

Sign tests for long-memory time series



Miguel A. Delgado*, Carlos Velasco

*Departamento de Economía, Universidad Carlos III de Madrid, Madrid 126-128,
Getafe, 28903 Madrid, Spain*

Abstract

This paper proposes sign-based tests for simple and composite hypotheses on the long-memory parameter of a time series process. The tests allow for nonstationary hypothesis, such as unit root, as well as for stationary hypotheses, such as weak dependence or no integration. The proposed generalized Lagrange multiplier sign tests for simple hypotheses on the long-memory parameter are exact and locally optimal among those in their class. We also propose tests for composite hypotheses on the parameters of ARFIMA(p, d, q) processes. The resulting tests statistics have a standard normal limiting distribution under the null hypothesis.

JEL classification: C12; C22

Keywords: Exact tests; Nonparametric tests; Infinite variance; Long-range dependence; Fractional processes; Nonstationarity; Optimal tests

1. Introduction

Statistical inferences based on signs are robust to a significant amount of gross errors and still valid in the presence of observations with infinite variance. At the same time, the resulting inferences have high relative asymptotic efficiency. However, the most attractive feature of this methodology is that the resulting generalized score

*Corresponding author. Tel.: +34-91-6249804; fax: +34-91-6249849.
E-mail address: miguelangel.delgado@ucm3.es (M.A. Delgado).

or Lagrange multiplier tests are exact, and also locally optimal among those in their class. This methodological approach has been systematically set out by Boldin and coauthors, and the state of the arts is lucidly presented in the monograph by Boldin et al. (1997).

Sign-based inference on the parameters of an autoregressive process have been developed by Boldin (1995, 1996) and Boldin and Tyurin (1994), the ARMA case has been studied by Boldin and Stute (2000). Chapters 6 and 7 of the mentioned monograph by Boldin et al. (1997) discuss sign-based inference on the parameter of the AR(1) process in stationary, instable and explosive situations.

This methodology is especially appealing when applied to time series exhibiting infinite variance, which are likely in many fields, such as finance, economics, telecommunications and hydrology. However, it may be difficult to distinguish between a long-memory behavior and a short-memory process with a large variance. Some time series are expected to exhibit the Noah effect (high variability or infinite variance) and the Joseph effect (i.e. self-similarity or long-run dependence), like Ethernet traffic data (see Willinger et al., 1995, 1997; and the references in Beran, 1994, Section 11.1). Similar problems arise when testing for conditional volatility, which often requires fourth marginal moments of observations, that may not exist for many data sets, but which typically display long-range properties (see e.g. Robinson, 1991). Furthermore, outlier innovations in trending integrated time series may lead to spurious structural changes that explain the observed persistence in many situations (Parke, 1999).

Fractionally integrated ARMA (ARFIMA) models with symmetric stable innovations, or in the domain of attraction of a stable law, have been studied by Kokoszka and Taqqu (1995, 1996a, b) and an algorithm for simulating stable ARFIMA process is provided by Kokoszka and Taqqu (1999). These authors studied the asymptotic properties of periodogram-based estimates of parameters. Also robust estimates of regression models with stationary long-memory errors have received some attention (see e.g. Beran, 1991; Koul and Surgailis, 1997; and the references therein), but there has been no attempt yet of developing robust inference for the long-memory parameter of stationary or nonstationary fractional time series. On the other hand, for the particular case of the instable AR(1) process, there have been several proposals for using robust methods when testing the unit root hypothesis in the direction of autoregressive alternatives, see e.g. Boldin et al. (1997), Breitung and Gourieroux (1997), Campbell and Dufour (1995), Hasan and Koenker (1997), Phillips (1995) and Wright (2000).

In this paper, we propose tests on the long-memory parameter using signs of residuals. These tests are relatively more efficient than any other test based on signs. We will use a different definition of the ARFIMA model which allows us to consider simultaneously stationary and nonstationary processes, so any value of the long-memory parameter is allowed, including also some overdifferencing hypothesis. The resulting tests are still valid in the presence of infinite variances, and exact tests on simple hypothesis are provided. We also consider testing composite hypothesis on the long-memory parameter for general ARFIMA(p, d, q) models, which yields, as a particular case, tests for the composite unit root hypothesis in the direction of

long-memory alternatives. The basic result is a uniform approximation for the score statistic based on the sign of residuals, extending to long-range-dependent data results of Boldin (1996) and Boldin and Stute (2000) for weakly dependent processes.

The rest of the paper is organized as follows. An exact test for simple hypothesis is presented in Section 2, where it is also discussed the asymptotic behavior of the test and its power in the direction of Pitman's alternatives. Section 3 is devoted to tests for simple linear hypotheses on the parameter vector of a possibly nonstationary ARFIMA model. Section 4 proposes a test on the composite hypothesis of the long-memory parameter. The results of a Monte Carlo study are placed in Section 5. Mathematical proofs can be found in Section 6, and they are based in some instrumental results in a Lemmata at the end of the paper.

2. Testing simple hypotheses

Consider the model

$$(1 - L)^d u_t = \varepsilon_t, \quad t = 1, 2, \dots, \quad u_s = 0 \text{ all } s \leq 0, \quad (1)$$

where d is an unknown parameter, and we assume, at the moment, that $\{\varepsilon_t\}_{t \geq 1}$ are iid with a (nondegenerate) unknown distribution function G , L denotes the lag operator, i.e. $Lu_t = u_{t-1}$, and the fractional difference operator is given by the formal binomial expansion

$$(1 - L)^d = \sum_{j=0}^{\infty} \psi_j(d) L^j, \quad \psi_j(d) := \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}, \quad j = 0, 1, \dots,$$

where $\Gamma(z) = \int_0^{\infty} x^{z-1} e^{-x} dx$ so, using Stirling's formula, the coefficients $\psi_j(d)$ behave as $\Gamma(-d)^{-1} j^{-d-1}$ as $j \rightarrow \infty$. Thus $\psi_0(d) = 1$ and $\psi_j(d) = \psi_{j-1}(d)(j-d-1)/j, j \geq 1$. When d is a positive integer, only the first $d+1$ terms are nonzero and we obtain the usual definition of the d th difference operator.

Model (1) can be interpreted as a truncated AR(∞) process, and it is said that $\{u_t\}$ is fractionally integrated of order d or that it is an ARFIMA(0, d , 0) process. The parameter d determines the long-range properties of u_t , while in Sections 3 and 4 we consider that the short run behavior is further determined by some ARMA parameters. The process $\{u_t\}$ has also a moving average (MA) representation. Inverting the fractional difference operator in (1) we obtain

$$u_t = (1 - L)^{-d} \varepsilon_t = \sum_{j=0}^{t-1} \psi_j(-d) \varepsilon_{t-j}, \quad t = 1, 2, \dots, \quad \varepsilon_s = 0, \quad s \leq 0. \quad (2)$$

The weights $\psi_j(-d), j = 0, 1, \dots$, are not absolute summable for any $d > 0$, which is a major source of technical difficulties, compared to the exponentially decreasing coefficients in the MA(∞) representation of ARMA processes. Asymptotically equivalent operators to fractional differencing, cf. Robinson and Marinucci (2001), could be considered within the methodology developed in this paper, as long as the coefficients of their MA representation are monotone for values of d in an interval of

zero, as are the $\psi_j(-d), j = 1, 2, \dots$, see Lemma 1(a). Note that the process u_t is not stationary for any value $d \neq 0$, but is increasingly instable as d grows. This set up allows us to consider series for any value of d , not limited to the usual stationary and invertible range, $(-0.5, 0.5)$, as when considering an infinite series in (2) with ε_t of finite variance, or to some open interval inside $(0, 0.5)$ as, for instance, in Kokoszka and Taqqu (1996b).

Given observations $\mathcal{U}_n := \{u_t, t = 1, \dots, n\}$, we are interested in testing the hypothesis

$$H_0 : d = d_0,$$

in the direction of one- and two-sided alternatives under the following minimal assumptions:

- A1. $E(\varepsilon_1) = 0$ and $E|\varepsilon_1|^{1+\delta} < \infty$, some $\delta > 0$.
- A2. $G(0) = \frac{1}{2}$.
- A3. There exists a density $g(x) = G'(x)$ such that g is Hölder continuous of order $\gamma \in (0, 1]$ at $x = 0$, $g(0) > 0$, and $\sup_x g(x) < \infty$.

Assumption A1 is fairly weak, and A2 is often required in robust inference. It can be easily tested, noticing that under A2, $n^{1/2}(G_n(0) - 1/2)$ converges in distribution to a normal with zero mean and variance $1/4$, where G_n is the empirical distribution of $\{u_t\}$. The composite hypothesis of equality between mean and median can be tested in the lines of the symmetry test proposed by Bai and Ng (2002). Assumption A3 is frequently required for obtaining central limit theorems in robust procedures, e.g. in order to derive the asymptotic distribution of the least absolute deviation estimator.

We propose an exact test based on signs of innovations, computed under the restriction on the null. These innovations have the form

$$\varepsilon_t(d_0) := (1 - L)^{d_0} u_t = (1 - L)^{d_0 - d} \varepsilon_t, \quad t \geq 1,$$

so

$$\varepsilon_t(d_0) = \sum_{j=0}^{t-1} \psi_j(d_0) u_{t-j} = \sum_{j=0}^{t-1} \psi_j(d_0 - d) \varepsilon_{t-j}, \quad t = 1, 2, \dots$$

Therefore, under H_0 , $\varepsilon_t(d_0) \equiv \varepsilon_t(d) \equiv \varepsilon_t$. Consider the sign functions

$$S_t(\cdot) := \text{sign}(\varepsilon_t(\cdot)) = 2 \cdot 1(\varepsilon_t(\cdot) > 0) - 1,$$

where $1(A)$ is the indicator function of the event A . Since $\{S_t(d)\}$ are distributed as a uniform random variable taking values $+1$ and -1 , and under the alternative hypothesis $\{S_t(d_0)\}$ are possibly autocorrelated with mean different than zero, $\{S_t(d_0)\}$ form a basis for constructing test statistics. These tests are nonparametric, in the sense that they are locally most powerful (LMP) within a certain class, without any knowledge of the G functional form. The test statistic is

$$T_n(d_0) = n \sum_{j=1}^{n-1} \frac{1}{j} \tilde{\gamma}_{nj}(d_0), \quad (3)$$

where

$$\tilde{\gamma}_{nj}(\cdot) = \frac{1}{n} \sum_{t=j+1}^n S_t(\cdot) S_{t-j}(\cdot)$$

is the j th order autocorrelation estimate of $\{S_t(\cdot)\}$. Notice that $\{S_t(d)\}$ have mean zero and variance one under H_0 , and $\tilde{\gamma}_{nj}(d)$ is a consistent estimator of the j th order autocorrelation of $S_t(d)$.

We show that the resulting test is LMP, in the sense that the curvature of the power function (i.e. its mean rate of increase) is maximum, amongst all the tests based on signs, in a neighborhood of d_0 ; also we provide the distribution of $T_n(d_0)$ for each n . In fact, $T_n(d_0)$ is a generalized Lagrange multiplier (GLM) test statistic, based on the same principle as the tests proposed by Robinson (1994) and Tanaka (1999), which are LMP under Gaussianity. The test statistic proposed by Tanaka (1999) is related to ours; it has form (3), substituting $\tilde{\gamma}_{nj}(\cdot)$ by

$$\hat{\gamma}_{nj}(\cdot) = \frac{\sum_{t=j+1}^n \varepsilon_t(\cdot) \varepsilon_{t-j}(\cdot)}{\sum_{t=1}^n \varepsilon_t(\cdot)^2},$$

the j th order autocorrelation estimate of $\{\varepsilon_t(\cdot)\}$. Robinson's (1994) test has also the form in (3), using autocorrelation estimates in the frequency domain. Robinson's (1994) and Tanaka's (1999) tests are LMP when innovations are Gaussian, but they can be arbitrarily inefficient under departures from Gaussianity, and their properties have not been investigated when innovations variances are infinite.

Also, we provide the asymptotic distribution of the test statistic suitably standardized, and prove the consistency of the test. In particular, we show that $n^{-1/2} T_n(d_0)$ converges in distribution under the null hypothesis to a normal random variable centered at zero and with variance $\pi^2/6$, as Robinson's (1994) and Tanaka's (1999) tests. Finally, we provide an uniform asymptotic first-order expansion, which forms a basis for studying the behavior of the test under local alternatives, as well as for studying its asymptotic relative efficiency (ARE).

2.1. Local optimality

Given observations \mathcal{U}_n , consider the vector of signs

$$\mathbf{S}^n(\mathcal{U}_n, d_0) = (S_1(d_0), \dots, S_n(d_0))'.$$

The possible observed values of the random vector $\mathbf{S}^n(\mathcal{U}_n, d_0)$ are vectors of the form

$$\mathbf{s}^n = (s_1, \dots, s_n)', \quad (4)$$

where s_i can take the values $+1$ and -1 . A test can be based on a critical region \mathcal{Q} , which is a set of $n \times 1$ vectors of form (4). That is, H_0 is rejected when the following event occurs:

$$\{\mathcal{U}_n : \mathbf{S}^n(\mathcal{U}_n, d_0) \in \mathcal{Q}\}.$$

The power of the test is a function of d , of the form,

$$\Pr\{\mathbf{S}^n(\mathcal{U}_n, d_0) \in \mathcal{Q} \mid d\}.$$

First, fix an arbitrary significance level α , such that, under H_0 ,

$$\Pr\{\mathbf{S}^n(\mathcal{U}_n, d_0) \in \mathcal{Q} \mid d = d_0\} = \alpha. \quad (5)$$

When A2 holds, $\Pr\{\mathbf{S}^n(\mathcal{U}_n, d_0) = \mathbf{s}^n\} = 2^{-n}$ under H_0 , where \mathbf{s}^n is a vector of form (4), only multiples of 2^{-n} can be taken for the level α in (5), and the critical region \mathcal{Q} contains exactly $K = 2^n \alpha$ points of form (4). The LMP test is the one exceeding the power function of any other test in a vicinity of d_0 . Consider for definiteness the alternative $H_1: d > d_0$. Amongst the tests with a given significance level, we are interested in the one whose power function has the highest rate of increase in a vicinity of d_0 . In other words, we look for the test with critical region \mathcal{Q} , which maximizes

$$\frac{\partial \Pr\{\mathbf{S}^n(\mathcal{U}_n, d_0) \in \mathcal{Q} \mid d = d_0\}}{\partial d} \quad (6)$$

subject to (5). Since

$$\frac{\partial \Pr\{\mathbf{S}^n(\mathcal{U}_n, d_0) \in \mathcal{Q} \mid d = d_0\}}{\partial d} = \sum_{\mathbf{s}^n \in \mathcal{Q}} \frac{\partial \Pr\{\mathbf{S}^n(\mathcal{U}_n, d_0) = \mathbf{s}^n \mid d = d_0\}}{\partial d},$$

we should include into \mathcal{Q} the vectors \mathbf{s}^n for which (6) is as large as possible. That is, the critical region will be of the form,

$$\mathcal{Q} = \left\{ \mathbf{s} : \frac{\partial \Pr\{\mathbf{S}^n(\mathcal{U}_n, d_0) = \mathbf{s}^n \mid d = d_0\}}{\partial d} \geq \text{const} \right\},$$

where the constant is chosen such that (5) is satisfied. The value of the constant, that is, the critical value can be obtained by Monte Carlo as accurately as desired.

