Queen's Economics Department Working Paper No. 1376

# Truncated sum of squares estimation of fractional time series models with deterministic trends

Javier Hualde
Universidad Pública de Navarra

Morten Ørregaard Nielsen
Queen's University and CREATES

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

3-2019

# Truncated sum of squares estimation of fractional time series models with deterministic trends*

Javier Hualde[†]
Universidad Pública de Navarra

Morten Ørregaard Nielsen[‡]
Queen's University and CREATES

March 21, 2019

**Abstract**

We consider truncated (or conditional) sum of squares estimation of a parametric model composed of a fractional time series and an additive generalized polynomial trend. Both the memory parameter, which characterizes the behaviour of the stochastic component of the model, and the exponent parameter, which drives the shape of the deterministic component, are considered not only unknown real numbers, but also lying in arbitrarily large (but finite) intervals. Thus, our model captures different forms of nonstationarity and noninvertibility. As in related settings, the proof of consistency (which is a prerequisite for proving asymptotic normality) is challenging due to non-uniform convergence of the objective function over a large admissible parameter space, but, in addition, our framework is substantially more involved due to the competition between stochastic and deterministic components. We establish consistency and asymptotic normality under quite general circumstances, finding that results differ crucially depending on the relative strength of the deterministic and stochastic components. Finite-sample properties are illustrated by means of a Monte Carlo experiment.

*JEL classification*: C22.

*Key words and phrases*: Asymptotic normality, consistency, deterministic trend, fractional process, generalized polynomial trend, noninvertibility, nonstationarity, truncated sum of squares estimation.

*We are grateful to the Co-Editor, Rob Taylor, three anonymous referees, Søren Johansen, Peter M. Robinson, and seminar participants at Universidad Carlos III, Universitat Pompeu Fabra, Universidad de Alicante, University of Nottingham, and at the 3rd CREATES Long Memory Symposium for useful comments. Javier Hualde's research is supported by the Spanish Ministerio de Economía y Competitividad through project ECO2015-64330-P. Morten Ø. Nielsen's research is supported by the Canada Research Chairs program and the Social Sciences and Humanities Research Council of Canada (SSHRC). Both authors are thankful to the Center for Research in Econometric Analysis of Time Series (CREATES, funded by the Danish National Research Foundation, DNRF78) for financial support.

[†]Email: `javier.hualde@unavarra.es`

[‡]Corresponding author. Email: `mon@econ.queensu.ca`

## 1 Introduction

In time series analysis the most traditional approach to modeling stochastic components is by stationary and invertible autoregressive moving average (ARMA) processes, but unit root nonstationary and noninvertible processes have also been extensively considered. Additionally, supplementing the random component, the presence of a low-order polynomial term, such as a constant or a linear deterministic trend is usually assumed.

More recently, the relatively simple ARMA modeling framework has been generalized in various directions. Here, one of the main developments is that of fractionally integrated ARMA (FARIMA) models which bridge the behavioral gap between stationary and invertible ARMA, which has "memory parameter" $\delta_0$ equal to zero, and unit root nonstationary process, where $\delta_0 = 1$. A zero-mean FARIMA$(p_1, \delta_0, p_2)$ process $z_t$ is given by

$$z_t = \Delta_+^{-\delta_0} u_t, \;\; t = 0, \pm 1, \pm 2, \ldots, \tag{1}$$

$$\alpha(L) u_t = \beta(L) \varepsilon_t, \tag{2}$$

where $\Delta = 1 - L$ and $L$ are the difference and lag operators, respectively, and where $\delta_0$ can take any value within an arbitrarily large compact interval. For any series $v_t$, real number $\zeta$ and $t \geq 1$, we define the operator $\Delta_+^{\zeta}$ by

$$\Delta_+^{\zeta} v_t = \Delta^{\zeta} \{ v_t \mathbb{I}(t \geq 1) \} = \sum_{i=0}^{t-1} \pi_i \left( -\zeta \right) v_{t-i}$$

with $\mathbb{I}(\cdot)$ denoting the indicator function, $\pi_i(v) = 0$ for $i < 0$, $\pi_0(v) = 1$, and

$$\pi_i \left( v \right) = \frac{\Gamma \left( v + i \right)}{\Gamma \left( v \right) \Gamma \left( 1 + i \right)} = \frac{v(v+1) \ldots (v + i - 1)}{i!}, \;\; i \geq 1, \tag{3}$$

denoting the coefficients in the usual binomial expansion of $(1 - z)^{-v}$, where $\Gamma(\cdot)$ is the gamma function with the convention $\Gamma(i) = 0$ for $i = 0, -1, -2, ...$ Thus, in particular, (1) implies the initial condition $z_t = 0$ for $t \leq 0$, which is common in the fractional time series literature implying a so-called "Type II" fractional model. Additionally, $\alpha(L)$ and $\beta(L)$ are real polynomials of degrees $p_1$ and $p_2$, which share no common zeros and have all their zeros outside the unit circle in the complex plane, and $\varepsilon_t$ is a zero-mean, serially uncorrelated and homoskedastic sequence. More precise conditions will be imposed below.

For the sake of greater generality, we retain (1) but generalize (2) to

$$u_t = \omega(L; \boldsymbol{\varphi}_0) \varepsilon_t, \;\; t = 0, \pm 1, \pm 2, \ldots, \tag{4}$$

where $\boldsymbol{\varphi}_0$ is an unknown $p \times 1$ vector and $\omega(s; \boldsymbol{\varphi}) = \sum_{j=0}^{\infty} \omega_j(\boldsymbol{\varphi}) s^j$ with $\omega_0(\boldsymbol{\varphi}) = 1$, $|\omega(s; \boldsymbol{\varphi})| \neq 0$ for $|s| \leq 1$, and $\sum_{j=0}^{\infty} |\omega_j(\boldsymbol{\varphi})| < \infty$ (these conditions on $\omega(s; \boldsymbol{\varphi})$ will be made precise in Assumption A1 below). Like $\alpha$ and $\beta$ in (2), $\omega$ in (4) thus characterizes parametric short memory autocorrelation.

Although many theoretical works have exclusively assumed a purely random process (see, e.g., the discussion in Hualde and Robinson, 2011, or Nielsen, 2015), in practice, model (1), (4) (or a semiparametric version of it, where $u_t$ is a nonparametric invertible weakly dependent process, that is with spectrum which is bounded and bounded away from zero at all frequencies) is usually complemented with deterministic components. To model deterministic behavior, the literature stresses low-order polynomials in $t$, such as a constant intercept or a linear function, but this seems arbitrary in light of the fractional behaviour of $z_t$. In fact, these cases are nested by the so-called generalized polynomial (or power law) model considered by, e.g., Wu (1981), Robinson (2005, 2012), and Phillips (2007). Thus, we extend the fractional model (1), (4) by letting the observable process $x_t$ be generated by

$$x_t = \mu_0 t_+^{\gamma_0} + z_t, \tag{5}$$

where $t_+ = t\mathbb{I}(t \geq 1)$, $\mu_0$ and $\gamma_0$ are unknown real-valued parameters, and $z_t$ is generated by (1) and (4). The truncation in $t_+$ in (5) has been employed in related settings, see e.g. (1.8) in Robinson (2005) and (19)–(20) in Robinson and Iacone (2005), and like that in (1), it imposes zero initial conditions. Obviously, (5) covers standard cases like a constant ($\gamma_0 = 0$) or a linear trend ($\gamma_0 = 1$), but since $\gamma_0$ is real, model (5) characterizes a wide range of situations.

Of course, when $\gamma_0$ is known, (5) reduces to a fractional time series model with an additive deterministic component such as a constant term, $\mu_0$, or a linear trend term, $\mu_0 t_+$ (for the cases $\gamma_0 = 0$ and $\gamma_0 = 1$, respectively). These special cases are also covered by our results. For example, letting $\gamma_0 = 0$ be known, $\mu_0$ would be the so-called "level parameter" analyzed by Johansen and Nielsen (2016) to alleviate bias issues arising from the imposition of zero initial conditions. Of course, this is especially relevant in applications since typically (economic) time series do not start at zero unless some preliminary recentering or reinitialization has been performed, e.g., by subtracting the first observation. We note that, even for these special cases of the model, our results appear to be novel. However, in general, we consider $\gamma_0$ to be an unknown, real-valued parameter to be estimated jointly with the remaining parameters.

Letting the deterministic component be of a fractional order mimics the fractional behavior of the stochastic component. This can be made more precise using the terminology of White and Granger (2011), where different definitions of trends appear. Because $Var(x_t)$ grows at rate $t^{2\delta_0 - 1}$ when $\delta_0 > 1/2$, whereas $E(x_t) = \mu_0 t_+^{\gamma_0}$, according to White and Granger (2011, pp. 9–10), if $\delta_0 > 1/2$, the process (5) has a "stochastic trend in variance", whereas if $\mu_0 \neq 0$ and $\gamma_0 \neq 0$, it also has a "stochastic trend in mean". We note that the evolution of these two trends is governed by the parameters $\delta_0$ and $\gamma_0$, respectively, and hence letting $\gamma_0$ be real-valued appears as natural as letting $\delta_0$ be real-valued; the two parameters just affect different aspects of the distribution of $x_t$.

As will be seen below, apart from belonging to a compact set, the only restriction we place on $\gamma_0$ is that $\gamma_0 > -1$. The reason is mainly technical: most of our results are rooted on the analysis of fractional differences of time trends, that is for $t \geq 1$, $\Delta_+^d t^c$ for real values of $c$ and $d$. As Robinson (2005) justifies (see also Lemma S.13 below), $\Delta_+^d t^c$ is driven by the exact rate $t^{c-d}$, but only if $c > -1$. Thus, while considering the case $\gamma_0 \leq -1$ in (5) is possible, this would increase substantially the technical difficulties in our proofs without adding much, if any, relevance from an empirical perspective.

In similar contexts to ours, several authors have considered the same type of generalized polynomial trends as in (5), with $\gamma_0$ being an unknown real-valued parameter. For the same type of truncated/conditional sum-of-squares estimator that we analyze in this paper, Wu (1981) noted in his Example 4 that model (5) does not satisfy his assumptions for the asymptotic analysis, even when $z_t$ is an independent sequence, because of the asymptotic singularity of the Hessian and the requirement that the parameters $\mu$ and $\gamma$ have different normalizations. The analysis of Wu (1981) was generalized by Phillips (2007) to allow such singularity of the Hessian and hence accommodate model (5), but assuming at most weakly dependent errors $z_t$. Robinson and Marinucci (2000) and Robinson and Iacone (2005) consider semiparametric frequency domain estimators in models that include both a generalized polynomial trend and a nonstationary FARIMA stochastic component. In a spatial setting, Robinson (2012) considers a general version of (5) involving more deterministic terms, but with only weakly dependent errors, $z_t$, and explicitly excludes the situation where the deterministic component is dominated by the stochastic component, i.e. where $\gamma_0 + 1/2 < \delta_0$ in our notation. The latter situation is discussed in Johansen and Nielsen (2016), who consider truncated/conditional sum-of-squares estimation of (5) with *known* $\gamma_0 = 0$ and $\delta_0 > 1/2$, and hence $\gamma_0 + 1/2 < \delta_0$, although with $u_t$ being an independent sequence. They prove consistency and asymptotic normality of the standard estimator which ignores the deterministic term. Finally, Robinson (2005) considers $M$-estimation of a model like (5) with *known* $\gamma_0$ and allowing for fractional $z_t$. We note that if $\gamma_0$ were known in (5) our estimation problem would be simplified greatly. More importantly, since $\gamma_0$ is allowed to take any real value, it seems natural to consider it an unknown parameter as in, e.g., Phillips (2007) and Robinson (2012).

In this paper we analyze the model (5) with the stochastic term $z_t$ given by (1) and (4), and prove consistency and asymptotic normality of the parameter estimators. Note that (5) with $\mu_0 = 0$ assumed known reduces to the model discussed in Hualde and Robinson (2011), where the behaviour of the observable process is entirely driven by $\delta_0$ and $\boldsymbol{\varphi}_0$. As will be seen, the complication added by the consideration of unknown deterministic parameters is substantial, especially because we let both the memory ($\delta_0$) and exponent ($\gamma_0$) parameters lie in arbitrarily large, but finite, intervals, so $x_t$ can display many different behaviours. As in related works, e.g. Hualde and Robinson (2011) and Nielsen (2015), the proof of consistency (a prerequisite for proving asymptotic normality) is challenging due to non-uniform convergence of the objective function over a large admissible parameter space. However, in addition to this well-known complication, our framework is substantially more involved due to the competition between the stochastic and deterministic components, and this competition needs to be explicitly taken advantage of in the proof of consistency.

We establish consistency and asymptotic normality of the parameter estimators under quite general circumstances, finding that results differ substantially depending on the relative strength of the deterministic and stochastic components. In particular, when $\gamma_0 + 1/2 > \delta_0$ the estimators of all parameters in the model are consistent and asymptotically normally distributed. On the other hand, when $\gamma_0 + 1/2 < \delta_0$ the parameters related to the deterministic part of the model, $\mu_0$ and $\gamma_0$, cannot be consistently estimated, but, interestingly, those related to the stochastic part of the model, $\delta_0$ and $\boldsymbol{\varphi}_0$, are still consistently estimated and their asymptotic normal distribution is unaffected by the presence of the remaining unestimable parameters. This latter result resembles that of Heyde and Dai (1996), who provided conditions under which small trends do not affect the properties of Whittle estimators applied to short- or long-range dependent processes. Similar results were derived by Abadir, Distaso, and Giraitis (2007) and Iacone (2010) for different versions of the local Whittle estimator. In comparison to these works, our main results can be view as a step forward in the difficult task of disentangling the stochastic memory properties of observed time series from the low-frequency effect of deterministic components; a problem that goes back to, at least, Künsch (1986). In fact, our proposed estimators of the stochastic components of the model retain their limiting properties regardless the intensity of the deterministic signal, which seems an advantage of our approach in this context. Finally, we include a Monte Carlo simulation study which supports our theoretical results and illustrates the findings.

The next section formalizes the model and assumptions. In Section 3 we present the estimator and our main results on consistency and asymptotic normality. The results of some Monte Carlo simulations are reported in Section 4, and concluding remarks are presented in Section 5. All proofs are given in the supplementary material Hualde and Nielsen (2019).

## 2 Model and assumptions

As usual, we let the true values of the parameters be denoted by subscript zero. We consider the model (5), where $z_t$ is generated by (1) and (4), and $\mu_0$, $\gamma_0$, $\delta_0$, and $\boldsymbol{\varphi}_0$ are unknown parameters to be estimated.

We first impose an assumption on the short memory component, $\omega$, where $\boldsymbol{\varphi}_0$ is assumed to lie in $\Psi$, which is a compact and convex subset of $\mathbb{R}^p$.

**A1. (i)** for all $\boldsymbol{\varphi} \in \Psi \backslash \{\boldsymbol{\varphi}_0\}$, $|\omega(s;\boldsymbol{\varphi})| \neq |\omega(s;\boldsymbol{\varphi}_0)|$ on a set $S \subset \{s : |s| = 1\}$ of positive Lebesgue measure;

**(ii)** for all $\boldsymbol{\varphi} \in \Psi$, $\omega\left(e^{i\lambda};\boldsymbol{\varphi}\right)$ is differentiable in $\lambda$ with derivative in $\mathrm{Lip}\left(\varsigma\right)$ for $1/2 < \varsigma \leq 1$;

**(iii)** for all $\lambda$, $\omega\left(e^{i\lambda};\boldsymbol{\varphi}\right)$ is continuous in $\boldsymbol{\varphi}$;

**(iv)** for all $\boldsymbol{\varphi} \in \Psi$, $|\omega\left(s;\boldsymbol{\varphi}\right)| \neq 0, |s| \leq 1$.

Assumption A1 is identical to A1 in Hualde and Robinson (2011). In particular, (i) ensures identification, (ii) and (iv) imply that $u_t$ is an invertible weakly dependent process, while by (ii) and (iii) $\sup_{\boldsymbol{\varphi}\in\Psi} |\omega_j\left(\boldsymbol{\varphi}\right)| = O\left(j^{-1-\varsigma}\right)$ as $j \to \infty$ (see Hualde and Robinson, 2011) and hence $\sup_{|s|=1,\boldsymbol{\varphi}\in\Psi} |\omega(s;\boldsymbol{\varphi})| < \infty$. Also, writing $\omega^{-1}(s;\boldsymbol{\varphi}) = \phi(s;\boldsymbol{\varphi}) = \sum_{j=0}^{\infty} \phi_j(\boldsymbol{\varphi})s^j$, it holds that $\phi_0\left(\boldsymbol{\varphi}\right) = 1$ for all $\boldsymbol{\varphi}$, and (ii), (iii), and (iv) imply that

$$\sup_{\boldsymbol{\varphi}\in\Psi} |\phi_j\left(\boldsymbol{\varphi}\right)| = O\left(j^{-1-\varsigma}\right) \text{ as } j \to \infty \tag{6}$$

for $\varsigma > 1/2$ given in (ii), and

$$\inf_{|s|=1,\boldsymbol{\varphi}\in\Psi} |\phi\left(s;\boldsymbol{\varphi}\right)| = \frac{1}{\sup_{|s|=1,\boldsymbol{\varphi}\in\Psi} |\omega(s;\boldsymbol{\varphi})|} > 0. \tag{7}$$

A1 is easily satisfied in the stationary and invertible ARMA case. Another model covered by A1 is the exponential spectrum model of Bloomfield (1973), which leads to a relatively simple variance matrix formula in the context of fractional time series models, see Robinson (1994). More generally, A1 is similar to conditions employed in asymptotic theory for conditional sum-of-squares estimates, e.g. Hualde and Robinson (2011) and Nielsen (2015), as well as Whittle estimators that restrict to stationarity, e.g. Fox and Taqqu (1986), Dahlhaus (1989), and Giraitis and Surgailis (1990). Assumption A1 can be readily verified because $\omega$ is a known parametric function. In fact, $\omega$ satisfying A1 are invariably employed by practitioners.

**A2.** The $\varepsilon_t$ in (4) are stationary and ergodic with finite fourth moment, $E\left(\varepsilon_t|\,\mathcal{F}_{t-1}\right) = 0$, $E\left(\varepsilon_t^2|\,\mathcal{F}_{t-1}\right) = \sigma_0^2$, a.s., where $\mathcal{F}_t$ is the $\sigma$-field of events generated by $\varepsilon_s$, $s \leq t$, and conditional (on $\mathcal{F}_{t-1}$) third and fourth moments of $\varepsilon_t$ equal the corresponding unconditional moments.

Assumption A2 is identical to A2 in Hualde and Robinson (2011). It does not impose independence or identity of distribution of $\varepsilon_t$, but rules out conditional heteroskedasticity. It is standard in the time series asymptotics literature since Hannan (1973).

Finally, we impose an assumption on the parameter space. Let prime denote transposition.

**A3.** The parameter space for $\boldsymbol{\vartheta} = (\delta, \boldsymbol{\varphi}', \gamma)'$ in (1), (4), and (5) is given by $\Xi = [\triangledown_1, \triangledown_2] \times \Psi \times [\Box_1, \Box_2]$ with $\triangledown_1 < \triangledown_2$ and $-1 < \Box_1 < \Box_2$, where $\Psi$ is compact and convex and $\boldsymbol{\vartheta}_0 = (\delta_0, \boldsymbol{\varphi}_0', \gamma_0)' \in \Xi$. For $\mu$ the parameter space is $\mathbb{R}$, and if $\gamma_0 + 1/2 > \delta_0$, we also assume that $\mu_0 \neq 0$.

We assume that $\mu_0 \neq 0$ when $\gamma_0 + 1/2 > \delta_0$ since otherwise $\gamma_0$ is not identified in (5). Note that even vanishing trends (negative $\gamma_0$) can be identified when $\delta_0$ is sufficiently small ($\delta_0 < \gamma_0 + 1/2$) because then, in a sense, the value of $\delta_0$ in (1) helps the identification of the deterministic part.

Finally, note that the model where $\gamma_0$ is known in (5), e.g., the model with a constant term or a linear trend, is a special case of our model with unknown $\gamma_0$. Hence, the asymptotic results can easily be specialized to this situation, see Corollary 1 below. In general, though, $[\triangledown_1, \triangledown_2]$ and $[\Box_1, \Box_2]$ are allowed to be arbitrarily large, with the only restriction that $\Box_1 > -1$, which implies $\gamma_0 > -1$.

