Essex Finance Centre
Working Paper Series

Working Paper No19: 05-2017

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

“Fabrizio Iacone, Stephen J. Leybourne and A.M. Robert Taylor”

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

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], Stephen J. Leybourne[b] and A.M. Robert Taylor[c]

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

[b] School of Economics, University of Nottingham

[c] Essex Business School, University of Essex

May 9, 2017

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. 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, and unlike conventional unit root and stationarity tests, 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.

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

*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

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. This class of tests therefore has significant advantages over procedures that consider either the integer null of $d = 1$ against the integer alternative $d = 0$, the so-called *unit root* tests such as that of Dickey and Fuller (1979), or the null of $d = 0$ against the alternative $d = 1$, the so-called *stationarity* tests such as that of Kwiatkowski, Phillips, Schmidt and Shin (1992). In particular, the limiting null distributions of these unit root and stationarity statistics are 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. 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. Compared to unit root or stationarity tests, that would also now be based on a corresponding estimated trend break points, the advantages of our approach become further emphasised. The asymptotics underlying the unit root and stationarity tests are critically dependent on whether breaks occurs or not; see, for example, Perron and Rodríguez (2003) in the context of unit root testing, and Busetti and Harvey (2001,2003) in the context of stationarity testing. Crucially, in addition to the power losses noted above, the limiting null distributions of these tests in the cases where breaks are fitted differ according to the number of breaks fitted, the number of breaks actually present and the locations of those, making correctly sized inference rather problematic.

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 an evaluation of the finite sample size and power properties of our LM-type test. Section 5 concludes. Proofs are contained in a mathematical appendix.

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 $\xrightarrow{p}$ and 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 shock, e_t, is a zero mean, fractionally integrated process of order d, denoted $e_t \in I(d)$, and we will assume that $d \in (-0.5, 0.5) \cup (0.5, 1.5)$. Both stationary, non-stationary, and fractionally over-differenced time series are therefore permitted within our set-up. Precise conditions on the shocks will be given below. In (2.1), the deterministic trend break term, $DT_t(\tau^*)$, is defined for a generic τ as $DT_t(\tau) := (t - \lfloor \tau T \rfloor)\mathbb{I}(t \geq \lfloor \tau T \rfloor)$. 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.

Writing $d =: d_0 + \theta$, our interest in this paper focuses on testing the null hypothesis $H_0 : \theta = 0$ in (2.1); that is, $e_t \in I(d_0)$ under H_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 II" 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 $\sigma_\infty^2 := \sum_{h=-\infty}^{\infty} E(\eta_t \eta_{t+h})$. 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 $\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2)$ with $0 < \sigma_\varepsilon^2 < \infty$; and (b) the following higher-order moment conditions hold on ε_t, $E|\varepsilon_t|^{\overline{q}} < \infty$ for $\overline{q} > \max(2, 2/(1+2d))$ if $d \in (-0.5, 0.5)$, $\overline{q} > \max(2, 2/(2d-1))$ 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). 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 theorem result will be needed on the estimates of the β_j, $j = 1, 2, 3$, parameters charaterising 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.

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*. Defining $g(z;\psi) := a(z)\,b^{-1}(z)$, we can estimate the parameter vector $\psi^* := (a_1, ..., a_p, b_1 ..., b_q)'$ under H_0 as

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

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), a consistent estimator of ω^2 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 2. A one-sided test could also be considered, using 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 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 3. As discussed in Nielsen (2004, p.126) the foregoing LM test for the null hypothesis H_0, is asymptotically equivalent under H_c to the corresponding Wald and Likelihood Ratio tests for testing H_0. Moreover, as discussed in Robinson (1994) and Nielsen (2004), these tests 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 so 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 holds 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 4. 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 the $(d_0 - 1)$th differences of the de-trended residuals $\widehat{u}_t(\tau)$, defined for Model B in (3.6) below, will annihilate the estimated trend component when $d_0 > 1.5$. However, for $d_0 > 1.5$ the trend component will have no impact on the large sample behaviour of these statistics and they will therefore have the same large sample behaviour as given for LM in (3.4).

Remark 5. 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 determinisitic 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

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

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 6. 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); cf. Chang and Perron (2016) and Lavielle and Moulines (2000).

Remark 7. 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 functional central limit theorem; see Chang and Perron (2016). Secondly, in the case where $\beta_3 = 0$, then a functional central limit theorem is used to establish that the rate given in (3.14) holds uniformly in τ.

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 8. 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 9. 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 10. 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 $d_0 = 0.5$. When $\beta_3 \neq 0$, as noted in Remark 9, 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 $\sum_{t=1}^{T}(\sum_{j=1}^{t-1} j^{-1}\Delta_+^{d_0}(DT_{t-j}(\widehat{\tau}) - DT_{t-j}(\tau^*)))\widehat{\varepsilon}_t = o_p(T^{-1/2})$. 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. We will nonetheless include $d_0 = 1/2$ in the Monte Carlo exercise in section 4. Here we find that the finite sample properties of $LM(\widehat{\tau})$ for $d_0 = 0.5$ do not appear inconsistent with Theorem 1 also being valid for $d_0 = 0.5$.

Remark 11. In parallel with the discussion in Remark 2 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 12. 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

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

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 B, 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 13. The large sample results in Theorem 1 are in sharp contrast to 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. In particular, asymptotic null critical values for these tests differ between the no trend break and trend break cases, and in the latter case also depend on the true location(s) of the trend break(s). 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 14. Consider the case where an observed time seriers x_t satisfies the DGP

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

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.

Remark 15. 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. 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 are relate to a nominal asymptotic 0.05 level using the relevant critical value from the χ^2_1 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 cover the case of $d_0 = 0.5$. 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 $\hat{\omega}^2 := \pi^2/6 - \left(1 - \hat{a}^2\right)\left(\ln\left(1 - \hat{a}\right)\right)^2/\hat{a}^2$.[1]

Empirical size results are reported in Table 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 $d_0 = 0$.

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

[1] 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 $\overline{u}_1 = 0$.

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 $LM(\hat{\tau})$ 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 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 $LM(\hat{\tau})$, 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 $LM(\hat{\tau})$ 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 $LM(\hat{\tau})$ 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 $LM(\hat{\tau})$, 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 $LM(\hat{\tau})$ are based on a richer deterministic specification than $\overline{LM}$, and correspondingly appear to show the greater degree of asymmetry. Comparing $LM(\tau^*)$ and $LM(\hat{\tau})$, 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, $\hat{\tau}$ 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 $LM(\hat{\tau})$ 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 $d = 1.25$ are shown in Figures 4, 5 and 6 respectively. Interestingly, when $d_0 = 0.75$ the asymmetry of the $LM(\tau^*)$ and $LM(\hat{\tau})$ 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.[2]

5 Conclusions

In this paper 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. Unlike conventional unit root and stationarity tests, these results hold regardless of whether a trend break actually occurs in the data or not. 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.

References

Andrews, D.W.K., 1993. Generic uniform convergence, *Econometric Theory* 8, 241-257.

[2]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.

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.

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.

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, 427–31.

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

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

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.

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.

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, 1999. Alternative forms of fractional Brownian motion, *Journal of Statistical Planning and Inference* 80, 111-122.

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

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

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

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

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

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.

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 functional central limit theorem [FCLT], if $T^{-1/2}\sum_{t=1}^{\lfloor \tau T \rfloor}\varepsilon_t \Rightarrow \sigma_\varepsilon W(\tau)$, where "$\Rightarrow$" indicates convergence in the uniform metric, and $W(\tau)$ is a standard Brownian motion, we will write $\sum_{t=1}^{\lfloor \tau T \rfloor}\varepsilon_t = O_p\left(T^{1/2}\right)$. The operator $-\ln(\Delta)$ that appears in the appendix admits the expansion $-\ln(\Delta) = \sum_{j=1}^{\infty} j^{-1}L^j$ (as in a Taylor expansion for $-\ln(1-x)$ around $x=1$), and for a generic series ξ_t we introduce the operator $\{-\ln(\Delta)\}_+$ so that $\{-\ln(\Delta)\}_+\xi_t := -\ln(\Delta)\{\xi_t \mathbb{I}(t>0)\}$ and therefore $\{-\ln(\Delta)\}_+\xi_t = \sum_{j=1}^{\infty} j^{-1}\xi_{t-j}$. 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. Finally, we prove Lemma 1 by putting these three lemmas together.

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,

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

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 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 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 $\sum_{s\geq 0}|E(u_t u_{t+s})| < \infty$ 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 $\delta \geq 0$, (A.4) is given in Bai (1994), Proposition 4, when $\delta = 0$, or in Lavielle and Moulines (2000), Theorem 8. Lavielle and Moulines (2000) do not explicitly consider $\delta < 0$, 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

$$\widehat{\beta}_2(\widehat{\tau}) - \widehat{\beta}_2(\tau^*) = \left(\frac{\lfloor \tau^* T\rfloor - \lfloor \widehat{\tau} T\rfloor}{\lfloor \tau^* T\rfloor \lfloor \widehat{\tau} T\rfloor}\sum_{t=1}^{\lfloor \tau^* T\rfloor} u_t - \frac{1}{\lfloor \widehat{\tau} T\rfloor}\sum_{t=1+\lfloor \widehat{\tau} T\rfloor}^{\lfloor \tau^* T\rfloor} u_t\right)\mathbb{I}(\widehat{\tau} \leq \tau^*) \tag{A.5}$$

$$+\left(\frac{\lfloor \tau^* T\rfloor - \lfloor \widehat{\tau} T\rfloor}{\lfloor \tau^* T\rfloor \lfloor \widehat{\tau} T\rfloor}\sum_{t=1}^{\lfloor \tau^* T\rfloor} u_t + \frac{1}{\lfloor \widehat{\tau} T\rfloor}\sum_{t=1+\lfloor \tau^* T\rfloor}^{\lfloor \widehat{\tau} T\rfloor} u_t + \beta_3 \frac{\lfloor \tau^* T\rfloor - \lfloor \widehat{\tau} T\rfloor}{\lfloor \widehat{\tau} T\rfloor}\right)\mathbb{I}(\widehat{\tau} > \tau^*). \tag{A.6}$$

Because $\lfloor \tau^* T\rfloor - \lfloor \widehat{\tau} T\rfloor = O_p(1)$ and $\sum_{t=1}^{\lfloor \tau^* T\rfloor} u_t = O_p(T^{1/2+\delta})$, the first term on the right hand side of (A.5) is $O_p(1 \times T^{-2} \times T^{1/2+\delta}) = O_p(T^{-3/2+\delta}) = o_p(T^{-1/2+\delta})$. As for the second term, it follows from Equation (8) of Lavielle and Moulines (2000) that for $\varepsilon > 0$,

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

so that

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

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 (A.5) is such that

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

Proceeding in the same way, we can also show that the first two terms in (A.6) 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,

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

then $K_T(d)\left(\widehat{\beta}(\widehat{\tau}) - \widehat{\beta}(\tau^*)\right) = o_p(1)$ and $K_T(d)\left(\widehat{\beta}(\tau^*) - \beta\right) = O_p(1)$, which establishes (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),

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

for $m \in (-\alpha - 1/2, -\alpha + 1/2)$. Noting that, given the range of α, it is always possible to choose $m \in (-1/2, 1/2)$, so that so $E\left(\Delta_+^m \eta_t\right)^2 = O(1)$ it therefore follows that

$$\sum_{j=1}^{t-1}\left(\Delta^m\left\{\eta_{t-j}\mathbb{I}(t>j)\right\}\right)^2 = \sum_{j=1}^{t-1}\left(\Delta_+^m \eta_t\right)^2 = O_p(t).$$

