Article:

Iacone, Fabrizio orcid.org/0000-0002-2681-9036, Leybourne, Stephen J. and Taylor, A. M. Robert (2019) Testing the order of fractional integration of a time series in the possible presence of a trend break at an unknown point. Econometric Theory. 1201–1233. ISSN: 0266-4666

https://doi.org/10.1017/S0266466618000361

Testing the Order of Fractional Integration of a Time Series in the Possible Presence of a Trend Break at an Unknown Point*

Fabrizio Iacone[a,b], Stephen J. Leybourne[c] and A.M. Robert Taylor[d]

[a] Department of Economics and Related Studies, University of York

[b] Dipartimento di Economia, Management e Metodi Quantitativi, Università degli Studi di Milano

[c] School of Economics, University of Nottingham

[d] Essex Business School, University of Essex

August 15, 2018

Abstract

We develop a test, based on the Lagrange multiplier [LM] testing principle, for the value of the long memory parameter of a univariate time series that is composed of a fractionally integrated shock around a potentially broken deterministic trend. Our proposed test is constructed from data which are de-trended allowing for a trend break whose (unknown) location is estimated by a standard residual sum of squares estimator applied either to the levels or first differences of the data, depending on the value specified for the long memory parameter under the null hypothesis. We demonstrate that the resulting LM-type statistic has a standard limiting null chi-squared distribution with one degree of freedom, and attains the same asymptotic local power function as an infeasible LM test based on the true shocks. Our proposed test therefore attains the same asymptotic local optimality properties as an oracle LM test in both the trend break and no trend break environments. Moreover, this asymptotic local power function does not alter between the break and no break cases and so there is no loss in asymptotic local power from allowing for a trend break at an unknown point in the sample, even in the case where no break is present. We also report the results from a Monte Carlo study into the finite-sample behaviour of our proposed test.

Keywords: Fractional integration; trend break; Lagrange multiplier test; asymptotically locally most powerful test.

JEL classification: C22.

*We are grateful to the Editor, Peter Phillips, three anonymous referees and the Co-Editor, Anna Mikusheva, for their very helpful and constructive comments. Taylor gratefully acknowledges financial support provided by the Economic and Social Research Council of the United Kingdom under research grant ES/M01147X/1.

Correspondence to: Robert Taylor, Essex Business School, University of Essex, Colchester, CO4 3SQ, U.K. *E-mail*: `robert.taylor@essex.ac.uk`

1 Introduction

In this paper we consider the problem of testing for the order of integration, d say, of a fractionally integrated time series process that may be stationary or non-stationary around a deterministic trend function. Our point of departure from the extant literature is to allow for the possibility that the trend function is broken and, moreover, that the change in trend, should it occur, takes place at an unknown point in time. We follow the approach of Robinson (1994), Tanaka (1999) and Nielsen (2004) who construct Lagrange Multiplier [LM] test statistics in the frequency domain and time domain, respectively. These statistics are computationally convenient in that they avoid having to estimate the order of integration under the alternative.

For the case where the form of the deterministic kernel is known (which in the current context we interpret to mean that any putative break point in the deterministic trend function is taken as known, and that it is known whether a trend break is present or not), Robinson (1994), Tanaka (1999) and Nielsen (2004) show that residual-based variants of these LM tests are asymptotically locally most powerful against a class of (local) alternatives under Gaussianity and have asymptotic critical values given by the chi-squared distribution with one degree of freedom [χ_1^2], regardless of the value of the long memory parameter being tested. Although based on different and hence not directly comparable models, these large sample properties contrast with those of most popular *unit root* tests, such as that of Dickey and Fuller (1979), and *stationarity* tests, such as that of Kwiatkowski, Phillips, Schmidt and Shin (1992). In particular, the limiting null distributions of unit root and stationarity statistics tend to be non-standard and depend on the functional form of the fitted deterministic, differing between the no trend break and trend break cases, and dependent on the location of the trend break. Moreover, where a trend break is fitted but not actually present in the data, these tests show a considerable decline in asymptotic local power relative to the case where a break is not fitted.

In practice, both the location of a putative break point and, indeed, whether or not a trend break has even occurred will typically be unknown to the investigator. As a result, we therefore consider a residual-based LM-type test which allows for the possibility that a deterministic trend break occurs at an unknown point in the sample. The timing of the (putative) trend break is estimated by applying a conventional minimum residual sum of squares [RSS] criterion across all candidate break points to either the levels or first differences of the data depending on the value specified for d by the null hypothesis. Specifically, where $d < 0.5$ the levels data are used, while for $d > 0.5$ the first differences of the data are employed because, where a trend break occurs, this delivers an estimator for the trend break location whose rate of consistency is strictly faster than that of the levels based estimator.

Focussing our attention on the time domain approach of Tanaka (1999) and Nielsen (2004), we establish that, regardless of whether a trend break actually occurs or not, our proposed LM-type test inherits all of the desirable properties of the original LM test in the known deterministic case; that is, asymptotic local optimality together with asymptotic critical values from the χ_1^2 distribution. We demonstrate that this holds because where a trend break occurs, the location of the break is estimated at a sufficiently fast rate that it may be treated as known in large samples and, hence, reduces in the limit to the known deterministics case. Where a break does not occur, yet we fit a redundant trend break to the data, we show that this does not impact upon the asymptotic distribution of the statistic either under the null or under local alternatives. Although we consider the possibility of a single level break here, we conjecture that our asymptotic results will also pertain for the case of multiple possible trend breaks occurring at unknown points in the data.

The remainder of the paper is organised as follows. Section 2 sets out the fractionally integrated trend break model within which we work. Our proposed LM-type statistic for the case of an unknown trend break is described in section 3, where we also establish its large sample properties via comparison to an infeasible LM statistic based on the true errors rather than regression residuals. In section 4 we present a Monte Carlo simulation-based evaluation of the finite sample size and power properties of our LM-type test. An illustrative empirical application of our proposed tests to data on the monthly U.S. inflation rate is reported in section 5. Conclusions with some directions for future research are given in section 6. Proofs of our main results in Lemma 1 and Theorem 1 are provided in a mathematical appendix. Additional Monte Carlo results together with detailed proofs of preparatory lemmas used in the proofs of Lemma 1 and Theorem 1 are given in an accompanying online supplementary appendix associated with this article, available at Cambridge Journals Online (journals.cambridge.org/ect)

In what follows we use the following notation: '$x := y$' to indicate that x is defined by y ; '$\sim$' to denote that the ratio of the quantity on the left hand side to that on the right hand side of the symbol tends to 1 as the sample size tends to infinity; the operator '$\lfloor . \rfloor$' is used to denote the integer part of its argument; $\mathbb{I}(.)$ denotes the indicator function; L is used to denote the standard lag operator. Finally, we use $\xrightarrow{d}$ and $\xrightarrow{p}$ to denote convergence in distribution and in probability, respectively, in each case as the sample size diverges.

2 The Fractionally Integrated Trend Break Model

We consider the following model for the scalar random variable x_t,

$$x_t = \beta_1 + \beta_2 t + \beta_3 DT_t(\tau^*) + e_t, \quad t = 1, ..., T. \tag{2.1}$$

The shocks, e_t, are assumed to follow a zero mean, type 2 fractionally integrated process of order d, denoted $e_t \in I(d)$; precise assumptions will be stated below. We will assume that $d \in (-0.5, 0.5) \cup (0.5, 1.5)$.[1] Both (asymptotically) stationary, non-stationary, and fractionally over-differenced time series are therefore permitted within our set-up. In (2.1), the deterministic trend break term, $DT_t(\tau^*)$, is defined for a generic τ as $DT_t(\tau) := (t - \lfloor \tau T \rfloor)\mathbb{I}(t \geq \lfloor \tau T \rfloor)$. This corresponds to the deterministic kernel considered in Model B of Perron (1989), the so-called "changing growth" model, which allows

[1]For technical reasons, discussed further in Remark 11 below, we will not formally derive the large sample properties of tests of the null hypothesis that $d = 0.5$ in this paper and for that reason $d = 0.5$ is excluded from the range of values d that we consider in (2.1). We do, however, investigate the finite sample behaviour of our proposed methodology when applied to the case of testing the null hypothesis of $d = 0.5$ via Monte Carlo simulation methods in section 4.

for a change in the slope of the trend function without a change in the level at the time of the break.[2] Where a trend break occurs, i.e. where $\beta_3 \neq 0$, we assume that the true trend break fraction is such that $\tau^* \in [\tau_L, \tau_U] =: \Lambda \subset [0,1]$, where the quantities τ_L and τ_U are trimming parameters below and above which, respectively, a trend break is deemed not to occur. A negative (positive) trend break occurs when $\beta_3 < 0$ ($\beta_3 > 0$).

Writing $d =: d_0 + \theta$, where d is the true (unknown) value of the long memory parameter in (2.1), our interest in this paper focuses on testing the null hypothesis that $d = d_0$; that is, $H_0 : \theta = 0$ in (2.1). Under H_0 we therefore have that $e_t \in I(d_0)$. As in Robinson (1994) and Tanaka (1999), we will focus attention on local alternatives whereby $H_c : \theta := \theta_T = c/\sqrt{T}$, with c a constant. Notice that H_c reduces to H_0 when $c = 0$. More generally, c is the Pitman drift for this testing problem and, as we will later demonstrate, will determine the asymptotic local power of the test. Unless otherwise stated, all of the large sample results provided in this paper are based on the assumption that H_c holds on (2.1) for some value of the constant c.

Our model is completed by formalising the properties of e_t. For $t > 0$, e_t is taken to follow the fractionally integrated process

$$e_t := \sum_{s=1}^{t} \Delta_{t-s}^{(d)} \eta_s \tag{2.2}$$

where, for any $d \in (-0.5, 0.5) \cup (0.5, 1.5)$, $\Delta_t^{(d)} := \Gamma(t+d) / (\Gamma(d) \Gamma(t+1))$, with $\Gamma(\cdot)$ denoting the Gamma function, with the convention that $\Gamma(0) := \infty$ and $\Gamma(0)/\Gamma(0) := 1$. In view of the expansion $(1-L)^{-d} = \sum_{t=0}^{\infty} \Delta_t^{(d)} L^t$, the definition in (2.2) can also be written as $e_t = \Delta^{-d}\{\eta_t \mathbb{I}(t > 0)\}$. To simplify notation, and following Johansen and Nielsen (2010), we also introduce the operator Δ_+^{α} so that, for a generic α and a generic series ξ_t, $\Delta_+^{\alpha} \xi_t := \Delta^{\alpha}\{\xi_t \mathbb{I}(t > 0)\}$, and therefore $e_t = \Delta_+^{-d} \eta_t$. The model for e_t is completed by assuming $e_t = 0$ when $t \leq 0$. In common with the earlier contributions to this literature in Robinson (1994), Tanaka (1999) and Nielsen (2004), we therefore assume that e_t is a so-called "type 2" fractionally integrated process.

Finally, η_t in (2.2) is assumed to be a zero mean, stationary process with spectral density that is absolutely continuous and strictly positive at all frequencies with long run variance $\sigma_\infty^2 := \sum_{h=-\infty}^{\infty} \mathrm{E}\left(\eta_t \eta_{t+h}\right)$. More precisely, we make the following assumption regarding η_t.

Assumption 1 *Let $\{\eta_t\}$ be generated by the finite-order ARMA(p,q) process, $a(L)\eta_t = b(L)\varepsilon_t$, satisfying the following conditions: (a) the polynomials $a(z) := 1 - a_1 z - \cdots - a_p z^p$ and $b(z) := 1 - b_1 z - \cdots - b_q z^q$ contain no common factors and are such that $a(z) \neq 0$ and $b(z) \neq 0$ for $|z| \leq 1$, and the innovation process ε_t is such that $\varepsilon_t \sim i.i.d.\big(0, \sigma_\varepsilon^2\big)$ with $0 < \sigma_\varepsilon^2 < \infty$; and (b) the following*

[2] We exclude the possibility of a simultaneous break in level (cf Model C of Perron, 1989). Where testing the null value of d larger that 0.5, our test procedure will be based on the first differences of the data, and hence a simultaneous level break would be reduced to an outlier, which would have no effect on the asymptotic properties of the test when the null hypothesis holds. However, for testing null values less than 0.5, the presence of a simultaneous level break reduces the rate of convergence of the estimate of the break location under the null to below that required for the test to have a pivotal limiting distribution; see Chang and Perron (2016) for details and further discussion.

higher-order moment conditions hold on ε_t*,* $\mathrm{E}\left|\varepsilon_t\right|^{\overline{q}} < \infty$ *for* $\overline{q} > \max\left(2, 2/\left(1+2d\right)\right)$ *if* $d \in (-0.5, 0.5)$*,* $\overline{q} > \max\left(2, 2/\left(2d-1\right)\right)$ *if* $d \in (0.5, 1.5)$*.*

Remark 1. The requirement in part (a) of Assumption 1 that η_t follows a stationary and invertible finite-ordered ARMA process with no common factors is fairly standard in this literature; see, for example, Tanaka (1999) or Nielsen (2004). Under these conditions, η_t has strictly positive and bounded spectral density at all frequencies and it is therefore $I(0)$. The higher-order moment conditions placed on ε_t in part (b) of Assumption 1 would not be required in cases where the true trend break date, τ^*, was known. However, where τ^* is unknown and must be estimated from the data then, as we shall see below, a functional central limit [FCLT] theorem result will be needed on the estimates of the β_j, $j = 1, 2, 3$, parameters characterising the deterministic component. As Johansen and Nielsen (2012) show, this requires moment conditions like those given in part (b) of Assumption 1 to hold on ε_t.

Remark 2. Assumption 1 imposes an i.i.d. condition on the innovations, ε_t. This assumption is in common with many of the published papers in the trend break and long memory literatures, a number of which we draw upon for auxiliary results in establishing the large sample properties of our proposed testing procedures. In particular, the result given in part (i) of our main result in Theorem 1 below for the case where a trend break occurs, $\beta_3 \neq 0$, relies on the convergence rates for the trend break fraction estimator used in our procedure and these have been derived under an i.i.d. assumption on the innovations in Lavielle and Moulines (2000) and Cheung and Perron (2016). Similarly, the result in part (ii) of Theorem 1 relating to the no break case, $\beta_3 = 0$, relies on an application of the FCLT established in Marinucci and Robinson (2000), again established under an i.i.d. assumption on the innovations. While it seems plausible that the results given in this paper would continue to hold under a weaker conditionally homoskedastic martingale difference assumption, formally establishing whether this is true or not is beyond the scope of the present paper as it would also require establishing that the results in these auxiliary papers also carry over to the case of martingale difference innovations.

3 Lagrange Multiplier Tests

As background motivation in section 3.1, we first briefly review the construction of the LM test for H_0 in cases where e_t in (2.1) is observable; that is, where the true values of β_i, $i = 1, 2, 3$, are all known and, where the true value of β_3 is non-zero, the trend break location τ^* is also known. In section 3.2 we then discuss how the LM testing principle can be generalised to the case where the true values of these parameters are not known and, hence, the test statistic needs to be based on regression residuals.

3.1 An Infeasible LM Test

Where e_t is observable, the LM statistic for testing H_0, under the assumption that η_t is Gaussian, obtains directly from Nielsen (2004), *inter alia*. We demonstrate the derivation of the LM statistic in the particular case in which $\eta_t = \varepsilon_t$ is a normally, independently distributed sequence. Then, for σ_ε^2,

θ, we consider the likelihood

$$\mathcal{L}\left(\sigma_{\varepsilon}^{2},\theta\right)=-\frac{T}{2}\ln\left(2\pi\right)-\frac{T}{2}\ln\left(\sigma_{\varepsilon}^{2}\right)-\frac{1}{2\sigma_{\varepsilon}^{2}}\sum_{t=1}^{T}\left(\Delta_{+}^{d_0+\theta}e_t\right)^2$$

see also Equation (38), and the subsequent discussion, in Tanaka (1999). Estimating $\widehat{\sigma}^2:=\frac{1}{T}\sum_{t=1}^{T}\left(\Delta_{+}^{d_0+\theta}e_t\right)^2$, the concentrated likelihood is given by

$$\mathcal{L}\left(\theta\right)=-\frac{T}{2}\ln\left(2\pi\right)-\frac{T}{2}\ln\left(\frac{1}{T}\sum_{t=1}^{T}\left(\Delta_{+}^{d_0+\theta}e_t\right)^2\right)-\frac{T}{2}$$

and

$$\frac{\partial\mathcal{L}\left(\theta\right)}{\partial\theta}=-\frac{T}{2}\frac{1}{\widehat{\sigma}^2}\frac{2}{T}\sum_{t=1}^{T}\left(\left\{\ln\left(\Delta\right)\Delta^{d_0+\theta}\right\}_{+}e_t\right)\left(\Delta_{+}^{d_0+\theta}e_t\right)$$

where the operator $-\ln\left(\Delta\right)$ admits the expansion $-\ln\left(\Delta\right)=\sum_{j=1}^{\infty}j^{-1}L^j$, as for a Taylor-series expansion for $-\ln\left(1-x\right)$ around $x=1$, and where for a generic series ξ_t we introduce the operator $\left\{-\ln\left(\Delta\right)\right\}_{+}$ so that $\left\{-\ln\left(\Delta\right)\right\}_{+}\xi_t:=-\ln\left(\Delta\right)\left\{\xi_t\mathbb{I}\left(t>0\right)\right\}$ and therefore $\left\{-\ln\left(\Delta\right)\right\}_{+}\xi_t=\sum_{j=1}^{t-1}j^{-1}\xi_{t-j}$. Consequently, defining $r_j:=\widehat{\sigma}^{-2}T^{-1}\sum_{t=1}^{T-j}\left(\Delta_{+}^{d_0+\theta}e_t\right)\left(\Delta_{+}^{d_0+\theta}e_{t+j}\right)$, we have that $\frac{\partial\mathcal{L}(\theta)}{\partial\theta}=T\sum_{t=1}^{T-1}j^{-1}r_j$. Moreover, defining $v_t:=\sum_{j=1}^{t-1}j^{-1}\varepsilon_{t-j}$, we have, under H_0, that

$$\left.\frac{\partial\mathcal{L}\left(\theta\right)}{\partial\theta}\right|_{\theta=0}=\frac{1}{\widehat{\sigma}^2}\sum_{t=1}^{T}v_t\varepsilon_t\qquad\text{and}\qquad\frac{\sqrt{T}}{T}\left.\frac{\partial\mathcal{L}\left(\theta\right)}{\partial\theta}\right|_{\theta=0}\xrightarrow{d}N\left(0,\frac{\pi^2}{6}\right)$$

by a central limit theorem [CLT] for martingale difference sequences. Finally, following Tanaka (1999,p.561), $\lim_{T\to\infty}\mathrm{E}\left(-\frac{1}{T}\left.\frac{\partial^2\mathcal{L}(\theta)}{\partial\theta^2}\right|_{\theta=0}\right)=\frac{\pi^2}{6}$ and so, in view of the asymptotic orthogonality between the estimates of σ^2 and of θ (see also Nielsen, 2004, p.125), the LM statistic can be written as $6T/\pi^2\left(\sum_{t=1}^{T-1}j^{-1}r_j\right)^2$.

The LM statistic above is derived under the assumption that η_t is independently distributed, such that ε_t is observable under H_0. In the more realistic case in which a generic ARMA structure is assumed for η_t, its parameters must be estimated and the test statistic corrected to take these into account. To that end, defining $g\left(z;\psi\right):=a\left(z\right)b^{-1}\left(z\right)$, we can estimate the parameter vector $\psi^*:=\left(a_1,...,a_p,b_1...,b_q\right)'$ under H_0 as

$$\widehat{\psi}:=\arg\min_{\psi\in\Theta}\sum_{t=1}^{T}\left(g\left(L;\psi\right)\Delta_{+}^{d_0}e_t\right)^2. \tag{3.1}$$

Throughout the paper the regularity condition that Θ is a $\mathbb{R}^{p+q}$ compact space of parameters for an $ARMA(p,q)$ model, such that the $ARMA$ processes corresponding to parameters in Θ are stationary and invertible with no common factors, will be taken to hold. Then, based on the estimate $\widehat{\psi}$, we construct the quantities

$$\widehat{\varepsilon}_t:=g\left(L;\widehat{\psi}\right)\Delta_{+}^{d_0}e_t,\ \widehat{s}^2:=T^{-1}\sum_{t=1}^{T}\widehat{\varepsilon}_t^2,\ \widehat{r}_j:=\widehat{s}^{-2}T^{-1}\sum_{t=1}^{T-j}\widehat{\varepsilon}_t\widehat{\varepsilon}_{t+j},\ \widehat{A}:=\sum_{j=1}^{T-1}j^{-1}\widehat{r}_j. \tag{3.2}$$

Defining g_j as the coefficient on z^j in the expansion of $\partial \ln g(z;\psi)/\partial\psi|_{\psi=\psi^*}$, and setting

$$\kappa := \sum_{j=1}^{\infty} g_j j^{-1}, \quad \Phi := \sum_{j=0}^{\infty} g_j g_j', \quad \omega^2 := \pi^2/6 - \kappa'\Phi^{-1}\kappa$$

then, as demonstrated in Theorem 3.3 of Tanaka (1999), under H_c and the conditions given in part (a) of Assumption 1 we have that $T^{1/2}\widehat{A} \xrightarrow{d} N\left(c\omega^2, \omega^2\right)$. As discussed in Nielsen (2004, p.132), an estimator of ω^2 which is consistent under H_c is obtained on substituting the estimates from $\widehat{\psi}$ into the expressions for κ and Φ above; we denote this estimator by $\widehat{\omega}^2$. The resulting LM statistic is then given by

$$LM := T\frac{\widehat{A}^2}{\widehat{\omega}^2}. \tag{3.3}$$

Under the conditions of part (a) of Assumption 1 and the local alternative H_c,

$$LM \xrightarrow{d} \chi_1^2\left(c^2\omega^2\right) \tag{3.4}$$

where $\chi_1^2\left(c^2\omega^2\right)$ indicates a χ_1^2 distribution with non-centrality parameter $c^2\omega^2$; see, *inter alia*, Theorem 4.2 of Nielsen (2004, p.132).

Remark 3. A one-sided test could also be considered, based on the one-sided score statistic $S := \left(\frac{T}{\widehat{\omega}^2}\right)^{1/2}\widehat{A}$, as in Robinson (1994, pp. 1424,1426). This would allow testing, for example, the unit root null hypothesis, $d_0 = 1$, against the alternative $d_0 < 1$. Such tests will be more powerful than the two-sided LM test based on LM, against one-sided alternatives (in the correct tail). Indeed, under Gaussianity, the one-sided score test is asymptotically uniformly most powerful (UMP). Under H_0, $S \xrightarrow{d} N(0,1)$.

Remark 4. As discussed in Nielsen (2004, p.126) the foregoing LM statistic for the null hypothesis H_0, is asymptotically equivalent under H_c to the corresponding Wald and Likelihood Ratio statistics for testing H_0. Moreover, as discussed in Robinson (1994) and Nielsen (2004), the tests based on these statistics are (locally) optimal in the sense that under Gaussianity they achieve a limiting non-central χ_1^2 distribution with the maximal available non-centrality parameter and are therefore locally most powerful. However, it should be stressed that Gaussianity is not required as part of the conditions stated in part (a) of Assumption 1 to establish the large sample convergence result in (3.4).

3.2 Feasible LM-type Tests Based on Regression Residuals

We now consider the case of practical relevance where e_t is unobserved and so the LM statistic must be constructed from regression residuals, rather than from e_t. We will show that a feasible statistic can still be designed, and that it is asymptotically equivalent to the infeasible LM statistic in (3.3).

Where the true (potential) trend break location, τ^*, in (2.1) is known, then the form of the deterministic component is known to the practitioner, up to the unknown parameters β_j, $j = 1, 2, 3$, and, hence, lies within the non-stochastic regressors set-up considered by Robinson (1994) and Nielsen (2004). These authors show how to construct a feasible LM statistic for H_0 in this case which attains a $\chi_1^2\left(c^2\omega^2\right)$ limiting distribution under H_c provided the conditions of part (a) of Assumption 1 hold, with this result holding regardless of the true values of β_j, $j = 1, 2, 3$, so that, in particular, the same limiting results hold in both the trend break and no trend break environments. Our focus in this paper is, however, the more realistic setting where τ^* is unknown to the practitioner. In place of τ^* we will therefore need to build our test statistic around a suitable estimate of τ^*. An immediate implication of doing so, however, is that the assumption of non-stochastic regressors required by Robinson (1994) and Nielsen (2004) is no longer met. Indeed, accounting for this difference is the primary purpose of this paper.

An obvious estimator of τ^* to use is the minimum RSS estimator, $\widehat{\tau}$ say, which minimises the RSS over the sequence of levels regressions of x_t on $(1, t, DT_t(\tau))'$, taken across all $\tau \in \Lambda$. Where a trend break occurs, so that the true value of β_3 is non-zero, at time τ^*, then the properties of $\widehat{\tau}$ depend on the order of integration of e_t. In particular, Chang and Perron (2016) show that when $e_t \in I(d)$, $d \in (-0.5, 0.5) \cup (0.5, 1.5)$ then $\widehat{\tau} - \tau^* = O_p\left(T^{-3/2+d}\right)$. However, for the equivalent problem of searching for a level break in the first differences of the data, we obtain from Lavielle and Moulines (2000) that when $d \in (0.5, 1.5)$ and $\widehat{\tau}$ is now defined as the estimator which minimises the RSS over the sequence of regressions in first differences of Δx_t on $(1, DU_t(\tau))'$, where $DU_t(\tau) := \mathbb{I}(t \geq \lfloor \tau T \rfloor)$, then $\widehat{\tau} - \tau^* = O_p\left(T^{-1}\right)$. A faster rate of consistency can therefore be obtained by using the first differences-based RSS estimator when $d > 1/2$. In view of these rates of consistency, we will undertake the estimation of τ^*, and the consequent estimation of β_1, β_2 and β_3 and, hence, e_t, using two different regression models, whose form depends on the value of d_0 specified under the null hypothesis, as follows:

Model A: For $d_0 \in (-0.5, 0.5)$, we let $y_t := x_t$ and use the levels form representation of (2.1):

$$y_t = \beta_1 + \beta_2 t + \beta_3 DT_t(\tau^*) + u_t, \quad t = 1, ..., T, \quad u_t \in I(d)$$

where $u_t := e_t$ and, under H_0, $d = d_0$.

Model B: For $d_0 \in (0.5, 1.5)$, we let $y_t := \Delta x_t$ and use the first-differenced transformation of (2.1):

$$y_t = \beta_2 + \beta_3 DU_t(\tau^*) + u_t, \quad t = 2, ..., T, \quad u_t \in I(d-1)$$

where $u_t := \Delta e_t$, and, under H_0, $d = d_0$.

Remark 5. Taken together, Models A and B allow us to consider inference on the long memory parameter in (2.1) in the presence of a possibly broken trend for hypothesised values of the long memory parameter in the range $d_0 \in (-0.5, 0.5) \cup (0.5, 1.5)$. It is worth noting that we will not explicitly consider tests for null hypotheses which impose $d_0 > 1.5$ in (2.1). Here the resulting test statistics would be identical to the statistics of the form given in section 3.1 on substituting $\Delta_+^{d_0} e_t$ for $\Delta_+^{d_0} x_t$. This is the case because taking d_0th differences annihilates the deterministic trend component in (2.1) when $d_0 > 1.5$. For $d_0 > 1.5$ the deterministic trend component will therefore have no impact on the large sample behaviour of these statistics which coincide with that given for LM in (3.4).

Remark 6. It is also worth commenting that although Robinson (1994) and Nielsen (2004) do not restrict d_0 to lie in a particular interval, they instead assume that sufficient rate conditions hold on the estimates of the parameters characterising the deterministic trend function; see Robinson (1994,p.1434) and Equation (12) of Nielsen (2004). In these papers, the fractional differences of the disturbances from (2.1) taken under the null hypothesis, that is $\Delta_+^{d_0} e_t$, are estimated using the residuals from the regression of $\Delta_+^{d_0} x_t$ onto the $\Delta_+^{d_0}$ differences of the deterministic kernel. Replacing $\Delta_+^{d_0} e_t$ by these residuals in (3.2), yields an estimate of $\widehat{\varepsilon}_t$ and, proceeding as in (3.2) and (3.3), it is then possible to compute a feasible version of the LM statistic based on these residuals. Under the regularity conditions detailed in Robinson (1994) or Nielsen (2004), doing so yields a feasible LM statistic that has the same limiting distribution as the infeasible LM statistic. Establishing such regularity conditions is straightforward in many cases, such as where the deterministic component is a polynomial trend, but is considerably more complicated in the case considered in this paper where we allow for the possibility that a trend break occurs at an unknown point in the sample. Here we need to establish the uniform (in τ) rate result for the estimated coefficients of the deterministic trend function given in (3.11) of Lemma 1 in the case where no trend break occurs, and the corresponding rate result in (3.14) of Lemma 1 for where a break does occur. Moreover, where a trend break occurs, we also need to ensure that the estimate of τ^* is consistent at a sufficiently fast rate, as is done in (3.12) and (3.13) of Lemma 1 below. Establishing the results stated in Lemma 1 requires a functional central limit theorem to hold, which in turn requires that $d > -0.5$. We note that the restriction that $d > -0.5$ is also imposed in Chang and Perron (2016) when establishing properties for the estimates of τ^* and of β_1, β_2 and β_3 which they consider.

In each of Model A and B we will also need to consider two scenarios, depending on whether the trend break is in fact present or not; that is, whether $\beta_3 = 0$ or $\beta_3 \neq 0$. To that end, and in order to discuss Models A and B simultaneously, we now introduce some common notation, noting that in the case of Model B, β_1 is not estimated. This notation is indexed by a generic value of $\tau \in \Lambda$. In the context of Model A we define $z_t(\tau) := (1, t, DT_t(\tau))'$ and $\beta := (\beta_1, \beta_2, \beta_3)'$, whereas in the context of Model B we define $z_t(\tau) := (1, DU_t(\tau))'$ and $\beta := (\beta_2, \beta_3)'$. Finally, we define the OLS estimate of β (under Model A or Model B, as appropriate) as

$$\widehat{\beta}(\tau) := \left(\sum_{t=j}^{T} z_t(\tau) z_t(\tau)'\right)^{-1} \left(\sum_{t=j}^{T} z_t(\tau) y_t\right) \tag{3.5}$$

where $j = 1$ in the case of Model A, and $j = 2$ for Model B. We then define the corresponding de-trended residuals as

$$\widehat{u}_t(\tau) := y_t - z_t(\tau)' \widehat{\beta}(\tau) \tag{3.6}$$

for $t = 1, ..., T$ in the case of Model A, and for $t = 2, ..., T$ in the case of Model B. For Model B, we set $\widehat{u}_1(\tau) := 0$, so that $\widehat{u}_t(\tau)$ is defined for $t = 1, ..., T$ in both cases.

Under H_0, we can estimate η_t by taking the corresponding fractional differences of these OLS de-trended residuals, as $\Delta_+^{\delta_0} \widehat{u}_t(\tau)$, for $\delta_0 := d_0$ when Model A is used, and for $\delta_0 := d_0 - 1$ when

Model B is used, for a specific value of τ. Proceeding as in the infeasible case, for any τ we can then estimate $\widehat{\psi}(\tau)$ via

$$\widehat{\psi}(\tau) := \arg\min_{\psi\in\Theta} \sum_{t=1}^{T} \left(g(L;\psi)\,\Delta_{+}^{\delta_0}\widehat{u}_t(\tau) \right)^2 \tag{3.7}$$

and use this to compute the quantities

$$\widehat{\varepsilon}_t(\tau) := g\left(L;\widehat{\psi}(\tau)\right)\Delta_{+}^{\delta_0}\widehat{u}_t(\tau) \tag{3.8}$$

and

$$\widehat{s}(\tau)^2 := T^{-1}\sum_{t=1}^{T}\widehat{\varepsilon}_t^2(\tau)\,,\ \widehat{r}_j(\tau) := \widehat{s}(\tau)^{-2}\,\frac{1}{T}\sum_{t=1}^{T-j}\widehat{\varepsilon}_t(\tau)\,\widehat{\varepsilon}_{t+j}(\tau)\,,\quad \widehat{A}(\tau) := \sum_{j=1}^{T-1} j^{-1}\widehat{r}_j(\tau)\,.$$

Given $\widehat{\psi}(\tau)$, we also compute $\widehat{\omega}^2(\tau)$ yielding the LM-type statistic

$$LM(\tau) := T\frac{\widehat{A}^2(\tau)}{\widehat{\omega}^2(\tau)}. \tag{3.9}$$

If the true break fraction, τ^*, was known then one would simply evaluate $LM(\tau)$ of (3.9) at $\tau = \tau^*$; the resulting statistic, $LM(\tau^*)$, would for either $d_0 = 0$ or $d_0 = 1$ coincide with the statistic from Robinson (1994), discussed at the start of this subsection. Our focus, however, is on the case where τ^* is unknown and, following the earlier discussion, our proposed test will be based on evaluating $LM(\tau)$ at $\widehat{\tau}$, the minimum RSS estimate

$$\widehat{\tau} := \arg\min_{\tau\in\Lambda} \textstyle\sum_{t=1}^{T}\left(\widehat{u}_t(\tau)\right)^2 \tag{3.10}$$

whose form is determined according to the value of d_0 being tested under the null hypothesis, H_0. Specifically, if d_0 lies in the region $(-0.5, 0.5)$ then we estimate τ^* using the levels of the data and test the null hypothesis that the long memory parameter in the levels data is d_0, whereas if d_0 lies in the range $(0.5, 1.5)$ we instead estimate τ^* using the first differences of the data and test the null hypothesis that the long memory parameter in the first differenced data is $d_0 - 1$.