The next proposition provides an expression for $\Pr\{\mathbf{S}^n(\mathcal{U}_n, d_0) = \mathbf{s}^n\}$.

Proposition 1. *If A1–A3 holds, for each $n \geq 2$,*

$$\Pr\{\mathbf{S}^n(\mathcal{U}_n, d_0) = \mathbf{s}^n\} = \left(\frac{1}{2}\right)^n \left(1 + 2g(0)E|\varepsilon_1| \sum_{j=1}^{n-1} \psi_j(d_0 - d) \sum_{m=j+1}^n s_m s_{m-j} \right) + o(|d - d_0|) \text{ as } d \rightarrow d_0.$$

Therefore, under the same assumptions,

$$\begin{aligned} \frac{\partial \Pr\{\mathbf{S}^n(\mathcal{U}_n, d_0) = \mathbf{s}^n \mid d = d_0\}}{\partial d} &= \left(\frac{1}{2}\right)^{n-1} g(0)E|\varepsilon_1| \sum_{j=1}^{n-1} \frac{1}{j} \sum_{m=j+1}^n s_m s_{m-j} \\ &\propto \sum_{j=1}^{n-1} \frac{1}{j} \sum_{m=j+1}^n s_m s_{m-j}, \end{aligned}$$

noticing that $\partial\psi_j(d_0 - d)/\partial d|_{d=d_0} = j^{-1}$, which shows that the test based on $T_n(d_0)$ is LMP, as stated in next theorem, which also states that $T_n(d_0)$ enjoys a pivotal distribution under the null hypothesis, and therefore, the quantiles of the test statistic,

$$q_{\theta n} = \frac{1}{2} \{(\inf\{t: \Pr(T_n \leq t | d = d_0) \geq \theta\} + \sup\{t: \Pr(t: T_n \leq t | d = d_0) \leq \theta\})\},$$

can be computed for each n .

Theorem 1. *Under A1–A3, the test with critical region, for each $n \geq 2$,*

$$\mathcal{Q} = \{\mathcal{U}_n: T_n(d_0) > q_{1-\alpha n}\}$$

is a locally optimal sign test at the α -level of significance (i.e. a locally unbiased sign test of maximal mean curvature) for testing H_0 in the direction of $H_1: d > d_0$. The test with critical region

$$\mathcal{Q} = \{\mathcal{U}_n: T_n(d_0) < q_{\alpha n}\}$$

is a locally optimal sign test at the α -level of significance for testing H_0 in the direction of $H_1: d < d_0$. The test with critical region

$$\mathcal{Q} = \{\mathcal{U}_n: T_n(d_0) < q_{\alpha n} \text{ and } T_n(d_0) > q_{1-\alpha n}\}$$

is a locally optimal sign test at the 2α -level of significance for testing H_0 in the direction of $H_1: d \neq d_0$. Under H_0 , the test statistic $T_n(d_0)$ is distribution free for each n .

The $T_n(d)$ quantiles can be calculated exactly by enumeration, trying all 2^n equiprobable combinations of signs, but it is computationally very expensive. Alternatively, they can also be computed very accurately by Monte Carlo (see Table 1 below).

2.2. Asymptotic test

The asymptotic distribution as $n \rightarrow \infty$ of the test statistic T_n , suitably standardized, is an immediate consequence of Anderson (1971) Theorem 7.7.5.

Table 1
Exact critical values of $T_n(d)/\sqrt{n}$

	0.5%	1%	2.5%	5%	10%	90%	95%	97.5%	99%	99.5%
$n = 50$	-1.90	-1.77	-1.55	-1.36	-1.13	1.26	1.76	2.23	2.81	3.27
$n = 100$	-2.06	-1.90	-1.90	-1.44	-1.17	1.27	1.74	2.17	2.70	3.13
$n = 200$	-2.19	-2.01	-1.74	-1.49	-1.20	1.28	1.73	2.13	2.63	2.99
$n = 400$	-2.27	-2.09	-1.80	-1.54	-1.23	1.29	1.72	2.11	2.57	2.91
$n = 1000$	-2.39	-2.16	-1.86	-1.58	-1.26	1.29	1.68	2.04	2.48	2.77
$n = 2000$	-2.42	-2.23	-1.88	-1.60	-1.27	1.29	1.68	2.02	2.42	2.70
$n = \infty$	-2.58	-2.33	-1.96	-1.64	-1.28	1.28	1.64	1.96	2.33	2.58

Theorem 2. *If A1–A3 hold, as $n \rightarrow \infty$,*

$$\frac{1}{\sqrt{n}} T_n(d) \xrightarrow{d} N\left(0, \frac{\pi^2}{6}\right).$$

Then, asymptotic critical values based on the standard normal distribution can be used for large sample sizes.

2.3. Power under local alternatives

Consider Pitman's alternatives of the form,

$$H_{1n}(a): d = d_n := d_0 + \frac{a}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right),$$

for some constant a . The next proposition is essential for providing a Central Limit theorem (CLT) for $n^{-1/2}T_n(d_0)$ under $H_{1n}(a)$.

Proposition 2. *If A1–A3 hold, under $H_{1n}(a)$, as $n \rightarrow \infty$,*

$$\sup_{|\theta| \leq Cn^\nu} \left| \frac{1}{\sqrt{n}} T_n(d_n + n^{-1/2}\theta) - \frac{1}{\sqrt{n}} T_n(d_n) - \lambda\theta \right| = o_p(1),$$

where θ is an arbitrary positive and finite constant, C is a positive constant, $\nu < \gamma^*/[2(1 + 2\gamma^*)]$, $\gamma^* = \min\{\delta, \gamma\}$ and

$$\lambda = -2g(0)E|\varepsilon_1|\pi^2/6.$$

Note that the drift value parameter λ does not depend on d_0 , unlike, for instance, when testing for AR or ARMA parameters (cf. Boldin, 1995). As an immediate consequence of the above proposition, the next theorem states that the test does not have trivial power in the direction of $H_{1n}(a)$ for $a \neq 0$, and shows that, asymptotically, the power of the test does not depend on d_0 .

Theorem 3. *If A1–A3 hold, under $H_{1n}(a)$, as $n \rightarrow \infty$,*

$$\frac{1}{\sqrt{n}} T_n(d_0) \xrightarrow{d} N\left(2ag(0)E|\varepsilon_1| \frac{\pi^2}{6}, \frac{\pi^2}{6}\right).$$

2.4. ARE

We now compare the ARE of our sign test with statistic $T_n(d_0)$ with respect to known tests, such as Tanaka (1999) or Robinson (1994). These latter tests are Gaussian score tests constructed in the time and frequency domain, respectively, and

share the same asymptotic distribution. Tanaka's test has the form,

$$L_n^{\text{Tan}}(d_0) = n \sum_{j=1}^{n-1} \frac{1}{j} \hat{\gamma}_{nj}(d_0),$$

just replacing in $T_n(d_0)$ the sign sample autocorrelations of the residuals $\varepsilon_t(d_0)$ under H_0 by the usual autocorrelations. On the other hand, Robinson's tests are designed in the frequency domain. A version of them for testing H_0 is

$$L_n^{\text{Rob1}}(d_0) := \frac{2\pi}{\hat{\sigma}_\varepsilon^2(d_0)n^{1/2}} \sum_{j=1}^{n-1} \log |2 \sin(\lambda_j/2)| I_{n\varepsilon}(\lambda_j, d_0),$$

where $I_{n\varepsilon}(\lambda_j, d_0) = (2\pi n)^{-1} |\sum_{t=1}^n \varepsilon_t(d_0) \exp(i\lambda_j t)|^2$ is the periodogram of the residuals $\{\varepsilon_t(d_0)\}$ and $\lambda_j = 2\pi j/n$ are the Fourier frequencies, with

$$\hat{\sigma}_\varepsilon^2(d_0) := \frac{1}{n} \sum_{t=1}^n \varepsilon_t(d_0)^2 = \frac{2\pi}{n} \sum_{j=1}^n I_{n\varepsilon}(\lambda_j, d_0).$$

A different version which does not use fractional differentiation is

$$L_n^{\text{Rob2}}(d_0) := \frac{2\pi}{\hat{\sigma}_\varepsilon^2(d_0)n^{1/2}} \sum_{j=1}^{n-1} \log |2 \sin(\lambda_j/2)| \frac{I_{nu}(\lambda_j)}{|2 \sin(\lambda_j/2)|^{-2d_0}},$$

where $I_{nu}(\lambda_j) = (2\pi n)^{-1} |\sum_{t=1}^n u_t \exp(i\lambda_j t)|^2$ is the periodogram of the data \mathcal{U}_n and

$$\hat{\sigma}_\varepsilon^2(d_0) := \frac{1}{n} \sum_{j=1}^{n-1} \frac{2\pi I_{nu}(\lambda_j)}{|2 \sin(\lambda_j/2)|^{-2d_0}}.$$

These authors obtained that under $H_{1n}(a)$, both $L_n^{\text{Tan}}(d_0)$ and $L_n^{\text{Rob1}}(d_0)$ converge in distribution to a normal random variable centered at $a\pi^2/6$ and with variance $\pi^2/6$. A similar result can be expected for $L_n^{\text{Rob2}}(d_0)$ under (1) and some restrictions on d (see Velasco and Robinson, 2000).

Therefore the ARE of the sign test based on $T_n(d_0)$ with respect to the tests based on scores $L_n(d)$ equals

$$e_{T,L} = (2g(0)E|\varepsilon_1|)^2.$$

This ARE has the same expression as that of sign tests for finite AR(p) models against least squares tests and does not depend on d_0 nor on the scale parameter of $G(x)$, see the discussion in Boldin et al. (1997, p. 157).

3. Testing multiple hypotheses

We consider now the problem of hypothesis testing on parameters of general ARFIMA models extending Boldin and Stute's (2000) treatment for ARMA processes. Consider the ARFIMA(p, d, q) model

$$A_d(L)(1-L)^d u_t = B_b(L)\varepsilon_t, \quad (7)$$

where the ε_t satisfy the same conditions as in Section 2, $\varepsilon_s = u_s = 0$, all $s \leq 0$, $A_a(L) := 1 - a_1L - \dots - a_pL^p$ and $B_b(L) := 1 + b_1L + \dots + b_qL^q$, with $a = (a_1, \dots, a_p)'$, $b = (b_1, \dots, b_q)'$ and $\theta = (d, a', b')'$.

The nonparametric sign tests of Section 2 are extended in this section for multiple hypotheses $H_0 : \theta = \theta_0$. These tests are based on the signs of the residuals computed under the restriction on the null, $\{S_t(\theta_0)\}$, where, for $\theta_0 = (d_0, a'_0, b'_0)'$,

$$\varepsilon_t(\theta_0) = A_{a_0}(L)(1 - L)^{d_0}u_t + B_{b_0}^*(L)\varepsilon_t(\theta_0), \quad t \geq 1,$$

with $B^*(z) = 1 - B(z)$ and $\varepsilon_t = \varepsilon_t(\theta) = 0$ all $t \leq 0$. We assume the following conditions on the autoregressive polynomials.

- A4. The polynomials $A_a(z)$ and $B_b(z)$ have roots outside the unit circle.
- A5. The polynomials $A_a(z)$ and $B_b(z)$ have no roots in common.

As in Section 2, the test statistics are based on $\tilde{\gamma}_{nj}(\theta_0)$, the j th order autocorrelation estimate of $\{S_t(\theta_0)\}$, with $S_t(\cdot) = \text{sign}(\varepsilon_t(\cdot))$. We follow here the same strategy.

3.1. Local optimality

Considering the vector of signs

$$\mathbf{S}^n(\mathcal{U}_n, \theta_0) = (S_1(\theta_0), \dots, S_n(\theta_0))'$$

our test is based on a critical region \mathcal{Q} , which is a set of $n \times 1$ vectors of form (4), such that H_0 is rejected when the $\{\mathcal{U}_n : \mathbf{S}^n(\mathcal{U}_n, \theta_0) \in \mathcal{Q}\}$ occurs. The power of the test is then $\Pr\{\mathbf{S}^n(\mathcal{U}_n, \theta_0) \in \mathcal{Q} | \theta\}$. Now fix an arbitrary significance level α , such that, under H_0 ,

$$\Pr\{\mathbf{S}^n(\mathcal{U}_n, \theta_0) \in \mathcal{Q} | \theta = \theta_0\} = \alpha. \quad (8)$$

Since when A2 holds, $\Pr\{\mathbf{S}^n(\mathcal{U}_n, \theta_0) = \mathbf{s}^n\} = 2^{-n}$ under H_0 and the critical region \mathcal{Q} contains exactly $K = 2^n \alpha$ points of form (4). The LMP test is the one exceeding the power function of any other test in a vicinity of θ_0 .

For the alternative $H_1 : d > d_0$ the discussion follows as in Section 2. We look for the test with critical region \mathcal{Q} , which maximizes

$$\frac{\partial \Pr\{\mathbf{S}^n(\mathcal{U}_n, \theta_0) \in \mathcal{Q} | \theta = \theta_0\}}{\partial d}$$

subject to (5). Since

$$\frac{\partial \Pr\{\mathbf{S}^n(\mathcal{U}_n, \theta_0) \in \mathcal{Q} | \theta = \theta_0\}}{\partial d} = \sum_{\mathbf{s}^n \in \mathcal{Q}} \frac{\partial \Pr\{\mathbf{S}^n(\mathcal{U}_n, \theta_0) = \mathbf{s}^n | \theta = \theta_0\}}{\partial d},$$

we should include into \mathcal{Q} the vectors \mathbf{s}^n for which (6) is as large as possible. That is, the critical region will be,

$$\mathcal{Q} = \left\{ \mathbf{S} : \frac{\partial \Pr\{\mathbf{S}^n(\mathcal{U}_n, \theta_0) = \mathbf{s}^n | \theta = \theta_0\}}{\partial d} \geq \text{const} \right\},$$

where the constant is chosen such that (8) is satisfied. The value of the constant, which is the critical value, can be obtained by Monte Carlo as accurately as desired. Similar arguments can be derived to justify optimal tests for alternatives $a_i > a_{i0}$, $i = 1, \dots, p$, and $b_k > b_{k0}$, $k = 1, \dots, q$.

Let the sequence $\{\alpha_t = \alpha_t(\mathbf{a})\}$ be defined by the linear recursion

$$\alpha_t = a_1 \alpha_{t-1} + \dots + a_p \alpha_{t-p}, \quad t = 1, 2, \dots$$

with $\alpha_t = 0$ for $t < 0$ and $\alpha_0 = 1$ and the sequence $\{\beta_t = \beta_t(\mathbf{b})\}$ be defined by

$$\beta_t = -b_1 \beta_{t-1} - \dots - b_q \beta_{t-q}, \quad t = 1, 2, \dots$$

with $\beta_t = 0$ for $t < 0$ and $\beta_0 = 1$. The next proposition provides expressions for $\Pr\{\mathbf{S}^n(\mathcal{U}_n, \theta_0) = \mathbf{s}^n\}$.

