1,\ldots,T} \left\| a_T^{-2} \sum_{t=1}^{s} (\tilde{Z}_t - Z_t) \right\| \geq \lambda \right) = 0,$$

which by Theorem 4.2 of Billingsley (1968) implies that also $a_T^{-2}\sum_{t=1}^{[Tr]} Z_t \Rightarrow L(r)$ in $D_5[0,1]$. In fact, let

$$e_t = \{Var(\varepsilon_{2t}^2 \mathbb{I}_{\{|a_T^{-1}\varepsilon_{2,t-2}|\leq\delta\}})\}^{-1/2} \{\varepsilon_{2t}^2 \mathbb{I}_{\{|a_T^{-1}\varepsilon_{2,t-2}|\leq\delta\}} - E(\varepsilon_{2t}^2 \mathbb{I}_{\{|a_T^{-1}\varepsilon_{2,t-2}|\leq\delta\}})\};$$

since $\sum_{t=1}^{T} \iota_{t\delta} = \sum_{t=1}^{T} \mathbb{I}_{\{|a_T^{-1}\varepsilon_{2t}|>\delta\}}$ with probability approaching one as $T \to \infty$, with the same probability it holds that

$$\begin{aligned}
\max_{s=1,\ldots,T} \left\| a_T^{-2} \sum_{t=1}^{s} (\tilde{Z}_t - Z_t) \right\| &= \max_{s=1,\ldots,T} \left\| a_T^{-2} \sum_{t=1}^{[Tr]} \varepsilon_{2t}^2 \mathbb{I}_{\{|a_T^{-1}\varepsilon_{2t}|<\delta\}} (\boldsymbol{\varepsilon}_t^* - \boldsymbol{\varepsilon}_t) \right\| \\
&\leq T^{1/2} E(a_T^{-2}\varepsilon_{2t}^2 \mathbb{I}_{\{|a_T^{-1}\varepsilon_{2t}|\leq\delta\}}) \max_{r\in[0,1]} \left\| T^{-1/2} \sum_{t=1}^{[Tr]} (\boldsymbol{\varepsilon}_t^* - \boldsymbol{\varepsilon}_t) \right\| \\
&\quad + \{TE(a_T^{-4}\varepsilon_{2t}^4 \mathbb{I}_{\{|a_T^{-1}\varepsilon_{2,t-2}|\leq\delta\}})\}^{1/2} \max_{s=1,\ldots,T} \left\| T^{-1/2} \sum_{t=1}^{[Tr]} e_t (\boldsymbol{\varepsilon}_t^* - \boldsymbol{\varepsilon}_t) \right\|,
\end{aligned}$$

where $T^{1/2}E(a_T^{-2}\varepsilon_{2t}^2 \mathbb{I}_{\{|a_T^{-1}\varepsilon_{2,t}|\leq\delta\}}) \to 0$ as $T \to \infty$ and $TE(a_T^{-4}\varepsilon_{2t}^4 \mathbb{I}_{\{|a_T^{-1}\varepsilon_{2,t}|\leq\delta\}}) \to \delta^{4-\alpha}\alpha/(4-\alpha) \to 0$ as $T \to \infty$ followed by $\delta \to 0$, both by Karamata's theorem, whereas the maximum over r does not depend on δ and converges weakly as $T \to \infty$ to the maximum on $[0,1]$ of a Wiener processes, while the maximum over s is $O_P(1)$ as $T \to \infty$, uniformly in δ, by Kolmogorov's maximal inequality.

Approximation can again be used to argue that the convergence $a_T^{-2}\sum_{t=1}^{[Tr]} Z_t \Rightarrow L(r)$ is joint with $W_T(r) \Rightarrow \sigma_1 W(r)$ in $D_6[0,1]$. To this end, consider additionally an IID sequence $\{\varepsilon_{1t}^{**}\}$ independent of the random elements introduced so far and with ε_{1t}^{**} distributed like ε_{11}. Next, in W_T replace ε_{1t} by ε_{1t}^{**} whenever ε_{1t} was retained in $\{\tilde{Z}_t\}$:

$$W_{T,\delta}(r) = W_T(r) + T^{-1/2} \sum_{t=1}^{[Tr]} \iota_{t\delta} (\varepsilon_{t-1}^{**} + \varepsilon_{t+1}^{**} - \varepsilon_{t-1} - \varepsilon_{t+1}).$$

Then $W_{T,\delta}$ is distributed like W_T, so $W_{T,\delta}(r) \Rightarrow W(r)$. Since $W_{T,\delta}(r)$ and $a_T^{-2}\sum_{t=1}^{[Tr]} \tilde{Z}_t$ are independent, their convergence is joint and to independent limits. On the other hand, since

$$\begin{aligned}
\max_{r\in[0,1]} \|W_{T,\delta}(r) - W_T(r)\| &\leq (E\iota_{t\delta})^{1/2} \max_{r\in[0,1]} \left\| T^{-1/2} \sum_{t=1}^{[Tr]} \frac{\iota_{t\delta} - E\iota_{t\delta}}{\{Var(\iota_{t\delta})\}^{1/2}} (\varepsilon_{t-1}^{**} + \varepsilon_{t+1}^{**} - \varepsilon_{t-1} - \varepsilon_{t-2}) \right\| \\
&\quad + E\iota_{t\delta} \max_{r\in[0,1]} \left\| T^{-1/2} \sum_{t=1}^{[Tr]} (\varepsilon_{t-1}^{**} + \varepsilon_{t+1}^{**} - \varepsilon_{t-1} - \varepsilon_{t-2}) \right\| \xrightarrow{P} 0
\end{aligned}$$

as $T \to \infty$, because $E\iota_{t\delta} \to 0$ and the maxima over r converge weakly to maxima of Wiener processes with variances independent of δ, we can conclude that $a_T^{-2}\sum_{t=1}^{[Tr]} Z_t$ and $W_T(r)$ converge like $a_T^{-2}\sum_{t=1}^{[Tr]} \tilde{Z}_t$ and $W_{T,\delta}(r)$:

$$\left(a_T^{-2} \sum_{t=1}^{\lfloor Tr \rfloor} Z_t', W_T(r) \right)' \Rightarrow \left(\sum_{i=1}^{\infty} \mathbb{I}_{\{U_i \leq r\}} \Gamma_i^{-2/\alpha} \theta_i', \sigma_1 W(r) \right)' \tag{A.21}$$

in $D_6[0,1]$ as $T \to \infty$, with independent jump and Wiener components.

To obtain the unbalanced-tail convergence (A.20), for every $\delta > 0$ consider the J_1-continuous (Jacod and Shiryaev (2003), Corollary VI.2.8) map $G_\delta : D_6[0,1] \mapsto D_7[0,1]$ defined for $V \in D_6[0,1]$ by

$$V \overset{G_\delta}{\longmapsto} \left(\sum_{s \le r} g_\delta(\Delta V_1(s)), V'(r) \right)'_{r \in [0,1]},$$

where $g_\delta(x) = \operatorname{sgn}(x)\sqrt{|x|}\{\mathbb{I}_{|x| \ge 2\delta} + \mathbb{I}_{|x| \in [\delta, 2\delta)}(\delta^{-1}|x| - 1)\}$ is continuous and vanishes on $(-\delta, \delta)$. From the CMT it follows that, jointly with (A.21),

$$\sum_{t=1}^{\lfloor Tr \rfloor} g_\delta(a_T^{-2} \operatorname{sgn}(\varepsilon_{2t})\varepsilon_{2t}^2) \Rightarrow \sum_{i=1}^{\infty} \mathbb{I}_{\{U_i \le r\}} g_\delta(\Gamma_i^{-2/\alpha}\theta_{i1}).$$