We then rewrite

$$
\begin{aligned}
& T^{-(1/2+\alpha)}\left(\ln\left(T\right)\right)^{-r}\left|\sum_{t=1}^{\lfloor\tau T\rfloor}\left(\left(\left(\ln\left(\Delta\right)\right)^{r}\Delta^{-(\alpha+\theta_T)}\right)_{+}\eta_t-\left(\left(\ln\left(\Delta\right)\right)^{r}\Delta^{-\alpha}\right)_{+}\eta_t\right)\right| \\
\leq\ & \left|\theta_T\right|T^{-(1/2+\alpha)}\left(\ln\left(T\right)\right)^{-r}\left|\sum_{t=1}^{\lfloor\tau T\rfloor}\left(\left(\ln\left(\Delta\right)\right)^{r+1}\Delta^{-\alpha}\right)_{+}\eta_t\right| && \text{(A.7)} \\
& +\frac{1}{2}\theta_T^2T^{-(1/2+\alpha)}\left(\ln\left(T\right)\right)^{-r}\left|\sum_{t=1}^{\lfloor\tau T\rfloor}\left(\left(\ln\left(\Delta\right)\right)^{r+2}\Delta^{-\alpha}\right)_{+}\eta_t\right| && \text{(A.8)} \\
& +O_p\left(T^{-(1/2+\alpha)}\sum_{t=1}^{T}t^{1/2}\left|\theta_T\right|^3\right). && \text{(A.9)}
\end{aligned}
$$

From Marinucci and Robinson (2000) and the rate for θ_T, the term in (A.7) is $O_p\left(T^{-1/2}\ln\left(T\right)\right)=o_p\left(1\right)$, and the term in (A.8) can be treated in the same way. The remainder (A.9) is $O_p\left(T^{-(1/2+\alpha)}\right)=o_p\left(1\right)$.

Proof of Lemma 1:

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

• Under H_c, from Lemma C1, setting $r=0$ and $\alpha=\delta$, we can conclude that, when $\beta_3=0$, the result in (A.1) still holds. 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, this limit coincides with that of $T^{-(1/2+\delta)}\sum_{t=1}^{\lfloor\tau T\rfloor}\Delta_{+}^{-(\delta+\theta_T)}\eta_t$ under H_c. Therefore, (A.2) is also valid under H_c for Model A. For Model B, (A.3) holds for any $\delta\in(-1/2,1/2)$. For T sufficiently large, $(\delta+\theta_T)\in(-1/2,1/2)$ still holds, so (A.3) still holds. Consequently, (A.4) 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.

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

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

$$
\begin{aligned}
T^{-1/2}\sum_{t=1}^{T}\widehat{\varepsilon}_t\left(\tau\right)\widehat{v}_t\left(\tau\right)-T^{-1/2}\sum_{t=1}^{T}\widehat{\varepsilon}_t\widehat{v}_t &= o_p\left(1\right) && \text{(A.10)} \\
\widehat{s}^2\left(\tau\right)-\widehat{s}^2 &= o_p\left(1\right) && \text{(A.11)} \\
\widehat{\omega}^2\left(\tau\right)-\widehat{\omega}^2 &= o_p\left(1\right). && \text{(A.12)}
\end{aligned}
$$

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

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 $\widehat{\omega}^2(\widehat{\tau}) - \widehat{\omega}^2 = o_p(1)$.

Proof of Lemma A2:

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

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

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

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

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

We may then write the loss functions in (3.1) and (3.7) as $\sum_{t=1}^{T}(\widehat{\varepsilon}_t(\psi))^2$ and $\sum_{t=1}^{T}(\widehat{\varepsilon}_t(\psi;\tau))^2$, 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 $T^{-1}\sum_{t=1}^{T}(\widehat{\varepsilon}_t(\psi))^2 \xrightarrow{p} E(g(L;\psi)\eta_t)^2$, 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 $T^{-1}\sum_{t=1}^{T}(\widehat{\varepsilon}_t(\psi))^2$ to the limit, and stochastic equicontinuity; see page 224 of Andrews (1992). Sufficient conditions for stochastic equiconuity 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 used in Theorem A1 of Andrews (1993), to establish part (i) of the lemma we need to verify that $T^{-1}\left(\sum_{t=1}^{T}(\widehat{\varepsilon}_t(\psi;\tau))^2 - \sum_{t=1}^{T}(\widehat{\varepsilon}_t(\psi))^2\right) = o_p(1)$ 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 $\widehat{\varepsilon}_t(\psi;\tau)$, we have that when $d_0 < 0.5$,

$$\widehat{\varepsilon}_t(\psi;\tau) = g(L;\psi)\Delta_+^{\delta}\left(y_t - z_t(\tau)'\widehat{\beta}(\tau)\right)$$
$$= g(L;\psi)\Delta_+^{\delta}u_t + g(L;\psi)\Delta_+^{\delta}z_t(\tau)'\left(\beta - \widehat{\beta}(\tau)\right) \quad \text{(A.13)}$$
$$= \widehat{\varepsilon}_t(\psi) + g(L;\psi)\Delta_+^{\delta}z_t(\tau)'\left(\beta - \widehat{\beta}(\tau)\right) \quad \text{(A.14)}$$

and that

$$\sum_{t=1}^{T}\left(\widehat{\varepsilon}_{t}\left(\psi;\tau\right)\right)^{2}-\sum_{t=1}^{T}\left(\widehat{\varepsilon}_{t}\left(\psi\right)\right)^{2} = \sum_{t=1}^{T}\left(g\left(L;\psi\right)\Delta_{+}^{\delta}z_{t}\left(\tau\right)'\left(\beta-\widehat{\beta}\left(\tau\right)\right)\right)^{2} \tag{A.15}$$

$$+2\sum_{t=1}^{T}\left(\widehat{\varepsilon}_{t}\left(\psi\right)\right)\left(g\left(L;\psi\right)\Delta_{+}^{\delta}z_{t}\left(\tau\right)'\left(\beta-\widehat{\beta}\left(\tau\right)\right)\right). \tag{A.16}$$

Where $d_0 > 0.5$, imposing $\widehat{u}_1(\tau) = 0$ adds the remainder term $-g(L;\psi)\Delta_t^{(-\delta)}u_1$ to (A.13) and (A.14), and notice that as $e_t = 0$ if $t < 0$, then $u_1 = e_1 = \eta_1$, and so we can write this remainder term as

$$-\Delta_{t}^{(-\delta)}\widehat{\varepsilon}_{1}\left(\psi\right). \tag{A.17}$$

Consider Model A first. Using $(a+b)^2 \leq 2a^2 + 2b^2$, the left hand side of (A.15) 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 $g(1;\psi)^2 < C$. Then, using the fact that $\sum_{t=1}^{T}\mu_{3,t}(\tau)^2 = \sum_{t=1+\lfloor\tau T\rfloor}^{T}\mu_{3,t}(\tau)^2 \leq \sum_{t=1}^{T}\mu_{2,t}^2$, the expression above is seen to be of $O_p(1)$ using Lemma 1 of Robinson (2005) and Lemma A1. The term in (A.16) is $O_p\left(T^{1/2}\right)$ by the Cauchy-Schwarz inequality.

Next we consider Model B. Here the left hand side of (A.15) 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 (A.16) is $O_p\left(T^{1/2}\right)$. For Model B we also have to account for the additional remainder term in (A.17), so that we also need to analyse

$$\sum_{t=1}^{T}\left(\Delta_{t}^{(-\delta)}\right)^{2}\left(\widehat{\varepsilon}_{1}\left(\psi\right)\right)^{2}-2\sum_{t=1}^{T}\Delta_{t}^{(-\delta)}\widehat{\varepsilon}_{1}\left(\psi\right)\widehat{\varepsilon}_{t}\left(\psi\right)-2\sum_{t=1}^{T}\Delta_{t}^{(-\delta)}\widehat{\varepsilon}_{1}\left(\psi\right)g\left(L;\psi\right)\Delta_{+}^{\delta}z_{t}\left(\tau\right)'\left(\beta-\widehat{\beta}\left(\tau\right)\right). \tag{A.18}$$

Noting that $(\widehat{\varepsilon}_1(\psi))^2 = O_p(1)$, 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 $\sum_{t=1}^{T}(\Delta_t^{(-\delta)})^2(\widehat{\varepsilon}_1(\psi))^2 = O_p(\sum_{t=1}^{T}t^{2(-\delta-1)}) = O_p(1)$. 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 $O_p(1)$ 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 (A.18) is $O_p(1)$, so that the whole expression in (A.18) is of $o_p\left(T^{1/2}\right)$.

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 $\widehat{\psi}(\tau) - \widehat{\psi} = o_p(1)$, 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) yield

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

respectively, where

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

Recalling (A.14), we have that

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

As with the treatment of (A.13) and (A.14) above, these expressions should properly be augmented by additional remainder terms under Model B. However, proceeding as in the derivation of (A.18) 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

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

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 $\widehat{\psi}$,

$$\sum_{t=1}^{T} \widehat{\varepsilon}_t(\psi) \left.\frac{\partial \widehat{\varepsilon}_t(\psi)}{\partial \psi_i}\right|_{\psi=\psi^*} + \left[D\left(\widetilde{\psi}^i\right)\right]_i \left(\widehat{\psi}_i - \psi_i^*\right) = 0 \tag{A.19}$$

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 (A.19) for each i and multiplying by $T^{1/2}$, we get

$$T^{-1/2}\sum_{t=1}^{T} \widehat{\varepsilon}_t(\psi) \left.\frac{\partial \widehat{\varepsilon}_t(\psi)}{\partial \psi}\right|_{\psi=\psi^*} + \widetilde{D}\left(\widehat{\psi}\right) T^{1/2}\left(\widehat{\psi} - \psi^*\right) = 0. \tag{A.20}$$

Notice that $\widetilde{D}(\widehat{\psi}) \to_p \Phi\sigma_\varepsilon^2$; see, for example, Nielsen (2004), part (iii) 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 Theorem 4.1. This therefore implies that $T^{1/2}(\widehat{\psi}-\psi^*)=O_p(1)$ (indeed it is clear from part (ii) 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 (A.20) for the feasible estimate $\widehat{\psi}(\tau)$, from which we can obtain a formula for $\widehat{\psi}(\tau)$. Then, define

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

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 $\widehat{\psi}(\tau)$,

$$
\sum_{t=1}^{T}\widehat{\varepsilon}_t(\psi;\tau)\left.\frac{\partial\widehat{\varepsilon}_t(\psi;\tau)}{\partial\psi_i}\right|_{\psi=\psi^*}+\left[D\left(\widetilde{\psi}^i(\tau);\tau\right)\right]_i\left(\widehat{\psi}_i(\tau)-\psi_i^*\right)=0 \tag{A.21}
$$

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

$$
T^{-1/2}\sum_{t=1}^{T}\widehat{\varepsilon}_t(\psi;\tau)\left.\frac{\partial\widehat{\varepsilon}_t(\psi;\tau)}{\partial\psi}\right|_{\psi=\psi^*}+\widetilde{D}\left(\widehat{\psi}(\tau);\tau\right)T^{1/2}\left(\widehat{\psi}(\tau)-\psi^*\right)=0. \tag{A.22}
$$

In order 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:

$$
\sup_{\tau}\left\|\widetilde{D}\left(\widehat{\psi}\right)-\widetilde{D}\left(\widehat{\psi}(\tau);\tau\right)\right\|\xrightarrow{p}0. \tag{A.23}
$$

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 (A.23) and using the triangle inequality, the expression in (A.23) is bounded by $\left\|\widetilde{D}\left(\widehat{\psi}\right)-\Phi\sigma_\varepsilon^2\right\|+\sup_\tau\left\|\widetilde{D}\left(\widehat{\psi}\left(\tau\right);\tau\right)-\Phi\sigma_\varepsilon^2\right\|$, recalling that $\widetilde{D}\left(\widehat{\psi}\right)\to_p\Phi\sigma_\varepsilon^2$ so that $\left\|\widetilde{D}\left(\widehat{\psi}\right)-\Phi\sigma_\varepsilon^2\right\|=o_p\left(1\right)$.