In Theorem 1 below we will determine the large sample behaviour of $LM(\widehat{\tau})$ by comparing it to the infeasible LM statistic, LM of (3.3). Inherent in doing so will be to analyse the distance between $\widehat{\varepsilon}_t$ and $\widehat{\varepsilon}_t(\widehat{\tau})$, the latter given by $\widehat{\varepsilon}_t(\tau)$ in (3.8) evaluated at $\tau = \widehat{\tau}$, and establish how this affects the distance between $LM(\widehat{\tau})$ and LM. The behaviour of $LM(\widehat{\tau})$ clearly depends on the large sample properties of the estimates $\widehat{\tau}$ of (3.10) and $\widehat{\beta}(\widehat{\tau})$, the latter given by $\widehat{\beta}(\tau)$ of (3.5) evaluated at $\tau = \widehat{\tau}$. Consequently, in Lemma 1 we first establish these results under H_c both for the case where a trend break occurs ($\beta_3 \neq 0$) and where a trend break does not occur ($\beta_3 = 0$). Theorem 1 will then subsequently establish that these properties are sufficient to allow us to show that the difference, $LM(\widehat{\tau}) - LM$, is asymptotically negligible, regardless of whether or not a trend break occurs.

Lemma 1 *Let x_t be generated by (2.1) under $H_c : \theta := \theta_T = c/\sqrt{T}$, and let Assumption 1 hold. For $d_0 \in (-0.5, 0.5)$, define, for generic α, the diagonal matrix $K_T(\alpha) := diag\left\{T^{1/2-\alpha}, T^{3/2-\alpha}, T^{3/2-\alpha}\right\}$, whereas for $d_0 \in (0.5, 1.5)$, define the diagonal matrix $K_T(\alpha) := diag\left\{T^{3/2-\alpha}, T^{3/2-\alpha}\right\}$. Then the following results hold:*

(i) Where $\beta_3 \neq 0$, the estimates $\widehat{\tau}$ of (3.10) and $\widehat{\beta}(\widehat{\tau})$, the latter given by (3.5) evaluated at $\tau = \widehat{\tau}$, are such that

$$K_T(d_0)\left(\widehat{\beta}(\widehat{\tau}) - \beta\right) = O_p(1) \tag{3.11}$$

and

$$\widehat{\tau} - \tau^* = O_p\left(T^{d_0-3/2}\right) \text{ if } d_0 \in (-0.5, 0.5) \tag{3.12}$$

$$\widehat{\tau} - \tau^* = O_p\left(T^{-1}\right) \text{ if } d_0 \in (0.5, 1.5) \tag{3.13}$$

(ii) Where $\beta_3 = 0$, the estimate $\widehat{\beta}(\tau)$ of (3.5) is such that, for $d_0 \in (-0.5, 0.5) \cup (0.5, 1.5)$,

$$K_T(d_0)\left(\widehat{\beta}(\tau) - \beta\right) = O_p(1), \tag{3.14}$$

uniformly in τ.

Remark 7. The result in part (ii) of Lemma 1 shows that when no break occurs, the (centred and appropriately scaled) OLS estimator of β from (3.5) converges to a well-defined limiting distribution and that this holds uniformly in τ. This uniform convergence then implies that it must also hold on replacing τ with $\widehat{\tau}$, even though the latter is a random variable (even asymptotically).

Remark 8. The additional higher order moment conditions stipulated in part (b) of Assumption 1 are required for two reasons. Firstly, when $\beta_3 \neq 0$, estimation of τ^* exploits a FCLT; see Chang and Perron (2016). Secondly, in the case where $\beta_3 = 0$, then a FCLT theorem is used to establish that the rate given in (3.14) holds uniformly in τ. Monte Carlo simulation results are reported in the accompanying on-line supplement which investigate the consequences of violating these conditions. The results suggest that violation of the moment condition can inflate the empirical size of the test, the more so the greater the degree of departure from the stated assumption.

In Theorem 1 we now state our main result, establishing the large sample behaviour of the LM-type statistic $LM(\widehat{\tau})$.

Theorem 1 *Let the conditions of Lemma 1 hold. Then, for $d_0 \in (-0.5, 0.5) \cup (0.5, 1.5)$:*

(i) If $\beta_3 \neq 0$, then $LM(\widehat{\tau}) - LM = o_p(1)$.

(ii) If $\beta_3 = 0$, then $LM(\tau) - LM = o_p(1)$, uniformly in τ.

Some remarks are in order.

Remark 9. An immediate consequence of Theorem 1 is that $LM(\widehat{\tau}) - LM = o_p(1)$ irrespective of whether $\beta_3 \neq 0$ or $\beta_3 = 0$. Consequently, regardless of the value of β_3, $LM(\widehat{\tau}) \xrightarrow{d} \chi_1^2(c^2\omega^2)$ under H_c, thereby retaining asymptotic optimality. Moreover, since $LM(\widehat{\tau}) \xrightarrow{d} \chi_1^2$ under H_0, standard critical values can still be used.

Remark 10. The result given in part (i) of Theorem 1 demonstrates that when $\beta_3 \neq 0$, such that a trend break does occur, the difference between the LM-type statistics based on $\widehat{\varepsilon}_t$ and $\widehat{\varepsilon}_t(\widehat{\tau})$ is asymptotically negligible. This arises because $\widehat{\tau} \xrightarrow{p} \tau^*$ at a sufficiently fast rate; cf. part (i) of Lemma 1. Part (ii) of Theorem 1 shows that when no break occurs, the difference between the LM-type statistics based on $\widehat{\varepsilon}_t$ and $\widehat{\varepsilon}_t(\tau)$ is asymptotically negligible, and that this holds uniformly in τ and, hence, holds for $\widehat{\tau}$.

Remark 11. It is important to acknowledge that, in common with the results given in Lavielle and Moulines (2000) and Chang and Perron (2016), Theorem 1 does not cover the case of $d_0 = 0.5$ and so testing the null hypothesis that the long memory parameter is equal to 0.5 is formally excluded from our analysis; cf. footnote 1. When $\beta_3 \neq 0$, as noted in Remark 10, the proof of Theorem 1 is based on establishing that the difference between the LM-type statistics based on $\widehat{\varepsilon}_t$ and $\widehat{\varepsilon}_t(\widehat{\tau})$ is asymptotically negligible. A key part of the derivation of the theorem is proving that $\widehat{A} - \widehat{A}(\widehat{\tau}) = o_p(T^{-1/2})$ and, as the difference $\widehat{\varepsilon}_t - \widehat{\varepsilon}_t(\widehat{\tau})$ depends on the term $\Delta_+^{d_0}(DT_t(\widehat{\tau}) - DT_t(\tau^*))$, on showing that $\sum_{t=1}^{T}(\sum_{j=1}^{t-1} j^{-1}\Delta_+^{d_0}(DT_{t-j}(\widehat{\tau}) - DT_{t-j}(\tau^*)))\widehat{\varepsilon}_t = o_p\left(T^{1/2}\right)$. The remainder term $\Delta_+^{d_0}(DT_t(\widehat{\tau}) - DT_t(\tau^*))$ is a random variable which is potentially correlated with ε_t and, hence, with $\widehat{\varepsilon}_t$. In order to allow for this correlation, we exploit the fact that $DT_t(\widehat{\tau}) - DT_t(\tau^*)$ follows a (broken) trend, and we use a method of proof based on summation by parts. However, the bound that we can establish on $\widehat{A} - \widehat{A}(\widehat{\tau})$ in this way is weaker the larger is d_0, until for $d_0 = 0.5$ it is not sufficient to establish the required $o_p(T^{1/2})$ bound; we refer the reader to Lemma C2 and Lemma D2 in the proof for further details.[3] We will nonetheless include $d_0 = 1/2$ in the Monte Carlo exercise in section 4.

Remark 12. In parallel with the discussion in Remark 3 above, a one-sided test could also be considered based on the score-type statistic $S(\widehat{\tau}) := \left(\frac{T}{\widehat{\omega}^2(\widehat{\tau})}\right)^{1/2}\widehat{A}(\widehat{\tau})$. The large sample theory for $S(\widehat{\tau})$ follows from the results given in this paper; in particular, under H_0, $S(\widehat{\tau}) \xrightarrow{d} N(0,1)$.

Remark 13. Theorem 1 demonstrates that the $LM(\widehat{\tau})$ test has non-trivial asymptotic local power. We conjecture that the test is also consistent against fixed alternatives, even in cases where the null value d_0 results in selecting Model A when in fact Model B applies, or vice-versa. Finite sample Monte Carlo simulation provided in section S.2.1 of the accompanying on-line supplement are supportive of this conjecture.

Remark 14. The single trend break model (2.1) could be extended to allow for multiple trend breaks. Specifically, we replace (2.1) with an (up to) m break model specification

$$x_t = \beta_1 + \beta_2 t + \boldsymbol{\beta}_3' \mathbf{DT}_t(\boldsymbol{\tau}^*) + e_t$$

[3] It is worth noting that this issue does not arise in the context of the frequency domain tests of Robinson (1994), or the analogous time-domain tests of Nielsen (2004), because they assume that the location of any trend break is known and, hence, they do not need to estimate τ^*.

where, $\mathbf{DT}_t(\boldsymbol{\tau}^*) := [DT_t(\tau_1^*), ..., DT_t(\tau_m^*)]'$. Here $\boldsymbol{\tau}^* := [\tau_1^*, ..., \tau_m^*]'$ is the vector of (unknown) putative trend break fractions, $\boldsymbol{\beta}_3 := [\beta_{3,1}, ..., \beta_{3,m}]'$ the associated break magnitude parameters such that a trend break occurs at time $\lfloor \tau_i^* T \rfloor$ when $\beta_{3,i} \neq 0$, $i = 1, ..., m$. The break fractions are assumed to be such that $\tau_i^* \in \Lambda$ for all $i = 1, ..., m$. A standard assumption in such a model is that $|\tau_i^* - \tau_j^*| \geq \eta > 0$, for all i, j, $i \neq j$, such that the DGP admits (up to) m level breaks occurring at unknown points across the interval Λ, with a sample fraction of at least $\lfloor \eta T \rfloor$ observations between breaks (note that m and η must satisfy the relation $m \leq 1 + \lfloor (\tau_U - \tau_L)/\eta \rfloor$). Provided that m breaks are estimated using the obvious m-dimensional analogue of (3.10), yielding the vector of estimates, $\widehat{\boldsymbol{\tau}}$ say, then we conjecture that the corresponding LM statistic, $LM(\widehat{\boldsymbol{\tau}})$ say, will have precisely the same properties as $LM(\widehat{\tau})$ in Theorem 1. That is, we conjecture that $LM(\widehat{\boldsymbol{\tau}}) \xrightarrow{d} \chi_1^2(c^2\omega^2)$ under H_1 and $LM(\widehat{\boldsymbol{\tau}}) \xrightarrow{d} \chi_1^2$ under H_0 irrespective of whether $\beta_{3,i} = 0$ or $\beta_{3,i} \neq 0$ for any particular i. For Model B Lavielle and Moulines (2000) demonstrate that $\widehat{\tau}_i \xrightarrow{p} \tau_i^*$ whenever $\beta_{3,i} \neq 0$ at the same rate as $\widehat{\tau} \xrightarrow{p} \tau^*$ in the single break case considered above. For Model A, it would seem likely that the same parallel with the single break case would hold, but formally Chang and Perron (2016) only consider the case of a single break in trend. For both Models A and B one would also need to formally establish that analogous uniformity arguments to those made in the proof of Theorem 1 can also be made in those cases where $\beta_{3,i} = 0$.

Remark 15. Although based on different models, it is nonetheless worth noting an important difference between the large sample results in Theorem 1 and those which hold for autoregressive unit root tests and stationarity tests which allow for the possibility of trend break(s). The limiting distributions of these, under both the null and the relevant local alternatives, depend on the number of trend breaks fitted, the number of breaks present in the data and the locations of these; see, for example, Perron and Rodríguez (2003) in the context of unit root tests, and Busetti and Harvey (2001,2003) in the context of stationarity tests. Moreover, their asymptotic local power functions depend on the number of trend breaks fitted, decreasing the more breaks are fitted, other things equal. This is not the case in our setting where, as the results in Theorem 1 demonstrate, the limiting distribution of our feasible $LM(\widehat{\tau})$ statistic is independent of any nuisance parameters arising from the deterministic kernel under both the null hypothesis and local alternatives. However, it is important to emphasise that this is an asymptotic result and so it will be important to investigate how well this asymptotic prediction holds up in finite samples. This we will investigate by Monte Carlo simulation methods in section 4.

Remark 16. Consider the case where an observed time series x_t satisfies the DGP

$$x_t = \beta_2 + \beta_3 DU_t(\tau^*) + e_t,\ t = 1, ..., T \tag{3.15}$$

where $e_t \in I(d)$, $d \in (-0.5, 0.5)$. In this case, x_t may be subject to a change in the mean but it is otherwise asymptotically stationary and invertible. It should be clear that inference on d in this model is equivalent to inference on δ in Model B in the context of DGP (2.1). Consequently, the results in Theorem 1 are also appropriate to this testing problem. As a leading example consider testing the null hypothesis that $d = 0$ in (3.15). It is well known that the model (3.15) with $d = 0$ and $\beta_3 \neq 0$ can generate spurious evidence of long memory when the break is not accounted for; see, for example, Diebold and Inoue (2001), Gourieroux and Jasiak (2001), Granger and Hyung (2004), Mikosch and Stărică (2004) and Qu (2011), and also the simulation results reported in section 4 below for the $\overline{LM}$ test in cases where $\beta_3 \neq 0$. Tests of the form proposed in this paper would allow for valid inference on d in (3.15), regardless of the value of β_3.

Remark 17. Observe that under H_0, $\widehat{\psi}$ defined in (3.1) and $\widehat{\psi}(\widehat{\tau})$ defined for (3.7) evaluated at $\tau = \widehat{\tau}$ are infeasible and feasible estimates, respectively, of the parameters characterising the (stationary and invertible) ARMA process, η_t. It is well known that, in the infeasible case, $\sqrt{T}\left(\widehat{\psi} - \psi^*\right) \rightarrow_d N\left(0, \Phi^{-1}\right)$; see, for example, Hamilton (1994), Chapter 5, and Harvey (1993), Chapter 3. This large sample result also holds when deterministic trend kernels, containing elements such as 1 (a constant), t (a linear trend), a broken intercept, $DU_t(\tau^*)$, or a broken trend, $DT_t(\tau^*)$, (τ^* assumed known in the latter two cases), are accounted for, so that ψ^* is estimated using de-trended residuals. This asymptotic equivalence, formally established in Theorem 4.1 of Nielsen (2004), holds because deterministic regressors such as these meet condition (12) of Nielsen (2004) or the similar condition given in Robinson (1994) page 1434. Crucially, however, the stochastic trend break regressors $DT_t(\widehat{\tau})$ and $DU_t(\widehat{\tau})$ do not meet these conditions. Nonetheless, as we demonstrate in Lemma A2, if $\beta_3 = 0$ then $\widehat{\psi}(\tau) - \widehat{\psi} = o_p\left(T^{-1/2}\right)$, uniformly in τ; moreover, as shown in Lemma C2, if $\beta_3 \neq 0$ then $\widehat{\psi}(\widehat{\tau}) - \widehat{\psi} = o_p\left(T^{-1/2}\right)$. Inference on ψ^* can therefore be made under H_0 using the result that $\sqrt{T}\left(\widehat{\psi}(\widehat{\tau}) - \psi^*\right) \rightarrow_d N\left(0, \Phi^{-1}\right)$. Consequently, an immediate corollary of Lemmas A2 and C2 is that using the appropriately de-trended residuals instead of η_t does not change the limiting distribution of the resulting estimate of ψ^* even when one includes the stochastic regressors $DT_t(\widehat{\tau})$ or $DU_t(\widehat{\tau})$.

4 Monte Carlo Simulations

We now present the results from a Monte Carlo simulation study investigating the finite sample performance of our proposed test based on the $LM(\widehat{\tau})$ statistic, exploring cases where no trend break occurs and where a trend break occurs. We investigate both finite sample size under the null hypothesis and finite sample power under local alternatives. Additional simulation experiments are reported in the accompanying on-line supplement with the results from these experiments summarised at the end of this section.

As benchmarks for comparison, we also simulate the (infeasible) tests based on: (i) the LM statistic in (3.3), (ii) the $LM(\tau^*)$ statistic given by (3.9) evaluated at $\tau = \tau^*$, and (iii) the statistic $\overline{LM}$, which is calculated as for the $LM(\widehat{\tau})$ statistic in section 3.2 but replacing $z_t(\tau)$ by z_t throughout, where for Model A, $z_t := (1, t)'$ and for Model B, $z_t := 1$. Recall that the first benchmark test is based on the unobservable e_t, while the second requires knowledge of the true (putative) break location, τ^*. The third benchmark test is based on the assumption that $\beta_3 = 0$ in (2.1). Its behaviour when $\beta_3 \neq 0$ allows us to quantify the finite sample consequences of neglecting a trend break when one is present in the DGP. When $\beta_3 = 0$ it quantifies the finite sample power losses that are incurred by unnecessarily allowing for a trend break.

All reported experiments are run over 10,000 Monte Carlo replications using the RNDN function of Gauss 13. Our simulation DGP is given by (2.1) with $\beta_1 = \beta_2 = 0$ (this is without loss of generality because all of the tests considered are exact invariant to β_1 and β_2) and $\beta_3 \in \{0, 0.1, 1\}$, with the break fraction set as $\tau^* = 0.5$. Notice that LM and $LM(\tau^*)$ are also exact invariant with respect to β_3. Excepting the tests based on LM and $\overline{LM}$, all tests are computed setting $\Lambda = [0.15, 0.85]$. All reported results relate to a nominal asymptotic 0.05 level using the relevant critical value from the χ_1^2 distribution.

We first consider the empirical size of these four tests across a range of values of d_0 and for sample sizes $T \in \{256, 512, 1024\}$. We generate $\{\eta_t\}$ according to $\eta_t = a\eta_{t-1} + \varepsilon_t$, $t = 1, ..., T$, with $\eta_0 = 0$, for $a \in \{-0.5, 0, 0.5\}$ and with $\{\varepsilon_t\}$ generated as an i.i.d. $N(0,1)$ sequence of variables. Consequently, η_t is also i.i.d. $N(0,1)$ when $a = 0$ and is a weakly stationary $AR(1)$ process when $a = \pm 0.5$. The shocks, e_t, $t = 1, ...T$, are then generated according to (2.2) to be such that $e_t \in I(d_0)$, for $d_0 \in \{0, 0.25, 0.5, 0.75, 1, 1.25\}$. Recall that Theorem 1 does not formally cover the case of $d_0 = 0.5$. In the simulation results reported here we take Model A to apply for the case where $d_0 = 0.5$. Finally, we simulate x_t, $t = 1, ..., T$, according to (2.1) for the values of β_1, β_2, β_3 and τ^* specified as above. In calculating the four test statistics we assumed knowledge of the autoregressive order (either zero or one) for η_t, but not of the parameter a in the case where η_t is an $AR(1)$. Notice that when η_t is i.i.d., then $\omega^2 = \pi^2/6$, otherwise ω^2 must be estimated. Following Tanaka (1999,p.564), we used $\hat{\omega}^2 := \pi^2/6 - \left(1 - \hat{a}^2\right)\left(\ln\left(1 - \hat{a}\right)\right)^2/\hat{a}^2$.[4]

Empirical size results are reported in Tables 1, 2 and 3 for $a = 0$, -0.5, 0.5 respectively. Consider first the results for the (infeasible) LM test. Due to the exact invariance of the LM test to d_0, results are only reported for $d_0 = 0$. We see that the LM test has size close to the nominal 0.05 level throughout, which we might expect given that it is calculated using the true e_t. Turning to the (infeasible) $LM(\tau^*)$ test (which is exact invariant to β_3), its empirical sizes are also in general reasonably close to the nominal level for $a = 0$ and $a = -0.5$; however, for $a = 0.5$ it can be significantly undersized for the smaller values of T considered. For our feasible $LM(\widehat{\tau})$ test, a degree of finite sample oversize is seen for $\beta_3 = 0$ and $\beta_3 = 0.1$, for both $a = 0$ and $a = -0.5$. For $a = 0.5$, similarly to what we observe for the $LM(\tau^*)$ test, $LM(\widehat{\tau})$ displays a tendency to undersize for the smaller sample sizes considered, though generally to a lesser extent than is seen for $LM(\tau^*)$. We believe the empirical size results for $LM(\widehat{\tau})$ are quite encouraging in that they would appear to show that relatively little in the way of size control is lost when moving from an LM-type test that requires knowledge of the (putative) break point to one which makes no such concession. It is also worth noting that the empirical size results in Tables 1, 2 and 3 for $LM(\widehat{\tau})$ differ very little for the case of $d_0 = 0.5$ *vis-à-vis* those for either $d_0 = 0.25$ or $d_0 = 0$.

[4] In the case of LM, $\widehat{a} := (\sum_{t=2}^T \widehat{\eta}_t\widehat{\eta}_{t-1})/(\sum_{t=2}^T \widehat{\eta}_{t-1})^2$ with $\widehat{\eta}_t := \Delta_+^{d_0} e_t$. For $LM(\tau)$, evaluated at either $\tau = \tau^*$ or $\tau = \widehat{\tau}$, $\widehat{a}(\tau) := \sum_{t=2}^T \widehat{\eta}_t(\tau)\widehat{\eta}_{t-1}(\tau)/\sum_{t=2}^T \left(\widehat{\eta}_{t-1}(\tau)\right)^2$, with $\widehat{\eta}_t(\tau) := \Delta_+^{d_0}\widehat{u}_t(\tau)$ under Model A, and $\widehat{\eta}_t(\tau) := \Delta_+^{d_0-1}\widehat{u}_t(\tau)$ under Model B. Finally, for $\overline{LM}$, $\widehat{a} := \sum_{t=2}^T \overline{\eta}_t\overline{\eta}_{t-1}/\sum_{t=2}^T \left(\overline{\eta}_{t-1}\right)^2$, where: for Model A, $\overline{\eta}_t := \Delta_+^{d_0}\overline{u}_t$ with $\overline{u}_t$ the OLS residuals from the regression of x_t on $(1,t)'$ for $t = 1, ..., T$; for Model B, $\overline{\eta}_t(\tau) := \Delta_+^{d_0-1}\overline{u}_t(\tau)$, with $\overline{u}_t$ the residuals from the regression of Δx_t on 1 for $t = 2, ..., T$, setting $\overline{u}_1 = 0$.

Next consider the results for the $\overline{LM}$ test which show the effect on empirical size of not allowing for a trend break, both where one occurs in the data ($\beta_3 \neq 0$) and where one does not ($\beta_3 = 0$). When $\beta_3 = 0$ the $\overline{LM}$ test, similarly to $LM\,(\tau^*)$, demonstrates reasonable size control for $a = 0$ and $a = -0.5$ but is rather undersized when $a = 0.5$ for the smaller T. However, where $\beta_3 \neq 0$, the $\overline{LM}$ test is seen to be completely unreliable, with empirical size reaching 1.0 in many cases. Unsurprisingly, the degree of size distortion becomes more serious as $|\beta_3|$ increases, this being a measure of the degree to which the model which omits the trend break is misspecified. The magnitude of the size distortions in $\overline{LM}$ are also seen to be larger the smaller is d_0, other things equal. This reflects the fact that omitting the broken trend in the deterministic specification renders the residuals contaminated by both a broken trend proportional to $(t - \lfloor \tau^* T \rfloor)^{1-d_0}$ and a linear trend proportional to t^{1-d_0}. Because (broken) trends have features similar to the properties of an integrated time series, see for example Iacone (2010), inference on d_0 is more heavily contaminated the larger is the exponent $(1 - d_0)$ on these contaminating trend terms in the residuals. Thus, inference when $d_0 = 0$ and more generally for lower values of d_0 is heavily distorted, whereas the contaminating effect when $d_0 = 1.25$ is seen to be much less dramatic.

We next turn to an examination of the finite sample local power properties of the tests. In order to save space, we restrict attention to the single sample size $T = 512$ for the case where η_t is i.i.d. $N(0, 1)$. In Figures 1-6, results are reported for $d_0 \in \{0, 0.25, 0.5, 0.75, 1, 1.25\}$. We consider an interval of local alternative values for c chosen as $c \in \{-5.0, -4.75, -4.50, ..., -0.25, 0, 0.25, ..., 4.50, 4.75, 5\}$ which is symmetric about the null value, $c = 0$. Local powers of $LM\,(\widehat{\tau})$ for each of $\beta_3 = 0$, $\beta_3 = 0.1$ and $\beta_3 = 1$ are plotted graphically against c, once more using the 0.05 χ_1^2 critical value. Also shown, again for benchmarking purposes, are the local powers of the LM, $LM\,(\tau^*)$ and $\overline{LM}$ tests, the latter is only reported for the case where $\beta_3 = 0$ because of its very poor size control for non-zero values of β_3 observed in Tables 1-3. Also shown is the relevant asymptotic local power function of the tests; that is, rejection frequencies for the $\chi_1^2\left(c^2\pi^2/6\right)$ distribution, denoted Asy. This asymptotic power function is invariant to d_0, as is the finite sample local power function of LM. We see that the local power function for LM lies very close to the symmetric (around $c = 0$) local power function of Asy.

Figure 1 graphs the local power functions of the tests for $d_0 = 0$. For both $LM\,(\tau^*)$ and $LM\,(\widehat{\tau})$, for a given value $v > 0$ finite sample powers are higher for $c = -v$ than for $c = v$. This is also true for $\overline{LM}$, though to a lesser extent. For $c < 0$, the powers of $LM\,(\tau^*)$ and $LM\,(\widehat{\tau})$ can exceed the corresponding asymptotic local power, but this is partly attributable to the slight oversizing of these tests seen in Table 1. For $c > 0$, however, these powers fall some way below the corresponding asymptotic local power values. Indeed, for small values of $c > 0$, power falls below the nominal level, albeit fairly modestly. It gives the impression that the finite sample power curves for $LM\,(\tau^*)$ and $LM\,(\widehat{\tau})$ are rightward shifted relative to the centering of their common asymptotic local power function. We have no ready explanation as to why such finite sample asymmetry (around $c = 0$) should occur, but that it arises for both $LM\,(\tau^*)$ and $LM(\widehat{\tau})$, and also for $\overline{LM}$, but not for LM, clearly suggests it is connected to the fact that the first three tests are based on estimated deterministic trend terms; indeed, of these three tests $LM\,(\tau^*)$ and $LM\,(\widehat{\tau})$ are based on a richer deterministic specification than $\overline{LM}$, and correspondingly appear to show the greater degree of asymmetry. Comparing $LM(\tau^*)$ and $LM(\widehat{\tau})$, we see that they generally have fairly similar levels of power, particularly when $\beta_3 = 1$; this might be expected since, for a large break magnitude of this kind, $\widehat{\tau}$ should be in close proximity to τ^*.

In Figure 2, where $d_0 = 0.25$, most of the same comments made for Figure 1 apply here also. However, $LM(\tau^*)$ does now appear slightly more powerful than $LM(\widehat{\tau})$ when $\beta_3 = 1$. The results for $d_0 = 0.5$ in Figure 3 appear qualitatively very similar to those for $d_0 = 0.25$.

The corresponding results for $d_0 = 0.75$, $d_0 = 1$ and $d_0 = 1.25$ are shown in Figures 4, 5 and 6 respectively. Interestingly, when $d_0 = 0.75$ the asymmetry of the $LM(\tau^*)$ and $LM(\widehat{\tau})$ power curves (and indeed of $\overline{LM}$), appears somewhat less evident than for the three cases discussed above, with $LM(\tau^*)$ and $LM(\widehat{\tau})$ once more demonstrating similar power when $\beta_3 = 1$. For $d_0 = 1.0$ and $d_0 = 1.25$ the asymmetries in the power functions of $LM(\tau^*)$ and $LM(\widehat{\tau})$ reappear to some extent; in the latter case with $LM(\tau^*)$ appearing slightly more powerful than $LM(\widehat{\tau})$, which suggests that $\widehat{\tau}$ is struggling to estimate τ^* particularly well by this point.

The overall power performance of $LM(\widehat{\tau})$ test should be gauged in context. Expecting it to always closely replicate the power behaviour of $\overline{LM}$ or $LM(\tau^*)$ tests (let alone the infeasible LM test) in finite samples represents something of an unrealistic challenge. Respectively, these tests need to correctly assume that no trend break occurs, or if one does occur, that the true break date is known in order for their size to controlled, and their powers to be in any way meaningful. As such, they require prior information that is simply never made available to a practitioner. Conversely, the $LM(\widehat{\tau})$ test does not place any reliance on the veracity of such information. Judged on this basis, we consider that the relative finite sample power performance of $LM(\widehat{\tau})$ across our range of values for d_0 is actually more than acceptable.[5]

Additional finite sample simulation results can be found in the accompanying on-line supplement to this paper. For the tests discussed in this section, these investigate: (i) the empirical power properties of the tests against fixed alternatives; (ii) the impact of innovation distributions which violate the moment conditions stated in Assumption 1; and (iii) the use of model selection methods to select the autoregressive lag order for the short memory component of the model. Concentrating on the results for $LM(\widehat{\tau})$, the findings of these simulation experiments can be summarised as follows. Under (i), even for "distant" fixed alternatives where the $LM(\widehat{\tau})$ statistic is based on the wrong model (e.g. $d = 0.6$ and $d_0 = 0.4$ where Model A is used to construct the test statistic but in fact Model B holds for the true DGP) the $LM(\widehat{\tau})$ test appears consistent (its power approaches 1 with increasing T) and, other things being equal, its power also increases with the distance between d and d_0; for example, $d = 0.6$, $d_0 = 0.4$ shows lower power than $d = 0.75$, $d_0 = 0.25$. For (ii), when moment conditions are significantly violated e.g. when $d = d_0 = 0.51$ and the innovations are t_5 distributed, $LM(\widehat{\tau})$ is badly oversized. Indeed its size appears to diverge with T when $\beta_3 = 0$. On the basis of these

[5] Unreported simulations we have conducted for larger T confirm that the local power curves of $LM(\widehat{\tau})$ do indeed converge towards their asymptotic counterparts. However, this convergence appears to be rather slow. For example, in the case where $d_0 = 1$ and $c = 2$, the power of the test based on $LM(\widehat{\tau})$ for $T = n \times 512$, $n = 1, 4, 16$ is $0.42, 0.57, 0.64$, while the corresponding asymptotic local power at $c = 2$ is 0.73.

simulation results, we do not recommend using $LM(\widehat{\tau})$ for testing null hypotheses close to $d = 0.5$ especially if fat-tailed behaviour is suspected to be present in the data. Under (iii), we found that using a standard Bayes Information Criterion to select the lag order, assuming a maximal order of two, made only minor differences to the sizes reported in Tables 1-3, where the lag order is assumed known.

5 Empirical Example

In this section we use the $LM(\widehat{\tau})$ testing procedure developed in this paper to examine the persistence properties of the logarithm of the CPI price index in the U.S. over the period January 1970 to January 2018. The data were obtained from the OECD database over a total of 577 monthly observations. The log CPI data are graphed in Figure 7. Figure 8 plots the annual inflation rate (calculated as 100 times the annual differences of the log CPI data). Figure 7 suggests that the log CPI data would appear to be well characterised by the "changing growth" model, displaying an apparent negative change in the slope of the trend function in the early 1980s but without any sudden change in the level at that point. A negative change in the slope of the trend of log CPI implies a level shift in the inflation rate (either the annual inflation rate as in Figure 8, or a corresponding monthly inflation rate based on the first differences of the log CPI data) in the early 1980s from a regime of relatively high inflation to one of relatively low inflation, and again this can clearly be seen in Figure 8. Clarida, Gali and Gertler (2000) argue that the appointment of Paul Volcker as chairman of the Board of Governors of the Federal Reserve System in July 1979 brought about a more aggressive stance on monetary policy, which may have caused the apparent level change in inflation; see also Boivin and Giannoni (2006).

Investigating the degree of persistence of inflation is important because it has implications for the appropriate timing and intensity of monetary policy intervention; see, for example, Fuhrer (2011) and Angeloni *et al.* (2006). Although effective monetary policy requires that the central bank understands the structural origins of inflation dynamics, a summary, or reduced form, measure of inflation persistence is nevertheless important as it provides a benchmark that should be kept under consideration when designing economic models. The order of integration, being informative about the strength of the autocorrelation at long range, seems particularly well suited to this end. Early applications of fractionally integrated models to inflation data include, among others, Hassler and Wolters (1995) and Baillie, Chung and Tieslau (1996). These studies find statistically significant evidence of positive fractional integration (with estimated values of the long memory parameter, for various estimation methods, found to lie in the range 0.40 to 0.47) in U.S. inflation rate data. Fixing the slope of the deterministic trend function to be constant across the sample, using the exact local Whittle [ELW] method of Shimotsu and Phillips (2005), with the modification to allow for deterministic trends developed in Shimotsu (2010), and a bandwidth of $m = \lfloor T^{0.65} \rfloor$, we estimate the long memory parameter for the log CPI data in Figure 7 to be $\hat{d} = 1.30$, again implying relatively strong positive fractional integration in inflation.