Further, by Kolmogorov's maximal inequality and Karamata's theorem, for all $\lambda > 0$ it holds that

$$P\left(\sup_{r \in [0,1]} \left| \sum_{t=1}^{\lfloor Tr \rfloor} (g_\delta(a_T^{-2} \operatorname{sgn}(\varepsilon_{2t})\varepsilon_{2t}^2)) - a_T^{-1}\varepsilon_{2t}) \right| \ge \lambda \right) \le \lambda^{-2} a_T^{-2} T E(\varepsilon_{21}^2 \mathbb{I}_{|a_T^{-1}\varepsilon_{21}| \le \sqrt{2\delta}})$$

$$\underset{T \to \infty}{\to} \frac{1}{\lambda^2} \frac{\alpha}{2-\alpha} (2\delta)^{1-\frac{\alpha}{2}} \underset{\delta \to 0}{\to} 0,$$

whereas

$$\sup_{r \in [0,1]} \left| \sum_{i=1}^{\infty} \mathbb{I}_{\{U_i \le r\}} \{ g_\delta(\Gamma_i^{-2/\alpha}\theta_{i1}) - \Gamma_i^{-2/\alpha}\theta_{i1} \} \right| \le \sum_{i=1}^{\infty} \mathbb{I}_{|\Gamma_i^{-2/\alpha}| \le 2\delta} \Gamma_i^{-2/\alpha} = \sum_{j = \min\{i : |\Gamma_i^{-2/\alpha}| \le 2\delta\}}^{\infty} \Gamma_j^{-2/\alpha} \underset{\delta \to 0}{\overset{a.s.}{\to}} 0$$

because $\Gamma_j^{-2/\alpha} \sim j^{-2/\alpha}$ a.s. implies $\min\{i : |\Gamma_i^{-2/\alpha}| \le 2\delta\} \to \infty$ a.s. when $\delta \to 0$. Therefore, by Theorem 4.2 of Billingsley (1968),

$$\left(a_T^{-1} \sum_{t=1}^{\lfloor Tr \rfloor} \varepsilon_{2t}, a_T^{-2} \sum_{t=1}^{\lfloor Tr \rfloor} Z_t', W_T(r) \right)' \Rightarrow \left(\sum_{i=1}^{\infty} \mathbb{I}_{\{U_i \le r\}} (\Gamma_i^{-1/\alpha}\theta_{i1}, \Gamma_i^{-2/\alpha}\theta_i')', \sigma_1 W(r) \right),$$

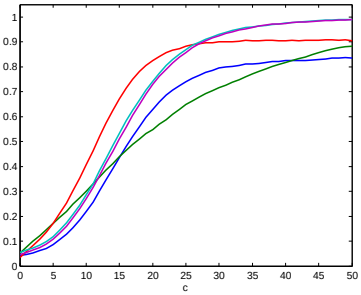
which by the CMT implies (A.20), and hence, (23), with

$$(\mathcal{U}_\alpha(r), [\mathcal{U}_\alpha]_r, S(r), Q_-(r), Q_+(r))' = \sum_{i=1}^{\infty} \mathbb{I}_{\{U_i \le r\}} (\Gamma_i^{-1/\alpha}\theta_{i1}, \Gamma_i^{-2/\alpha}(1, \theta_{i3}, \theta_{i4}, \theta_{i5})')'.$$

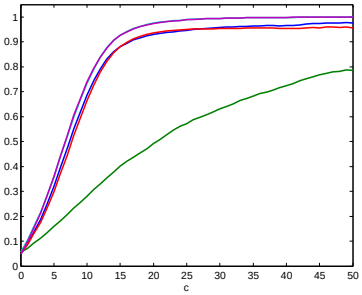
From here (A.19) follows directly. As to (A.18), it follows from Theorerm 2.7 of Kurtz and Protter (1991) since the UT (uniform tightness) condition can be checked as on pp.784-786 of Paulauskas and Rachev (1998). ■

OLS de-meaning

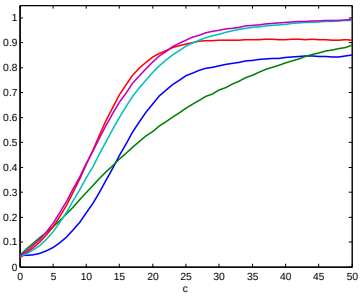
GLS de-meaning



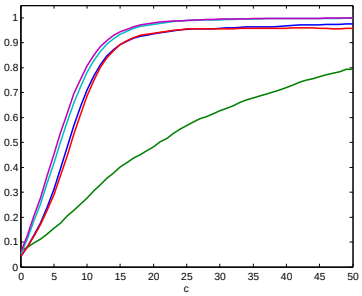
(A) $\alpha = 1.75, \gamma = 0.1$



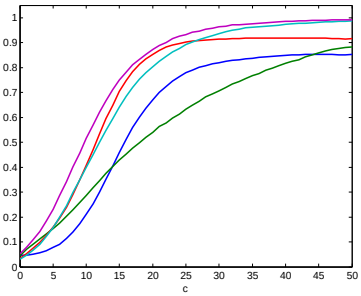
(B) $\alpha = 1.75, \gamma = 0.1$



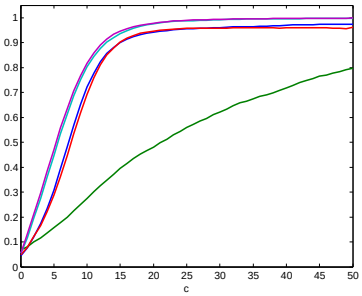
(C) $\alpha = 1.75, \gamma = 1$



(D) $\alpha = 1.75, \gamma = 1$



(E) $\alpha = 1.75, \gamma = 10$

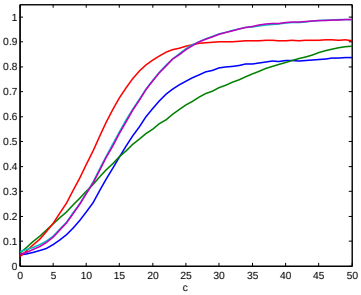


(F) $\alpha = 1.75, \gamma = 10$

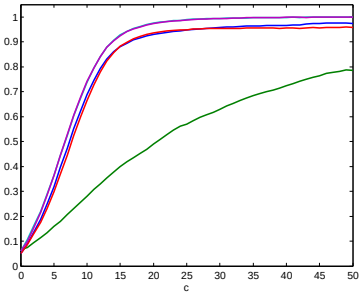
FIGURE 1: Local power of unit root tests under OLS and local GLS de-meaning when $T = 200$. The DGP is (25) and (26) with $c \in \{0, 1, 2, ..., 50\}$ and $\varphi = \theta = 0$.
Key: $t_{\hat{\rho}}$ ——; $\mathcal{VRT}$ ——; $\mathcal{MSB}$ ——; $t^{W}_{\hat{\rho},1}$ ——; $t^{W,\alpha_u}_{\hat{\rho},1}$ ——.

OLS de-meaning

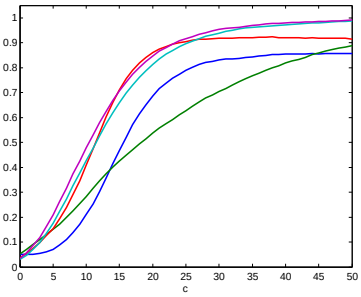
GLS de-meaning



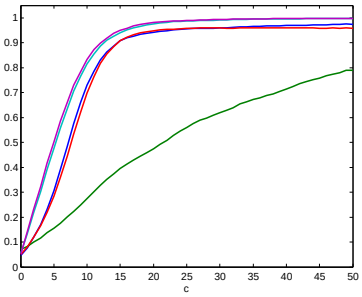
(A) $\alpha = 1.5, \gamma = 0.1$



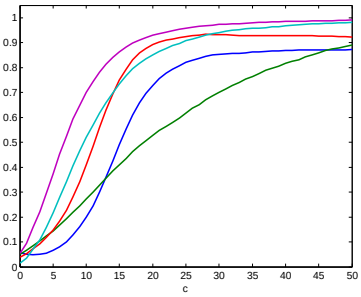
(B) $\alpha = 1.5, \gamma = 0.1$



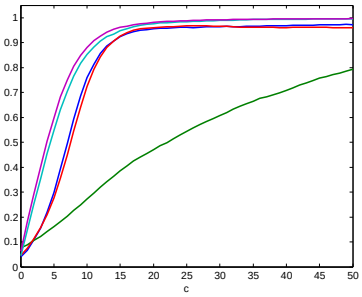
(C) $\alpha = 1.5, \gamma = 1$



(D) $\alpha = 1.5, \gamma = 1$



(E) $\alpha = 1.5, \gamma = 10$

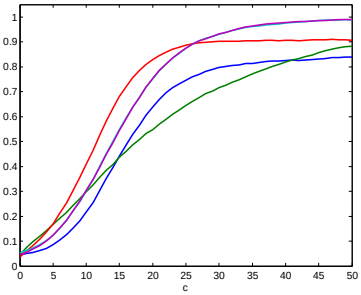


(F) $\alpha = 1.5, \gamma = 10$

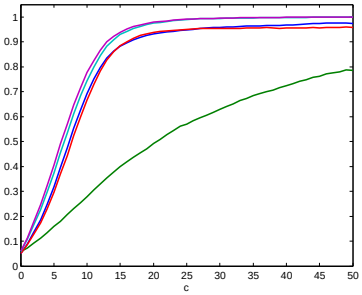
FIGURE 2: Local power of unit root tests under OLS and local GLS de-meaning when $T = 200$. The DGP is (25) and (26) with $c \in 0, 1, 2, ..., 50$ and $\varphi = \theta = 0$. Key: $t_{\hat{\rho}}$ ——; $\mathcal{VRT}$ ——; $\mathcal{MSB}$ ——; $t^{W}_{\hat{\rho},1}$ ——; $t^{W,\alpha_u}_{\hat{\rho},1}$ ——.