## 3 Truncated sum of squares estimation

We collect the parameters for the stochastic component in $\boldsymbol{\tau} = (\delta, \boldsymbol{\varphi}')'$ with true value $\boldsymbol{\tau}_0 = (\delta_0, \boldsymbol{\varphi}_0')'$, and denote the estimator (to be defined below) by $\widehat{\boldsymbol{\tau}} = (\widehat{\delta}, \widehat{\boldsymbol{\varphi}}')'$. We also use the notation $\boldsymbol{\vartheta} = (\boldsymbol{\tau}', \gamma)'$, $\boldsymbol{\vartheta}_0 = (\boldsymbol{\tau}_0', \gamma_0)'$, and $\widehat{\boldsymbol{\vartheta}} = (\widehat{\boldsymbol{\tau}}', \widehat{\gamma})'$. The Gaussian log-likelihood, conditional on $x_t = 0$ for $t \leq 0$, is, apart from a constant, given by

$$
\begin{aligned}
L_T(\boldsymbol{\vartheta}, \mu, \sigma^2) &= -\frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(\phi(L;\boldsymbol{\varphi})\Delta_+^{\delta}(x_t - \mu t^{\gamma})\right)^2 \\
&= -\frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(\phi(L;\boldsymbol{\varphi})\Delta_+^{\delta}x_t - \mu c_t(\gamma, \delta, \boldsymbol{\varphi})\right)^2, \qquad (8)
\end{aligned}
$$

where we define the (convolution) coefficients

$$
c_t(\gamma, \delta, \boldsymbol{\varphi}) = \phi(L;\boldsymbol{\varphi})\Delta_+^{\delta}t^{\gamma} = \sum_{j=0}^{t-1} a_j\left(-\delta, \boldsymbol{\varphi}\right)(t-j)^{\gamma}, \qquad (9)
$$

and

$$
a_j\left(d, \boldsymbol{\varphi}\right) = \sum_{k=0}^{j}\phi_k(\boldsymbol{\varphi})\pi_{j-k}\left(d\right), \qquad (10)
$$

Alternatively,

$$
c_t(\gamma, \delta, \boldsymbol{\varphi}) = \sum_{j=1}^{t} b_j\left(\gamma, \delta\right)\phi_{t-j}(\boldsymbol{\varphi}),
$$

with

$$b_j(\gamma,\delta) = \Delta_+^\delta j^\gamma = \sum_{k=0}^{j-1} \pi_k(-\delta)(j-k)^\gamma . \tag{11}$$

Clearly, the likelihood function (8) is quadratic in $\mu$, so for any given $\boldsymbol{\vartheta}$ we concentrate with respect to $\mu$ and find

$$\widehat{\mu}(\boldsymbol{\vartheta}) = \frac{\sum_{t=1}^T c_t(\gamma,\delta,\boldsymbol{\varphi})\,\phi(L;\boldsymbol{\varphi})\Delta_+^\delta x_t}{\sum_{t=1}^T c_t^2(\gamma,\delta,\boldsymbol{\varphi})}, \tag{12}$$

and we then propose the estimator

$$\widehat{\boldsymbol{\vartheta}} = \arg\min_{\boldsymbol{\vartheta}\in\Xi} R_T(\boldsymbol{\vartheta}), \quad R_T(\boldsymbol{\vartheta}) = \frac{1}{T}\sum_{t=1}^T \left(\phi(L;\boldsymbol{\varphi})\Delta_+^\delta x_t - \widehat{\mu}(\boldsymbol{\vartheta})\,c_t(\gamma,\delta,\boldsymbol{\varphi})\right)^2, \tag{13}$$

along with $\widehat{\mu} = \widehat{\mu}(\widehat{\boldsymbol{\vartheta}})$. Finally, an estimator of $\sigma_0^2 = Var(\varepsilon_t)$ is given by $\widehat{\sigma}^2 = R_T(\widehat{\boldsymbol{\vartheta}})$.

The estimator (13) is often termed nonlinear least squares or conditional sum-of-squares, although we prefer the term truncated sum-of-squares as suggested by Hualde and Robinson (2011). It is motivated by the Gaussian likelihood function (8) and is therefore expected to be asymptotically efficient under Gaussianity (though we do not assume Gaussianity anywhere in the analysis). The estimator goes back to, at least, Box and Jenkins (1971) for estimation of nonfractional ARMA models (where $\delta_0$ is a known integer). In the context of fractional time series, the estimator was first analyzed by Li and McLeod (1986) in stationary FARIMA models with $0 < \delta_0 < 1/2$, Beran (1995), and later by Hualde and Robinson (2011) and Nielsen (2015) for $\delta_0$ lying in arbitrarily large compact intervals.

**Theorem 1** *Let (1), (4), (5), and Assumptions A1–A3 hold.*

**(i)** *If* $\gamma_0 + 1/2 > \delta_0$ *then* $\widehat{\boldsymbol{\vartheta}} \to_p \boldsymbol{\vartheta}_0$ *as* $T\to\infty$.

**(ii)** *If* $\gamma_0 + 1/2 < \delta_0$ *then* $\widehat{\boldsymbol{\tau}} \to_p \boldsymbol{\tau}_0$ *as* $T\to\infty$.

We note that the result in part (i) of Theorem 1 shows consistency of the estimator of the parameter vector $\boldsymbol{\vartheta}_0$. Under (i), $\mu_0$ can also be consistently estimated, but we do not report this result here. Because $\widehat{\mu} = \widehat{\mu}(\widehat{\boldsymbol{\vartheta}})$ is given in explicit form in (12), consistency is not a prerequisite to justifying its limiting distribution, which we provide in Theorem 2. For the same reason we do not report a consistency result for $\widehat{\sigma}^2 = R_T(\widehat{\boldsymbol{\vartheta}})$, which, under the conditions of Theorem 2, can be easily justified using the same techniques as those employed in the proof of that theorem. However, justifying that $\widehat{\mu}$ and $\widehat{\sigma}^2$ are consistent under the milder conditions of Theorem 1 is not immediate. In fact, it may require establishing uniform convergence in probability (in a neighborhood of $\boldsymbol{\vartheta}_0$) of the functionals $\widehat{\mu}(\boldsymbol{\vartheta})$ and $R_T(\boldsymbol{\vartheta})$, which is sufficient by Lemma A.3 of Johansen and Nielsen (2010).

The result in part (ii) only includes consistency of the estimator of $\boldsymbol{\tau}_0$. In fact, $\gamma_0$ and $\mu_0$ cannot possibly be consistently estimated in the case in part (ii), where the deterministic signal is not strong enough. This is easily seen by considering for example $\delta_0 = 1$ (a random walk) in which case the deterministic parameters cannot be estimated consistently when

$\gamma_0 < 1/2$ because the deterministic signal is drowned by the stochastic noise. This is the well-known result that a level ($\gamma_0 = 0$) cannot be estimated consistently for a unit root process ($\delta_0 = 1$), whereas a linear trend ($\gamma_0 = 1$) can be consistently estimated. Another example is $\delta_0 = 0$ (short memory) in which case trends of order $\gamma_0 < -1/2$ cannot be estimated consistently. In other words, suppose the unknown deterministic component is dominated by a constant ($\gamma_0 = 0$), then $\widehat{\gamma}$ and $\widehat{\mu}$ are consistent as long as $\delta_0 < 1/2$, i.e. $z_t$ is (asymptotically) stationary. Thus, a notable feature about part (ii) of Theorem 1 is that, even though $\gamma_0$ and $\mu_0$ cannot be consistently estimated, the remaining parameters $\boldsymbol{\tau}_0$ can still be consistently estimated. Indeed, in part (ii) of Theorem 1 it is not assumed that $\mu_0 \neq 0$; that is, $\mu_0 = 0$ is allowed, in which case $\gamma_0$ is not even identified, but the remaining parameters $\boldsymbol{\tau}_0$ can still be consistently estimated. The intuition here is that when $\mu_0 = 0$ there is no deterministic term in the model, and estimating one does not render the remaining estimates inconsistent.

As in related work, e.g. Hualde and Robinson (2011) and Nielsen (2015), the proof of Theorem 1 is challenging due to non-uniform convergence of the objective function over a large admissible parameter space for $\delta$. However, in addition to this well-known complication, our framework is substantially more involved due to the competition between the stochastic and deterministic components. In our proof of Theorem 1(i), this competition is used explicitly for some parts of the parameter space; for details of the proof strategy please see Subsection S.2.1.1 of the supplementary material Hualde and Nielsen (2019).

Next, we discuss the asymptotic distribution of our estimators, which requires an additional regularity condition.

**A4.** **(i)** $\boldsymbol{\vartheta}_0 \in \mathrm{int}(\Xi)$;

**(ii)** for all $\lambda$, $\omega\left(e^{i\lambda};\boldsymbol{\varphi}\right)$ is thrice continuously differentiable in $\boldsymbol{\varphi}$ on a closed neighbourhood $\mathcal{N}_\epsilon\left(\boldsymbol{\varphi}_0\right)$ of radius $\epsilon \in (0, 1/2)$ about $\boldsymbol{\varphi}_0$, and for all $\boldsymbol{\varphi} \in \mathcal{N}_\epsilon(\boldsymbol{\varphi}_0)$ these partial derivatives with respect to $\boldsymbol{\varphi}$ are themselves differentiable in $\lambda$ with derivative in $\mathrm{Lip}\left(\varsigma\right)$ for $1/2 < \varsigma \leq 1$;

**(iii)** the matrix

$$\boldsymbol{A} = \begin{pmatrix} \pi^2/6 & -\sum_{j=1}^{\infty} \boldsymbol{b}_j'\left(\boldsymbol{\varphi}_0\right)/j \\ -\sum_{j=1}^{\infty} \boldsymbol{b}_j\left(\boldsymbol{\varphi}_0\right)/j & \sum_{j=1}^{\infty} \boldsymbol{b}_j\left(\boldsymbol{\varphi}_0\right)\boldsymbol{b}_j'\left(\boldsymbol{\varphi}_0\right) \end{pmatrix}$$

is nonsingular, where $\boldsymbol{b}_j\left(\boldsymbol{\varphi}_0\right) = \sum_{k=0}^{j-1} \omega_k\left(\boldsymbol{\varphi}_0\right) \partial\phi_{j-k}\left(\boldsymbol{\varphi}_0\right)/\partial\boldsymbol{\varphi}$.

This assumption is almost identical to A3 in Hualde and Robinson (2011), with the only difference that our A4(ii) is slightly stronger than their A3(ii) in imposing thrice instead of twice continuously differentiable $\omega(e^{i\lambda};\boldsymbol{\varphi})$ and corresponding $\mathrm{Lip}\left(\varsigma\right)$ conditions, which appear to be necessary to derive the bounds in (14) below and the corresponding bounds in Hualde and Robinson (2011, p. 3169). The main reason for strengthening the assumption in Hualde and Robinson (2011) is to obtain the bounds (14) and also that, in our proof, third derivatives of $\phi_j(\boldsymbol{\varphi})$ are involved in the proof of convergence of the Hessian matrix below. As in (6), letting $\varphi_i$ denote the $i$-th element of $\boldsymbol{\varphi}$, A1(ii), A1(iv) and A4(ii) imply that, as $j \to \infty$,

$$\sup_{\boldsymbol{\varphi}\in\mathcal{N}_\epsilon(\boldsymbol{\varphi}_0)} \left|\frac{\partial\phi_j(\boldsymbol{\varphi})}{\partial\varphi_i}\right| = O(j^{-1-\varsigma}), \ \sup_{\boldsymbol{\varphi}\in\mathcal{N}_\epsilon(\boldsymbol{\varphi}_0)} \left|\frac{\partial^2\phi_j(\boldsymbol{\varphi})}{\partial\varphi_i\partial\varphi_l}\right| = O(j^{-1-\varsigma}),$$

$$\sup_{\boldsymbol{\varphi}\in\mathcal{N}_\epsilon(\boldsymbol{\varphi}_0)} \left|\frac{\partial^3\phi_j(\boldsymbol{\varphi})}{\partial\varphi_i\partial\varphi_l\partial\varphi_k}\right| = O(j^{-1-\varsigma}), \tag{14}$$

where $\varsigma > 1/2$ is given in A4(ii). Again A4 is satisfied in the ARMA case.

Letting $I_q$ denote the $q$-rowed identity matrix, define the scaling matrix

$$\boldsymbol{P}_T = \mathrm{diag}(I_{p+1}, T^{\delta_0-\gamma_0}, T^{\delta_0-\gamma_0}\log T), \tag{15}$$

and, letting $\mathbf{0}_q$ denote a $q$-dimensional vector of zeros, also define

$$\boldsymbol{V} = \begin{pmatrix} \sigma_0^2\boldsymbol{A} & \mathbf{0}_{p+1} \\ \mathbf{0}'_{p+1} & \frac{\mu_0^2\phi^2(1;\boldsymbol{\varphi}_0)\Gamma^2(\gamma_0+1)}{\Gamma^2(\gamma_0-\delta_0+1)(2(\gamma_0-\delta_0)+1)^3} \end{pmatrix}.$$

**Theorem 2** *Let (1), (4), (5), and Assumptions A1–A4 hold.*

**(i)** *If $\gamma_0 + 1/2 > \delta_0$, then, as $T \to \infty$,*

$$T^{1/2}\boldsymbol{P}_T^{-1}\begin{pmatrix} \widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0 \\ \widehat{\mu} - \mu_0 \end{pmatrix} \to_d \begin{pmatrix} I_{p+2} \\ \mathbf{0}'_{p+1} \;\; -\mu_0 \end{pmatrix}\boldsymbol{N}, \tag{16}$$

*where $\boldsymbol{N}$ is a random variable distributed as $N\left(\mathbf{0}_{p+2}, \sigma_0^2\boldsymbol{V}^{-1}\right)$.*

**(ii)** *If $\gamma_0 + 1/2 < \delta_0$, then, as $T \to \infty$,*

$$T^{1/2}(\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) \to_d N\left(\mathbf{0}_{p+1}, \boldsymbol{A}^{-1}\right). \tag{17}$$

As discussed earlier, an interesting special case arises when $\gamma_0$ is known in (5). For this case, we let $R_T(\boldsymbol{\tau},\gamma) = R_T(\boldsymbol{\vartheta})$, $\widehat{\mu}(\boldsymbol{\tau},\gamma) = \widehat{\mu}(\boldsymbol{\vartheta})$, $\Xi_\gamma = [\triangledown_1, \triangledown_2] \times \Psi$, c.f. (12), (13), and A3, and consider the estimates $\widehat{\boldsymbol{\tau}}_\gamma = \arg\min_{\boldsymbol{\tau}\in\Xi_\gamma} R_T(\boldsymbol{\tau},\gamma_0)$ and $\widehat{\mu}_\gamma = \widehat{\mu}(\widehat{\boldsymbol{\tau}}_\gamma, \gamma_0)$. We state the results corresponding to Theorem 2 for this special case as a corollary. To this end, define

$$\boldsymbol{P}_{\gamma,T} = \mathrm{diag}(I_{p+1}, T^{\delta_0-\gamma_0}) \text{ and } \boldsymbol{V}_\gamma = \begin{pmatrix} \sigma_0^2\boldsymbol{A} & \mathbf{0}_{p+1} \\ \mathbf{0}'_{p+1} & \frac{\phi^2(1;\boldsymbol{\varphi}_0)\Gamma^2(\gamma_0+1)}{\Gamma^2(\gamma_0-\delta_0+1)2(\gamma_0-\delta_0+1/2)} \end{pmatrix}.$$

**Corollary 1** *Let (1), (4), (5), and Assumptions A1–A4 hold, and suppose $\gamma_0$ is known.*

**(i)** *If $\gamma_0 + 1/2 > \delta_0$, then, as $T \to \infty$,*

$$T^{1/2}\boldsymbol{P}_{\gamma,T}^{-1}\begin{pmatrix} \widehat{\boldsymbol{\tau}}_\gamma - \boldsymbol{\tau}_0 \\ \widehat{\mu}_\gamma - \mu_0 \end{pmatrix} \to_d \boldsymbol{N}_\gamma, \tag{18}$$

*where $\boldsymbol{N}_\gamma$ is a random variable distributed as $N\left(\mathbf{0}_{p+2}, \sigma_0^2\boldsymbol{V}_\gamma^{-1}\right)$.*

**(ii)** *If $\gamma_0 + 1/2 < \delta_0$, then, as $T \to \infty$,*

$$T^{1/2}(\widehat{\boldsymbol{\tau}}_\gamma - \boldsymbol{\tau}_0) \to_d N\left(\mathbf{0}_{p+1}, \boldsymbol{A}^{-1}\right). \tag{19}$$

A notable feature of the results in Theorem 2 and Corollary 1 is that the asymptotic distribution of the estimators of $\boldsymbol{\tau}$ is unaffected by the presence of the deterministic component in (5), and $\widehat{\boldsymbol{\tau}}$, $\widehat{\boldsymbol{\tau}}_\gamma$ have the same asymptotic distribution as in, e.g., Theorem 2.2 of Hualde and Robinson (2011). As with the consistency result in Theorem 1, the asymptotic distribution result for $\widehat{\boldsymbol{\tau}}$ in Theorem 2 is also unaffected by the relative magnitudes of the stochastic and deterministic components. In particular, when $\gamma_0+1/2>\delta_0$, the estimates of the parameters of the deterministic and stochastic parts of the model are asymptotically independently distributed, noting also that the factor $\phi^2(1;\boldsymbol{\varphi}_0)$ in $\boldsymbol{V}$ is related to the long run variance of $u_t$. Both these results are common in the case when $\gamma_0$ is known, and we therefore see that estimation of $\gamma_0$ does not alter these findings. Moreover, even when $\gamma_0+1/2<\delta_0$, so that $\gamma_0$ and $\mu_0$ cannot be consistently estimated, the asymptotic distribution of $\widehat{\boldsymbol{\tau}}$, $\widehat{\boldsymbol{\tau}}_\gamma$ is unaffected.

The variance $\boldsymbol{A}^{-1}$ in the asymptotic distribution of $\widehat{\boldsymbol{\tau}}$, $\widehat{\boldsymbol{\tau}}_\gamma$ in (16), (19) is equal to the inverse Fisher information under Gaussianity, see also Dahlhaus (1989) and Hualde and Robinson (2011). Because the estimates $\widehat{\boldsymbol{\tau}}$, $\widehat{\boldsymbol{\tau}}_\gamma$are also asymptotically independent of the remaining coefficient estimates, it therefore follows that $\widehat{\boldsymbol{\tau}}$, $\widehat{\boldsymbol{\tau}}_\gamma$ are asymptotically efficient under the additional assumption of Gaussianity, regardless of the value of $\gamma_0$.

We also remark that $\boldsymbol{A}^{-1}$ does not depend on the true value $\delta_0$ nor on the parameter space of $\delta$ which can be arbitrarily large (see Assumption A3). This finding is expected from the results of Hualde and Robinson (2011) and the block-diagonality of the variance matrix $\boldsymbol{V}^{-1}$, and it is shared with other parametric estimators, see e.g. Dahlhaus (1989) and Hualde and Robinson (2011). However, it is not shared by all parametric estimators that allow arbitrarily large parameter space for $\delta$. For example, the tapered parametric Whittle estimate of Velasco and Robinson (2000) has asymptotic variance that depends on the order of the taper, which in turn must be chosen as a function of the parameter space for $\delta$, and in that sense the asymptotic variance of their estimator depends on the parameter space chosen for $\delta$.

We notice from (16) in Theorem 2 that $\widehat{\gamma}$ is $T^{\gamma_0-\delta_0+1/2}$-consistent whereas $\widehat{\mu}$ is only $T^{\gamma_0-\delta_0+1/2}/\log T$-consistent. On the other hand, when $\gamma_0$ is known as in Corollary 1, then $\widehat{\mu}$ is $T^{\gamma_0-\delta_0+1/2}$-consistent, and hence there is a rate-of-convergence loss, albeit small, in not knowing $\gamma_0$. To conduct inference on $\gamma_0$ or $\mu_0$ using Theorem 2(i) or Corollary 1(i), the previously introduced estimator of $\sigma_0^2$, $\widehat{\sigma}^2=R_T(\widehat{\boldsymbol{\vartheta}})$, is needed.

Finally, the joint asymptotic distribution of $\widehat{\boldsymbol{\vartheta}}$ and $\widehat{\mu}$ given in (16) is singular, i.e. the asymptotic variance matrix is singular, due to the fact that the expansion of $\widehat{\mu}$ (see (S.83)) has a term linear in $\widehat{\gamma}-\gamma_0$ that dominates. Actually, in a somewhat different context, the same singularity problem arises in Robinson (2012). This makes testing joint hypotheses on $\gamma_0$ and $\mu_0$ impossible. However, separate inference can be conducted on $\gamma_0$ and $\mu_0$. For example, it is straightforward given (16) to construct confidence intervals and test hypotheses such as $\gamma_0=0$ (the deterministic term is constant).

## 4 Monte Carlo evidence

With the aim of investigating several finite sample issues we carried out a Monte Carlo experiment. This is divided in three main parts. First we analyze the finite sample performance of the estimates (13). Next, the second part is devoted to illustrating the behaviour of a testing procedure designed to assess whether we face a strong deterministic component situation (with $\gamma_0+1/2>\delta_0$) or not. Because the limiting properties of our proposed estimators rely on the relative strength of the deterministic and stochastic components, this is an important issue. We present a heuristic testing strategy and justify its use by means of several finite sample results covering various scenarios. A formal theoretical treatment of our proposal should be possible, but we do not pursue it here because it would mainly require very lengthy repetitions of the steps and techniques that we have already employed

Table 1: Monte Carlo bias of $\widehat{\delta}$

| | $\mu_0 = -5$ | | | $\mu_0 = 1$ | | | $\mu_0 = 10$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $\gamma_0 - \delta_0 \setminus T$ | 64 | 128 | 256 | 64 | 128 | 256 | 64 | 128 | 256 |
| 0.2 | −0.122 | −0.059 | −0.031 | −0.122 | −0.059 | −0.031 | −0.122 | −0.059 | −0.031 |
| 0.1 | −0.119 | −0.058 | −0.030 | −0.119 | −0.057 | −0.030 | −0.119 | −0.058 | −0.030 |
| 0.0 | −0.115 | −0.056 | −0.029 | −0.115 | −0.056 | −0.029 | −0.115 | −0.056 | −0.029 |
| −0.1 | −0.110 | −0.053 | −0.028 | −0.110 | −0.053 | −0.028 | −0.110 | −0.053 | −0.028 |
| −0.2 | −0.103 | −0.050 | −0.026 | −0.104 | −0.050 | −0.026 | −0.103 | −0.050 | −0.026 |
| −0.3 | −0.096 | −0.046 | −0.024 | −0.098 | −0.047 | −0.025 | −0.096 | −0.046 | −0.024 |
| −0.4 | −0.089 | −0.042 | −0.022 | −0.093 | −0.045 | −0.023 | −0.088 | −0.042 | −0.022 |
| −0.5 | −0.081 | −0.039 | −0.020 | −0.089 | −0.043 | −0.022 | −0.081 | −0.039 | −0.020 |
| −0.6 | −0.073 | −0.035 | −0.018 | −0.086 | −0.042 | −0.022 | −0.073 | −0.035 | −0.018 |
| −0.7 | −0.065 | −0.031 | −0.015 | −0.083 | −0.041 | −0.022 | −0.066 | −0.031 | −0.016 |
| −0.8 | −0.058 | −0.027 | −0.013 | −0.081 | −0.041 | −0.021 | −0.058 | −0.027 | −0.014 |
| −0.9 | −0.051 | −0.023 | −0.012 | −0.079 | −0.040 | −0.021 | −0.051 | −0.023 | −0.011 |
| −1.0 | −0.044 | −0.020 | −0.010 | −0.078 | −0.039 | −0.021 | −0.044 | −0.020 | −0.010 |
| −1.1 | −0.039 | −0.018 | −0.009 | −0.077 | −0.039 | −0.020 | −0.038 | −0.017 | −0.009 |

Note: Based on 10,000 Monte Carlo replications.

to justify the main theoretical results of the paper. Finally, the third part will be devoted to describing the behaviour of the estimator which assumes $\mu_0 = 0$ in (5), which is precisely that considered by Hualde and Robinson (2011) and Nielsen (2015). This estimator omits the presence of possible deterministic terms and it is therefore misspecified whenever $\mu_0 \neq 0$. However, this estimator could be sensible in cases where the deterministic component is dominated by the stochastic one, and, again, without providing a formal analysis, we provide some finite sample evidence which corroborates our conjecture.