Hassler and Wolters (1995) find no significant evidence of conditional heteroskedasticity in the data but, like Figure 7 above, Figure 1 of Hassler and Wolters (1995,p.38) clearly suggests the presence of a negative trend break in the early 1980s in the U.S. log CPI data. Given the apparent change in the level of inflation in the early 1980s, and following the discussion in Remark 16, we therefore need to be alert to the possibility that the methods of inference used in the empirical studies of the U.S. inflation rate discussed above could be suggesting the presence of long memory in inflation because of an unmodelled level break in the data rather than genuine long memory. The $LM(\widehat{\tau})$ test developed in this paper may therefore be useful in the context of these data.

We will test the key null hypothesis that inflation is a short memory, $I(0)$, process. This corresponds to testing $H_0 : d = 1$ in the log CPI data which therefore entails the use of Model B. A second hypothesis of interest could be testing $H_0 : d = 0$, in which case the log CPI data is a short memory process about a broken deterministic trend. In this latter case, Model A is therefore used. In both cases, the short memory component of the series is selected using the Bayes Information Criterion (BIC) of Schwarz (1978), choosing between AR(0), AR(1) and AR(2) models. A summary of the results for the $LM(\widehat{\tau})$ test for these two null hypotheses is as follows:

	$H_0 : d = 0$	$H_0 : d = 1$
$\widehat{\tau}$	0.32	0.26
BIC	AR(2)	AR(1)
$LM(\widehat{\tau})$	19.80	1.74
p-value	0.000	0.187

where BIC denotes the model selected using the information criterion, $LM(\widehat{\tau})$ is the outcome of the test statistic for H_0 in each case and p-value is the associated (asymptotic) p-value for the test (obtained from the χ_1^2 distribution), and $\widehat{\tau}$ is the estimated break fraction obtained using Model A in the case of testing $H_0 : d = 0$, and Model B in the case of $H_0 : d = 1$. It can therefore be seen that the null hypothesis that log CPI is a short memory process (allowing for a trend break) is rejected at any conventional significance level. In contrast, the null hypothesis that log CPI is an I(1) process about a deterministic trend subject to a break, and hence that inflation is $I(0)$ about a level change, cannot be rejected using conventional significance levels (although, of course, one cannot conclude for certain from this test result that inflation is not a long memory process).[6] The estimated break fraction $\widehat{\tau} = 0.26$ corresponds to a break date of June 1982, which seems not inconsistent with the likely timescale needed for the impact of the changes in monetary policy adopted by the Fed after the appointment of Volcker to feed through into the recorded inflation rate.[7] The dashed line in Figure 7 shows the fitted broken trend function for $\widehat{\tau} = 0.26$, while the dashed line in Figure 8 depicts the corresponding fitted broken level function for the inflation rate data.

For comparative purposes, we also tested $H_0 : d = 1$ in the log CPI data using the $\overline{LM}$ test which does not allow for a trend break in the data. Here the BIC selects an AR(1) model and the outcome of

[6]The ELW estimate of d in the log CPI data, allowing for a break in trend at $\widehat{\tau} = 0.26$, is $\hat{d} = 1.137$.

[7]That said, the 1981/1982 recession in the US, which is known to have been particularly sharp over this period, could also have been a relevant factor.

the statistic is $\overline{LM} = 72.33$, thereby leading to an overwhelming rejection of the $I(1)$ null hypothesis. The outcome of the $\overline{LM}$ test is therefore consistent with the findings of the previous empirical studies discussed above that inflation displays long memory persistence, contrasting with the result of the $LM(\widehat{\tau})$ test which suggests that inflation is a short memory ($I(0)$) process about a changing level. To shed further light on this matter, we test for a trend break in the log CPI data using the SW and MW tests of Iacone, Leybourne and Taylor (2013b) and Iacone, Leybourne and Taylor (2013a), respectively, both of which are robust to the order of integration of the data. The outcome of the SW and MW statistics were 14.33 and 83.80, respectively, with the respective 5% critical values given by 12.08 and 61.98. Both tests therefore find significant evidence of the presence of a trend break in the data, lending further weight to the conclusions drawn from the $LM(\widehat{\tau})$ test that inflation is a short memory process, and that the evidence of positive long memory found in some earlier studies might be attributable to an unmodelled level shift in the data occurring in the early 1980s.

Finally, recall that the ELW estimate of the long memory parameter in the log CPI data when a trend break is not allowed for is $\hat{d} = 1.30$. The outcomes of the $\overline{LM}$ and $LM(\widehat{\tau})$ statistics for testing the null hypothesis $H_0 : d = 1.3$ are 1.187 and 7.715, respectively, with associated p-values of 0.275 and 0.005, respectively. As such, while we can easily reject the null hypothesis that $d = 1.3$ at any conventional significance level when we allow for a trend break in the data, we cannot when we impose a fixed slope on the trend function.

6 Conclusions

We have been concerned with the problem of conducting inference on the long memory parameter in the context of a series which is fractionally integrated around a potentially broken deterministic trend. To that end, we have extended the LM-based testing approach of Robinson (1994), Tanaka (1999) and Nielsen (2004), which assumes a known functional form for the deterministic kernel, to the unknown trend break case we consider. This was achieved by basing the LM-type tests on data which have been de-trended allowing for a trend break with the location of the break estimated by a residual sum of squares estimator. This estimator was based either on the levels or first differences of the data dependent on the value imposed on the long memory parameter under the null hypothesis. We have demonstrated that the resulting LM-type test shares the same large sample asymptotic local optimality properties as are obtained in the known deterministic kernel case of Robinson (1994), Tanaka (1999) and Nielsen (2004) and, again like those tests, has asymptotic null critical values given by the χ^2_1 distribution. Results were reported from a Monte Carlo study into the finite-sample behaviour of our proposed test and it was found that the test performs well in terms of size control and local power levels. An empirical application to U.S. inflation data suggested that some previous findings of positive long memory in the inflation rate might be attributable to an unmodelled level change in inflation in the early 1980s.

We conclude with two suggestions for future research. First, we have here considered the case where the trend break magnitude, β_3, in DGP (2.1) is either exactly zero, such that no trend break occurs, or is a non-zero constant, such that a trend break of fixed magnitude occurs. It would also be interesting to investigate the behaviour of our proposed $LM(\widehat{\tau})$ test in cases where β_3 is local-to-zero at some polynomial rate in T; viz., $\beta_3 = KT^{-\alpha}$ for some $\alpha > 0$ and where K is a constant. We are currently working on a detailed analytical investigation into this local trend break case which we hope to report in separate work in due course. Our analysis so far leads us to conjecture that our main result from Theorem 1 that $LM(\widehat{\tau}) - LM = o_p(1)$ will continue to hold in this case, regardless of the value of α. This would have the implication that this result holds uniformly in β_3. Second, we have focussed here on the use of time-domain methods for developing tests on the long memory parameter, d. It would also be interesting to develop tests in the frequency-domain, along the lines of the LM tests of Robinson (1994). Following Iacone (2010) and Perron and McCloskey (2013), it might be feasible that, for certain values of d, these statistics could be implemented with trimming of some low frequency periodogram ordinates, enabling inference on d to be carried out without the need to parametrically account for any breaks present in the deterministic trend function. The degree of trimming that would be required depends on d, and when d is small it may require eliminating a relatively large number of the low frequency periodogram ordinates. As such, this approach would likely be better suited to cases falling under our Model B than under Model A.

References

Angeloni, I., L. Aucremanne, M. Ehrmann, J. Gali, A. Levin and F. Smets, 2006. New evidence on inflation persistence and price stickiness in the Euro area: implications for macro modeling, *Journal of the European Economic Association* 4, 562–574.

Baillie, R.T., C.-F. Chung and M.A. Tieslau, 1996. Analysing inflation by the fractionally integrated ARFIMA–GARCH model, *Journal of Applied Econometrics* 11, 23–40.

Boivin, J., and M. P. Giannoni, 2006. Has monetary policy become more effective?, *Review of Economics and Statistics* 88, 445–462.

Busetti, F. and A.C. Harvey, 2001. Testing for the presence of a random walk in series with structural breaks. *Journal of Time Series Analysis* 22, 127–150.

Busetti, F. and A.C. Harvey 2003. Further comments on stationarity tests in series with structural breaks at unknown points, *Journal of Time Series Analysis* 24, 137–140.

Chang, S.Y., and P. Perron, 2016. Inference on a structural break in trend with fractionally integrated errors, *Journal of Time Series Analysis* 37, 555–574.

Clarida, R., J. Gali and M. Gertler, 2000. Monetary policy rules and macroeconomic stability: evidence and some theory, *The Quarterly Journal of Economics* 115, 147–180.

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.

Diebold, F.X., and A. Inoue, 2001. Long memory and structural change, *Journal of Econometrics* 105, 131–159.

Fuhrer, J.C., 2011. Inflation persistence, *Handbook of Monetary Economics* 3, 423–486.

Gourieroux, C., and J. Jasiak, 2001. Memory and infrequent breaks, *Economics Letters* 70, 29–41.

Granger, C.W.J., and N. Hyung, 2004. Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns, *Journal of Empirical Finance* 11, 399–421.

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

Harvey, A.C., 1993. *Time Series Models*, Harvester Wheatsheaf.

Hassler, U., and J. Wolters, 1995. Long memory in inflation rates: international evidence, *Journal of Business and Economic Statistics* 13, 37–45.

Iacone, F., 2010. Local Whittle estimation of the memory parameter in presence of deterministic components, *Journal of Time Series Analysis* 31, 37–49.

Iacone, F., S.J. Leybourne and A.M.R. Taylor, 2013a. On the behavior of fixed-b trend break tests under fractional integration, *Econometric Theory* 29, 393–418.

Iacone, F., S.J. Leybourne and A.M.R. Taylor, 2013b. Testing for a break in trend when the order of integration is unknown, *Journal of Econometrics* 176, 30–45.

Johansen, S., and M.Ø. Nielsen, 2010. Likelihood inference for a nonstationary fractional autoregressive model, *Journal of Econometrics* 158, 51–66.

Johansen, S., and M.Ø. Nielsen, 2012. A necessary moment condition for the fractional functional central limit theorem, *Econometric Theory* 28, 671–679.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin, 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? *Journal of Econometrics* 54, 159–178.

Lavielle, M., and E. Moulines, 2000. Least squares estimation of an unknown number of shifts in a time series, *Journal of Time Series Analysis* 21, 33–59.

Marinucci, D., and P.M. Robinson, 2000. Weak convergence of multivariate fractional processes, *Stochastic Processes and their Applications* 86, 103–120.

McCloskey A., and P. Perron, 2013. Memory parameter estimation in the presence of level shifts and deterministic trend, *Econometric Theory* 29, 1196–1237.

Mikosch, T., and K. Starica, 2004. Nonlinearity in financial time series, the long range dependence and the I-GARCH effect, *Review of Economics and Statistics* 86, 378–390.

Nielsen, M.Ø., 2004. Efficient likelihood inference in nonstationary univariate models, *Econometric Theory* 20, 116–146.

Perron, P., 1989. The great crash, the oil price shock, and the unit root hypothesis, *Econometrica* 57, 1361–1401.

Perron, P. and G. Rodríguez, 2003. GLS detrending, efficient unit root tests and structural change, *Journal of Econometrics* 115, 1–27.

Qu, Z., 2011. A test against spurious long memory. *Journal of Business and Economic Statistics* 29, 423–438.

Robinson, P.M., 1994. Efficient tests of nonstationary hypotheses, *Journal of the American Statistical Association* 89, 1420–1437.

Schwarz, G., 1978. Estimating the dimension of a model, *Annals of Statistics* 6, 461–464.

Shimotsu, K., 2010. Exact local Whittle estimation of fractional integration with unknown mean and time trend. *Econometric Theory* 26, 501–540.

Shimotsu, K. and P.C.B. Phillips, 2005. Exact local Whittle estimation of fractional integration. *Annals of Statistics* 33, 1890–1933.

Tanaka, K., 1999. The nonstationary fractional unit root, *Econometric Theory* 15, 549–582.

A Appendix

The conditions stated in Lemma 1 are assumed to hold throughout this appendix. We will use the nomenclature C throughout to denote a generic positive bound. For a generic matrix B, we denote by $\overline{ei}\,(B)$ the largest eigenvalue of B, and define the norm of B as $\|B\| := \left\{\overline{ei}\,(B'B)\right\}^{1/2}$. Where a function of τ is considered, the stochastic orders $O_p\,(.)$ and $o_p\,(.)$ will be assumed to hold for the function using a suitable metric, and, unless specified otherwise, we will use the uniform distance. For example, from the standard FCLT, if $T^{-1/2}\sum_{t=1}^{\lfloor \tau T\rfloor}\varepsilon_t \Rightarrow \sigma_\varepsilon W\,(\tau)$, where "$\Rightarrow$" indicates weak convergence in the uniform metric, and $W\,(\tau)$ is a standard Brownian motion, we will write $\sum_{t=1}^{\lfloor \tau T\rfloor}\varepsilon_t = O_p\left(T^{1/2}\right)$. To abbreviate notation (and mirroring the definition of δ_0) we define $\delta := d$ if $d \in (-0.5, 0.5)$ and $\delta := d-1$ if $d \in (0.5, 1.5)$.

A.1 Proof of Lemma 1

We first detail results under H_0; here it holds that $d = d_0$ and $\delta = \delta_0$. We consider the cases $\beta_3 = 0$ and $\beta_3 \neq 0$ separately, and for each case we divide the proof into Lemma A1 and Lemma B1, to make it easier to follow. We then detail in Lemma C1 how to account for the local alternative, H_c. We prove

Lemma 1 by putting these three lemmas together. Proofs of Lemmas A1, B1 and C1 are provided in the accompanying on-line supplementary appendix.

Lemma A1. Let $\widehat{\beta}(\tau)$ be the OLS estimate in (3.5). For $\beta_3 = 0$, under H_0,

$$K_T(d)\left(\widehat{\beta}(\tau) - \beta\right) = O_p(1). \tag{A.1}$$

□

Lemma B1. Let $\widehat{\beta}(\tau)$ be the OLS estimate in (3.5) and $\widehat{\tau}$ the minimum RSS estimate in (3.10). For $\beta_3 \neq 0$ and under H_0:

(i) if $d_0 \in (-0.5, 0.5)$, then

$$\widehat{\tau} \xrightarrow{p} \tau^* \text{ and } \widehat{\tau} - \tau^* = O_p\left(T^{-3/2+\delta}\right) \tag{A.2}$$

(ii) if $d_0 \in (0.5, 1.5)$, then

$$\widehat{\tau} \xrightarrow{p} \tau^* \text{ and } \widehat{\tau} - \tau^* = O_p\left(T^{-1}\right) \tag{A.3}$$

(iii) for $d_0 \in (-0.5, 0.5) \cup (0.5, 1.5)$,

$$K_T(d)\left(\widehat{\beta}(\widehat{\tau}) - \beta\right) = O_p(1). \tag{A.4}$$

□

Lemma C1. For $\alpha \in (-1/2, 1/2)$, $r \geq 0$, r integer,

$$T^{-(1/2+\alpha)}(\ln(T))^{-r}\sum_{t=1}^{\lfloor \tau T\rfloor}\left((\ln(\Delta))^r \Delta^{-(\alpha+\theta_T)}\right)_+ \eta_t = T^{-(1/2+\alpha)}(\ln(T))^{-r}\sum_{t=1}^{\lfloor \tau T\rfloor}\left((\ln(\Delta))^r \Delta^{-\alpha}\right)_+ \eta_t + o_p(1).$$

□

Using Lemmas A1, B1 and C1, the proof of Lemma 1 is completed as follows:

- Under H_0, Lemma 1 follows directly from Lemmas A1 and B1.

- Under H_c, from Lemma C1, setting $r = 0$ and $\alpha = \delta_0$, the FCLT $T^{-(1/2+\delta_0)}\sum_{t=1}^{\lfloor \tau T\rfloor} u_t \Rightarrow \sigma_\infty W(\tau; \delta_0)$ still holds. Therefore, when $\beta_3 = 0$, the result in (3.14) follows using arguments similar to those used in Lemma A.1. For the proof under $\beta_3 \neq 0$, we observe that Chang and Perron (2016) derived (A.2) using the FCLT for $T^{-(1/2+\delta)}\sum_{t=1}^{\lfloor \tau T\rfloor} \Delta_+^{-\delta}\eta_t$. However, from Lemma C1, we can replace this with $T^{-(1/2+\delta_0)}\sum_{t=1}^{\lfloor \tau T\rfloor} \Delta_+^{-(\delta_0+\theta_T)}\eta_t$. Therefore, (A.2) is also valid under H_c for Model A. For Model B, (A.3) holds for any $\delta \in (-1/2, 1/2)$ and, since for T sufficiently large $(\delta_0 + \theta_T) \in (-1/2, 1/2)$ still holds, then (3.11) is still met.

A.2 Proof of Theorem 1

We organise the proof of Theorem 1 in a similar way to the proof of Lemma 1 above. That is, we derive results under H_0 first, considering the cases $\beta_3 = 0$ and $\beta_3 \neq 0$ separately, and then subsequently discuss the corresponding results under H_c. We first state some preparatory results in Lemmas A2, B2, C2 and D2, each of whose proof is again provided in the accompanying on-line supplementary appendix.

Lemma A2. Under $\beta_3 = 0$ and H_0: (i) $\widehat{\psi}(\tau) - \widehat{\psi} = o_p(1)$, and (ii) $T^{1/2}(\widehat{\psi}(\tau) - \widehat{\psi}) = o_p(1)$. □

Lemma B2. Recalling that $\widehat{\varepsilon}_t(\tau) = g(L; \widehat{\psi}(\tau))\Delta_+^\delta \widehat{u}_t(\tau)$ and $\widehat{\varepsilon}_t = g(L; \widehat{\psi})\Delta_+^\delta u_t$, and defining $\widehat{v}_t(\tau) := \sum_{j=1}^{t-1} j^{-1}\widehat{\varepsilon}_{t-j}(\tau)$ and $\widehat{v}_t := \sum_{j=1}^{t-1} j^{-1}\widehat{\varepsilon}_{t-j}$, then under $\beta_3 = 0$ and H_0,

$$
\begin{aligned}
T^{-1/2}\sum_{t=1}^{T}\widehat{\varepsilon}_t(\tau)\widehat{v}_t(\tau) - T^{-1/2}\sum_{t=1}^{T}\widehat{\varepsilon}_t\widehat{v}_t &= o_p(1) && \text{(A.5)}\\
\widehat{s}^2(\tau) - \widehat{s}^2 &= o_p(1) && \text{(A.6)}\\
\widehat{\omega}^2(\tau) - \widehat{\omega}^2 &= o_p(1). && \text{(A.7)}
\end{aligned}
$$

□

Lemma C2. When $\beta_3 \neq 0$, under H_0, $T^{1/2}\left(\widehat{\psi}(\widehat{\tau}) - \widehat{\psi}\right) = o_p(1)$. □

Lemma D2. When $\beta_3 \neq 0$, under H_0, $T^{-1/2}\sum_{t=1}^{T}\widehat{\varepsilon}_t(\widehat{\tau})\widehat{v}_t(\widehat{\tau}) - T^{-1/2}\sum_{t=1}^{T}\widehat{\varepsilon}_t\widehat{v}_t = o_p(1)$, $\widehat{s}^2(\widehat{\tau}) - \widehat{s}^2 = o_p(1)$, and $\widehat{\omega}^2(\widehat{\tau}) - \widehat{\omega}^2 = o_p(1)$. □

Using Lemmas A2, B2, C2 and D2, the proof of Theorem 1 is then completed as follows. We derive the result under H_0 and $\beta_3 = 0$ first. Re-write $\widehat{A}(\tau) = T^{-1/2}\sum_{t=1}^{T}\widehat{\varepsilon}_t(\tau)\widehat{v}_t(\tau)/\widehat{s}^2(\tau)$ and, in view of Lemma B2 and continuity, $\widehat{A}(\tau) - \widehat{A} = o_p\left(T^{-1/2}\right)$; in the same way, $LM(\tau) - LM = o_p(1)$. The proof for $\beta_3 \neq 0$ is similar, but uses Lemma C2 and Lemma D2 instead. Where H_c holds, the results in Lemma A2, Lemma B2, Lemma C2 and Lemma D2 can be straightforwardly extended, applying the mean value theorem expansion used in Lemma C1, to show that the rate is not affected under the alternative.

Table 1. Empirical size of tests, $a = 0$

	T	LM		$\overline{LM}$		$LM(\tau^*)$		$LM(\widehat{\tau})$	
d_0			$\beta_3 = 0$	$\beta_3 = 0.1$	$\beta_3 = 1$		$\beta_3 = 0$	$\beta_3 = 0.1$	$\beta_3 = 1$
	256	0.048	0.041	1.000	1.000	0.050	0.069	0.065	0.050
0	512	0.047	0.043	1.000	1.000	0.054	0.069	0.064	0.054
	1024	0.047	0.046	1.000	1.000	0.052	0.060	0.059	0.052
	256		0.041	1.000	1.000	0.050	0.072	0.065	0.058
0.25	512		0.044	1.000	1.000	0.054	0.069	0.065	0.058
	1024		0.045	1.000	1.000	0.053	0.060	0.059	0.055
	256		0.039	0.857	1.000	0.048	0.069	0.065	0.057
0.5	512		0.042	1.000	1.000	0.050	0.065	0.060	0.057
	1024		0.044	1.000	1.000	0.051	0.059	0.055	0.054
	256		0.036	0.122	1.000	0.038	0.039	0.047	0.040
0.75	512		0.040	0.372	1.000	0.042	0.045	0.047	0.045
	1024		0.044	0.886	1.000	0.046	0.048	0.049	0.046
	256		0.036	0.042	1.000	0.041	0.060	0.059	0.044
1	512		0.039	0.051	1.000	0.043	0.063	0.063	0.044
	1024		0.044	0.063	1.000	0.045	0.059	0.057	0.046
	256		0.037	0.038	0.316	0.042	0.068	0.069	0.055
1.25	512		0.039	0.039	0.429	0.043	0.071	0.070	0.050
	1024		0.044	0.045	0.546	0.045	0.064	0.062	0.052

Table 2. Empirical size of tests, $a = -0.5$

	T	LM	$\overline{LM}$			$LM(\tau^*)$	$LM(\widehat{\tau})$		
d_0			$\beta_3 = 0$	$\beta_3 = 0.1$	$\beta_3 = 1$		$\beta_3 = 0$	$\beta_3 = 0.1$	$\beta_3 = 1$
	256	0.051	0.042	0.966	1.000	0.055	0.074	0.072	0.055
0	512	0.052	0.047	1.000	1.000	0.058	0.073	0.068	0.058
	1024	0.047	0.044	1.000	1.000	0.052	0.062	0.058	0.052
	256		0.042	1.000	1.000	0.057	0.079	0.075	0.060
0.25	512		0.046	1.000	1.000	0.057	0.076	0.069	0.057
	1024		0.045	1.000	1.000	0.052	0.063	0.061	0.052
	256		0.039	0.998	0.998	0.055	0.078	0.074	0.064
0.5	512		0.044	1.000	1.000	0.054	0.073	0.064	0.061
	1024		0.045	1.000	1.000	0.050	0.062	0.056	0.055
	256		0.037	0.338	1.000	0.039	0.034	0.034	0.039
0.75	512		0.042	0.869	1.000	0.043	0.040	0.044	0.043
	1024		0.041	1.000	1.000	0.042	0.043	0.045	0.046
	256		0.037	0.050	1.000	0.041	0.059	0.061	0.041
1	512		0.042	0.078	1.000	0.045	0.064	0.061	0.045
	1024		0.042	0.126	1.000	0.045	0.058	0.056	0.044
	256		0.035	0.035	0.757	0.040	0.071	0.071	0.044
1.25	512		0.043	0.043	0.905	0.046	0.072	0.071	0.048
	1024		0.042	0.045	0.976	0.045	0.065	0.064	0.046

Table 3. Empirical size of tests, $a = 0.5$

	T	LM		$\overline{LM}$		$LM(\tau^*)$		$LM(\widehat{\tau})$	
d_0			$\beta_3 = 0$	$\beta_3 = 0.1$	$\beta_3 = 1$		$\beta_3 = 0$	$\beta_3 = 0.1$	$\beta_3 = 1$
	256	0.048	0.010	0.949	0.997	0.018	0.032	0.032	0.025
0	512	0.050	0.023	1.000	1.000	0.036	0.059	0.056	0.040
	1024	0.048	0.032	1.000	1.000	0.044	0.066	0.061	0.044
	256		0.010	0.351	0.452	0.018	0.034	0.034	0.031
0.25	512		0.023	0.243	1.000	0.037	0.062	0.059	0.054
	1024		0.032	0.999	1.000	0.046	0.071	0.064	0.060
	256		0.015	0.081	1.000	0.019	0.031	0.036	0.031
0.5	512		0.025	0.725	1.000	0.036	0.062	0.058	0.055
	1024		0.032	1.000	1.000	0.045	0.072	0.063	0.061
	256		0.010	0.019	0.092	0.010	0.011	0.013	0.011
0.75	512		0.019	0.077	0.112	0.022	0.021	0.024	0.023
	1024		0.026	0.302	0.458	0.030	0.031	0.034	0.030
	256		0.011	0.013	0.517	0.012	0.021	0.021	0.014
1	512		0.021	0.024	0.961	0.024	0.039	0.039	0.025
	1024		0.026	0.035	1.000	0.033	0.052	0.050	0.036
	256		0.012	0.012	0.033	0.013	0.028	0.027	0.025
1.25	512		0.021	0.021	0.075	0.023	0.050	0.049	0.041
	1024		0.026	0.027	0.123	0.034	0.061	0.060	0.051

Figure 1. Local power of tests, $T = 512$, $d_0 = 0$.

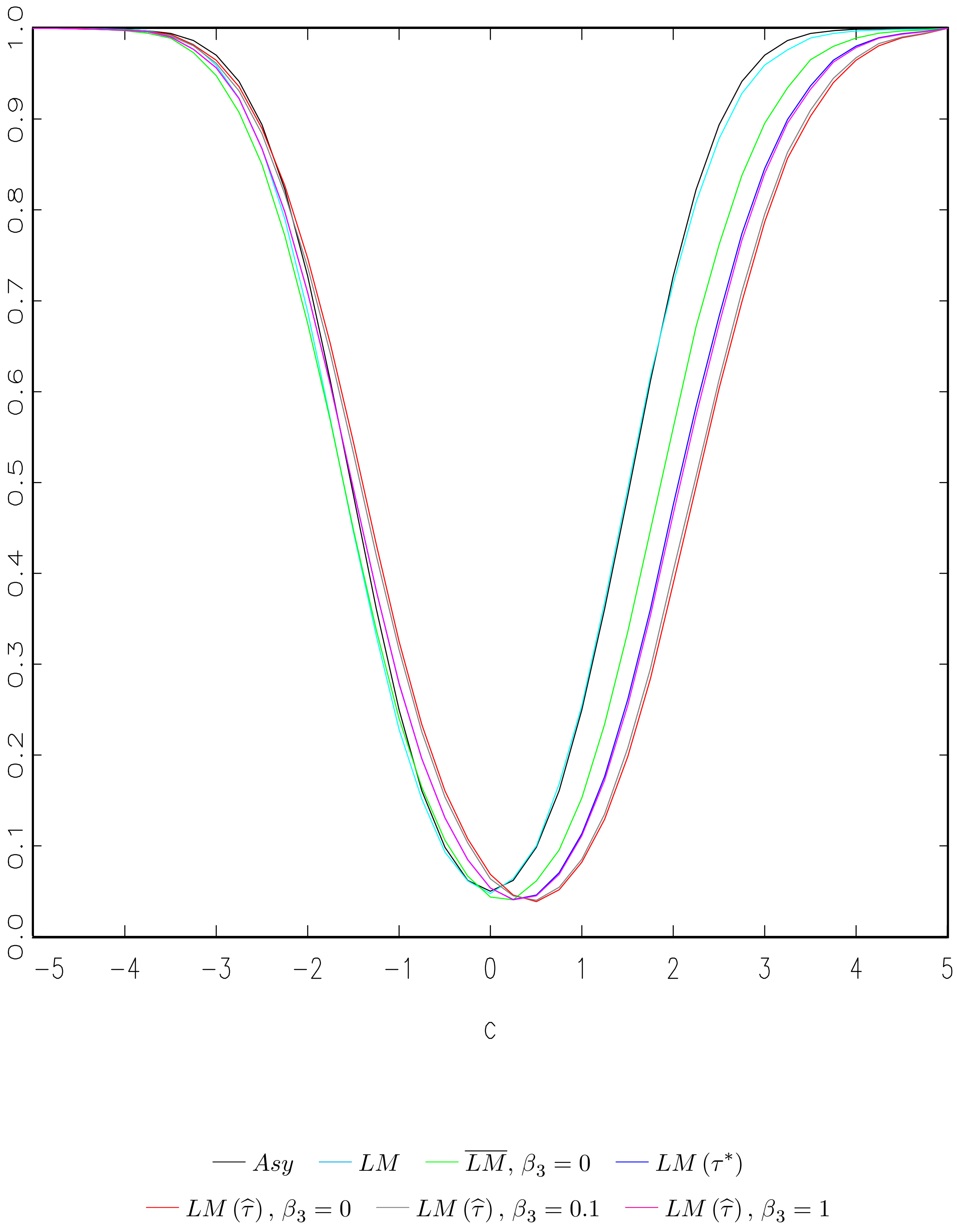


Figure 2. Local power of tests, $T = 512$, $d_0 = 0.25$.