OLS de-meaning

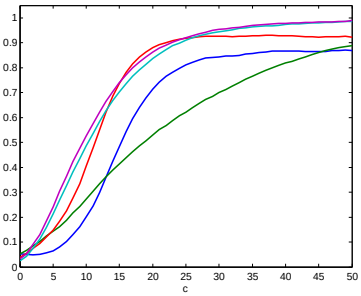
GLS de-meaning



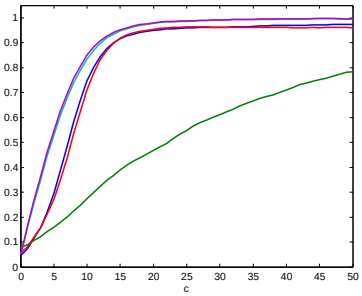
(A) $\alpha = 1.25, \gamma = 0.1$



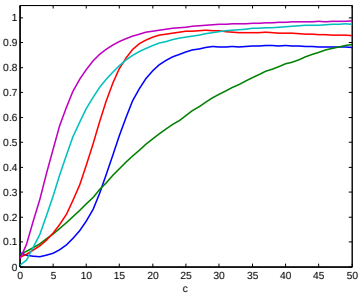
(B) $\alpha = 1.25, \gamma = 0.1$



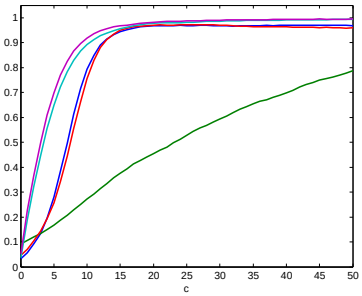
(C) $\alpha = 1.25, \gamma = 1$



(D) $\alpha = 1.25, \gamma = 1$



(E) $\alpha = 1.25, \gamma = 10$

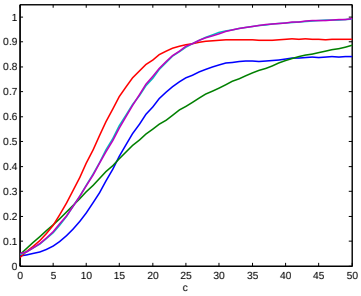


(F) $\alpha = 1.25, \gamma = 10$

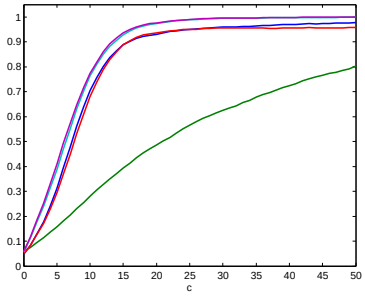
FIGURE 3: Local power of unit root tests under OLS and local GLS de-meaning when $T = 200$. The DGP is (25) and (26) with $c \in 0, 1, 2, ..., 50$ and $\varphi = \theta = 0$. Key: $t_{\hat{\rho}}$ ——; $\mathcal{VRT}$ ——; $\mathcal{MSB}$ ——; $t^{W}_{\hat{\rho},1}$ ——; $t^{W,\alpha_u}_{\hat{\rho},1}$ ——.

OLS de-meaning

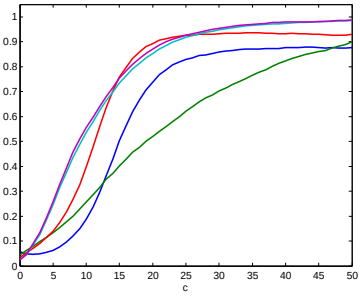
GLS de-meaning



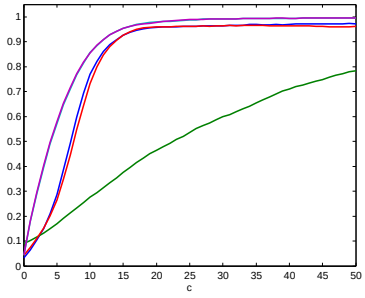
(A) OLS de-meaning, $\alpha = 1.0$,$\gamma = 0.1$



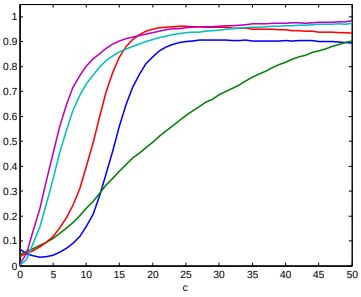
(B) GLS de-meaning, $\alpha = 1.0$,$\gamma = 0.1$



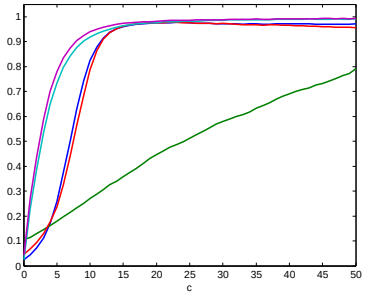
(C) OLS de-meaning, $\alpha = 1.0$,$\gamma = 1$



(D) GLS de-meaning, $\alpha = 1.0$,$\gamma = 1$



(E) OLS de-meaning, $\alpha = 1.0$,$\gamma = 10$

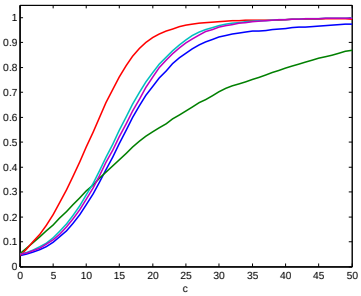


(F) GLS de-meaning, $\alpha = 1.0$,$\gamma = 10$

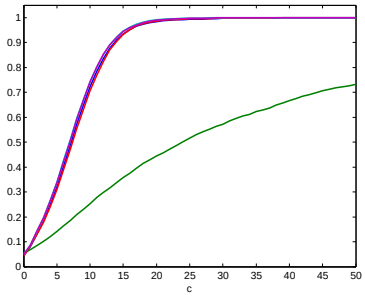
FIGURE 4: Local power of unit root tests under OLS and local GLS de-meaning when $T = 200$. The DGP is (25) and (26) with $c \in 0, 1, 2, ..., 50$ and $\varphi = \theta = 0$. Key: $t_{\hat{\rho}}$; $\mathcal{VRT}$; $\mathcal{MSB}$; $t^{W}_{\hat{\rho},1}$; $t^{W,\alpha_u}_{\hat{\rho},1}$.

OLS de-meaning

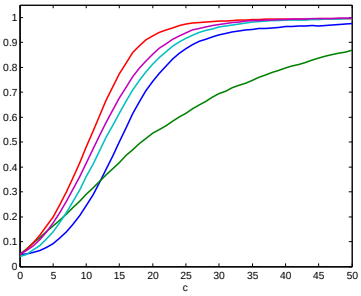
GLS de-meaning



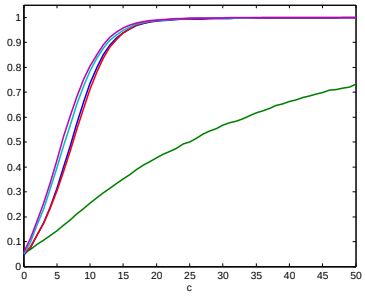
(A) $\alpha = 1.75, \gamma = 0.1$



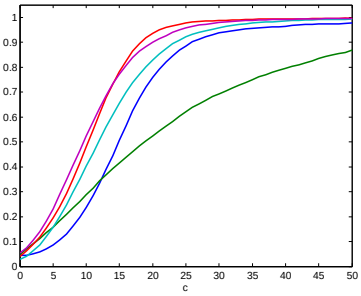
(B) $\alpha = 1.75, \gamma = 0.1$



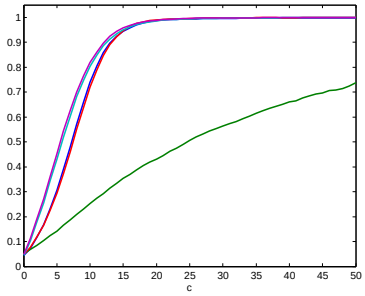
(C) $\alpha = 1.75, \gamma = 1$



(D) $\alpha = 1.75, \gamma = 1$



(E) $\alpha = 1.75, \gamma = 10$

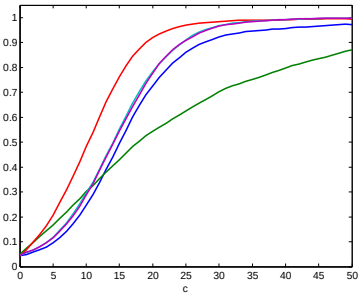


(F) $\alpha = 1.75, \gamma = 10$

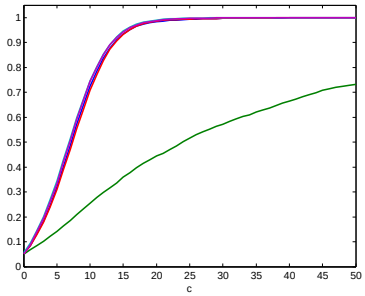
FIGURE 5: Local power of unit root tests under OLS and local GLS de-meaning when $T = 500$. The DGP is (25) and (26) with $c \in 0, 1, 2, ..., 50$ and $\varphi = \theta = 0$. Key: $t_{\hat{\rho}}$ ——; $\mathcal{VRT}$ ——; $\mathcal{MSB}$ ——; $t^{W}_{\hat{\rho},1}$ ——; $t^{W,\alpha_u}_{\hat{\rho},1}$ ——.