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

$$\frac{1}{T}\sum_{t=1}^T\widehat{\varepsilon}_t\left(\widetilde{\psi}\left(\tau\right)\right)^2-\frac{1}{T}\sum_{t=1}^T\widehat{\varepsilon}_t\left(\psi^*\right)^2=o_p\left(1\right) \quad (A.24)$$

$$\frac{1}{T}\sum_{t=1}^T\left.\frac{\partial\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi_i}\frac{\partial\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi_j}\right|_{\psi=\widetilde{\psi}(\tau)}-\frac{1}{T}\sum_{t=1}^T\left.\frac{\partial\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi_i}\frac{\partial\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi_j}\right|_{\psi=\psi^*}=o_p\left(1\right) \quad (A.25)$$

$$\frac{1}{T}\sum_{t=1}^T\left.\frac{\partial^2\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi_i\partial\psi_j}\right|_{\psi=\widetilde{\psi}(\tau)}-\frac{1}{T}\sum_{t=1}^T\left.\frac{\partial^2\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi_i\partial\psi_j}\right|_{\psi=\psi^*}=o_p\left(1\right). \quad (A.26)$$

Because $\eta_t=\frac{b(L;\psi^*)}{a(L;\psi^*)}\varepsilon_t$ is a stationary and invertible ARMA process, then $g\left(L;\psi\right)\eta_t=\frac{a(L;\psi)}{b(L;\psi)}\frac{b(L;\psi^*)}{a(L;\psi^*)}\varepsilon_t$ is also an ARMA process. For ψ_i, the *i-th* element of ψ, $\frac{\partial}{\partial\psi_i}g\left(L;\psi\right)\eta_t$ and $\frac{\partial^2}{\partial\psi_i\partial\psi_j}g\left(L;\psi\right)\eta_t$ are also ARMA processes, and so $\left|\frac{\partial}{\partial\psi_i}g\left(1;\psi\right)\right|<C$ and $\left|\frac{\partial^2}{\partial\psi_i\partial\psi_j}g\left(1;\psi\right)\right|<C$ uniformly in ψ. Proceeding as in Bai (1993), we illustrate (A.24)-(A.26) for the ARMA(1,1) case, $\left(1-\psi_1^*L\right)\eta_t=\left(1+\psi_2^*L\right)\varepsilon_t$.

Consider first (A.24). Because $\widehat{\varepsilon}_t\left(\psi^*\right)=\varepsilon_t$, we rewrite

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

As in Equation (3) of Bai (1993),

$$\begin{aligned}\widehat{\varepsilon}_t\left(\widetilde{\psi}\left(\tau\right)\right)-\varepsilon_t &= \left(-1\right)^{t-1}\left(\widetilde{\psi}_2\left(\tau\right)\right)^t\varepsilon_0-\left(\widetilde{\psi}_1\left(\tau\right)-\psi_1^*\right)\sum_{j=0}^{t-1}\left(-1\right)^j\left(\widetilde{\psi}_2\left(\tau\right)\right)^j\eta_{t-j-1}\\ &\quad -\left(\widetilde{\psi}_2\left(\tau\right)-\psi_2^*\right)\sum_{j=0}^{t-1}\left(-1\right)^j\left(\widetilde{\psi}_2\left(\tau\right)\right)^j\varepsilon_{t-j-1}.\end{aligned}$$

The compactness of Θ means that there exists $\overline{c}<1-\varepsilon$, where $\varepsilon>0$ depends on Θ, such that $\sup\left|\psi_2\right|<\overline{c}<1$, and so

$$\begin{aligned}\left|\widehat{\varepsilon}_t\left(\widetilde{\psi}\left(\tau\right)\right)-\varepsilon_t\right| &\leq \overline{c}^t\left|\varepsilon_0\right|+\left(\widetilde{\psi}_1\left(\tau\right)-\psi_1^*\right)\sum_{j=0}^{t-1}\overline{c}^j\left|\eta_{t-j-1}\right|\\ &\quad +\left(\widetilde{\psi}_2\left(\tau\right)-\psi_2^*\right)\sum_{j=0}^{t-1}\overline{c}^j\left|\varepsilon_{t-j-1}\right|.\end{aligned}$$

and $\sum_{j=0}^{t-1}\overline{c}^j\left|\eta_{t-j-1}\right|=O_p\left(1\right)$ because $E\left(\left|\eta_{t-j-1}\right|\right)<C$ and $\sum_{j=0}^{t-1}\overline{c}^j<\sum_{j=0}^{\infty}\overline{c}^j<C$. In the same way, $\sum_{j=0}^{t-1}\overline{c}^j\left|\varepsilon_{t-j-1}\right|=O_p\left(1\right)$. Rewriting $sk_t:=\sum_{j=0}^{t-1}\overline{c}^j\left|\eta_{t-j-1}\right|$, $sl_t:=\sum_{j=0}^{t-1}\overline{c}^j\left|\varepsilon_{t-j-1}\right|$, then $sk_t=O_p\left(1\right)$ and $sl_t=O_p\left(1\right)$, and $sk_t^2=O_p\left(1\right)$, $sl_t^2=O_p\left(1\right)$. Then,

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^T\left(\widehat{\varepsilon}_t\left(\widetilde{\psi}\left(\tau\right)\right)-\varepsilon_t\right)^2 &\leq \frac{C}{T}\sum_{t=1}^T\overline{c}^{2t}\varepsilon_0^2+C\left(\widetilde{\psi}_1\left(\tau\right)-\psi_1^*\right)^2\frac{1}{T}\sum_{t=1}^T\left(sk_t\right)^2\\ &\quad +C\left(\widetilde{\psi}_2\left(\tau\right)-\psi_2^*\right)^2\frac{1}{T}\sum_{t=1}^T\left(sl_t\right)^2.\end{aligned}$$

The first term in the foregoing bound is $\frac{C}{T}\sum_{t=1}^{T}\overline{c}^{2t}\varepsilon_0^2 = O_p\left(\frac{1}{T}\right) = o_p(1)$. As for the second term, $\frac{1}{T}\sum_{t=1}^{T}(sk_t)^2 = O_p(1)$ and, using the rate for $\left(\widetilde{\psi}_1(\tau)-\psi_1^*\right)^2$, this is seen to be of $o_p(1)$. The last term follows in the same way. Therefore,

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

and $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t\left(\widehat{\varepsilon}_t\left(\widetilde{\psi}(\tau)\right)-\varepsilon_t\right) = o_p(1)$ by the Cauchy-Schwarz inequality, which concludes the demonstration of (A.24) 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 (A.25). Proceeding in the same way as for (A.24), it is sufficient to show that the following results hold:

$$\frac{1}{T}\sum_{t=1}^{T}\left(\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_i}\right|_{\psi=\widetilde{\psi}(\tau)}-\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_i}\right|_{\psi=\psi^*}\right)^2 = o_p(1) \tag{A.27}$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\left(\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_i}\right|_{\psi=\psi^*}\right)^2 = O_p(1)\,. \tag{A.28}$$

Consider first the result in (A.27). Again we illustrate this in the ARMA(1,1) case, noting that these results hold for the more general for the ARMA(p,q) case. In the ARMA(1,1) case, considering $\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_2}$ first,

$$\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_2} = -\widehat{\varepsilon}_{t-1}(\psi)-\psi_2\frac{\partial\widehat{\varepsilon}_{t-1}(\psi)}{\partial\psi_2}$$

and we observe that

$$\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_2}\right|_{\psi=\psi^*} = -\varepsilon_{t-1}-\psi_2^*\left.\frac{\partial\widehat{\varepsilon}_{t-1}(\psi)}{\partial\psi_2}\right|_{\psi=\psi^*} = -\sum_{j=0}^{t-1}\left(-\psi_2^*\right)^j\varepsilon_{t-j-1} \tag{A.29}$$

is a AR(1). Taking differences,

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

and, iterating,

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

so that

$$\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}\left(\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_2}\right|_{\psi=\widetilde{\psi}(\tau)}-\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_2}\right|_{\psi=\psi^*}\right)^2 \\
\leq\ & \frac{2}{T}\sum_{t=1}^{T}\left(\sum_{j=0}^{t-1}\left(-\widetilde{\psi}_2(\tau)\right)^j\left(\widehat{\varepsilon}_{t-j-1}\left(\widetilde{\psi}(\tau)\right)-\varepsilon_{t-j-1}\right)\right)^2 \qquad \text{(A.30)} \\
& +2\left(\widetilde{\psi}_2(\tau)-\psi_2^*\right)^2\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{j=0}^{t-1}\left(-\widetilde{\psi}_2(\tau)\right)^j\left.\frac{\partial\widehat{\varepsilon}_{t-j-1}(\psi)}{\partial\psi_2}\right|_{\psi=\psi^*}\right)^2. \qquad \text{(A.31)}
\end{aligned}$$

Consider first the term in (A.30). Using the bound for $\left|\widehat{\varepsilon}_t\left(\widetilde{\psi}(\tau)\right)-\varepsilon_t\right|$, this is again bounded by

$$\begin{aligned}
\leq\ & \frac{C}{T}\sum_{t=1}^{T}\left(\sum_{j=0}^{t-1}\overline{c}^j\ \overline{c}^{t-j}\left|\varepsilon_0\right|\right)^2 \qquad \text{(A.32)} \\
& +\frac{C}{T}\sum_{t=1}^{T}\left(\sum_{j=0}^{t-1}\overline{c}^j\left(\widetilde{\psi}_1(\tau)-\psi_1^*\right)sk_{t-j}\right)^2 \qquad \text{(A.33)} \\
& +\frac{C}{T}\sum_{t=1}^{T}\left(\sum_{j=0}^{t-1}\overline{c}^j\left(\widetilde{\psi}_2(\tau)-\psi_2^*\right)sl_{t-j}\right)^2. \qquad \text{(A.34)}
\end{aligned}$$