Proposition 3. *If A1–A4 holds, as $d \rightarrow d_0$, for each $n \geq 2$,*

$$\begin{aligned} & \Pr\{\mathbf{S}^n(\mathcal{U}_n, \theta_0) = \mathbf{s}^n\} \\ &= \left(\frac{1}{2}\right)^n \left(1 + 2g(0)E|\varepsilon_1| \sum_{j=1}^{n-1} \psi_j(d_0 - d) \sum_{m=j+1}^n s_m s_{m-j}\right) \\ &+ o(|d - d_0|), \end{aligned}$$

as $a_i \rightarrow a_{i0}$, $i = 1, \dots, p$, $n \geq i + 1$,

$$\begin{aligned} & \Pr\{\mathbf{S}^n(\mathcal{U}_n, \theta_0) = \mathbf{s}^n\} \\ &= \left(\frac{1}{2}\right)^n \left(1 + 2g(0)E|\varepsilon_1| \sum_{j=1}^{n-i-1} \alpha_j(\theta_0) \sum_{m=i+j+1}^n s_m s_{m-j-i}(a_i - a_{i0})\right) \\ &+ o(|a_i - a_{i0}|) \end{aligned}$$

and, as $b_k \rightarrow b_{k0}$, $k = 1, \dots, q$, $n \geq k + 1$,

$$\begin{aligned} & \Pr\{\mathbf{S}^n(\mathcal{U}_n, \theta_0) = \mathbf{s}^n\} \\ &= \left(\frac{1}{2}\right)^n \left(1 + 2g(0)E|\varepsilon_1| \sum_{j=1}^{n-k-1} \beta_j(\theta_0) \sum_{m=k+j+1}^n s_m s_{m-j-k}(b_k - b_{k0})\right) \\ &+ o(|b_k - b_{k0}|). \end{aligned}$$

Therefore, under the same assumptions of the proposition,

$$\frac{\partial \Pr\{\mathbf{S}^n(\mathcal{U}_n, d_0) = \mathbf{s}^n \mid \theta = \theta_0\}}{\partial d} \propto \sum_{j=1}^{n-1} \frac{1}{j} \sum_{m=j+1}^n s_m s_{m-j},$$

which shows that the test based on

$$T_n^{(0)}(\theta_0) := n \sum_{j=1}^{n-1} \frac{1}{j} \tilde{\gamma}_{nj}(\theta_0),$$

is LMP for $H_1 : d > d_0$ as stated in next Theorem, with a pivotal distribution. Similar to Boldin and Stute (2000), we set

$$T_n^{(i)}(\theta_0) := n \sum_{j=1}^{n-i-1} \alpha_j(\theta_0) \tilde{\gamma}_{n,i+j}(\theta_0), \quad T_n^{(p+k)}(\theta_0) := n \sum_{j=1}^{n-k-1} \beta_j(\theta_0) \tilde{\gamma}_{n,k+j}(\theta_0).$$

Theorem 4. *Under A1–A4, the test with critical region, for each $n \geq 2$,*

$$\mathcal{Q} = \{\mathcal{U}_n : T_n^{(0)}(\theta_0) > q_{1-\alpha n}\}$$

is a locally optimal sign test at the α -level of significance for testing H_0 in the direction of $H_1 : d > d_0$ among the tests based on $S^n(\theta_0)$.

The tests with critical region

$$\mathcal{Q} = \{\mathcal{U}_n : T_n^{(i)}(\theta_0) > q_{1-\alpha n}\}, \quad i = 1, \dots, p,$$

are locally optimal sign test at the α -level of significance for testing H_0 in the direction of $H_1 : a_i > a_{i0}$.

The test with critical region

$$\mathcal{Q} = \{\mathcal{U}_n : T_n^{(p+k)}(\theta_0) > q_{1-\alpha n}\}, \quad k = 1, \dots, q,$$

are locally optimal sign test at the α -level of significance for testing H_0 in the direction of $H_1 : b_k > b_{k0}$. Under H_0 , the test statistics $T_n^{(i)}(\theta_0)$ are distribution free.

Reversing the inequalities, similar results hold for sign tests against $H_1 : \theta_j < \theta_{j0}$, $j = 0, \dots, 1 + p + q$. The $T_n^{(i)}(\theta)$ quantiles can be calculated by enumeration or by Monte Carlo as those of $T_n(d)$.

3.2. Asymptotic test

The asymptotic distribution of the (suitably standardized) sign score statistic

$$l_n(\theta) = (T_n^{(0)}(\theta), T_n^{(1)}(\theta), \dots, T_n^{(p+q)}(\theta))', \quad n \geq \max\{p, q\} + 1,$$

is given in the next theorem extending the results of Theorem 2.

Theorem 5. *If A1–A4 hold, as $n \rightarrow \infty$,*

$$\frac{1}{\sqrt{n}} l_n(\theta) \xrightarrow{d} N(0, \mathbf{A}(\theta)),$$

where $\mathbf{A}(\theta)$ is a matrix with components,

$$\begin{aligned}
A_{11}(\theta) &= \sum_{t=1}^{\infty} \frac{1}{t^2} = \frac{\pi^2}{6}, \\
A_{1,1+i}(\theta) &= \sum_{t=1}^{\infty} \frac{\alpha_{t+i-1}}{t}, \quad i = 1, \dots, p, \\
A_{1,1+p+k}(\theta) &= \sum_{t=1}^{\infty} \frac{\beta_{t+k-1}}{t}, \quad k = 1, \dots, q, \\
A_{1+i,1+j}(\theta) &= \sum_{t=0}^{\infty} \alpha_t \alpha_{t+|i-j|}, \quad i, j = 1, \dots, p, \\
A_{1+p+k,1+p+j}(\theta) &= \sum_{t=0}^{\infty} \beta_t \beta_{t+|k-j|}, \quad k, j = 1, \dots, q, \\
A_{1+i,1+p+k}(\theta) &= 1(i \geq k) \sum_{t=0}^{\infty} \beta_t \alpha_{t+i-k} + 1(i < k) \sum_{t=0}^{\infty} \alpha_t \beta_{t+k-i}, \\
& \quad i = 1, \dots, p, \quad k = 1, \dots, q.
\end{aligned}$$

Then, asymptotic critical values based on the standard normal distribution can be used for large sample sizes.

For testing the null hypothesis $H_0: \theta = \theta_0$ versus $H_1: \theta \neq \theta_0$ we can employ the quadratic form

$$K_n(\theta_0) = l_n(\theta_0)' \mathbf{A}(\theta_0)^{-1} l_n(\theta_0)$$

if additionally A5 holds, so if the conditions of Theorem 5 and A5 are satisfied, under H_0 ,

$$\frac{1}{n} K_n(\theta_0) \rightarrow_d \chi_{p+q+1}^2.$$

3.3. Power under local alternatives

Consider Pitman's alternatives of the form,

$$H_{1n}(\mathbf{r}): \theta = \theta_n := \theta_0 + \frac{\mathbf{r}}{\sqrt{n}} + o(n^{-1/2}),$$

for some constant $(p+q+1) \times 1$ vector \mathbf{r} . Next proposition is essential for providing a CLT for $n^{-1/2} l_n(\theta_0)$ under $H_{1n}(\mathbf{r})$.

Proposition 4. *If A1–A4 hold, under $H_{1n}(\mathbf{r})$, as $n \rightarrow \infty$,*

$$\sup_{|\rho| \leq Cn^\gamma} |n^{-1/2} l_n(\theta_n + n^{-1/2} \rho) - n^{-1/2} l_n(\theta_n) - \lambda(\theta_0) \rho| = o_p(1),$$

where $\boldsymbol{\rho}$ is an arbitrary positive and finite constant vector, C is a positive constant, $v < \gamma^*/[2(1 + 2\gamma^*)]$, $\gamma^* = \min\{\delta, \gamma\}$ and

$$\lambda(\theta) = -2g(0)E|\varepsilon_1|\mathbf{A}(\theta).$$

The following theorem states that the test does not have trivial power in the direction of $\mathbf{H}_{1n}(\mathbf{r})$ for $\mathbf{r} \neq \mathbf{0}$.

Theorem 6. *If A1–A4 hold, under $\mathbf{H}_{1n}(\mathbf{r})$, as $n \rightarrow \infty$,*

$$\frac{1}{\sqrt{n}} l_n(\theta_0) \xrightarrow{d} N(2g(0)E|\varepsilon_1|\mathbf{A}(\theta_0)\mathbf{r}, \mathbf{A}(\theta_0)).$$

If in addition A5 holds, then under $\mathbf{H}_{1n}(\mathbf{r})$,

$$\frac{1}{n} K_n(\theta_0) \xrightarrow{d} \chi_{p+q+1}^2(\Lambda(\theta_0))$$

where $\Lambda(\theta_0) := (2g(0)E|\varepsilon_1|)^2 \mathbf{r}' \mathbf{A}(\theta_0) \mathbf{r}$ is the noncentrality parameter of the chi-square distribution.

As for simple hypothesis, the power does not depend on d_0 asymptotically, though it does depend on short-memory ARMA parameters. Similar ARE comparisons of our sign tests based on $l_n(\theta_0)$ with respect to known tests, such as Tanaka (1999) or Robinson's (1994), can be discussed as in Section 2.

4. Testing composite hypotheses

Frequently we are interested in testing hypothesis about d ,

$$H'_0 : d = d_0$$

where $\theta^{(2)} := (a', b)'$ is a nuisance parameter, since d determines the main features of the long run dynamics of u_t , including whether the series is a weak dependent process ($d = 0$) or is a unit root process, ($d = 1$). Also one side tests can be set, such as whether the processes is trending nonstationary ($d = 0.5$) against stationarity ($d < 0.5$). For other composite hypotheses concerning ARMA parameters, see Boldin and Stute (2000).

The power of the signs tests for composite hypotheses will be studied under the sequence of alternatives

$$H'_{1n}(\mathbf{h}) : \theta = \theta_n := \theta_0 + \mathbf{h}n^{-1/2},$$

where $\theta_0 = (d_0, \theta^{(2)})'$ and $\mathbf{h} = (h^{(1)}, \mathbf{h}^{(2)})'$ is a fixed $(1 + p + q)$ -vector. Thus the alternatives $H'_{1n}(\mathbf{h})$ allow the nuisance parameters $\theta^{(2)}$ to vary in a range of order $O(n^{-1/2})$.

We can partition the matrix $\mathbf{A}(\theta)$ as

$$\mathbf{A}(\theta) = \begin{pmatrix} J & \mathbf{B}'(\theta) \\ \mathbf{B}(\theta) & \mathbf{M}(\theta) \end{pmatrix},$$

where $\mathbf{B}(\theta)$ is a column vector and $J = \sum_{t=1}^{\infty} t^{-2} = \pi^2/6$. Set

$$\mathbf{V}(\theta) = \begin{pmatrix} c^{-1/2}(\theta) & \mathbf{0}' \\ -c^{-1/2}(\theta)\mathbf{M}^{-1}(\theta)\mathbf{B}(\theta) & \mathbf{M}^{-1/2}(\theta) \end{pmatrix},$$

where $c(\theta) = J - \mathbf{B}'(\theta)\mathbf{M}^{-1}(\theta)\mathbf{B}(\theta)$, so $\mathbf{V}(\theta)\mathbf{V}'(\theta) = \mathbf{A}^{-1}(\theta)$.

Let $\{\hat{\theta}_n\}$ be a sequence of estimates of θ which may depend on the conjectured value d_0 and on the observations \mathcal{U}_n , such that

$$n^{1/2}(\hat{\theta}_n - \theta_n) = O_p(1),$$

under $H'_{1n}(\mathbf{h})$. In particular $\hat{\theta}_n^{(1)}$ might be taken to be d_0 , so in this case $\hat{\theta}_{0n} = (d_0, \hat{\theta}_n^{(2)})'$, where $\hat{\theta}_n^{(2)}$ is a $n^{1/2}$ -consistent estimate of $\hat{\theta}^{(2)}$. There are many robust estimates for AR and ARMA models available in the literature which are $n^{1/2}$ -consistent for infinite variance innovations and that could be used for $\hat{\theta}_n^{(2)}$. See e.g. Yohai and Maronna (1977), Bloomfield and Steiger (1983), Pollard (1991), Davis and Resnik (1986), Mikosch et al. (1995) and Boldin et al. (1997). Note that $\theta^{(2)}$ does not include the long-memory parameter d which is given by H'_0 , so these robust ARMA estimates are applied to the prefiltered series $(1 - L)^{d_0}u_t$.

For testing H'_0 we will use the test statistic

$$L_n^\pi(\hat{\theta}_{n0}) := n^{-1}(\pi \circ \hat{\mathbf{V}}_n' l_n(\hat{\theta}_{n0}))^2$$

where π stands for the projection onto the subspace spanned by the first coordinate vector.

Theorem 7. *If A1–A5 hold, under $H'_{1n}(\mathbf{h})$, as $n \rightarrow \infty$,*

$$L_n^\pi(\hat{\theta}_{n0}) \xrightarrow{d} \chi_1^2(\lambda_\pi(\theta_0)),$$

where $\lambda_\pi(\theta_0) = (2g(0)E|\varepsilon_1|h^{(1)})^2 c(\theta_0)$.

In particular, under H'_0 ,

$$L_n^\pi(\hat{\theta}_{n0}) \xrightarrow{d} \chi_1^2.$$

One side tests, useful for testing against more stationarity or nonstationary alternatives, can be based on the fact that $n^{-1/2}\pi \circ \hat{\mathbf{V}}_n' l_n(\hat{\theta}_{n0}) \xrightarrow{d} N(0, 1)$ and interpreted as in Robinson (1994).

5. Monte Carlo

For sample sizes $n = 50, 100, 200, 400$ we consider Pitman's local fractional hypothesis

$$H_{1n}(-a): d = d_n := d_0 - \frac{a}{\sqrt{n}}$$

for $a = 0, 0.5, 1, 1.5, 2$. Data is generated as $u_t = (1 - L)^{-d_n} \varepsilon_t$. Setting $d_0 = 1$, we are able to compare sign and Tanaka's score tests directed against fractional alternatives ($d < 1$) with the Dickey-Fuller test for an unit root ($d = 1$) directed against weak stationarity ($d = 0$). We also investigate the finite sample properties of a robustified version of Dickey-Fuller test based on ranks introduced by Breitung and Gourieroux (1997). We consider four different distributions: $\varepsilon_t \sim \text{IID}, N(0,1), t_4, t_2$ and t_1 . The first one has all moments, but the t_m distribution only has $m - \varepsilon$ bounded absolute moments, $\varepsilon > 0$. The different tables report the percentage of rejections by one-sided tests of the null hypothesis $H_0: d = 1$ at the 5% of significance in 100,000 experiments. In power comparisons, Dickey-Fuller and Tanaka tests use asymptotic critical values, while our sign test and Breitung and Gourieroux (1997) test are exact, and critical values are obtained by Monte Carlo for each sample size.

Table 1 report the exact critical values of our scaled test statistic $T_n(d_0)/\sqrt{n}$ under the null $d = d_0$, for one- and two-sided hypothesis and several sample sizes.

The exact critical values are far away from the asymptotic ones, even for fairly large sample sizes. This point is confirmed in Table 2, where we report the empirical sizes for the different asymptotic tests, based on the standard normal approximation for the sign and Tanaka's tests.