OLS de-meaning

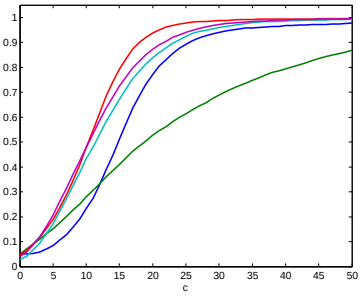
GLS de-meaning



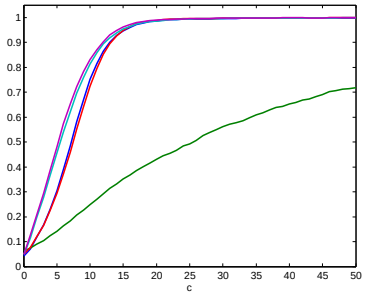
(A) $\alpha = 1.5, \gamma = 0.1$



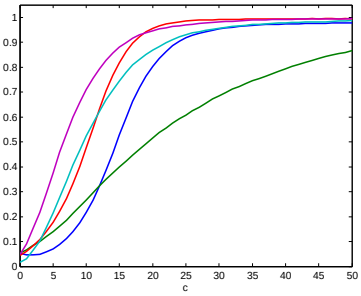
(B) $\alpha = 1.5, \gamma = 0.1$



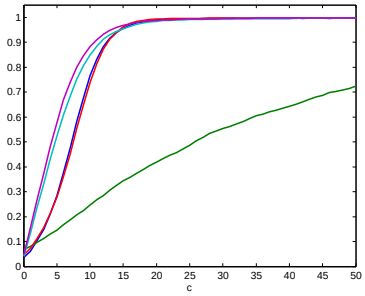
(C) $\alpha = 1.5, \gamma = 1$



(D) $\alpha = 1.5, \gamma = 1$



(E) $\alpha = 1.5, \gamma = 10$

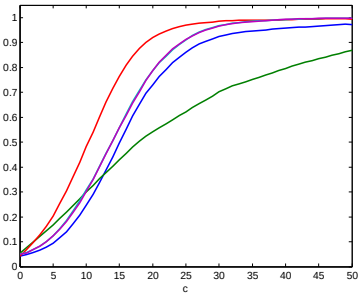


(F) $\alpha = 1.5, \gamma = 10$

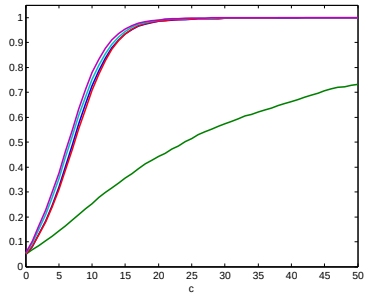
FIGURE 6: Local power of unit root tests under OLS and local GLS de-meaning when $T = 500$. The DGP is (25) and (26) with $c \in 0, 1, 2, ..., 50$ and $\varphi = \theta = 0$. Key: $t_{\hat{\rho}}$; $\mathcal{VRT}$; $\mathcal{MSB}$; $t^{W}_{\hat{\rho},1}$; $t^{W,\alpha_u}_{\hat{\rho},1}$.

OLS de-meaning

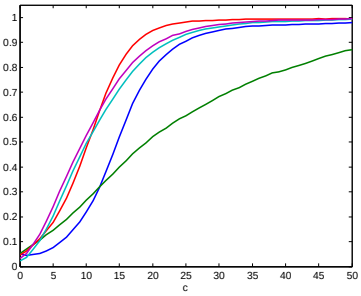
GLS de-meaning



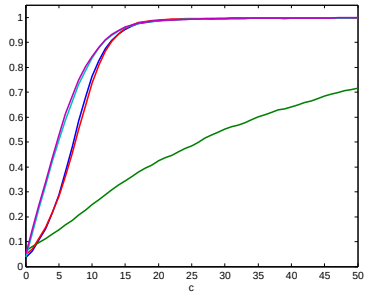
(A) $\alpha = 1.25, \gamma = 0.1$



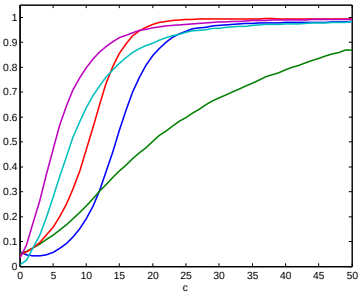
(B) $\alpha = 1.25, \gamma = 0.1$



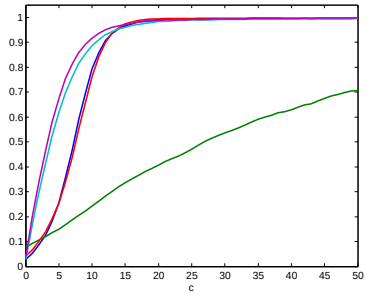
(C) $\alpha = 1.25, \gamma = 1$



(D) $\alpha = 1.25, \gamma = 1$



(E) $\alpha = 1.25, \gamma = 10$

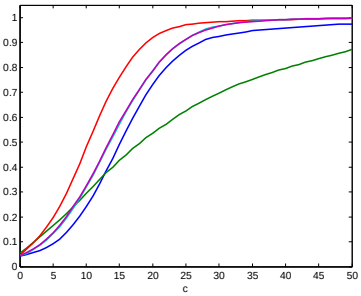


(F) $\alpha = 1.25, \gamma = 10$

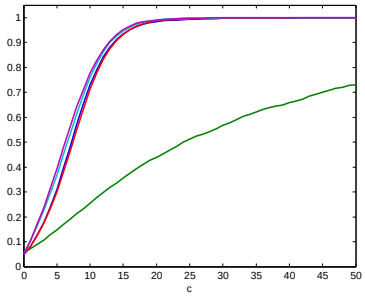
FIGURE 7: Local power of unit root tests under OLS and local GLS de-meaning when $T = 500$. The DGP is (25) and (26) with $c \in 0, 1, 2, ..., 50$ and $\varphi = \theta = 0$. Key: $t_{\hat{\rho}}$ ——; $\mathcal{VRT}$ ——; $\mathcal{MSB}$ ——; $t^{W}_{\hat{\rho},1}$ ——; $t^{W,\alpha_u}_{\hat{\rho},1}$ ——.