The term in (A.33) is bounded by $C\left(\widetilde{\psi}_1(\tau)-\psi_1^*\right)^2\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{j=0}^{t-1}\overline{c}^j sk_{t-j}\right)^2$, where it is recalled that $\left|\widetilde{\psi}_1(\tau)-\psi_1^*\right|=o_p(1)$. Using $sk_t=O_p(1)$, then $\left(\sum_{j=0}^{t-1}\overline{c}^j sk_{t-j}\right)^2=O_p(1)$ and it follows that (A.33) is of $o_p(1)$. In the same way, using $sl_t=O_p(1)$ we establish $\left(\sum_{j=0}^{t-1}\overline{c}^j sl_{t-j}\right)^2=O_p(1)$ and then, recalling that $\left|\widetilde{\psi}_2(\tau)-\psi_2^*\right|=o_p(1)$, we conclude that (A.34) is also of $o_p(1)$. Finally, (A.32) has order

$$\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{j=0}^{t-1}\overline{c}^j\overline{c}^{t-j}\right)^2\leq\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{j=0}^{\lfloor t/2\rfloor}\overline{c}^j\overline{c}^{t-j}+\sum_{j=\lfloor t/2\rfloor+1}^{t}\overline{c}^j\overline{c}^{t-j}\right)^2 \\
\leq\ & C\frac{1}{T}\sum_{t=1}^{T}\left(\overline{c}^{t/2}\sum_{j=0}^{\lfloor t/2\rfloor}\overline{c}^j+\overline{c}^{t/2}\sum_{j=\lfloor t/2\rfloor+1}^{t}\overline{c}^{t-j}\right)^2\leq\frac{C}{T}\sum_{t=1}^{T}\left(\overline{c}^{t/2}\right)^2\leq\frac{C}{T}.
\end{aligned}$$

Next consider the term in (A.31). Recalling (A.29), (A.31) is

$$\begin{aligned}
& \left(\widetilde{\psi}_2(\tau)-\psi^*\right)^2\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{j=0}^{t-1}\left(-\widetilde{\psi}_2(\tau)\right)^j\sum_{k=0}^{t-j-1}\left(-\psi_2^*\right)^k\varepsilon_{t-k-1}\right)^2 \\
\leq\ & \left(\widetilde{\psi}_2(\tau)-\psi^*\right)^2\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{j=0}^{t-1}\overline{c}^j\sum_{k=0}^{t-j-1}\overline{c}^k\left|\varepsilon_{t-k-1}\right|\right)^2 \\
\leq\ & \left(\widetilde{\psi}_2(\tau)-\psi^*\right)^2\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{j=0}^{t-1}\overline{c}^j sl_{t-j}\right)^2
\end{aligned}$$

which is seen to be of $o_p(1)$ using the bound (A.34). These results together establish the result in (A.27).

Consider next the result in (A.28). Again using (A.29), $\left|\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_2}\right|_{\psi=\psi^*}\right|\leq\sum_{j=0}^{t-1}\overline{c}^j\left|\varepsilon_{t-j-1}\right|=sl_t$ and the fact that $sl_t^2=O_p(1)$, it also follows that $T^{-1}\sum_{t=1}^{T}(\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_2}\right|_{\psi=\psi^*})^2=O_p(1)$. This establishes (A.28) and therefore completes the proof of the result in (A.25) for $\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_2}\right|_{\psi=\psi^*}$.

The proofs of (A.27) and (A.28) for $\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi_1}\right|_{\psi=\psi^*}$ follow in the same way. The result in (A.26) can be obtain in a similar fashion and is omitted in the interest of brevity.

Continuing, we next need to show that

$$\frac{1}{T}\sum_{t=1}^{T}\widehat{\varepsilon}_t\left(\psi\right)\frac{\partial^2\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi_i\partial\psi_j}\bigg|_{\psi=\widetilde{\psi}(\tau)}-\frac{1}{T}\sum_{t=1}^{T}\widehat{\varepsilon}_t\left(\psi\right)\frac{\partial^2\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi_i\partial\psi_j}\bigg|_{\psi=\psi^*}=o_p\left(1\right). \tag{A.35}$$

The right hand side of (A.35) can be written as

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

in which each term can be seen to be of $o_p\left(1\right)$, using the limits for (A.24), (A.26) and the Cauchy-Schwarz inequality.

We next move to an analysis of the contribution of the terms sa_t, $(\widetilde{\psi}\left(\tau\right))$, ..., $sh_t(\widetilde{\psi}(\tau);\tau)$ to (A.23). Recalling that $T^{-1}\sum_{t=1}^{T}(sa_t(\psi^*)+se_t(\psi^*))\rightarrow_p\Phi\sigma_\varepsilon^2$, using (A.25), (A.35) and $\|\widetilde{\psi}\left(\tau\right)-\psi^*\|\leq\|\widehat{\psi}\left(\tau\right)-\psi^*\|$, it also holds that $T^{-1}\sum_{t=1}^{T}sa_t\left(\widetilde{\psi}\left(\tau\right)\right)+se_t\left(\widetilde{\psi}\left(\tau\right)\right)\rightarrow_p\Phi\sigma_\varepsilon^2$. Next, $T^{-1}\sum_{t=1}^{T}sd_t(\widetilde{\psi}\left(\tau\right);\tau)=o_p\left(1\right)$ and $T^{-1}\sum_{t=1}^{T}sh_t(\widetilde{\psi}(\tau);\tau)=o_p\left(1\right)$, proceeding as in the discussion of (A.15). 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 as in the discussion of (A.16). This completes the proof of (A.23).

For the next step of the proof, equating the left hand sides of the two expansions in (A.22) and (A.20) and re-arranging yields,

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

Noting that $T^{-1/2}\sum_{t=1}^{T}\widehat{\varepsilon}_t\left(\psi\right)\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi}\Big|_{\psi=\psi^*}=O_p\left(1\right)$ 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

$$T^{-1/2}\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\psi;\tau\right)\frac{\partial\widehat{\varepsilon}_t\left(\psi;\tau\right)}{\partial\psi}-\widehat{\varepsilon}_t\left(\psi\right)\frac{\partial\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi}\right)\bigg|_{\psi=\psi^*} \tag{A.36}$$

is of $o_p\left(1\right)$. To do so, first re-write (A.36) as

$$T^{-1/2}\sum_{t=1}^{T}\left(g\left(L;\psi\right)\Delta_+^\delta z_t\left(\tau\right)'\left(\beta-\widehat{\beta}\left(\tau\right)\right)\right)\frac{\partial\widehat{\varepsilon}_t\left(\psi\right)}{\partial\psi}\bigg|_{\psi=\psi^*} \tag{A.37}$$

$$+T^{-1/2}\sum_{t=1}^{T}\widehat{\varepsilon}_t\left(\psi\right)\frac{\partial}{\partial\psi}\left(g\left(L;\psi\right)\Delta_+^\delta z_t\left(\tau\right)'\left(\beta-\widehat{\beta}\left(\tau\right)\right)\right)\bigg|_{\psi=\psi^*} \tag{A.38}$$

$$+T^{-1/2}\sum_{t=1}^{T}\left(g\left(L\ \psi\right)\Delta_+^\delta z_t\left(\tau\right)'\left(\beta-\widehat{\beta}\left(\tau\right)\right)\right)\frac{\partial}{\partial\psi}\left(g\left(L;\psi\right)\Delta_+^\delta z_t\left(\tau\right)'\left(\beta-\widehat{\beta}\left(\tau\right)\right)\right)\bigg|_{\psi=\psi^*} \tag{A.39}$$

In view of Lemma 3 of Robinson (2005), the order of (A.39) is the same as the order of

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

Proceeding as in the discussion of (A.15), 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 (A.37), using summation by parts the absolute value of this is bounded by

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

and, in view of Lemma 3 of Robinson (2005), this bound has the same order of

$$\leq T^{-1/2}\sum_{t=1}^{T-1}\left|\left(\Delta_{+}^{\delta}z_{t+1}(\tau)-\Delta_{+}^{\delta}z_t(\tau)\right)\right|'\left|\left(\beta-\widehat{\beta}(\tau)\right)\right|\left|\sum_{s=1}^{t}\frac{\partial\widehat{\varepsilon}_s(\psi)}{\partial\psi}\right| \tag{A.40}$$

$$+T^{-1/2}\left|\left(\Delta_{+}^{\delta}z_T(\tau)\right)\right|'\left|\beta-\widehat{\beta}(\tau)\right|\left|\sum_{t=1}^{T}\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi}\right|. \tag{A.41}$$

The term in (A.40) 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,

$$\begin{aligned}&\sum_{t=1}^{T-1}\left|\left(\Delta^{\delta}z_{t+1}(\tau)-\Delta^{\delta}z_t(\tau)\right)'\left(\beta-\widehat{\beta}(\tau)\right)\right|\\ \leq\ & \sum_{t=1}^{T-1}\left|\Delta\mu_{1,t+1}\right|\left|\beta_1-\widehat{\beta}_1(\tau)\right|+\sum_{t=1}^{T-1}\left|\Delta\mu_{2,t+1}\right|\left|\beta_2-\widehat{\beta}_2(\tau)\right|+\sum_{t=1}^{T-1}\left|\Delta\mu_{3,t+1}(\tau)\right|\left|\widehat{\beta}_3(\tau)\right|.\end{aligned} \tag{A.42}$$

If $\delta>0$, the terms in (A.42) 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 (A.40) is of order $o_p\left(T^{-1/2}\times 1\times T^{1/2}\right)$ $=o_p(1)$.[3] The remainder term in (A.41) can be shown to be of order

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

[3]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.

If, on the other hand, $\delta < 0$ then the first term in (A.42) 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 (A.42) are unaffected by the sign of δ, and it is easily verified that (A.41) remains of $O_p\left(T^{-1/2}\right)$ so that both (A.40) and (A.41) are of $O_p\left(T^{-1/2}\right)$.

When model B is used,

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

and 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

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

whereas, when $\delta < 0$,

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

In both cases, it is again easy to show that the remainder, (A.41), is of order $O_p\left(T^{-1/2}\right)$.

Combining the orders established for (A.37) when either Model A or Model is used, it then follows that (A.37) is of $o_p(1)$. By similar arguments as used for (A.37), the term in (A.38) 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.10) first. To that end, re-write

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

Then it can be seen that (A.10) follows if we can show the following:

$$\sum_{t=1}^{T} \widehat{\varepsilon}_t\left(\widehat{\psi}(\tau);\tau\right)\left(\widehat{v}_t\left(\widehat{\psi}(\tau);\tau\right) - \widehat{v}_t\left(\widehat{\psi}\right)\right) = o_p\left(T^{1/2}\right) \tag{A.43}$$

$$\sum_{t=1}^{T} \left(\widehat{\varepsilon}_t\left(\widehat{\psi}(\tau);\tau\right) - \widehat{\varepsilon}_t\left(\widehat{\psi}\right)\right)\widehat{v}_t\left(\widehat{\psi}\right) = o_p\left(T^{1/2}\right). \tag{A.44}$$

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

$$\begin{aligned}\widehat{\varepsilon}_t\left(\widehat{\psi}(\tau)\right)-\widehat{\varepsilon}_t\left(\widehat{\psi}\right) &= \left(\widehat{\psi}(\tau)-\widehat{\psi}\right)'\left.\frac{\partial\widehat{\varepsilon}_t(\psi)}{\partial\psi}\right|_{\psi=\psi^*}\\ &\quad+\frac{1}{2}\left(\widehat{\psi}(\tau)-\widehat{\psi}\right)'\left.\frac{\partial^2\widehat{\varepsilon}_t(\psi)}{\partial\psi\partial\psi'}\right|_{\psi=\widetilde{\psi}}\left(\widehat{\psi}(\tau)-\widehat{\psi}\right) \qquad \text{(A.45)}\end{aligned}$$

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 (A.45) is only correct if ψ is a scalar; otherwise, a row by row espansion should be derived, similarly to (A.20), and then stacked as in (A.22), but this approximation does not affect the results). Consequently, the last term of (A.45) 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 $\delta\in(0,1/2)$,