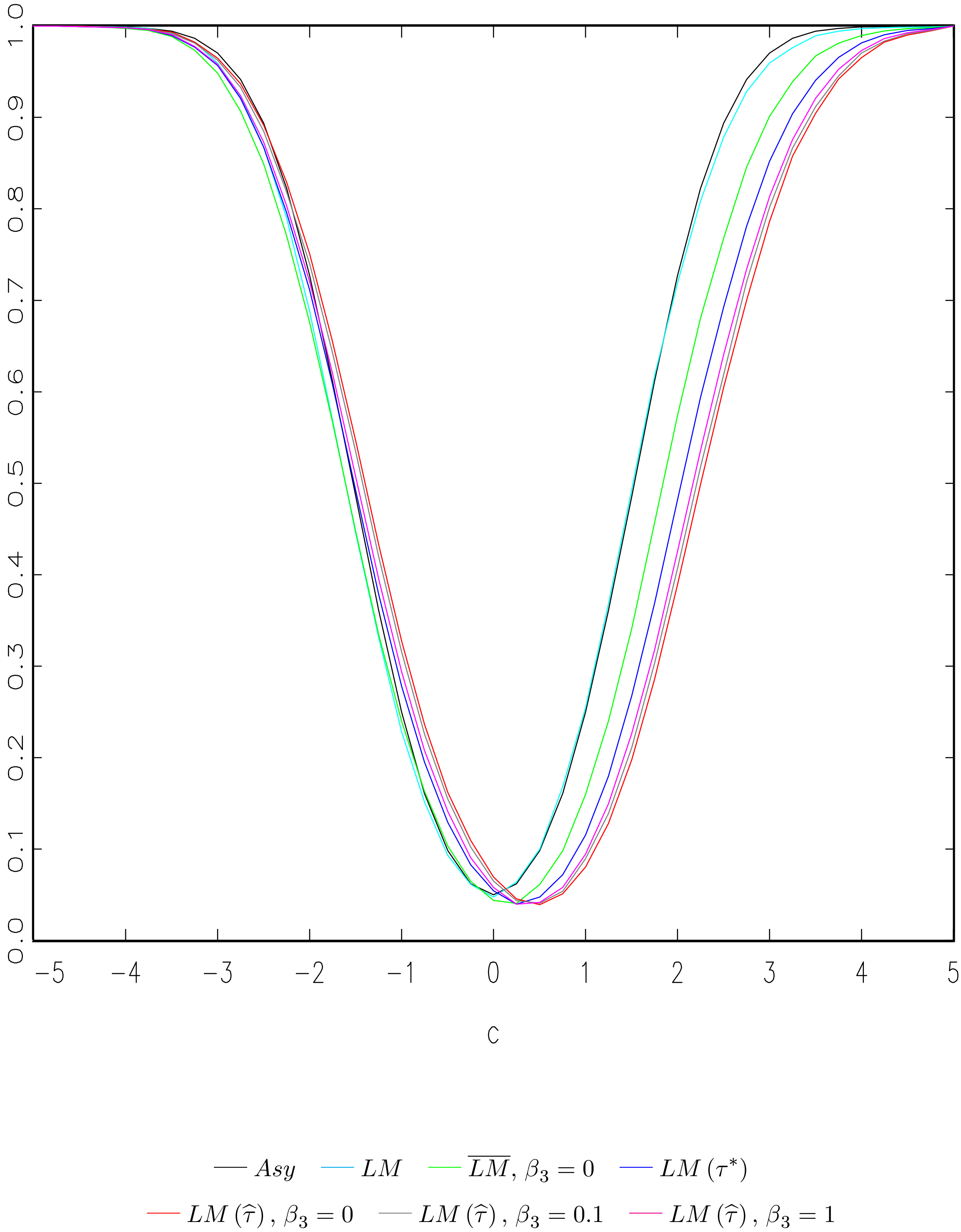
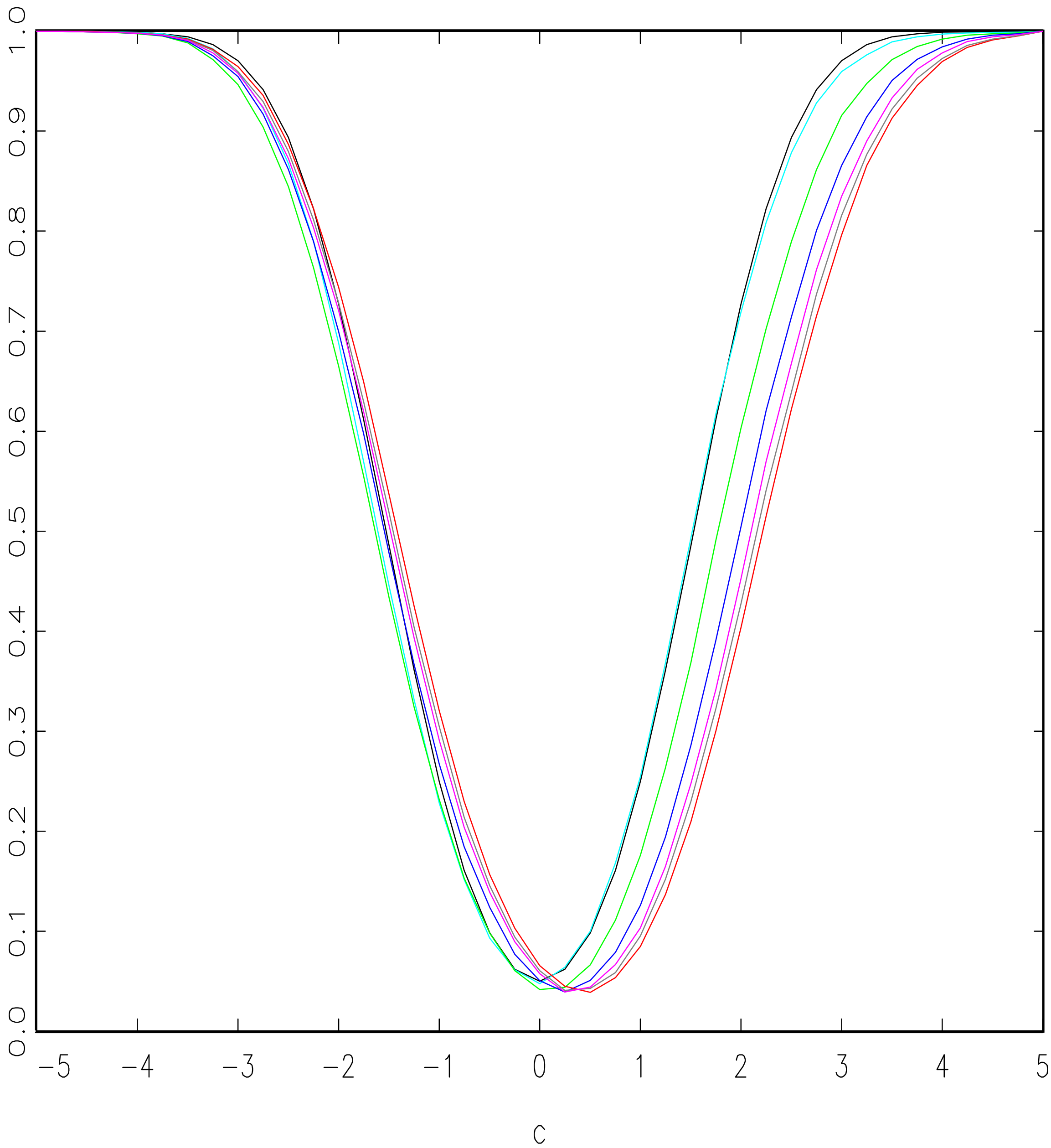


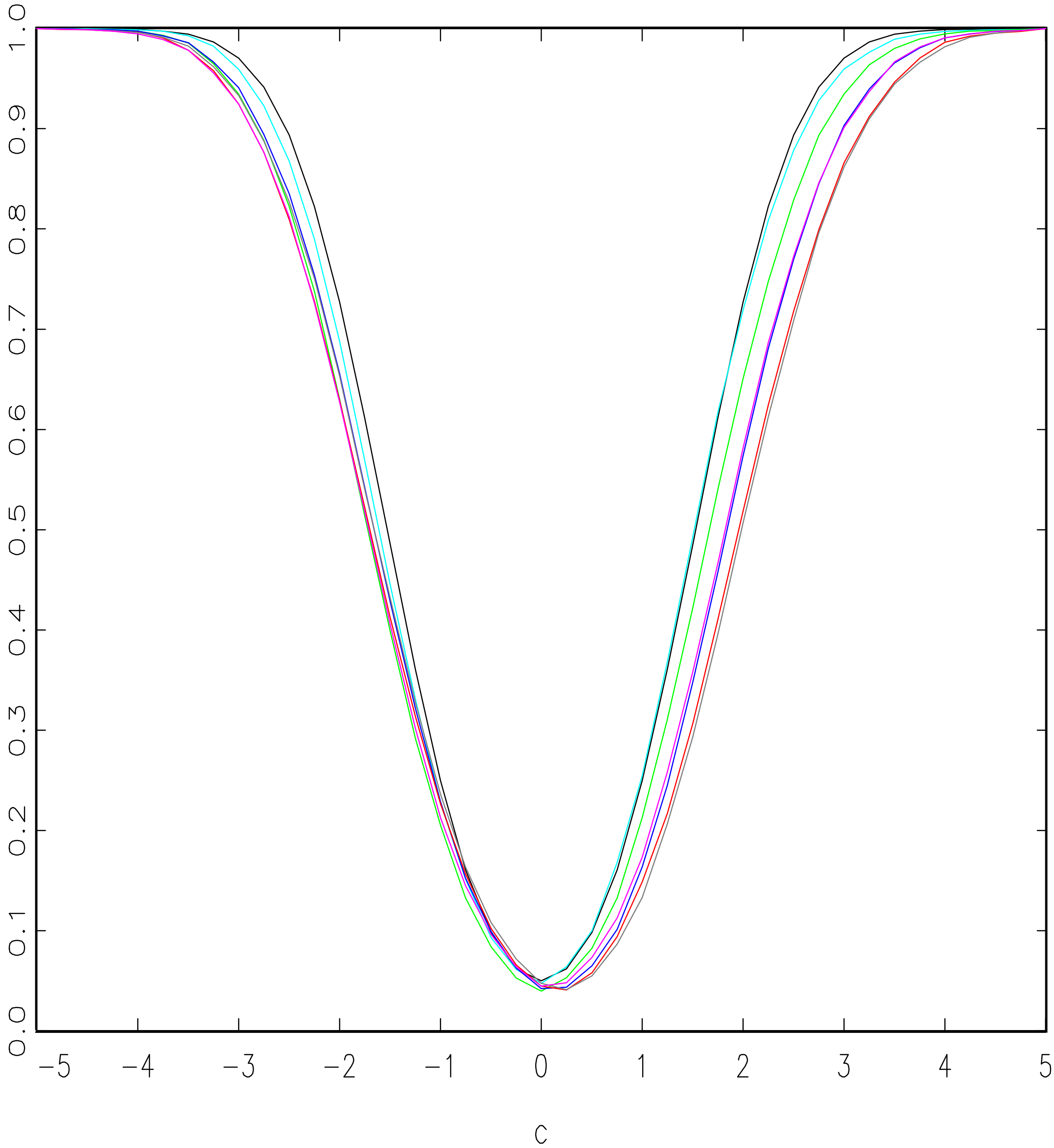
Figure 3. Local power of tests, $T = 512$, $d_0 = 0.5$.



Asy LM $\overline{LM}$, $\beta_3 = 0$ $LM(\tau^*)$

$LM(\hat{\tau})$, $\beta_3 = 0$ $LM(\hat{\tau})$, $\beta_3 = 0.1$ $LM(\hat{\tau})$, $\beta_3 = 1$

Figure 4. Local power of tests, $T = 512$, $d_0 = 0.75$.



Asy LM $\overline{LM}$, $\beta_3 = 0$ $LM\,(\tau^*)$

$LM\,(\widehat{\tau})\,,\ \beta_3 = 0$ $LM\,(\widehat{\tau})\,,\ \beta_3 = 0.1$ $LM\,(\widehat{\tau})\,,\ \beta_3 = 1$

Figure 5. Local power of tests, $T = 512$, $d_0 = 1$.

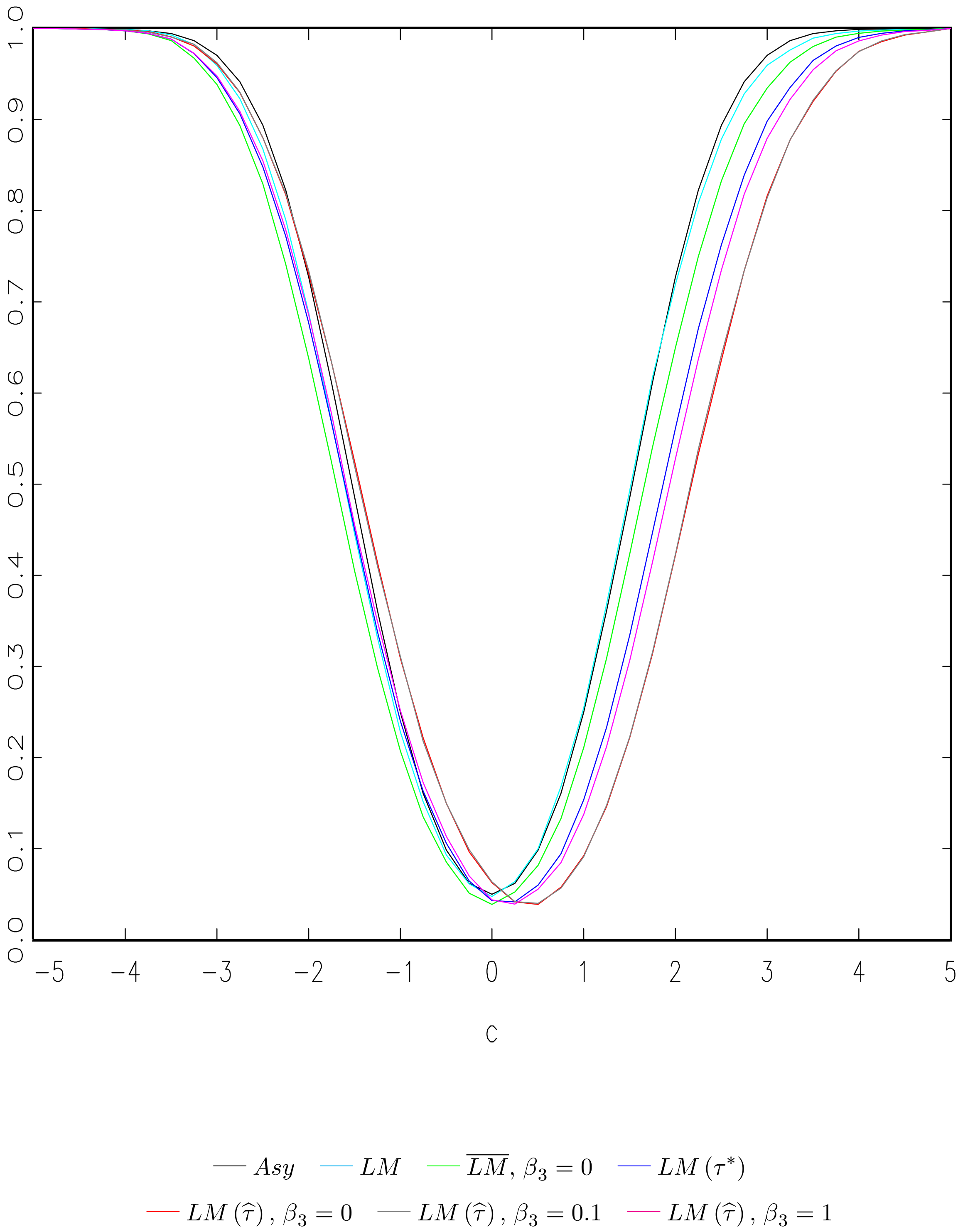


Figure 6. Local power of tests, $T = 512$, $d_0 = 1.25$.

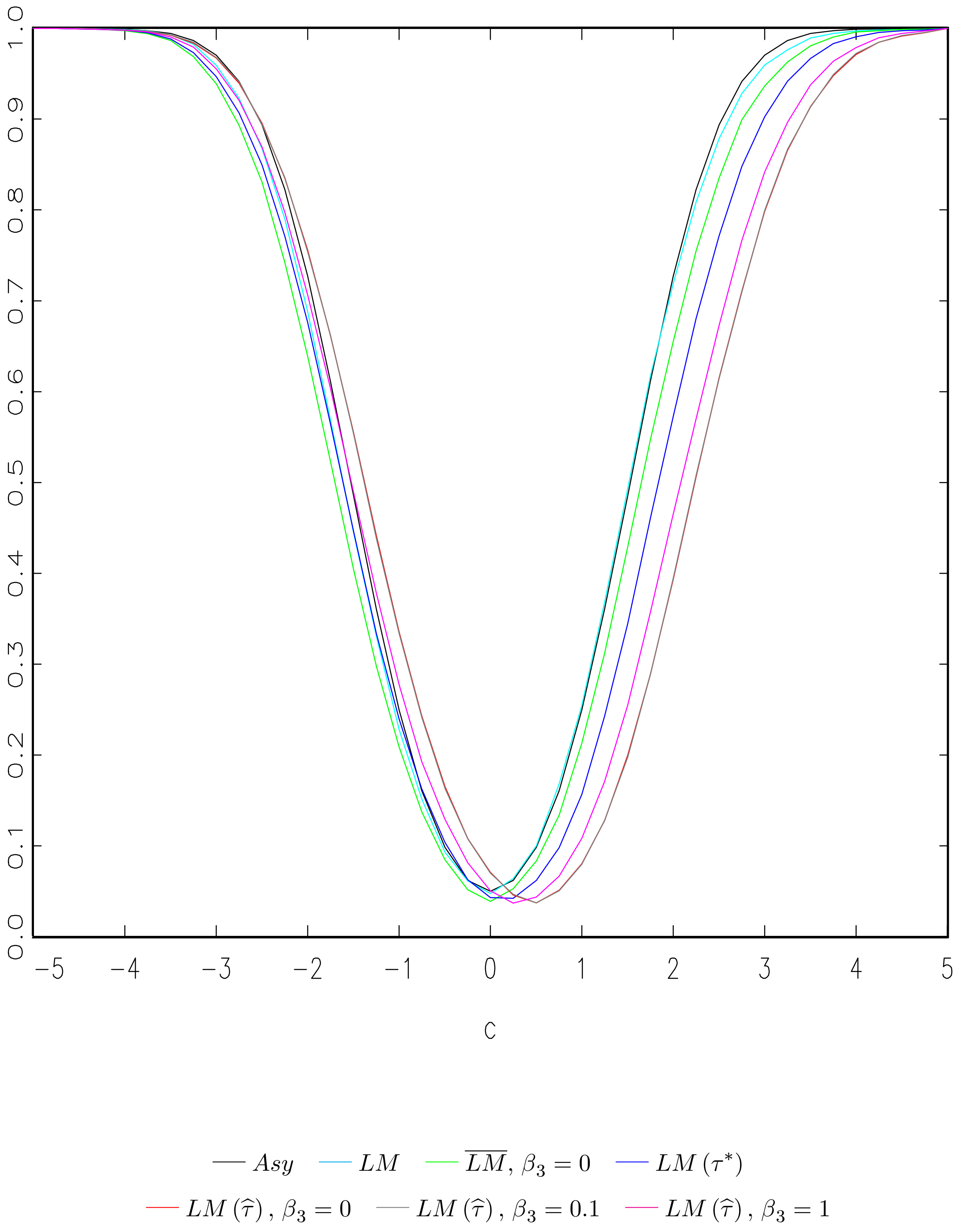


Figure 7. U.S. CPI price index, January 1970 - January 2018.

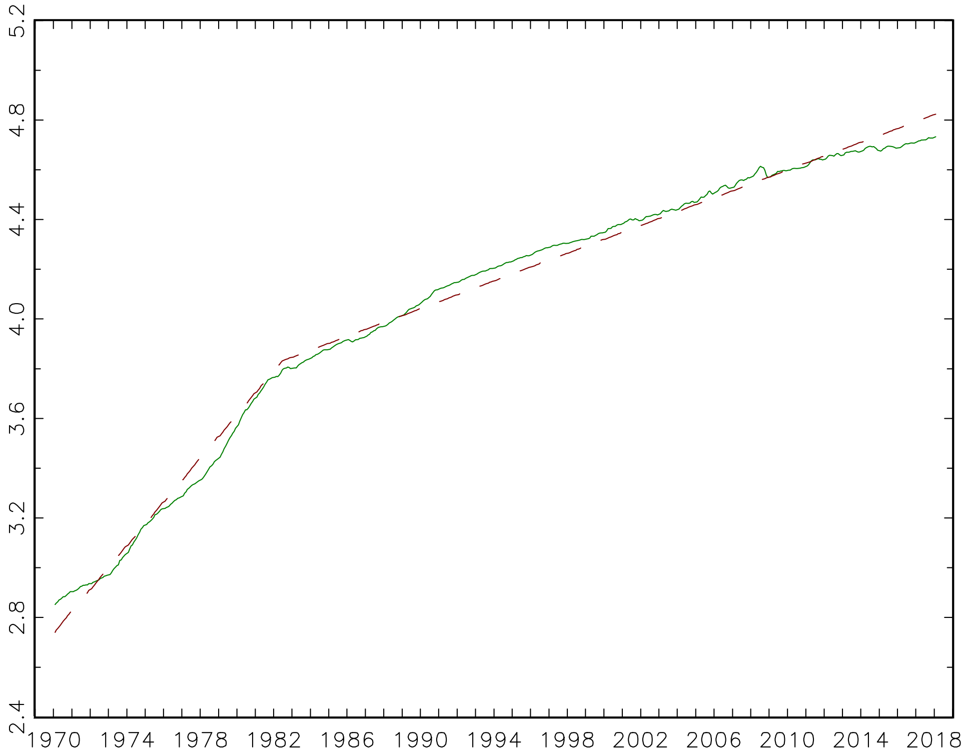
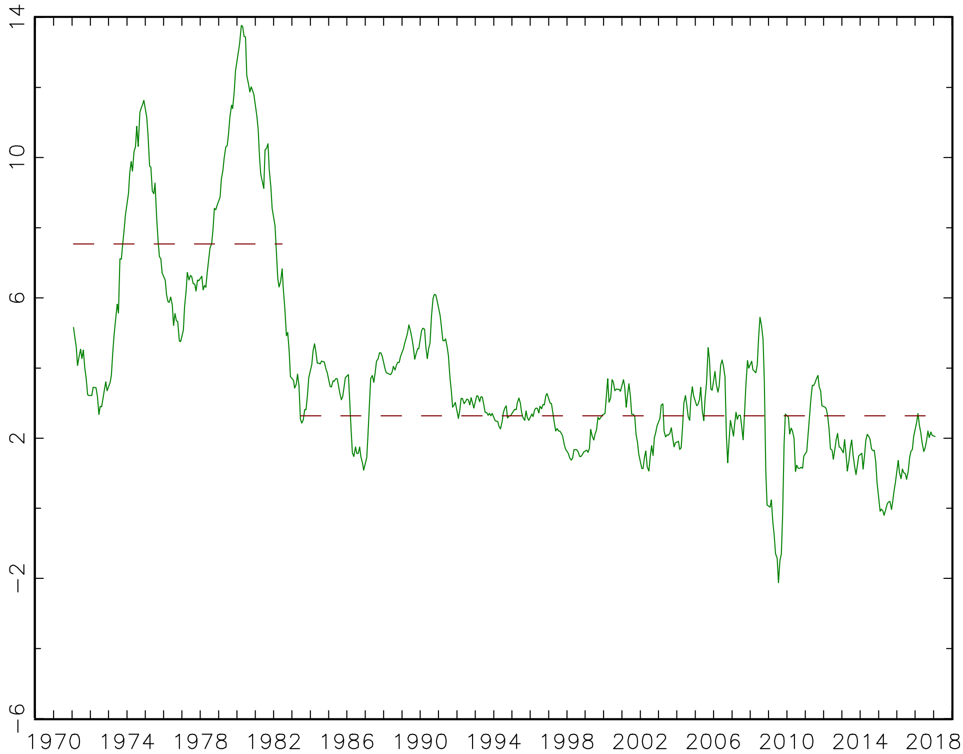


Figure 8. U.S. CPI annual inflation (%), January 1971 - January 2018.



Supplementary Online Appendix

to

"Testing the Order of Fractional Integration of a Time Series in the Possible Presence of a Trend Break at an Unknown Point"

by

F. Iacone, S.J. Leybourne and A.M.R. Taylor

Date: August 15, 2018

Contents: Section S.1 of this supplement contains proofs of Lemmas A1, B1 and C1 used in the proof of Lemma 1 and Lemmas A2, B2, C2 and D2 used in the proof of Theorem 1. Section S.2 provides additional Monte Carlo simulation results relating to empirical power properties against fixed alternatives, the impact of innovation distributions which violate the moment conditions in Assumption 1, and the use of model selection methods to select the ARMA component of the model. Additional references not cited in the main article are included at the end of the supplement.

S.1 Mathematical Proofs

Proof of Lemma A1:

For Model A, (A.1) is established, in the Skorohod measure, for example, by Iacone, Leybourne and Taylor (2013a), page 417. For Model B, rate (A.1) in the Skorohod measure is established for the type 1 version of the fractionally integrated process, for example, by Iacone, Leybourne and Taylor (2014); however, the same result can be derived for the type 2 version using the FCLT in Marinucci and Robinson (2000). Both results are established using the FCLT $T^{-1/2+\delta}\sum_{t=1}^{\lfloor \tau T\rfloor} u_t \Rightarrow \sigma_\infty W(\tau;\delta)$ where $W(\tau;\delta)$ is a Type 2 fractional Browinan motion, and the weak convergence is in the Skorohod measure. To show that this convergence also holds in the uniform metric, we follow Billingsley (1968), page 153; for the weak convergence $X_n \Rightarrow X$ it is possible to go from the Skorohod to the uniform metric if: (i) the limit object X lies in $C[0,1]$, the space of continuous function in $[0,1]$ with the uniform metric, with probability 1, and (ii) the jumps of X_n occur at fixed time points rather than at time points with random position. This applies not only to the standard Brownian motion, but also to both type 1 and type 2 fractional Brownian motions; see Shao (2011) page 604 for an application of this result for type 1 processes. For condition (i), notice that the type 2 fractional Brownian motion also has almost surely continuous sample paths, see Marinucci and Robinson (1999) page 116. Condition (ii) is immediately met.

Proof of Lemma B1:

For Model A, (A.2) follows from Chang and Perron (2016), Theorem 1 and Theorem 2, part i (case for $m = 0$). Chang and Perron (2016) derive their results for type 1 fractionally integrated processes, but the same results can be derived for the type 2 version using the FCLT in Marinucci and Robinson (2000) and bounds from Lavielle and Moulines (2000); in particular, the Hájek-Rényi type inequality in Lavielle and Moulines (2000) holds for both type 1 and type 2 processes.

For Model B, Theorem 3 and Theorem 7 of Lavielle and Moulines (2000) yield (A.3) for $\tau^* \in [\tau_U, \tau_L] \subset (0,1)$. Regarding the case $\delta < 0$ for Model B, notice that, although Lavielle and Moulines (2000) focus attention on $\delta > 0$, their condition $H1(\phi)$ is still met when $\delta < 0$, with $\phi = 1$; see Lavielle and Moulines (2000) page 35, where the sufficient condition $\sum_{s\geq 0} |\mathrm{E}(u_t u_{t+s})| < \infty$ is given.

Finally, for Model A, rate (A.4) again follows by adapting results from Theorem 4 of Chang and Perron (2016). For Model B with $\delta = 0$, (A.4) is given in Bai (1994), Proposition 4; Lavielle and Moulines (2000), Theorem 8 establish (A.4), focusing on the case of a shrinking break, and $\delta > 0$. Lavielle and Moulines (2000) do not explicitly consider $\delta < 0$ altogether, but we show below that the result follows applying the bound in Corollary 2.1 of Lavielle and Moulines (2000) to the expression in Proposition 4 of Bai (1994). Using our notation, the expression in the proof of Proposition 4 of Bai (1994) is given by

$$\widehat{\beta}_2(\widehat{\tau}) - \widehat{\beta}_2(\tau^*) = \left(\frac{\lfloor\tau^*T\rfloor - \lfloor\widehat{\tau}T\rfloor}{\lfloor\tau^*T\rfloor\lfloor\widehat{\tau}T\rfloor}\sum_{t=1}^{\lfloor\tau^*T\rfloor} u_t - \frac{1}{\lfloor\widehat{\tau}T\rfloor}\sum_{t=1+\lfloor\widehat{\tau}T\rfloor}^{\lfloor\tau^*T\rfloor} u_t\right)\mathbb{I}\left(\lfloor\widehat{\tau}T\rfloor \leq \lfloor\tau^*T\rfloor\right) \tag{S.1}$$

$$+\left(\frac{\lfloor\tau^*T\rfloor - \lfloor\widehat{\tau}T\rfloor}{\lfloor\tau^*T\rfloor\lfloor\widehat{\tau}T\rfloor}\sum_{t=1}^{\lfloor\tau^*T\rfloor} u_t + \frac{1}{\lfloor\widehat{\tau}T\rfloor}\sum_{t=1+\lfloor\tau^*T\rfloor}^{\lfloor\widehat{\tau}T\rfloor} u_t + \beta_3\frac{\lfloor\tau^*T\rfloor - \lfloor\widehat{\tau}T\rfloor}{\lfloor\widehat{\tau}T\rfloor}\right)\mathbb{I}\left(\lfloor\widehat{\tau}T\rfloor > \lfloor\tau^*T\rfloor\right). \tag{S.2}$$

Because $\lfloor\tau^*T\rfloor - \lfloor\widehat{\tau}T\rfloor = O_p(1)$ and $\sum_{t=1}^{\lfloor\tau^*T\rfloor} u_t = O_p\left(T^{1/2+\delta}\right)$, the first term on the right hand side of (S.1) is $O_p\left(1 \times T^{-2} \times T^{1/2+\delta}\right) = O_p\left(T^{-3/2+\delta}\right) = o_p\left(T^{-1/2+\delta}\right)$. As for the second term of (S.1), we now show that, for $\varepsilon > 0$, $\sum_{t=1+\lfloor\widehat{\tau}T\rfloor}^{\lfloor\tau^*T\rfloor} u_t = O_p(T^{\varepsilon})$. It follows from Equation (8) of Lavielle and Moulines (2000) that for $\varepsilon > 0$,

$$\sup_{i\in\mathbb{Z}} P\left(\max_{k+i\geq m+i} k^{-(1/2+\varepsilon)}\left|\sum_{t=i}^{i+k} u_t\right| \geq c\right) \leq C(1,\varepsilon)\, m^{1-2(1/2+\varepsilon)}$$

if $\delta < 0$ and

$$\sup_{i\in\mathbb{Z}} P\left(\max_{k+i\geq m+i} k^{-(1/2+\delta+\varepsilon)}\left|\sum_{t=i}^{i+k} u_t\right| \geq c\right) \leq C(1,\varepsilon)\, m^{1-2(1/2+\delta+\varepsilon)}$$

if $\delta > 0$. Either way, then,

$$\sup_{i\in\mathbb{Z}} P\left(\max_{k+i\geq m+i} k^{-1}\left|\sum_{t=i}^{i+k} u_t\right| \geq c\right) \leq C(1,1)\, m^{-1}.$$

Taking $i = \lfloor \tau T \rfloor$, $k = \lfloor \tau^* T \rfloor - \lfloor \tau T \rfloor + T^{\varepsilon}$ for $\varepsilon > 0$ we can then allow for $m \to \infty$ and therefore, uniformly in τ, $\left(\lfloor \tau^* T \rfloor - \lfloor \tau T \rfloor + T^{\varepsilon}\right)^{-1} \left| \sum_{t=1+\lfloor \tau T \rfloor}^{\lfloor \tau^* T \rfloor + T^{\varepsilon}} u_t \right| = O_p(1)$. Next, notice that

$$\begin{aligned} \left| \sum_{t=1+\lfloor \tau T \rfloor}^{\lfloor \tau^* T \rfloor} u_t \right| &= \left| \sum_{t=1+\lfloor \tau T \rfloor}^{\lfloor \tau^* T \rfloor + T^{\varepsilon}} u_t - \sum_{t=1+\lfloor \tau^* T \rfloor}^{\lfloor \tau^* T \rfloor + T^{\varepsilon}} u_t \right| \\ &\leq \left| \sum_{t=1+\lfloor \tau T \rfloor}^{\lfloor \tau^* T \rfloor + T^{\varepsilon}} u_t \right| + \left| \sum_{t=1+\lfloor \tau^* T \rfloor}^{\lfloor \tau^* T \rfloor + T^{\varepsilon}} u_t \right| \\ &= O_p\left(\left(\lfloor \tau^* T \rfloor - \lfloor \tau T \rfloor + T^{\varepsilon}\right) + T^{\varepsilon}\right) \end{aligned}$$

and that $\left| \sum_{t=1+\lfloor \widehat{\tau} T \rfloor}^{\lfloor \tau^* T \rfloor} u_t \right| = O_p(T^{\varepsilon})$, using $\lfloor \tau^* T \rfloor - \lfloor \widehat{\tau} T \rfloor = O_p(1)$. Finally, therefore we have that the second term on the right hand side of (S.1) is such that

$$\frac{1}{\lfloor \widehat{\tau} T \rfloor} \sum_{t=1+\lfloor \widehat{\tau} T \rfloor}^{\lfloor \tau^* T \rfloor} u_t = O_p\left(T^{\varepsilon - 1}\right) = o_p\left(T^{-1/2+\delta}\right).$$

Proceeding in the same way, we can also show that the first two terms in (S.2) are of $o_p\left(T^{-1/2+\delta}\right)$. Finally, the remainder term $\beta_3 \frac{\lfloor \tau^* T \rfloor - \lfloor \widehat{\tau} T \rfloor}{\lfloor \widehat{\tau} T \rfloor} = O_p\left(T^{-1}\right) = o_p\left(T^{-1/2+\delta}\right)$ using (A.3). As in Proposition 4 of Bai (1994), the proof for $\widehat{\beta}_3(\widehat{\tau}) - \widehat{\beta}_3(\tau^*) = o_p\left(T^{-1/2+\delta}\right)$ proceeds in the same way, and we can then conclude that $\widehat{\beta}(\widehat{\tau}) - \widehat{\beta}(\tau^*) = o_p\left(T^{-1/2+\delta}\right)$. Rearranging,

$$K_T(d)\left(\widehat{\beta}(\widehat{\tau}) - \beta\right) = K_T(d)\left(\widehat{\beta}(\widehat{\tau}) - \widehat{\beta}(\tau^*) + \widehat{\beta}(\tau^*) - \beta\right) = K_T(d)\left(\widehat{\beta}(\widehat{\tau}) - \widehat{\beta}(\tau^*)\right) + K_T(d)\left(\widehat{\beta}(\tau^*) - \beta\right).$$

Using the rate for $\widehat{\beta}(\widehat{\tau}) - \widehat{\beta}(\tau^*)$, then $K_T(d)\left(\widehat{\beta}(\widehat{\tau}) - \widehat{\beta}(\tau^*)\right) = o_p(1)$; the rate $K_T(d)\left(\widehat{\beta}(\tau^*) - \beta\right) = O_p(1)$ follows because τ^* is not random and therefore $\widehat{\beta}(\tau^*)$ is a standard regression estimate with non-random regressors, also see in Robinson (1994) and Nielsen (2004). These two rates are sufficient to establish (A.4).

Proof of Lemma C1:

By a third order expansion and the mean value theorem,

$$\begin{aligned} \left((\ln(\Delta))^r \Delta^{-(\alpha+\theta_T)}\right)_+ \eta_t &= \left((\ln(\Delta))^r \Delta^{-\alpha}\right)_+ \eta_t - \theta_T \left((\ln(\Delta))^{r+1} \Delta^{-\alpha}\right)_+ \eta_t \\ &\quad + 1/2\,(\theta_T)^2 \left((\ln(\Delta))^{r+2} \Delta^{-\alpha}\right)_+ \eta_t \\ &\quad - 1/6\,(\theta_T)^3 \left((\ln(\Delta))^{r+3} \Delta^{-\left(\alpha+\theta_{m,T}\right)}\right)_+ \eta_t \end{aligned}$$

for $|\theta_{m,T}| \leq |\theta_T|$. Then proceeding as in Lemma 4 of Robinson (2005), we write

$$\left((\ln(\Delta))^{r+3} \Delta^{-\left(\alpha+\theta_{m,T}\right)}\right)\{\eta_t \mathbb{I}(t>0)\} = \sum_{j=1}^{t-1} c_j \eta_{t-j}$$

where c_j is the coefficient of s^j in the Taylor expansion of $\{\ln(1-s)\}^{r+3} \times (1-s)^{-\left(\alpha+\theta_{m,T}\right)}$. From Stirling's approximation, also see (7.3) of Robinson (2005), $c_j \sim (\ln(j))^{r+3} \times j^{-\left(\alpha+\theta_{m,T}\right)-1}$. As $|\alpha| < 1/2$, then, for T large enough, $-\left(\alpha+\theta_{m,T}\right) - 1 < -1/2$, and $\sum_{j=1}^{\infty} c_j^2 < C$. Then, by the

Cauchy-Schwarz inequality, $\sum_{j=1}^{t-1} c_j \eta_{t-j} \leq \left(\sum_{j=1}^{t-1} c_j^2 \sum_{j=1}^{t-1} \eta_{t-j}^2 \right)^{1/2} \leq C \left(\sum_{j=1}^{t-1} \eta_j^2 \right)^{1/2}$ and we established the bound

$$\left(\left(\ln\left(\Delta\right)\right)^{r+3} \Delta^{-\left(\alpha+\theta_{m,T}\right)} \right) \left\{ \eta_t \mathbb{I}\left(t>0\right) \right\} = O\left(\left\{ \sum_{j=1}^{t-1} \eta_j^2 \right\}^{1/2} \right) = O_p\left(t\right)$$

as $\mathrm{E}\left(\eta_t\right)^2 = O\left(1\right)$.