OLS de-meaning

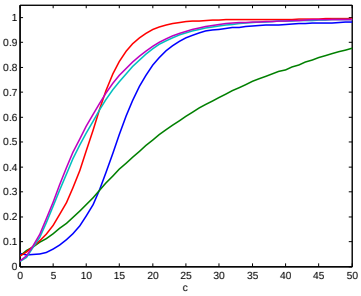
GLS de-meaning



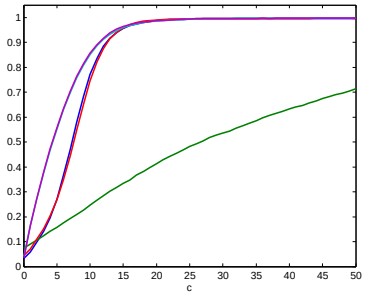
(A) $\alpha = 1.0, \gamma = 0.1$



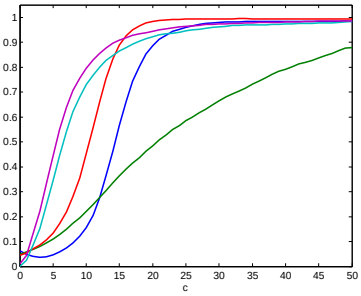
(B) $\alpha = 1.0, \gamma = 0.1$



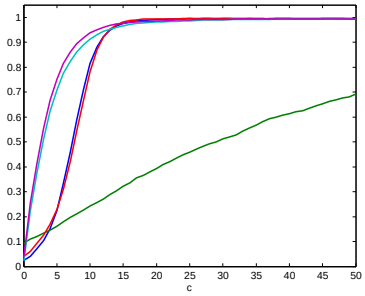
(C) $\alpha = 1.0, \gamma = 1$



(D) $\alpha = 1.0, \gamma = 1$



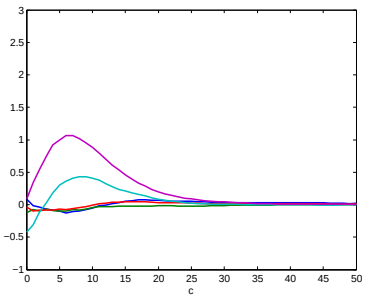
(E) $\alpha = 1.0, \gamma = 10$



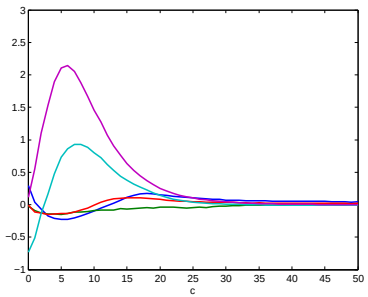
(F) $\alpha = 1.0, \gamma = 10$

FIGURE 8: Local power of unit root tests under OLS and local GLS de-meaning when $T = 500$. The DGP is (25) and (26) with $c \in 0, 1, 2, ..., 50$ and $\varphi = \theta = 0$. Key: $t_{\hat{\rho}}$ ——; $\mathcal{VRT}$ ——; $\mathcal{MSB}$ ——; $t^{W}_{\hat{\rho},1}$ ——; $t^{W,\alpha_u}_{\hat{\rho},1}$ ——.

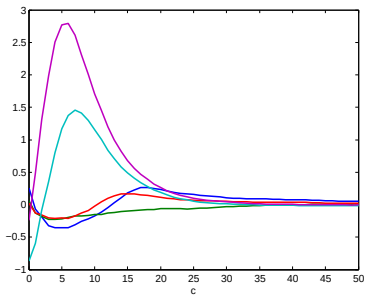
OLS de-meaning



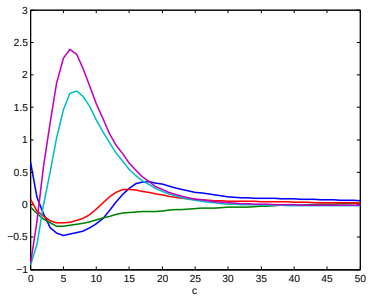
(A) $\alpha = 1.75$



(B) $\alpha = 1.50$

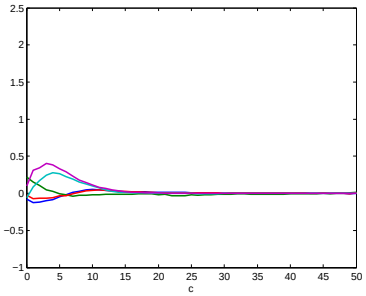


(C) $\alpha = 1.25$

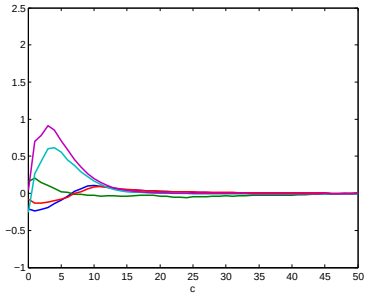


(D) $\alpha = 1.00$

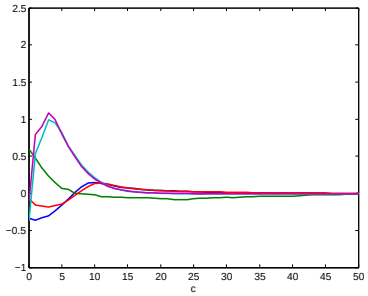
GLS de-meaning



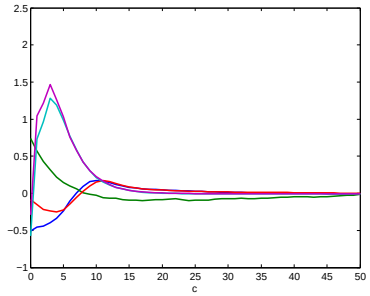
(E) $\alpha = 1.75$



(F) $\alpha = 1.50$



(G) $\alpha = 1.25$

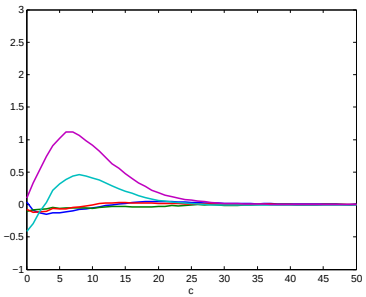


(H) $\alpha = 1.00$

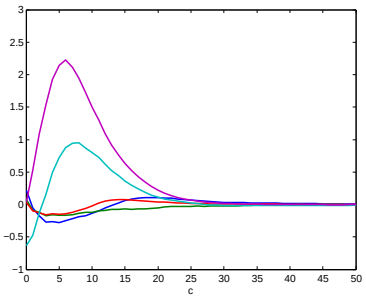
FIGURE 9: Local power variation of the unit root tests. The DGP is (25) and (26) with $c \in 0, 1, 2, ..., 50$, $\varphi = \theta = 0$ and $T = 200$.

Key: $t_{\hat{\rho}}$ ——; $\mathcal{VRT}$ ——; $\mathcal{MSB}$ ——; $t^{W}_{\hat{\rho},1}$ ——; $t^{W,\alpha_u}_{\hat{\rho},1}$ ——.

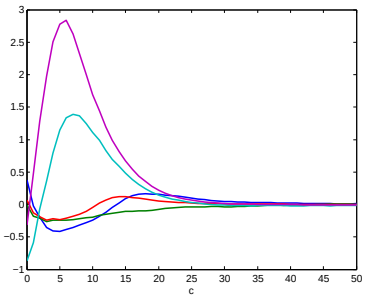
OLS de-meaning



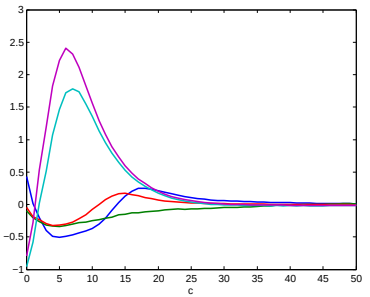
(A) $\alpha = 1.75$



(B) $\alpha = 1.50$

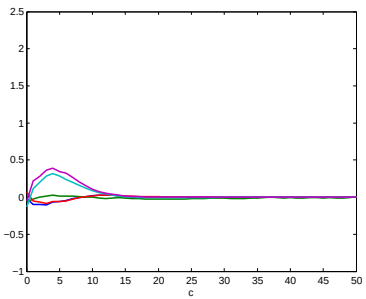


(C) $\alpha = 1.25$

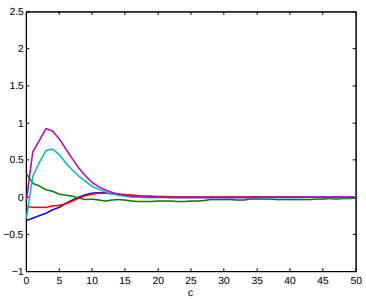


(D) $\alpha = 1.00$

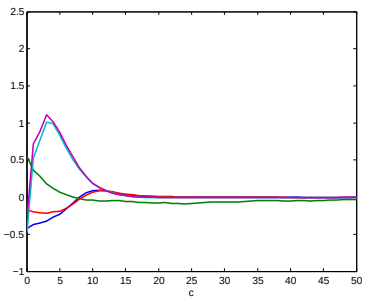
GLS de-meaning



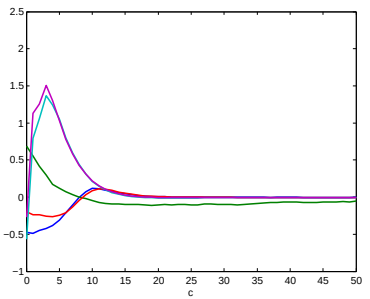
(E) $\alpha = 1.75$



(F) $\alpha = 1.50$



(G) $\alpha = 1.25$



(H) $\alpha = 1.00$

FIGURE 10: Local power variation of the unit root tests. The DGP is (25) and (26) with $c \in 0, 1, 2, ..., 50$, $\varphi = \theta = 0$ and $T = 500$.

Key: $t_{\hat{\rho}}$ ——; $\mathcal{VRT}$ ——; $\mathcal{MSB}$ ——; $t^{W}_{\hat{\rho},1}$ ——; $t^{W,\alpha_u}_{\hat{\rho},1}$ ——.