Table 2
Empirical size of asymptotic tests at 5% of significance

		sign	sign*	Tanaka	Tanaka*	Dickey-Fuller
N(0, 1)	$n = 50$	1.67	3.66	1.52	2.96	5.36
	$n = 100$	2.58	4.77	2.54	4.35	5.23
	$n = 200$	3.23	5.36	3.20	5.12	4.96
	$n = 400$	3.75	5.60	3.81	5.61	5.01
t_4	$n = 50$	1.67	3.82	1.38	2.73	5.23
	$n = 100$	2.62	4.86	2.40	4.18	5.03
	$n = 200$	3.36	5.42	3.16	5.05	4.94
	$n = 400$	3.81	5.65	3.70	5.50	4.78
t_2	$n = 50$	1.63	3.58	1.22	2.28	4.52
	$n = 100$	2.60	4.80	1.99	3.38	4.38
	$n = 200$	3.31	5.37	2.63	4.04	4.23
	$n = 400$	3.82	5.63	2.98	4.27	4.37
t_1	$n = 50$	1.68	3.60	1.16	1.67	3.13
	$n = 100$	2.60	4.81	1.53	2.12	2.86
	$n = 200$	3.33	5.44	1.66	2.15	2.81
	$n = 400$	3.83	5.02	1.65	2.11	2.75

Despite the statistic T_n discussed in the text and the original Tanaka's test, we also report results, in the column sign*, for the asymptotic test based on the statistic,

$$\bar{T}_n(\cdot) = \frac{1}{n} \sum_{j=1}^n \tilde{\rho}_{nj}(\cdot),$$

where

$$\tilde{\rho}_{nj}(\cdot) = \frac{\sum_{t=j+1}^n [S_t(\cdot) - \bar{S}_n(\cdot)][S_{t-j}(\cdot) - \bar{S}_n(\cdot)]}{\sum_{t=1}^n [S_t(\cdot) - \bar{S}_n(\cdot)]^2}$$

is the j th sample autocorrelation of $S_t(\cdot)$. Using standard arguments, it is straightforward to show that $T_n(d_0) - \bar{T}_n(d_0) = o_p(n^{-1/2})$ under H_0 . Likewise, in the column Tanaka*, we report the proportion of rejections for the test statistic based on the sample autocorrelations of $\varepsilon_t(\cdot)$ instead of $\hat{\gamma}_{nj}$. The asymptotic size accuracy of sign and Tanaka tests is quite poor for the sample sizes considered, compared with the accuracy of the Dickey-Fuller test, though the modified versions (sign* and Tanaka*) are more accurate. As expected, Tanaka's test performs very poorly in the presence of errors with fat distribution tails.

Tables 3–6 report the empirical power of the tests under local alternatives. The sign test rejects more frequently the null hypothesis than Tanaka's test under small local departures ($a = 0.5, 1$) when the error distribution exhibit fat tails. As expected,

Table 3
Empirical power at 5% of significance of the exact sign test

		a			
		0.5	1	1.5	2
N(0, 1)	$n = 50$	8.75	13.99	21.01	29.57
	$n = 100$	8.60	13.60	20.66	29.33
	$n = 200$	8.39	13.36	20.08	28.56
	$n = 400$	8.33	13.20	19.75	28.01
t_4	$n = 50$	9.39	15.83	24.29	33.94
	$n = 100$	9.22	15.64	24.54	34.93
	$n = 200$	8.99	15.33	23.90	34.54
	$n = 400$	9.08	15.34	23.91	33.94
t_2	$n = 50$	10.97	19.20	29.24	39.90
	$n = 100$	10.79	19.62	30.85	42.94
	$n = 200$	10.82	19.75	31.39	44.73
	$n = 400$	10.70	19.68	31.79	45.52
t_1	$n = 50$	7.08	14.23	23.05	31.40
	$n = 100$	10.92	22.22	35.03	47.78
	$n = 200$	13.73	28.60	45.00	59.74
	$n = 400$	16.19	34.14	52.76	69.72

Table 4
Empirical power at 5% of significance of Tanaka's asymptotic test

		a			
		0.5	1	1.5	2
N(0, 1)	$n = 50$	4.13	9.91	10.10	36.31
	$n = 100$	6.14	13.73	26.15	43.39
	$n = 200$	7.75	15.92	29.35	45.88
	$n = 400$	8.72	17.24	30.54	46.70
t_4	$n = 50$	3.85	9.21	19.68	35.72
	$n = 100$	5.99	13.38	25.65	42.92
	$n = 200$	7.35	15.49	28.59	45.63
	$n = 400$	8.35	16.71	30.06	46.68
t_2	$n = 50$	3.23	7.66	16.52	33.13
	$n = 100$	4.78	10.54	21.91	41.09
	$n = 200$	5.87	12.50	24.94	43.94
	$n = 400$	6.74	13.70	26.50	45.30
t_1	$n = 50$	2.07	4.16	9.34	22.52
	$n = 100$	2.71	5.30	11.61	29.86
	$n = 200$	2.88	5.54	12.51	33.34
	$n = 400$	2.73	5.51	12.01	34.10

the empirical power does not change much when the innovations have at least one moment (i.e. Normal, t_4 and t_2) for sample sizes relatively large (n greater than 100). Interestingly, when the error does not have first moments, the empirical power of the test increases with n , which may indicate that the test is able to detect local alternatives converging to the null faster than $n^{-1/2}$ under these circumstances.

Tables 5 and 6 report the empirical power for the asymptotic Dickey-Fuller test, and the exact Dickey-Fuller test based on ranks. The asymptotic tests is less powerful than the robust exact test under departures from normality. However, the percentage of rejections converges to the nominal size for increasing n . That is, the simulations indicates that the two versions of the Dickey-Fuller test seem to have trivial power under local alternatives to the unit root null hypothesis, converging to the null at the rate $n^{-1/2}$.

6. Proofs

Proof of Proposition 1. Henceforth, $S^n(\mathcal{U}_n, d_0) := S^n(d_0)$. Apply Lemma 3 recursively with the initial condition in Lemma 2, as in (17), and applying Lemma 4, for

Table 5
Empirical power at 5% of significance of Dickey-Fuller exact test based on ranks

		a			
		0.5	1	1.5	2
N(0, 1)	$n = 50$	7.71	11.22	16.52	23.53
	$n = 100$	6.95	9.42	12.68	16.98
	$n = 200$	6.07	7.87	9.65	12.44
	$n = 400$	6.00	7.23	8.58	9.92
t_4	$n = 50$	7.49	11.29	16.43	23.38
	$n = 100$	6.86	9.24	12.70	16.81
	$n = 200$	8.40	10.33	12.85	15.56
	$n = 400$	5.86	7.04	8.27	10.04
t_2	$n = 50$	7.65	11.33	16.87	23.87
	$n = 100$	6.89	9.53	13.05	17.57
	$n = 200$	6.33	8.12	10.33	12.89
	$n = 400$	5.95	7.10	8.54	10.04
t_1	$n = 50$	9.65	14.69	20.71	27.76
	$n = 100$	8.51	12.49	16.64	21.68
	$n = 200$	7.60	10.43	13.70	16.84
	$n = 400$	7.19	9.09	11.46	13.65

$\gamma^* = \min(\delta, \gamma)$, we obtain, as $d \rightarrow d_0$,

$$\begin{aligned}
& \Pr[1(\mathbf{S}^n(d_0) = \mathbf{s}^n)] \\
&= \left(\frac{1}{2}\right)^n + g(0) \sum_{m=2}^n \sum_{j=1}^{m-1} s_m \psi_j(d_0 - d) \left[\left(\frac{1}{2}\right)^{n-1} \mathbb{E}|\varepsilon_1| s_{m-j} + \mathcal{O}(|d_0 - d|^{\delta/(1+\delta)}) \right] \\
&\quad + \mathcal{O}(|d_0 - d|^{1+\gamma^*}) \\
&= \left(\frac{1}{2}\right)^n \left[1 + 2g(0)\mathbb{E}|\varepsilon_1| \sum_{j=1}^{n-1} \psi_j(d_0 - d) \sum_{m=j+1}^n s_m s_{m-j} \right] \\
&\quad + \mathcal{O}(|d_0 - d|^{1+\gamma^*}) + \mathcal{O}(|d_0 - d|^{1+\delta/(1+\delta)}),
\end{aligned}$$

after noticing that $|\psi_j(d_0 - d)|/|d_0 - d| = \mathcal{O}(1)$ as $d \rightarrow d_0, j = 1, 2, \dots$, by Lemma 1 (b), and that $\delta, \gamma^* > 0$. \square

Proof of Theorem 1. It follows from Proposition 1. \square

Proof of Theorem 2. It follows from Anderson's (1971) Theorems 7.7.1 and 7.7.5; see also the proof of Theorem 6.2.3 of Boldin et al. (1997). \square

Table 6
Empirical power at 5% of significance of Dickey-Fuller asymptotic test

		a			
		0.5	1	1.5	2
N(0, 1)	$n = 50$	7.48	10.35	14.10	19.12
	$n = 100$	6.65	8.43	10.66	13.45
	$n = 200$	5.93	7.05	8.40	9.96
	$n = 400$	5.65	6.39	7.23	8.14
t_4	$n = 50$	7.21	10.02	13.78	18.71
	$n = 100$	6.45	8.29	10.54	13.36
	$n = 200$	5.79	6.97	8.29	9.73
	$n = 400$	5.51	6.28	7.04	8.07
t_2	$n = 50$	6.33	8.88	12.27	16.76
	$n = 100$	5.65	7.20	9.15	11.71
	$n = 200$	5.15	6.08	7.32	8.92
	$n = 400$	4.96	5.51	6.40	7.31
t_1	$n = 50$	4.32	6.08	16.25	12.44
	$n = 100$	3.71	4.89	6.36	8.32
	$n = 200$	3.38	4.14	4.99	6.06
	$n = 400$	3.24	3.59	4.15	4.74

Proof of Proposition 2. It follows from the definition of T_n that

$$\begin{aligned}
 & n^{-1/2}T_n(d_n + n^{-1/2}\theta) - n^{-1/2}T_n(d_n) \\
 &= n^{-1/2} \sum_{t=1}^{n-1} t^{-1} \{ \tilde{\gamma}_n(d_n + n^{-1/2}\theta) - \tilde{\gamma}_n(d_n) \}. \tag{9}
 \end{aligned}$$

Then defining $\Delta_k(x)$ as

$$\Delta_k(x) := \begin{cases} 1, & \varepsilon_k < x, \\ 1/2, & \varepsilon_k = x, \\ 0, & \varepsilon_k > x \end{cases}$$

and setting $\eta_n := n^{-1/2}\theta$ we have under $H_{1n}(a)$, that

$$\begin{aligned}
 S_k(d_n + \eta_n) &= \text{sign}(\Delta^{d_n + \eta_n} u_k) \\
 &= \text{sign}(\Delta^{\eta_n} \varepsilon_k) \\
 &= 1 - 2\Delta_k(\eta_n u_{k-1}(\eta_n)),
 \end{aligned}$$

where $u_{k-1}(\eta_n) := \eta_n^{-1}(1 - \Delta^{\eta_n})\varepsilon_k$, and $\Delta = \Delta(L) = 1 - L$ is the difference operator. Then (9) is

$$\sum_{t=1}^{n-1} t^{-1} \{4x_{tn}(\eta_n) + 4y_{tn}(\eta_n) - 2z_{tn}(\eta_n)\},$$

where

$$\begin{aligned} x_{tn}(\eta_n) &= n^{-1/2} \sum_{k=t+1}^n \begin{pmatrix} \Delta_{k-t}(\eta_n) u_{k-t-1}(\eta_n) \Delta_k(\eta_n) u_{k-1}(\eta_n) \\ -\Delta_{k-t}(\eta_n) u_{k-t-1}(\eta_n) G(\eta_n) u_{k-1}(\eta_n) \\ -\Delta_{k-t}(0) \Delta_k(0) + \Delta_{k-t}(0) G(0) \end{pmatrix}, \\ y_{tn}(\eta_n) &= n^{-1/2} \sum_{k=t+1}^n (\Delta_{k-t}(\eta_n) u_{k-t-1}(\eta_n) G(\eta_n) u_{k-1}(\eta_n) - \Delta_{k-t}(0) G(0)), \\ z_{tn}(\eta_n) &= n^{-1/2} \sum_{k=t+1}^n \begin{pmatrix} \Delta_{k-t}(\eta_n) u_{k-t-1}(\eta_n) - \Delta_{k-t}(0) \\ +\Delta_k(\eta_n) u_{k-1}(\eta_n) - \Delta_k(0) \end{pmatrix}. \end{aligned}$$

Now the proof of the proposition is immediate from Lemmas 5–9 with $\eta_n = \theta n^{-1/2}$. \square

Proof of Theorem 3. It follows from the CLT of Theorem 2 with $d = d_n$ and Proposition 2 with $\theta = -a$. \square

Proof of Proposition 3. The proposition follows as Proposition 1, adapting the proofs of Lemmas 3 and 4 accordingly, noticing that under model (7) we have that

$$1(\varepsilon_t(\theta_0) < 0) = 1(\varepsilon_t < [\Delta^d(L) a_\theta(L) - \Delta^{d_0}(L) a_{\theta_0}(L)] u_t + [b_\theta^*(L) \varepsilon_t - b_{\theta_0}^*(L) \varepsilon_t(\theta_0)]).$$

Then, if only one a_j changes, $j = 1, \dots, p$, we have that

$$\begin{aligned} 1(\varepsilon_t(\theta_0) < 0) &= 1 \left(\varepsilon_t < [a_{\theta^1, j} - a_{\theta_0^1, j}] \Delta^d(L) u_{t-j} + B_{\theta^2}^*(L) \frac{\Delta^d(L)}{B_{\theta^2}(L)} [a_{\theta^1}(L) - a_{\theta_0^1}(L)] u_t \right) \\ &= 1(\varepsilon_t < [a_{\theta^1, j} - a_{\theta_0^1, j}] \alpha_\theta(L) \varepsilon_{t-j}), \end{aligned}$$

whereas if only d changes,

$$\begin{aligned} 1(\varepsilon_t(\theta_0) < 0) &= 1 \left(\varepsilon_t < [1 - \Delta^{d_0-d}(L)] \Delta^d(L) A_{\theta^1}(L) u_t + B_{\theta^2}^*(L) \frac{a_{\theta^1}(L)}{b_{\theta^2}(L)} [\Delta^d(L) - \Delta^{d_0}(L)] u_t \right) \\ &= 1 \left(\varepsilon_t < \alpha_\theta(L) \sum_{j=1}^{t-1} \psi_j(d - d_0) \varepsilon_{t-j} \right). \end{aligned}$$

If only one b_k changes, $k = 1, \dots, q$, we have that

$$1(\varepsilon_t(\theta_0) < 0) = 1\left(\varepsilon_t < [b_{\theta,k} - b_{\theta_0,k}] \varepsilon_{t-k} + \sum_{r=1}^q b_{\theta_0,r} [\varepsilon_{t-r} - \varepsilon_{t-r}(\theta_0)]\right),$$

where $\varepsilon_{t-r} - \varepsilon_{t-r}(\theta_0) = [b_{\theta_0,k} - b_{\theta,k}] B_{\theta_0}^{-1}(L) \varepsilon_{t-r-k}$. Then $1(\varepsilon_t(\theta_0) < 0) = 1(\varepsilon_t < [b_{\theta,k} - b_{\theta_0,k}] \beta_{\theta_0}(L) \varepsilon_{t-k})$. \square

Proof of Theorem 4. It follows directly from Proposition 3. \square

Proof of Theorem 5. It follows from Anderson's (1971) Theorem 7.7.5 as Theorem 2. \square