We then rewrite

$$\begin{aligned}
& T^{-(1/2+\alpha)} \left(\ln\left(T\right)\right)^{-r} \left| \sum_{t=1}^{\lfloor \tau T \rfloor} \left(\left(\left(\ln\left(\Delta\right)\right)^{r} \Delta^{-(\alpha+\theta_T)} \right)_+ \eta_t - \left(\left(\ln\left(\Delta\right)\right)^{r} \Delta^{-\alpha} \right)_+ \eta_t \right) \right| \\
\leq\ & \left|\theta_T\right| T^{-(1/2+\alpha)} \left(\ln\left(T\right)\right)^{-r} \left| \sum_{t=1}^{\lfloor \tau T \rfloor} \left(\left(\ln\left(\Delta\right)\right)^{r+1} \Delta^{-\alpha} \right)_+ \eta_t \right| \qquad \text{(S.3)} \\
& + \frac{1}{2} \theta_T^2 T^{-(1/2+\alpha)} \left(\ln\left(T\right)\right)^{-r} \left| \sum_{t=1}^{\lfloor \tau T \rfloor} \left(\left(\ln\left(\Delta\right)\right)^{r+2} \Delta^{-\alpha} \right)_+ \eta_t \right| \qquad \text{(S.4)} \\
& + O_p\left(T^{-(1/2+\alpha)} \sum_{t=1}^{T} t^{1/2} \left|\theta_T\right|^3 \right). \qquad \text{(S.5)}
\end{aligned}$$

From Marinucci and Robinson (2000) and the rate for θ_T, the term in (S.3) is $O_p\left(T^{-1/2} \ln\left(T\right)\right) = o_p\left(1\right)$, and the term in (S.4) can be treated in the same way. The remainder (S.5) is $O_p\left(T^{-(1/2+\alpha)}\right) = o_p\left(1\right)$.

Proof of Lemma A2:

We first need to introduce some additional notation, as in Iacone, Leybourne and Taylor (2013b). To that end, we define

$$\begin{aligned}
\mu_{1,t} :=\ & \Delta^{\delta} \left\{ 1\mathbb{I}\left(t>0\right) \right\}, \ \mu_{2,t} := \Delta^{\delta} \left\{ t\mathbb{I}\left(t>0\right) \right\}, \\
\mu_{3,t}\left(\tau\right) :=\ & \begin{cases} \Delta^{\delta} \left\{ \left(t - \lfloor \tau T \rfloor\right) \mathbb{I}\left(t > \lfloor \tau T \rfloor\right) \right\} & \text{for Model A} \\ \Delta^{\delta} \left\{ 1\mathbb{I}\left(t > \lfloor \tau T \rfloor\right) \right\} & \text{for Model B} \end{cases}
\end{aligned}$$

where, for $\delta \in (-1/2, 0) \cup (0, 1/2)$, we observe from Lemma 1 of Robinson (2005) and Iacone, Leybourne and Taylor (2013b), page 40, that

$$\begin{aligned}
\mu_{1,t} &= \frac{1}{\Gamma\left(1-\delta\right)} t^{-\delta} + O\left(t^{-1-\delta} + t^{-1} \mathbb{I}\left(\delta>0\right) \right), \ \Delta\mu_{1,t} = \Delta_t^{(-\delta)} \\
\mu_{2,t} &= \frac{1}{\Gamma\left(2-\delta\right)} t^{1-\delta} + \left(t^{-\delta} + 1\mathbb{I}\left(\delta>0\right) \right), \ \Delta\mu_{2,t} = \mu_{1,t}.
\end{aligned}$$

Next we define $\widehat{\varepsilon}_t(\psi) := g(L;\psi)\Delta_+^{\delta} u_t$ and $\widehat{\varepsilon}_t\left(\psi;\tau\right) := g\left(L;\psi\right) \Delta_+^{\delta} \widehat{u}_t\left(\tau\right)$. Notice therefore that, under H_0, $\widehat{\varepsilon}_t(\widehat{\psi})$ and $\widehat{\varepsilon}_t(\widehat{\psi}(\tau);\tau)$ coincide with $\widehat{\varepsilon}_t$ defined in (3.2) and $\widehat{\varepsilon}_t\left(\tau\right)$ defined in (3.8), respectively. Moreover, under H_0, $\widehat{\varepsilon}_t\left(\psi^*\right) = \varepsilon_t$.

We may then write the loss functions in (3.1) and (3.7) as $\sum_{t=1}^{T} \left(\widehat{\varepsilon}_t\left(\psi\right)\right)^2$ and $\sum_{t=1}^{T} \left(\widehat{\varepsilon}_t\left(\psi;\tau\right)\right)^2$, respectively. Consistency of $\widehat{\psi}$ is well known in this context, and can be readily established using a routine consistency argument for implicitly defined extremum estimates; see, for example, Newey and McFadden (1994). This requires uniform (in ψ) convergence of a suitably scaled version of the loss function so that $T^{-1}\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\psi\right)\right)^2 \xrightarrow{p} \mathrm{E}\left(g\left(L;\psi\right)\eta_t\right)^2$, together with identification of the parameters ψ_0. The former is established as a uniform weak law of large numbers, that is obtained using pointwise convergence of the scaled loss function $T^{-1}\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\psi\right)\right)^2$ to the limit, and stochastic equicontinuity; see page 244 of Andrews (1992). Sufficient conditions for stochastic equicontinuity to hold in this case are that the loss function is differentiable with first derivative bounded in probability; see Assumptions (b) and (c) on page 246 of Andrews (1992).

Using the same approach as in Theorem A1 of Andrews (1993), to establish part (i) of the lemma we need to verify that $T^{-1}\left(\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\psi;\tau\right)\right)^2 - \sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\psi\right)\right)^2\right) = o_p\left(1\right)$ uniformly in both ψ and τ. Uniformity in ψ can be established using the same arguments outlined above for the case of estimating $\widehat{\psi}$. We therefore focus here on establishing uniform convergence in τ.

Substituting (3.6) into the definition for $\widehat{\varepsilon}_t\left(\psi;\tau\right)$, we have that when $d_0 < 0.5$,

$$\begin{aligned}
\widehat{\varepsilon}_t\left(\psi;\tau\right) &= g\left(L;\psi\right)\Delta_+^{\delta}\left(y_t - z_t\left(\tau\right)'\widehat{\beta}\left(\tau\right)\right) \\
&= g\left(L;\psi\right)\Delta_+^{\delta}u_t + g\left(L;\psi\right)\Delta_+^{\delta}z_t\left(\tau\right)'\left(\beta - \widehat{\beta}\left(\tau\right)\right) && \text{(S.6)} \\
&= \widehat{\varepsilon}_t\left(\psi\right) + g\left(L;\psi\right)\Delta_+^{\delta}z_t\left(\tau\right)'\left(\beta - \widehat{\beta}\left(\tau\right)\right) && \text{(S.7)}
\end{aligned}$$

and that

$$\begin{aligned}
\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\psi;\tau\right)\right)^2 - \sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\psi\right)\right)^2 &= \sum_{t=1}^{T}\left(g\left(L;\psi\right)\Delta_+^{\delta}z_t\left(\tau\right)'\left(\beta - \widehat{\beta}\left(\tau\right)\right)\right)^2 && \text{(S.8)} \\
&\quad + 2\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\psi\right)\right)\left(g\left(L;\psi\right)\Delta_+^{\delta}z_t\left(\tau\right)'\left(\beta - \widehat{\beta}\left(\tau\right)\right)\right). && \text{(S.9)}
\end{aligned}$$

When $d_0 > 0.5$, imposing $\widehat{u}_1\left(\tau\right) = 0$ adds the remainder term

$$-g\left(L;\psi\right)\Delta_t^{(-\delta)}\left\{u_1 + \widehat{r}_1\left(\tau\right)\right\} \tag{S.10}$$

where $\widehat{r}_t\left(\tau\right) := \beta_1 + \left(\beta_2 - \widehat{\beta}_2\left(\tau\right)\right)\left\{\mathbb{I}\left(t > 0\right)\right\}$.

Consider Model A first. Using $\left(a+b\right)^2 \leq 2a^2 + 2b^2$, the right hand side of (S.8) is bounded by

$$\begin{aligned}
& C\sum_{t=1}^{T}\left(g\left(L;\psi\right)\mu_{1,t}\right)^2\left(\beta_1 - \widehat{\beta}_1\left(\tau\right)\right)^2 + C\sum_{t=1}^{T}\left(g\left(L;\psi\right)\mu_{2,t}\right)^2\left(\beta_2 - \widehat{\beta}_2\left(\tau\right)\right)^2 \\
& + C\sum_{t=1}^{T}\left(g\left(L;\psi\right)\mu_{3,t}\left(\tau\right)\right)^2\left(\widehat{\beta}_3\left(\tau\right)\right)^2 \\
\leq\ & C\sum_{t=1}^{T}\mu_{1,t}^2\left(\beta_1 - \widehat{\beta}_1\left(\tau\right)\right)^2 + C\sum_{t=1}^{T}\mu_{2,t}^2\left(\beta_2 - \widehat{\beta}_2\left(\tau\right)\right)^2 + C\sum_{t=1}^{T}\mu_{3,t}\left(\tau\right)^2\left(\widehat{\beta}_3\left(\tau\right)\right)^2
\end{aligned}$$

using Lemma 3 of Robinson (2005) and $g\left(1;\psi\right)^2 < C$. Then, using the fact that $\sum_{t=1}^{T}\mu_{3,t}\left(\tau\right)^2 = \sum_{t=1+\lfloor\tau T\rfloor}^{T}\mu_{3,t}\left(\tau\right)^2 \leq \sum_{t=1}^{T}\mu_{2,t}^2$, the expression above is seen to be of $O_p\left(1\right)$ using Lemma 1 of Robinson (2005) and Lemma A1. The term in (S.9) is $O_p\left(T^{1/2}\right)$ by the Cauchy-Schwarz inequality.

Next we consider Model B. Here the right hand side of (S.8) is bounded by

$$C\sum_{t=1}^{T}\left(g\left(L;\psi\right)\mu_{1,t}\right)^{2}\left(\beta_{2}-\widehat{\beta}_{2}\left(\tau\right)\right)^{2}+C\sum_{t=1}^{T}\left(g\left(L;\psi\right)\mu_{3,t}\left(\tau\right)\right)^{2}\left(\widehat{\beta}_{3}\left(\tau\right)\right)^{2}$$

which is again $O_p(1)$. Another application of the Cauchy-Schwarz inequality yields that (S.9) is $O_p\left(T^{1/2}\right)$. For Model B we also have to account for the additional remainder term in (S.10): notice that as $e_t=0$ if $t<0$, then $u_1=e_1=\eta_1$, and we can therefore write $g\left(L;\psi\right)\Delta_t^{(-\delta)}u_1=\Delta_t^{(-\delta)}\widehat{\varepsilon}_1\left(\psi\right)$. To account for this term we need to add it to the summations in (S.6) and (S.7): we then analyse

$$\sum_{t=1}^{T}\left(\Delta_t^{(-\delta)}\right)^{2}\left(\widehat{\varepsilon}_1\left(\psi\right)\right)^{2}-2\sum_{t=1}^{T}\Delta_t^{(-\delta)}\widehat{\varepsilon}_1\left(\psi\right)\widehat{\varepsilon}_t\left(\psi\right)-2\sum_{t=1}^{T}\Delta_t^{(-\delta)}\widehat{\varepsilon}_1\left(\psi\right)g\left(L;\psi\right)\Delta_{+}^{\delta}z_t\left(\tau\right)'\left(\beta-\widehat{\beta}\left(\tau\right)\right). \tag{S.11}$$

Noting that $\left(\widehat{\varepsilon}_1\left(\psi\right)\right)^2=O_p\left(1\right)$, uniformly in ψ, and, in view of the fact that $|\Delta_t^{(-\delta)}|\sim Ct^{-\delta-1}$ when $\delta\neq 0$, and that $|\Delta_t^{(-\delta)}|<Ct^{-\delta-1}$, it follows that $\sum_{t=1}^{T}(\Delta_t^{(-\delta)})^2(\widehat{\varepsilon}_1(\psi))^2=O_p(\sum_{t=1}^{T}t^{2(-\delta-1)})=O_p\left(1\right)$. As for the second term, $\sum_{t=1}^{T}\Delta_t^{(-\delta)}\widehat{\varepsilon}_1(\psi)\widehat{\varepsilon}_t(\psi)=O_p(\sum_{t=1}^{T}t^{-\delta-1})$, which is $O_p\left(1\right)$ if $\delta>0$ and $O_p(T^{-\delta})=o_p\left(T^{1/2}\right)$ if $\delta<0$, recalling that $\delta>-0.5$. Finally, by the Cauchy-Schwarz inequality the third term in (S.11) is $O_p\left(1\right)$, so that the whole expression in (S.11) is of $o_p\left(T^{1/2}\right)$. In view of Lemma 3 of Robinson (2005), Lemma A.1 and bound for $|\Delta_t^{(-\delta)}|$, it also holds that the contribution of the remainder $g\left(L;\psi\right)\Delta_t^{(-\delta)}\widehat{r}_1\left(\tau\right)$ is also of order $o_p\left(T^{1/2}\right)$.

Combining the foregoing results we therefore have that

$$\sup_{\tau}\left|\frac{1}{T}\left(\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\psi;\tau\right)\right)^2-\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\psi\right)\right)^2\right)\right|\xrightarrow{p}0.$$

As noted before, this is sufficient to establish that $\widehat{\psi}\left(\tau\right)-\widehat{\psi}=o_p\left(1\right)$, which therefore completes the proof of part (i) of the lemma.

We now turn to the proof of part (ii) of the lemma. Minimisation of the loss functions in (3.1) and (3.7) yields

$$\sum_{t=1}^{T}\widehat{\varepsilon}_t\left(\psi\right)\left.\frac{\partial\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi}\right|_{\psi=\widehat{\psi}}=0\ \text{ and }\ \sum_{t=1}^{T}\widehat{\varepsilon}_t\left(\psi;\tau\right)\left.\frac{\partial\widehat{\varepsilon}_t\left(\psi;\tau\right)}{\partial\psi}\right|_{\psi=\widehat{\psi}(\tau)}=0$$

respectively, where

$$\begin{aligned}\frac{\partial\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi}&:=\frac{\partial}{\partial\psi}g(L;\psi)\Delta_{+}^{\delta}u_t\\ \frac{\partial^2\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi\partial\psi'}&:=\frac{\partial^2}{\partial\psi\partial\psi'}g(L;\psi)\Delta_{+}^{\delta}u_t.\end{aligned}$$

Recalling (S.7), we have that

$$\begin{aligned}\frac{\partial\widehat{\varepsilon}_t\left(\psi;\tau\right)}{\partial\psi}&=\frac{\partial\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi}+\frac{\partial}{\partial\psi}\left(g\left(L;\psi\right)\Delta_{+}^{\delta}z_t\left(\tau\right)'\left(\beta-\widehat{\beta}\left(\tau\right)\right)\right)\\ \frac{\partial^2\widehat{\varepsilon}_t\left(\psi;\tau\right)}{\partial\psi\partial\psi'}&=\frac{\partial^2\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi\partial\psi'}+\frac{\partial^2}{\partial\psi\partial\psi'}\left(g\left(L;\psi\right)\Delta_{+}^{\delta}z_t\left(\tau\right)'\left(\beta-\widehat{\beta}\left(\tau\right)\right)\right).\end{aligned}$$

As with the treatment of (S.6) and (S.7) above, these expressions should properly be augmented by additional remainder terms under Model B. However, proceeding as in the derivation of (S.11) above, these can be ignored with no loss of asymptotic generality and we shall therefore do so hereafter in the interests in brevity. Next, we define

$$\begin{aligned} D_1(\psi) :=& \quad \frac{1}{T}\sum_{t=1}^{T} \frac{\partial \widehat{\varepsilon}_t(\psi)}{\partial \psi}\frac{\partial \widehat{\varepsilon}_t(\psi)}{\partial \psi'},\ D_2(\psi) := \frac{1}{T}\sum_{t=1}^{T} \widehat{\varepsilon}_t(\psi)\frac{\partial^2 \widehat{\varepsilon}_t(\psi)}{\partial \psi \partial \psi'} \\ D(\psi) :=& \quad D_1(\psi) + D_2(\psi) \end{aligned}$$

and we denote by $[D(\psi)]_i$ the i-th row of matrix $D(\psi)$. A mean value theorem expansion of the first order conditions from loss function (3.1) for the infeasible estimate $\widehat{\psi}$ yields, for the i-th element, $\widehat{\psi}_i$, of $\widehat{\psi}$,

$$\sum_{t=1}^{T} \widehat{\varepsilon}_t(\psi)\left.\frac{\partial \widehat{\varepsilon}_t(\psi)}{\partial \psi_i}\right|_{\psi=\psi^*} + \left[D\left(\widetilde{\psi}^i\right)\right]_i\left(\widehat{\psi}-\psi^*\right) = 0 \tag{S.12}$$

where $\widetilde{\psi}^i$ is a $(p+q)$ dimensional vector such that $\|\widetilde{\psi}^i - \psi^*\| \leq \|\widehat{\psi}-\psi^*\|$. Stacking the rows $\left[D\left(\widetilde{\psi}^i\right)\right]_i$ for all i, denote

$$\widetilde{D}\left(\widehat{\psi}\right) := \begin{pmatrix} \left[D\left(\widetilde{\psi}^1\right)\right]_1 \\ \dots \\ \left[D\left(\widetilde{\psi}^{p+q}\right)\right]_{p+q} \end{pmatrix}$$

and, stacking rows of (S.12) for each i and multiplying by $T^{1/2}$, we get

$$T^{-1/2}\sum_{t=1}^{T} \widehat{\varepsilon}_t(\psi)\left.\frac{\partial \widehat{\varepsilon}_t(\psi)}{\partial \psi}\right|_{\psi=\psi^*} + \widetilde{D}\left(\widehat{\psi}\right)T^{1/2}\left(\widehat{\psi}-\psi^*\right) = 0. \tag{S.13}$$

Notice that $\widetilde{D}(\widehat{\psi}) \to_p \Phi\sigma_\varepsilon^2$; see, for example, Nielsen (2004), part (iii) of the proof of Theorem 4.1 (the limit for $\widetilde{D}(\widehat{\psi})$ is included in the limit in Nielsen, 2004, as it is a $(p+q)$ sub-matrix of the matrix in the limit in (iii)), and that $T^{-1/2}\sum_{t=1}^{T} \widehat{\varepsilon}_t(\psi)\left.\frac{\partial \widehat{\varepsilon}_t(\psi)}{\partial \psi}\right|_{\psi=\psi^*} = O_p(1)$; see, for example, Nielsen (2004), part (ii) of the proof of Theorem 4.1. This therefore implies that $T^{1/2}(\widehat{\psi}-\psi^*) = O_p(1)$ (indeed it is clear from part (ii) of the proof of Theorem 4.1 of Nielsen (2004) that $T^{1/2}(\widehat{\psi}-\psi^*)$ has a limiting normal distribution with mean zero under H_0).

To prove (ii) in Lemma A2, we derive an expression similar to (S.13) for the feasible estimate $\widehat{\psi}(\tau)$, from which we can obtain a formula for $\widehat{\psi}(\tau)$. Then, define

$$\begin{aligned} D_1(\psi;\tau) :=& \quad \frac{1}{T}\sum_{t=1}^{T} \frac{\partial \widehat{\varepsilon}_t(\psi;\tau)}{\partial \psi}\frac{\partial \widehat{\varepsilon}_t(\psi;\tau)}{\partial \psi'},\ D_2(\psi;\tau) := \frac{1}{T}\sum_{t=1}^{T} \widehat{\varepsilon}_t(\psi;\tau)\frac{\partial^2 \widehat{\varepsilon}_t(\psi;\tau)}{\partial \psi \partial \psi'} \\ D(\psi;\tau) :=& \quad D_1(\psi;\tau) + D_2(\psi;\tau) \end{aligned}$$

and apply the mean value theorem expansion of the first order conditions from loss function (3.7) as we did for (3.1) beforehand. We then obtain, for the i-th element, $\widehat{\psi}_i(\tau)$, of $\widehat{\psi}(\tau)$,

$$\sum_{t=1}^{T} \widehat{\varepsilon}_t(\psi;\tau)\left.\frac{\partial \widehat{\varepsilon}_t(\psi;\tau)}{\partial \psi_i}\right|_{\psi=\psi^*} + \left[D\left(\widetilde{\psi}^i(\tau);\tau\right)\right]_i\left(\widehat{\psi}(\tau)-\psi^*\right) = 0$$

where $[D(\widetilde{\psi}^i(\tau);\tau)]_i$ denotes the i-th row of the matrix $D(\psi;\tau)$ and $\widetilde{\psi}^i(\tau)$ is such that $\|\widetilde{\psi}^i(\tau)-\psi^*\| \leq \|\widehat{\psi}(\tau)-\psi^*\|$. Denoting by $\widetilde{D}(\widehat{\psi}(\tau);\tau)$ the matrix obtained by stacking of the rows $[D(\widetilde{\psi}^i(\tau);\tau)]_i$, and multiplying by $T^{1/2}$, we obtain that

$$T^{-1/2}\sum_{t=1}^T \widehat{\varepsilon}_t(\psi;\tau)\left.\frac{\partial\widehat{\varepsilon}_t(\psi;\tau)}{\partial\psi}\right|_{\psi=\psi^*} + \widetilde{D}\left(\widehat{\psi}(\tau);\tau\right)T^{1/2}\left(\widehat{\psi}(\tau)-\psi^*\right) = 0. \tag{S.14}$$

To prove part (ii) of the lemma, we will show that the distance $\left\|\widehat{\psi}-\widehat{\psi}(\tau)\right\|$ is $o_p\left(T^{-1/2}\right)$ so $\widehat{\psi}$ and $\widehat{\psi}(\tau)$ have the same limit distribution. To that end, we first need to establish that the following result holds:

$$\sup_\tau\left\|\widetilde{D}\left(\widehat{\psi}\right)-\widetilde{D}\left(\widehat{\psi}(\tau);\tau\right)\right\| \xrightarrow{p} 0. \tag{S.15}$$

To do so, we first expand the summands in $D(\psi(\tau);\tau)$ as follows:

$$\begin{aligned}
sa_t(\psi) &:= \frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi}\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi'}\\
sb_t(\psi;\tau) &:= \frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi}\left(\frac{\partial}{\partial\psi'}\left(g(L;\psi)\Delta_+^\delta z_t(\tau)'\left(\beta-\widehat{\beta}(\tau)\right)\right)\right)\\
sc_t(\psi;\tau) &:= \left(\frac{\partial}{\partial\psi}\left(g(L;\psi)\Delta_+^\delta z_t(\tau)'\left(\beta-\widehat{\beta}(\tau)\right)\right)\right)\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi'}\\
sd_t(\psi;\tau) &:= \left(\frac{\partial}{\partial\psi}\left(g(L;\psi)\Delta_+^\delta z_t(\tau)'\left(\beta-\widehat{\beta}(\tau)\right)\right)\right)\left(\frac{\partial}{\partial\psi'}\left(g(L;\psi)\Delta_+^\delta z_t(\tau)'\left(\beta-\widehat{\beta}(\tau)\right)\right)\right)\\
se_t(\psi) &:= \widehat{\varepsilon}_t(\psi)\frac{\partial^2\widehat{\varepsilon}_t(\psi)}{\partial\psi\partial\psi'}\\
sf_t(\psi;\tau) &:= \widehat{\varepsilon}_t(\psi)\frac{\partial^2}{\partial\psi\partial\psi'}\left(g(L;\psi)\Delta_+^\delta z_t(\tau)'\left(\beta-\widehat{\beta}(\tau)\right)\right)\\
sg_t(\psi;\tau) &:= \left(g(L;\psi)\Delta_+^\delta z_t(\tau)'\left(\beta-\widehat{\beta}(\tau)\right)\right)\frac{\partial^2\widehat{\varepsilon}_t(\psi)}{\partial\psi\partial\psi'}\\
sh_t(\psi;\tau) &:= \left(g(L;\psi)\Delta_+^\delta z_t(\tau)'\left(\beta-\widehat{\beta}(\tau)\right)\right)\frac{\partial^2}{\partial\psi\partial\psi'}\left(g(L;\psi)\Delta_+^\delta z_t(\tau)'\left(\beta-\widehat{\beta}(\tau)\right)\right).
\end{aligned}$$

Adding and subtracting $\Phi\sigma_\varepsilon^2$ in (S.15) and using the triangle inequality, the expression in (S.15) is bounded by $\left\|\widetilde{D}\left(\widehat{\psi}\right)-\Phi\sigma_\varepsilon^2\right\|+\sup_\tau\left\|\widetilde{D}\left(\widehat{\psi}(\tau);\tau\right)-\Phi\sigma_\varepsilon^2\right\|$, where recall that $\widetilde{D}\left(\widehat{\psi}\right)\to_p\Phi\sigma_\varepsilon^2$ so that $\left\|\widetilde{D}\left(\widehat{\psi}\right)-\Phi\sigma_\varepsilon^2\right\|=o_p(1)$.

We then have to show that $\frac{1}{T}\sum_{t=1}^T\left(sa_t\left(\widetilde{\psi}(\tau)\right)+se_t\left(\widetilde{\psi}(\tau)\right)\right)-\Phi\sigma_\varepsilon^2=o_p(1)$ and that the averages taken over $t=1,...,T$ of $sb_t\left(\widetilde{\psi}(\tau);\tau\right)$, $sc_t\left(\widetilde{\psi}(\tau);\tau\right)$, $sd_t\left(\widetilde{\psi}(\tau);\tau\right)$, $sf_t\left(\widetilde{\psi}(\tau);\tau\right)$, $sg_t\left(\widetilde{\psi}(\tau);\tau\right)$ and $sh_t\left(\widetilde{\psi}(\tau);\tau\right)$ are all of $o_p(1)$ for $\left\|\widetilde{\psi}(\tau)-\psi^*\right\|\leq\left\|\widehat{\psi}(\tau)-\psi^*\right\|$. To that end, we first show that the following results hold:

$$\frac{1}{T}\sum_{t=1}^T\widehat{\varepsilon}_t\left(\widetilde{\psi}(\tau)\right)^2-\frac{1}{T}\sum_{t=1}^T\widehat{\varepsilon}_t(\psi^*)^2=o_p(1) \tag{S.16}$$

$$\frac{1}{T}\sum_{t=1}^T\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_i}\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_j}\right|_{\psi=\widetilde{\psi}(\tau)}-\frac{1}{T}\sum_{t=1}^T\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_i}\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_j}\right|_{\psi=\psi^*}=o_p(1) \tag{S.17}$$

$$\frac{1}{T}\sum_{t=1}^T\left.\frac{\partial^2\widehat{\varepsilon}_t(\psi)}{\partial\psi_i\partial\psi_j}\right|_{\psi=\widetilde{\psi}(\tau)}-\frac{1}{T}\sum_{t=1}^T\left.\frac{\partial^2\widehat{\varepsilon}_t(\psi)}{\partial\psi_i\partial\psi_j}\right|_{\psi=\psi^*}=o_p(1). \tag{S.18}$$

Because $\eta_t = \frac{b(L;\psi^*)}{a(L;\psi^*)}\varepsilon_t$ is a stationary and invertible ARMA process, then $g(L;\psi)\,\eta_t = \frac{a(L;\psi)}{b(L;\psi)}\frac{b(L;\psi^*)}{a(L;\psi^*)}\varepsilon_t$ is also an ARMA process. For ψ_i, the *i-th* element of ψ, $\frac{\partial}{\partial\psi_i} g(L;\psi)\,\eta_t$ and $\frac{\partial^2}{\partial\psi_i\partial\psi_j} g(L;\psi)\,\eta_t$ are also ARMA processes, and so $\left|\frac{\partial}{\partial\psi_i} g(1;\psi)\right| < C$ and $\left|\frac{\partial^2}{\partial\psi_i\partial\psi_j} g(1;\psi)\right| < C$ uniformly in ψ. Proceeding as in Bai (1993), we illustrate (S.16)-(S.18) for the ARMA(1,1) case, $(1-\psi_1^* L)\,\eta_t = (1+\psi_2^* L)\,\varepsilon_t$.

Consider first (S.16). Because $\widehat{\varepsilon}_t(\psi^*) = \varepsilon_t$, we rewrite

$$\widehat{\varepsilon}_t\left(\widetilde{\psi}(\tau)\right)^2 - \varepsilon_t^2 = \left(\widehat{\varepsilon}_t\left(\widetilde{\psi}(\tau)\right) - \varepsilon_t\right)^2 + 2\varepsilon_t\left(\widehat{\varepsilon}_t\left(\widetilde{\psi}(\tau)\right) - \varepsilon_t\right).$$

From

$$\begin{aligned}
\varepsilon_t &= \eta_t - \psi_1^*\eta_{t-1} - \psi_2^*\varepsilon_{t-1} \\
\widehat{\varepsilon}_t\left(\widetilde{\psi}(\tau)\right) &= \eta_t - \widetilde{\psi}_1(\tau)\,\eta_{t-1} - \widetilde{\psi}_2(\tau)\,\widehat{\varepsilon}_{t-1}\left(\widetilde{\psi}(\tau)\right)
\end{aligned}$$

then

$$\begin{aligned}
\widehat{\varepsilon}_t\left(\widetilde{\psi}(\tau)\right) - \varepsilon_t &= -\left(\widetilde{\psi}_1(\tau) - \psi_1^*\right)\eta_{t-1} - \left(\widetilde{\psi}_2(\tau) - \psi_2^*\right)\varepsilon_{t-1} - \widetilde{\psi}_2(\tau)\left(\widehat{\varepsilon}_t\left(\widetilde{\psi}(\tau)\right) - \varepsilon_{t-1}\right) \\
&= -\left(\widetilde{\psi}_1(\tau) - \psi_1^*\right)\textstyle\sum_{j=0}^{\infty}(-1)^j\left(\widetilde{\psi}_2(\tau)\right)^j\eta_{t-j-1} \\
&\quad -\left(\widetilde{\psi}_2(\tau) - \psi_2^*\right)\textstyle\sum_{j=0}^{\infty}(-1)^j\left(\widetilde{\psi}_2(\tau)\right)^j\varepsilon_{t-j-1}
\end{aligned}$$

using repeated substitution, also see Equation (3) of Bai (1993). To abbreviate notation, denote

$$sk_t\left(\widetilde{\psi}(\tau)\right) := \textstyle\sum_{j=0}^{\infty}(-1)^j\left(\widetilde{\psi}_2(\tau)\right)^j\eta_{t-j-1},\ sl_t\left(\widetilde{\psi}(\tau)\right) := \textstyle\sum_{j=0}^{\infty}(-1)^j\left(\widetilde{\psi}_2(\tau)\right)^j\varepsilon_{t-j-1},$$

then

$$\left|\widehat{\varepsilon}_t\left(\widetilde{\psi}(\tau)\right) - \varepsilon_t\right| \le C\left|\widetilde{\psi}_1(\tau) - \psi_1^*\right|\left|sk_t\left(\widetilde{\psi}(\tau)\right)\right| + C\left|\widetilde{\psi}_2(\tau) - \psi_2^*\right|\left|sl_t\left(\widetilde{\psi}(\tau)\right)\right| \tag{S.19}$$

Notice that $sk_t\left(\widetilde{\psi}(\tau)\right)$ is ARMA(2,1) and $sl_t\left(\widetilde{\psi}(\tau)\right)$ is AR(1). The compactness of Θ means that there exists $0 < \overline{c} < 1 - \varepsilon$, where $\varepsilon > 0$ depends on Θ, such that $\sup|\psi_2| < \overline{c} < 1$, and so

$$\left|sk_t\left(\widetilde{\psi}(\tau)\right)\right| \le \textstyle\sum_{j=0}^{\infty}\overline{c}^j\left|\eta_{t-j-1}\right|,\ \left|sl_t\left(\widetilde{\psi}(\tau)\right)\right| \le \textstyle\sum_{j=0}^{\infty}\overline{c}^j\left|\varepsilon_{t-j-1}\right|,$$

and $\sum_{j=0}^{\infty}\overline{c}^j\left|\eta_{t-j-1}\right| = O_p(1)$ because $\mathrm{E}\left(\left|\eta_{t-j-1}\right|\right) < C$ and $\sum_{j=0}^{\infty}\overline{c}^j < C$, so $\left|sk_t\left(\widetilde{\psi}(\tau)\right)\right| = O_p(1)$. In the same way we also establish $sl_t\left(\widetilde{\psi}(\tau)\right) = O_p(1)$. Therefore, the first term in the bound (S.19) is $o_p(1)$ because $sk_t = O_p(1)$ and $\left|\widetilde{\psi}_1(\tau) - \psi_1^*\right| = o_p(1)$; the second term can be discussed in the same way. Therefore, $\left|\widehat{\varepsilon}_t\left(\widetilde{\psi}(\tau)\right) - \varepsilon_t\right| = o_p(1)$ and $\left(\widehat{\varepsilon}_t\left(\widetilde{\psi}(\tau)\right) - \varepsilon_t\right)^2 = o_p(1)$ and

$$\frac{1}{T}\textstyle\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\widetilde{\psi}(\tau)\right) - \varepsilon_t\right)^2 = o_p(1).$$

Finally, $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t\left(\widehat{\varepsilon}_t\left(\widetilde{\psi}(\tau)\right) - \varepsilon_t\right) = o_p(1)$ by the Cauchy-Schwarz inequality, which concludes the demonstration of (S.16) for the ARMA(1,1) case. The result holds for the more general ARMA(p,q) case using a similar but more tedious treatment.