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

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

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

We first consider (A.43). The left hand side of (A.43) is such that,

$$\begin{aligned}&\sum_{t=1}^{T}\widehat{\varepsilon}_t\left(\widehat{\psi}(\tau);\tau\right)\left(\widehat{v}_t\left(\widehat{\psi}(\tau);\tau\right)-\widehat{v}_t\left(\widehat{\psi}\right)\right)\\ =&\sum_{t=1}^{T}\widehat{\varepsilon}_t\left(\widehat{\psi}(\tau);\tau\right)\left(\widehat{v}_t\left(\widehat{\psi}(\tau)\right)-\widehat{v}_t\left(\widehat{\psi}\right)\right) \qquad \text{(A.46)}\\ &+\sum_{t=1}^{T}\widehat{\varepsilon}_t\left(\widehat{\psi}(\tau);\tau\right)g\left(L;\widehat{\psi}(\tau)\right)\left\{-\ln(\Delta)\Delta^{\delta}\right\}z_t(\tau)'\left(\beta-\widehat{\beta}(\tau)\right). \qquad \text{(A.47)}\end{aligned}$$

The stochastic order of (A.46) is the same as that 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 (A.47),

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

Noting that

$$\begin{aligned}
\sup_{\rho}\left|\sum_{s=1}^{\lfloor\rho T\rfloor}\widehat{\varepsilon}_s\left(\widehat{\psi}(\tau);\tau\right)\right| \;\leq\;& \sup_{\rho}\left|\sum_{s=1}^{\lfloor\rho T\rfloor}\widehat{\varepsilon}_t\left(\widehat{\psi}(\tau)\right)\right| \\
&+\sup_{\rho}\left|\sum_{s=1}^{\lfloor\rho T\rfloor}g\left(L;\widehat{\psi}(\tau)\right)\Delta_{+}^{\delta}z_s(\tau)'\left(\beta-\widehat{\beta}(\tau)\right)\right| \qquad (A.48)
\end{aligned}$$

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 (A.45) 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). As for the term (A.48), using the again the fact that $\Delta_{+}^{\delta_0}u_t$ is ARMA and Lemma 3 of Robinson (2005), as was done in the proof of Lemma A2, this term is seen to have the same stochastic order as

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

so we conclude that $\sup_{\rho}\left|\sum_{s=1}^{\lfloor\rho T\rfloor}\widehat{\varepsilon}_s\left(\widehat{\psi}(\tau);\tau\right)\right|=O_p\left(T^{1/2}\right)$. The term

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

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

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

and then proceeding as in the discussion of (A.40), this is seen to be of $O_p\left((\ln(T))^2T^{-1/2+\delta}\right)$ 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. In all cases the remainder

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

The stated result in (A.43) is therefore established.

The proof for (A.44) is similar, and we discuss it below. The expression in (A.44) can be written as

$$\sum_{t=1}^{T} \widehat{v}_t\left(\widehat{\psi}\right)\left(\widehat{\varepsilon}_t\left(\widehat{\psi}(\tau)\right)-\widehat{\varepsilon}_t\left(\widehat{\psi}\right)\right) \tag{A.49}$$

$$+\sum_{t=1}^{T} \widehat{v}_t\left(\widehat{\psi}\right) g\left(L;\widehat{\psi}(\tau)\right)\Delta_+^{\delta} z_t(\tau)'\left(\beta-\widehat{\beta}(\tau)\right). \tag{A.50}$$

As in the case (A.46), is the same as that of

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

Again the discussion of (A.50) is similar to the discussion of (A.47): we apply summation by parts to (A.50) and discuss the role of the terms $g\left(L;\widehat{\psi}(\tau)\right)\Delta_+^{\delta} z_t(\tau)'\left(\beta-\widehat{\beta}(\tau)\right)$ as in the discussion of (A.47), 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 $\left\|\widetilde{\psi}-\psi\right\| \le \left\|\left(\widehat{\psi}-\psi\right)\right\|$

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

so $\sup_\rho\left|\sum_{t=1}^{\lfloor\rho T\rfloor}\widehat{v}_t\left(\widehat{\psi}\right)\right| = O_p\left(\ln(T)\,T^{1/2}\right)$ again in view of the FCLT in Marinucci and Robinson (2000) and (A.50) is $o_p(1)$. The result in (A.10) is thereby established.

For (A.11),

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

the two terms of which are of $o_p\left(T^{1/2}\right)$ and of $o_p(T)$, respectively, proceeding in the same way as in the discussion of (A.43) and (A.44).

Finally, since κ and Φ are continuous function of ψ, (A.12) follows by an application of Slutzky's Theorem.

Proof of Lemma C2.

We have that,

$$\begin{aligned}
\widehat{\varepsilon}_t(\psi;\widehat{\tau}) &= g(L;\psi)\Delta_+^{\delta}\left(y_t - z_t(\widehat{\tau})'\widehat{\beta}(\widehat{\tau})\right) = g(L;\psi)\Delta_+^{\delta}\left(u_t + z_t(\tau^*)'\beta - z_t(\widehat{\tau})'\widehat{\beta}(\widehat{\tau})\right)\\
&= g(L;\psi)\Delta_+^{\delta}\left(u_t + z_t(\tau^*)'\beta - z_t(\tau^*)'\widehat{\beta}(\widehat{\tau}) + z_t(\tau^*)'\widehat{\beta}(\widehat{\tau}) - z_t(\widehat{\tau})'\widehat{\beta}(\widehat{\tau})\right)\\
&= \widehat{\varepsilon}_t(\psi) + g(L;\psi)\Delta_+^{\delta} z_t(\tau^*)'\left(\beta-\widehat{\beta}(\widehat{\tau})\right) + g(L;\psi)\Delta_+^{\delta}\left(z_t(\tau^*)-z_t(\widehat{\tau})\right)'\widehat{\beta}(\widehat{\tau}).
\end{aligned} \tag{A.51}$$

Therefore,

$$
\begin{aligned}
&\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\psi;\widehat{\tau}\right)\right)^2-\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\psi\right)\right)^2\\
=\ &\sum_{t=1}^{T}\left(g\left(L;\psi\right)\Delta_+^{\delta}z_t\left(\tau^*\right)'\left(\beta-\widehat{\beta}\left(\widehat{\tau}\right)\right)\right)^2+2\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\psi\right)\right)\left(g\left(L;\psi\right)\Delta_+^{\delta}z_t\left(\tau^*\right)'\left(\beta-\widehat{\beta}\left(\widehat{\tau}\right)\right)\right) && \text{(A.52)}\\
&+2\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\psi\right)\right)g\left(L;\psi\right)\Delta_+^{\delta}\left(z_t\left(\tau^*\right)-z_t\left(\widehat{\tau}\right)\right)'\widehat{\beta}\left(\widehat{\tau}\right) && \text{(A.53)}\\
&+2\sum_{t=1}^{T}\left(g\left(L;\psi\right)\Delta_+^{\delta}z_t\left(\tau^*\right)'\left(\beta-\widehat{\beta}\left(\widehat{\tau}\right)\right)\right)\left(g\left(L;\psi\right)\Delta_+^{\delta}\left(z_t\left(\tau^*\right)-z_t\left(\widehat{\tau}\right)\right)'\widehat{\beta}\left(\widehat{\tau}\right)\right) && \text{(A.54)}\\
&+\sum_{t=1}^{T}\left(g\left(L;\psi\right)\Delta_+^{\delta}\left(z_t\left(\tau^*\right)-z_t\left(\widehat{\tau}\right)\right)'\widehat{\beta}\left(\widehat{\tau}\right)\right)^2 && \text{(A.55)}
\end{aligned}
$$

where the two terms in (A.52) are $O_p\left(T^{1/2}\right)$ uniformly in ψ using (3.11) and proceeding as in Lemma A2.

As for (A.55), 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

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

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

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

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

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

while in the context of the second term,

$$
\begin{aligned}
\left(\mu_{3,t}\left(\tau^*\right)-\mu_{3,t}\left(\widehat{\tau}\right)\right)&=\left(\mu_{3,t}\left(\tau^*\right)-\mu_{3,t-1}\left(\tau^*\right)+\mu_{3,t-1}\left(\tau^*\right)-...-\mu_{3,t}\left(\widehat{\tau}\right)\right)\\
\left|\mu_{3,t}\left(\tau^*\right)-\mu_{3,t}\left(\widehat{\tau}\right)\right|&<C\left(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor\right)\left(t-\lfloor\widehat{\tau}T\rfloor\right)^{-\delta} && \text{(A.56)}
\end{aligned}
$$

and

$$
\begin{aligned}
&\sum_{t=1+\lfloor\widehat{\tau}T\rfloor}^{T}\left(\mu_{3,t}\left(\tau^*\right)-\mu_{3,t}\left(\widehat{\tau}\right)\right)^2\leq C\left(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor\right)^2\sum_{t=1+\lfloor\widehat{\tau}T\rfloor}^{T}\left(t-\lfloor\widehat{\tau}T\rfloor\right)^{-2\delta}\\
\leq\ &C\left(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor\right)^2\sum_{t=1}^{T}t^{-2\delta}\leq C\left(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor\right)^2T^{1-2\delta}=O_p\left(T^{(1-3/2+\delta)\times2}T^{1-2\delta}\right)=O_p\left(1\right).
\end{aligned}
$$

When Model B is used,

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

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

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

recalling $-2\delta-2<-1$ as $\delta>-1/2$; $\sum_{t=1+\lfloor\widehat{\tau}T\rfloor}^{T}\left(\mu_{3,t}\left(\tau^{*}\right)-\mu_{3,t}\left(\widehat{\tau}\right)\right)^{2}=O_{p}\left(1\right)$; 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

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

It therefore, follows that

$$\sum_{t=1}^{T}\left(g\left(L;\psi\right)\Delta_{+}^{\delta}\left(z_{t}\left(\tau^{*}\right)-z_{t}\left(\widehat{\tau}\right)\right)'\widehat{\beta}\left(\widehat{\tau}\right)\right)^{2}=O_{p}\left(1\right) \tag{A.57}$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\left(g\left(L;\psi\right)\Delta_{+}^{\delta}\left(z_{t}\left(\tau^{*}\right)-z_{t}\left(\widehat{\tau}\right)\right)'\widehat{\beta}\left(\widehat{\tau}\right)\right)^{2}\xrightarrow{p}0 \tag{A.58}$$

uniformly in ψ, thereby accounting for (A.55). The two remaining cross products in the expansion of $\sum_{t=1}^{T}\left(\widehat{\varepsilon}_{t}\left(\psi;\widehat{\tau}\right)\right)^{2}-\sum_{t=1}^{T}\left(\widehat{\varepsilon}_{t}\left(\psi\right)\right)^{2}$, (A.53) and (A.54), can be dealt with by applications of the Cauchy-Schwarz inequality. Consequently, $\frac{1}{T}\left|\sum_{t=1}^{T}\left(\widehat{\varepsilon}_{t}\left(\psi;\widehat{\tau}\right)\right)^{2}-\sum_{t=1}^{T}\left(\widehat{\varepsilon}_{t}\left(\psi\right)\right)^{2}\right|\xrightarrow{p}0$ uniformly in ψ, and we can conclude that $\widehat{\psi}\left(\widehat{\tau}\right)-\widehat{\psi}\xrightarrow{p}0$.