TABLE 1. Unconditional '(u)' and conditional '(c)' cdfs of $\tau_{c,\alpha}$ at $x \in \{0.5, 1, 1.5, 2, 3\}$

x, α	1 (u)	1 (c)	1.5 (u)	1.5 (c)	1.75 (u)	1.75 (c)	1 (u)	1 (c)	1.5 (u)	1.5 (c)	1.75 (u)	1.75 (c)
	$c = 0$						$c = 15$					
0.5	0.13	0.80	0.12	0.53	0.08	0.31	0.13	0.80	0.12	0.53	0.07	0.30
1	**0.23**	**0.89**	**0.32**	**0.77**	**0.37**	**0.67**	**0.23**	**0.89**	**0.32**	**0.77**	**0.36**	**0.67**
1.5	0.31	0.95	0.51	0.92	0.70	0.91	0.32	0.95	0.51	0.92	0.69	0.91
2	0.37	0.98	0.62	0.98	0.81	0.98	0.38	0.98	0.62	0.98	0.80	0.98
3	0.47	1.00	0.73	1.00	0.88	0.99	0.46	1.00	0.73	1.00	0.87	0.99

TABLE 2. Average estimate of $\hat{\alpha}_u$

		$T = 100$	$T = 1000$	$T = 10000$	$T = 100000$
	α_2	$\widehat{\alpha}_u$	$\widehat{\alpha}_u$	$\widehat{\alpha}_u$	$\widehat{\alpha}_u$
$\gamma = 0.1$	2.00	1.97	1.99	2.00	2.00
	1.75	1.97	1.99	2.00	2.00
	1.50	1.96	1.99	2.00	2.00
	1.25	1.96	1.99	2.00	2.00
	1.00	1.96	1.99	2.00	2.00
$\gamma = 0.316$	2.00	1.97	1.99	2.00	2.00
	1.75	1.96	1.98	1.99	1.99
	1.50	1.94	1.97	1.99	1.99
	1.25	1.93	1.97	1.99	1.99
	1.00	1.92	1.97	1.99	2.00
$\gamma = 1$	2.00	1.98	1.99	2.00	2.00
	1.75	1.89	1.91	1.93	1.94
	1.50	1.84	1.89	1.93	1.96
	1.25	1.82	1.90	1.95	1.98
	1.00	1.81	1.92	1.97	1.99
$\gamma = 3.16$	2.00	1.98	1.99	2.00	2.00
	1.75	1.79	1.80	1.81	1.82
	1.50	1.62	1.68	1.75	1.82
	1.25	1.55	1.71	1.83	1.92
	1.00	1.58	1.81	1.92	1.97
$\gamma = 10$	2.00	1.98	1.99	2.00	2.00
	1.75	1.76	1.76	1.76	1.76
	1.50	1.52	1.53	1.56	1.61
	1.25	1.31	1.43	1.59	1.74
	1.00	1.24	1.57	1.80	1.92
$\gamma = 31.6$	2.00	1.98	2.00	2.00	2.00
	1.75	1.76	1.75	1.75	1.75
	1.50	1.51	1.50	1.51	1.52
	1.25	1.25	1.28	1.35	1.47
	1.00	1.03	1.26	1.58	1.81

TABLE 3. Empirical size of unit root tests under OLS and local GLS de-meaning. The DGP is (25) and (26) with $T = 200$ and $\gamma = 0.1$.

			OLS de-meaning							Local GLS de-meaning						
φ	θ	α_2	$t_{\widehat{\rho}}$	$\mathcal{VRT}$	$\mathcal{MSB}$	$\mathcal{MZ}_\varphi$	$\mathcal{MZ}_t$	$t^W_{\widehat{\rho},1}$	$t^{W,\widehat{\alpha}_u}_{\widehat{\rho},1}$	$t_{\widehat{\rho}}$	$\mathcal{VRT}$	$\mathcal{MSB}$	$\mathcal{MZ}_\varphi$	$\mathcal{MZ}_t$	$t^W_{\widehat{\rho},1}$	$t^{W,\widehat{\alpha}_u}_{\widehat{\rho},1}$
0	0	1.75	0.042	0.052	0.035	0.038	0.037	0.054	0.047	0.049	0.056	0.051	0.046	0.047	0.057	0.049
		1.5	0.044	0.052	0.039	0.040	0.039	0.058	0.049	0.051	0.066	0.051	0.049	0.050	0.060	0.051
		1.25	0.045	0.047	0.037	0.038	0.036	0.053	0.047	0.050	0.058	0.051	0.049	0.047	0.059	0.067
		1	0.040	0.046	0.036	0.035	0.035	0.046	0.051	0.048	0.061	0.050	0.047	0.048	0.057	0.065
0.5	0	1.75	0.040	0.044	0.048	0.046	0.041	0.055	0.050	0.045	0.052	0.058	0.052	0.051	0.056	0.053
		1.5	0.038	0.045	0.049	0.048	0.043	0.056	0.054	0.051	0.062	0.062	0.058	0.058	0.064	0.052
		1.25	0.044	0.041	0.050	0.049	0.043	0.057	0.051	0.051	0.055	0.063	0.059	0.057	0.060	0.066
		1	0.040	0.040	0.043	0.043	0.039	0.045	0.056	0.046	0.058	0.058	0.053	0.052	0.059	0.067
-0.5	0	1.75	0.041	0.060	0.029	0.030	0.032	0.067	0.061	0.046	0.059	0.044	0.042	0.043	0.065	0.057
		1.5	0.043	0.060	0.033	0.033	0.032	0.069	0.065	0.050	0.070	0.049	0.046	0.046	0.070	0.056
		1.25	0.042	0.056	0.034	0.036	0.031	0.066	0.058	0.048	0.062	0.045	0.044	0.046	0.066	0.076
		1	0.036	0.054	0.030	0.031	0.032	0.058	0.065	0.047	0.064	0.043	0.039	0.043	0.064	0.073
0	0.5	1.75	0.034	0.050	0.039	0.037	0.040	0.057	0.050	0.040	0.057	0.054	0.047	0.047	0.057	0.050
		1.5	0.032	0.050	0.041	0.042	0.038	0.056	0.053	0.046	0.064	0.057	0.054	0.053	0.065	0.054
		1.25	0.036	0.045	0.047	0.046	0.040	0.055	0.050	0.043	0.057	0.054	0.050	0.051	0.062	0.067
		1	0.035	0.044	0.041	0.039	0.039	0.048	0.054	0.040	0.061	0.049	0.046	0.044	0.058	0.065
0	-0.5	1.75	0.045	0.075	0.044	0.044	0.039	0.108	0.097	0.064	0.071	0.059	0.055	0.057	0.100	0.097
		1.5	0.041	0.077	0.046	0.045	0.038	0.102	0.104	0.064	0.072	0.058	0.057	0.057	0.101	0.094
		1.25	0.049	0.071	0.048	0.048	0.045	0.106	0.094	0.062	0.070	0.055	0.054	0.054	0.099	0.112
		1	0.045	0.068	0.041	0.044	0.039	0.097	0.109	0.061	0.067	0.057	0.054	0.054	0.101	0.114

TABLE 4. Empirical size of unit root tests under OLS and local GLS de-meaning. The DGP is (25) and (26) with $T = 200$ and $\gamma = 1$.

			OLS de-meaning							Local GLS de-meaning						
φ	θ	α_2	$t_{\widehat{\rho}}$	$\mathcal{VRT}$	$\mathcal{MSB}$	$\mathcal{MZ}_\varphi$	$\mathcal{MZ}_t$	$t^W_{\widehat{\rho},1}$	$t^{W,\widehat{\alpha}_u}_{\widehat{\rho},1}$	$t_{\widehat{\rho}}$	$\mathcal{VRT}$	$\mathcal{MSB}$	$\mathcal{MZ}_\varphi$	$\mathcal{MZ}_t$	$t^W_{\widehat{\rho},1}$	$t^{W,\widehat{\alpha}_u}_{\widehat{\rho},1}$
0	0	1.75	0.046	0.047	0.036	0.036	0.039	0.042	0.044	0.043	0.067	0.045	0.039	0.040	0.052	0.060
		1.5	0.050	0.052	0.036	0.038	0.043	0.032	0.040	0.046	0.070	0.051	0.044	0.042	0.057	0.055
		1.25	0.053	0.052	0.039	0.039	0.047	0.025	0.033	0.047	0.079	0.056	0.046	0.044	0.053	0.046
		1	0.056	0.045	0.034	0.040	0.053	0.026	0.023	0.032	0.090	0.043	0.031	0.033	0.043	0.047
0.5	0	1.75	0.045	0.040	0.041	0.045	0.044	0.043	0.045	0.041	0.063	0.052	0.045	0.046	0.055	0.062
		1.5	0.050	0.045	0.050	0.049	0.051	0.035	0.041	0.044	0.067	0.062	0.054	0.050	0.056	0.059
		1.25	0.049	0.046	0.050	0.049	0.052	0.027	0.031	0.049	0.076	0.068	0.058	0.057	0.054	0.048
		1	0.055	0.038	0.042	0.046	0.057	0.026	0.024	0.034	0.086	0.051	0.040	0.038	0.044	0.048
-0.5	0	1.75	0.042	0.053	0.029	0.030	0.032	0.052	0.055	0.041	0.070	0.038	0.034	0.036	0.059	0.066
		1.5	0.048	0.060	0.033	0.035	0.039	0.039	0.055	0.045	0.073	0.045	0.038	0.039	0.067	0.065
		1.25	0.050	0.059	0.034	0.035	0.042	0.035	0.043	0.045	0.082	0.052	0.043	0.042	0.059	0.056
		1	0.056	0.052	0.029	0.037	0.051	0.032	0.034	0.033	0.093	0.039	0.030	0.032	0.048	0.057
0	0.5	1.75	0.039	0.043	0.038	0.040	0.039	0.043	0.048	0.038	0.060	0.049	0.043	0.043	0.057	0.060
		1.5	0.042	0.050	0.041	0.042	0.048	0.036	0.042	0.038	0.069	0.054	0.047	0.045	0.054	0.049
		1.25	0.044	0.051	0.046	0.045	0.050	0.027	0.032	0.043	0.080	0.060	0.051	0.050	0.052	0.048
		1	0.051	0.043	0.035	0.042	0.056	0.025	0.023	0.030	0.086	0.045	0.035	0.035	0.042	0.049
0	-0.5	1.75	0.052	0.066	0.043	0.045	0.042	0.089	0.101	0.056	0.080	0.052	0.049	0.049	0.097	0.108
		1.5	0.057	0.072	0.050	0.055	0.051	0.081	0.098	0.061	0.077	0.059	0.053	0.053	0.116	0.108
		1.25	0.056	0.071	0.051	0.056	0.053	0.070	0.084	0.061	0.093	0.064	0.056	0.056	0.112	0.106
		1	0.064	0.068	0.051	0.057	0.061	0.069	0.065	0.050	0.104	0.059	0.045	0.043	0.111	0.123