We turn next to the result in (S.17). Proceeding in the same way as for (S.16), it is sufficient to show that the following results hold:

$$\frac{1}{T}\sum_{t=1}^{T}\left(\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_i}\right|_{\psi=\widetilde{\psi}(\tau)}-\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_i}\right|_{\psi=\psi^*}\right)^2=o_p(1) \tag{S.20}$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\left(\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_i}\right|_{\psi=\psi^*}\right)^2=O_p(1). \tag{S.21}$$

Consider first the result in (S.20). Again we illustrate this in the ARMA(1,1) case, noting that these results hold for the more general ARMA(p,q) case. In the ARMA(1,1) case, considering $\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_2}$ first,

$$\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_2}=-\widehat{\varepsilon}_{t-1}(\psi)-\psi_2\frac{\partial\widehat{\varepsilon}_{t-1}(\psi)}{\partial\psi_2}, \tag{S.22}$$

and notice that, using repeated substitutions,

$$\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_2}\right|_{\psi=\psi^*}=-\sum_{j=0}^{\infty}(-\psi_2^*)^j\,\varepsilon_{t-j-1}$$

is AR(1) and therefore $\left|\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_2}\right|_{\psi=\psi^*}\right|=O_p(1)$, which is sufficient to establish the result in (S.21).

Moreover, using (S.22) again,

$$\begin{aligned}
&\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_2}\right|_{\psi=\widetilde{\psi}(\tau)}-\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_2}\right|_{\psi=\psi^*}\\
=&-\left(\widehat{\varepsilon}_{t-1}\left(\widetilde{\psi}(\tau)\right)-\varepsilon_{t-1}\right)-\left(\widetilde{\psi}_2(\tau)\left.\frac{\partial\widehat{\varepsilon}_{t-1}(\psi)}{\partial\psi_2}\right|_{\psi=\widetilde{\psi}(\tau)}-\psi_2^*\left.\frac{\partial\widehat{\varepsilon}_{t-1}(\psi)}{\partial\psi_2}\right|_{\psi=\psi^*}\right)\\
=&-\left(\widehat{\varepsilon}_{t-1}\left(\widetilde{\psi}(\tau)\right)-\varepsilon_{t-1}\right)-\left(\widetilde{\psi}_2(\tau)-\psi_2^*\right)\left.\frac{\partial\widehat{\varepsilon}_{t-1}(\psi)}{\partial\psi_2}\right|_{\psi=\psi^*}\\
&-\widetilde{\psi}_2(\tau)\left(\left.\frac{\partial\widehat{\varepsilon}_{t-1}(\psi)}{\partial\psi_2}\right|_{\psi=\widetilde{\psi}(\tau)}-\left.\frac{\partial\widehat{\varepsilon}_{t-1}(\psi)}{\partial\psi_2}\right|_{\psi=\psi^*}\right)
\end{aligned}$$

and, using repeated substitutions,

$$\begin{aligned}
\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_2}\right|_{\psi=\widetilde{\psi}(\tau)}-\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_2}\right|_{\psi=\psi^*}=-\quad&\sum_{j=0}^{\infty}\left(-\widetilde{\psi}_2(\tau)\right)^j\left(\widehat{\varepsilon}_{t-j-1}\left(\widetilde{\psi}(\tau)\right)-\varepsilon_{t-j-1}\right)\\
&-\left(\widetilde{\psi}_2(\tau)-\psi_2^*\right)\sum_{j=0}^{\infty}\left(-\widetilde{\psi}_2(\tau)\right)^j\left.\frac{\partial\widehat{\varepsilon}_{t-j-1}(\psi)}{\partial\psi_2}\right|_{\psi=\psi^*}.
\end{aligned}$$

Thus, bounding

$$\begin{aligned}
&\left|\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_2}\right|_{\psi=\widetilde{\psi}(\tau)}-\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_2}\right|_{\psi=\psi^*}\right|\\
\leq&\sum_{j=0}^{\infty}\bar{c}^j\left|\widehat{\varepsilon}_{t-j-1}\left(\widetilde{\psi}(\tau)\right)-\varepsilon_{t-j-1}\right|+\left|\widetilde{\psi}_2(\tau)-\psi_2^*\right|\sum_{j=0}^{\infty}\bar{c}^j\left|\left.\frac{\partial\widehat{\varepsilon}_{t-j-1}(\psi)}{\partial\psi_2}\right|_{\psi=\psi^*}\right|
\end{aligned}$$

this is $o_p(1)$, which is sufficient to establish the result in (S.20).

The result in (S.18) can be obtained in a similar fashion and the proof is omitted in the interest of brevity.

Continuing, we next show that

$$\frac{1}{T}\sum_{t=1}^{T}\widehat{\varepsilon}_t\left(\psi\right)\left.\frac{\partial^2\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi_i\partial\psi_j}\right|_{\psi=\widetilde{\psi}(\tau)}-\frac{1}{T}\sum_{t=1}^{T}\widehat{\varepsilon}_t\left(\psi\right)\left.\frac{\partial^2\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi_i\partial\psi_j}\right|_{\psi=\psi^*}=o_p\left(1\right). \tag{S.23}$$

The left hand side of (S.23) can be written as

$$\frac{1}{T}\sum_{t=1}^{T}\left(\left(\widehat{\varepsilon}_t\left(\psi\right)-\varepsilon_t\right)\left.\frac{\partial^2\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi_i\partial\psi_j}\right|_{\psi=\widetilde{\psi}(\tau)}\right)+\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t\left(\left.\frac{\partial^2\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi_i\partial\psi_j}\right|_{\psi=\widetilde{\psi}(\tau)}-\left.\frac{\partial^2\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi_i\partial\psi_j}\right|_{\psi=\psi^*}\right)$$

in which each term can be seen to be of $o_p\left(1\right)$, using the limits for (S.16), (S.18) and the Cauchy-Schwarz inequality.

We can now move to the contribution of the terms sa_t, $(\widetilde{\psi}\left(\tau\right))$, ..., $sh_t(\widetilde{\psi}(\tau);\tau)$ to (S.15). Using (S.17), then $T^{-1}\sum_{t=1}^{T}\left(sa_t(\psi^*)-sa_t\left(\widetilde{\psi}\left(\tau\right)\right)\right)\rightarrow_p 0$, and using (S.23) then $T^{-1}\sum_{t=1}^{T}\left(se_t(\psi^*)-se_t\left(\widetilde{\psi}\left(\tau\right)\right)\right)\rightarrow_p 0$. Thus, recalling that $T^{-1}\sum_{t=1}^{T}(sa_t(\psi^*)+se_t(\psi^*))\rightarrow_p\Phi\sigma_\varepsilon^2$, it also holds that $T^{-1}\sum_{t=1}^{T}\left(sa_t\left(\widetilde{\psi}\left(\tau\right)\right)+se_t\left(\widetilde{\psi}\left(\tau\right)\right)\right)\rightarrow_p\Phi\sigma_\varepsilon^2$. Next, $T^{-1}\sum_{t=1}^{T}sd_t(\widetilde{\psi}\left(\tau\right);\tau)=o_p\left(1\right)$ and $T^{-1}\sum_{t=1}^{T}sh_t(\widetilde{\psi}(\tau);\tau)=o_p\left(1\right)$ using arguments similar to those in the discussion of the right hand side of (S.8). Finally, the contribution of the terms $sb_t(\widetilde{\psi}\left(\tau\right);\tau)$, $sc_t(\widetilde{\psi}\left(\tau\right);\tau)$, $sf_t(\widetilde{\psi}\left(\tau\right);\tau)$ and $sg_t(\widetilde{\psi}\left(\tau\right);\tau)$ is of $o_p\left(1\right)$, using the Cauchy Schwarz inequality, again as in the discussion of (S.9). This completes the proof of (S.15).

For the next step of the proof, equating the left hand sides of the two expansions in (S.14) and (S.13) and re-arranging, yields

$$\begin{aligned}
T^{1/2}\left(\widehat{\psi}\left(\tau\right)-\widehat{\psi}\right) &= -\widetilde{D}\left(\widehat{\psi}\left(\tau\right);\tau\right)^{-1}T^{-1/2}\sum_{t=1}^{T}\widehat{\varepsilon}_t\left(\psi;\tau\right)\left.\frac{\partial\widehat{\varepsilon}_t\left(\psi;\tau\right)}{\partial\psi}\right|_{\psi=\psi^*}\\
&\quad+\left\{\widetilde{D}\left(\widehat{\psi}\right)^{-1}-\widetilde{D}\left(\widehat{\psi}\left(\tau\right);\tau\right)^{-1}+\widetilde{D}\left(\widehat{\psi}\left(\tau\right);\tau\right)^{-1}\right\}T^{-1/2}\sum_{t=1}^{T}\widehat{\varepsilon}_t\left(\psi\right)\left.\frac{\partial\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi}\right|_{\psi=\psi^*}\\
&= -\widetilde{D}\left(\widehat{\psi}\left(\tau\right);\tau\right)^{-1}T^{-1/2}\sum_{t=1}^{T}\left.\left(\widehat{\varepsilon}_t\left(\psi;\tau\right)\frac{\partial\widehat{\varepsilon}_t\left(\psi;\tau\right)}{\partial\psi}-\widehat{\varepsilon}_t\left(\psi\right)\frac{\partial\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi}\right)\right|_{\psi=\psi^*}\\
&\quad+\left\{\widetilde{D}\left(\widehat{\psi}\right)^{-1}-\widetilde{D}\left(\widehat{\psi}\left(\tau\right);\tau\right)^{-1}\right\}T^{-1/2}\sum_{t=1}^{T}\widehat{\varepsilon}_t\left(\psi\right)\left.\frac{\partial\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi}\right|_{\psi=\psi^*}.
\end{aligned}$$

Noting that $T^{-1/2}\sum_{t=1}^{T}\widehat{\varepsilon}_t\left(\psi\right)\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi}\right|_{\psi=\psi^*}=O_p\left(1\right)$ and that $\widetilde{D}(\widehat{\psi})^{-1}-\widetilde{D}(\widehat{\psi}\left(\tau\right);\tau)^{-1}=o_p\left(1\right)$, the second term in the expression above is seen to be of $o_p\left(1\right)$. As for the first term, since $\widetilde{D}(\widehat{\psi}\left(\tau\right);\tau)^{-1}\overset{p}{\rightarrow}(\Phi\sigma_\varepsilon^2)^{-1}$, we need to show that the function of τ given by

$$T^{-1/2}\sum_{t=1}^{T}\left.\left(\widehat{\varepsilon}_t\left(\psi;\tau\right)\frac{\partial\widehat{\varepsilon}_t\left(\psi;\tau\right)}{\partial\psi}-\widehat{\varepsilon}_t\left(\psi\right)\frac{\partial\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi}\right)\right|_{\psi=\psi^*} \tag{S.24}$$

is of $o_p\left(1\right)$.

Recalling (S.7) $\widehat{\varepsilon}_t(\psi;\tau)=\widehat{\varepsilon}_t(\psi)+g(L;\psi)\,\Delta_+^{\delta}z_t(\tau)'\left(\beta-\widehat{\beta}(\tau)\right)$ then

$$\frac{\partial\widehat{\varepsilon}_t(\psi;\tau)}{\partial\psi}=\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi}+\frac{\partial\left[g(L;\psi)\,\Delta_+^{\delta}z_t(\tau)'\left(\beta-\widehat{\beta}(\tau)\right)\right]}{\partial\psi}$$

and

$$\begin{aligned}\widehat{\varepsilon}_t(\psi;\tau)\frac{\partial\widehat{\varepsilon}_t(\psi;\tau)}{\partial\psi}&=\left(\widehat{\varepsilon}_t(\psi)+g(L;\psi)\,\Delta_+^{\delta}z_t(\tau)'\left(\beta-\widehat{\beta}(\tau)\right)\right)\\ &\times\left(\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi}+\frac{\partial\left[g(L;\psi)\,\Delta_+^{\delta}z_t(\tau)'\left(\beta-\widehat{\beta}(\tau)\right)\right]}{\partial\psi}\right)\end{aligned}$$

and we therefore rewrite elements in (S.24) as

$$\begin{aligned}&\widehat{\varepsilon}_t(\psi;\tau)\frac{\partial\widehat{\varepsilon}_t(\psi;\tau)}{\partial\psi}=\widehat{\varepsilon}_t(\psi)\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi}\\ &+\widehat{\varepsilon}_t(\psi)\frac{\partial\left[g(L;\psi)\,\Delta_+^{\delta}z_t(\tau)'\left(\beta-\widehat{\beta}(\tau)\right)\right]}{\partial\psi}\\ &+g(L;\psi)\,\Delta_+^{\delta}z_t(\tau)'\left(\beta-\widehat{\beta}(\tau)\right)\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi}\\ &+\left(g(L;\psi)\,\Delta_+^{\delta}z_t(\tau)'\left(\beta-\widehat{\beta}(\tau)\right)\right)\left(\frac{\partial\left[g(L;\psi)\,\Delta_+^{\delta}z_t(\tau)'\left(\beta-\widehat{\beta}(\tau)\right)\right]}{\partial\psi}\right).\end{aligned}$$

Thus, (S.24) is

$$T^{-1/2}\textstyle\sum_{t=1}^{T}\left(g(L;\psi)\,\Delta_+^{\delta}z_t(\tau)'\left(\beta-\widehat{\beta}(\tau)\right)\right)\dfrac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi}\Bigg|_{\psi=\psi^*}\tag{S.25}$$

$$+T^{-1/2}\textstyle\sum_{t=1}^{T}\widehat{\varepsilon}_t(\psi)\dfrac{\partial}{\partial\psi}\left(g(L;\psi)\,\Delta_+^{\delta}z_t(\tau)'\left(\beta-\widehat{\beta}(\tau)\right)\right)\Bigg|_{\psi=\psi^*}\tag{S.26}$$

$$+T^{-1/2}\textstyle\sum_{t=1}^{T}\left(g(L\;\psi)\,\Delta_+^{\delta}z_t(\tau)'\left(\beta-\widehat{\beta}(\tau)\right)\right)\dfrac{\partial}{\partial\psi}\left(g(L;\psi)\,\Delta_+^{\delta}z_t(\tau)'\left(\beta-\widehat{\beta}(\tau)\right)\right)\Bigg|_{\psi=\psi^*}\tag{S.27}$$

In view of Lemma 3 of Robinson (2005), the order of (S.27) is the same as the order of

$$T^{-1/2}\textstyle\sum_{t=1}^{T}\left(\Delta_+^{\delta}z_t(\tau)'\left(\beta-\widehat{\beta}(\tau)\right)\right)^2.$$

Proceeding as in the discussion of (S.8), when Model A is used, this term is of $O_p\left(T^{-1/2}\right)=o_p(1)$. Similarly, when Model B is used, it is again of $O_p\left(T^{-1/2}\right)=o_p(1)$. Regarding the term (S.25), using summation by parts the absolute value of this term is bounded by

$$\begin{aligned}\leq\;&T^{-1/2}\textstyle\sum_{t=1}^{T-1}\left|\left(g(L;\psi)\,\Delta_+^{\delta}z_{t+1}(\tau)-g(L;\psi)\,\Delta_+^{\delta}z_t(\tau)\right)\right|'\left|\left(\beta-\widehat{\beta}(\tau)\right)\right|\left|\textstyle\sum_{s=1}^{t}\dfrac{\partial\widehat{\varepsilon}_s(\psi)}{\partial\psi}\right|\\ &+T^{-1/2}\left|\left(g(L;\psi)\,\Delta_+^{\delta}z_T(\tau)\right)\right|'\left|\beta-\widehat{\beta}(\tau)\right|\left|\textstyle\sum_{t=1}^{T}\dfrac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi}\right|\end{aligned}$$

for $\psi=\psi^*$ and, in view of Lemma 3 of Robinson (2005), this bound has the same order as

$$\leq\;T^{-1/2}\textstyle\sum_{t=1}^{T-1}\left|\left(\Delta_+^{\delta}z_{t+1}(\tau)-\Delta_+^{\delta}z_t(\tau)\right)\right|'\left|\left(\beta-\widehat{\beta}(\tau)\right)\right|\left|\textstyle\sum_{s=1}^{t}\dfrac{\partial\widehat{\varepsilon}_s(\psi)}{\partial\psi}\right|\tag{S.28}$$

$$+T^{-1/2}\left|\left(\Delta_+^{\delta}z_T(\tau)\right)\right|'\left|\beta-\widehat{\beta}(\tau)\right|\left|\textstyle\sum_{t=1}^{T}\dfrac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi}\right|\tag{S.29}$$

for $\psi = \psi^*$.

The term in (S.28) can be bounded as

$$T^{-1/2} \sum_{t=1}^{T-1} \left| \left(\Delta^\delta z_{t+1}(\tau) - \Delta^\delta z_t(\tau) \right) \right|' \left| \left(\beta - \widehat{\beta}(\tau) \right) \right| \sup_\rho \left| \sum_{s=1}^{\lfloor \rho T \rfloor} \frac{\partial \widehat{\varepsilon}_s(\psi)}{\partial \psi} \right|$$

where it holds that $\sup_\rho \left| \sum_{s=1}^{\lfloor \rho T \rfloor} \frac{\partial \widehat{\varepsilon}_s(\psi)}{\partial \psi} \Big|_{\psi=\psi^*} \right| = O_p\left(T^{1/2}\right)$, because this is a ARMA process.

When Model A is used,

$$\begin{aligned} & \sum_{t=1}^{T-1} \left| \left(\Delta^\delta z_{t+1}(\tau) - \Delta^\delta z_t(\tau) \right)' \left(\beta - \widehat{\beta}(\tau) \right) \right| \\ \leq \; & \sum_{t=1}^{T-1} \left| \Delta\mu_{1,t+1} \right| \left| \beta_1 - \widehat{\beta}_1(\tau) \right| + \sum_{t=1}^{T-1} \left| \Delta\mu_{2,t+1} \right| \left| \beta_2 - \widehat{\beta}_2(\tau) \right| + \sum_{t=1}^{T-1} \left| \Delta\mu_{3,t+1}(\tau) \right| \left| \widehat{\beta}_3(\tau) \right|. \end{aligned} \tag{S.30}$$

If $\delta > 0$, the terms in (S.30) are such that

$$\begin{aligned} \sum_{t=1}^{T-1} \left| \Delta\mu_{1,t+1} \right| \left| \beta_1 - \widehat{\beta}_1(\tau) \right| &= O_p\left(\sum_{t=1}^{T-1} t^{-1} T^{-1/2+\delta} \right) = O_p\left((\ln(T)) T^{-1/2+\delta} \right) = o_p(1) \\ \sum_{t=1}^{T-1} \left| \Delta\mu_{2,t+1} \right| \left| \beta_2 - \widehat{\beta}_2(\tau) \right| &= O_p\left(\sum_{t=1}^{T-1} t^{-\delta} T^{-3/2+\delta} \right) = O_p\left(T^{-1/2} \right) = o_p(1) \\ \sum_{t=1}^{T-1} \left| \Delta\mu_{3,t+1}(\tau) \right| \left| \widehat{\beta}_3(\tau) \right| &\leq \sum_{t=1}^{T-1} \left| \Delta\mu_{2,t+1} \right| \left| \widehat{\beta}_3(\tau) \right| = O_p\left(T^{-1/2} \right) = o_p(1) \end{aligned}$$

where we have used the rates from (3.14), and in the last bound we have used the result that $\sup_\tau \sum_{t=1}^{T-1} \left| \Delta\mu_{3,t+1}(\tau) \right| \leq \sum_{t=1}^{T-1} \left| \Delta\mu_{2,t+1} \right|$. It then follows that (S.28) is of order $o_p\left(T^{-1/2} \times 1 \times T^{1/2}\right)$ $= o_p(1)$.[8] The remainder term in (S.29) can be shown to be of order

$$T^{-1/2} \times T^{-\delta} \times T^{-1/2+\delta} \times T^{1/2} + T^{-1/2} \times T^{1-\delta} \times T^{-3/2+\delta} \times T^{1/2} = O_p\left(T^{-1/2} \right).$$

If, on the other hand, $\delta < 0$ then the first term in (S.30) is bounded as

$$\sum_{t=1}^{T-1} \left| \Delta\mu_{1,t+1} \right| \left| \beta_1 - \widehat{\beta}_1(\tau) \right| = O_p\left(\sum_{t=1}^{T-1} t^{-1-\delta} T^{-1/2+\delta} \right) = O_p\left(T^{-1/2} \right) = o_p(1).$$

The bounds of the other two terms in (S.30) are unaffected by the sign of δ, and it is easily verified that (S.29) remains of $O_p\left(T^{-1/2}\right)$ so that both (S.28) and (S.29) are of $O_p\left(T^{-1/2}\right)$.

When model B is used we may proceed in the same way, again using bounds (S.28) and (S.29) but instead of (S.30) we have

$$\begin{aligned} & \sum_{t=1}^{T-1} \left| \left(\Delta^\delta z_{t+1}(\tau) - \Delta^\delta z_t(\tau) \right)' \left(\beta - \widehat{\beta}(\tau) \right) \right| \\ \leq \; & \sum_{t=1}^{T-1} \left| \Delta\mu_{1,t+1} \right| \left| \beta_2 - \widehat{\beta}_2(\tau) \right| + \sum_{t=1}^{T-1} \left| \Delta\mu_{3,t+1}(\tau) \right| \left| \widehat{\beta}_3(\tau) \right| \end{aligned}$$

[8] Notice that we bound $\left|\Delta\mu_{1,t+1}\right| = O\left(t^{-1}\right)$ even though the stronger bound $\left|\Delta\mu_{1,t+1}\right| = O\left(t^{-1-\delta}\right)$ holds. We do so because this bound will be needed in a similar proof in Lemma B2. We therefore prefer to use the weaker bound here so as to shorten the subsequent proof of Lemma B2.

where notice that $\sup_\tau \sum_{t=1}^{T-1} \left|\Delta\mu_{3,t+1}(\tau)\right| \leq \sum_{t=1}^{T-1} \left|\Delta\mu_{1,t+1}\right|$. Then, when $\delta > 0$, the functions of τ have stochastic orders

$$\begin{array}{rcl} \sum_{t=1}^{T-1} \left|\Delta\mu_{1,t+1}\right| \left|\beta_2 - \widehat{\beta}_2(\tau)\right| & = & O_p\left(\sum_{t=1}^{T-1} t^{-1} T^{-1/2+\delta}\right) = O_p\left((\ln(T)) T^{-1/2+\delta}\right) = o_p(1) \\ \sum_{t=1}^{T-1} \left|\Delta\mu_{3,t+1}(\tau)\right| \left|\widehat{\beta}_3(\tau)\right| & \leq & \sum_{t=1}^{T-1} \left|\Delta\mu_{1,t+1}\right| \left|\widehat{\beta}_3(\tau)\right| = O_p\left((\ln(T)) T^{-1/2+\delta}\right) = o_p(1) \end{array}$$

whereas, when $\delta < 0$,

$$\begin{array}{rcl} \sum_{t=1}^{T-1} \left|\Delta\mu_{1,t+1}\right| \left|\beta_2 - \widehat{\beta}_2(\tau)\right| & = & O_p\left(\sum_{t=1}^{T-1} t^{-1-\delta} T^{-1/2+\delta}\right) = O_p\left(T^{-\delta} T^{-1/2+\delta}\right) = o_p(1) \\ \sum_{t=1}^{T-1} \left|\Delta\mu_{3,t+1}(\tau)\right| \left|\widehat{\beta}_3(\tau)\right| & \leq & \sum_{t=1}^{T-1} \left|\Delta\mu_{1,t+1}\right| \left|\widehat{\beta}_3(\tau)\right| = O_p\left(T^{-\delta} T^{-1/2+\delta}\right) = o_p(1). \end{array}$$

We have therefore verified that the bound for (S.28) still holds. Proceeding as before, it is also easy to show that the remainder, (S.29), is of order $O_p\left(T^{-1/2}\right)$.

Combining the orders established for (S.28) and (S.29), it then follows that (S.25) is of $o_p(1)$. By similar arguments, the term in (S.26) can also be shown to be of $o_p(1)$, thereby completing the proof of Lemma A2.

Proof of Lemma B2:

Recall that $\widehat{\varepsilon}_t$ and $\widehat{\varepsilon}_t(\tau)$ are shorthand notations for $\widehat{\varepsilon}_t\left(\widehat{\psi}\right)$ and $\widehat{\varepsilon}_t\left(\widehat{\psi}(\tau);\tau\right)$, respectively, and define $\widehat{v}_t\left(\widehat{\psi}\right) := \sum_{j=1}^{t-1} j^{-1} \widehat{\varepsilon}_{t-j}\left(\widehat{\psi}\right)$ and $\widehat{v}_t\left(\widehat{\psi}(\tau);\tau\right) := \sum_{j=1}^{t-1} j^{-1} \widehat{\varepsilon}_{t-j}\left(\widehat{\psi}(\tau);\tau\right)$, so that $\widehat{v}_t$ and $\widehat{v}_t(\tau)$ are correspondingly shorthand notations for $\widehat{v}_t\left(\widehat{\psi}\right)$ and $\widehat{v}_t\left(\widehat{\psi}(\tau);\tau\right)$, respectively.

We consider (A.5) first. To that end, re-write

$$\begin{array}{rcl} \widehat{\varepsilon}_t\left(\widehat{\psi}(\tau);\tau\right)\widehat{v}_t\left(\widehat{\psi}(\tau);\tau\right) - \widehat{\varepsilon}_t\left(\widehat{\psi}\right)\widehat{v}_t\left(\widehat{\psi}\right) & = & \widehat{\varepsilon}_t\left(\widehat{\psi}(\tau);\tau\right)\widehat{v}_t\left(\widehat{\psi}(\tau);\tau\right) - \widehat{\varepsilon}_t\left(\widehat{\psi}(\tau);\tau\right)\widehat{v}_t\left(\widehat{\psi}\right) \\ & & +\widehat{\varepsilon}_t\left(\widehat{\psi}(\tau);\tau\right)\widehat{v}_t\left(\widehat{\psi}\right) - \widehat{\varepsilon}_t\left(\widehat{\psi}\right)\widehat{v}_t\left(\widehat{\psi}\right) \\ & = & \widehat{\varepsilon}_t\left(\widehat{\psi}(\tau);\tau\right)\left(\widehat{v}_t\left(\widehat{\psi}(\tau);\tau\right) - \widehat{v}_t\left(\widehat{\psi}\right)\right) \\ & & + \left(\widehat{\varepsilon}_t\left(\widehat{\psi}(\tau);\tau\right) - \widehat{\varepsilon}_t\left(\widehat{\psi}\right)\right)\widehat{v}_t\left(\widehat{\psi}\right) \end{array}$$

Then it can be seen that (A.5) follows if we can show the following:

$$\sum_{t=1}^{T} \widehat{\varepsilon}_t\left(\widehat{\psi}(\tau);\tau\right)\left(\widehat{v}_t\left(\widehat{\psi}(\tau);\tau\right) - \widehat{v}_t\left(\widehat{\psi}\right)\right) = o_p\left(T^{1/2}\right) \tag{S.31}$$

$$\sum_{t=1}^{T} \left(\widehat{\varepsilon}_t\left(\widehat{\psi}(\tau);\tau\right) - \widehat{\varepsilon}_t\left(\widehat{\psi}\right)\right)\widehat{v}_t\left(\widehat{\psi}\right) = o_p\left(T^{1/2}\right). \tag{S.32}$$

To that end, observe first that

$$\widehat{\varepsilon}_t\left(\widehat{\psi}(\tau);\tau\right) - \widehat{\varepsilon}_t\left(\widehat{\psi}\right) = \widehat{\varepsilon}_t\left(\widehat{\psi}(\tau)\right) - \widehat{\varepsilon}_t\left(\widehat{\psi}\right) + g\left(L;\widehat{\psi}(\tau)\right)\Delta_+^{\delta} z_t(\tau)'\left(\beta - \widehat{\beta}(\tau)\right)$$

where

$$g\left(L;\widehat{\psi}(\tau)\right)\Delta_+^{\delta} z_t(\tau)'\left(\beta - \widehat{\beta}(\tau)\right) = o_p(1)$$

and

$$\widehat{\varepsilon}_t\left(\widehat{\psi}(\tau)\right)-\widehat{\varepsilon}_t\left(\widehat{\psi}\right) = \left(\widehat{\psi}(\tau)-\widehat{\psi}\right)'\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi}\right|_{\psi=\psi^*} + \frac{1}{2}\left(\widehat{\psi}(\tau)-\widehat{\psi}\right)'\left.\frac{\partial^2\widehat{\varepsilon}_t(\psi)}{\partial\psi\partial\psi'}\right|_{\psi=\widetilde{\psi}}\left(\widehat{\psi}(\tau)-\widehat{\psi}\right) \tag{S.33}$$

where $\left\|\widetilde{\psi}-\widehat{\psi}\right\|\leq\left\|\widehat{\psi}(\tau)-\widehat{\psi}\right\|$ and $\sup_\psi \frac{\partial^2\widehat{\varepsilon}_t(\psi)}{\partial\psi\partial\psi'}=O_p(1)$, as $\frac{\partial^2\widehat{\varepsilon}_t(\psi)}{\partial\psi\partial\psi'}$ is still ARMA (strictly speaking, the term in (S.33) is only correct if ψ is a scalar; otherwise, a row by row expansion should be derived, similarly to (S.13), and then stacked as in (S.14), but this approximation does not affect the results). Consequently, the last term of (S.33) is $o_p\left(T^{-1}\right)$, and notice that this holds uniformly in τ. It then follows that $\widehat{\varepsilon}_t\left(\widehat{\psi}(\tau)\right)-\widehat{\varepsilon}_t\left(\widehat{\psi}\right)=o_p\left(T^{-1/2}\right)$ and $\widehat{\varepsilon}_t\left(\widehat{\psi}(\tau)\right)=O_p(1)$, and finally that $\widehat{\varepsilon}_t\left(\widehat{\psi}(\tau);\tau\right)=O_p(1)$.

In the same way, observe that

$$\widehat{v}_t\left(\widehat{\psi}(\tau);\tau\right)-\widehat{v}_t\left(\widehat{\psi}\right)=\widehat{v}_t\left(\widehat{\psi}(\tau)\right)-\widehat{v}_t\left(\widehat{\psi}\right)+g\left(L;\widehat{\psi}(\tau)\right)\left\{-\ln(\Delta)\Delta^\delta\right\}_+ z_t(\tau)'\left(\beta-\widehat{\beta}(\tau)\right)$$

where

$$\widehat{v}_t\left(\widehat{\psi}(\tau)\right)-\widehat{v}_t\left(\widehat{\psi}\right)=\left(\widehat{\psi}(\tau)-\widehat{\psi}\right)'\sum_{j=1}^{t-1}j^{-1}\left.\frac{\partial\widehat{\varepsilon}_{t-j}(\psi)}{\partial\psi}\right|_{\psi=\psi^*}+o_p\left((\ln(t))T^{-1}\right).$$

It then follows that $\widehat{v}_t\left(\widehat{\psi}(\tau)\right)-\widehat{v}_t\left(\widehat{\psi}\right)=o_p\left(T^{-1/2}\right)$ and $\widehat{v}_t\left(\widehat{\psi}(\tau)\right)=O_p(1)$ and $\widehat{v}_t\left(\widehat{\psi}\right)=O_p(1)$.