In all settings covered by our experiment we generated the observable series $x_t, t = 1, \ldots, T$, from (5) for $T = 64, 128, 256$. Regarding the stochastic component $z_t$, we fix $\delta_0 = 1$ and set $u_t = \varepsilon_t$, where $\varepsilon_t$ is an independent $N(0, 1)$ sequence. Essentially there is no loss of generality in fixing $\delta_0 = 1$: our estimates are approximately invariant to the particular values $\delta_0$, $\gamma_0$, as long as the difference $\gamma_0 - \delta_0$ is kept fixed. The reason for this approximate invariance is that one of the key ingredients of the loss function is the difference operator applied to a time trend, namely $b_t(\gamma, \delta)$, see (11), and this behaves approximately as a multiple of $t^{\gamma-\delta}$, so it is the difference $\gamma - \delta$ (and not the particular values $\gamma$ or $\delta$) that is the relevant quantity. This is confirmed by our experiment, so we do not report results for other values of $\delta_0$, although they are available from the authors upon request.

Regarding the deterministic part, we let $\gamma_0$ and $\mu_0$ vary. In particular, we set $\gamma_0 = \delta_0 - 0.1i$, for $i$ taking all integer values between $-2$ and 11, and $\mu_0 = -5, 1, 10$. Note that $\mu_0$ does not have any effect on the limiting distribution of $\widehat{\boldsymbol{\tau}}$, but it does indeed affect that of $\widehat{\gamma}$. In particular, in view of Theorem 2, it is expected that larger (in absolute value) values of $\mu_0$ lead to better estimates $\widehat{\gamma}$, the intuitive reason being that the deterministic signal is stronger.

We computed $\widehat{\delta}$ and $\widehat{\gamma}$ using the optimizing intervals $\delta \in [\delta_0 - 5, \delta_0 + 5]$ and $\gamma \in [-0.99999, \gamma_0 + 5]$, respectively, and for each we report Monte Carlo bias and standard deviation (SD). All results are based on 10,000 replications.

Results for Monte Carlo bias of $\widehat{\delta}$ and $\widehat{\gamma}$ are presented in Tables 1 and 2, respectively. Here, the performance of $\widehat{\delta}$ reflects the limiting theory developed in Theorems 1 and 2. The

Table 2: Monte Carlo bias of $\widehat{\gamma}$

| | $\mu_0=-5$ | | | $\mu_0=1$ | | | $\mu_0=10$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $\gamma_0-\delta_0 \setminus T$ | 64 | 128 | 256 | 64 | 128 | 256 | 64 | 128 | 256 |
| 0.2 | 0.000 | 0.000 | 0.000 | 0.003 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 |
| 0.1 | 0.000 | 0.000 | 0.000 | 0.007 | 0.002 | 0.001 | 0.000 | 0.000 | 0.000 |
| 0.0 | 0.001 | 0.001 | 0.000 | 0.015 | 0.007 | 0.004 | 0.000 | 0.000 | 0.000 |
| −0.1 | 0.001 | 0.001 | 0.000 | 0.036 | 0.018 | 0.010 | 0.000 | 0.000 | 0.000 |
| −0.2 | 0.002 | 0.001 | 0.001 | 0.088 | 0.046 | 0.026 | 0.000 | 0.000 | 0.000 |
| −0.3 | 0.002 | 0.002 | 0.001 | 0.205 | 0.148 | 0.097 | 0.001 | 0.000 | 0.000 |
| −0.4 | 0.002 | 0.001 | 0.001 | 0.363 | 0.322 | 0.267 | 0.001 | 0.001 | 0.001 |
| −0.5 | −0.001 | −0.001 | −0.001 | 0.496 | 0.502 | 0.437 | 0.000 | 0.000 | 0.000 |
| −0.6 | −0.007 | −0.007 | −0.005 | 0.605 | 0.610 | 0.565 | −0.001 | 0.000 | 0.000 |
| −0.7 | −0.018 | −0.017 | −0.016 | 0.685 | 0.707 | 0.673 | −0.002 | −0.002 | −0.002 |
| −0.8 | −0.033 | −0.032 | −0.029 | 0.744 | 0.780 | 0.740 | −0.005 | −0.004 | −0.005 |
| −0.9 | −0.046 | −0.047 | −0.042 | 0.809 | 0.836 | 0.800 | −0.008 | −0.007 | −0.008 |
| −1.0 | −0.055 | −0.057 | −0.052 | 0.869 | 0.878 | 0.856 | −0.011 | −0.010 | −0.011 |
| −1.1 | −0.059 | −0.062 | −0.057 | 0.940 | 0.946 | 0.906 | −0.013 | −0.012 | −0.014 |

Note: Based on 10,000 Monte Carlo replications.

Table 3: Monte Carlo standard deviation of $\widehat{\delta}$

| | $\mu_0=-5$ | | | $\mu_0=1$ | | | $\mu_0=10$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $\gamma_0-\delta_0 \setminus T$ | 64 | 128 | 256 | 64 | 128 | 256 | 64 | 128 | 256 |
| 0.2 | 0.143 | 0.087 | 0.057 | 0.143 | 0.087 | 0.057 | 0.143 | 0.087 | 0.057 |
| 0.1 | 0.143 | 0.087 | 0.056 | 0.144 | 0.087 | 0.056 | 0.143 | 0.087 | 0.057 |
| 0.0 | 0.144 | 0.087 | 0.057 | 0.144 | 0.087 | 0.057 | 0.144 | 0.087 | 0.057 |
| −0.1 | 0.144 | 0.087 | 0.057 | 0.144 | 0.087 | 0.057 | 0.144 | 0.087 | 0.056 |
| −0.2 | 0.143 | 0.087 | 0.056 | 0.145 | 0.087 | 0.057 | 0.143 | 0.087 | 0.056 |
| −0.3 | 0.142 | 0.086 | 0.056 | 0.144 | 0.087 | 0.057 | 0.142 | 0.086 | 0.056 |
| −0.4 | 0.142 | 0.085 | 0.056 | 0.142 | 0.086 | 0.056 | 0.141 | 0.085 | 0.056 |
| −0.5 | 0.140 | 0.084 | 0.055 | 0.140 | 0.085 | 0.056 | 0.139 | 0.084 | 0.055 |
| −0.6 | 0.138 | 0.084 | 0.055 | 0.138 | 0.084 | 0.055 | 0.138 | 0.083 | 0.055 |
| −0.7 | 0.136 | 0.083 | 0.055 | 0.137 | 0.084 | 0.055 | 0.136 | 0.083 | 0.055 |
| −0.8 | 0.133 | 0.082 | 0.054 | 0.138 | 0.084 | 0.055 | 0.134 | 0.081 | 0.054 |
| −0.9 | 0.130 | 0.080 | 0.053 | 0.138 | 0.084 | 0.055 | 0.131 | 0.080 | 0.053 |
| −1.0 | 0.127 | 0.079 | 0.053 | 0.138 | 0.084 | 0.055 | 0.127 | 0.079 | 0.053 |
| −1.1 | 0.125 | 0.078 | 0.053 | 0.138 | 0.084 | 0.055 | 0.124 | 0.078 | 0.052 |

Note: Based on 10,000 Monte Carlo replications.

bias of $\widehat{\delta}$ is clearly decreasing in absolute value as $T$ increases, even for the boundary case $\gamma_0-\delta_0=-1/2$, which is not covered by our theory. It is also noticeable that when the deterministic signal gets stronger (so $\gamma_0-\delta_0$ is higher) results worsen substantially, although higher values of $|\mu_0|$ improve $\widehat{\delta}$ when $\gamma_0-\delta_0\leq-1/2$. As expected, the behaviour of $\widehat{\gamma}$ is qualitatively different. When $\gamma_0-\delta_0\leq-1/2$ and $\mu_0=1$, the bias of $\widehat{\gamma}$ is very large (in absolute value) and does not decrease as $T$ increases. On the other hand, the picture changes dramatically when $\gamma_0-\delta_0>-1/2$, with very small biases as $\gamma_0-\delta_0$ gets larger, reflecting the fast convergence rates in those cases implied by Theorem 2. As anticipated, higher values of $|\mu_0|$ lead to smaller bias, although when $\gamma_0-\delta_0\leq-1/2$ biases do not decrease as $T$ increases.

Table 4: Monte Carlo standard deviation of $\widehat{\gamma}$

| | $\mu_0 = -5$ | | | $\mu_0 = 1$ | | | $\mu_0 = 10$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $\gamma_0 - \delta_0 \setminus T$ | 64 | 128 | 256 | 64 | 128 | 256 | 64 | 128 | 256 |
| 0.2 | 0.015 | 0.009 | 0.006 | 0.077 | 0.046 | 0.029 | 0.008 | 0.005 | 0.003 |
| 0.1 | 0.020 | 0.013 | 0.009 | 0.112 | 0.066 | 0.043 | 0.010 | 0.007 | 0.004 |
| 0.0 | 0.026 | 0.018 | 0.013 | 0.150 | 0.094 | 0.064 | 0.013 | 0.009 | 0.006 |
| −0.1 | 0.033 | 0.024 | 0.018 | 0.248 | 0.150 | 0.095 | 0.016 | 0.012 | 0.009 |
| −0.2 | 0.040 | 0.030 | 0.024 | 0.534 | 0.317 | 0.178 | 0.020 | 0.015 | 0.012 |
| −0.3 | 0.047 | 0.038 | 0.031 | 0.938 | 0.761 | 0.552 | 0.023 | 0.019 | 0.015 |
| −0.4 | 0.057 | 0.046 | 0.039 | 1.295 | 1.207 | 1.106 | 0.027 | 0.023 | 0.019 |
| −0.5 | 0.071 | 0.060 | 0.051 | 1.508 | 1.505 | 1.435 | 0.033 | 0.028 | 0.025 |
| −0.6 | 0.100 | 0.091 | 0.091 | 1.619 | 1.636 | 1.599 | 0.042 | 0.037 | 0.034 |
| −0.7 | 0.142 | 0.128 | 0.132 | 1.660 | 1.681 | 1.670 | 0.055 | 0.050 | 0.047 |
| −0.8 | 0.190 | 0.180 | 0.178 | 1.679 | 1.707 | 1.692 | 0.073 | 0.068 | 0.065 |
| −0.9 | 0.235 | 0.227 | 0.225 | 1.694 | 1.712 | 1.702 | 0.093 | 0.088 | 0.087 |
| −1.0 | 0.269 | 0.266 | 0.268 | 1.698 | 1.696 | 1.691 | 0.115 | 0.110 | 0.109 |
| −1.1 | 0.302 | 0.295 | 0.300 | 1.687 | 1.693 | 1.677 | 0.136 | 0.132 | 0.132 |

Note: Based on 10,000 Monte Carlo replications.

The Monte Carlo standard deviation (SD) of $\widehat{\delta}$ and of $\widehat{\gamma}$ are reported in Tables 3 and 4, respectively. Results for $\widehat{\delta}$ are as expected, but now larger $\gamma_0 - \delta_0$ lead to slightly larger SD. Also, as the theory predicts, results for $\widehat{\delta}$ are hardly affected by the value of $\mu_0$. Regarding $\widehat{\gamma}$, for $\gamma_0 - \delta_0 \leq -1/2$ and $\mu_0 = 1$ the SD of $\widehat{\gamma}$ is very large and quite stable for different values of $T$. As anticipated from Theorem 2, for $\gamma_0 - \delta_0 > -1/2$ the SD of $\widehat{\gamma}$ decreases with $T$ and is very small for larger $\gamma_0 - \delta_0$ due to the rate of convergence $T^{\gamma_0 - \delta_0 + 1/2}$. Note also that the SD of $\widehat{\gamma}$ for $\mu_0 = 1$ and $\gamma_0 - \delta_0 = 0$ (so $\widehat{\gamma}$ has the standard $T^{1/2}$ rate) is comparable to that corresponding the the $T^{1/2}$-consistent estimator $\widehat{\delta}$ (see Table 3). A similar pattern occurs for $\mu_0 = -5$ and $\mu_0 = 10$, although when $|\mu_0|$ increases the SD of $\widehat{\gamma}$ is very small. This is due to the effect of $\mu_0$ on the limiting distribution of $\widehat{\gamma}$ in the (2,2) block of $\boldsymbol{V}$ and (16). Roughly speaking, this effect implies that the SD of $\widehat{\gamma}$ for cases $\mu_0 = -5$, 10, should be 5 and 10 times smaller than the corresponding one for $\mu_0 = 1$. This effect can be clearly observed in Table 4 for larger values of $\gamma_0 - \delta_0$, the effect being less clear when $\gamma_0 - \delta_0$ is closer to (but above) $-1/2$. In the latter case, the SD of $\widehat{\gamma}$ for $\mu_0 = 1$ is pretty large, the $\widehat{\gamma}$ estimator being quite unreliable here due to the slow rate of convergence.

We next examine the performance of a one-sided Lagrange multiplier (LM) testing procedure which tests the boundary condition $H_0 : \gamma_0 - \delta_0 = -1/2$ against the one-sided alternative $H_1 : \gamma_0 - \delta_0 > -1/2$. This testing strategy is mainly based on the restricted estimator $\widetilde{\boldsymbol{\vartheta}} = \arg\min_{\boldsymbol{\vartheta} \in \widetilde{\Xi}} R_T(\boldsymbol{\vartheta})$, where $\widetilde{\Xi} = \{\boldsymbol{\vartheta} \in \Xi, \gamma - \delta = -1/2\}$, and on the LM statistic

$$LM = \frac{T}{2\widetilde{\sigma}^2} \frac{\partial R_T(\widetilde{\boldsymbol{\vartheta}})}{\partial \boldsymbol{\vartheta}'} \left( \frac{\partial^2 R_T(\widetilde{\boldsymbol{\vartheta}})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \right)^{-1} \frac{\partial R_T(\widetilde{\boldsymbol{\vartheta}})}{\partial \boldsymbol{\vartheta}},$$

where $\widetilde{\sigma}^2 = R_T(\widetilde{\boldsymbol{\vartheta}})$. As mentioned before, we do not pursue here a formal analysis, but we conjecture that, as $T \to \infty$,

$$LM \to_d \chi_1^2 \qquad \text{under } H_0. \tag{20}$$

Table 5: Monte Carlo bias of $\widetilde{\delta}$

| | $\mu_0 = -5$ | | | $\mu_0 = 1$ | | | $\mu_0 = 10$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $\gamma_0 - \delta_0 \setminus T$ | 64 | 128 | 256 | 64 | 128 | 256 | 64 | 128 | 256 |
| 0.2 | 0.635 | 0.617 | 0.599 | 0.414 | 0.406 | 0.398 | 0.678 | 0.668 | 0.656 |
| 0.1 | 0.538 | 0.521 | 0.503 | 0.323 | 0.315 | 0.307 | 0.579 | 0.570 | 0.558 |
| 0.0 | 0.442 | 0.426 | 0.408 | 0.236 | 0.228 | 0.220 | 0.481 | 0.472 | 0.461 |
| −0.1 | 0.347 | 0.332 | 0.315 | 0.154 | 0.148 | 0.141 | 0.383 | 0.375 | 0.364 |
| −0.2 | 0.253 | 0.239 | 0.224 | 0.080 | 0.078 | 0.073 | 0.285 | 0.278 | 0.269 |
| −0.3 | 0.161 | 0.150 | 0.138 | 0.021 | 0.024 | 0.024 | 0.188 | 0.183 | 0.175 |
| −0.4 | 0.072 | 0.066 | 0.059 | −0.018 | −0.007 | −0.002 | 0.092 | 0.089 | 0.084 |
| −0.5 | −0.012 | −0.009 | −0.007 | −0.037 | −0.020 | −0.011 | −0.003 | −0.003 | −0.002 |
| −0.6 | −0.086 | −0.066 | −0.046 | −0.044 | −0.024 | −0.013 | −0.096 | −0.089 | −0.079 |
| −0.7 | −0.139 | −0.094 | −0.056 | −0.047 | −0.025 | −0.014 | −0.186 | −0.166 | −0.134 |
| −0.8 | −0.164 | −0.097 | −0.052 | −0.047 | −0.025 | −0.013 | −0.266 | −0.216 | −0.149 |
| −0.9 | −0.169 | −0.092 | −0.046 | −0.047 | −0.024 | −0.013 | −0.322 | −0.231 | −0.139 |
| −1.0 | −0.165 | −0.085 | −0.042 | −0.046 | −0.024 | −0.013 | −0.346 | −0.222 | −0.123 |
| −1.1 | −0.159 | −0.080 | −0.039 | −0.046 | −0.024 | −0.013 | −0.347 | −0.208 | −0.112 |

Note: Based on 10,000 Monte Carlo replications.

Justifying (20) formally requires the use of the proof techniques employed in the proofs of Theorems 1 and 2. The result (20) is crucial for our testing strategy but, unfortunately, this test statistic by itself is not informative enough: as expected, $LM$ tends to take large values whenever $\gamma_0 - \delta_0 > -1/2$ or $\gamma_0 - \delta_0 < -1/2$, so rejecting $H_0$ based solely on $LM$ is not conclusive about whether we face a strong or weak deterministic component situation, with $\gamma_0 - \delta_0 > -1/2$ or $\gamma_0 - \delta_0 < -1/2$, respectively. We implement a solution to this problem based on the behaviour of $\widehat{\delta}$ and $\widetilde{\delta}$, that is the unrestricted and restricted estimators of $\delta_0$, respectively. As illustrated in Tables 1 and 3, our Monte Carlo evidence indicates that $\widehat{\delta}$ is consistent regardless of whether or not $H_0$ is true; see also Theorem 1. On the contrary, our Monte Carlo evidence indicates that $\widetilde{\delta}$ overestimates $\delta_0$ if $\gamma_0 - \delta_0 > -1/2$, but underestimates $\delta_0$ if $\gamma_0 - \delta_0 < -1/2$. This is illustrated in Tables 5 and 6, where we report Monte Carlo bias and SD of $\widetilde{\delta}$, respectively. Here, it is clear that the positive and negative bias of $\widetilde{\delta}$ is exacerbated for larger $|\mu_0|$. It is also noticeable that the overestimation effect when $\gamma_0 - \delta_0 > -1/2$ is stronger than the underestimation effect when $\gamma_0 - \delta_0 < -1/2$. As expected, under $H_0$, both $\widehat{\delta}$ and $\widetilde{\delta}$ appear to be consistent.