TABLE 5. Empirical size of unit root tests under OLS and local GLS de-meaning. The DGP is (25) and (26) with $T = 200$ and $\gamma = 10$.

			OLS de-meaning							Local GLS de-meaning						
φ	θ	α_2	$t_{\widehat{\rho}}$	$\mathcal{VRT}$	$\mathcal{MSB}$	$\mathcal{MZ}_\varphi$	$\mathcal{MZ}_t$	$t^W_{\widehat{\rho},1}$	$t^{W,\widehat{\alpha}_u}_{\widehat{\rho},1}$	$t_{\widehat{\rho}}$	$\mathcal{VRT}$	$\mathcal{MSB}$	$\mathcal{MZ}_\varphi$	$\mathcal{MZ}_t$	$t^W_{\widehat{\rho},1}$	$t^{W,\widehat{\alpha}_u}_{\widehat{\rho},1}$
0	0	1.75	0.046	0.045	0.034	0.036	0.039	0.031	0.052	0.045	0.068	0.049	0.044	0.044	0.054	0.054
		1.5	0.056	0.051	0.039	0.042	0.052	0.016	0.057	0.040	0.076	0.047	0.040	0.038	0.045	0.057
		1.25	0.056	0.050	0.038	0.039	0.057	0.007	0.037	0.033	0.093	0.047	0.035	0.033	0.037	0.068
		1	0.067	0.044	0.039	0.044	0.066	0.003	0.011	0.024	0.106	0.046	0.027	0.026	0.024	0.031
0.5	0	1.75	0.044	0.038	0.046	0.045	0.044	0.033	0.052	0.044	0.065	0.056	0.050	0.051	0.055	0.057
		1.5	0.056	0.045	0.046	0.048	0.058	0.019	0.058	0.041	0.073	0.058	0.049	0.047	0.044	0.061
		1.25	0.056	0.046	0.044	0.047	0.061	0.007	0.037	0.032	0.089	0.057	0.043	0.040	0.038	0.067
		1	0.065	0.040	0.043	0.050	0.070	0.004	0.011	0.021	0.103	0.055	0.033	0.029	0.024	0.031
-0.5	0	1.75	0.046	0.051	0.029	0.031	0.036	0.044	0.063	0.044	0.072	0.044	0.040	0.040	0.063	0.064
		1.5	0.056	0.057	0.033	0.038	0.049	0.025	0.069	0.039	0.081	0.043	0.038	0.036	0.054	0.069
		1.25	0.056	0.057	0.033	0.036	0.053	0.013	0.049	0.031	0.098	0.045	0.031	0.030	0.048	0.074
		1	0.067	0.050	0.036	0.045	0.065	0.007	0.019	0.023	0.108	0.041	0.027	0.023	0.030	0.041
0	0.5	1.75	0.038	0.043	0.037	0.036	0.043	0.032	0.054	0.036	0.067	0.049	0.045	0.044	0.053	0.056
		1.5	0.051	0.049	0.040	0.045	0.056	0.019	0.057	0.033	0.074	0.050	0.040	0.039	0.043	0.061
		1.25	0.049	0.049	0.037	0.041	0.058	0.007	0.039	0.028	0.091	0.049	0.038	0.036	0.034	0.061
		1	0.061	0.043	0.035	0.043	0.069	0.006	0.013	0.019	0.103	0.042	0.027	0.025	0.020	0.028
0	-0.5	1.75	0.050	0.063	0.044	0.044	0.044	0.077	0.113	0.059	0.078	0.055	0.050	0.049	0.101	0.109
		1.5	0.063	0.070	0.052	0.055	0.060	0.060	0.126	0.052	0.084	0.056	0.049	0.049	0.108	0.131
		1.25	0.063	0.069	0.050	0.055	0.065	0.044	0.114	0.048	0.104	0.062	0.047	0.047	0.114	0.173
		1	0.074	0.065	0.055	0.061	0.074	0.033	0.062	0.035	0.114	0.056	0.038	0.035	0.118	0.150

TABLE 6. Empirical size of unit root tests under OLS and local GLS de-meaning. The DGP is (25) and (26) with $T = 500$ and $\gamma = 0.1$.

			OLS de-meaning							Local GLS de-meaning						
φ	θ	α_2	$t_{\widehat{\rho}}$	$\mathcal{VRT}$	$\mathcal{MSB}$	$\mathcal{MZ}_\varphi$	$\mathcal{MZ}_t$	$t^W_{\widehat{\rho},1}$	$t^{W,\widehat{\alpha}_u}_{\widehat{\rho},1}$	$t_{\widehat{\rho}}$	$\mathcal{VRT}$	$\mathcal{MSB}$	$\mathcal{MZ}_\varphi$	$\mathcal{MZ}_t$	$t^W_{\widehat{\rho},1}$	$t^{W,\widehat{\alpha}_u}_{\widehat{\rho},1}$
0	0	1.75	0.044	0.053	0.045	0.045	0.042	0.050	0.050	0.045	0.054	0.045	0.046	0.045	0.051	0.050
		1.5	0.044	0.050	0.043	0.044	0.041	0.046	0.049	0.054	0.051	0.053	0.052	0.053	0.056	0.048
		1.25	0.042	0.054	0.043	0.046	0.042	0.050	0.044	0.049	0.051	0.051	0.049	0.049	0.054	0.051
		1	0.043	0.054	0.047	0.046	0.042	0.044	0.040	0.048	0.056	0.050	0.047	0.047	0.051	0.063
0.5	0	1.75	0.045	0.051	0.052	0.052	0.046	0.051	0.054	0.045	0.053	0.047	0.046	0.046	0.053	0.050
		1.5	0.043	0.047	0.051	0.050	0.043	0.045	0.051	0.052	0.050	0.055	0.053	0.053	0.054	0.048
		1.25	0.043	0.051	0.051	0.052	0.043	0.048	0.044	0.049	0.050	0.054	0.052	0.052	0.054	0.052
		1	0.044	0.051	0.052	0.051	0.045	0.044	0.040	0.049	0.055	0.051	0.052	0.051	0.053	0.062
-0.5	0	1.75	0.043	0.056	0.042	0.042	0.040	0.057	0.057	0.047	0.055	0.045	0.045	0.046	0.055	0.055
		1.5	0.042	0.053	0.041	0.041	0.038	0.049	0.056	0.051	0.053	0.051	0.048	0.048	0.059	0.051
		1.25	0.041	0.057	0.039	0.042	0.038	0.054	0.047	0.049	0.052	0.049	0.046	0.046	0.057	0.055
		1	0.044	0.056	0.044	0.044	0.040	0.048	0.046	0.047	0.057	0.047	0.045	0.046	0.053	0.063
0	0.5	1.75	0.043	0.052	0.051	0.054	0.048	0.056	0.056	0.048	0.053	0.052	0.049	0.049	0.055	0.053
		1.5	0.040	0.049	0.052	0.054	0.047	0.049	0.052	0.049	0.051	0.054	0.052	0.052	0.056	0.049
		1.25	0.041	0.053	0.050	0.051	0.046	0.053	0.044	0.047	0.047	0.051	0.051	0.052	0.053	0.052
		1	0.036	0.053	0.053	0.050	0.045	0.045	0.045	0.046	0.056	0.054	0.049	0.048	0.054	0.062
0	-0.5	1.75	0.050	0.062	0.057	0.056	0.049	0.077	0.078	0.058	0.054	0.056	0.055	0.054	0.073	0.074
		1.5	0.049	0.060	0.056	0.055	0.047	0.075	0.075	0.059	0.057	0.056	0.054	0.056	0.074	0.068
		1.25	0.047	0.061	0.058	0.059	0.049	0.073	0.068	0.058	0.054	0.057	0.052	0.055	0.073	0.072
		1	0.048	0.061	0.057	0.057	0.044	0.067	0.065	0.057	0.060	0.052	0.052	0.055	0.074	0.081

TABLE 7. Empirical size of unit root tests under OLS and local GLS de-meaning. The DGP is (25) and (26) with $T = 500$ and $\gamma = 1$.

