# Supplementary Online Appendix

to

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

by

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

Date: August 15, 2018

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

## S.1 Mathematical Proofs

**Proof of Lemma A1:**

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

**Proof of Lemma B1:**

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

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

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

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

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

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

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

if $\delta < 0$ and

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

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

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

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

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

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

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

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

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

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

**Proof of Lemma C1:**

By a third order expansion and the mean value theorem,

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

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

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

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

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

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

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

We then rewrite

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

$$\leq \quad |\theta_T| \, T^{-(1/2+\alpha)} (\ln(T))^{-r} \left| \sum_{t=1}^{\lfloor \tau T \rfloor} \left( (\ln(\Delta))^{r+1} \Delta^{-\alpha} \right)_+ \eta_t \right| \tag{S.3}$$

$$+ \frac{1}{2} \theta_T^2 T^{-(1/2+\alpha)} (\ln(T))^{-r} \left| \sum_{t=1}^{\lfloor \tau T \rfloor} \left( (\ln(\Delta))^{r+2} \Delta^{-\alpha} \right)_+ \eta_t \right| \tag{S.4}$$

$$+ O_p\left( T^{-(1/2+\alpha)} \sum_{t=1}^{T} t^{1/2} |\theta_T|^3 \right). \tag{S.5}$$

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

**Proof of Lemma A2:**

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

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

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

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

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 $\psi$) convergence of a suitably scaled version of the loss function so that $T^{-1}\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\psi\right)\right)^2 \xrightarrow{p} \mathrm{E}\left(g\left(L;\psi\right)\eta_t\right)^2$, together with identification of the parameters $\psi_0$. The former is established as a uniform weak law of large numbers, that is obtained using pointwise convergence of the scaled loss function $T^{-1}\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\psi\right)\right)^2$ to the limit, and stochastic equicontinuity; see page 244 of Andrews (1992). Sufficient conditions for stochastic equicontinuity to hold in this case are that the loss function is differentiable with first derivative bounded in probability; see Assumptions (b) and (c) on page 246 of Andrews (1992).

Using the same approach as in Theorem A1 of Andrews (1993), to establish part (i) of the lemma we need to verify that $T^{-1}\left(\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\psi;\tau\right)\right)^2-\sum_{t=1}^{T}\left(\widehat{\varepsilon}_t\left(\psi\right)\right)^2\right)=o_p\left(1\right)$ uniformly in both $\psi$ and $\tau$. Uniformity in $\psi$ 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 $\tau$.

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

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

and that

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

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

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

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

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

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

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

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

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

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

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

Noting that $\left(\widehat{\varepsilon}_1\left(\psi\right)\right)^2 = O_p(1)$, uniformly in $\psi$, 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 (S.11) is $O_p(1)$, so that the whole expression in (S.11) is of $o_p\left(T^{1/2}\right)$. In view of Lemma 3 of Robinson (2005), Lemma A.1 and bound for $|\Delta_t^{(-\delta)}|$, it also holds that the contribution of the remainder $g\left(L;\psi\right)\Delta_t^{(-\delta)}\widehat{r}_1\left(\tau\right)$ is also of order $o_p\left(T^{1/2}\right)$.

Combining the foregoing results we therefore have that

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

As noted before, this is sufficient to establish that $\widehat{\psi}\left(\tau\right)-\widehat{\psi} = o_p(1)$, 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) yields

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

respectively, where

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

Recalling (S.7), we have that

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

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

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

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 $\widehat{\psi}$,

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

From

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

then

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

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

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

then

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

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

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

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

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

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

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

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

and

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

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

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

and notice that, using repeated substitutions,

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

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

Moreover, using (S.22) again,

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

and, using repeated substitutions,

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

Thus, bounding

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

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

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

Continuing, we next show that

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

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

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

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

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

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

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

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

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

is of $o_p(1)$.

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

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

and

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

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

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

Thus, (S.24) is

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

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

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

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

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

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

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

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

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

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

for $\psi = \psi^*$.

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

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

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

When Model A is used,

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

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

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

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

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

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

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

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

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

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

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

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

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

whereas, when $\delta < 0$,

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

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

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

**Proof of Lemma B2:**

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

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

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

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

$$
\begin{aligned}
\sum_{t=1}^{T} \widehat{\varepsilon}_t\left(\widehat{\psi}(\tau);\tau\right)\left(\widehat{v}_t\left(\widehat{\psi}(\tau);\tau\right) - \widehat{v}_t\left(\widehat{\psi}\right)\right) &= o_p\left(T^{1/2}\right) && \text{(S.31)} \\
\sum_{t=1}^{T} \left(\widehat{\varepsilon}_t\left(\widehat{\psi}(\tau);\tau\right) - \widehat{\varepsilon}_t\left(\widehat{\psi}\right)\right)\widehat{v}_t\left(\widehat{\psi}\right) &= o_p\left(T^{1/2}\right). && \text{(S.32)}
\end{aligned}
$$