Next, let

$$\lambda_{1,t}:=\sum_{j=1}^{t-1}j^{-1}\mu_{1,t-j},\ \lambda_{2,t}:=\sum_{j=1}^{t-1}j^{-1}\mu_{2,t-j},\ \lambda_{3,t}(\tau):=\sum_{j=1}^{t-1}j^{-1}\mu_{3,t-j}(\tau),$$

and notice that, by Lemma 2 of Robinson (2005),

$$\lambda_{1,t}=O\left(\ln(t)\,t^{-\delta}\right),\ \lambda_{2,t}=O\left(\ln(t)\,t^{1-\delta}\right),\ \Delta\lambda_{2,t+1}=O\left(\ln(t+1)(t+1)^{-\delta}\right)$$

and, when $\delta\in(0,1/2)$,

$$\Delta\lambda_{1,t+1}=O\left(\ln(t+1)(t+1)^{-1}\right), \tag{S.34}$$

whereas, when $\delta\in(-1/2,0)$,

$$\Delta\lambda_{1,t+1}=O\left(\ln(t+1)(t+1)^{-1-\delta}\right). \tag{S.35}$$

We now move to the discussion of (S.31) and (S.32). The left hand side of (S.31) is

$$\sum_{t=1}^T\widehat{\varepsilon}_t\left(\widehat{\psi}(\tau);\tau\right)\left(\widehat{v}_t\left(\widehat{\psi}(\tau)\right)-\widehat{v}_t\left(\widehat{\psi}\right)\right) \tag{S.36}$$

$$+\sum_{t=1}^T\widehat{\varepsilon}_t\left(\widehat{\psi}(\tau);\tau\right)g\left(L;\widehat{\psi}(\tau)\right)\left\{-\ln(\Delta)\Delta^\delta\right\}_+ z_t(\tau)'\left(\beta-\widehat{\beta}(\tau)\right). \tag{S.37}$$

The stochastic order of (S.36) is bounded by the stochastic order of

$$\sum_{t=1}^T\left|\widehat{\varepsilon}_t\left(\widehat{\psi}(\tau);\tau\right)\right|\left|\widehat{v}_t\left(\widehat{\psi}(\tau)\right)-\widehat{v}_t\left(\widehat{\psi}\right)\right|=o_p\left(T\times T^{-1/2}\right)=o_p\left(T^{1/2}\right).$$

For (S.37),

$$
\begin{aligned}
& \left|\sum_{t=1}^{T} \widehat{\varepsilon}_t\left(\widehat{\psi}(\tau);\tau\right)\left(\widehat{v}_t\left(\widehat{\psi}(\tau);\tau\right)-\widehat{v}_t\left(\widehat{\psi}(\tau)\right)\right)\right| \\
\leq\ & \sum_{t=1}^{T-1}\left|\left(\widehat{v}_{t+1}\left(\widehat{\psi}(\tau);\tau\right)-\widehat{v}_{t+1}\left(\widehat{\psi}(\tau)\right)\right)-\left(\widehat{v}_t\left(\widehat{\psi}(\tau);\tau\right)-\widehat{v}_t\left(\widehat{\psi}(\tau)\right)\right)\right| \\
& \times\sup_{\rho}\left|\sum_{s=1}^{\lfloor\rho T\rfloor}\widehat{\varepsilon}_s\left(\widehat{\psi}(\tau);\tau\right)\right| \\
& +\left|\left(\widehat{v}_T\left(\widehat{\psi}(\tau);\tau\right)-\widehat{v}_T\left(\widehat{\psi}(\tau)\right)\right)\right|\left|\sum_{s=1}^{T}\widehat{\varepsilon}_s\left(\widehat{\psi}(\tau);\tau\right)\right|.
\end{aligned}
$$

Noting that

$$
\begin{aligned}
\sup_{\rho}\left|\sum_{s=1}^{\lfloor\rho T\rfloor}\widehat{\varepsilon}_s\left(\widehat{\psi}(\tau);\tau\right)\right| \ \leq\ & \sup_{\rho}\left|\sum_{s=1}^{\lfloor\rho T\rfloor}\widehat{\varepsilon}_t\left(\widehat{\psi}(\tau)\right)\right| \\
& +\sup_{\rho}\left|\sum_{s=1}^{\lfloor\rho T\rfloor} g\left(L;\widehat{\psi}(\tau)\right)\Delta_+^{\delta} z_s(\tau)'\left(\beta-\widehat{\beta}(\tau)\right)\right| \qquad \text{(S.38)}
\end{aligned}
$$

the term $\sup_{\rho}\left|\sum_{s=1}^{\lfloor\rho T\rfloor}\widehat{\varepsilon}_t\left(\widehat{\psi}(\tau)\right)\right|$ is seen to be of $O_p\left(T^{1/2}\right)$ in view of (S.33) and

$$
\widehat{\varepsilon}_t\left(\widehat{\psi}\right)=\varepsilon_t+\left(\widehat{\psi}-\psi\right)'\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi}\right|_{\psi=\psi^*}+\frac{1}{2}\left(\widehat{\psi}-\psi\right)'\left.\frac{\partial^2\widehat{\varepsilon}_t(\psi)}{\partial\psi\partial\psi'}\right|_{\psi=\widetilde{\psi}}\left(\widehat{\psi}-\psi\right)
$$

for $\left\|\widetilde{\psi}-\psi\right\| \leq \left\|\left(\widehat{\psi}-\psi\right)\right\|$; also see Theorem 1 of Bai (1993). Using again Lemma 3 of Robinson (2005) as was done in the proof of Lemma A2, the term (S.38) is seen to have stochastic order as

$$
\begin{aligned}
& \sup_{\rho}\left|\sum_{s=1}^{\lfloor\rho T\rfloor}\Delta_+^{\delta} z_s(\tau)'\left(\beta-\widehat{\beta}(\tau)\right)\right| \\
\leq\ & C\sum_{t=1}^{T}\mu_{1,t}\left|\beta_1-\widehat{\beta}_1(\tau)\right|+\sum_{t=1}^{T}\mu_{2,t}\left|\beta_2-\widehat{\beta}_2(\tau)\right|+\sum_{t=1}^{T}\mu_{2,t}\left|\widehat{\beta}_3(\tau)\right|=O_p\left(T^{1/2}\right).
\end{aligned}
$$

We therefore conclude that $\sup_{\rho}\left|\sum_{s=1}^{\lfloor\rho T\rfloor}\widehat{\varepsilon}_s\left(\widehat{\psi}(\tau);\tau\right)\right| = O_p\left(T^{1/2}\right)$. To complete the discussion of (S.37) we now consider the term

$$
\sum_{t=1}^{T-1}\left|\left(\widehat{v}_{t+1}\left(\widehat{\psi}(\tau);\tau\right)-\widehat{v}_{t+1}\left(\widehat{\psi}(\tau)\right)\right)-\left(\widehat{v}_t\left(\widehat{\psi}(\tau);\tau\right)-\widehat{v}_t\left(\widehat{\psi}(\tau)\right)\right)\right|
$$

and notice that this has the same stochastic order as

$$
\sum_{t=1}^{T-1}\left|\left(\left\{(\ln(\Delta))\Delta^{\delta}\right\}_+ z_{t+1}(\tau)-\left\{(\ln(\Delta))\Delta^{\delta}\right\}_+ z_t(\tau)\right)'\left(\beta-\widehat{\beta}(\tau)\right)\right|.
$$

When Model A is used, the latter is bounded by

$$
\sum_{t=1}^{T-1}\left|\Delta\lambda_{1,t+1}\right|\left|\beta_1-\widehat{\beta}_1(\tau)\right|+\sum_{t=1}^{T-1}\left|\Delta\lambda_{2,t+1}\right|\left|\beta_2-\widehat{\beta}_2(\tau)\right|+\sum_{t=1}^{T-1}\left|\Delta\lambda_{3,t+1}(\tau)\right|\left|\widehat{\beta}_3(\tau)\right|.
$$

Using (S.34) and (S.35) and proceeding as in the discussion of (S.30), this is seen to be of $O_p\left((\ln(T))^2 T^{-1/2+\delta}\right)$ when $\delta>0$ and of $O_p\left((\ln(T))T^{-1/2}\right)$ when $\delta<0$. When Model B is used, the same bounds may be established in the same way. Finally, in all cases,

$$
\left|\left(\widehat{v}_T\left(\widehat{\psi}(\tau);\tau\right)-\widehat{v}_T\left(\widehat{\psi}(\tau)\right)\right)\right|\left|\sum_{s=1}^{T}\widehat{\varepsilon}_s\left(\widehat{\psi}(\tau);\tau\right)\right|=O_p(\ln(T)).
$$

Combining these results, (S.37) has stochastic order $o_p\left(T^{1/2}\right)$. Together with the stochastic order obtained for (S.36), the stated result in (S.31) is therefore established.

The proof for (S.32) is similar, and we discuss it below. The expression in (S.32) can be written as

$$\sum_{t=1}^{T} \widehat{v}_t\left(\widehat{\psi}\right)\left(\widehat{\varepsilon}_t\left(\widehat{\psi}\left(\tau\right)\right)-\widehat{\varepsilon}_t\left(\widehat{\psi}\right)\right) \tag{S.39}$$

$$+\sum_{t=1}^{T} \widehat{v}_t\left(\widehat{\psi}\right) g\left(L;\widehat{\psi}\left(\tau\right)\right)\Delta_+^{\delta} z_t\left(\tau\right)'\left(\beta-\widehat{\beta}\left(\tau\right)\right). \tag{S.40}$$

As in the discussion of (S.36), the stochastic order of (S.39) is bounded by the stochastic order of

$$\sum_{t=1}^{T}\left|\widehat{v}_t\left(\widehat{\psi}\right)\right|\left|\widehat{\varepsilon}_t\left(\widehat{\psi}\left(\tau\right)\right)-\widehat{\varepsilon}_t\left(\widehat{\psi}\right)\right| = o_p\left(T\times T^{-1/2}\right)=o_p\left(T^{1/2}\right).$$

Again the discussion of (S.40) is similar to the discussion of (S.37): we apply summation by parts to (S.40) and discuss the role of the terms $g\left(L;\widehat{\psi}\left(\tau\right)\right)\Delta_+^{\delta} z_t\left(\tau\right)'\left(\beta-\widehat{\beta}\left(\tau\right)\right)$ as in the discussion of (S.37), but in this case notice that we must discuss the partial sums $\sum_{t=1}^{\lfloor\rho T\rfloor}\widehat{v}_t\left(\widehat{\psi}\right)$. Letting $v_t := \sum_{j=1}^{t-1} j^{-1}\varepsilon_{t-j}$, for $\left\|\widetilde{\psi}-\psi\right\|\le\left\|\left(\widehat{\psi}-\psi\right)\right\|$

$$\begin{aligned}\widehat{v}_t\left(\widehat{\psi}\right) &= v_t+\left(\widehat{\psi}-\psi\right)'\sum_{j=1}^{t-1} j^{-1}\left.\frac{\partial\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi}\right|_{\psi=\psi^*}+\frac{1}{2}\left(\widehat{\psi}-\psi\right)'\sum_{j=1}^{t-1} j^{-1}\left.\frac{\partial^2\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi\partial\psi'}\right|_{\psi=\widetilde{\psi}}\left(\widehat{\psi}-\psi\right)\\ &= v_t+\left(\widehat{\psi}-\psi\right)'\sum_{j=1}^{t-1} j^{-1}\left.\frac{\partial\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi}\right|_{\psi=\psi^*}+o_p\left(\ln\left(t\right)T^{-1}\right)\end{aligned}$$

so $\sup_\rho\left|\sum_{t=1}^{\lfloor\rho T\rfloor}\widehat{v}_t\left(\widehat{\psi}\right)\right| = O_p\left(\ln\left(T\right)T^{1/2}\right)$ again in view of the FCLT in Marinucci and Robinson (2000) and (S.40) is $o_p\left(T^{1/2}\right)$. The result in (A.5) is thereby established.

For (A.6),

$$\begin{aligned}&\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\widehat{\psi}\left(\tau\right);\tau\right)\right)^2-\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\widehat{\psi}\right)\right)^2\\ =\;&\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\widehat{\psi}\left(\tau\right);\tau\right)-\widehat{\varepsilon}_t\left(\widehat{\psi}\right)\right)\widehat{\varepsilon}_t\left(\widehat{\psi}\left(\tau\right);\tau\right)+\sum_{t=1}^{T}\widehat{\varepsilon}_t\left(\widehat{\psi}\right)\left(\widehat{\varepsilon}_t\left(\widehat{\psi}\left(\tau\right);\tau\right)-\widehat{\varepsilon}_t\left(\widehat{\psi}\right)\right)\end{aligned}$$

the two terms of which are $o_p\left(T^{1/2}\right)$ proceeding in the same way as in the discussion of (S.31) and (S.32).

Finally, since κ and Φ are continuous function of ψ, (A.7) follows by an application of Slutzky's Theorem.

Proof of Lemma C2.

We have that,

$$\begin{aligned}\widehat{\varepsilon}_t\left(\psi;\widehat{\tau}\right) &= g\left(L;\psi\right)\Delta_+^{\delta}\left(y_t-z_t\left(\widehat{\tau}\right)'\widehat{\beta}\left(\widehat{\tau}\right)\right)=g\left(L;\psi\right)\Delta_+^{\delta}\left(u_t+z_t\left(\tau^*\right)'\beta-z_t\left(\widehat{\tau}\right)'\widehat{\beta}\left(\widehat{\tau}\right)\right)\\ &= g\left(L;\psi\right)\Delta_+^{\delta}\left(u_t+z_t\left(\tau^*\right)'\beta-z_t\left(\tau^*\right)'\widehat{\beta}\left(\widehat{\tau}\right)+z_t\left(\tau^*\right)'\widehat{\beta}\left(\widehat{\tau}\right)-z_t\left(\widehat{\tau}\right)'\widehat{\beta}\left(\widehat{\tau}\right)\right)\\ &= \widehat{\varepsilon}_t\left(\psi\right)+g\left(L;\psi\right)\Delta_+^{\delta} z_t\left(\tau^*\right)'\left(\beta-\widehat{\beta}\left(\widehat{\tau}\right)\right)+g\left(L;\psi\right)\Delta_+^{\delta}\left(z_t\left(\tau^*\right)-z_t\left(\widehat{\tau}\right)\right)'\widehat{\beta}\left(\widehat{\tau}\right).\end{aligned} \tag{S.41}$$

We first show that $\left\|\widehat{\psi}\left(\widehat{\tau}\right)-\widehat{\psi}\right\| = o_p\left(1\right)$. For this purpose, we need to show that

$$\frac{1}{T}\left|\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\psi;\widehat{\tau}\right)\right)^2-\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\psi\right)\right)^2\right|\rightarrow_p 0 \tag{S.42}$$

uniformly in ψ, and notice that, in view of the stochastic equicontinuity discussed in Lemma A.2, it is sufficient to establish (S.42). We then rewrite

$$\begin{aligned}
&\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\psi;\widehat{\tau}\right)\right)^2-\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\psi\right)\right)^2\\
=\;&\sum_{t=1}^{T}\left(g\left(L;\psi\right)\Delta_{+}^{\delta}z_t\left(\tau^*\right)'\left(\beta-\widehat{\beta}\left(\widehat{\tau}\right)\right)\right)^2+2\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\psi\right)\right)\left(g\left(L;\psi\right)\Delta_{+}^{\delta}z_t\left(\tau^*\right)'\left(\beta-\widehat{\beta}\left(\widehat{\tau}\right)\right)\right) &\text{(S.43)}\\
&+2\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\psi\right)\right)g\left(L;\psi\right)\Delta_{+}^{\delta}\left(z_t\left(\tau^*\right)-z_t\left(\widehat{\tau}\right)\right)'\widehat{\beta}\left(\widehat{\tau}\right) &\text{(S.44)}\\
&+2\sum_{t=1}^{T}\left(g\left(L;\psi\right)\Delta_{+}^{\delta}z_t\left(\tau^*\right)'\left(\beta-\widehat{\beta}\left(\widehat{\tau}\right)\right)\right)\left(g\left(L;\psi\right)\Delta_{+}^{\delta}\left(z_t\left(\tau^*\right)-z_t\left(\widehat{\tau}\right)\right)'\widehat{\beta}\left(\widehat{\tau}\right)\right) &\text{(S.45)}\\
&+\sum_{t=1}^{T}\left(g\left(L;\psi\right)\Delta_{+}^{\delta}\left(z_t\left(\tau^*\right)-z_t\left(\widehat{\tau}\right)\right)'\widehat{\beta}\left(\widehat{\tau}\right)\right)^2 &\text{(S.46)}
\end{aligned}$$

where the two terms in (S.43) are $O_p\left(T^{1/2}\right)$ uniformly in ψ using (3.11) and proceeding as for (S.8) and (S.9) in Lemma A2.

As for (S.46), we can again apply Lemma 3 of Robinson (2005) to account for the polynomial $g\left(L;\psi\right)$. Assuming $\tau^*<\widehat{\tau}$ (the case $\tau^*>\widehat{\tau}$ works in the same way), notice that

$$\sum_{t=1}^{T}\left(\Delta_{+}^{\delta}\left(z_t\left(\tau^*\right)-z_t\left(\widehat{\tau}\right)\right)'\widehat{\beta}\left(\widehat{\tau}\right)\right)^2=\textstyle\sum_{t=1}^{T}\left(\mu_{3,t}\left(\tau^*\right)-\mu_{3,t}\left(\widehat{\tau}\right)\right)^2\widehat{\beta}_3\left(\widehat{\tau}\right)$$

and $\widehat{\beta}_3\left(\widehat{\tau}\right)\xrightarrow{p}\beta_3$ so $\widehat{\beta}_3\left(\widehat{\tau}\right)=O_p\left(1\right)$. Term (S.46) has therefore the same stochastic order as that of

$$\textstyle\sum_{t=1}^{T}\left(\mu_{3,t}\left(\tau^*\right)-\mu_{3,t}\left(\widehat{\tau}\right)\right)^2=\sum_{t=1+\lfloor\tau^*T\rfloor}^{\lfloor\widehat{\tau}T\rfloor}\left(\mu_{3,t}\left(\tau^*\right)\right)^2+\sum_{t=1+\lfloor\widehat{\tau}T\rfloor}^{T}\left(\mu_{3,t}\left(\tau^*\right)-\mu_{3,t}\left(\widehat{\tau}\right)\right)^2.$$

When Model A is used the first term on the right hand side of the foregoing equation is such that,

$$\textstyle\sum_{t=1+\lfloor\tau^*T\rfloor}^{\lfloor\widehat{\tau}T\rfloor}\left(\mu_{3,t}\left(\tau^*\right)\right)^2=\sum_{t=1}^{\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor}\mu_{2,t}^2\leq C\left(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor\right)^{3-2\delta}=O_p\left(T^{(\delta-1/2)\times(3-2\delta)}\right)=o_p\left(1\right)$$

while in the context of the second term,

$$\left(\mu_{3,t}\left(\tau^*\right)-\mu_{3,t}\left(\widehat{\tau}\right)\right)=\left(\mu_{3,t}\left(\tau^*\right)-\mu_{3,t-1}\left(\tau^*\right)+\mu_{3,t-1}\left(\tau^*\right)-\ldots-\mu_{3,t}\left(\widehat{\tau}\right)\right)$$

and, if $\delta>0$,

$$\left|\mu_{3,t}\left(\tau^*\right)-\mu_{3,t}\left(\widehat{\tau}\right)\right|<C\left(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor\right)\left(t-\lfloor\widehat{\tau}T\rfloor\right)^{-\delta} \tag{S.47}$$

and

$$\begin{aligned}
&\textstyle\sum_{t=1+\lfloor\widehat{\tau}T\rfloor}^{T}\left(\mu_{3,t}\left(\tau^*\right)-\mu_{3,t}\left(\widehat{\tau}\right)\right)^2\leq C\left(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor\right)^2\sum_{t=1+\lfloor\widehat{\tau}T\rfloor}^{T}\left(t-\lfloor\widehat{\tau}T\rfloor\right)^{-2\delta}\\
\leq\;&C\left(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor\right)^2\textstyle\sum_{t=1}^{T}t^{-2\delta}
\end{aligned}$$

whereas, if $\delta < 0$,

$$\left|\mu_{3,t}\left(\tau^*\right) - \mu_{3,t}\left(\widehat{\tau}\right)\right| < C\left(\lfloor\widehat{\tau}T\rfloor - \lfloor\tau^*T\rfloor\right)\left(t - \lfloor\tau^*T\rfloor\right)^{-\delta} \tag{S.48}$$

and

$$\begin{aligned} & \textstyle\sum_{t=1+\lfloor\widehat{\tau}T\rfloor}^{T}\left(\mu_{3,t}\left(\tau^*\right) - \mu_{3,t}\left(\widehat{\tau}\right)\right)^2 \leq C\left(\lfloor\widehat{\tau}T\rfloor - \lfloor\tau^*T\rfloor\right)^2 \sum_{t=1+\lfloor\widehat{\tau}T\rfloor}^{T}\left(t - \lfloor\tau^*T\rfloor\right)^{-2\delta} \\ \leq \; & C\left(\lfloor\widehat{\tau}T\rfloor - \lfloor\tau^*T\rfloor\right)^2 \textstyle\sum_{t=1+\lfloor\tau^*T\rfloor}^{T}\left(t - \lfloor\tau^*T\rfloor\right)^{-2\delta} \leq C\left(\lfloor\widehat{\tau}T\rfloor - \lfloor\tau^*T\rfloor\right)^2 \sum_{t=1}^{T} t^{-2\delta}. \end{aligned}$$

Either way, then,

$$\textstyle\sum_{t=1+\lfloor\widehat{\tau}T\rfloor}^{T}\left(\mu_{3,t}\left(\tau^*\right) - \mu_{3,t}\left(\widehat{\tau}\right)\right)^2 \leq C\left(\lfloor\widehat{\tau}T\rfloor - \lfloor\tau^*T\rfloor\right)^2 T^{1-2\delta} = O_p\left(T^{(1-3/2+\delta)\times 2}T^{1-2\delta}\right) = O_p\left(1\right).$$

When Model B is used,

$$\textstyle\sum_{t=1+\lfloor\tau^*T\rfloor}^{\lfloor\widehat{\tau}T\rfloor}\left(\mu_{3,t}\left(\tau^*\right)\right)^2 = \sum_{t=1}^{\lfloor\widehat{\tau}T\rfloor - \lfloor\tau^*T\rfloor}\mu_{1,t}^2 \leq C\left(\lfloor\widehat{\tau}T\rfloor - \lfloor\tau^*T\rfloor\right)^{1-2\delta} = O_p\left(1\right).$$

If $\delta < 0$, using $\left|\mu_{1,t+1} - \mu_{1,t}\right| < Ct^{-\delta-1}$,

$$\begin{aligned} & \textstyle\sum_{t=1+\lfloor\widehat{\tau}T\rfloor}^{T}\left(\mu_{3,t}\left(\tau^*\right) - \mu_{3,t}\left(\widehat{\tau}\right)\right)^2 \leq C\left(\lfloor\widehat{\tau}T\rfloor - \lfloor\tau^*T\rfloor\right)^2 \sum_{t=1+\lfloor\widehat{\tau}T\rfloor}^{T}\left(t - \lfloor\widehat{\tau}T\rfloor\right)^{-2\delta-2} \\ \leq \; & C\left(\lfloor\widehat{\tau}T\rfloor - \lfloor\tau^*T\rfloor\right)^2 \textstyle\sum_{t=1}^{T} t^{-2\delta-2} \leq C\left(\lfloor\widehat{\tau}T\rfloor - \lfloor\tau^*T\rfloor\right)^2 = O_p\left(1\right) \end{aligned}$$

recalling $-2\delta - 2 < -1$ as $\delta > -1/2$. When $\delta > 0$, using $\left|\mu_{1,t+1} - \mu_{1,t}\right| < Ct^{-1}$, the stochastic order of $\sum_{t=1+\lfloor\widehat{\tau}T\rfloor}^{T}\left(\mu_{3,t}\left(\tau^*\right) - \mu_{3,t}\left(\widehat{\tau}\right)\right)^2$ is

$$\begin{aligned} & \textstyle\sum_{t=1+\lfloor\widehat{\tau}T\rfloor}^{T}\left(\mu_{3,t}\left(\tau^*\right) - \mu_{3,t}\left(\widehat{\tau}\right)\right)^2 \leq C\left(\lfloor\widehat{\tau}T\rfloor - \lfloor\tau^*T\rfloor\right)^2 \sum_{t=1+\lfloor\widehat{\tau}T\rfloor}^{T}\left(t - \lfloor\widehat{\tau}T\rfloor\right)^{-2} \\ \leq \; & C\left(\lfloor\widehat{\tau}T\rfloor - \lfloor\tau^*T\rfloor\right)^2 \textstyle\sum_{t=1}^{T} t^{-2} \leq C\left(\lfloor\widehat{\tau}T\rfloor - \lfloor\tau^*T\rfloor\right)^2 = O_p\left(1\right). \end{aligned}$$

It therefore follows that

$$\sum_{t=1}^{T}\left(g\left(L;\psi\right)\Delta_+^{\delta}\left(z_t\left(\tau^*\right) - z_t\left(\widehat{\tau}\right)\right)'\widehat{\beta}\left(\widehat{\tau}\right)\right)^2 = O_p\left(1\right) \tag{S.49}$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\left(g\left(L;\psi\right)\Delta_+^{\delta}\left(z_t\left(\tau^*\right) - z_t\left(\widehat{\tau}\right)\right)'\widehat{\beta}\left(\widehat{\tau}\right)\right)^2 \xrightarrow{p} 0 \tag{S.50}$$

uniformly in ψ, thereby accounting for (S.46). The two remaining cross products in the expansion of $\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\psi;\widehat{\tau}\right)\right)^2 - \sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\psi\right)\right)^2$, (S.44) and (S.45), can be dealt with by applications of the Cauchy-Schwarz inequality. Consequently (S.42) holds, and we conclude that $\widehat{\psi}\left(\widehat{\tau}\right) - \widehat{\psi} \xrightarrow{p} 0$.

To complete the proof of Lemma C2, we need to show that $\left(\widehat{\psi} - \widehat{\psi}\left(\widehat{\tau}\right)\right) = o_p\left(T^{1/2}\right)$. Again we proceed as in the proof of Lemma A2 and account for the extra term $g\left(L;\psi\right)\Delta^{\delta}\left(z_t\left(\tau^*\right) - z_t\left(\widehat{\tau}\right)\right)'\widehat{\beta}\left(\widehat{\tau}\right)$. The result in (S.50) and additional applications of the Cauchy-Schwarz inequality are sufficient to extend the arguments used in establishing Lemma A2 to conclude that $\widetilde{D}\left(\widehat{\psi}\right)^{-1} - \widetilde{D}\left(\widehat{\psi}\left(\widehat{\tau}\right);\widehat{\tau}\right)^{-1} \xrightarrow{p} 0$

still holds. To complete the second part of Lemma C2 we need to check the stochastic order of (S.24) when $\tau = \widehat{\tau}$ and $\beta_3 \neq 0$. Here we need to demonstrate that

$$T^{-1/2} \sum_{t=1}^{T} \left(\Delta_{+}^{\delta} \left(z_t \left(\tau^* \right) - z_t \left(\widehat{\tau} \right) \right)' \widehat{\beta} \left(\widehat{\tau} \right) \right)^2 \xrightarrow{p} 0 \tag{S.51}$$

$$T^{-1/2} \sum_{t=1}^{T} \left(\Delta_{+}^{\delta} \left(z_t \left(\tau^* \right) - z_t \left(\widehat{\tau} \right) \right)' \widehat{\beta} \left(\widehat{\tau} \right) \right) \Delta_{+}^{\delta} z_t \left(\tau^* \right)' \left(\beta - \widehat{\beta} \left(\widehat{\tau} \right) \right) \xrightarrow{p} 0 \tag{S.52}$$

and

$$T^{-1/2} \sum_{t=1}^{T} \left(\Delta_{+}^{\delta} \left(z_t \left(\tau^* \right) - z_t \left(\widehat{\tau} \right) \right)' \widehat{\beta} \left(\widehat{\tau} \right) \right) \left. \frac{\partial \varepsilon_t \left(\psi \right)}{\partial \psi} \right|_{\psi = \psi^*} \xrightarrow{p} 0 \tag{S.53}$$

$$T^{-1/2} \sum_{t=1}^{T} \left(\Delta_{+}^{\delta} \left(z_t \left(\tau^* \right) - z_t \left(\widehat{\tau} \right) \right)' \widehat{\beta} \left(\widehat{\tau} \right) \right) \varepsilon_t \left(\psi^* \right) \xrightarrow{p} 0. \tag{S.54}$$

The first two limits are readily established, using (S.49) for (S.51) and, in the case (S.52), the bound for the right hand side of (S.8) and an application of the Cauchy-Schwarz inequality.

Assuming that $\widehat{\tau} > \tau^*$, the expression in (S.53) has the same order as that of

$$T^{-1/2} \sum_{t=1+\lfloor \tau^* T \rfloor}^{\lfloor \widehat{\tau} T \rfloor} \mu_{3,t} \left(\tau^* \right) \left. \frac{\partial \varepsilon_t \left(\psi \right)}{\partial \psi} \right|_{\psi = \psi^*} + T^{-1/2} \sum_{t=1+\lfloor \widehat{\tau} T \rfloor}^{T} \left(\mu_{3,t} \left(\tau^* \right) - \mu_{3,t} \left(\widehat{\tau} \right) \right) \left. \frac{\partial \varepsilon_t \left(\psi \right)}{\partial \psi} \right|_{\psi = \psi^*}$$

where we note that $\left. \frac{\partial \varepsilon_t (\psi)}{\partial \psi} \right|_{\psi = \psi^*}$ is still ARMA.

Using summation by parts,

$$\left| \sum_{t=1+\lfloor \tau^* T \rfloor}^{\lfloor \widehat{\tau} T \rfloor} \mu_{3,t} \left(\tau^* \right) \left. \frac{\partial \varepsilon_t \left(\psi \right)}{\partial \psi} \right|_{\psi = \psi^*} \right| \tag{S.55}$$

$$\leq \sum_{t=1+\lfloor \tau^* T \rfloor}^{\lfloor \widehat{\tau} T \rfloor - 1} \left| \Delta \mu_{3,t+1} \left(\tau^* \right) \right| \max_{1+\lfloor \tau^* T \rfloor \leq t \leq \lfloor \widehat{\tau} T \rfloor - 1} \left| \sum_{s=1+\lfloor \tau^* T \rfloor}^{t-1} \left. \frac{\partial \varepsilon_s \left(\psi \right)}{\partial \psi} \right|_{\psi = \psi^*} \right| \tag{S.56}$$

$$+ \mu_{3,\lfloor \widehat{\tau} T \rfloor} \left(\tau^* \right) \left| \sum_{t=1+\lfloor \tau^* T \rfloor}^{\lfloor \widehat{\tau} T \rfloor} \left. \frac{\partial \varepsilon_t \left(\psi \right)}{\partial \psi} \right|_{\psi = \psi^*} \right| \tag{S.57}$$

and

$$\left| \sum_{t=1+\lfloor \widehat{\tau} T \rfloor}^{T} \left(\mu_{3,t} \left(\tau^* \right) - \mu_{3,t} \left(\widehat{\tau} \right) \right) \left. \frac{\partial \varepsilon_t \left(\psi \right)}{\partial \psi} \right|_{\psi = \psi^*} \right| \tag{S.58}$$

$$\leq \sum_{t=1+\lfloor \widehat{\tau} T \rfloor}^{T-1} \left| \Delta \left(\mu_{3,t+1} \left(\tau^* \right) - \mu_{3,t+1} \left(\widehat{\tau} \right) \right) \right| \max_{1+\lfloor \widehat{\tau} T \rfloor \leq t \leq T-1} \left| \sum_{s=1+\lfloor \widehat{\tau} T \rfloor}^{t-1} \left. \frac{\partial \varepsilon_s \left(\psi \right)}{\partial \psi} \right|_{\psi = \psi^*} \right| \tag{S.59}$$

$$+ \left| \mu_{3,T} \left(\tau^* \right) - \mu_{3,T} \left(\widehat{\tau} \right) \right| \left| \sum_{t=1+\lfloor \widehat{\tau} T \rfloor}^{T} \left. \frac{\partial \varepsilon_t \left(\psi \right)}{\partial \psi} \right|_{\psi = \psi^*} \right|. \tag{S.60}$$

We discuss Model A first, beginning with the two components of the bound of (S.55). For (S.56), notice that

$$\sum_{t=1+\lfloor \tau^* T \rfloor}^{\lfloor \widehat{\tau} T \rfloor - 1} \left| \Delta \mu_{3,t+1} \left(\tau^* \right) \right| = \sum_{t=1}^{\lfloor \widehat{\tau} T \rfloor - \lfloor \tau^* T \rfloor} \left| \Delta \mu_{1,t} \right| \leq C \sum_{t=1}^{\lfloor \widehat{\tau} T \rfloor - \lfloor \tau^* T \rfloor} t^{-\delta} \leq C \left(\lfloor \widehat{\tau} T \rfloor - \lfloor \tau^* T \rfloor \right)^{1-\delta}$$

while

$$
\begin{aligned}
& \max_{1+\lfloor\tau^*T\rfloor\le t\le\lfloor\widehat{\tau}T\rfloor-1}\left|\sum_{s=1+\lfloor\tau^*T\rfloor}^{t-1}\frac{\partial\varepsilon_s(\psi)}{\partial\psi}\right| \qquad \text{(S.61)}\\
\le\ & \max_{1+\lfloor\tau^*T\rfloor\le t\le\lfloor\widehat{\tau}T\rfloor-1}\left|(t-\lfloor\tau^*T\rfloor)^{1/2}\right|\max_{1+\lfloor\tau^*T\rfloor\le t\le\lfloor\widehat{\tau}T\rfloor-1}\left|(t-\lfloor\tau^*T\rfloor)^{-1/2}\sum_{s=1+\lfloor\tau^*T\rfloor}^{t-1}\frac{\partial\varepsilon_s(\psi)}{\partial\psi}\right|\\
\le\ & (\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor)^{1/2}\max_{1+\lfloor\tau^*T\rfloor\le t\le\lfloor\widehat{\tau}T\rfloor-1}\left|(t-\lfloor\tau^*T\rfloor)^{-1/2}\sum_{s=1+\lfloor\tau^*T\rfloor}^{t-1}\frac{\partial\varepsilon_s(\psi)}{\partial\psi}\right|\\
\le\ & (\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor)^{1/2}\max_{1+\lfloor\tau^*T\rfloor\le t\le T}\left|(t-\lfloor\tau^*T\rfloor)^{-1/2}\sum_{s=1+\lfloor\tau^*T\rfloor}^{t-1}\frac{\partial\varepsilon_s(\psi)}{\partial\psi}\right|
\end{aligned}
$$

and, using Equation (8) of Bai (1994),

$$\max_{1+\lfloor\tau^*T\rfloor\le t\le T}\left|(t-\lfloor\tau^*T\rfloor)^{-1/2}\sum_{s=1+\lfloor\tau^*T\rfloor}^{t-1}\left.\frac{\partial\varepsilon_t(\psi)}{\partial\psi}\right|_{\psi=\psi^*}\right| = O_p(\ln(T))$$

so that the stochastic order of (S.61) is the same as $(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor)^{1/2}\ln(T)$ and the order of (S.56) is the same as,

$$(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor)^{1-\delta}(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor)^{1/2}\ln(T)$$

which is of $o_p(1)$ using (3.12).