			OLS de-meaning							Local GLS de-meaning						
φ	θ	α_2	$t_{\widehat{\rho}}$	$\mathcal{VRT}$	$\mathcal{MSB}$	$\mathcal{MZ}_\varphi$	$\mathcal{MZ}_t$	$t^W_{\widehat{\rho},1}$	$t^{W,\widehat{\alpha}_u}_{\widehat{\rho},1}$	$t_{\widehat{\rho}}$	$\mathcal{VRT}$	$\mathcal{MSB}$	$\mathcal{MZ}_\varphi$	$\mathcal{MZ}_t$	$t^W_{\widehat{\rho},1}$	$t^{W,\widehat{\alpha}_u}_{\widehat{\rho},1}$
0	0	1.75	0.049	0.050	0.046	0.047	0.047	0.041	0.051	0.046	0.054	0.051	0.047	0.047	0.051	0.058
		1.5	0.052	0.050	0.046	0.048	0.053	0.028	0.040	0.042	0.063	0.051	0.044	0.043	0.048	0.046
		1.25	0.050	0.053	0.046	0.047	0.048	0.022	0.029	0.038	0.064	0.048	0.039	0.037	0.039	0.048
		1	0.052	0.043	0.043	0.044	0.051	0.021	0.024	0.033	0.074	0.044	0.034	0.031	0.040	0.042
0.5	0	1.75	0.049	0.047	0.048	0.047	0.049	0.041	0.054	0.046	0.052	0.052	0.049	0.047	0.048	0.062
		1.5	0.053	0.046	0.049	0.053	0.056	0.028	0.046	0.044	0.063	0.051	0.045	0.047	0.049	0.050
		1.25	0.048	0.049	0.052	0.054	0.050	0.025	0.032	0.039	0.061	0.051	0.043	0.042	0.039	0.049
		1	0.052	0.041	0.048	0.046	0.055	0.023	0.026	0.032	0.073	0.045	0.037	0.036	0.041	0.043
-0.5	0	1.75	0.048	0.053	0.045	0.045	0.047	0.045	0.049	0.046	0.056	0.047	0.044	0.045	0.053	0.059
		1.5	0.052	0.054	0.044	0.045	0.050	0.031	0.042	0.042	0.064	0.049	0.041	0.041	0.051	0.046
		1.25	0.049	0.055	0.046	0.046	0.047	0.027	0.037	0.039	0.065	0.048	0.040	0.039	0.041	0.049
		1	0.051	0.044	0.042	0.042	0.051	0.025	0.024	0.032	0.075	0.042	0.033	0.031	0.044	0.043
0	0.5	1.75	0.044	0.049	0.052	0.052	0.051	0.043	0.048	0.043	0.056	0.051	0.046	0.046	0.050	0.064
		1.5	0.049	0.049	0.051	0.054	0.055	0.030	0.044	0.041	0.062	0.052	0.044	0.044	0.049	0.048
		1.25	0.047	0.051	0.051	0.054	0.050	0.025	0.030	0.038	0.062	0.055	0.045	0.043	0.043	0.049
		1	0.049	0.042	0.047	0.049	0.055	0.024	0.025	0.032	0.072	0.045	0.036	0.035	0.042	0.046
0	-0.5	1.75	0.053	0.059	0.052	0.056	0.055	0.067	0.074	0.053	0.055	0.054	0.050	0.051	0.070	0.085
		1.5	0.058	0.060	0.055	0.060	0.059	0.050	0.064	0.054	0.065	0.056	0.051	0.050	0.072	0.073
		1.25	0.055	0.062	0.061	0.062	0.055	0.044	0.049	0.047	0.067	0.060	0.049	0.046	0.074	0.079
		1	0.058	0.050	0.055	0.054	0.057	0.042	0.042	0.041	0.081	0.050	0.042	0.039	0.080	0.082

TABLE 8. Empirical size of unit root tests under OLS and local GLS de-meaning. The DGP is (25) and (26) with $T = 500$ and $\gamma = 10$.

			OLS de-meaning							Local GLS de-meaning						
φ	θ	α_2	$t_{\widehat{\rho}}$	$\mathcal{VRT}$	$\mathcal{MSB}$	$\mathcal{MZ}_\varphi$	$\mathcal{MZ}_t$	$t^W_{\widehat{\rho},1}$	$t^{W,\widehat{\alpha}_u}_{\widehat{\rho},1}$	$t_{\widehat{\rho}}$	$\mathcal{VRT}$	$\mathcal{MSB}$	$\mathcal{MZ}_\varphi$	$\mathcal{MZ}_t$	$t^W_{\widehat{\rho},1}$	$t^{W,\widehat{\alpha}_u}_{\widehat{\rho},1}$
0	0	1.75	0.045	0.048	0.041	0.044	0.044	0.029	0.056	0.044	0.050	0.048	0.043	0.043	0.044	0.048
		1.5	0.053	0.054	0.044	0.047	0.054	0.017	0.057	0.037	0.067	0.047	0.040	0.038	0.041	0.055
		1.25	0.058	0.053	0.046	0.049	0.059	0.007	0.027	0.028	0.079	0.042	0.030	0.028	0.028	0.047
		1	0.062	0.048	0.045	0.048	0.064	0.002	0.007	0.025	0.095	0.040	0.025	0.024	0.022	0.025
0.5	0	1.75	0.046	0.044	0.048	0.047	0.045	0.028	0.054	0.045	0.048	0.049	0.048	0.048	0.045	0.048
		1.5	0.055	0.050	0.047	0.049	0.058	0.017	0.058	0.038	0.066	0.051	0.043	0.041	0.042	0.057
		1.25	0.057	0.050	0.050	0.051	0.062	0.007	0.027	0.029	0.078	0.045	0.033	0.032	0.031	0.050
		1	0.063	0.044	0.048	0.050	0.066	0.003	0.007	0.026	0.092	0.045	0.032	0.029	0.022	0.023
-0.5	0	1.75	0.045	0.050	0.039	0.041	0.042	0.034	0.060	0.044	0.051	0.047	0.043	0.042	0.049	0.050
		1.5	0.054	0.057	0.043	0.044	0.052	0.019	0.062	0.037	0.068	0.044	0.038	0.037	0.043	0.059
		1.25	0.057	0.055	0.045	0.048	0.059	0.010	0.031	0.027	0.080	0.041	0.030	0.028	0.031	0.052
		1	0.062	0.052	0.043	0.049	0.063	0.005	0.010	0.025	0.096	0.041	0.025	0.025	0.026	0.027
0	0.5	1.75	0.042	0.047	0.044	0.046	0.048	0.031	0.056	0.042	0.052	0.049	0.044	0.044	0.046	0.052
		1.5	0.054	0.052	0.049	0.051	0.057	0.019	0.063	0.037	0.066	0.050	0.043	0.041	0.044	0.058
		1.25	0.057	0.053	0.052	0.055	0.063	0.007	0.032	0.029	0.078	0.047	0.035	0.034	0.034	0.050
		1	0.061	0.047	0.051	0.054	0.068	0.003	0.010	0.028	0.094	0.048	0.032	0.032	0.023	0.028
0	-0.5	1.75	0.052	0.056	0.050	0.053	0.050	0.052	0.088	0.056	0.053	0.056	0.053	0.052	0.071	0.076
		1.5	0.060	0.062	0.054	0.057	0.061	0.037	0.093	0.048	0.073	0.055	0.047	0.048	0.070	0.093
		1.25	0.064	0.060	0.055	0.061	0.066	0.022	0.058	0.039	0.083	0.052	0.040	0.040	0.067	0.097
		1	0.069	0.057	0.058	0.060	0.073	0.014	0.025	0.038	0.099	0.053	0.039	0.037	0.074	0.085