To that end, observe first that

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

where

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

and

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

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

In the same way, observe that

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

where

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

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

Next, let

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

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

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

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

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

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

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

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

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

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

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

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

For (S.37),

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

Noting that

$$
\begin{aligned}
\sup_\rho\left|\sum_{s=1}^{\lfloor\rho T\rfloor} \widehat{\varepsilon}_s\left(\widehat{\psi}(\tau) ; \tau\right)\right| \leq & \sup_\rho\left|\sum_{s=1}^{\lfloor\rho T\rfloor} \widehat{\varepsilon}_t\left(\widehat{\psi}(\tau)\right)\right| \\
& +\sup_\rho\left|\sum_{s=1}^{\lfloor\rho T\rfloor} g\left(L ; \widehat{\psi}(\tau)\right) \Delta_{+}^{\delta} z_s(\tau)^{\prime}\left(\beta-\widehat{\beta}(\tau)\right)\right| \qquad \text{(S.38)}
\end{aligned}
$$

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

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

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

$$
\begin{aligned}
& \sup_\rho\left|\sum_{s=1}^{\lfloor\rho T\rfloor} \Delta_{+}^{\delta} z_s(\tau)^{\prime}\left(\beta-\widehat{\beta}(\tau)\right)\right| \\
\leq & C \sum_{t=1}^{T} \mu_{1,t}\left|\beta_1-\widehat{\beta}_1(\tau)\right|+\sum_{t=1}^{T} \mu_{2,t}\left|\beta_2-\widehat{\beta}_2(\tau)\right|+\sum_{t=1}^{T} \mu_{2,t}\left|\widehat{\beta}_3(\tau)\right|=O_p\left(T^{1/2}\right) .
\end{aligned}
$$

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

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

and notice that this has the same stochastic order as

$$
\sum_{t=1}^{T-1}\left|\left(\left\{(\ln(\Delta)) \Delta^{\delta}\right\}_{+} z_{t+1}(\tau)-\left\{(\ln(\Delta)) \Delta^{\delta}\right\}_{+} z_t(\tau)\right)^{\prime}\left(\beta-\widehat{\beta}(\tau)\right)\right| .
$$

When Model A is used, the latter is bounded by

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

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

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

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

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

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

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

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

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

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

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

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

For (A.6),

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

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

Finally, since $\kappa$ and $\Phi$ are continuous function of $\psi$, (A.7) follows by an application of Slutzky's Theorem.

**Proof of Lemma C2.**

We have that,

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

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

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

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

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

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

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

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

and $\widehat{\beta}_{3}\left(\widehat{\tau}\right)\xrightarrow{p}\beta_{3}$ so $\widehat{\beta}_{3}\left(\widehat{\tau}\right)=O_{p}\left(1\right)$. Term (S.46) has therefore the same stochastic order as that of

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

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,

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

and, if $\delta>0$,

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

and

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

whereas, if $\delta < 0$,

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

and

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

Either way, then,

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

When Model B is used,

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

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

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

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

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

It therefore follows that

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

and

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

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

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

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

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

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

and

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

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

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

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

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

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

Using summation by parts,

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

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

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

and

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

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

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

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

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

while

$$\max_{1+\lfloor\tau^* T\rfloor\leq t\leq\lfloor\widehat{\tau}T\rfloor-1}\left|\sum_{s=1+\lfloor\tau^* T\rfloor}^{t-1}\frac{\partial\varepsilon_s(\psi)}{\partial\psi}\right| \tag{S.61}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

and

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

so that (S.59) is of order $O_p\left(T^{-1/2+\delta} \times \ln(T) \times T^{1/2}\right) = O_p\left(T^{\delta}\ln(T)\right) = o_p\left(T^{1/2}\right)$.
If $\delta < 0$,

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

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

and (S.59) has stochastic order as

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

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

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

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

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

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

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

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

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

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

If $\delta > 0$,

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

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

Turning to (S.58), recall first that

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

then

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

if $\delta < 0$, and

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

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

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

Using the fact, which will be established below, that

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

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

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

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

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

using the fact that $q > 2$.
We have therefore proved (S.53) for all cases. To complete the proof of Lemma C2, the bound for (S.54) can be established in the same way.

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

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

Moreover, for $p > 0$,

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

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

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

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

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

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

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

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

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

## S.2 Additional Monte Carlo Simulations

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

### S.2.1 Power against Fixed Magnitude Alternatives

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

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

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

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

Table S1. Empirical power of tests for distant alternatives

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

Table S2. Empirical power of tests for DF type alternative

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

### S.2.2 Moment Conditions

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

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

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

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

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

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

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

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

$d = 0.51$

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

$d = 0.55$

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

$d = 0.6$

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

$d = 0.75$

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

$d = 1$

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

### S.2.3 Model Selection

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

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

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

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

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

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

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

## Additional References

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


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

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

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

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

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

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

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

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

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

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

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