Proof of Proposition 4. The proof is based on a combination of the ideas of Proposition 2—for the fractional aspect and Theorem 3.1 (Boldin and Stute, 2000), for the MA part and the joint effect of the drift of short- and long-memory parameters. Note that, $\eta_n = \rho_1 n^{-1/2}$,

$$\varepsilon_t(\theta_n + n^{-1/2} \rho) - \varepsilon_t = (\Delta^{\eta_n}(L) - 1) \varepsilon_t + \left(\frac{A_{\theta_n + n^{-1/2} \rho}(L)}{A_{\theta_n}(L)} \frac{B_{\theta_n}(L)}{B_{\theta_n + n^{-1/2} \rho}(L)} - 1 \right) \varepsilon_t \quad (10)$$

$$+ (\Delta^{\eta_n}(L) - 1) \left(\frac{A_{\theta_n + n^{-1/2} \rho}(L)}{A_{\theta_n}(L)} \frac{B_{\theta_n}(L)}{B_{\theta_n + n^{-1/2} \rho}(L)} - 1 \right) \varepsilon_t. \quad (11)$$

The contribution of the two approximation terms on (10) to $n^{-1/2} I_n(\theta_n + n^{-1/2} \rho) - n^{-1/2} I_n(\theta_n)$ is dealt with as in Proposition 2 and as in Boldin and Stute (2000), respectively. The contribution of (11) is the joint contribution of the locally in the memory parameter and the ARMA parameters, and is of smaller order as we now argue. The idea is to include this contribution in the remainder term $\sigma_t(\rho)$ used in Boldin and Stute (2000) such that,

$$\sup |\sigma_t(\rho)| \leq \hat{\sigma}_t, \quad t = 1, \dots, n, \quad n > n_0,$$

where

$$\sup_n \frac{1}{\log n} \sum_{t=1}^n E \hat{\sigma}_t < \infty,$$

and $\sigma_t(\rho)$, $\hat{\sigma}_t$ are \mathcal{F}_{t-1} -measurable for $t = 1, \dots, n$. This follows because,

$$(\Delta^{\eta_n}(L) - 1) \left(1 - \frac{A_{\theta_n + n^{-1/2} \rho}(L)}{A_{\theta_n}(L)} \frac{B_{\theta_n}(L)}{B_{\theta_n + n^{-1/2} \rho}(L)} \right) = n^{-1} \sum_{j=2}^{\infty} L^j \varphi_{n,j}$$

where $\sup_n |\varphi_{n,j}| \leq C j^{-1}$, $j = 2, \dots, n$ as $n \rightarrow \infty$, an C denotes, henceforth, a generic constant. So, the truncated filter up to lag n satisfies $|\varphi_n(1)| = O(\log n)$. The bound on $\sum_{t=1}^n E \hat{\sigma}_t$ is enough for the methods of Boldin and Stute (2000) to go through for our set up. \square

Proof of Theorem 6. The first statement follows from Proposition 4 and Theorem 5 as in the proof of Theorem 3. The second follows at once from the first. \square

Proof of Theorem 7. By assumption $n^{1/2}(\hat{\theta}_n - \theta_n) = O_p(1)$ under $H'_{1n}(\mathbf{h})$, and hence $n^{1/2}(\hat{\theta}_{0n} - \theta_n) = O_p(1)$ as well. Therefore in view of Proposition 4, consistency of $\hat{\mathbf{V}}_n$ and definition of $\mathbf{V}(\theta)$, we have

$$\begin{aligned} n^{-1/2}\hat{\mathbf{V}}'_n l_n(\hat{\theta}_{0n}) &= n^{-1/2}\hat{\mathbf{V}}'_n l_n(\theta_0) - 2g(0)E|\varepsilon_1|\hat{\mathbf{V}}'_n \mathbf{A}(\theta_0)n^{1/2}(\hat{\theta}_{0n} - \theta_n) + o_p(1) \\ &= n^{-1/2}\mathbf{V}'(\theta_0)l_n(\theta_0) - 2g(0)E|\varepsilon_1|\mathbf{V}'(\theta_0)\mathbf{A}(\theta_0)n^{1/2}(\hat{\theta}_{0n} - \theta_n) + o_p(1) \\ &= n^{-1/2}\mathbf{V}'(\theta_0)l_n(\theta_0) - 2g(0)E|\varepsilon_1|\mathbf{V}^{-1}(\theta_0)(\mathbf{h}^{(1)}, n^{1/2}(\hat{\theta}_{0n}^{(2)} - \theta_n^{(2)}))' + o_p(1). \end{aligned}$$

Then, since

$$\mathbf{V}^{-1}(\theta_0) = \begin{pmatrix} c^{1/2}(\theta_0) & \mathbf{0}' \\ \mathbf{M}^{-1/2}(\theta_0)\mathbf{B}(\theta_0) & \mathbf{M}^{1/2}(\theta_0) \end{pmatrix}$$

the last equality implies

$$\pi \circ n^{-1/2}\hat{\mathbf{V}}'_n l_n(\hat{\theta}_{0n}) = \pi \circ n^{-1/2}\mathbf{V}'(\theta_0)l_n(\theta_0) - 2g(0)E|\varepsilon_1|c^{1/2}(\theta_0)\mathbf{h}^{(1)} + o_p(1).$$

Now, this expression and the fact that

$$\pi \circ n^{-1/2}\mathbf{V}'(\theta_0)l_n(\theta_0) \rightarrow_d \mathbf{N}(0, 1)$$

prove the theorem. \square

7. Lemmata

The first Lemma provides bounds on ψ_j , which are applied for bounding its derivatives at different places. The rest of the lemmas are applied for proving the different propositions and theorems in the text.

Lemma 1. Define $\alpha_j(\eta) := -\psi_j(\eta)/\eta$, $\eta \neq 0$ and $\alpha_j(0) := \lim_{\eta \rightarrow 0} -\psi_j(\eta)/\eta$, $j = 1, 2, \dots$. Then

- (a) $\alpha_j(\eta)$ is monotone decreasing for all η such that $|\eta| < 1$, $j = 2, 3, \dots$
- (b) There exists a neighborhood of zero, say \mathcal{N} , such that,

$$|\alpha_j(\eta) - j^{-1}| \leq C|\eta|j^{|\eta|-1} \log^2(j+1), \text{ for each } \eta \in \mathcal{N} \text{ and each } j = 1, 2, \dots$$

- (c) For any constant $K > 0$, $\sup_{|\eta| \leq Kn^{-r}} \sum_{j=1}^n |\alpha_j(\eta)| = O(\log n)$, for all $r > 0$, as $n \rightarrow \infty$.

Proof. (a) Since $\alpha_j(\eta) > 0$, we only need to check that

$$\frac{\partial}{\partial \eta} \log \alpha_j(\eta) < 0, \quad |\eta| < 1.$$

First, for $\eta > 0$, we have that $\log \alpha_j(\eta) = \log(-\psi_j(\eta)) - \log \eta$ or using for $\eta < 0$ that $\log \alpha_j(\eta) = \log(\psi_j(\eta)) - \log(-\eta)$. Then it is immediate to check that

$$\frac{\partial}{\partial \eta} \log \alpha_j(\eta) = \varphi(-\eta) - \varphi(j - \eta) - \frac{1}{\eta},$$

where $\varphi(\eta) = \partial \log \Gamma(\eta) / \partial \eta$ is the digamma function. Now, using that

$$\varphi(z+j) = \varphi(z) + \sum_{r=0}^{j-1} \frac{1}{z+r}$$

(e.g. Theorem 1.2.7 of Andrews et al., 1999) the claim follows.

(b) First notice that,

$$\psi'_j(\eta) = \frac{1}{\Gamma^2(-\eta)\Gamma(j+1)} (-\Gamma(-\eta)\Gamma'(j-\eta) + \Gamma(j-\eta)\Gamma'(-\eta)),$$

where $\Gamma'(\eta) = \partial \Gamma(\eta) / \partial \eta$ and $\psi'_j(\eta) = \partial \psi_j(\eta) / \partial \eta$. Similarly, with $\Gamma''(\eta) = \partial^2 \Gamma(\eta) / \partial \eta^2$ and $\psi''_j(\eta) = \partial^2 \psi_j(\eta) / \partial \eta^2$ we obtain that

$$\psi''_j(\eta) = \frac{\Gamma(-\eta)^{-4}}{\Gamma(j+1)} \begin{pmatrix} \Gamma^3(-\eta)\Gamma''(j-\eta) + \Gamma^2(-\eta)\Gamma'(-\eta)\Gamma'(j-\eta) \\ -\Gamma^2(-\eta)\Gamma(j-\eta)\Gamma''(-\eta) - \Gamma^2(-\eta)\Gamma'(j-\eta)\Gamma'(-\eta) \\ -2\Gamma^2(-\eta)\Gamma'(j-\eta)\Gamma'(-\eta) + 2\Gamma'(-\eta)^2\Gamma(j-\eta)\Gamma(-\eta) \end{pmatrix} \quad (12)$$

see expression (4.21) in p. 124 of Wright (1995). Then use that

$$\frac{\partial^h \Gamma(x+a)}{\partial x^h \Gamma(x+b)} = x^{a-b} \log^h x (1 + o(1)) \text{ as } x \rightarrow \infty \text{ for } h = 0, 1, 2, \quad (13)$$

proved in Lemma 2 of Wright (1995), to bound $\psi_j(\eta)$ and its derivatives for large j . Therefore, applying a mean value theorem argument for $\eta \in \mathcal{N}$

$$\psi_j(\eta) = \psi_j(0) + \eta \psi'_j(0) + \frac{1}{2} \eta^2 \psi''_j(\eta^*), \quad j = 1, 2, \dots \quad (14)$$

for some $|\eta^*| \leq |\eta|$. Now, take into account that

$$\psi_j(0) = 0, \quad \psi'_j(0) = -j^{-1}, \quad j = 1, 2, \dots$$

(so $\alpha_j(0) = j^{-1}$) and that, for all η^* with $|\eta^*| \leq |\eta|$ in \mathcal{N} ,

$$|\psi''_j(\eta^*)| \leq C j^{|\eta^*|-1} \log^2 j \leq C j^{|\eta|-1} \log^2(j+1), \quad j = 1, 2, \dots \quad (15)$$

using (12)–(13) (cf. bound for $a_{2j}(\eta)$ in Wright, 1995, p. 124). Then, (14) and (15) prove (b).

(c) is an straightforward consequence of (b), because using that $\alpha_j(\eta) > 0$, there exists a constant C such that,

$$\begin{aligned} \sup_{|\eta| \leq Kn^{-r}} \sum_{j=1}^n |\alpha_j(\eta) - j^{-1}| &\leq C \frac{1}{n^r} \sum_{j=1}^n j^{Kn^{-r}-1} \log^2 n \\ &= O\left(n^{Kn^{-r}} \frac{\log^3 n}{n^r}\right) \\ &= o(1) \text{ as } n \rightarrow \infty, \end{aligned}$$

after noticing that $n^{Kn^{-r}} = O(1)$ as $n \rightarrow \infty$ for all $r > 0$. Thus, applying the triangle inequality $\sup_{|\eta| \leq Kn^{-r}} \sum_{j=1}^n |\alpha_j(\eta)| = O\left(\sum_{j=1}^n j^{-1}\right) + o(1) = O(\log n)$. \square

Lemma 2. Under A1–A3,

$$\Pr\{S_1(d_0) = s_1\} = \frac{1}{2}.$$

Proof. It follows directly from A2, since $\varepsilon_1(d_0) = (1 - L)^{d_0 - d} \varepsilon_1 = \varepsilon_1$ by (2). \square

Lemma 3. Under A1–A3, for each $k = 2, 3, \dots$,

$$\begin{aligned} \Pr\{S^k(d_0) = s^k\} &= \frac{1}{2} \Pr\{S^{k-1}(d_0) = s^{k-1}\} \\ &\quad + g(0)s_k \sum_{j=1}^{k-1} \psi_j(d_0 - d) \mathbb{E}[\varepsilon_{k-j} \mathbf{1}(S^{k-1}(d_0) = s^{k-1})] \\ &\quad + O(|d_0 - d|^{\gamma^* + 1}), \end{aligned}$$

as $d \rightarrow d_0$, where $\gamma^* = \min(\delta, \gamma)$.

Proof. Let \mathcal{F}_m the sigma algebra of events generated by $\{\varepsilon_k, k \leq m\}$. Then, for each $k = 2, \dots, n$,

$$\begin{aligned} \Pr[S^k(d_0) = s^k] &= \mathbb{E} \left\{ \mathbf{1}(S^{k-1}(d_0) = s^{k-1}) \left([1(S_k(d_0) = 1)] \frac{1 + s_k}{2} + 1(S_k(d_0) = -1) \frac{1 - s_k}{2} \right) \right\} \\ &= \mathbb{E} \left\{ \mathbf{1}(S^{k-1}(d_0) = s^{k-1}) \left(\frac{1 + s_k}{2} - s_k \mathbf{1}(\varepsilon_k(d_0) < 0) \right) \right\} \\ &= \mathbb{E} \left\{ \mathbf{1}(S^{k-1}(d_0) = s^{k-1}) \left(\frac{1 + s_k}{2} - s_k \mathbb{E} \left[\mathbf{1} \left(\varepsilon_k < - \sum_{j=1}^{k-1} \psi_j(d_0 - d) \varepsilon_{k-j} \right) \middle| \mathcal{F}_{k-1} \right] \right) \right\} \\ &= \frac{1 + s_k}{2} \mathbb{E}[\mathbf{1}(S^{k-1}(d_0) = s^{k-1})] - s_k \mathbb{E} \left[\mathbf{1}(S^{k-1}(d_0) = s^{k-1}) G \left(- \sum_{j=1}^{k-1} \psi_j(d_0 - d) \varepsilon_{k-j} \right) \right]. \end{aligned}$$

Assumption A3 guarantees that $G(x) = G(0) + xg(\rho x)$ for some $\rho \in [-1, 1]$. Also, by A3 g is Hölder continuous at zero of order $\gamma^* \in (0, \gamma]$, i.e. $|g(x) - g(0)| \leq M|x|^{\gamma^*}$, where M is a constant that may depend on γ^* . Take $\gamma^* = \min(\delta, \gamma)$, then, $|g(\rho x) - g(0)| \leq C|x|^{\gamma^*}$, and $G(x) = G(0) + xg(0) + e(x)$, where $e(x) \leq C|x|^{\gamma^* + 1}$. Therefore,

$$\begin{aligned} &\mathbb{E} \left[\mathbf{1}(S^{k-1}(d_0) = s^{k-1}) G \left(- \sum_{j=1}^{k-1} \psi_j(d_0 - d) \varepsilon_{k-j} \right) \right] \\ &= \mathbb{E} \left\{ \mathbf{1}(S^{k-1}(d_0) = s^{k-1}) \left[\frac{1}{2} - g(0) \sum_{j=1}^{k-1} \psi_j(d_0 - d) \varepsilon_{k-j} \right. \right. \\ &\quad \left. \left. + e \left(- \sum_{j=1}^{k-1} \psi_j(d_0 - d) \varepsilon_{k-j} \right) \right] \right\}. \end{aligned}$$

And the lemma is proved after noticing that, by definition of e and von Bahr and Esseen (1965) Theorem 2,

$$\begin{aligned} \mathbb{E} \left| 1(\mathbf{S}^{k-1}(d_0) = \mathbf{s}^{k-1}) e \left(- \sum_{j=1}^{k-1} \psi_j(d_0 - d) \varepsilon_{k-j} \right) \right| &\leq \mathbb{C} \mathbb{E} \left| \sum_{j=1}^{k-1} \psi_j(d_0 - d) \varepsilon_{k-j} \right|^{1+\gamma^*} \\ &\leq 2\mathbb{C} \mathbb{E} |\varepsilon_1|^{1+\gamma^*} \sum_{j=1}^{k-1} |\psi_j(d_0 - d)|^{1+\gamma^*} \\ &= \mathcal{O}(|d_0 - d|^{1+\gamma^*}) \text{ as } d \rightarrow d_0, \end{aligned}$$

by Lemma 1(b). \square

Lemma 4. Under A1–A3, for each $k = 1, 2, \dots$ and $m = 1, \dots, k$, and some $\delta > 0$,

$$\mathbb{E}[\varepsilon_m 1(\mathbf{S}^k(d_0) = \mathbf{s}^k)] = s_m \left(\frac{1}{2}\right)^k \mathbb{E}|\varepsilon_1| + \mathcal{O}(|d_0 - d|^{\delta/(1+\delta)}),$$

as $d \rightarrow d_0$.