For the remainder term (S.57), $\mu_{3,\lfloor\widehat{\tau}T\rfloor}(\tau^*)\le C(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor)^{1-\delta}$. Again using Equation (8) of Bai (1994), (S.57) has the same stochastic order as

$$(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor)^{1-\delta}\times\ln(T)\times(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor)^{1/2}$$

which is of $o_p(1)$. Hence, the stochastic order of (S.55) is $o_p(1)$ if Model A is used.

Moving to the two components of the bound of (S.58), term in (S.59) is bounded by

$$\sum_{t=1+\lfloor\widehat{\tau}T\rfloor}^{T-1}\left|\Delta\left(\mu_{3,t+1}(\tau^*)-\mu_{3,t+1}(\widehat{\tau})\right)\right|\sup_{\rho_1,\,\rho_2}\left|\sum_{s=\lfloor\rho_1T\rfloor}^{\lfloor\rho_2T\rfloor}\left.\frac{\partial\varepsilon_s(\psi)}{\partial\psi}\right|_{\psi=\psi^*}\right|$$

where $\sup_{\rho_1,\,\rho_2}\left|\sum_{s=\lfloor\rho_1T\rfloor}^{\lfloor\rho_2T\rfloor}\left.\frac{\partial\varepsilon_s(\psi)}{\partial\psi}\right|_{\psi=\psi^*}\right|$. Noticing that

$$\Delta\left(\mu_{3,t}(\tau^*)-\mu_{3,t}(\widehat{\tau})\right)=\Delta\left(\mu_{3,t}(\tau^*)-\mu_{3,t-1}(\tau^*)+\mu_{3,t-1}(\tau^*)-...-\mu_{3,t}(\widehat{\tau})\right)$$

and the bound for $\Delta\mu_{1,t}$, then, if $\delta>0$,

$$\left|\Delta\left(\mu_{3,t}(\tau^*)-\mu_{3,t}(\widehat{\tau})\right)\right|<C(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor)(t-\lfloor\widehat{\tau}T\rfloor)^{-1}$$

and

$$
\begin{aligned}
& \sum_{t=1+\lfloor\widehat{\tau}T\rfloor}^{T-1}\left|\Delta\left(\mu_{3,t+1}(\tau^*)-\mu_{3,t+1}(\widehat{\tau})\right)\right|\le C(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor)\sum_{t=1+\lfloor\widehat{\tau}T\rfloor}^{T-1}(t-\lfloor\widehat{\tau}T\rfloor)^{-1}\\
\le\ & C(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor)\sum_{t=1}^{T}t^{-1}\le C(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor)\ln(T)=O_p\left(T^{-1/2+\delta}\ln(T)\right)
\end{aligned}
$$

so that (S.59) is of order $O_p\left(T^{-1/2+\delta}\times\ln(T)\times T^{1/2}\right)=O_p\left(T^{\delta}\ln(T)\right)=o_p\left(T^{1/2}\right)$.

If $\delta<0$,

$$\left|\Delta\left(\mu_{3,t}(\tau^*)-\mu_{3,t}(\widehat{\tau})\right)\right|<C\left(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor\right)\left(t-\lfloor\widehat{\tau}T\rfloor\right)^{-1-\delta}$$

$$\sum_{t=1+\lfloor\widehat{\tau}T\rfloor}^{T-1}\left|\Delta\left(\mu_{3,t+1}(\tau^*)-\mu_{3,t+1}(\widehat{\tau})\right)\right|\le C\left(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor\right)T^{-\delta}$$

and (S.59) has stochastic order as

$$\left(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor\right)T^{-\delta}T^{1/2}=O_p\left(T^{-1/2+\delta}T^{-\delta}T^{1/2}\right)=O_p(1)=o_p\left(T^{1/2}\right).$$

So, regardless of whether $\delta<0$ or $\delta>0$, (S.59) is of $o_p\left(T^{1/2}\right)$.

For the remainder term in (S.60), recalling (S.47) or (S.48),

$$\left|\mu_{3,T}(\tau^*)-\mu_{3,T}(\widehat{\tau})\right|<C\left(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor\right)T^{-\delta}=O_p\left(T^{-1/2+\delta}\times T^{-\delta}\right)=O_p\left(T^{-1/2}\right)$$

and (S.60) is therefore of order $O_p\left(T^{-1/2}\times T^{1/2}\right)=O_p(1)$. We can then conclude that, under Model A, (S.55) and (S.58) are $o_p\left(T^{1/2}\right)$ and (S.53) is $o_p(1)$.

We now discuss the case when Model B is used, again considering (S.55) and (S.58). Beginning with the two components of the bound of (S.55), if $\delta<0$,

$$\sum_{t=1+\lfloor\tau^*T\rfloor}^{\lfloor\widehat{\tau}T\rfloor-1}\left|\Delta\mu_{3,t+1}(\tau^*)\right|=\sum_{t=1}^{\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor}\left|\Delta\mu_{1,t}\right|\le C\sum_{t=1}^{\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor}t^{-1-\delta}\le C\left(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor\right)^{-\delta}$$

and, recalling the bound for (S.61), (S.56) has stochastic order

$$\left(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor\right)^{-\delta}\times\left(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor\right)^{1/2}\ln(T)=O_p(\ln(T))$$

where we have used the result that $\left(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor\right)=O_p(1)$, as in (3.13).

If $\delta>0$,

$$\sum_{t=1+\lfloor\tau^*T\rfloor}^{\lfloor\widehat{\tau}T\rfloor-1}\left|\Delta\mu_{3,t+1}(\tau^*)\right|\le C\sum_{t=1}^{\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor}t^{-1}\le C\ln(T)$$

and, recalling the bound for (S.61), then (S.56) has stochastic order $O_p((\ln(T))^2)$. Thus, regardless of δ, (S.56) has order $O_p((\ln(T))^2)$. For the remainder term in (S.57), $\mu_{3,\lfloor\widehat{\tau}T\rfloor}(\tau^*)\le C\left(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor\right)^{-\delta}$ and so (S.57) has the same stochastic order as that of $\left(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor\right)^{-\delta}\times\ln(T)\times\left(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor\right)^{1/2}=O_p(\ln(T))$. Consequently, (S.55) is of $O_p\left((\ln(T))^2\right)$.

Turning to (S.58), recall first that

$$\left(\mu_{3,t}(\tau^*)-\mu_{3,t}(\widehat{\tau})\right)=\left(\mu_{3,t}(\tau^*)-\mu_{3,t-1}(\tau^*)+\mu_{3,t-1}(\tau^*)-...-\mu_{3,t}(\widehat{\tau})\right)$$

then

$$\left|\mu_{3,t}(\tau^*)-\mu_{3,t}(\widehat{\tau})\right|<C\left(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor\right)\left(t-\lfloor\widehat{\tau}T\rfloor\right)^{-1-\delta}$$

if $\delta<0$, and

$$\left|\mu_{3,t}(\tau^*)-\mu_{3,t}(\widehat{\tau})\right|<C\left(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor\right)\left(t-\lfloor\widehat{\tau}T\rfloor\right)^{-1}$$

if $\delta > 0$. Where $\delta < 0$, (S.58) is therefore bounded by

$$\begin{aligned} & \sum_{t=1+\lfloor\widehat{\tau}T\rfloor}^{T} C\left(\lfloor\widehat{\tau}T\rfloor - \lfloor\tau^{*}T\rfloor\right)\left(t - \lfloor\widehat{\tau}T\rfloor\right)^{-\delta-1}\left|\left.\frac{\partial\varepsilon_{t}\left(\psi\right)}{\partial\psi}\right|_{\psi=\psi^{*}}\right| \\ \leq & \sum_{t=1+\lfloor\widehat{\tau}T\rfloor}^{T} C\left(\lfloor\widehat{\tau}T\rfloor - \lfloor\tau^{*}T\rfloor\right)\left(t - \lfloor\widehat{\tau}T\rfloor\right)^{-\delta-1}\sup_{t}\left|\left.\frac{\partial\varepsilon_{t}\left(\psi\right)}{\partial\psi}\right|_{\psi=\psi^{*}}\right| \\ \leq & C\left(\lfloor\widehat{\tau}T\rfloor - \lfloor\tau^{*}T\rfloor\right)\sum_{t=1}^{T} t^{-\delta-1}\sup_{t}\left|\left.\frac{\partial\varepsilon_{t}\left(\psi\right)}{\partial\psi}\right|_{\psi=\psi^{*}}\right|. \end{aligned}$$

Using the fact, which will be established below, that

$$\sup_{t}\left|\left.\frac{\partial\varepsilon_{t}\left(\psi\right)}{\partial\psi}\right|_{\psi=\psi^{*}}\right| = O_{p}\left(T^{1/q}\right) \tag{S.62}$$

the stochastic order of (S.58) when $\delta < 0$ is

$$O_{p}\left(\sum_{t=1}^{T} t^{-\delta-1}T^{1/q}\right) = O_{p}\left(T^{-\delta+1/q}\right) = o_{p}\left(T^{1/2}\right)$$

in view of the condition that $q > 1/\left(1/2+\delta\right)$ imposed by Assumption 1. Where $\delta > 0$, (S.58) is bounded by

$$C\left(\lfloor\widehat{\tau}T\rfloor - \lfloor\tau^{*}T\rfloor\right)\sum_{t=1}^{T} t^{-1}\sup_{t}\left|\left.\frac{\partial\varepsilon_{t}\left(\psi\right)}{\partial\psi}\right|_{\psi=\psi^{*}}\right| = O\left(\ln\left(T\right)T^{1/q}\right) = o_{p}\left(T^{1/2}\right)$$

using the fact that $q > 2$.

We have therefore proved (S.53) for all cases. To complete the proof of Lemma C2, the bound for (S.54) can be established in the same way.

We end this proof with a derivation of the result stated in (S.62). Let X_t and Y_t be two random variables such that $X_t = O_p\left(f_t\right)$ and $Y_t = O_p\left(g_t\right)$, where f_t and g_t are positive sequences in t. Then, as is well known, see for example White (2001,p.28), that

$$\begin{aligned} X_tY_t &= O_p\left(f_tg_t\right) && \text{(S.63)} \\ X_t + Y_t &= O_p\left(\max\left(f_t, g_t\right)\right). && \text{(S.64)} \end{aligned}$$

Moreover, for $p > 0$,

$$\left|X_t\right|^p = O_p\left(f_t^p\right). \tag{S.65}$$

Finally, let ξ_t be a process with $\left|\xi_t\right| = O_p\left(1\right)$ for any t. Then, for any $p > 0$,

$$\sup_{t=1,\ldots,T}\left|\xi_t\right| = O_p\left(T^{1/p}\right). \tag{S.66}$$

To establish (S.65), notice first that $X_t/f_t = O_p\left(1\right)$, letting $Y_t = f_t^{-1}$ in (S.63). Rewriting $\left|X_t\right|^p = f_t^p\left|X_t/f_t\right|^p$, in view of (S.63), the result follows if $\left|X_t/f_t\right|^p = O_p\left(1\right)$. To establish this, let $\lfloor p\rfloor$ be the integer part of p, and $\mathbb{I}\left(A\right)$ the indicator function, that takes value 1 if the event A is true, and 0 otherwise, and let $P := \lfloor p \rfloor + 1\mathbb{I}(p - \lfloor p \rfloor > 0)$, so $P = p$ if p is an integer, and $P = \lfloor p \rfloor + 1$ otherwise; that is, P is ceiling of p. Notice that, for any sequence x_t, it holds that $|x_t|^p \leq 1 + |x_t|^P$, and so

$$|X_t/f_t|^p \leq 1 + |X_t/f_t|^P . \tag{S.67}$$

For $p < 1$, using (S.67) with $P = 1$, $|X_t/f_t|^p \leq 1 + |X_t/f_t| = O_p(1)$ by (S.64). For $1 < p \leq 2$, first notice that $|X_t/f_t|^2 = |X_t/f_t| \times |X_t/f_t| = O_p(1)$ in view of (S.63). The result then follows using (S.67) with $P = 2$, $|X_t/f_t|^2 = O_p(1)$ and (S.64). Higher values of p, for any finite P, can be treated in the same way, thus establishing (S.65).

To establish (S.66), notice first that, in view of the (S.65), for any t it holds that $|\xi_t|^p = O_p(1)$. Next, notice that $\max_t |\xi_t|^p \leq \sum_{t=1}^{T} |\xi_t|^p = O_p(T)$, i.e., $|\xi_t|^p = O_p(T)$, uniformly in t. As the power is a monotone mapping, then $\max_t |\xi_t|^p = (\max_t |\xi_t|)^p$, and $\max_t |\xi_t| = (\max_t |\xi_t|^p)^{1/p}$. Thus, $|\xi_t| = O_p\left(T^{1/p}\right)$ uniformly in t.

In view of the fact that p in (S.66) is arbitrary, we can take $p \geq q$ to establish the result in (S.62).

Proof of Lemma D2.
Using the expansion in (S.41) again, the first two terms can be accounted for proceeding as in the proof of Lemma B2, using (3.11) in place of (3.14). The additional contribution of the term $g(L;\psi)\,\Delta_+^{\delta}\left(z_t(\tau^*) - z_t(\widehat{\tau})\right)' \widehat{\beta}(\widehat{\tau})$ is discussed proceeding as in Lemma C2.

S.2 Additional Monte Carlo Simulations

Throughout this supplement, the simulation DGPs used are as detailed in section 4 except for those changed aspects detailed in each case considered below.

S.2.1 Power against Fixed Magnitude Alternatives

In Theorem 1 we established that the test based on $LM(\widehat{\tau})$ has non-trivial asymptotic local power, achieving the Gaussian local power envelope. Finite sample simulations of power against local alternatives were reported in section 4. In the additional simulations reported here we investigate finite sample power against fixed alternatives; that is, where the distance between the true long memory parameter, d, and the value imposed under the null hypothesis, d_0, is not a function of the sample size, T. Of particular interest is the case where Model A is implied under H_0 ($d_0 < 1/2$), but in fact Model B should be used ($d > 1/2$), or vice-versa. We will also consider power in the classical set up of the unit-root test (as in the Dickey-Fuller test), testing $H_0 : d_0 = 1$ when in fact the true DGP is a stationary AR(1), with autoregressive parameter 0.9 (so that, in fact, $d = 0$).

For simplicity and for ease of exposition, for the first part of this exercise we consider the DGP $e_t = \Delta_+^{-d}\varepsilon_t$ for the following four cases for d_0 (the incorrect null value) and d (the true value): $(d, d_0) \in \{(0.6, 0.4), (0.75, 0.25), (0.4, 0.6), (0.25, 0.75)\}$. Table S1 below gives the results for nominal asymptotic 0.05 level tests.

Focussing on the results for $LM(\widehat{\tau})$, from this exercise, these results can be summarised as: (i) finite sample power for given (d, d_0), increases with T, and (ii) finite sample power for a given T increases with the distance $|d_0 - d|$. With regard to (ii), it is also worth commenting that over-differencing (i.e. basing the $LM(\widehat{\tau})$ test on Model A when Model B is in fact the correct choice for the true long memory parameter) leads to tests with lower power than under-differencing (basing the test on Model B when Model A is the correct choice), for a given value of $|d - d_0|$. When $d - d_0 > 0$ the autocorrelations are not summable, whereas when $d - d_0 < 0$ the autocorrelations sum to zero. The former is easier to detect using tests such as $LM(\widehat{\tau})$, which is based on a sum of weighted sample autocorrelations.

Our second set of simulations are concerned with conventional unit root testing, when the alternative is that of the traditional Dickey-Fuller (DF) type. The null hypothesis is $H_0 : d_0 = 1$ such that $e_t = e_{t-1} + \varepsilon_t$ when the true DGP is in fact $I(0)$ but with autoregressive root close to 1, $e_t = 0.9e_{t-1} + \varepsilon_t$, with $\varepsilon_t \sim$ i.i.d. $N(0, 1)$. Table S2 below reports the results of these experiments, again for tests run at the nominal asymptotic 0.05 level. We can observe from these results that the test based on $LM(\widehat{\tau})$ has power that increases in T, and has similar power to the infeasible LM test, regardless of whether a trend break occurs or not.

Table S1. Empirical power of tests for distant alternatives

$d = 0.6$, $d_0 = 0.4$					
	LM	$LM(\tau^*)$		$LM(\widehat{\tau})$	
T			$\beta_3 = 0.0$	$\beta_3 = 0.1$	$\beta_3 = 1.0$
256	1.000	0.825	0.731	0.727	0.763
512	1.000	0.996	0.992	0.993	0.994
1024	1.000	1.000	1.000	1.000	1.000
$d = 0.75$, $d_0 = 0.25$					
	LM	$LM(\tau^*)$		$LM(\widehat{\tau})$	
T			$\beta_3 = 0.0$	$\beta_3 = 0.1$	$\beta_3 = 1.0$
256	1.000	1.000	1.000	1.000	1.000
512	1.000	1.000	1.000	1.000	1.000
1024	1.000	1.000	1.000	1.000	1.000
$d = 0.4$, $d_0 = 0.6$					
	LM	$LM(\tau^*)$		$LM(\widehat{\tau})$	
T			$\beta_3 = 0.0$	$\beta_3 = 0.1$	$\beta_3 = 1.0$
256	0.928	0.895	0.742	0.802	0.826
512	1.000	0.999	0.991	0.996	0.991
1024	1.000	1.000	1.000	1.000	1.000
$d = 0.25$, $d_0 = 0.75$					
	LM	$LM(\tau^*)$		$LM(\widehat{\tau})$	
T			$\beta_3 = 0.0$	$\beta_3 = 0.1$	$\beta_3 = 1.0$
256	1.000	1.000	1.000	1.000	1.000
512	1.000	1.000	1.000	1.000	1.000
1024	1.000	1.000	1.000	1.000	1.000

Table S2. Empirical power of tests for DF type alternative

$d = 0$, $a = 0.9$					
	LM	$LM(\tau^*)$		$LM(\widehat{\tau})$	
T			$\beta_3 = 0.0$	$\beta_3 = 0.1$	$\beta_3 = 1.0$
256	0.309	0.312	0.324	0.334	0.315
512	0.678	0.678	0.680	0.689	0.679
1024	0.964	0.964	0.964	0.965	0.964

S.2.2 Moment Conditions

Assumption 1 imposes the moment conditions $\mathrm{E}\left|\varepsilon_t\right|^{\overline{q}} < \infty$ for $\overline{q} > \max\left(2, 2/\left(1+2d\right)\right)$ if $d \in (-0.5, 0.5)$, $\overline{q} > \max\left(2, 2/\left(2d-1\right)\right)$ if $d \in (0.5, 1.5)$. For $d \in (-0.5, 0)$ and $d \in (0.5, 1)$ these are stronger than, for example, the moment conditions in Nielsen (2004), who needed only $\overline{q} \geq 2$ to establish his results, and may be very strong; for example, when $d \to 0.5^+$, $2/\left(2d-1\right) \to \infty$. For the case where no trend break occurs, these conditions are required to establish uniformly in τ results for the $LM\left(\tau\right)$ statistic: our proof is based on the application of a functional central limit theorem for partial sums of fractionally integrated processes, and similar conditions are necessary; see Johansen and Nielsen (2012). Where a trend break occurs, similar conditions are used to derive a sufficient rate of convergence for the estimate $\widehat{\tau}$; see, for example, Condition A of Chang and Perron (2016).

To investigate the consequences of the required moment conditions not being met, we simulate the tests in the case of a fractional noise process, $e_t = \Delta_+^{-d}\varepsilon_t$, with $d = 0.51$, 0.55,0.6, 0.75, 1.0, for ε_t either standard normal or t_5 innovations. We summarize the minimum moment requirements $\mathrm{E}\left|\varepsilon_t\right|^{\overline{q}} < \infty$ with $\overline{q} > q_0$ for q_0 as given in the table:

d	0.51	0.55	0.60	0.75	1.00
q_0	100	20	10	4	2

We observe therefore that these conditions are always met in case of normally distributed innovations, but are only met when $d = 1$ in the case of t_5 innovations. The moment conditions of Nielsen (2004) are met by both of these innovation distributions. Alongside the $LM\left(\widehat{\tau}\right)$ test, we also simulated the $\overline{LM}$ and $LM(\tau^*)$ tests, to verify that the stronger moment conditions are not needed in these cases, in line with Nielsen (2004). We use $T = 256$, 512, 1024 and for values of d close to 0.5 we also consider $T = 2048$, 4096, 8192. The results are given in Table S3 below, again for nominal asymptotic 0.05 level tests. The main conclusions we can draw from the results in Table S3 are as follows:

(i) That the moment conditions of Assumption 1 are not needed for the $\overline{LM}$ and $LM(\tau^*)$ tests is clearly seen in the results. As a general pattern, empirical sizes appear to converge towards the nominal 0.05 level for all values of d for both innovation distributions for these tests.

(ii) The moment conditions for $LM(\widehat{\tau})$ are not met for the t_5 distributed innovations except for the $d = 1$ case, whereas these are always met for normally distributed innovations. We see from the results in Table S3 that for d up to $d = 0.75$ the empirical size of $LM(\widehat{\tau})$ is generally badly inflated for the case of t_5 innovations *vis-à-vis* normally distributed innovations.

(iii) Indeed, for t_5 distributed observations we find that for $d = 0.51$ or $d = 0.55$ (i.e. the most demanding moment conditions) the empirical size of the $LM(\widehat{\tau})$ test appears to be diverging when $\beta_3 = 0$ even for extremely large T. For $\beta_3 \neq 0$ empirical sizes appear to diverge at first, but then appear to be corrected, approaching 0.05 for the very large values of T considered. The case of $d = 0.6$ displays less acute size distortions but we still find that the $LM(\widehat{\tau})$ is unreliable, especially when $\beta_3 = 0$. Thus, for t_5 distributed innovations, the size properties deteriorate as the "gap" between the required moment condition and the actual moment is increased.

Table S3. Empirical size in presence of standard normal and t_5 distributed innovations

$d = 0.51$

	std. normal					t_5				
	$\overline{LM}$	$LM(\tau^*)$		$LM(\widehat{\tau})$		$\overline{LM}$	$LM(\tau^*)$		$LM(\widehat{\tau})$	
T	$\beta_3 = 0$		$\beta_3 = 0$	$\beta_3 = 0.1$	$\beta_3 = 1$			$\beta_3 = 0$	$\beta_3 = 0.1$	$\beta_3 = 1$
256	0.045	0.049	0.118	0.086	0.083	0.045	0.051	0.267	0.188	0.105
512	0.050	0.056	0.162	0.106	0.099	0.049	0.056	0.365	0.252	0.123
1024	0.050	0.057	0.201	0.109	0.099	0.052	0.059	0.456	0.231	0.121
2048	0.051	0.055	0.211	0.103	0.938	0.053	0.059	0.537	0.212	0.121
4096	0.054	0.056	0.211	0.102	0.085	0.059	0.060	0.580	0.180	0.106
8192	0.049	0.053	0.200	0.085	0.080	0.053	0.055	0.619	0.162	0.100

$d = 0.55$

	std. normal					t_5				
	$\overline{LM}$	$LM(\tau^*)$		$LM(\widehat{\tau})$		$\overline{LM}$	$LM(\tau^*)$		$LM(\widehat{\tau})$	
T	$\beta_3 = 0$		$\beta_3 = 0$	$\beta_3 = 0.1$	$\beta_3 = 1$	$\beta_3 = 0$		$\beta_3 = 0$	$\beta_3 = 0.1$	$\beta_3 = 1$
256	0.042	0.042	0.078	0.059	0.062	0.041	0.043	0.170	0.130	0.083
512	0.046	0.049	0.098	0.075	0.076	0.045	0.048	0.234	0.171	0.092
1024	0.047	0.048	0.116	0.076	0.077	0.048	0.053	0.282	0.156	0.092
2048	0.047	0.049	0.121	0.073	0.073	0.050	0.050	0.324	0.142	0.091
4096	0.052	0.052	0.116	0.075	0.068	0.055	0.054	0.344	0.120	0.088

$d = 0.6$

	std. normal					t_5				
	$\overline{LM}$	$LM(\tau^*)$		$LM(\widehat{\tau})$		$\overline{LM}$	$LM(\tau^*)$		$LM(\widehat{\tau})$	
T	$\beta_3 = 0$		$\beta_3 = 0$	$\beta_3 = 0.1$	$\beta_3 = 1$	$\beta_3 = 0$		$\beta_3 = 0$	$\beta_3 = 0.1$	$\beta_3 = 1$
256	0.041	0.039	0.051	0.044	0.050	0.037	0.039	0.093	0.079	0.062
512	0.043	0.044	0.061	0.059	0.059	0.042	0.040	0.123	0.096	0.069
1024	0.045	0.046	0.070	0.057	0.060	0.046	0.049	0.139	0.100	0.071
2048	0.046	0.046	0.072	0.056	0.059	0.048	0.048	0.154	0.090	0.070

$d = 0.75$

	std. normal					t_5				
	$\overline{LM}$	$LM(\tau^*)$		$LM(\widehat{\tau})$		$\overline{LM}$	$LM(\tau^*)$		$LM(\widehat{\tau})$	
T	$\beta_3 = 0$		$\beta_3 = 0$	$\beta_3 = 0.1$	$\beta_3 = 1$	$\beta_3 = 0$		$\beta_3 = 0$	$\beta_3 = 0.1$	$\beta_3 = 1$
256	0.036	0.038	0.039	0.047	0.040	0.033	0.037	0.039	0.042	0.041
512	0.040	0.042	0.045	0.047	0.045	0.040	0.042	0.044	0.047	0.044
1024	0.044	0.046	0.048	0.049	0.046	0.043	0.046	0.048	0.050	0.048

$d = 1$

	std. normal					t_5				
	$\overline{LM}$	$LM(\tau^*)$		$LM(\widehat{\tau})$		$\overline{LM}$	$LM(\tau^*)$		$LM(\widehat{\tau})$	
T	$\beta_3 = 0$		$\beta_3 = 0$	$\beta_3 = 0.1$	$\beta_3 = 1$	$\beta_3 = 0$		$\beta_3 = 0$	$\beta_3 = 0.1$	$\beta_3 = 1$
256	0.036	0.041	0.060	0.059	0.044	0.034	0.039	0.058	0.057	0.043
512	0.039	0.043	0.063	0.063	0.044	0.040	0.042	0.062	0.061	0.046
1024	0.044	0.045	0.059	0.057	0.046	0.043	0.047	0.060	0.058	0.048

S.2.3 Model Selection

In the Monte Carlo simulations in section 4 of the paper we assumed knowledge of the correct ARMA specification for the short memory component of the model. This is not usually known in practice and so here we investigate the consequences of selecting the short memory component of the model using the familiar Bayes Information Criterion (BIC) of Schwarz (1978). We will consider just the case of $d = 1$ in the interest of brevity. We simulated the same AR(1) with $a = 0.5$ as we did for the exercise summarised in Table 3 in the paper, but we now selected the lag of the AR model using the BIC, choosing between the i.i.d. model (underfitting), AR(1) model (correct fitting) and AR(2) model (overfitting). Thus, after simulating η_t as $\eta_t = 0.5\eta_{t-1} + \varepsilon_t$ and $e_t = \Delta_+^{-1}\eta_t$ and simulating $x_t = \beta_1 + \beta_2 t + \beta_3 DT_t(\tau^*) + e_t$, we estimated $\widehat{\tau}$ from Model B and then $\widehat{\beta}_2(\widehat{\tau})$, $\widehat{\beta}_3(\widehat{\tau})$, and computed the residuals $\widehat{u}_t(\widehat{\tau})$, see equation (3.6), and finally, noticing that under H_0 $u_t = \Delta e_t$ is $I(0)$, we computed $\widehat{\eta}_t(\widehat{\tau}) := \widehat{u}_t(\widehat{\tau})$. For comparison, we also repeated the exercise assuming that the true τ^* is known, again estimating $\widehat{\beta}_2(\tau^*)$, $\widehat{\beta}_3(\tau^*)$ from Model B, then computing residuals $\widehat{u}_t(\tau^*)$ and finally $\widehat{\eta}_t(\tau^*) := \widehat{u}_t(\tau^*)$. As a second comparison, for the case $\beta_3 = 0$ only, we also estimated $\overline{\beta}_2$ in the regression model $\Delta x_t = \beta_2 + u_t$ and computed residuals $\overline{u}_t$ and then $\overline{\eta}_t := \overline{u}_t$, as we would do with the knowledge that $\beta_3 = 0$. When the DGP for η_t is known, we can use $\widehat{\eta}_t(\widehat{\tau})$, $\widehat{\eta}_t(\tau^*)$ and $\overline{\eta}_t$ to compute the $LM(\widehat{\tau})$, $LM(\tau^*)$ and $\overline{LM}$ statistics, respectively: in this exercise, we first selected models for $\widehat{\eta}_t(\widehat{\tau})$, $\widehat{\eta}_t(\tau^*)$ and $\overline{\eta}_t$ using BIC. This information criterion yields consistent estimation of ARMA structure when the series η_t is used, and we are interested in particular in checking if the same holds when residuals $\widehat{\eta}_t(\widehat{\tau})$ are used instead, and what consequences estimating the orders has on the $LM(\widehat{\tau})$ test.

In our experiment, the i.i.d. model was never selected by the BIC in the 10,000 replications considered. The frequency with which the correct AR(1) model was chosen by the BIC is given in the table below. In the remaining cases BIC selected the AR(2) model.

	$\overline{\eta}_t$	$\eta_t(\tau^*)$		$\eta_t(\widehat{\tau})$	
T	$\beta_3 = 0$		$\beta_3 = 0$	$\beta_3 = 0.1$	$\beta_3 = 1$
256	0.9784	0.9777	0.9757	0.9766	0.9775
512	0.9877	0.9875	0.9868	0.9870	0.9873
1024	0.9911	0.9912	0.9907	0.9907	0.9915

We can therefore observe that the BIC correctly selects the AR(1) model in the vast majority of cases, and that this selection frequency is tending towards one as T increases. Moreover, estimation of the location of the break would appear to have almost no impact on the efficacy of the BIC to select the correct model for the shocks.

We then repeated the simulation experiment given in Table 3 of the main paper but where we now estimated the order of the short memory AR component using the BIC. These results are reported in the table below.

	$\overline{LM}$	$LM(\tau^*)$		$LM(\widehat{\tau})$	
T	$\beta_3 = 0$		$\beta_3 = 0$	$\beta_3 = 0.1$	$\beta_3 = 1$
256	0.010	0.011	0.020	0.020	0.014
512	0.020	0.024	0.039	0.039	0.025
1024	0.026	0.033	0.052	0.050	0.036

These results are observed to be basically identical to those reported in Table 3, with any changes only occurring at the third decimal place.

Additional References

These are the additional references cited in this supplementary appendix and not listed in the main paper.

Andrews, D.W.K., 1992. Generic uniform convergence, *Econometric Theory* 8, 241-257.

Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change point, *Econometrica* 61, 821-856.

Bai, J., 1993. On the partial sums of residuals in autoregressive and moving average models, *Journal of Time Series Analysis* 14, 247-260.

Bai, J., 1994. Least squares estimation of a shift in linear processes, *Journal of Time Series Analysis* 15, 453-472.

Billingsley, P., 1968. *Convergence of Probability Measures.* New York: Wiley.

Iacone, F., S.J. Leybourne and A.M.R. Taylor, 2014. A fixed-b test for a break in level at an unknown time under fractional integration, *Journal of Time Series Analysis* 35, 40-54.

Marinucci, D., and P.M. Robinson, 1999. Alternative forms of fractional Brownian motion, *Journal of Statistical Planning and Inference* 80, 111-122.

Newey , W., and D. McFadden, 1994. Large sample estimation and hypothesis testing, in *Handbook of Econometrics,* vol. IV, Ch. 36, 2111-2245.

Robinson, P.M., 2005. Efficiency improvements in inference on stationary and nonstationary time series, *Annals of Statistics* 33, 1800-1842.

Shao, X., 2011. A simple test of changes in mean in the possible presence of long-range dependence, *Journal of Time Series Analysis* 32, 598-606.

White, H., 2001. Asymptotic Theory for Econometricians, Academic Press.