To complete the proof of Lemma C2, we again 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 (A.58) 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 (A.36) when $\tau=\widehat{\tau}$ and $\beta_{3}\neq0$. Here we need to demonstrate that

$$T^{-1/2}\sum_{t=1}^{T}\left(\Delta_{+}^{\delta}\left(z_{t}\left(\tau^{*}\right)-z_{t}\left(\widehat{\tau}\right)\right)'\widehat{\beta}\left(\widehat{\tau}\right)\right)^{2}\xrightarrow{p}0 \tag{A.59}$$

$$T^{-1/2}\sum_{t=1}^{T}\left(\Delta_{+}^{\delta}\left(z_{t}\left(\tau^{*}\right)-z_{t}\left(\widehat{\tau}\right)\right)'\widehat{\beta}\left(\widehat{\tau}\right)\right)\Delta_{+}^{\delta}z_{t}\left(\tau^{*}\right)'\left(\beta-\widehat{\beta}\left(\widehat{\tau}\right)\right)\xrightarrow{p}0 \tag{A.60}$$

and

$$T^{-1/2}\sum_{t=1}^{T}\left(\Delta_{+}^{\delta}\left(z_{t}\left(\tau^{*}\right)-z_{t}\left(\widehat{\tau}\right)\right)'\widehat{\beta}\left(\widehat{\tau}\right)\right)\left.\frac{\partial\varepsilon_{t}\left(\psi\right)}{\partial\psi}\right|_{\psi=\psi^{*}}\xrightarrow{p}0 \tag{A.61}$$

$$T^{-1/2}\sum_{t=1}^{T}\left(\Delta_{+}^{\delta}\left(z_{t}\left(\tau^{*}\right)-z_{t}\left(\widehat{\tau}\right)\right)'\widehat{\beta}\left(\widehat{\tau}\right)\right)\varepsilon_{t}\left(\psi^{*}\right)\xrightarrow{p}0. \tag{A.62}$$

The first two limits are readily established, using (A.57) for (A.59) and, in the case (A.60), the bound for the right hand side of (A.15) and an application of the Cauchy-Schwarz inequality.

Assuming that $\widehat{\tau}>\tau^{*}$, the expression in (A.61) 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,

$$\left|\sum_{t=1+\lfloor\tau^*T\rfloor}^{\lfloor\widehat{\tau}T\rfloor}\mu_{3,t}\left(\tau^*\right)\left.\frac{\partial\varepsilon_t\left(\psi\right)}{\partial\psi}\right|_{\psi=\psi^*}\right| \tag{A.63}$$

$$\leq \sum_{t=1+\lfloor\tau^*T\rfloor}^{\lfloor\widehat{\tau}T\rfloor-1}\left|\Delta\mu_{3,t+1}\left(\tau^*\right)\right|\max_{1+\lfloor\tau^*T\rfloor\leq t\leq\lfloor\widehat{\tau}T\rfloor-1}\left|\sum_{s=1+\lfloor\tau^*T\rfloor}^{t-1}\left.\frac{\partial\varepsilon_s\left(\psi\right)}{\partial\psi}\right|_{\psi=\psi^*}\right| \tag{A.64}$$

$$+\mu_{3,\lfloor\widehat{\tau}T\rfloor}\left(\tau^*\right)\left|\sum_{t=1+\lfloor\tau^*T\rfloor}^{\lfloor\widehat{\tau}T\rfloor}\left.\frac{\partial\varepsilon_t\left(\psi\right)}{\partial\psi}\right|_{\psi=\psi^*}\right| \tag{A.65}$$

and

$$\left|\sum_{t=1+\lfloor\widehat{\tau}T\rfloor}^{T}\left(\mu_{3,t}\left(\tau^*\right)-\mu_{3,t}\left(\widehat{\tau}\right)\right)\left.\frac{\partial\varepsilon_t\left(\psi\right)}{\partial\psi}\right|_{\psi=\psi^*}\right| \tag{A.66}$$

$$\leq \sum_{t=1+\lfloor\widehat{\tau}T\rfloor}^{T-1}\left|\Delta\left(\mu_{3,t+1}\left(\tau^*\right)-\mu_{3,t+1}\left(\widehat{\tau}\right)\right)\right|\max_{1+\lfloor\widehat{\tau}T\rfloor\leq t\leq T-1}\left|\sum_{s=1+\lfloor\widehat{\tau}T\rfloor}^{t-1}\left.\frac{\partial\varepsilon_s\left(\psi\right)}{\partial\psi}\right|_{\psi=\psi^*}\right| \tag{A.67}$$

$$+\left|\mu_{3,T}\left(\tau^*\right)-\mu_{3,T}\left(\widehat{\tau}\right)\right|\left|\sum_{t=1+\lfloor\widehat{\tau}T\rfloor}^{T}\left.\frac{\partial\varepsilon_t\left(\psi\right)}{\partial\psi}\right|_{\psi=\psi^*}\right|. \tag{A.68}$$

We discuss Model A first, beginning with (A.63). Here,

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

while

$$\max_{1+\lfloor\tau^*T\rfloor\leq t\leq\lfloor\widehat{\tau}T\rfloor-1}\left|\sum_{s=1+\lfloor\tau^*T\rfloor}^{t-1}\frac{\partial\varepsilon_s\left(\psi\right)}{\partial\psi}\right| \tag{A.69}$$

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

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

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

so that the stochastic order of (A.69) is the same as $\left(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^*T\rfloor\right)^{1/2}\ln\left(T\right)$ and the order of (A.64) is the same as,

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

which is of $o_p\left(1\right)$ using (3.12).

For the remainder term (A.65), $\mu_{3,\lfloor\widehat{\tau}T\rfloor}(\tau^*) \leq C(\lfloor\widehat{\tau}T\rfloor - \lfloor\tau^*T\rfloor)^{1-\delta}$. Again using Equation (8) of Bai (1994), (A.65) 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 (A.63) is $o_p(1)$ if Model A is used.

Moving to (A.66), this is bounded by

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

where $\max_{\rho,\,\theta} \left| \sum_{s=\lfloor\rho T\rfloor}^{\lfloor\tau T\rfloor} \left.\frac{\partial\varepsilon_s(\psi)}{\partial\psi}\right|_{\psi=\psi^*} \right| = O_p\left(T^{1/2}\right)$. Noticing that

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

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

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

so that (A.67) 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 $\delta < 0$,

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

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

and (A.67) has stochastic order as

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

So, regardless of whether $\delta < 0$ or $\delta > 0$, (A.67) is of $o_p\left(T^{1/2}\right)$.

For the remainder term in (A.68), recalling (A.56),

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

then (A.68) is of order $O_p\left(T^{-1/2} \times T^{1/2}\right) = O_p(1)$. Therefore, under Model A, (A.63) and (A.66) are $o_p\left(T^{1/2}\right)$ and (A.61) is $o_p(1)$.

When Model B is used, if $\delta < 0$,

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

and, recalling the bound for (A.69), (A.64) has stochastic order

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

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

If $\delta>0$,

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

recalling the bound for (A.69), then (A.64) has stochastic order $O_p((\ln(T))^2)$. Thus, regardless of δ, (A.64) has order $O_p((\ln(T))^2)$. For the remainder term in (A.65), $\mu_{3,\lfloor\widehat{\tau}T\rfloor}\left(\tau^{*}\right)\leq C\left(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^{*}T\rfloor\right)^{-\delta}$ and so (A.65) has the same stochastic order as that of $\left(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^{*}T\rfloor\right)^{-\delta}\times\ln\left(T\right)\times\left(\lfloor\widehat{\tau}T\rfloor-\lfloor\tau^{*}T\rfloor\right)^{1/2}=O_{p}\left(\ln\left(T\right)\right)$. Consequently, (A.63) is of $O_{p}\left(\left(\ln\left(T\right)\right)^{2}\right)$.

Turning to (A.66), recall first that

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

then

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

if $\delta<0$, and

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

if $\delta>0$. Where $\delta<0$, (A.66) 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

$$\sup_{t}\left|\left.\frac{\partial\varepsilon_{t}\left(\psi\right)}{\partial\psi}\right|_{\psi=\psi^{*}}\right|=O_{p}\left(T^{1/q}\right)\tag{A.70}$$

the stochastic order 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 $\delta>0$, (A.66) 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 $q > 2$.

To complete the proof of Lemma C2, we now need only show (A.70). We first show that, for ε_t an independent and identically distributed sequence with $E\left|\varepsilon_t\right|^p < \infty$ for $p \geq 1$, then,

$$\sup_t \left|\varepsilon_t\right| = O_p\left(T^{1/p}\right).$$

Notice that $\max\left|\varepsilon_t\right|^p \leq \sum_{t=1}^T \left|\varepsilon_t\right|^p$, and, since $E\left|\varepsilon_t\right|^p < \infty$ implies $\left|\varepsilon_t\right|^p = O_p(1)$, it then follows that $\max\left|\varepsilon_t\right|^p = O_p(T)$, or $\left|\varepsilon_t\right|^p = O_p(T)$, uniformly in t. Note that $\max\left|\varepsilon_t\right|^p = (\max\left|\varepsilon_t\right|)^p$ as the power is a monotone mapping. Thus, $\left|\varepsilon_t\right| = O_p\left(T^{1/p}\right)$ uniformly in t. Next, for $\eta_t = \sum_{j=0}^\infty c_j \varepsilon_{t-j}$ with $\sum_{j=0}^\infty j\left|c_j\right| < \infty$ (notice that this condition is met in ARMA models) and with $p \geq 2$, we establish $\sup_t \left|\eta_t\right| = O_p\left(T^{1/p}\right)$. Let $\widetilde{\varepsilon}_t := \sum_{j=T+1}^\infty c_j \varepsilon_{t-j}$, so that $\eta_t = \sum_{j=0}^T c_j \varepsilon_{t-j} + \widetilde{\varepsilon}_t$, and

$$\sup_t \left|\eta_t\right| \leq \sum_{j=0}^T \left|c_j\right| \sup_t \left|\varepsilon_{t-j}\right| + \sup_t \left|\widetilde{\varepsilon}_t\right|$$

where $\sum_{j=0}^T \left|c_j\right| \sup_t \left|\varepsilon_{t-j}\right| \leq \sup_t \left|\varepsilon_t\right| \sum_{j=0}^\infty \left|c_j\right| = O_p\left(T^{1/p}\right)$. Also, notice that $\max\left|\widetilde{\varepsilon}_t\right| = O_p((E(\max \widetilde{\varepsilon}_t^2))^{1/2})$ and $\max \widetilde{\varepsilon}_t^2 \leq \sum_{t=1}^T \widetilde{\varepsilon}_t^2$ and $E\left(\widetilde{\varepsilon}_t^2\right) = \sum_{j=T+1}^\infty c_j^2 = O_p\left(T^{-1}\right)$ in view of the fact that $\sum_{j=1}^\infty j\left|c_j\right| < \infty$. Hence, $\max \widetilde{\varepsilon}_t^2 = O_p(1)$ and $\max\left|\widetilde{\varepsilon}_t\right| = O_p(1)$. Therefore,

$$\sup_t \left|\eta_t\right| = O_p\left(T^{1/p} + 1\right) = O_p\left(T^{1/p}\right). \tag{A.71}$$

The stated bound in (A.70) then follows from (A.71) with $p = q$, while the stated bound for (A.62) can be established in the same way.

Proof of Lemma D2.

Using the expansion in (A.51) 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.

Proof of Theorem 1.