Proof. Applying Lemma 3 recursively, with initial condition in Lemma 2,

$$\begin{aligned} &\mathbb{E}[1(\mathbf{S}^k(d_0) = \mathbf{s}^k)] \\ &= \left(\frac{1}{2}\right)^k + g(0) \sum_{m=2}^k \sum_{j=1}^{m-1} s_m \psi_j(d_0 - d) \left(\frac{1}{2}\right)^{k-m} \mathbb{E}[\varepsilon_{m-j} 1(\mathbf{S}^{m-1}(d_0) = \mathbf{s}^{m-1})] \\ &\quad + \mathcal{O}(|d_0 - d|^{\gamma^*+1}) \text{ as } d \rightarrow d_0 \\ &= \left(\frac{1}{2}\right)^k + \mathcal{O}(|d_0 - d|) \text{ as } d \rightarrow d_0. \end{aligned} \tag{16}$$

Now write,

$$\begin{aligned} \mathbb{E}[\varepsilon_m 1(\mathbf{S}^k(d_0) = \mathbf{s}^k)] &= \mathbb{E}[\varepsilon_m (1(\mathbf{S}^k(d_0) = \mathbf{s}^k) - 1(\mathbf{S}^k(d) = \mathbf{s}^k))] \\ &\quad + \mathbb{E}[\varepsilon_m 1(\mathbf{S}^k(d) = \mathbf{s}^k)]. \end{aligned} \tag{17}$$

Applying Hölders' inequality, the first term on the right-hand side of (17) is bounded by

$$\{\mathbb{E}|\varepsilon_1|^{1+\delta}\}^{1/(1+\delta)} \{\mathbb{E}|1(\mathbf{S}^k(d_0) = \mathbf{s}^k) - 1(\mathbf{S}^k(d) = \mathbf{s}^k)|^{(1+\delta)/\delta}\}^{\delta/(1+\delta)}.$$

Now,

$$\begin{aligned} &\mathbb{E}|1(\mathbf{S}^k(d_0) = \mathbf{s}^k) - 1(\mathbf{S}^k(d) = \mathbf{s}^k)| \\ &= \Pr(\mathbf{S}^k(d_0) = \mathbf{s}^k, \mathbf{S}^k(d) \neq \mathbf{s}^k) + \Pr(\mathbf{S}^k(d_0) \neq \mathbf{s}^k, \mathbf{S}^k(d) = \mathbf{s}^k) \\ &\leq \Pr(\mathbf{S}^k(d_0) \neq \mathbf{S}^k(d)) \\ &\leq \sum_{j=1}^k \Pr(S_j(d_0) \neq S_j(d)) \end{aligned} \tag{18}$$

for each $k = 1, 2, \dots$, with

$$\Pr(S_j(d_0) \neq S_j(d)) = \Pr(\varepsilon_j(d_0) > 0, \varepsilon_j(d) \leq 0) + \Pr(\varepsilon_j(d_0) \leq 0, \varepsilon_j(d) > 0).$$

Then, for each integer j ,

$$\begin{aligned} \Pr(\varepsilon_j(d_0) > 0, \varepsilon_j(d) \leq 0) &= \Pr\left(\varepsilon_j > -\sum_{t=1}^{j-1} \psi_t(d_0 - d)\varepsilon_{j-t}, \varepsilon_j \leq 0\right) \\ &= \Pr\left(-\sum_{t=1}^{j-1} \psi_t(d_0 - d)\varepsilon_{j-t} < \varepsilon_j \leq 0\right) \\ &\leq \mathbb{E}\left[\mathbb{E}\left[1\left(-\sum_{t=1}^{j-1} \psi_t(d_0 - d)\varepsilon_{j-t} < \varepsilon_j \leq 0\right) \middle| \mathcal{F}_{j-1}\right]\right] \\ &\leq C\mathbb{E}\left|\sum_{t=1}^{j-1} \psi_t(d_0 - d)\varepsilon_{j-t}\right| \\ &\leq C\mathbb{E}|\varepsilon_1| \sum_{t=1}^{j-1} |\psi_t(d_0 - d)|, \end{aligned}$$

because $G(0) - G(x) \leq C|x|$, $x \leq 0$, and the last expression is $O(|d_0 - d|)$ as $d \rightarrow d_0$ for each integer j because $\psi_t(0) = 0$, $t \geq 1$ and the mean value theorem, see Lemma 1 (b). Applying a similar bound for $\Pr(\varepsilon_j(d_0) \leq 0, \varepsilon_j(d) > 0)$ we obtain that

$$\mathbb{E}|1(\mathbf{S}^k(d_0) = \mathbf{s}^k) - 1(\mathbf{S}^k(d) = \mathbf{s}^k)| = O(|d_0 - d|) \text{ as } d \rightarrow d_0,$$

and (18) is $O(|d_0 - d|^{\delta/(1+\delta)})$.

The second term on the right-hand side of (17) is

$$\begin{aligned} \mathbb{E}[\varepsilon_m 1(\mathbf{S}^k(d) = \mathbf{s}^k)] &= \mathbb{E}\left[\left(\prod_{j \neq m}^k 1(\text{sign}(\varepsilon_j) = s_j)\right) \varepsilon_m 1(\text{sign}(\varepsilon_m) = s_m)\right] \\ &= \left(\frac{1}{2}\right)^{k-1} \mathbb{E}[\varepsilon_m 1(\text{sign}(\varepsilon_m) = s_m)] \\ &= \left(\frac{1}{2}\right)^{k-1} s_m \mathbb{E}[\varepsilon_1 1(\varepsilon_1 > 0)] = \left(\frac{1}{2}\right)^k s_m \mathbb{E}|\varepsilon_1|, \end{aligned} \quad (19)$$

after noticing that, since $\mathbb{E}(\varepsilon_1) = 0$,

$$\begin{aligned} \mathbb{E}[\varepsilon_1 1(\varepsilon_1 > 0)] &= \frac{1}{2} \mathbb{E}[\varepsilon_1 (2 \cdot 1(\varepsilon_1 > 0) - 1)] + \frac{1}{2} \mathbb{E}(\varepsilon_1) \\ &= \frac{1}{2} \mathbb{E}[\varepsilon_1 \text{sign}(\varepsilon_1)] \\ &= \frac{1}{2} \mathbb{E}|\varepsilon_1|. \end{aligned}$$

Thus, (17), (18) and (19) prove the lemma. \square

Lemma 5. Let $r > 1/4$, $\Theta_n = \Theta n^{-r}$. Then under $H_{1n}(a)$ and assuming the same conditions as in Proposition 2, as $n \rightarrow \infty$,

$$\sup_{|\eta_n| \leq \Theta_n} \left| \sum_{t=1}^{n-1} t^{-1} x_{tn}(\eta_n) \right| = o_p(1).$$

Proof. First, $(1 - \Delta^{\eta_n})\varepsilon_k = \eta_n u_{k-1}(\eta_n)$, where

$$u_{k-1}(\eta_n) = \sum_{j=1}^{k-1} \alpha_j(\eta_n) \varepsilon_{k-j},$$

with $\alpha_j(\eta_n)$ defined as in Lemma 1.

Divide the interval $[-\Theta_n, \Theta_n]$ into 3^{m_n} subintervals by the points

$$\eta_{sn} = -\Theta_n + 2\Theta_n 3^{-m_n} s, \quad s = 0, 1, \dots, 3^{m_n}.$$

Choose r_1 such that $0 < r_1 < r$ and $r + r_1 > 1/2$ and let $3^{m_n} \sim n^{r_1}$.

For an arbitrary $|\eta_n| \leq \Theta_n$ select the point $\eta_{\ell n}$ from among the points $\{\eta_{sn}\}$ such that

$$0 \leq \eta_{\ell n} - \eta_n \leq \Theta_n 3^{-m_n} = \frac{\nabla}{2},$$

$\nabla := 2\Theta_n 3^{-m_n}$, so $\eta_{\ell-1,n} \leq \eta_n \leq \eta_{\ell n} \leq \eta_n + \nabla/2 \leq \eta_{\ell+1,n}$ by Lemma 1 (a), and for all $j > 0$,

$$\alpha_j(\eta_{\ell-1,n}) \geq \alpha_j(\eta_n) \geq \alpha_j(\eta_{\ell n}). \quad (20)$$

Define the variables,

$$\begin{aligned} \widehat{\varepsilon}_k(\eta_{sn}) &:= \varepsilon_k \left\{ 1 - \frac{\nabla}{\eta_{sn}} 1(\varepsilon_k \leq 0) \right\}, \\ \widetilde{\varepsilon}_k(\eta_{sn}) &:= \varepsilon_k \left\{ 1 - \frac{\nabla}{\eta_{sn}} 1(\varepsilon_k \geq 0) \right\} \end{aligned}$$

and the variables

$$\begin{aligned} \widetilde{u}_{k-1}(\eta_{\ell n}) &:= \sum_{j=1}^{k-1} \alpha_j(\eta_{\ell n}) \widetilde{\varepsilon}_{k-j}(\eta_{\ell n}), \\ \widehat{u}_{k-1}(\eta_{\ell n}) &:= \sum_{j=1}^{k-1} \alpha_j(\eta_{\ell-1,n}) \widehat{\varepsilon}_{k-j}(\eta_{\ell n}). \end{aligned}$$

Then it follows that

$$\eta_{\ell n} \widetilde{\varepsilon}_k(\eta_{\ell n}) \leq \eta_n \varepsilon_k \leq \eta_{\ell n} \widehat{\varepsilon}_k(\eta_{\ell n})$$

and, using (20), that, $k = 0, 1, \dots, n-1$,

$$\eta_{\ell n} \widetilde{u}_k(\eta_{\ell n}) \leq \eta_n u_k(\eta_n) \leq \eta_{\ell n} \widehat{u}_k(\eta_{\ell n}). \quad (21)$$

Writing

$$x_{in}(\eta_n) := x_{in}(\eta_n, \mathbf{U}_n),$$

where $\mathbf{U}_n = (u_1, \dots, u_{n-1})$, $\widehat{\mathbf{U}}_{sn} = (\widehat{u}_1(\eta_{sn}), \dots, \widehat{u}_{n-1}(\eta_{sn}))$, $\widetilde{\mathbf{U}}_{sn} = (\widetilde{u}_1(\eta_{sn}), \dots, \widetilde{u}_{n-1}(\eta_{sn}))$, we obtain that

$$\begin{aligned} \sum_{t=1}^{n-1} t^{-1} x_{in}(\eta_n, \mathbf{U}_n) &\leq \sum_{t=1}^{n-1} t^{-1} x_{in}(\eta_n, \widehat{\mathbf{U}}_{sn}) \\ &\quad + n^{-1/2} \sum_{t=1}^{n-1} t^{-1} \sum_{k=t+1}^n \left\{ \begin{array}{l} \Delta_{k-t}(\eta_{sn} \widehat{u}_{k-t-1, \ell}(\eta_{sn})) G(\eta_{sn} \widehat{u}_{k-1, \ell}(\eta_{sn})) \\ - \Delta_{k-t}(\eta_{sn} \widetilde{u}_{k-t-1, \ell}(\eta_{sn})) G(\eta_{sn} \widetilde{u}_{k-1, \ell}(\eta_{sn})) \end{array} \right\}, \\ \sum_{t=1}^{n-1} t^{-1} x_{in}(\eta_n, \mathbf{U}_n) &\geq \sum_{t=1}^{n-1} t^{-1} x_{in}(\eta_n, \widetilde{\mathbf{U}}_{sn}) \\ &\quad - n^{-1/2} \sum_{t=1}^{n-1} t^{-1} \sum_{k=t+1}^n \left\{ \begin{array}{l} \Delta_{k-t}(\eta_{sn} \widehat{u}_{k-t-1, \ell}(\eta_{sn})) G(\eta_{sn} \widehat{u}_{k-1, \ell}(\eta_{sn})) \\ - \Delta_{k-t}(\eta_{sn} \widetilde{u}_{k-t-1, \ell}(\eta_{sn})) G(\eta_{sn} \widetilde{u}_{k-1, \ell}(\eta_{sn})) \end{array} \right\}. \end{aligned}$$

In order to discretize Θ_n , we see that these last two inequalities imply that

$$\begin{aligned} &\sup_{|\eta_n| \leq \Theta_n} \left| \sum_{t=1}^{n-1} t^{-1} x_{in}(\eta_n) \right| \\ &\leq \sup_{s \leq 3^{mn}} \left| \sum_{t=1}^{n-1} t^{-1} x_{in}(\eta_{sn}, \widehat{\mathbf{U}}_{sn}) \right| + \sup_{s \leq 3^{mn}} \left| \sum_{t=1}^{n-1} t^{-1} x_{in}(\eta_{sn}, \widetilde{\mathbf{U}}_{sn}) \right| \\ &\quad + \sup_{s \leq 3^{mn}} \sum_{t=1}^{n-1} \frac{t^{-1}}{n^{1/2}} \sum_{k=t+1}^n \left\{ \begin{array}{l} \Delta_{k-t}(\eta_{sn} \widehat{u}_{k-t-1, s}(\eta_{sn})) G(\eta_{sn} \widehat{u}_{k-1, s}(\eta_{sn})) \\ - \Delta_{k-t}(\eta_{sn} \widetilde{u}_{k-t-1, s}(\eta_{sn})) G(\eta_{sn} \widetilde{u}_{k-1, s}(\eta_{sn})) \end{array} \right\}. \end{aligned} \quad (22)$$

Consider now the first term in (22) depending on $x_{in}(\eta_{sn}, \widehat{\mathbf{U}}_{sn})$. Writing

$$\begin{aligned} \xi_k(t) &= \xi_k(n, s, t) \\ &:= \Delta_{k-t}(\eta_{sn} \widehat{u}_{k-t-1, s}(\eta_{sn})) \{ \Delta_k(\eta_{sn} \widehat{u}_{k-1, s}(\eta_{sn})) - G(\eta_{sn} \widehat{u}_{k-1, s}(\eta_{sn})) \} \\ &\quad - \Delta_{k-t}(0) \{ \Delta_k(0) - G(0) \}, \end{aligned}$$

we obtain that

$$x_{in}(\eta_n, \widehat{\mathbf{U}}_{sn}) = n^{-1/2} \sum_{k=t+1}^n \xi_k(t).$$