This evidence suggests the following testing strategy, which can be named as a one-sided $LM$ test:

$$\text{Reject } H_0 \text{ whenever } \widetilde{\delta} - \widehat{\delta} > 0 \text{ and } LM > cv_\alpha,$$

where $cv_\alpha$ is the critical value from the $\chi_1^2$ distribution corresponding to significance level $\alpha$. Note that this test is correctly sized in the sense that

$$\Pr(LM > cv_\alpha, \widetilde{\delta} - \widehat{\delta} > 0) \leq \Pr(LM > cv_\alpha) \to \alpha \text{ as } T \to \infty, \text{ under } H_0,$$

see (20). Thus, asymptotically, the probability of rejection under the null is bounded by the nominal size, although our test could be conservative. Unfortunately, given that the dependence between $LM$ and $\widetilde{\delta} - \widehat{\delta}$ is complex, obtaining the exact limit of $\Pr(LM > cv_\alpha, \widetilde{\delta} - \widehat{\delta} > 0)$ under $H_0$ seems to be very challenging. However, at least our testing strategy

Table 6: Monte Carlo standard deviation of $\widetilde{\delta}$

| | $\mu_0 = -5$ | | | $\mu_0 = 1$ | | | $\mu_0 = 10$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $\gamma_0 - \delta_0 \setminus T$ | 64 | 128 | 256 | 64 | 128 | 256 | 64 | 128 | 256 |
| 0.2 | 0.024 | 0.016 | 0.011 | 0.037 | 0.023 | 0.014 | 0.016 | 0.012 | 0.009 |
| 0.1 | 0.025 | 0.017 | 0.012 | 0.041 | 0.026 | 0.016 | 0.017 | 0.013 | 0.009 |
| 0.0 | 0.027 | 0.019 | 0.013 | 0.048 | 0.031 | 0.020 | 0.018 | 0.014 | 0.010 |
| −0.1 | 0.029 | 0.021 | 0.014 | 0.059 | 0.038 | 0.026 | 0.019 | 0.014 | 0.011 |
| −0.2 | 0.032 | 0.023 | 0.017 | 0.075 | 0.050 | 0.034 | 0.020 | 0.016 | 0.012 |
| −0.3 | 0.037 | 0.028 | 0.020 | 0.090 | 0.061 | 0.043 | 0.022 | 0.017 | 0.014 |
| −0.4 | 0.045 | 0.034 | 0.027 | 0.101 | 0.069 | 0.048 | 0.025 | 0.020 | 0.017 |
| −0.5 | 0.056 | 0.045 | 0.036 | 0.107 | 0.072 | 0.051 | 0.030 | 0.026 | 0.022 |
| −0.6 | 0.072 | 0.059 | 0.046 | 0.111 | 0.074 | 0.052 | 0.039 | 0.034 | 0.030 |
| −0.7 | 0.092 | 0.073 | 0.054 | 0.112 | 0.075 | 0.052 | 0.053 | 0.047 | 0.043 |
| −0.8 | 0.110 | 0.083 | 0.058 | 0.113 | 0.075 | 0.052 | 0.071 | 0.067 | 0.062 |
| −0.9 | 0.126 | 0.089 | 0.060 | 0.113 | 0.075 | 0.052 | 0.097 | 0.091 | 0.074 |
| −1.0 | 0.136 | 0.091 | 0.059 | 0.114 | 0.075 | 0.052 | 0.127 | 0.109 | 0.078 |
| −1.1 | 0.140 | 0.091 | 0.059 | 0.114 | 0.075 | 0.052 | 0.153 | 0.117 | 0.077 |

Note: Based on 10,000 Monte Carlo replications.

ensures that the probability of rejection is asymptotically bounded by $\alpha$ under the null, while, as our results below show, the test has power under the alternative.

Tables 7–9 present the proportion of rejections (out of 10,000 replications) corresponding to this testing strategy for $\mu_0 = 1$, $\mu_0 = -5$, and $\mu_0 = 10$, respectively, and in each case for nominal sizes $\alpha = 0.01, 0.05, 0.10$. Overall, the performance of our proposal is very satisfactory. First, under $H_1$, as expected, the test shows higher power as $\gamma_0 - \delta_0$ and/or $T$ get larger, results being substantially better as $|\mu_0|$ increases. Second, under $H_0$ the size behaviour is adequate, although some undersizing occurs when the deterministic signal is weak ($\mu_0 = 1$). However this is corrected as $|\mu_0|$ increases. Nicely, when $\gamma_0 - \delta_0 < -1/2$ the proportion of rejections is in all cases smaller than the nominal size.

Finally, the last part of our Monte Carlo evidence is devoted to analyzing the behaviour of the estimator studied by Hualde and Robinson (2011), which omits the presence of possible deterministic terms. This is defined as

$$\overline{\boldsymbol{\tau}} = \arg\min_{\boldsymbol{\tau}\in\Upsilon} Q_T(\boldsymbol{\tau}), \quad Q_T(\boldsymbol{\tau}) = \frac{1}{T}\sum_{t=1}^{T}\left(\phi(L;\boldsymbol{\varphi})\Delta_+^{\delta}x_t\right)^2,$$

where $\Upsilon = [\bigtriangledown_1, \bigtriangledown_2] \times \Psi$, so that, when $\mu_0 \neq 0$, the estimate $\overline{\boldsymbol{\tau}}$ is based on a misspecified loss function. In Tables 10 and 11 we report results for Monte Carlo bias and SD, respectively, of $\overline{\delta}$ for the same situations as before. As anticipated, $\overline{\delta}$ displays large bias whenever the deterministic component is strong, i.e. when $\gamma_0 - \delta_0 > -1/2$, with this effect being more noticeable for larger $|\mu_0|$. However, when $\gamma_0 - \delta_0 < -1/2$, the bias is generally small and we conjecture that $\overline{\delta}$ is consistent in this case. The SD behaves in an opposite way to the bias: as $\gamma_0 - \delta_0$ decreases, the SD gets larger with smaller values of $|\mu_0|$ also increasing the SD. However, in all cases, the SD decreases as $T$ increases.

To conclude our analysis, in Table 12 we compare the Monte Carlo bias and SD of the estimates $\widehat{\delta}$ and $\overline{\delta}$ when $\mu_0 = 0$. Note that, if there is no deterministic component, $\overline{\delta}$ is based on the correct information $\mu_0 = 0$, so it is not misspecified and could have some

Table 7: Proportion of rejections of the one-sided LM test, $\mu_0 = 1$

| | $\alpha = 0.01$ | | | $\alpha = 0.05$ | | | $\alpha = 0.10$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $\gamma_0 - \delta_0 \setminus T$ | 64 | 128 | 256 | 64 | 128 | 256 | 64 | 128 | 256 |
| 0.2 | 0.034 | 0.628 | 0.998 | 0.633 | 0.988 | 1.000 | 0.906 | 0.998 | 1.000 |
| 0.1 | 0.019 | 0.329 | 0.975 | 0.419 | 0.938 | 0.999 | 0.770 | 0.987 | 1.000 |
| 0.0 | 0.010 | 0.125 | 0.767 | 0.242 | 0.742 | 0.990 | 0.561 | 0.928 | 0.998 |
| −0.1 | 0.005 | 0.035 | 0.285 | 0.120 | 0.399 | 0.850 | 0.332 | 0.699 | 0.958 |
| −0.2 | 0.002 | 0.007 | 0.042 | 0.054 | 0.132 | 0.360 | 0.167 | 0.337 | 0.631 |
| −0.3 | 0.001 | 0.001 | 0.003 | 0.020 | 0.033 | 0.068 | 0.074 | 0.109 | 0.184 |
| −0.4 | 0.000 | 0.000 | 0.000 | 0.008 | 0.008 | 0.010 | 0.033 | 0.032 | 0.039 |
| −0.5 | 0.000 | 0.000 | 0.000 | 0.003 | 0.003 | 0.003 | 0.017 | 0.013 | 0.011 |
| −0.6 | 0.000 | 0.000 | 0.000 | 0.002 | 0.001 | 0.001 | 0.009 | 0.006 | 0.004 |
| −0.7 | 0.000 | 0.000 | 0.000 | 0.001 | 0.001 | 0.001 | 0.006 | 0.004 | 0.003 |
| −0.8 | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 | 0.001 | 0.004 | 0.002 | 0.003 |
| −0.9 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 | 0.003 | 0.002 | 0.003 |
| −1.0 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 | 0.003 | 0.002 | 0.003 |
| −1.1 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.002 | 0.002 | 0.002 |

Note: Based on 10,000 Monte Carlo replications.

Table 8: Proportion of rejections of the one-sided LM test, $\mu_0 = -5$

| | $\alpha = 0.01$ | | | $\alpha = 0.05$ | | | $\alpha = 0.10$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $\gamma_0 - \delta_0 \setminus T$ | 64 | 128 | 256 | 64 | 128 | 256 | 64 | 128 | 256 |
| 0.2 | 0.956 | 1.000 | 1.000 | 0.997 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 |
| 0.1 | 0.890 | 0.999 | 1.000 | 0.987 | 1.000 | 1.000 | 0.996 | 1.000 | 1.000 |
| 0.0 | 0.755 | 0.991 | 1.000 | 0.952 | 1.000 | 1.000 | 0.985 | 1.000 | 1.000 |
| −0.1 | 0.551 | 0.938 | 1.000 | 0.851 | 0.993 | 1.000 | 0.933 | 0.998 | 1.000 |
| −0.2 | 0.302 | 0.699 | 0.971 | 0.655 | 0.923 | 0.998 | 0.797 | 0.969 | 1.000 |
| −0.3 | 0.119 | 0.308 | 0.650 | 0.385 | 0.625 | 0.884 | 0.553 | 0.776 | 0.945 |
| −0.4 | 0.036 | 0.063 | 0.127 | 0.164 | 0.230 | 0.352 | 0.277 | 0.360 | 0.497 |
| −0.5 | 0.008 | 0.007 | 0.006 | 0.053 | 0.043 | 0.038 | 0.101 | 0.085 | 0.078 |
| −0.6 | 0.001 | 0.001 | 0.000 | 0.014 | 0.006 | 0.001 | 0.031 | 0.014 | 0.005 |
| −0.7 | 0.000 | 0.000 | 0.001 | 0.004 | 0.002 | 0.003 | 0.010 | 0.003 | 0.004 |
| −0.8 | 0.000 | 0.000 | 0.003 | 0.001 | 0.004 | 0.014 | 0.005 | 0.007 | 0.020 |
| −0.9 | 0.000 | 0.001 | 0.005 | 0.002 | 0.008 | 0.025 | 0.004 | 0.016 | 0.041 |
| −1.0 | 0.000 | 0.001 | 0.004 | 0.002 | 0.012 | 0.027 | 0.006 | 0.023 | 0.044 |
| −1.1 | 0.000 | 0.000 | 0.004 | 0.002 | 0.010 | 0.025 | 0.008 | 0.023 | 0.045 |

Note: Based on 10,000 Monte Carlo replications.

finite sample advantages over the less parsimonious $\widehat{\delta}$. However, in view of Theorem 2, this finite sample advantage should disappear asymptotically. As expected, the behaviour of $\widehat{\delta}$ is similar to that displayed in Tables 1 and 3 for other values of $\mu_0$. Also, as expected, $\overline{\delta}$ outperforms $\widehat{\delta}$, although the relative difference decreases as $T$ increases. However, keeping in mind the risk of misspecification, it appears that the price to pay due to estimating the inexistent deterministic component outweighs the severe consequences of misspecification seen in Tables 10 and 11. Although these results are not reported, when $\mu_0 = 0$, $\widehat{\gamma}$ behaves in a completely unpredictable way, displaying huge SD. Thus, the particular case $\mu_0 = 0$ raises a serious concern, that is, acting as if $\mu_0 \neq 0$ with the possibility of taking $\widehat{\gamma}$

Table 9: Proportion of rejections of the one-sided LM test, $\mu_0 = 10$

| | $\alpha = 0.01$ | | | $\alpha = 0.05$ | | | $\alpha = 0.10$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $\gamma_0 - \delta_0 \setminus T$ | 64 | 128 | 256 | 64 | 128 | 256 | 64 | 128 | 256 |
| 0.2 | 0.981 | 1.000 | 1.000 | 0.998 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 0.1 | 0.947 | 1.000 | 1.000 | 0.993 | 1.000 | 1.000 | 0.998 | 1.000 | 1.000 |
| 0.0 | 0.855 | 0.997 | 1.000 | 0.972 | 1.000 | 1.000 | 0.990 | 1.000 | 1.000 |
| −0.1 | 0.664 | 0.976 | 1.000 | 0.901 | 0.997 | 1.000 | 0.958 | 0.999 | 1.000 |
| −0.2 | 0.411 | 0.830 | 0.993 | 0.721 | 0.962 | 1.000 | 0.844 | 0.985 | 1.000 |
| −0.3 | 0.185 | 0.440 | 0.824 | 0.446 | 0.730 | 0.954 | 0.605 | 0.850 | 0.981 |
| −0.4 | 0.056 | 0.112 | 0.221 | 0.201 | 0.292 | 0.476 | 0.319 | 0.431 | 0.619 |
| −0.5 | 0.013 | 0.012 | 0.010 | 0.063 | 0.053 | 0.047 | 0.120 | 0.101 | 0.090 |
| −0.6 | 0.002 | 0.001 | 0.000 | 0.014 | 0.004 | 0.001 | 0.032 | 0.011 | 0.003 |
| −0.7 | 0.001 | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 | 0.007 | 0.000 | 0.000 |
| −0.8 | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 |
| −0.9 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| −1.0 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.001 |
| −1.1 | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 | 0.002 |

Note: Based on 10,000 Monte Carlo replications.

Table 10: Monte Carlo bias of $\overline{\delta}$

| | $\mu_0 = -5$ | | | $\mu_0 = 1$ | | | $\mu_0 = 10$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $\gamma_0 - \delta_0 \setminus T$ | 64 | 128 | 256 | 64 | 128 | 256 | 64 | 128 | 256 |
| 0.2 | 0.863 | 0.796 | 0.739 | 0.499 | 0.475 | 0.455 | 1.025 | 0.948 | 0.877 |
| 0.1 | 0.765 | 0.698 | 0.642 | 0.407 | 0.383 | 0.363 | 0.921 | 0.847 | 0.779 |
| 0.0 | 0.668 | 0.604 | 0.548 | 0.318 | 0.295 | 0.275 | 0.818 | 0.749 | 0.682 |
| −0.1 | 0.575 | 0.511 | 0.456 | 0.236 | 0.213 | 0.193 | 0.718 | 0.653 | 0.588 |
| −0.2 | 0.484 | 0.422 | 0.367 | 0.158 | 0.139 | 0.121 | 0.619 | 0.559 | 0.496 |
| −0.3 | 0.397 | 0.336 | 0.283 | 0.093 | 0.078 | 0.064 | 0.525 | 0.467 | 0.407 |
| −0.4 | 0.312 | 0.256 | 0.206 | 0.046 | 0.035 | 0.028 | 0.433 | 0.379 | 0.322 |
| −0.5 | 0.235 | 0.183 | 0.138 | 0.018 | 0.012 | 0.009 | 0.345 | 0.296 | 0.244 |
| −0.6 | 0.164 | 0.119 | 0.083 | 0.000 | 0.001 | 0.001 | 0.260 | 0.217 | 0.172 |
| −0.7 | 0.102 | 0.069 | 0.044 | −0.007 | −0.003 | −0.001 | 0.182 | 0.148 | 0.110 |
| −0.8 | 0.052 | 0.033 | 0.019 | −0.011 | −0.006 | −0.003 | 0.112 | 0.086 | 0.060 |
| −0.9 | 0.013 | 0.007 | 0.004 | −0.016 | −0.009 | −0.004 | 0.047 | 0.034 | 0.023 |
| −1.0 | −0.012 | −0.007 | −0.004 | −0.017 | −0.008 | −0.004 | −0.007 | −0.005 | −0.003 |
| −1.1 | −0.033 | −0.018 | −0.010 | −0.019 | −0.008 | −0.004 | −0.053 | −0.035 | −0.022 |

Note: Based on 10,000 Monte Carlo replications.

seriously. However, fortunately, when $\mu_0 = 0$, $LM$ takes very small values (specifically, the proportion of rejections when applying our testing strategy for sizes $\alpha = 0.01, 0.05, 0.10$ is $(0.000, 0.001, 0.003)$, $(0.000, 0.000, 0.002)$, and $(0.000, 0.000, 0.001)$, for $T = 64, 128, 256$, respectively), so in this case $H_0$ would hardly ever be rejected in favour of $H_1$. In other words, the probability of mistakenly believing that the estimates of the deterministic component are accurate, when it in fact does not exist, is very small.

Table 11: Monte Carlo standard deviation of $\overline{\delta}$

| | $\mu_0 = -5$ | | | $\mu_0 = 1$ | | | $\mu_0 = 10$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $\gamma_0 - \delta_0 \setminus T$ | 64 | 128 | 256 | 64 | 128 | 256 | 64 | 128 | 256 |
| 0.2 | 0.033 | 0.021 | 0.013 | 0.034 | 0.021 | 0.013 | 0.034 | 0.022 | 0.014 |
| 0.1 | 0.033 | 0.021 | 0.013 | 0.038 | 0.023 | 0.014 | 0.033 | 0.023 | 0.014 |
| 0.0 | 0.034 | 0.022 | 0.014 | 0.044 | 0.028 | 0.018 | 0.033 | 0.022 | 0.015 |
| −0.1 | 0.036 | 0.023 | 0.015 | 0.055 | 0.035 | 0.023 | 0.034 | 0.023 | 0.015 |
| −0.2 | 0.038 | 0.025 | 0.016 | 0.071 | 0.047 | 0.032 | 0.033 | 0.024 | 0.016 |
| −0.3 | 0.041 | 0.028 | 0.020 | 0.087 | 0.060 | 0.042 | 0.034 | 0.025 | 0.017 |
| −0.4 | 0.046 | 0.034 | 0.025 | 0.098 | 0.068 | 0.047 | 0.036 | 0.027 | 0.020 |
| −0.5 | 0.055 | 0.041 | 0.031 | 0.103 | 0.070 | 0.050 | 0.038 | 0.030 | 0.023 |
| −0.6 | 0.064 | 0.051 | 0.039 | 0.106 | 0.072 | 0.050 | 0.042 | 0.035 | 0.028 |
| −0.7 | 0.075 | 0.058 | 0.044 | 0.106 | 0.072 | 0.050 | 0.049 | 0.041 | 0.034 |
| −0.8 | 0.082 | 0.062 | 0.046 | 0.108 | 0.073 | 0.050 | 0.055 | 0.046 | 0.038 |
| −0.9 | 0.088 | 0.065 | 0.048 | 0.107 | 0.072 | 0.051 | 0.060 | 0.051 | 0.041 |
| −1.0 | 0.090 | 0.067 | 0.048 | 0.108 | 0.073 | 0.050 | 0.064 | 0.054 | 0.043 |
| −1.1 | 0.092 | 0.067 | 0.048 | 0.107 | 0.072 | 0.050 | 0.068 | 0.056 | 0.044 |

Note: Based on 10,000 Monte Carlo replications.

Table 12: Monte Carlo bias and standard deviation of $\widehat{\delta}$ and $\overline{\delta}$, $\mu_0 = 0$

| | bias | | | standard devation | | |
|---|---|---|---|---|---|---|
| $\delta \setminus T$ | 64 | 128 | 256 | 64 | 128 | 256 |
| $\widehat{\delta}$ | −0.095 | −0.048 | −0.025 | 0.138 | 0.085 | 0.056 |
| $\overline{\delta}$ | −0.017 | −0.009 | −0.005 | 0.108 | 0.072 | 0.051 |

Note: Based on 10,000 Monte Carlo replications.

## 5 Concluding remarks

We have proposed and analyzed a parametric model which covers a wide range of situations characterized by general deterministic and stochastic components. These are mainly driven by power law and memory parameters, $\gamma_0$ and $\delta_0$, respectively, which are assumed to lie in sets which can be arbitrarily large. Our model might display many different behaviours, including "stochastic trend in mean and/or variance" and various types of dependence (antipersistency, weak dependence, long memory). Our results depend crucially on whether the deterministic signal is sufficiently strong. If this is the case, that is if $\gamma_0 + 1/2 > \delta_0$, all parameters can be consistently estimated and their estimators are asymptotically normal. Interestingly, the limiting results for estimators corresponding to the stochastic part of the model ($\widehat{\boldsymbol{\tau}}$) are identical to those achieved in the simpler, purely stochastic, setting of Hualde and Robinson (2011). When the deterministic signal is weak, i.e., $\gamma_0 + 1/2 < \delta_0$, $\gamma_0$ and $\mu_0$ cannot be consistently estimated, but $\widehat{\boldsymbol{\tau}}$ retains identical limiting properties as when $\gamma_0 + 1/2 > \delta_0$.

There are several interesting issues which have not been addressed in the present paper, but which will the object of future research. First, one could argue that the deterministic part of our model, which contains a single term, is too simplistic. However, our methods of proof should be extendable to cover a richer setting, allowing for multiple deterministic terms characterized by different power law parameters, such as

$$x_t = \sum_{j=1}^{d} \mu_{0j} t_+^{\gamma_{0j}} + z_t, \tag{21}$$

where $z_t$ is given in (1), (4), and, without loss of generality, we set $-1 < \gamma_{01} < \gamma_{02} < \cdots < \gamma_{0d} < \infty$. Whenever it exists, let $\dagger = \min\{j = 1, \ldots, d : \gamma_{0j} + 1/2 > \delta_0\}$ and $\mu_{0j} \neq 0$ for any $j \geq \dagger$. Our estimator can be extended to accommodate this greater generality in an obvious way, and we conjecture that results qualitatively identical to those in Theorem 2 apply. In particular, the estimator of $\boldsymbol{\tau}_0$ would retain identical properties irrespective of the strength of the deterministic signal(s), and whenever $\dagger$ exists, the estimators of $\gamma_{0j}$ for $j \geq \dagger$ will be $T^{\gamma_{0j}-\delta_0+1/2}$-consistent and asymptotically normal. In the setup of (21), as discussed after (5), letting one term have $\gamma_{0j} = 0$ known, the corresponding $\mu_{0j}$ would be the so-called "level parameter", the estimation of which can alleviate bias issues arising from non-zero initial conditions as analyzed by Johansen and Nielsen (2016). However, considering formally this extension would come at the cost of greater complication, and given that our present setting is already quite involved, we preferred to keep things as simple as possible at this stage, so the proofs present in a clear way the essence of the problem of the competition between deterministic and stochastic terms.

Second, a semiparametric approach which focuses on estimating $\gamma_0$ and $\delta_0$ without making parametric assumptions about the structure of $z_t$ seems possible and interesting. Third, the fractional process which characterizes our model has been termed as "Type II". Nevertheless, it seems that our theory could also be developed for the so-called "Type I" fractional process. Finally, a formal treatment of a testing procedure which, like our heuristic proposal, is designed to assess the relative strength of deterministic and stochastic components in a general setting like (21) seems relevant. This is likely possible, but is beyond the scope of the present paper and it will be the object of future research.