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

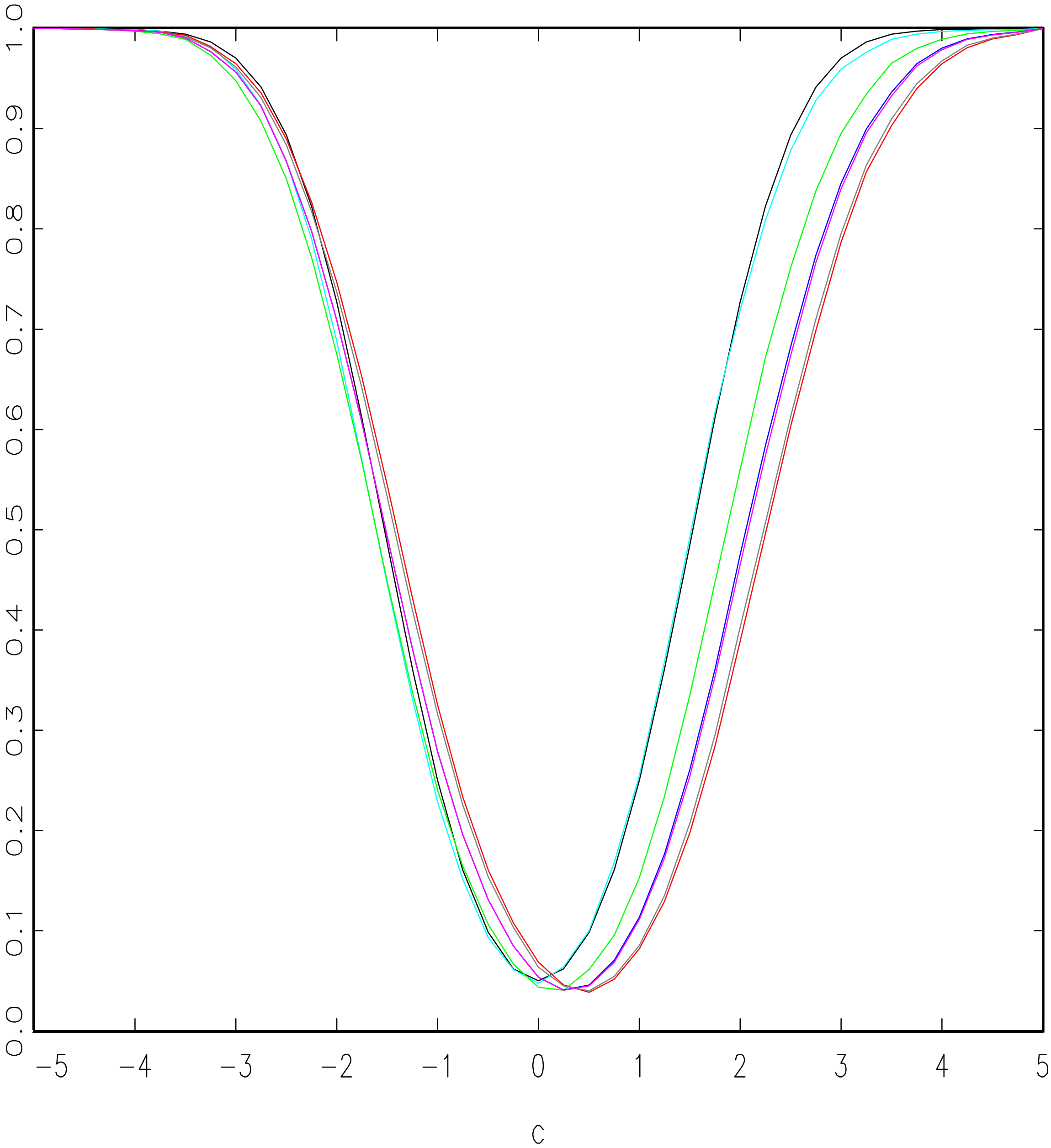
Table 2. Empirical size of tests, $a = -0.5$

	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, $T = 512$, $d_0 = 0$.



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

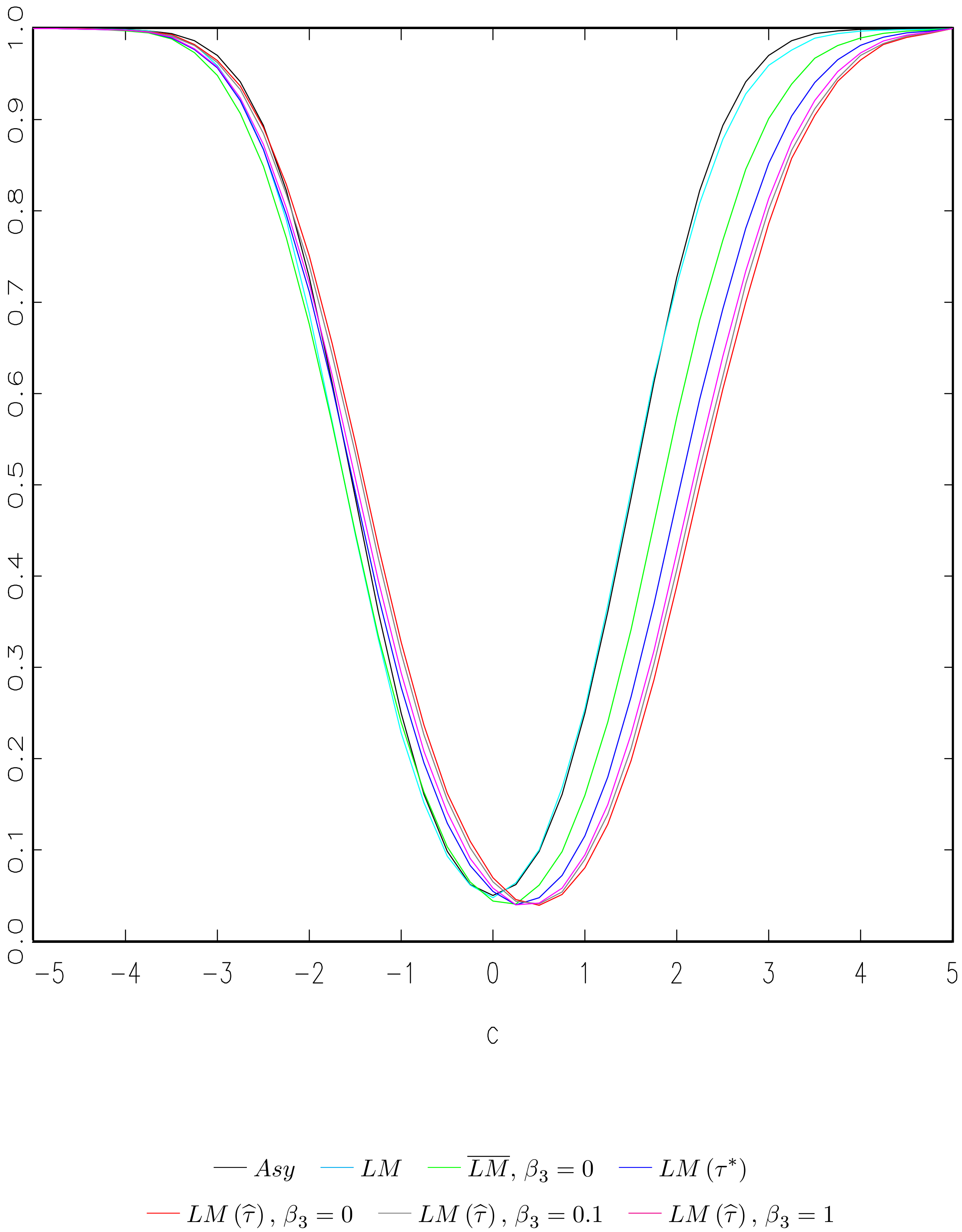


Figure 3. Local power of tests, $T = 512$, $d_0 = 0.5$.

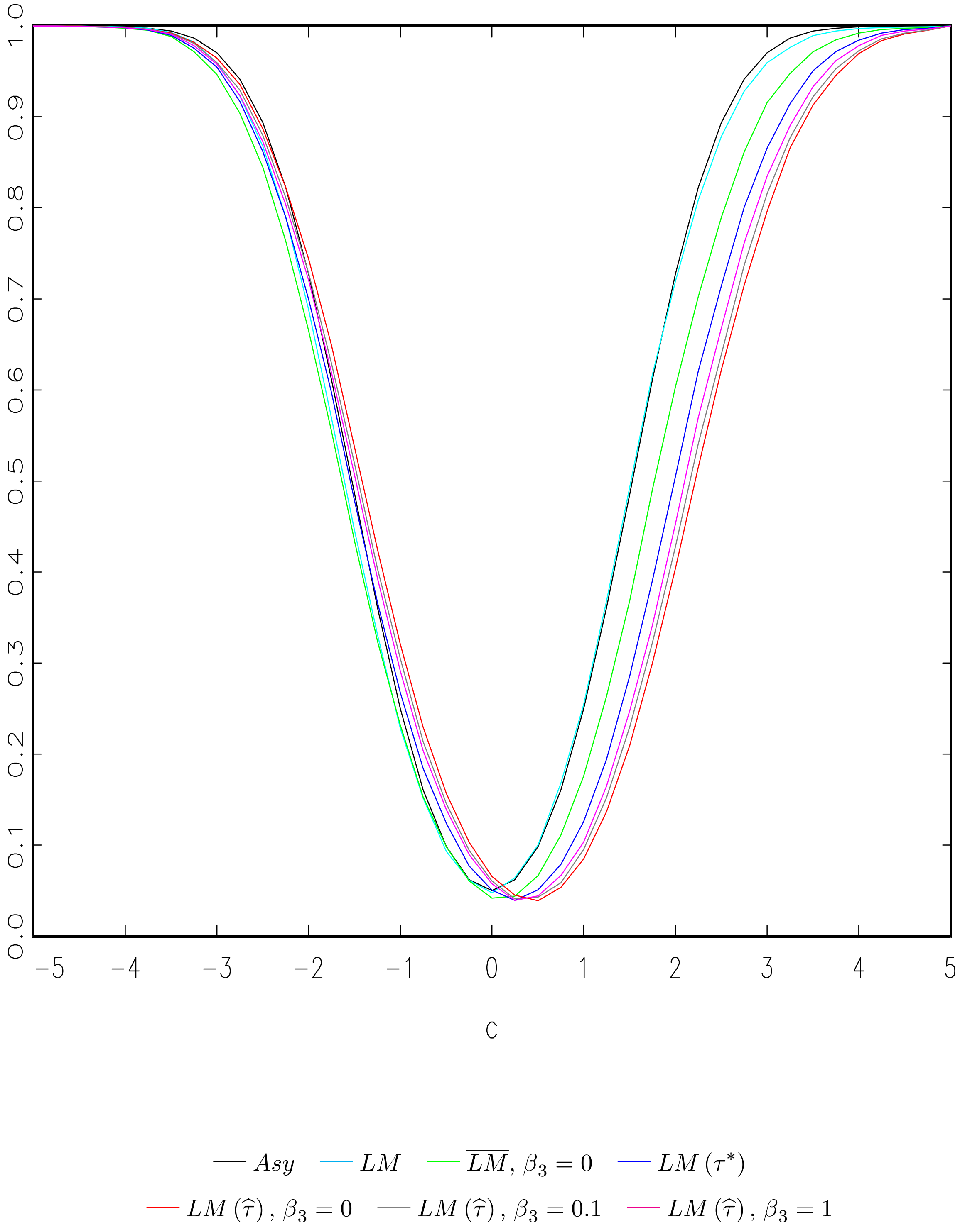
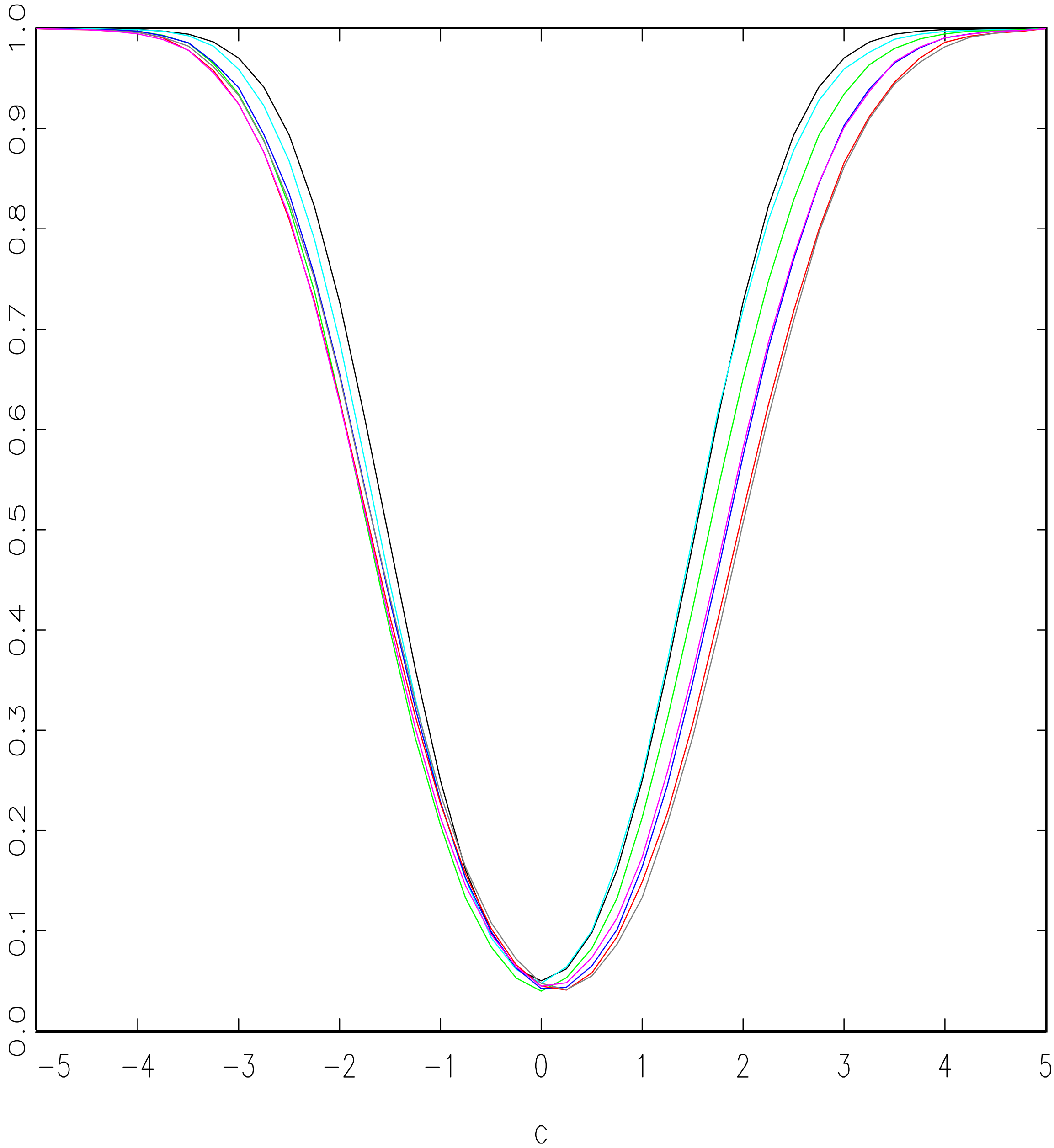