Consider the triangular array $\xi_k(t)$. For $k = t+1, \dots, n$, and each $t = 1, \dots, n-1$, $\xi_k(t)$ is a martingale difference sequence, so it has zero mean, $E\xi_k(t)\xi_j(t) = 0$, and

$$E\xi_k^2(t) \leq 2 \sup_x g(x) \{ E|\eta_{sn} \widehat{u}_{k-1, s}(\eta_{sn})| + E|\eta_{sn} \widehat{u}_{k-t-1, s}(\eta_{sn})| \} = O(n^{-r} \log n),$$

since $|\eta_{sn}| \leq \Theta_n = O(n^{-r})$, and using that, by Lemma 1(c), for each $k = 1, 2, \dots$

$$E|\widehat{u}_{k-1, s}(\eta_{sn})| \leq \sum_{j=1}^{k-2} |\alpha_j(\eta_{s-1, n})| E|\varepsilon_{k-1-j}| = O(\log n).$$

Furthermore, $E\xi_k(t)\xi_j(t') = 0$, for any t, t' , so the ξ_k are also uncorrelated between rows. Then by Chebyshev's inequality, for any $\varepsilon > 0$,

$$\begin{aligned} & \mathbf{P} \left\{ \sup_{s \leq 3^{m_n}} \left| \sum_{t=1}^{n-1} t^{-1} x_{tn}(\eta_{sn}, \widehat{\mathbf{U}}_{sn}) \right| > \varepsilon \right\} \\ & \leq \sum_{s=0}^{3^{m_n}} \mathbf{P} \left\{ \left| \sum_{t=1}^{n-1} t^{-1} x_{tn}(\eta_{sn}, \widehat{\mathbf{U}}_{sn}) \right| > \varepsilon \right\} \\ & \leq \varepsilon^2 \sum_{s=0}^{3^{m_n}} \sum_{t=1}^{n-1} \sum_{t'=1}^{n-1} t'^{-1} t^{-1} \mathbf{E}[x_{tn}(\eta_{sn}, \widehat{\mathbf{U}}_{sn}) x_{t'n}(\eta_{sn}, \widehat{\mathbf{U}}_{sn})]. \end{aligned}$$

This equals

$$\begin{aligned} & \varepsilon^2 \sum_{s=0}^{3^{m_n}} \sum_{t=1}^{n-1} \sum_{t'=1}^{n-1} t'^{-1} t^{-1} n^{-1} \sum_{k=t+1}^n \sum_{k'=t'+1}^n \mathbf{E}[\xi_k(t) \xi_{k'}(t')] \\ & \leq C 3^{m_n} \sum_{t=1}^{n-1} \sum_{t'=1}^{n-1} t'^{-1} t^{-1} n^{-1} \sum_{k=t+1}^n \mathbf{E}[\xi_k(t) \xi_k(t')] \\ & \leq C 3^{m_n} \sum_{t=1}^{n-1} \sum_{t'=1}^{n-1} t'^{-1} t^{-1} n^{-r} \log n = O(n^{r_1-r} \log^3 n) = o(1), \end{aligned}$$

using Cauchy inequality. By a similar argument

$$\sup_{s \leq 3^{m_n}} \left| \sum_{t=1}^{n-1} t^{-1} x_{tn}(\eta_{sn}, \widetilde{\mathbf{U}}_{sn}) \right| = o_p(1).$$

Consider now the last term on the right-hand side of (22), which is not greater than

$$\begin{aligned} & \sup_{s \leq 3^{m_n}} \sum_{t=1}^{n-1} \frac{t^{-1}}{n^{1/2}} \sum_{k=t+1}^n \{ \Delta_{k-t}(\eta_{sn} \widehat{u}_{k-t-1,s}(\eta_{sn})) \\ & \quad - \Delta_{k-t}(\eta_{sn} \widetilde{u}_{k-t-1,s}(\eta_{sn})) \} G(\eta_{sn} \widehat{u}_{k-1,s}(\eta_{sn})) \end{aligned} \quad (23)$$

$$\begin{aligned} & + \sup_{s \leq 3^{m_n}} \sum_{t=1}^{n-1} \frac{t^{-1}}{n^{1/2}} \sum_{k=t+1}^n \{ G(\eta_{sn} \widehat{u}_{k-1,s}(\eta_{sn})) \\ & \quad - G(\eta_{sn} \widetilde{u}_{k-1,s}(\eta_{sn})) \} \Delta_{k-t}(\eta_{sn} \widetilde{u}_{k-t-1,s}(\eta_{sn})). \end{aligned} \quad (24)$$

Now (24) does not exceed

$$\begin{aligned} & \sup_{s \leq 3^{m_n}} n^{-1/2} \sum_{t=1}^{n-1} t^{-1} \sum_{k=t+1}^n \{ G(\eta_{sn} \widehat{u}_{k-1,s}(\eta_{sn})) - G(\eta_{sn} \widetilde{u}_{k-1,s}(\eta_{sn})) \} \\ & \leq \sup_{s \leq 3^{m_n}} \sup_x g(x) n^{-1/2} \sum_{t=1}^{n-1} t^{-1} \sum_{k=t+1}^n \eta_{sn} \{ \widehat{u}_{k-1,s}(\eta_{sn}) - \widetilde{u}_{k-1,s}(\eta_{sn}) \} \\ & \leq 2\Theta_n 3^{-m_n} \sup_x g(x) n^{-1/2} \sum_{t=1}^{n-1} t^{-1} \sum_{k=t+1}^n |u_{k-1}| \\ & = O_p(n^{1/2-r_1-r} \log n) = o_p(1). \end{aligned} \quad (25)$$

It remains to bound (23). It does not exceed

$$\begin{aligned} & \sup_{s \leq 3^{m_n}} n^{-1/2} \sum_{t=1}^{n-1} t^{-1} \sum_{k=2}^n \{ \Delta_k(\eta_{sn} \widehat{u}_{k-1,s}(\eta_{sn})) - \Delta_k(\eta_{sn} \widetilde{u}_{k-1,s}(\eta_{sn})) \} \\ & \leq \sup_{s \leq 3^{m_n}} n^{-1/2} \sum_{t=1}^{n-1} t^{-1} \sum_{k=2}^n \left\{ \begin{array}{l} \Delta_k(\eta_{sn} \widehat{u}_{k-1,s}(\eta_{sn})) - G(\eta_{sn} \widehat{u}_{k-1,s}(\eta_{sn})) \\ - \Delta_k(\eta_{sn} \widetilde{u}_{k-1,s}(\eta_{sn})) + G(\eta_{sn} \widetilde{u}_{k-1,s}(\eta_{sn})) \end{array} \right\} \end{aligned} \quad (26)$$

$$+ \sup_{s \leq 3^{m_n}} n^{-1/2} \sum_{t=1}^{n-1} t^{-1} \sum_{k=2}^n \{ G(\eta_{sn} \widehat{u}_{k-1,s}(\eta_{sn})) - G(\eta_{sn} \widetilde{u}_{k-1,s}(\eta_{sn})) \} \quad (27)$$

Now (27) can be shown to be $o_p(1)$ as (25). Consider (26). Put

$$\begin{aligned} v_k &= v_k(n, s) \\ &:= \Delta_k(\eta_{sn} \widehat{u}_{k-1,s}(\eta_{sn})) - G(\eta_{sn} \widehat{u}_{k-1,s}(\eta_{sn})) - \Delta_k(\eta_{sn} \widetilde{u}_{k-1,s}(\eta_{sn})) + G(\eta_{sn} \widetilde{u}_{k-1,s}(\eta_{sn})). \end{aligned}$$

Then (26) becomes

$$\sup_{s \leq 3^{m_n}} \left| n^{-1/2} \sum_{t=1}^{n-1} t^{-1} \sum_{k=2}^n v_k(n, s) \right|.$$

Now the variables v_k are centered, uncorrelated, and $\sup_{s \leq 3^{m_n}} \text{Ev}_k^2 = O(n^{-r} 3^{-m_n}) = O(n^{-r-r_1})$ uniformly in k . Therefore by Chebyshev's inequality for any $\varepsilon > 0$,

$$\begin{aligned} & \mathbb{P} \left\{ \sup_{s \leq 3^{m_n}} \left| n^{-1/2} \sum_{t=1}^{n-1} t^{-1} \sum_{k=2}^n v_k \right| \geq \varepsilon \right\} \\ & \leq \varepsilon^{-2} n^{-1} \sum_{s=0}^{3^{m_n}} \sum_{t=1}^{n-1} \sum_{t'=1}^{n-1} t^{-1} t'^{-1} \sum_{k=2}^n \text{Ev}_k^2 = O(n^{-r} \log^2 n) = o(1), \end{aligned}$$

which completes the proof of the lemma. \square

Lemma 6. Let $r > 1/4$, $\Theta_n = \Theta n^{-r}$. Then under the assumptions of Proposition 2 and $H_{1n}(a)$, as $n \rightarrow \infty$,

$$\sup_{|\eta_n| \leq \Theta_n} \left| \sum_{t=1}^{n-1} t^{-1} z_{tn}(\eta_n) \right| = o_p(1).$$

Proof. Note that

$$\sup_{|\eta_n| \leq \Theta_n} \left| \sum_{t=1}^{n-1} t^{-1} z_{tn}(\eta_n) \right| \leq 2 \sup_{|\eta_n| \leq \Theta_n} \left| \sum_{t=1}^{n-1} t^{-1} n^{-1/2} \sum_{k=1}^n \{ \Delta_k(\eta_n u_{k-1}) - \Delta_k(0) \} \right|,$$

which can be shown to be $o_p(1)$ using the methods of the previous lemma. \square

Lemma 7. Let $\gamma^* = \min\{\delta, \gamma\}$ and $\Theta_n = \Theta n^{-r}$, $r > 0$, $0 < \Theta < \infty$. Then under the assumptions of Proposition 2 and $H_{1n}(a)$, as $n \rightarrow \infty$,

$$\begin{aligned} \mathbb{E} \sup_{|\eta_n| \leq \Theta_n} & \left| n^{-1} \sum_{t=1}^{n-1} t^{-1} \sum_{k=1}^n \{ \Delta_{k-t}(\eta_n u_{k-t-1}) - \Delta_{k-t}(0) \} u_{k-1}(\eta_n) \right| \\ & = O(n^{-r\gamma^*/(1+\gamma^*)} \log^3 n) = o(1). \end{aligned}$$

Proof. Under the set up of Lemma 5, but now with $3^{m_n} \sim \log n$, inequality (21) and the monotonicity of $\Delta_k(x)$ imply that

$$\begin{aligned} \Delta_{k-t}(\eta_n u_{k-t-1}) - \Delta_{k-t}(0) & \leq \Delta_{k-t}(\eta_{\ell_n} \widehat{u}_{k-t-1, \ell}(\eta_{\ell_n})) - \Delta_{k-t}(0) \\ & \leq |\Delta_{k-t}(\eta_{\ell_n} \widehat{u}_{k-t-1, \ell}(\eta_{\ell_n})) - \Delta_{k-t}(0)| \\ & \quad + |\Delta_{k-t}(\eta_{\ell_n} \widetilde{u}_{k-t-1, \ell}(\eta_{\ell_n})) - \Delta_{k-t}(0)| \end{aligned}$$

and

$$\begin{aligned} \Delta_{k-t}(\eta_n u_{k-t-1}) - \Delta_{k-t}(0) & \geq \Delta_{k-t}(\eta_{\ell_n} \widetilde{u}_{k-t-1, \ell}(\eta_{\ell_n})) - \Delta_{k-t}(0) \\ & \geq |\Delta_{k-t}(\eta_{\ell_n} \widetilde{u}_{k-t-1, \ell}(\eta_{\ell_n})) - \Delta_{k-t}(0)| \\ & \quad + |\Delta_{k-t}(\eta_{\ell_n} \widehat{u}_{k-t-1, \ell}(\eta_{\ell_n})) - \Delta_{k-t}(0)| \end{aligned}$$

Therefore

$$\begin{aligned} \sup_{|\eta_n| \leq \Theta_n} |\Delta_{k-t}(\eta_n u_{k-t-1}) - \Delta_{k-t}(0)| & \leq \sup_{s \leq 3^{m_n}} |\Delta_{k-t}(\eta_{ns} \widehat{u}_{k-t-1, s}(\eta_{ns})) - \Delta_{k-t}(0)| \quad (28) \\ & \quad + \sup_{s \leq 3^{m_n}} |\Delta_{k-t}(\eta_{ns} \widetilde{u}_{k-t-1, s}(\eta_{ns})) - \Delta_{k-t}(0)| \quad (29) \end{aligned}$$

We obtain that the expectation of (28) is

$$\begin{aligned} \mathbb{E} \sup_{s \leq 3^{m_n}} |\Delta_{k-t}(\eta_{ns} \widehat{u}_{k-t-1, s}(\eta_{ns})) - \Delta_{k-t}(0)| \\ \leq \sum_{s=0}^{3^{m_n}} \mathbb{E} |\Delta_{k-t}(\eta_{ns} \widehat{u}_{k-t-1, s}(\eta_{ns})) - \Delta_{k-t}(0)| \\ \leq \sum_{s=0}^{3^{m_n}} [\mathbb{P}\{0 < \varepsilon_{k-t} < |\eta_{ns} \widehat{u}_{k-t-1, s}(\eta_{ns})|\} + \mathbb{P}\{-|\eta_{ns} \widehat{u}_{k-t-1, s}(\eta_{ns})| < \varepsilon_{k-t} < 0\}]. \quad (30) \end{aligned}$$

The first probability is equal to,

$$\mathbb{E}[G(|\eta_{ns} \widehat{u}_{k-t-1, s}(\eta_{ns})|) - G(0)] \leq \sup_x g(x) \mathbb{E} |\eta_{ns} \widehat{u}_{k-t-1, s}(\eta_{ns})| = O(n^{-r} \log n),$$

applying Lemma 1(c). The second probability is estimated in the same way, so that (30) is $O(n^{-r} \log^2 n)$. Similarly,

$$\mathbb{E} \sup_{|\eta_n| \leq \Theta_n} |\Delta_{k-t}(u_{k-t-1}) - \Delta_{k-t}(0)| = O(n^{-r} \log^2 n).$$

Hence applying Hölder's inequality we obtain that

$$\begin{aligned}
& \mathbb{E} \sup_{|\eta_n| \leq \Theta_n} \left| n^{-1} \sum_{t=1}^{n-1} t^{-1} \sum_{k=1}^n \{ \Delta_{k-t}(\eta_n u_{k-t-1}) - \Delta_{k-t}(0) \} u_{k-1}(\eta_n) \right| \\
& \leq n^{-1} \sum_{t=1}^{n-1} t^{-1} \sum_{k=1}^n \mathbb{E} \sup_{|\eta_n| \leq \Theta_n} | \Delta_{k-t}(\eta_n u_{k-t-1}) - \Delta_{k-t}(0) | | u_{k-1}(\eta_n) | \\
& \leq n^{-1} \sum_{t=1}^{n-1} t^{-1} \sum_{k=1}^n \left[\mathbb{E} \sup_{|\eta_n| \leq \Theta_n} | \Delta_{k-t}(\eta_n u_{k-t-1}) - \Delta_{k-t}(0) |^{(1+\gamma^*)/\gamma^*} \right]^{\gamma^*/(1+\gamma^*)} \\
& \quad \times \left(\mathbb{E} \sup_{|\eta_n| \leq \Theta_n} | u_{k-1}(\eta_n) |^{1+\gamma^*} \right)^{1/(1+\gamma^*)} \\
& = O(n^{-\tau} \log^2 n)^{\gamma^*/(1+\gamma^*)} \log n = O(n^{-\tau\gamma^*/(1+\gamma^*)} \log^3 n),
\end{aligned}$$

completing the proof of the lemma, after noticing that, by Lemma 1(a), (c),

$$\begin{aligned}
\mathbb{E} \sup_{|\eta_n| \leq \Theta_n} | u_k(\eta_n) |^{1+\gamma^*} & \leq \mathbb{E} \left(\sum_{j=1}^{k-1} \sup_{|\eta_n| \leq \Theta_n} | \alpha_j(\eta_n) | | \varepsilon_{k-j} | \right)^{1+\gamma^*} \\
& \leq C \mathbb{E} \left(\sum_{j=1}^k | \alpha_j(-\Theta_n) | | \varepsilon_{k-j} | \right)^{1+\gamma^*} \\
& \leq C \left(\sum_{j=1}^k | \alpha_j(-\Theta_n) | \right)^{\gamma^*} \sum_{j=1}^k | \alpha_j(-\Theta_n) | \mathbb{E} | \varepsilon_{k-j} |^{1+\gamma^*} \\
& = O(\log^{1+\gamma^*} n),
\end{aligned}$$

using again Hölder's inequality. \square

Lemma 8. *Then under the assumptions of Proposition 2 and $H_{1n}(a)$, as $n \rightarrow \infty$,*

$$n^{-1} \sum_{t=1}^n t^{-1} \sum_{k=t+1}^n \Delta_{k-t}(0) u_{k-1}(\eta_n) = -\frac{1}{2} \mathbb{E} | \varepsilon_1 | \frac{\pi^2}{6} + \varsigma_n,$$

where $\sup_n \mathbb{E} | \varsigma_n | = o(1)$.