## 6 References

1. Abadir, K.M., Distaso, W. and Giraitis, L. (2007). Nonstationarity-extended local Whittle estimation. *Journal of Econometrics* **141**, 1353–1384.
2. Beran, J. (1995). Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *Journal of the Royal Statistical Society, Series B* **57**, 659–672.
3. Bloomfield, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika* **60**, 217–226.
4. Box, G.E.P. and Jenkins, G.M. (1971). *Time Series Analysis, Forecasting and Control.* Holden-Day, San Fracisco.
5. Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *Annals of Statistics* **17**, 1749–1766.
6. Fox, R. and Taqqu, M.S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Annals of Statistics* **14**, 517–532.
7. Giraitis, L. and Surgailis, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate. *Probability Theory and Related Fields* **86**, 87–104.

8. Hannan, E.J. (1973). The asymptotic theory of linear time series models. *Journal of Applied Probability* **10**, 130–145.
9. Heyde, C.C. and Dai, W. (1996). On the robustness to small trends of estimation based on the smoothed periodogram. *Journal of Time Series Analysis* **17**, 141–150.
10. Hualde, J. and Nielsen, M.Ø. (2019). Supplementary appendix to “Truncated sum of squares estimation of fractional time series models with deterministic trends”.
11. Hualde, J. and Robinson, P.M. (2011). Gaussian pseudo-maximum likelihood estimation of fractional time series models. *Annals of Statistics* **39**, 3152–3181.
12. Iacone, F. (2010). Local Whittle estimation of the memory parameter in presence of deterministic components. *Journal of Time Series Analysis* **31**, 37–49.
13. Johansen, S. and Nielsen, M.Ø. (2010). Likelihood inference for a nonstationary fractional autoregressive model. *Journal of Econometrics* **158**, 51–66.
14. Johansen, S. and Nielsen, M.Ø. (2016). The role of initial values in conditional sum-of-squares estimation of nonstationary fractional time series models. *Econometric Theory* **32**, 1095–1139.
15. Künsch, H. (1986). Discrimination between monotonic trends and long-range dependence. *Journal of Applied Probability* **23**, 1025–1030.
16. Li, W.K. and McLeod, A.I. (1986). Fractional time series modelling. *Biometrika* **73**, 217–221.
17. Nielsen, M.Ø. (2015). Asymptotics for the conditional-sum-of-squares estimator in multivariate fractional time series models. *Journal of Time Series Analysis* **36**, 154–188.
18. Phillips, P.C.B. (2007). Regression with slowly varying regressors and nonlinear trends. *Econometric Theory* **23**, 557–614.
19. Robinson, P.M. (1994). Efficient tests of nonstationary hypotheses. *Journal of the American Statistical Association* **89**, 1420–1437.
20. Robinson, P.M. (2005). Efficiency improvements in inference on stationary and nonstationary fractional time series. *Annals of Statistics* **33**, 1800–1842.
21. Robinson, P.M. (2012). Inference on power law spatial trends. *Bernoulli* **18**, 644–677.
22. Robinson, P.M. and Iacone, F. (2005). Cointegration in fractional systems with deterministic trends. *Journal of Econometrics* **129**, 263–298.
23. Robinson, P.M. and Marinucci, D. (2000). The averaged periodogram for nonstationary vector time series. *Statistical Inference for Stochastic Processes* **3**, 149–160.
24. Velasco, C. and Robinson, P.M. (2000). Whittle pseudo-maximum likelihood estimation for nonstationary time series. *Journal of the American Statistical Association* **95**, 1229–1243.
25. White, H. and Granger, C.W.J. (2011). Consideration of trends in time series. *Journal of Time Series Econometrics* **3**, issue 1, article 2.
26. Wu, C.-F. (1981). Asymptotic theory of nonlinear least squares estimation. *Annals of Statistics* **9**, 501–513.

# Supplementary Appendix

to

Truncated sum of squares estimation of fractional time series models with deterministic trends

by

Javier Hualde
Universidad Pública de Navarra

and

Morten Ørregaard Nielsen
Queen's University and CREATES

Date: March 21, 2019

# S.1 Introduction

In this supplement to Hualde and Nielsen (2019) we provide proofs of all technical results. This includes proofs of the main theorems and also some auxiliary and technical lemmas and their respective proofs. Note that the proofs of the auxiliary lemmas rely on the technical lemmas, but not vice versa. Equation references (S.$n$) for $n \geq 1$ refer to equations in this supplement and other equation references are to the main paper, Hualde and Nielsen (2019).

# S.2 Proofs of theorems

## S.2.1 Proof of Theorem 1(i): the $\gamma_0 + 1/2 > \delta_0$ case

### S.2.1.1 Overall design of the proof

Throughout, $\epsilon$ will denote a generic arbitrarily small positive constant, and $K$ a generic arbitrarily large positive constant. Fix $\varepsilon > 0$ and let $M_\varepsilon = \{\boldsymbol{\vartheta} \in \Xi : \|\boldsymbol{\tau} - \boldsymbol{\tau}_0\| < \varepsilon\}$, $\overline{M}_\varepsilon = \{\boldsymbol{\vartheta} \in \Xi : \|\boldsymbol{\tau} - \boldsymbol{\tau}_0\| \geq \varepsilon\}$, $N_\varepsilon = \{\boldsymbol{\vartheta} \in \Xi : |\gamma - \gamma_0| < \varepsilon\}$ and $\overline{N}_\varepsilon = \{\boldsymbol{\vartheta} \in \Xi : |\gamma - \gamma_0| \geq \varepsilon\}$. Then $\Pr(||\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0|| \geq \varepsilon) \to 0$ as $T \to \infty$, is implied by

$$\Pr(\widehat{\boldsymbol{\vartheta}} \in \overline{M}_\varepsilon) \to 0 \text{ as } T \to \infty, \tag{S.1}$$

$$\Pr(\widehat{\boldsymbol{\vartheta}} \in \overline{N}_\varepsilon \cap M_\varepsilon) \to 0 \text{ as } T \to \infty. \tag{S.2}$$

Strictly, $\varepsilon$ should be $\varepsilon/\sqrt{2}$ in (S.1) and (S.2), but since $\varepsilon$ is arbitrary this is irrelevant and we continue without the $\sqrt{2}$ factor.

We decompose the objective function as $R_T(\boldsymbol{\vartheta}) = \frac{1}{T}\sum_{t=1}^{T}(d_t(\boldsymbol{\vartheta}) + s_t(\boldsymbol{\vartheta}))^2$ with

$$d_t(\boldsymbol{\vartheta}) = \mu_0 \left( c_t(\gamma_0, \delta, \boldsymbol{\varphi}) - h_{t,T}(\gamma, \delta, \boldsymbol{\varphi}) \sum_{j=1}^{T} c_j(\gamma_0, \delta, \boldsymbol{\varphi}) h_{j,T}(\gamma, \delta, \boldsymbol{\varphi}) \right),$$

$$s_t(\boldsymbol{\vartheta}) = \varepsilon_t(\boldsymbol{\tau}) - h_{t,T}(\gamma, \delta, \boldsymbol{\varphi}) \sum_{j=1}^{T} \varepsilon_j(\boldsymbol{\tau}) h_{j,T}(\gamma, \delta, \boldsymbol{\varphi}),$$

where, as in Hualde and Robinson (2011),

$$\varepsilon_t(\boldsymbol{\tau}) = \sum_{j=0}^{t-1} a_j(\delta_0 - \delta, \boldsymbol{\varphi}) u_{t-j},$$

and where we also defined the coefficient

$$h_{t,T}(d_1, d_2, \boldsymbol{\varphi}) = \frac{c_t(d_1, d_2, \boldsymbol{\varphi})}{(\sum_{j=1}^{T} c_j^2(d_1, d_2, \boldsymbol{\varphi}))^{1/2}}, \tag{S.3}$$

which clearly satisfies $\sum_{t=1}^{T} h_{t,T}^2(d_1, d_2, \boldsymbol{\varphi}) = 1$.

The strategy of proof relies on recognizing the competition between the stochastic term $s_t(\boldsymbol{\vartheta})$ and deterministic term $d_t(\boldsymbol{\vartheta})$ in $R_T(\boldsymbol{\vartheta})$, taking into account that when considering (S.1), just $\boldsymbol{\tau}$ is for sure "far" from $\boldsymbol{\tau}_0$, whereas when dealing with (S.2), just $\gamma$ is "far" from $\gamma_0$. As will be seen, an important feature of the problem is that when $\gamma = \gamma_0$ we have $d_t(\boldsymbol{\vartheta}) = 0$, which complicates the treatment of (S.1). In any case, as in Hualde and Robinson (2011), we need to carefully consider the cases where $R_T(\boldsymbol{\vartheta})$ shows distinct behaviours, noting that either the deterministic or the stochastic term might dominate, and below we partition the parameter space accordingly.

#### S.2.1.2 Proof of (S.1)

To prove (S.1) we use

$$\Pr(\widehat{\boldsymbol{\vartheta}} \in \overline{M}_{\varepsilon}) = \Pr\left(\inf_{\boldsymbol{\vartheta}\in\overline{M}_{\varepsilon}} R_T(\boldsymbol{\vartheta}) \leq \inf_{\boldsymbol{\vartheta}\in M_{\varepsilon}} R_T(\boldsymbol{\vartheta})\right) \leq \Pr\left(\inf_{\boldsymbol{\vartheta}\in\overline{M}_{\varepsilon}} S_T(\boldsymbol{\vartheta}) \leq 0\right), \tag{S.4}$$

where $S_T(\boldsymbol{\vartheta}) = R_T(\boldsymbol{\vartheta}) - R_T(\boldsymbol{\vartheta}_0)$. Fix an arbitrarily small $\eta > 0$ such that $\eta < (\gamma_0 - \delta_0 + 1/2)/2$ and suppose that $\bigtriangledown_1 < \delta_0 - 1/2 - \eta$ and $\bigtriangledown_2 > \gamma_0 - \eta$. Our proof will cover trivially the situation where any of these conditions does not hold, in which case some of the steps below are superfluous. Let $\mathcal{I}_1 = \{\delta : \bigtriangledown_1 \leq \delta \leq \delta_0 - 1/2 - \eta\}$, $\mathcal{I}_2 = \{\delta : \delta_0 - 1/2 - \eta \leq \delta \leq \delta_0 - 1/2\}$, $\mathcal{I}_3 = \{\delta : \delta_0 - 1/2 \leq \delta \leq \delta_0 - 1/2 + \eta\}$, $\mathcal{I}_4 = \{\delta : \delta_0 - 1/2 + \eta \leq \delta \leq \gamma_0 - \eta\}$, and $\mathcal{I}_5 = \{\delta : \gamma_0 - \eta \leq \delta \leq \bigtriangledown_2\}$, noting that the upper bound for $\eta$ guarantees that $\mathcal{I}_4$ is non-empty. Correspondingly define $\mathcal{T}_i = \mathcal{I}_i \times \Psi$ and, fixing $\xi > 0$ and $\varrho > 0$, such that $\varrho < \eta/2$, also define $\mathcal{H}_i = \{\boldsymbol{\vartheta} \in \overline{M}_{\varepsilon} : \boldsymbol{\tau} \in \mathcal{T}_i, |\gamma - \gamma_0| \leq \xi T^{-\varkappa_i}\}$, $\overline{\mathcal{H}}_i = \{\boldsymbol{\vartheta} \in \overline{M}_{\varepsilon} : \boldsymbol{\tau} \in \mathcal{T}_i, \xi T^{-\varkappa_i} \leq |\gamma - \gamma_0| \leq \varrho\}$ and $\overline{\overline{\mathcal{H}}}_i = \{\boldsymbol{\vartheta} \in \overline{M}_{\varepsilon} : \boldsymbol{\tau} \in \mathcal{T}_i, |\gamma - \gamma_0| \geq \varrho\}$, $i = 1, \ldots, 5$, where $\varkappa_i > 0$ will be defined subsequently, noting that $\overline{\mathcal{H}}_i$ is non-empty for any $\xi$, $\varrho$, for $T$ large enough . Then, by (S.4), (S.1) is justified by showing

$$\Pr\left(\inf_{\overline{\overline{\mathcal{H}}}_i} S_T(\boldsymbol{\vartheta}) \leq 0\right) \to 0 \text{ as } T \to \infty \text{ for } i = 1, \ldots, 5, \tag{S.5}$$

$$\Pr\left(\inf_{\overline{\mathcal{H}}_i} S_T(\boldsymbol{\vartheta}) \leq 0\right) \to 0 \text{ as } T \to \infty \text{ for } i = 1, \ldots, 5, \tag{S.6}$$

$$\Pr\left(\inf_{\mathcal{H}_i} S_T(\boldsymbol{\vartheta}) \leq 0\right) \to 0 \text{ as } T \to \infty \text{ for } i = 1, \ldots, 5. \tag{S.7}$$

We note that $\mathcal{H}_i$, $\overline{\mathcal{H}}_i$, and $\overline{\overline{\mathcal{H}}}_i$ are designed exactly such that in $\mathcal{H}_i$ the stochastic term dominates $S_T(\boldsymbol{\vartheta})$, while in $\overline{\mathcal{H}}_i \cup \overline{\overline{\mathcal{H}}}_i$ it is the deterministic term that dominates. As will be seen, the analysis on $\overline{\overline{\mathcal{H}}}_i$ is much simpler because $\gamma$ is "far" from $\gamma_0$, whereas a much more delicate treatment is necessary for $\overline{\mathcal{H}}_i$. This motivates a separate analysis of (S.5), (S.6) and (S.7), at least for $i = 1, \ldots, 4$.

**Proof of (S.5), (S.6), and (S.7) for** $i = 5$ In this case, we give just one proof that covers the whole set $\mathcal{H}_5 \cup \overline{\mathcal{H}}_5 \cup \overline{\overline{\mathcal{H}}}_5$, where $\delta_0 - \delta \leq \delta_0 - \gamma_0 + \eta < 1/2$, so $\Delta_+^{\delta-\delta_0} u_t$ is asymptotically stationary. Let

$$S_T(\boldsymbol{\vartheta}) = U(\boldsymbol{\tau}) - r_T(\boldsymbol{\vartheta}), \tag{S.8}$$

where $U(\boldsymbol{\tau}) = E((\phi(L;\boldsymbol{\varphi})\Delta^{\delta-\delta_0}u_t)^2) - \sigma_0^2$ and

$$
\begin{aligned}
r_T(\boldsymbol{\vartheta}) = & \frac{1}{T}\sum_{t=1}^{T}\left((\phi(L;\boldsymbol{\varphi}_0)\{u_t\mathbb{I}((t\geq 1)\})^2 - \sigma_0^2\right) \\
& - \frac{1}{T}\sum_{t=1}^{T}\left(\varepsilon_t^2(\boldsymbol{\tau}) - E\left((\phi(L;\boldsymbol{\varphi})\Delta^{\delta-\delta_0}u_t)^2\right)\right) \\
& - \frac{1}{T}\left(\sum_{t=1}^{T}\phi(L;\boldsymbol{\varphi}_0)\{u_t\mathbb{I}(t\geq 1)\}h_{t,T}(\gamma_0,\delta_0,\boldsymbol{\varphi}_0)\right)^2 \\
& + \frac{1}{T}\left(\sum_{t=1}^{T}\varepsilon_t(\boldsymbol{\tau})h_{t,T}(\gamma,\delta,\boldsymbol{\varphi})\right)^2 \\
& - \frac{2}{T}\sum_{t=1}^{T}d_t(\boldsymbol{\vartheta})s_t(\boldsymbol{\vartheta}) - \frac{1}{T}\sum_{t=1}^{T}d_t^2(\boldsymbol{\vartheta}),
\end{aligned}
$$

noting that $\varepsilon_t(\boldsymbol{\tau}_0) = \phi(L;\boldsymbol{\varphi}_0)\{u_t\mathbb{I}(t\geq 1)\}$. It follows that (S.5), (S.6), and (S.7) for $i=5$ hold if we show that

$$\inf_{\|\boldsymbol{\tau}-\boldsymbol{\tau}_0\|\geq\varepsilon,\boldsymbol{\tau}\in\mathcal{T}_5} U(\boldsymbol{\tau}) > \epsilon, \tag{S.9}$$

$$\frac{1}{T}\sum_{t=1}^{T}\left((\phi(L;\boldsymbol{\varphi}_0)\{u_t\mathbb{I}(t\geq 1)\})^2 - \sigma_0^2\right) = o_p(1), \tag{S.10}$$

$$\sup_{\|\boldsymbol{\tau}-\boldsymbol{\tau}_0\|\geq\varepsilon,\boldsymbol{\tau}\in\mathcal{T}_5}\frac{1}{T}\sum_{t=1}^{T}\left(\varepsilon_t^2(\boldsymbol{\tau}) - E\left((\phi(L;\boldsymbol{\varphi})\Delta^{\delta-\delta_0}u_t)^2\right)\right) = o_p(1), \tag{S.11}$$

$$\sup_{\mathcal{H}_5\cup\overline{\mathcal{H}}_5\cup\overline{\overline{\mathcal{H}}}_5}\frac{1}{T}\left|\sum_{t=1}^{T}d_t(\boldsymbol{\vartheta})s_t(\boldsymbol{\vartheta})\right| = o_p(1), \tag{S.12}$$

$$\sup_{\mathcal{H}_5\cup\overline{\mathcal{H}}_5\cup\overline{\overline{\mathcal{H}}}_5}\frac{1}{T}\left(\sum_{t=1}^{T}\varepsilon_t(\boldsymbol{\tau})h_{t,T}(\gamma,\delta,\boldsymbol{\varphi})\right)^2 = o_p(1). \tag{S.13}$$

First, (S.9), (S.10), and (S.11) follow by identical arguments to those in the proofs of (2.8) and (2.9) in Hualde and Robinson (2011). Next, by (S.227) of Lemma S.19 with $\gamma_0-\delta\leq\eta$ and $\delta_0-\delta\leq\delta_0-\gamma_0+\eta$, the left-hand side of (S.12) is $O_p(T^{\max\{\theta,\delta_0-\gamma_0+\eta\}+2\theta-1/2+\eta})$, and by (S.222) of Lemma S.18, the left-hand side of (S.13) is $O_p(T^{2\max\{\theta,\delta_0-\gamma_0+\eta\}-1})$. Both are $o_p(1)$ for $\theta$ and $\eta$ sufficiently small, to conclude the proof of (S.5), (S.6), and (S.7) for $i=5$.

**Proof of (S.5) for** $i=1,\ldots,4$ First we show (S.5) which, in view of Lemma S.1 and that $d_t(\boldsymbol{\vartheta}_0)=0$, holds if, for $i=1,\ldots,4$,

$$\Pr\left(\inf_{\overline{\overline{\mathcal{H}}}_i}\frac{1}{T}\sum_{t=1}^{T}(d_t(\boldsymbol{\vartheta})+s_t(\boldsymbol{\vartheta}))^2\leq\sigma_0^2+\epsilon\right)\to 0 \text{ as } T\to\infty.$$

For $\delta \in \cup_{i=1}^{4}\mathcal{I}_i$ it holds that $\gamma_0 - \delta \geq \eta$, so the probability above is bounded by

$$\Pr\left(\inf_{\overline{\overline{\mathcal{H}}}_i} \frac{T^{2(\gamma_0-\delta)+1}}{T} \inf_{\overline{\overline{\mathcal{H}}}_i} \frac{1}{T^{2(\gamma_0-\delta)+1}} \sum_{t=1}^{T} \left(d_t(\boldsymbol{\vartheta}) + s_t(\boldsymbol{\vartheta})\right)^2 \leq \sigma_0^2 + \epsilon\right)$$
$$= \Pr\left(\inf_{\overline{\overline{\mathcal{H}}}_i} \frac{1}{T^{2(\gamma_0-\delta)+1}} \sum_{t=1}^{T} \left(d_t(\boldsymbol{\vartheta}) + s_t(\boldsymbol{\vartheta})\right)^2 \leq \frac{\sigma_0^2+\epsilon}{T^{2\eta}}\right)$$
$$\leq \Pr\left(\inf_{\overline{\overline{\mathcal{H}}}_i} \frac{1}{T^{2(\gamma_0-\delta)+1}} \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta}) - \sup_{\overline{\overline{\mathcal{H}}}_i} \frac{2}{T^{2(\gamma_0-\delta)+1}} \left|\sum_{t=1}^{T} d_t(\boldsymbol{\vartheta})\, s_t(\boldsymbol{\vartheta})\right| \leq \frac{\sigma_0^2+\epsilon}{T^{2\eta}}\right).$$

Thus, (S.5) for $i = 1, \ldots, 4$ follows for $\theta$ small enough by (S.229) of Lemma S.19, noting also that when $\delta \in \cup_{i=1}^{4}\mathcal{I}_i$, $\delta_0 - \delta \geq \delta_0 - \gamma_0 + \eta$, and by Lemma S.2.