Figure 4. Local power of tests, $T = 512$, $d_0 = 0.75$.



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, $T = 512$, $d_0 = 1$.

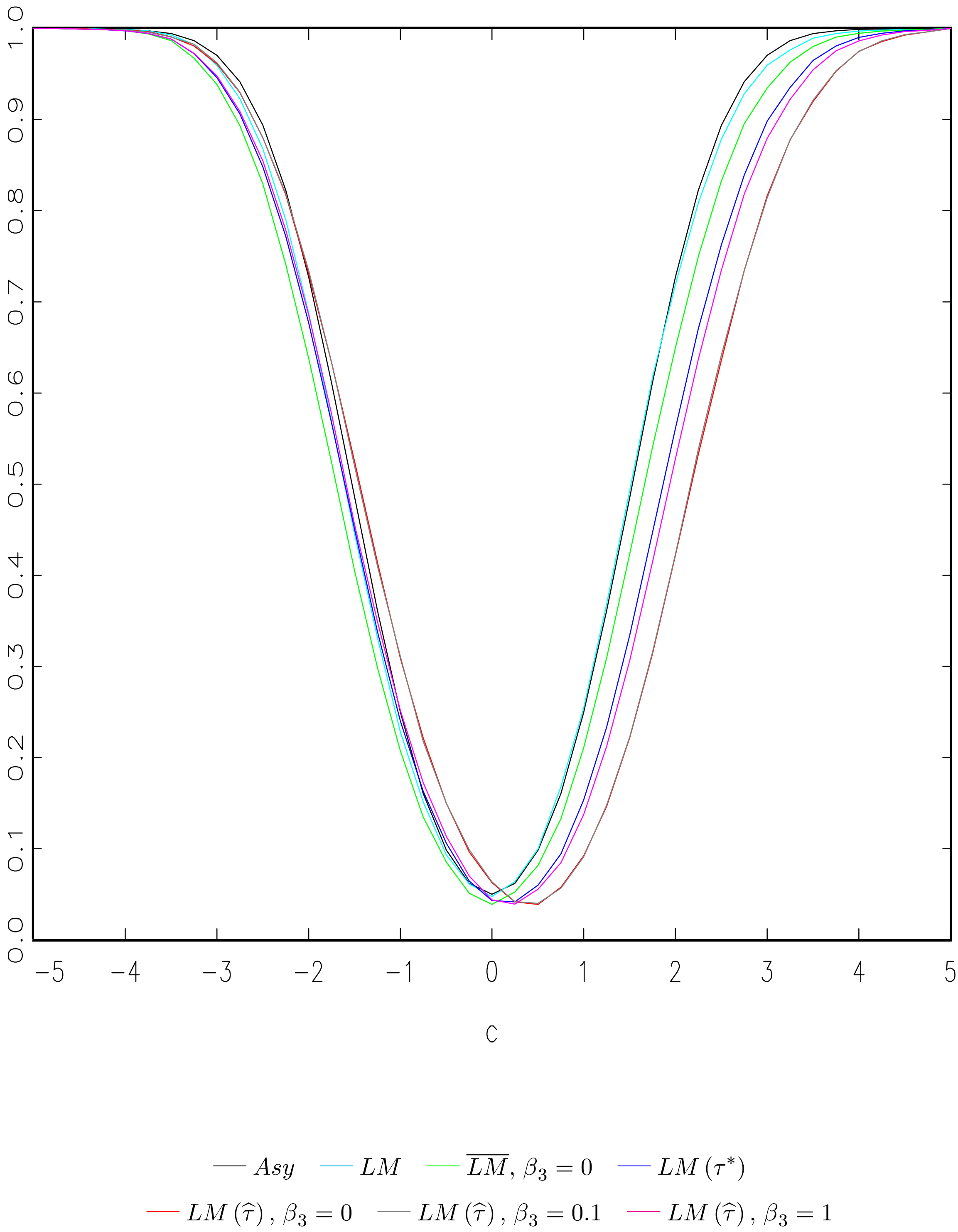
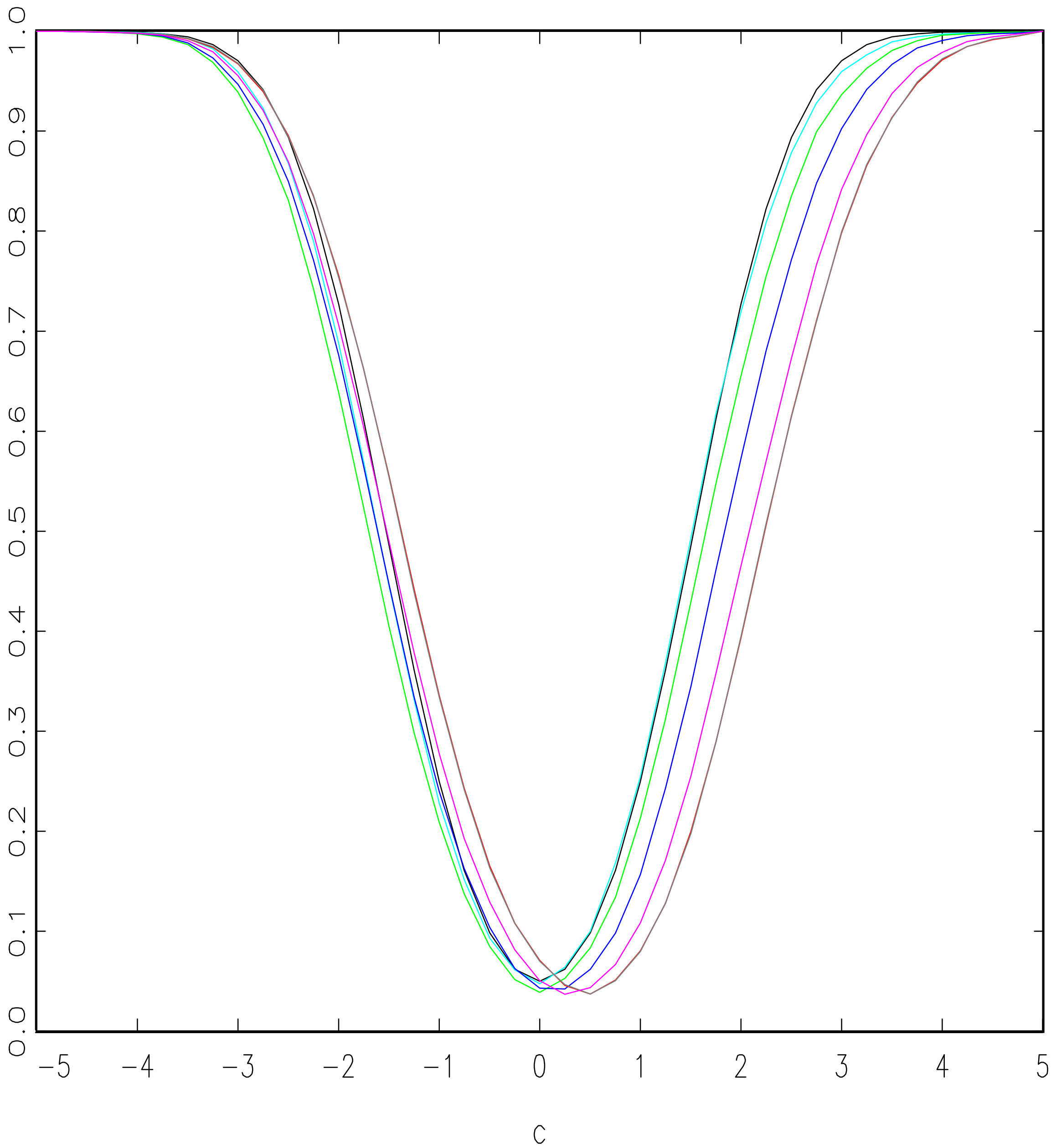


Figure 6. Local power of tests, $T = 512$, $d_0 = 1.25$.



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

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