Proof. We consider only the case $0 < \delta < 1$. We write

$$u_{k-1}(\eta_n) = \sum_{s=1}^{t-1} \alpha_s(\eta_n) \varepsilon_{k-s} + \alpha_t(\eta_n) \varepsilon_{k-t} + \sum_{s=t+1}^k \alpha_s(\eta_n) \varepsilon_{k-s},$$

so that

$$\begin{aligned} & n^{-1} \sum_{t=1}^{n-1} t^{-1} \sum_{k=t+1}^n \Delta_{k-t}(0) u_{k-1}(\eta_n) \\ &= n^{-1} \sum_{t=1}^{n-1} t^{-1} \sum_{k=t+1}^n \Delta_{k-t}(0) \sum_{s=1}^{t-2} \alpha_s(\eta_n) \varepsilon_{k-s} \end{aligned} \quad (31)$$

$$+ n^{-1} \sum_{t=1}^{n-1} t^{-1} \sum_{k=t+1}^n \Delta_{k-t}(0) \alpha_t(\eta_n) \varepsilon_{k-t} \quad (32)$$

$$+ n^{-1} \sum_{t=1}^{n-1} t^{-1} \sum_{k=t+1}^n \Delta_{k-t}(0) \sum_{s=t+1}^k \alpha_s(\eta_n) \varepsilon_{k-s}. \quad (33)$$

Consider (31). The expectation of its absolute value is bounded by

$$\begin{aligned} & \sum_{t=1}^{n-1} t^{-1} \sum_{s=1}^{t-1} \alpha_s(\eta_n) \mathbb{E} \left| n^{-1} \sum_{k=1}^{n-t} \Delta_k(0) \varepsilon_{k+t-s} \right| \\ &= \sum_{t=1}^{n-1} t^{-1} \sum_{s=1}^{t-1} \alpha_s(\eta_n) \frac{n-t}{n} \mathbb{E} \left| (n-t)^{-1} \sum_{k=1}^{n-t} \Delta_k(0) \varepsilon_{k+t-s} \right| \\ &= o \left(\sum_{t=1}^{n-1} t^{-1} \sum_{s=1}^{t-1} s^{|\eta_n|-1} n^{-1} (n-t)^{1/(1+\gamma^*)} \right) \\ &= o \left(\sum_{t=1}^{n-1} t^{-1} n^{-\gamma^*/(1+\gamma^*)} \log n \right) = O(n^{-\gamma^*/(1+\gamma^*)} \log^2 n), \end{aligned} \quad (34)$$

using that $\Delta_k(0) \varepsilon_{k+t-s}$ are iid with $1 + \gamma^*$ th finite moment, $t - s > 0$ (see von Bahr and Esseen, 1965).

Next, using the same argument, (32) equals $-\frac{1}{2} \mathbb{E} |\varepsilon_1|^2 / 6 + \delta_{4n}$, where $\mathbb{E} |\delta_{4n}| = o(1)$, because by Lemma 1(b)

$$\begin{aligned} n^{-1} \sum_{t=1}^{n-1} t^{-1} \alpha_{t-1}(\eta_n) \sum_{k=t+1}^n 1 &= \sum_{t=1}^{n-1} t^{-1} \alpha_t(\eta_n) \frac{n-t}{n} \\ &= \sum_{t=1}^{n-1} t^{-1} \alpha_t(\eta_n) - n^{-1} \sum_{t=1}^{n-1} \alpha_t(\eta_n) \\ &= \sum_{t=1}^{n-1} t^{-2} + O(|\eta_n|) + O(n^{|\eta_n|-1} + n^{-1} \log n) \\ &= \frac{\pi^2}{6} + o(1) \end{aligned}$$

and

$$\begin{aligned} E|\delta_{4n}| &= \sum_{t=1}^{n-1} t^{-1} \sum_{s=0}^{t-1} \alpha_s(\eta_n) \{O(n^{-\gamma^*/(1+\gamma^*)} + n^{-1} \log n)\} \\ &= O(n^{-\gamma^*/(1+\gamma^*)} \log^2 n) = o(1). \end{aligned}$$

Finally (33) is also $O(n^{-\gamma^*/(1+\gamma^*)} \log^3 n)$, and this completes the proof of the lemma. \square

Lemma 9. *Let $\gamma^* = \min(\delta, \gamma)$ and let $r > (1 + \gamma^*)/[2(1 + 2\gamma^*)]$, $\Theta_n = \Theta n^{-r}$, $0 < \Theta < \infty$. Then under the assumptions of Proposition 2 and $H_{1n}(a)$*

$$\sup_{|\eta_n| \leq \Theta_n} \left| \sum_{t=1}^{n-1} t^{-1} y_{tn}(\eta_n) + \frac{1}{2} g(0) E|\varepsilon_1| \frac{\pi^2}{6} \theta \right| = o_p(1).$$

Proof. We have that

$$\begin{aligned} &\sum_{t=1}^{n-1} t^{-1} y_{tn}(\eta_n) \\ &= G(0) \sum_{t=1}^{n-1} t^{-1} n^{-1/2} \sum_{k=t+1}^n (\Delta_{k-t}(\eta_n u_{k-t-1}(\eta_n)) - \Delta_{k-t}(0)) \end{aligned} \quad (35)$$

$$+ \sum_{t=1}^{n-1} \frac{t^{-1}}{n^{1/2}} \sum_{k=t+1}^n \Delta_{k-t}(\eta_n u_{k-t-1}(\eta_n)) g(\theta_k \eta_n u_{k-1}(\eta_n)) \eta_n u_{k-1}(\eta_n), \quad (36)$$

where $\theta_k \in (0, 1)$. By the proof of Lemma 6, (35) is $o_p(1)$ uniformly over $|\eta_n| \leq \Theta_n$ since $r > (1 + \gamma^*)/[2(1 + 2\gamma^*)] > \frac{1}{4}$.

By the condition $|g(x) - g(0)| \leq C|x|^{\gamma^*}$, $\gamma^* = \min(\delta, \gamma)$, we can rewrite (36) as

$$g(0) \eta_n \sum_{t=1}^{n-1} t^{-1} n^{-1/2} \sum_{k=t+1}^n \Delta_{k-t}(\eta_n u_{k-t-1}(\eta_n)) u_{k-1}(\eta_n) + \rho_n, \quad (37)$$

where for $|\eta_n| \leq \Theta_n$

$$\begin{aligned} |\rho_n| &\leq C n^{1/2} |\eta_n|^{1+\gamma^*} \sum_{t=1}^{n-1} t^{-1} n^{-1} \sum_{k=1}^n |u_{k-1}(\eta_n)|^{1+\gamma^*} \\ &= O_p(n^{1/2} |\eta_n|^{1+\gamma^*} \log n) = o_p(1) \end{aligned}$$

for $r > (1 + \gamma^*)/[2(1 + 2\gamma^*)]$.

The main term of (37) can be represented as

$$g(0)\eta_n n^{1/2} \sum_{t=1}^{n-1} t^{-1} n^{-1} \sum_{k=t+1}^n \{\Delta_{k-t}(\eta_n) u_{k-t-1}(\eta_n) - \Delta_{k-t}(0)\} u_{k-1}(\eta_n) \quad (38)$$

$$+ g(0)\eta_n n^{1/2} \sum_{t=1}^{n-1} t^{-1} n^{-1} \sum_{k=t+1}^n \Delta_{k-t}(0) u_{k-1}(\eta_n). \quad (39)$$

By Lemma 7, the expectation of the supremum of (38) over $|\eta_n| \leq \Theta_n$ is $o(1)$.

Finally (39) is, under $H_{1n}(a)$,

$$-\frac{1}{2} g(0) E|\varepsilon_1| \frac{\pi^2}{6} \theta + \varsigma_n,$$

where $\sup |\varsigma_n| = o_p(1)$ by Lemma 8. \square

Acknowledgements

We thank the comments of an Associate Editor and two referees, which have led to an improved version of the paper. Research funded by the Ministerio de Educación y Ciencia reference numbers BEC2001-1270 and SEJ2004-04583/ECON.

References

- Anderson, T.W., 1971. *The Statistical Analysis of Time Series*. Wiley, New York.
- Andrews, G.E., Askey, R., Roy, R., 1999. *Special functions*. Encyclopedia of Mathematics and its Applications, vol. 71. Cambridge University Press, Cambridge.
- von Bahr, B., Esseen, C.G., 1965. Inequalities for the r th absolute moment of a sum of random variables, $1 \leq r \leq 2$. *Annals of Mathematical Statistics* 36, 299–303.
- Bai, J., Ng, S., 2002. A consistent test for conditional symmetry in time series models. *Journal of Econometrics* 103, 225–258.
- Beran, J., 1991. M-estimates of location for data with slowly decaying correlations. *Journal of the American Statistical Association* 86, 704–708.
- Beran, J., 1994. *Statistics for Long-Memory Processes*. Chapman & Hall, New York.
- Bloomfield, P., Steiger, W.L., 1983. *Least Absolute Deviation. Theory, Applications and Algorithms*. Birkhäuser, Boston.
- Boldin, M.V., 1995. On nonparametric sign tests in multiparameter autoregression. *Mathematical Methods of Statistics* 3, 279–305.
- Boldin, M.V., 1996. On nonparametric sign estimates in multiparameter autoregression. *Mathematical Methods of Statistics* 5, 154–172.
- Boldin, M.V., Stute, W., 2000. On nonparametric sign tests in ARMA models with possibly infinite error variances. *Theory of Probability and its Applications*, forthcoming.
- Boldin, M.V., Tyurin, Yu.N., 1994. On nonparametric sign estimates in autoregression models. *Mathematical Methods of Statistics* 3, 279–305.
- Boldin, M.V., Simonova, G.I., Tyurin, Yu.N., 1997. Sign-based methods in linear statistical models. *Translations of Mathematical Monographs*, vol. 163. American Mathematical Society, Providence.
- Breitung, J., Gourieroux, C., 1997. Rank tests for unit roots. *Journal of Econometrics* 81, 7–27.
- Campbell, B., Dufour, J.M., 1995. Exact nonparametric orthogonality and random walk tests. *Review of Economics and Statistics* 77, 1–16.

- Davis, R.A., Resnik, S.I., 1986. Limit theory for the sample covariance and correlation functions of moving averages. *Annals of Statistics* 14, 533–558.
- Hasan, M.N., Koenker, R.W., 1997. Robust rank tests of the unit root hypothesis. *Econometrica* 65, 133–161.
- Kokoszka, P.S., Taqqu, M.S., 1995. Fractional ARIMA with stable innovations. *Stochastic Processes and their Applications* 60, 19–47.
- Kokoszka, P.S., Taqqu, M.S., 1996a. The asymptotic behavior of quadratic forms in heavy-tailed strongly dependent random variables. *Stochastic Processes and their Applications* 66, 21–40.
- Kokoszka, P.S., Taqqu, M.S., 1996b. Parameter estimation for infinite variance fractional ARIMA. *Annals of Statistics* 24, 1880–1913.
- Kokoszka, P.S., Taqqu, M.S., 1999. Discrete time parametric models with long memory and infinite variance. *Mathematical and Computer Modelling* 29, 203–215.
- Koul, H.L., Surgailis, D., 1997. Asymptotic expansion of M-estimators with long-memory errors. *Annals of Statistics* 25, 818–850.
- Mikosch, T., Gadrich, C., Klüppelberg, C., Adler, R.J., 1995. Parameter estimation for ARMA models with infinite variance innovations. *Annals of Statistics* 23, 305–326.
- Parke, W.R., 1999. What is fractional integration? *The Review of Economics and Statistics* 81, 632–638.
- Phillips, P.C.B., 1995. Robust nonstationary regression. *Econometric Theory* 12, 912–951.
- Pollard, D., 1991. Asymptotics for least absolute deviation regression estimators. *Econometric Theory* 7, 186–199.
- Robinson, P.M., 1991. Tests for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics* 47, 67–84.
- Robinson, P.M., 1994. Efficient tests of nonstationary hypothesis. *Journal of the American Statistical Association* 89, 1420–1437.
- Robinson, P.M., Marinucci, D., 2001. Narrow-band analysis of nonstationary processes. *Annals of Statistics* 29, 947–986.
- Tanaka, K., 1999. The nonstationary fractional unit root. *Econometric Theory* 14, 549–582.
- Velasco, C., Robinson, P.M., 2000. Whittle pseudo-maximum likelihood estimates for non-stationary time series. *Journal of the American Statistical Association* 95, 1229–1243.
- Willinger, W., Taqqu, M.S., Leland, W.E., Wilson, D.V., 1995. Self-similarity in high-speed packet traffic: analysis and modeling of Ethernet traffic measurements. *Statistical Science* 10, 67–85.
- Willinger, W., Taqqu, M.S., Sherman, R., Wilson, D.V., 1997. Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level. *IEEE ACM Transactions on Networking* 5, 71–86.
- Wright, J.H., 1995. Stochastic orders of magnitude associated with two-stage estimators of fractional arima systems. *Journal of Time Series Analysis* 16, 119–126.
- Wright, J.H., 2000. Alternative variance-ratio tests using ranks and signs. *Journal of Business and Economics Statistics* 18, 1–9.
- Yohai, V.J., Maronna, R.A., 1977. Asymptotic behavior of least-squares estimates for autoregressive processes with infinite variances. *Annals of Statistics* 5, 554–560.