**Proof of (S.6) and (S.7) for $i = 4$** Fix $\zeta$ such that $0 < \zeta < \eta$ and let $\varkappa_4 = \gamma_0 - \delta - \zeta$, noting that $\varkappa_4 \geq \eta - \zeta > 0$ when $\delta \in \mathcal{I}_4$. Then, because $d_t(\boldsymbol{\vartheta}_0) = 0$, (S.6) holds if

$$\Pr\left(\inf_{\overline{\mathcal{H}}_4} \frac{T^{2\varkappa_4}}{T^{2(\gamma_0-\delta)+1}} \left(\sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta}) - 2\left|\sum_{t=1}^{T} d_t(\boldsymbol{\vartheta})\, s_t(\boldsymbol{\vartheta})\right| - \sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}_0)\right) \leq 0\right) \to 0 \qquad \text{(S.14)}$$

as $T \to \infty$, noting the change in the normalization from (S.6) to (S.14), which is justified because the right-hand side of the inequality inside the probability in (S.6) is 0, so multiplying the left- and right-hand sides of the inequality by the same positive number does not alter the probability. Because $\sum_{t=1}^{T} d_t(\boldsymbol{\vartheta})\, c_t(\gamma, \delta, \boldsymbol{\varphi}) = 0$, it holds that

$$\sum_{t=1}^{T} d_t(\boldsymbol{\vartheta})\, s_t(\boldsymbol{\vartheta}) = \sum_{t=1}^{T} d_t(\boldsymbol{\vartheta})\, \varepsilon_t(\boldsymbol{\tau}). \qquad \text{(S.15)}$$

By the Cauchy-Schwarz inequality and (S.15), the probability in (S.14) is bounded by

$$\Pr\left(\inf_{\overline{\mathcal{H}}_4} \frac{T^{2\varkappa_4}}{T^{2(\gamma_0-\delta)+1}} \left(\sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta})\left(1 - 2v_T(\boldsymbol{\vartheta})\right) - \sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}_0)\right) \leq 0\right), \qquad \text{(S.16)}$$

where $v_T(\boldsymbol{\vartheta}) = \left(\sum_{t=1}^{T} \varepsilon_t^2(\boldsymbol{\tau}) / \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta})\right)^{1/2}$. Then (S.14) holds if

$$\sup_{\overline{\mathcal{H}}_4} \frac{T^{2\varkappa_4}}{T^{2(\gamma_0-\delta)+1}} \sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}_0) = o_p(1), \qquad \text{(S.17)}$$

$$\Pr\left(\inf_{\overline{\mathcal{H}}_4} \frac{T^{2\varkappa_4}}{T^{2(\gamma_0-\delta)+1}} \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta})\left(1 - 2v_T(\boldsymbol{\vartheta})\right) \leq \epsilon\right) \to 0 \text{ as } T \to \infty. \qquad \text{(S.18)}$$

First, given that $T^{2\varkappa_4-2(\gamma_0-\delta)-1}=T^{-1-2\zeta}$, (S.17) follows immediately by Lemma S.1. Next, fixing $c$ such that $0<c<1/2$, the probability in (S.18) equals

$$\begin{aligned}
&\Pr\left(\inf_{\overline{\mathcal{H}}_4}\frac{T^{2\varkappa_4}}{T^{2(\gamma_0-\delta)+1}}\sum_{t=1}^{T}d_t^2\left(\boldsymbol{\vartheta}\right)\left(1-2v_T\left(\boldsymbol{\vartheta}\right)\right)\leq\epsilon,\sup_{\overline{\mathcal{H}}_4}v_T\left(\boldsymbol{\vartheta}\right)\leq c\right)\\
&+\Pr\left(\inf_{\overline{\mathcal{H}}_4}\frac{T^{2\varkappa_4}}{T^{2(\gamma_0-\delta)+1}}\sum_{t=1}^{T}d_t^2\left(\boldsymbol{\vartheta}\right)\left(1-2v_T\left(\boldsymbol{\vartheta}\right)\right)\leq\epsilon,\sup_{\overline{\mathcal{H}}_4}v_T\left(\boldsymbol{\vartheta}\right)>c\right)\\
&\leq\Pr\left(\inf_{\overline{\mathcal{H}}_4}\frac{T^{2\varkappa_4}}{T^{2(\gamma_0-\delta)+1}}\sum_{t=1}^{T}d_t^2\left(\boldsymbol{\vartheta}\right)\left(1-2c\right)\leq\epsilon\right)+\Pr\left(\sup_{\overline{\mathcal{H}}_4}v_T\left(\boldsymbol{\vartheta}\right)>c\right), \qquad \text{(S.19)}
\end{aligned}$$

so (S.18) holds on showing

$$\lim_{T\to\infty}\inf_{\overline{\mathcal{H}}_4}\frac{T^{2\varkappa_4}}{T^{2(\gamma_0-\delta)+1}}\sum_{t=1}^{T}d_t^2\left(\boldsymbol{\vartheta}\right)>\epsilon, \qquad \text{(S.20)}$$

$$\sup_{\overline{\mathcal{H}}_4}v_T\left(\boldsymbol{\vartheta}\right)=o_p\left(1\right). \qquad \text{(S.21)}$$

By the Cauchy-Schwarz inequality, $\sum_{t=1}^{T}d_t^2\left(\boldsymbol{\vartheta}\right)\geq T^{-1}\overline{d}_T^2\left(\boldsymbol{\vartheta}\right)$, where $\overline{d}_T\left(\boldsymbol{\vartheta}\right)=\sum_{t=1}^{T}d_t\left(\boldsymbol{\vartheta}\right)$, so that (S.20) holds by (S.109) of Lemma S.3. To show (S.21), note that

$$\sup_{\overline{\mathcal{H}}_4}v_T\left(\boldsymbol{\vartheta}\right)\leq\left(\frac{\sup_{\overline{\mathcal{H}}_4}T^{-1-2\zeta}\sum_{t=1}^{T}\varepsilon_t^2\left(\boldsymbol{\tau}\right)}{\inf_{\overline{\mathcal{H}}_4}T^{2\varkappa_4-2(\gamma_0-\delta)-1}\sum_{t=1}^{T}d_t^2\left(\boldsymbol{\vartheta}\right)}\right)^{1/2} \qquad \text{(S.22)}$$

using $\varkappa_4=\gamma_0-\delta-\zeta$, where $\sup_{\overline{\mathcal{H}}_4}T^{-1-2\zeta}\sum_{t=1}^{T}\varepsilon_t^2\left(\boldsymbol{\tau}\right)=o_p\left(1\right)$ by Lemma S.17 because $\delta_0-\delta\leq1/2-\eta$. Then (S.22) is $o_p(1)$ by (S.20), which concludes the proof of (S.6) for $i=4$.

Next we show (S.7) for $i=4$. A potential problem here is that $\gamma=\gamma_0$ is admissible, so we cannot directly exploit the lower bound for the normalized $\sum_{t=1}^{T}d_t^2\left(\boldsymbol{\vartheta}\right)$ as in (S.20) because $d_t\left(\boldsymbol{\vartheta}\right)=0$ when $\gamma=\gamma_0$. However, we can instead take advantage of $|\gamma-\gamma_0|\leq\xi T^{-\varkappa_4}$ in $\mathcal{H}_4$ and apply the mean value theorem. First note that $\delta\in\mathcal{I}_4$ implies that $\delta_0-\delta\leq1/2-\eta$ and $\gamma_0-\delta\geq\eta$, so that $\Delta_+^{\delta-\delta_0}u_t$ is asymptotically stationary as in the proof for $i=5$. Then, given (S.8), the result follows by (S.9), (S.10), (S.11) (whose proofs apply also for $\delta\in\mathcal{I}_4$), and showing also that

$$\sup_{\mathcal{H}_4}\frac{1}{T}\left(\sum_{t=1}^{T}\varepsilon_t\left(\boldsymbol{\tau}\right)h_{t,T}\left(\gamma,\delta,\boldsymbol{\varphi}\right)\right)^2=o_p\left(1\right), \qquad \text{(S.23)}$$

$$\sup_{\mathcal{H}_4}\frac{1}{T}\left|\sum_{t=1}^{T}d_t\left(\boldsymbol{\vartheta}\right)s_t\left(\boldsymbol{\vartheta}\right)\right|=o_p\left(1\right). \qquad \text{(S.24)}$$

From (S.222) of Lemma S.18, the left-hand side of (S.23) is $O_p(T^{-2\eta})=o_p\left(1\right)$ by choosing $\theta<1/2-\eta$. Next, because $|\gamma-\gamma_0|<\xi T^{-\varkappa_4}$ in $\mathcal{H}_4$, by (S.226) and (S.228) of Lemma S.19 the left-hand side of (S.24) is $O_p\left(T^{\zeta-\eta+2\theta}\right)=o_p(1)$ for $\theta$ small enough because $\zeta<\eta$.

**Proof of (S.6) and (S.7) for $i = 3$** Fix $\varkappa_3 = \gamma_0 - \delta$, so noting that $\delta \in \mathcal{I}_3$, $\varkappa_3 \geq \gamma_0 - \delta_0 + 1/2 - \eta > 0$. Then, by the Cauchy-Schwarz inequality,

$$\Pr\left(\inf_{\overline{\mathcal{H}}_3} S_T(\boldsymbol{\vartheta}) \leq 0\right) \leq \Pr\left(\inf_{\overline{\mathcal{H}}_3} \frac{1}{T^2}\left(\overline{d}_T(\boldsymbol{\vartheta}) + \overline{s}_T(\boldsymbol{\vartheta})\right)^2 - \frac{1}{T}\sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}_0) \leq 0\right), \tag{S.25}$$

where $\overline{d}_t(\boldsymbol{\vartheta}) = \sum_{j=1}^{t} d_j(\boldsymbol{\vartheta})$ and $\overline{s}_t(\boldsymbol{\vartheta}) = \sum_{j=1}^{t} s_j(\boldsymbol{\vartheta})$, so

$$\overline{s}_T(\boldsymbol{\vartheta}) = \varepsilon_T(\delta_0 - \delta + 1, \boldsymbol{\varphi}) - \sum_{t=1}^{T} h_{t,T}(\gamma, \delta, \boldsymbol{\varphi}) \sum_{j=1}^{T} \varepsilon_j(\boldsymbol{\tau}) h_{j,T}(\gamma, \delta, \boldsymbol{\varphi}), \tag{S.26}$$

where, denoting $\varepsilon_t(\delta_0 - \delta, \boldsymbol{\varphi}) = \varepsilon_t(\boldsymbol{\tau})$,

$$\varepsilon_t(\delta_0 - \delta + 1, \boldsymbol{\varphi}) = \sum_{j=1}^{t} \varepsilon_j(\boldsymbol{\tau}) = \sum_{j=1}^{t}\sum_{k=0}^{j-1} a_k(\delta_0 - \delta, \boldsymbol{\varphi}) u_{j-k} = \sum_{j=0}^{t-1} a_j(\delta_0 - \delta + 1, \boldsymbol{\varphi}) u_{t-j}, \tag{S.27}$$

because

$$\pi_{j+1}(d) - \pi_j(d) = \pi_{j+1}(d-1). \tag{S.28}$$

The right-hand side of (S.25) is thus bounded by

$$\Pr\left(\inf_{\overline{\mathcal{H}}_3} \frac{1}{T^2}\overline{d}_T^2(\boldsymbol{\vartheta})\left(1 - 2\left|\overline{v}_T(\boldsymbol{\vartheta})\right|\right) - \frac{1}{T}\sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}_0) \leq 0\right), \tag{S.29}$$

where $\overline{v}_T(\boldsymbol{\vartheta}) = \overline{s}_T(\boldsymbol{\vartheta})/\overline{d}_T(\boldsymbol{\vartheta})$. Applying Lemma S.1, (S.6) for $i = 3$ would then hold if

$$\Pr\left(\inf_{\overline{\mathcal{H}}_3} \frac{1}{T^2}\overline{d}_T^2(\boldsymbol{\vartheta})\left(1 - 2\left|\overline{v}_T(\boldsymbol{\vartheta})\right|\right) \leq K\right) \to 0 \text{ as } T \to \infty, \tag{S.30}$$

for an arbitrarily large $K$. As in (S.19), fixing $c$ such that $0 < c < 1/2$, the probability in (S.30) is bounded by

$$\Pr\left(\inf_{\overline{\mathcal{H}}_3} \frac{1}{T^2}\overline{d}_T^2(\boldsymbol{\vartheta})(1 - 2c) \leq K\right) + \Pr\left(\sup_{\overline{\mathcal{H}}_3}\left|\overline{v}_T(\boldsymbol{\vartheta})\right| > c\right), \tag{S.31}$$

so, as in (S.22), (S.30) holds if

$$\sup_{\overline{\mathcal{H}}_3} \frac{1}{T}\left|\overline{s}_T(\boldsymbol{\vartheta})\right| = O_p(1), \tag{S.32}$$

$$\lim_{T\to\infty} \inf_{\overline{\mathcal{H}}_3} \frac{1}{T^2}\overline{d}_T^2(\boldsymbol{\vartheta}) > K. \tag{S.33}$$

For $\delta \in \mathcal{I}_3$ it holds that $\delta_0 - \delta \leq 1/2$, so in view of (S.26) the proof of (S.32) is immediate using (S.213) in Lemma S.16 together with Lemmas S.17 and S.18 with $\theta < 1/2$. Finally the proof of (S.33) follows by Lemma S.3, to conclude the proof of (S.6) for $i = 3$.

Next we show (S.7) for $i = 3$, which holds if

$$\Pr\left(\inf_{\mathcal{H}_3} \frac{1}{T}\left(\sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}) - 2\left|\sum_{t=1}^{T} d_t(\boldsymbol{\vartheta}) s_t(\boldsymbol{\vartheta})\right| - \sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}_0)\right) \leq 0\right) \to 0 \text{ as } T \to \infty,$$

where

$$\sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}) = \sum_{t=1}^{T} \varepsilon_t^2(\boldsymbol{\tau}) - \left(\sum_{t=1}^{T} \varepsilon_t(\boldsymbol{\tau}) h_{t,T}(\gamma, \delta, \boldsymbol{\varphi})\right)^2. \quad \text{(S.34)}$$

In the proof of their (2.7) for $i = 3$, Hualde and Robinson (2011) showed that

$$\Pr\left(\inf_{\|\boldsymbol{\tau}-\boldsymbol{\tau}_0\|\geq\varepsilon, \boldsymbol{\tau}\in\mathcal{T}_3} \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t^2(\boldsymbol{\tau}) > K\right) \to 1 \text{ as } T \to \infty \quad \text{(S.35)}$$

for any arbitrarily large fixed constant $K$ (for small enough $\eta$). Thus, noting (S.34), (S.7) for $i = 3$ holds by (S.35) and Lemma S.1 on showing

$$\sup_{\mathcal{H}_3} \frac{1}{T}\left(\sum_{t=1}^{T} \varepsilon_t(\boldsymbol{\tau}) h_{t,T}(\gamma, \delta, \boldsymbol{\varphi})\right)^2 = O_p(1), \quad \text{(S.36)}$$

$$\sup_{\mathcal{H}_3} \frac{1}{T}\left|\sum_{t=1}^{T} d_t(\boldsymbol{\vartheta}) s_t(\boldsymbol{\vartheta})\right| = O_p(1). \quad \text{(S.37)}$$

In (S.37), the bound will depend on $\xi$, which had to be set very large in the proof of (S.33) (see the proof of Lemma S.3). However, this can be dominated by the constant $K$ in (S.35), which can be chosen arbitrarily large by setting $\eta$ small enough. First, by (S.222) of Lemma S.18, the left-hand side of (S.36) holds by choosing $\theta < 1/2$ because $\delta_0 - \delta \leq 1/2$ when $\delta \in \mathcal{I}_3$. Next, noting that $\sup_{\mathcal{H}_3} |\gamma - \gamma_0| \leq \xi T^{-\varkappa_3}$ and that $\delta \in \mathcal{I}_3$ implies $\gamma_0 - \delta \geq \gamma_0 - \delta_0 + 1/2 - \eta > 0$ and $\delta_0 - \delta \leq 1/2$, it follows by (S.226) and (S.230) of Lemma S.19 (noting that for $T$ large enough $\gamma - \delta > 0$) that the left-hand side of (S.37) is $O_p(1)$ by choosing $\theta < 1/2$.

**Proof of (S.6) and (S.7) for** $i = 2$ Fix $\varkappa_2 = \gamma_0 - \delta_0 + 1/2 > 0$. Changing the normalization ($T^{2(\delta_0-\delta)}$ instead of $T$), by the Cauchy-Schwarz inequality as in (S.25), and proceeding as in (S.29), the left-hand side of (S.6) is bounded by

$$\Pr\left(\inf_{\overline{\mathcal{H}}_2} \frac{1}{T^{2(\delta_0-\delta)+1}} \overline{d}_T^2(\boldsymbol{\vartheta})\left(1 - 2\left|\overline{v}_T(\boldsymbol{\vartheta})\right|\right) - \sup_{\overline{\mathcal{H}}_2} \frac{1}{T^{2(\delta_0-\delta)}} \sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}_0) \leq 0\right).$$

Then, given Lemma S.1,

$$\sup_{\overline{\mathcal{H}}_2} \frac{1}{T^{2(\delta_0-\delta)}} \sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}_0) = \sup_{\overline{\mathcal{H}}_2} \frac{T}{T^{2(\delta_0-\delta)}} \frac{1}{T}\sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}_0) = O_p(1), \quad \text{(S.38)}$$

because when $\delta \in \mathcal{I}_2$, $\delta_0 - \delta \geq 1/2$. Thus (S.6) for $i = 2$ would hold if

$$\Pr\left(\inf_{\overline{\mathcal{H}}_2} \frac{1}{T^{2(\delta_0-\delta)+1}} \overline{d}_T^2(\boldsymbol{\vartheta})\left(1 - 2\left|\overline{v}_T(\boldsymbol{\vartheta})\right|\right) \leq K\right) \to 0 \text{ as } T \to \infty \quad \text{(S.39)}$$

for an arbitrarily large $K$, which, as in (S.31), follows if

$$\sup_{\overline{\mathcal{H}}_2} \frac{1}{T^{\delta_0-\delta+1/2}} \left|\overline{s}_T(\boldsymbol{\vartheta})\right| = O_p(1), \quad \text{(S.40)}$$

$$\lim_{T\to\infty} \inf_{\overline{\mathcal{H}}_2} \frac{1}{T^{2(\delta_0-\delta)+1}} \overline{d}_T^2(\boldsymbol{\vartheta}) > K. \quad \text{(S.41)}$$

The proof of (S.40) is almost identical to that of (S.32), again applying Lemmas S.16, S.17, and S.18. Finally (S.41) follows by Lemma S.3, to conclude the proof of (S.6) for $i = 2$.

Next we show (S.7) for $i = 2$, which holds if

$$\Pr\left(\inf_{\mathcal{H}_2} \frac{1}{T^{2(\delta_0-\delta)}}\left(\sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}) - 2\left|\sum_{t=1}^{T} d_t(\boldsymbol{\vartheta})\, s_t(\boldsymbol{\vartheta})\right| - \sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}_0)\right) \leq 0\right) \to 0 \qquad \text{(S.42)}$$

as $T \to \infty$. In the proof of their (2.7) for $i = 2$, Hualde and Robinson (2011) showed that

$$\Pr\left(\inf_{\|\boldsymbol{\tau}-\boldsymbol{\tau}_0\|\geq\varepsilon, \boldsymbol{\tau}\in\mathcal{T}_2} \frac{1}{T^{2(\delta_0-\delta)}} \sum_{t=1}^{T} \varepsilon_t^2(\boldsymbol{\tau}) > K\right) \to 1 \qquad \text{(S.43)}$$

as $T \to \infty$ for any arbitrarily large fixed constant $K$ (for small enough $\eta$). Thus, in view of (S.34), (S.38), and (S.43), it follows that (S.42) holds if

$$\sup_{\mathcal{H}_2} \frac{1}{T^{2(\delta_0-\delta)}}\left(\sum_{t=1}^{T} \varepsilon_t(\boldsymbol{\tau}) h_{t,T}(\gamma, \delta, \boldsymbol{\varphi})\right)^2 = O_p(1), \qquad \text{(S.44)}$$

$$\sup_{\mathcal{H}_2} \frac{1}{T^{2(\delta_0-\delta)}}\left|\sum_{t=1}^{T} d_t(\boldsymbol{\vartheta})\, s_t(\boldsymbol{\vartheta})\right| = O_p(1). \qquad \text{(S.45)}$$

Again, (S.44) holds by (S.223) of Lemma S.18 with $\theta < 1/2$ noting that $\delta \in \mathcal{I}_2$ implies $\delta_0 - \delta \geq 1/2$. Since $\sup_{\mathcal{H}_2} |\gamma - \gamma_0| \leq \xi T^{-\varkappa_2}$ and $\delta \in \mathcal{I}_2$ implies $\gamma_0 - \delta \geq \gamma_0 - \delta_0 + 1/2 > 0$ and $\delta_0 - \delta \geq 1/2$, (S.45) follows from (S.226) and (S.231) of Lemma S.19 setting $\theta < 1/2$, noting that for $T$ large enough $\gamma - \delta > 0$.

**Proof of (S.6) and (S.7) for** $i = 1$  Fix $\varkappa_1 = \gamma_0 - \delta_0 + 1/2 > 0$. As in the treatment of (S.14), (S.6) for $i = 1$ holds if

$$\sup_{\mathcal{I}_1} \frac{T^{2\varkappa_1}}{T^{2(\gamma_0-\delta)+1}} \sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}_0) = o_p(1), \qquad \text{(S.46)}$$

$$\Pr\left(\inf_{\overline{\mathcal{H}}_1} \frac{T^{2\varkappa_1}}{T^{2(\gamma_0-\delta)+1}} \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta})\,(1 - 2v_T(\boldsymbol{\vartheta})) \leq \epsilon\right) \to 0 \text{ as } T \to \infty, \qquad \text{(S.47)}$$

for an arbitrarily small $\epsilon$. First, (S.46) holds by Lemma S.1, noting that $2\varkappa_1 - 2(\gamma_0 - \delta) - 1 = 2(\delta - \delta_0)$ and $\sup_{\mathcal{I}_1} 2(\delta - \delta_0) = -1 - 2\eta < -1$. Next, as in the proof of (S.39), see also (S.18) and (S.22), (S.47) follows if

$$\sup_{\overline{\mathcal{H}}_1} \frac{1}{T^{2(\delta_0-\delta)}} \sum_{t=1}^{T} \varepsilon_t^2(\boldsymbol{\tau}) = O_p(1), \qquad \text{(S.48)}$$

$$\lim_{T\to\infty} \inf_{\overline{\mathcal{H}}_1} \frac{1}{T^{2(\delta_0-\delta)}} \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta}) > K. \qquad \text{(S.49)}$$

First, (S.48) follows immediately from (S.219) of Lemma S.17, noting that $\delta_0 - \delta \geq 1/2 + \eta$. Next, by the Cauchy-Schwarz inequality, (S.49) follows by Lemma S.3.

Finally we show (S.7) for $i = 1$, which holds if

$$\Pr\left(\inf_{\mathcal{H}_1} \frac{1}{T^{2(\delta_0-\delta)}} \left(\sum_{t=1}^{T} \left(d_t(\boldsymbol{\vartheta}) + s_t(\boldsymbol{\vartheta})\right)^2 - \sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}_0)\right) \le 0\right) \to 0 \tag{S.50}$$

as $T \to \infty$. By (S.46) and Lemma S.12 with $Z_t = d_t(\boldsymbol{\vartheta}) + s_t(\boldsymbol{\vartheta})$, (S.50) follows if there exists an $\epsilon > 0$ such that

$$\Pr\left(\inf_{\mathcal{H}_1} \frac{1}{T} \sum_{t=1}^{T} \left(\frac{1}{T^{\delta_0-\delta+1/2}} \overline{d}_t(\boldsymbol{\vartheta}) + \frac{1}{T^{\delta_0-\delta+1/2}} \overline{s}_t(\boldsymbol{\vartheta})\right)^2 > \epsilon\right) \to 1 \tag{S.51}$$

as $T \to \infty$. Note that in $\mathcal{H}_1$, $\gamma_0 - \delta \ge \eta$, so there exists $\alpha > 0$ such that for $T$ sufficiently large it holds that $\gamma - \delta \ge \alpha$. Define the set $\mathcal{G}_1 = \{\boldsymbol{\vartheta} : \boldsymbol{\tau} \in \mathcal{T}_1, \gamma - \delta \ge \alpha, \gamma \in [\Box_1, \Box_2]\}$. Let $[\cdot]$ denote the integer part of the argument and consider $S_T(r, \boldsymbol{\vartheta}) = T^{\delta-\delta_0-1/2}\overline{s}_{[Tr]}(\boldsymbol{\vartheta})$ a process indexed by $(r, \boldsymbol{\vartheta})$ that is càdlàg in $r$ and continuous in $\boldsymbol{\vartheta}$. We next show that

$$\begin{aligned} S_T(r, \boldsymbol{\vartheta}) \Rightarrow S(r, \boldsymbol{\vartheta}) =& \phi(1; \boldsymbol{\varphi})\, \omega(1; \boldsymbol{\varphi}_0)\, W(r; 1 + \delta_0 - \delta) \\ &- \frac{\phi(1; \boldsymbol{\varphi})\, \omega(1; \boldsymbol{\varphi}_0)\left(2(\gamma - \delta) + 1\right)}{\gamma - \delta + 1} r^{\gamma-\delta+1} \\ &\times \left(W(1; 1 + \delta_0 - \delta) - \int_0^1 u^{\gamma-\delta-1} W(u; 1 + \delta_0 - \delta)\, du\right), \end{aligned} \tag{S.52}$$

where $\Rightarrow$ means weak convergence in the product space of functions that are càdlàg in $r \in [0, 1]$ and continuous in $\boldsymbol{\vartheta} \in \mathcal{G}_1$ endowed with the Skorokhod topology in $r$ and the uniform topology in $\boldsymbol{\vartheta}$, and where $W(r; d) = \Gamma(d)^{-1} \int_0^r (1 - s)^{d-1} dB(s)$ and $B(s)$ denote fractional (Type II) and regular scalar Brownian motions, respectively, both with variance $\sigma_0^2$. Because $\overline{d}_t(\boldsymbol{\vartheta})$ is deterministic and $\overline{s}_t(\boldsymbol{\vartheta})$ is stochastic, and in view of the square in (S.51), (S.51) and hence (S.50) follows from (S.52) (also note Assumption A1(iv) and (7)). We note that a different approach for the case $i = 1$ was taken by Hualde and Robinson (2011) in their eqn. (2.36) based on the Cauchy-Schwarz inequality, but that approach does not appear sufficient; see Johansen and Nielsen (2018) for details on this point and for an argument very similar to that for our (S.50)–(S.52). In our (S.126) below, we give the result of Hualde and Robinson (2011) with an alternative proof based on our Lemma S.12.

We thus need to prove (S.52). Note (S.27) and

$$c_t(d_1, d_2, \boldsymbol{\varphi}) - c_{t-1}(d_1, d_2, \boldsymbol{\varphi}) = c_t(d_1, d_2 + 1, \boldsymbol{\varphi}). \tag{S.53}$$

By summation by parts and (S.53) we find

$$\begin{aligned} \sum_{j=1}^{T} c_j(\gamma, \delta, \boldsymbol{\varphi})\, \varepsilon_j(\boldsymbol{\tau}) =& c_T(\gamma, \delta, \boldsymbol{\varphi})\, \varepsilon_T(\delta_0 - \delta + 1, \boldsymbol{\varphi}) \\ &- \sum_{t=1}^{T-1} c_{t+1}(\gamma, \delta + 1, \boldsymbol{\varphi})\, \varepsilon_t(\delta_0 - \delta + 1, \boldsymbol{\varphi}). \end{aligned} \tag{S.54}$$

Also, by summation by parts on (10), noting (S.28),

$$a_j(d, \boldsymbol{\varphi}) = \phi(1; \boldsymbol{\varphi})\, \pi_j(d) - \pi_j(d) \sum_{k=j+1}^{\infty} \phi_k(\boldsymbol{\varphi}) - \sum_{k=0}^{j-1} \pi_{k+1}(d - 1) \sum_{l=0}^{k} \phi_{j-l}(\boldsymbol{\varphi}), \tag{S.55}$$

where for $j=0$ the last term on the right-hand side of (S.55) is 0. Thus, noting (S.27),

$$\begin{aligned}\varepsilon_t(\delta_0-\delta+1,\boldsymbol{\varphi}) &= \sum_{j=0}^{t-1} a_j\left(\delta_0-\delta+1,\boldsymbol{\varphi}\right)u_{t-j} \\ &= \phi\left(1;\boldsymbol{\varphi}\right)\omega\left(1;\boldsymbol{\varphi}_0\right)\sum_{j=0}^{t-1}\pi_j\left(\delta_0-\delta+1\right)\varepsilon_{t-j}+m_{1t}\left(\boldsymbol{\tau}\right), \qquad \text{(S.56)}\end{aligned}$$

where

$$\begin{aligned}m_{1t}\left(\boldsymbol{\tau}\right)=&\phi\left(1;\boldsymbol{\varphi}\right)\sum_{j=0}^{t-1}\pi_j\left(\delta_0-\delta+1\right)\left(u_{t-j}-\omega\left(1;\boldsymbol{\varphi}_0\right)\varepsilon_{t-j}\right) \\ &-\sum_{j=0}^{t-1}\pi_j\left(\delta_0-\delta+1\right)\sum_{k=j+1}^{\infty}\phi_k\left(\boldsymbol{\varphi}\right)u_{t-j}-\sum_{j=1}^{t-1}\sum_{k=0}^{j-1}\pi_{k+1}\left(\delta_0-\delta\right)\sum_{l=0}^{k}\phi_{j-l}\left(\boldsymbol{\varphi}\right)u_{t-j}. \qquad \text{(S.57)}\end{aligned}$$

Substituting (S.27), (S.54), (S.56) and (S.200) into $s_j(\boldsymbol{\vartheta})$, we get

$$\begin{aligned}\frac{1}{T^{\delta_0-\delta+1/2}}\sum_{j=1}^{t}s_j\left(\boldsymbol{\vartheta}\right)=&\frac{1}{T^{\delta_0-\delta+1/2}}\varepsilon_t(\delta_0-\delta+1,\boldsymbol{\varphi}) \\ &-\frac{c_T\left(\gamma,\delta,\boldsymbol{\varphi}\right)\varepsilon_T\left(\delta_0-\delta+1,\boldsymbol{\varphi}\right)\sum_{j=1}^{t}c_j\left(\gamma,\delta,\boldsymbol{\varphi}\right)}{T^{\delta_0-\delta+1/2}\sum_{j=1}^{T}c_j^2\left(\gamma,\delta,\boldsymbol{\varphi}\right)} \\ &+\frac{\sum_{j=1}^{t}c_j\left(\gamma,\delta,\boldsymbol{\varphi}\right)\sum_{k=1}^{T-1}c_{k+1}\left(\gamma,\delta+1,\boldsymbol{\varphi}\right)\varepsilon_k\left(\delta_0-\delta+1,\boldsymbol{\varphi}\right)}{T^{\delta_0-\delta+1/2}\sum_{j=1}^{T}c_j^2\left(\gamma,\delta,\boldsymbol{\varphi}\right)} \\ =&\phi\left(1;\boldsymbol{\varphi}\right)\omega\left(1;\boldsymbol{\varphi}_0\right)\widetilde{V}_t\left(\gamma,\delta\right)+m_{2t}\left(\boldsymbol{\vartheta}\right), \qquad \text{(S.58)}\end{aligned}$$

where

$$\begin{aligned}\widetilde{V}_t\left(\gamma,\delta\right)=&\frac{1}{T^{\delta_0-\delta+1/2}}\sum_{j=0}^{t-1}\pi_j\left(\delta_0-\delta+1\right)\varepsilon_{t-j} \\ &-\frac{1}{T^{\delta_0-\delta+1/2}}\frac{b_T\left(\gamma,\delta\right)\sum_{k=0}^{T-1}\pi_k\left(\delta_0-\delta+1\right)\varepsilon_{T-k}}{\sum_{k=1}^{T}b_k^2\left(\gamma,\delta\right)}\sum_{j=1}^{t}b_j\left(\gamma,\delta\right) \\ &+\frac{1}{T^{\delta_0-\delta+1/2}}\frac{\sum_{k=1}^{T-1}b_{k+1}\left(\gamma,\delta+1\right)\sum_{l=0}^{k-1}\pi_l\left(\delta_0-\delta+1\right)\varepsilon_{k-l}}{\sum_{k=1}^{T}b_k^2\left(\gamma,\delta\right)}\sum_{j=1}^{t}b_j\left(\gamma,\delta\right),\end{aligned}$$

and $m_{2t}\left(\boldsymbol{\vartheta}\right)$ collects remainder terms arising from (S.56) and (S.200). By relatively straightforward arguments, it can be shown that

$$\sup_{\mathcal{G}_1}\frac{1}{T}\sum_{t=1}^{T}m_{2t}^2\left(\boldsymbol{\vartheta}\right)=o_p\left(1\right), \qquad \text{(S.59)}$$

$$\sup_{\mathcal{G}_1}\frac{1}{T}\sum_{t=1}^{T}\left|m_{2t}\left(\boldsymbol{\vartheta}\right)\widetilde{V}_t\left(\gamma,\delta\right)\right|=o_p\left(1\right), \qquad \text{(S.60)}$$

The proofs of (S.59) and (S.60) involve several results. First, in order to deal with the first term on the right-hand side of (S.57), note that $u_t = \omega(1;\boldsymbol{\varphi}_0)\varepsilon_t + \widetilde{\varepsilon}_{t-1} - \widetilde{\varepsilon}_t$, where $\widetilde{\varepsilon}_t = \sum_{j=0}^{\infty} \widetilde{\omega}_j(\boldsymbol{\varphi}_0)\varepsilon_{t-j}$, $\widetilde{\omega}_j(\boldsymbol{\varphi}_0) = \sum_{k=j+1}^{\infty} \omega_k(\boldsymbol{\varphi}_0)$. By Assumptions A1 and A2, $\widetilde{\varepsilon}_t$ is well defined in the mean-square sense and $|\widetilde{\omega}_j(\boldsymbol{\varphi}_0)| = O(j^{-\varsigma})$. We also apply Lemmas S.9, S.11, S.13, S.14, S.15, S.16, and S.17.

Because $S_T(r,\boldsymbol{\vartheta})$ is defined on a product space, we can prove weak convergence in $r \in [0,1]$ and $\boldsymbol{\vartheta} \in \mathcal{G}_1$ separately. Thus, suppose first that $\boldsymbol{\vartheta} \in \mathcal{G}_1$ is fixed. Letting $t = [Tr]$, weak convergence of $\widetilde{V}_{[Tr]}(\gamma,\delta)$ as $T \to \infty$ then follows by first applying Theorem 2 of Hosoya (2005), noting that our Assumption A2 implies conditions A(i), A(ii) and A(iii) in Hosoya (2005). We then apply Lemmas S.10, S.11, and S.13 and the continuous mapping theorem, as in Robinson and Marinucci (2000), noting that for fixed $\boldsymbol{\vartheta} \in \mathcal{G}_1$ and in particular $\gamma - \delta \geq \alpha > 0$, it holds that $\int_0^1 u^{\gamma-\delta-1} W(u;1+\delta_0-\delta)\,du$ is a well-defined random variable with zero mean and finite variance (e.g., for $\delta = \delta_0$ this variance is $2\sigma_0^2((\gamma-\delta+1)(2(\gamma-\delta)+1))^{-1}$). To prove weak convergence in $\boldsymbol{\vartheta} \in \mathcal{G}_1$, we note that finite-dimensional convergence follows by weak convergence in $r \in [0,1]$. Tightness of the process $\widetilde{V}_{[Tr]}(\gamma,\delta)$ on the compact set $\mathcal{G}_1$ follows from Lemmas A.2 and C.3 of Johansen and Nielsen (2010) noting, in particular, that $\widetilde{V}_{[Tr]}(\gamma,\delta)$ is continuously differentiable for $\gamma - \delta \geq \alpha > 0$. Because $\boldsymbol{\varphi}$ only enters through the multiplicative function $\phi(1;\boldsymbol{\varphi})$, which is bounded and bounded away from zero for $\boldsymbol{\vartheta} \in \mathcal{G}_1$, tightness in $\boldsymbol{\varphi}$ follows straightforwardly. This proves (S.52) to conclude the proof of (S.7) for $i = 1$ and therefore that of (S.1).

#### S.2.1.3 Proof of (S.2)

Here, let $R_T(\boldsymbol{\tau},\gamma) = R_T(\boldsymbol{\vartheta}), d_t(\boldsymbol{\tau},\gamma) = d_t(\boldsymbol{\vartheta})$, and $s_t(\boldsymbol{\tau},\gamma) = s_t(\boldsymbol{\vartheta}) = s_{1t}(\boldsymbol{\tau}) - s_{2t}(\boldsymbol{\vartheta})$ with $s_{1t}(\boldsymbol{\tau}) = \varepsilon_t(\boldsymbol{\tau})$ and $s_{2t}(\boldsymbol{\tau},\gamma) = s_{2t}(\boldsymbol{\vartheta}) = h_{t,T}(\gamma,\delta,\boldsymbol{\varphi})\sum_{j=1}^{T} s_{1j}(\boldsymbol{\tau})h_{j,T}(\gamma,\delta,\boldsymbol{\varphi})$, so that, noting $\sum_{t=1}^{T} h_{t,T}^2(\gamma,\delta,\boldsymbol{\varphi}) = 1$,

$$\sum_{t=1}^{T} s_{2t}^2(\boldsymbol{\vartheta}) = \sum_{t=1}^{T} s_{1t}(\boldsymbol{\tau})s_{2t}(\boldsymbol{\vartheta}) = \left(\sum_{j=1}^{T} s_{1j}(\boldsymbol{\tau})h_{j,T}(\gamma,\delta,\boldsymbol{\varphi})\right)^2 .$$

Noting also (S.15),

$$\begin{aligned} R_T(\boldsymbol{\vartheta}) =& \frac{1}{T}\sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta}) + \frac{1}{T}\sum_{t=1}^{T} s_{1t}^2(\boldsymbol{\tau}) \\ & - \frac{1}{T}\left(\sum_{j=1}^{T} s_{1j}(\boldsymbol{\tau})h_{j,T}(\gamma,\delta,\boldsymbol{\varphi})\right)^2 + \frac{2}{T}\sum_{t=1}^{T} d_t(\boldsymbol{\vartheta})s_{1t}(\boldsymbol{\vartheta}). \end{aligned} \tag{S.61}$$

Clearly, if $\widehat{\boldsymbol{\vartheta}} \in \overline{N}_\varepsilon \cap M_\varepsilon$, then $\inf_{\overline{N}_\varepsilon \cap M_\varepsilon} R_T(\widehat{\boldsymbol{\tau}},\gamma) \leq R_T(\widehat{\boldsymbol{\tau}},\gamma_0)$, so that

$$\Pr(\widehat{\boldsymbol{\vartheta}} \in \overline{N}_\varepsilon \cap M_\varepsilon) \leq \Pr\left(\widehat{\boldsymbol{\vartheta}} \in \overline{N}_\varepsilon \cap M_\varepsilon, \inf_{\overline{N}_\varepsilon \cap M_\varepsilon} R_T(\widehat{\boldsymbol{\tau}},\gamma) - R_T(\widehat{\boldsymbol{\tau}},\gamma_0) \leq 0\right). \tag{S.62}$$

Recalling that $d_t(\boldsymbol{\tau},\gamma_0)=0$, $R_T(\widehat{\boldsymbol{\tau}},\gamma_0)=T^{-1}\sum_{t=1}^{T}s_{1t}(\widehat{\boldsymbol{\tau}})$ and this cancels with the corresponding term in $R_T(\widehat{\boldsymbol{\tau}},\gamma)$, see (S.61). Thus, (S.2) holds if

$$\lim_{T\to\infty}\inf_{\boldsymbol{\vartheta}\in\overline{N}_\varepsilon\cap M_\varepsilon}\frac{1}{T^{2(\gamma_0-\delta)+1}}\sum_{t=1}^{T}d_t^2(\boldsymbol{\vartheta})>\epsilon, \tag{S.63}$$

$$\sup_{\boldsymbol{\vartheta}\in\overline{N}_\varepsilon\cap M_\varepsilon}\frac{1}{T^{2(\gamma_0-\delta)+1}}\left|\sum_{t=1}^{T}d_t(\boldsymbol{\vartheta})\,s_{1t}(\boldsymbol{\tau})\right|=o_p(1), \tag{S.64}$$

$$\sup_{\boldsymbol{\vartheta}\in\overline{N}_\varepsilon\cap M_\varepsilon}\frac{1}{T^{2(\gamma_0-\delta)+1}}\left(\sum_{j=1}^{T}s_{1j}(\boldsymbol{\tau})\,h_{j,T}(\gamma,\delta,\boldsymbol{\varphi})\right)^2=o_p(1), \tag{S.65}$$

noting the change in the normalization compared with (S.62) ($T^{2(\gamma_0-\delta_0)+1}$ instead of $T$), which is justified because the right-hand side of the inequality inside the probability in (S.62) is 0, so multiplying the left- and right-hand sides of the inequality by a positive number does not alter the probability.

First, (S.63) follows from Lemma S.2, noting that in $\overline{N}_\varepsilon\cap M_\varepsilon$, $\gamma_0-\delta\geq\gamma_0-\delta_0-\varepsilon>-1/2$ setting $\varepsilon$ small enough. Next, letting both $\varepsilon$ and $\theta$ be sufficiently small and noting that in $\overline{N}_\varepsilon\cap M_\varepsilon$, $\delta_0-\delta\geq-\varepsilon$, by (S.229) of Lemma S.19 the left-hand side of (S.64) is $O_p(T^{-1/2+\delta_0-\gamma_0+3\theta+\varepsilon})=o_p(1)$. Finally, by (S.223) of Lemma S.18 the left-hand side of (S.65) is $O_p(T^{-2(\gamma_0-\delta_0+1/2-\theta-\varepsilon)})=o_p(1)$, to conclude the proof of (S.2) and therefore that of consistency of $\widehat{\boldsymbol{\vartheta}}$.

### S.2.2 Proof of Theorem 1(ii): the $\gamma_0+1/2<\delta_0$ case

Clearly

$$\Pr\left(\|\widehat{\boldsymbol{\tau}}-\boldsymbol{\tau}_0\|\geq\varepsilon\right)=\Pr\left(\inf_{\boldsymbol{\vartheta}\in\overline{M}_\varepsilon}R_T(\boldsymbol{\vartheta})\leq\inf_{\boldsymbol{\vartheta}\in M_\varepsilon}R_T(\boldsymbol{\vartheta})\right),$$

so, as in the proof for $\gamma_0+1/2>\delta_0$, the result follows by showing that the right-hand side of (S.4) is $o(1)$, which, in view of Lemma S.1, holds if

$$\Pr\left(\inf_{\boldsymbol{\vartheta}\in\overline{M}_\varepsilon}R_T(\boldsymbol{\vartheta})\leq\sigma_0^2+\epsilon\right)\to0\text{ as }T\to\infty.$$

This result is given in Lemma S.4, whose proof uses very similar techniques to those employed in the proof of (S.1). This completes the proof of Theorem 1.

### S.2.3 Proof of Theorem 2(i): the $\gamma_0+1/2>\delta_0$ case

Let $\boldsymbol{M}_T=\operatorname{diag}(I_{p+1},T^{\delta_0-\gamma_0})$, c.f. (15). We first show that

$$T^{1/2}\boldsymbol{M}_T^{-1}(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0)\to_d N\left(0,\sigma_0^2\boldsymbol{V}^{-1}\right). \tag{S.66}$$

By the mean value theorem,

$$\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0=-\left(\frac{\partial^2R_T(\overline{\boldsymbol{\vartheta}})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\right)^{-1}\frac{\partial R_T(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}}, \tag{S.67}$$

where $\overline{\boldsymbol{\vartheta}}$ represents an intermediate point which is allowed to vary across the different rows of $\partial^2R_T(\cdot)/\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'$.

We first analyze the score in (S.67). It can be easily seen that $\partial d_t(\boldsymbol{\vartheta}_0)/\partial\boldsymbol{\tau} = 0$ and $\partial s_{1t}(\boldsymbol{\tau})/\partial\gamma = 0$, so, recalling that $d_t(\boldsymbol{\vartheta}_0) = 0$ and the decomposition (S.61),

$$\frac{\partial R_T(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}} = \frac{2}{T}\sum_{t=1}^{T} s_t(\boldsymbol{\vartheta}_0)\left(\begin{pmatrix} \frac{\partial s_{1t}(\boldsymbol{\tau}_0)}{\partial\boldsymbol{\tau}} \\ \frac{\partial d_t(\boldsymbol{\vartheta}_0)}{\partial\gamma} \end{pmatrix} - \frac{\partial s_{2t}(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}}\right).$$

Then, by Lemma S.5(a) it holds that

$$\frac{T^{1/2}}{2}\boldsymbol{M}_T\frac{\partial R_T(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}} = \begin{pmatrix} T^{-1/2}I_{p+1} & 0 \\ 0 & T^{-1/2-(\gamma_0-\delta_0)} \end{pmatrix}\sum_{t=1}^{T}\varepsilon_t\begin{pmatrix} \frac{\partial s_{1t}(\boldsymbol{\tau}_0)}{\partial\boldsymbol{\tau}} \\ \frac{\partial d_t(\boldsymbol{\vartheta}_0)}{\partial\gamma} \end{pmatrix} + o_p(1). \quad \text{(S.68)}$$

Next, as in (2.54) of Hualde and Robinson (2011),

$$\frac{1}{T^{1/2}}\sum_{t=1}^{T}\varepsilon_t\frac{\partial s_{1t}(\boldsymbol{\tau}_0)}{\partial\boldsymbol{\tau}} = \frac{1}{T^{1/2}}\sum_{t=2}^{T}\varepsilon_t\sum_{j=1}^{\infty}\boldsymbol{m}_j(\boldsymbol{\varphi}_0)\varepsilon_{t-j} + o_p(1),$$

where $\boldsymbol{m}_j(\boldsymbol{\varphi}_0) = (-j^{-1}, \boldsymbol{b}_j'(\boldsymbol{\varphi}_0))'$. Also,

$$\begin{aligned}\frac{\partial d_t(\boldsymbol{\vartheta}_0)}{\partial\gamma} =& -\mu_0 c_t^{(1)}(\gamma_0,\delta_0,\boldsymbol{\varphi}_0) \\ &+ \mu_0 c_t(\gamma_0,\delta_0,\boldsymbol{\varphi}_0)\frac{\sum_{j=1}^{T} c_j(\gamma_0,\delta_0,\boldsymbol{\varphi}_0)c_j^{(1)}(\gamma_0,\delta_0,\boldsymbol{\varphi}_0)}{\sum_{j=1}^{T} c_j^2(\gamma_0,\delta_0,\boldsymbol{\varphi}_0)},\end{aligned}$$

where $c_t^{(1)}(\cdot,\cdot,\cdot)$ is the derivative of $c_t(\cdot,\cdot,\cdot)$ with respect to the first argument, so that

$$\begin{aligned}&\frac{1}{T^{\gamma_0-\delta_0+1/2}}\sum_{t=1}^{T}\varepsilon_t\frac{\partial d_t(\boldsymbol{\vartheta}_0)}{\partial\gamma} \\ &= \frac{\mu_0}{T^{\gamma_0-\delta_0+1/2}}\sum_{t=1}^{T}\varepsilon_t c_t(\gamma_0,\delta_0,\boldsymbol{\varphi}_0)\frac{\sum_{j=1}^{T} c_j(\gamma_0,\delta_0,\boldsymbol{\varphi}_0)c_j^{(1)}(\gamma_0,\delta_0,\boldsymbol{\varphi}_0)}{\sum_{j=1}^{T} c_j^2(\gamma_0,\delta_0,\boldsymbol{\varphi}_0)} \\ &\quad - \frac{\mu_0}{T^{\gamma_0-\delta_0+1/2}}\sum_{t=1}^{T}\varepsilon_t c_t^{(1)}(\gamma_0,\delta_0,\boldsymbol{\varphi}_0). \end{aligned} \quad \text{(S.69)}$$

By (11), $c_t^{(1)}(d_1,d_2,\boldsymbol{\varphi}) = \sum_{j=0}^{t-1}\phi_j(\boldsymbol{\varphi})b_{t-j}^{(1)}(d_1,d_2)$, where $b_j^{(1)}(\cdot,\cdot)$ is the derivative of $b_j(\cdot,\cdot)$ with respect to the first argument. Then, noting that $\gamma_0 + 1/2 > \delta_0$, by a similar analysis to that in the proof of Lemma S.15, the right-hand side of (S.69) equals

$$\begin{aligned}&\frac{\mu_0\phi(1;\boldsymbol{\varphi}_0)}{T^{\gamma_0-\delta_0+1/2}}\frac{\sum_{t=1}^{T}\varepsilon_t b_t(\gamma_0,\delta_0)\sum_{j=1}^{T} b_j(\gamma_0,\delta_0)b_j^{(1)}(\gamma_0,\delta_0)}{\sum_{j=1}^{T} b_j^2(\gamma_0,\delta_0)} \\ &- \frac{\mu_0\phi(1;\boldsymbol{\varphi}_0)}{T^{\gamma_0-\delta_0+1/2}}\sum_{t=1}^{T}\varepsilon_t b_t^{(1)}(\gamma_0,\delta_0) + o_p(1).\end{aligned} \quad \text{(S.70)}$$

Substituting (S.185) (evaluated at $(\gamma_0,\delta_0)$) into (S.70), the first two terms of (S.70) become

$$\begin{aligned}&\frac{\mu_0\phi(1;\boldsymbol{\varphi}_0)}{T^{\gamma_0-\delta_0+1/2}}\frac{\sum_{t=1}^{T}\varepsilon_t b_t(\gamma_0,\delta_0)\sum_{j=1}^{T}\log(\frac{j}{T})b_j^2(\gamma_0,\delta_0)}{\sum_{j=1}^{T} b_j^2(\gamma_0,\delta_0)} \\ &- \frac{\mu_0\phi(1;\boldsymbol{\varphi}_0)}{T^{\gamma_0-\delta_0+1/2}}\sum_{t=1}^{T}\log(\frac{t}{T})b_t(\gamma_0,\delta_0)\varepsilon_t + o_p(1).\end{aligned} \quad \text{(S.71)}$$

Using (S.184) and (S.185), (S.71) equals $\mu_0\phi(1;\boldsymbol{\varphi}_0)\Gamma(\gamma_0+1)T^{-1/2+\delta_0-\gamma_0}\sum_{t=1}^{T}\varepsilon_t g_{t,T}(\gamma_0-\delta_0+1)+$ $o_p(1)$ with

$$g_{t,T}(d)=\frac{\pi_t(d)\sum_{j=1}^{T}\log(\frac{j}{T})\pi_j^2(d)-\log(\frac{t}{T})\pi_t(d)\sum_{j=1}^{T}\pi_j^2(d)}{\sum_{j=1}^{T}\pi_j^2(d)}.$$

Collecting these terms shows that

$$\frac{T^{1/2}}{2}\boldsymbol{M}_T\frac{\partial R_T(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}}=\sum_{t=1}^{T}\varepsilon_t\boldsymbol{\eta}_{t,T}+o_p(1), \tag{S.72}$$

where

$$\boldsymbol{\eta}_{t,T}=\begin{pmatrix}\frac{1}{T^{1/2}}\sum_{j=1}^{\infty}\boldsymbol{m}_j(\boldsymbol{\varphi}_0)\varepsilon_{t-j}\\ \frac{1}{T^{\gamma_0-\delta_0+1/2}}\mu_0\phi(1;\boldsymbol{\varphi}_0)\Gamma(\gamma_0+1)g_{t,T}(\gamma_0-\delta_0+1)\end{pmatrix}.$$

Defining $\mathcal{F}_{t,T}=\mathcal{F}_t$ for any $1\le t\le T$, Assumption A2 implies that $\{\varepsilon_t\boldsymbol{\eta}_{t,T},\mathcal{F}_{t,T},1\le t\le T,T\ge 1\}$ is a martingale difference array. For any $(p+2)$-dimensional vector $\boldsymbol{\xi}$, define $\xi_{t,T}=\varepsilon_t\boldsymbol{\xi}'\boldsymbol{\eta}_{t,T}/\sigma_0(\boldsymbol{\xi}'\boldsymbol{V}\boldsymbol{\xi})^{1/2}$ and $B_T^2=\sum_{t=2}^{T}E(\xi_{t,T}^2|\mathcal{F}_{t-1,T})$. Then, by Corollary 3.1 of Hall and Heyde (1980), if

$$B_T^2\to_p 1, \tag{S.73}$$

and, for all $\epsilon>0$,

$$\sum_{t=2}^{T}E\left(\xi_{t,T}^2\mathbb{I}(|\xi_{t,T}|>\epsilon)|\mathcal{F}_{t-1,T}\right)\to_p 0, \tag{S.74}$$

it holds that $\sum_{t=2}^{T}\xi_{t,T}\to_d N(0,1)$, and hence

$$\sum_{t=2}^{T}\varepsilon_t\boldsymbol{\eta}_{t,T}\to_d N\left(0,\sigma_0^2\boldsymbol{V}\right) \tag{S.75}$$

by direct application of the Cramer-Wold device. First we note that

$$E\left(\xi_{t,T}^2|\mathcal{F}_{t-1,T}\right)=\frac{\boldsymbol{\xi}'\boldsymbol{\eta}_{t,T}\boldsymbol{\eta}_{t,T}'\boldsymbol{\xi}}{\boldsymbol{\xi}'\boldsymbol{V}\boldsymbol{\xi}},$$

so that (S.73) holds if $\sum_{t=2}^{T}\boldsymbol{\eta}_{t,T}\boldsymbol{\eta}_{t,T}'\to_p\boldsymbol{V}$. However, this follows straightforwardly by the same arguments as in the proof of (2.55) of Hualde and Robinson (2011) and Lemma S.10 because $\sum_{l=1}^{t}\sum_{j=1}^{\infty}\boldsymbol{m}_j(\boldsymbol{\varphi}_0)\varepsilon_{l-j}=O_p(t^{1/2})$, which implies, by summation by parts, that

$$\frac{1}{T^{\gamma_0-\delta_0+1}}\sum_{t=2}^{T}g_{t,T}(\gamma_0-\delta_0+1)\sum_{j=1}^{\infty}\boldsymbol{m}_j(\boldsymbol{\varphi}_0)\varepsilon_{t-j}=o_p(1). \tag{S.76}$$

Now (S.74) holds if, e.g., $\sum_{t=2}^{T}E(\xi_{t,T}^4|\mathcal{F}_{t-1,T})\to_p 0$, which, given that the fourth moment of $\varepsilon_t$ is finite, holds if $\sum_{t=2}^{T}(\boldsymbol{\xi}'\boldsymbol{\eta}_{t,T}\boldsymbol{\eta}_{t,T}'\boldsymbol{\xi})^2\to_p 0$, and this can be easily justified by previous arguments. This completes the proof of (S.75).

Next, noting (S.67), (S.72), and (S.75), the proof of (S.66) is completed by showing

$$\boldsymbol{M}_T\left(\frac{\partial^2 R_T(\overline{\boldsymbol{\vartheta}})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}-\frac{\partial^2 R_T(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\right)\boldsymbol{M}_T=o_p(1) \tag{S.77}$$

and

$$\frac{1}{2}\boldsymbol{M}_T\frac{\partial^2 R_T(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\boldsymbol{M}_T \to_p \boldsymbol{V}. \tag{S.78}$$

By Lemma S.6 it holds that, for some fixed $\varkappa > 0$, $T^{\varkappa}(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) \to_p 0$, and in light of this the proof of (S.77) is relatively straightforward. It consists of deriving all terms in $\partial^2 R_T(\overline{\boldsymbol{\vartheta}})/\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'$ and checking that the differences with respect the corresponding ones in $\partial^2 R_T(\boldsymbol{\vartheta}_0)/\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'$ satisfy (S.77). This requires the use of the mean value theorem and Assumption A4(ii), where typically the derivatives involve additional $\log T$ factors which are compensated by the factor $T^{-\varkappa}$ that arises because $\overline{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0 = O_p(T^{-\varkappa})$.

Now we show (S.78). Recalling $d_t(\boldsymbol{\vartheta}_0) = 0$ and noting $\partial^2 d_t(\boldsymbol{\vartheta}_0)/\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}' = 0$, Lemma S.5(b) implies that

$$\frac{1}{2}\boldsymbol{M}_T\frac{\partial^2 R_T(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\boldsymbol{M}_T = \boldsymbol{M}_T\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}\frac{\partial s_{1t}(\boldsymbol{\tau}_0)}{\partial\boldsymbol{\tau}}\frac{\partial s_{1t}(\boldsymbol{\tau}_0)}{\partial\boldsymbol{\tau}'} & 0\\ 0 & \left(\frac{\partial d_t(\boldsymbol{\vartheta}_0)}{\partial\gamma}\right)^2\end{pmatrix}\boldsymbol{M}_T + o_p(1), \tag{S.79}$$

so (S.78) holds by showing that, as $T \to \infty$,

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial s_{1t}(\boldsymbol{\tau}_0)}{\partial\boldsymbol{\tau}}\frac{\partial s_{1t}(\boldsymbol{\tau}_0)}{\partial\boldsymbol{\tau}'} \to_p \sigma_0^2\boldsymbol{A}, \tag{S.80}$$

$$\frac{1}{T^{2(\gamma_0-\delta_0)+1}}\sum_{t=1}^{T}\left(\frac{\partial d_t(\boldsymbol{\vartheta}_0)}{\partial\gamma}\right)^2 \to \frac{\mu_0^2\phi^2(1;\boldsymbol{\varphi}_0)\Gamma^2(\gamma_0+1)}{\Gamma^2(\gamma_0-\delta_0+1)(2(\gamma_0-\delta_0)+1)^3}. \tag{S.81}$$

Here, (S.80) follows from (2.53) of Hualde and Robinson (2011) and (S.81) follows by arguments used in the proof of (S.73).

Next, given (S.66), the remaining part of (16) is justified as follows. Noting

$$\phi(L;\boldsymbol{\varphi})\Delta_+^{\delta}x_t = \mu_0 c_t(\gamma_0,\delta,\boldsymbol{\varphi}) + \varepsilon_t(\boldsymbol{\tau}), \tag{S.82}$$

it follows from (12) that

$$\widehat{\mu} = \widehat{\mu}(\widehat{\boldsymbol{\vartheta}}) = \mu_0\sum_{t=1}^{T}c_t(\gamma_0,\widehat{\delta},\widehat{\boldsymbol{\varphi}})k_{t,T}(\widehat{\gamma},\widehat{\delta},\widehat{\boldsymbol{\varphi}}) + \sum_{t=1}^{T}\varepsilon_t(\widehat{\boldsymbol{\tau}})k_{t,T}(\widehat{\gamma},\widehat{\delta},\widehat{\boldsymbol{\varphi}}),$$

where $k_{t,T}(\gamma,\delta,\boldsymbol{\varphi}) = c_t(\gamma,\delta,\boldsymbol{\varphi})/\sum_{t=1}^{T}c_t^2(\gamma,\delta,\boldsymbol{\varphi})$. By straightforward application of the mean value theorem,

$$k_{t,T}(\widehat{\gamma},\widehat{\delta},\widehat{\boldsymbol{\varphi}}) = k_{t,T}(\gamma_0,\widehat{\delta},\widehat{\boldsymbol{\varphi}}) + k_{t,T}^{(1)}(\overline{\gamma},\widehat{\delta},\widehat{\boldsymbol{\varphi}})(\widehat{\gamma}-\gamma_0),$$

where $k_{t,T}^{(1)}(\cdot,\cdot,\cdot)$ is the derivative of $k_{t,T}(\cdot,\cdot,\cdot)$ with respect to the first argument and $|\overline{\gamma}-\gamma_0| \le |\widehat{\gamma}-\gamma_0|$. Thus,

$$\widehat{\mu} = \mu_0 + \mu_0(\widehat{\gamma}-\gamma_0)\sum_{t=1}^{T}c_t(\gamma_0,\widehat{\delta},\widehat{\boldsymbol{\varphi}})k_{t,T}^{(1)}(\overline{\gamma},\widehat{\delta},\widehat{\boldsymbol{\varphi}}) + \sum_{t=1}^{T}\varepsilon_t(\widehat{\boldsymbol{\tau}})k_{t,T}(\widehat{\gamma},\widehat{\delta},\widehat{\boldsymbol{\varphi}}),$$

which implies that

$$\begin{aligned}\frac{T^{\gamma_0-\delta_0+1/2}}{\log T}(\widehat{\mu}-\mu_0) =& \mu_0 T^{\gamma_0-\delta_0+1/2}(\widehat{\gamma}-\gamma_0)\frac{1}{\log T}\sum_{t=1}^{T}c_t(\gamma_0,\widehat{\delta},\widehat{\boldsymbol{\varphi}})k_{t,T}^{(1)}(\overline{\gamma},\widehat{\delta},\widehat{\boldsymbol{\varphi}})\\ &+\frac{T^{\gamma_0-\delta_0+1/2}}{\log T}\sum_{t=1}^{T}\varepsilon_t(\widehat{\boldsymbol{\tau}})k_{t,T}(\widehat{\gamma},\widehat{\delta},\widehat{\boldsymbol{\varphi}}).\end{aligned} \tag{S.83}$$

Then, the remaining part of (16) holds on showing

$$\frac{1}{\log T}\sum_{t=1}^{T} c_t(\gamma_0, \widehat{\delta}, \widehat{\boldsymbol{\varphi}}) k_{t,T}^{(1)}(\overline{\gamma}, \widehat{\delta}, \widehat{\boldsymbol{\varphi}}) \to_p -1, \tag{S.84}$$

$$\frac{T^{\gamma_0-\delta_0+1/2}}{\log T}\sum_{t=1}^{T} \varepsilon_t(\widehat{\boldsymbol{\tau}}) k_{t,T}(\widehat{\gamma}, \widehat{\delta}, \widehat{\boldsymbol{\varphi}}) = o_p(1), \tag{S.85}$$

noting that joint asymptotic distribution of $\hat{\mu}$ and $\hat{\gamma}$ follows easily from (S.83)–(S.85) by the Cramér-Wold device.

Clearly, (S.84) follows if

$$\sum_{t=1}^{T} c_t(\gamma_0, \widehat{\delta}, \widehat{\boldsymbol{\varphi}}) k_{t,T}^{(1)}(\overline{\gamma}, \widehat{\delta}, \widehat{\boldsymbol{\varphi}}) - \sum_{t=1}^{T} c_t(\gamma_0, \delta_0, \boldsymbol{\varphi}_0) k_{t,T}^{(1)}(\gamma_0, \delta_0, \boldsymbol{\varphi}_0) = o_p(\log T), \tag{S.86}$$

$$\frac{1}{\log T}\sum_{t=1}^{T} c_t(\gamma_0, \delta_0, \boldsymbol{\varphi}_0) k_{t,T}^{(1)}(\gamma_0, \delta_0, \boldsymbol{\varphi}_0) \to -1. \tag{S.87}$$

First, (S.86) can be easily justified by applying Lemmas S.7 and S.14, noting that

$$k_{t,T}^{(1)}(\gamma, \delta, \boldsymbol{\varphi}) = \frac{c_t^{(1)}(\gamma, \delta, \boldsymbol{\varphi})}{\sum_{j=1}^{T} c_j^2(\gamma, \delta, \boldsymbol{\varphi})} - \frac{2c_t(\gamma, \delta, \boldsymbol{\varphi}) \sum_{j=1}^{T} c_j^{(1)}(\gamma, \delta, \boldsymbol{\varphi}) c_j(\gamma, \delta, \boldsymbol{\varphi})}{\left(\sum_{j=1}^{T} c_j^2(\gamma, \delta, \boldsymbol{\varphi})\right)^2}.$$

Next, the left-hand side of (S.87) is

$$-\frac{1}{\log T}\frac{\sum_{t=1}^{T} c_t^{(1)}(\gamma_0, \delta_0, \boldsymbol{\varphi}_0) c_t(\gamma_0, \delta_0, \boldsymbol{\varphi}_0)}{\sum_{t=1}^{T} c_t^2(\gamma_0, \delta_0, \boldsymbol{\varphi}_0)} = -\frac{1}{\log T}\frac{\sum_{t=1}^{T} b_t^2(\gamma_0, \delta_0)\log t}{\sum_{t=1}^{T} b_t^2(\gamma_0, \delta_0)} + o(1),$$

by (S.185), noting that the remainder is of smaller order. Thus (S.87) follows immediately using (S.184), Lemma S.11, and noting that by simple application of summation by parts, the mean value theorem and Lemma S.10 for $d > -1/2$, it can be easily shown that

$$\frac{1}{T^{2d+1}\log T}\sum_{t=1}^{T} \log(t)\, t^{2d} \to \frac{1}{2d+1}.$$

Next, the left-hand side of (S.85) is

$$\frac{T^{\gamma_0-\delta_0+1/2}}{\log T}\sum_{t=1}^{T} \phi(L; \boldsymbol{\varphi}_0)\{u_t \mathbb{I}(t \geq 1)\} k_{t,T}(\gamma_0, \delta_0, \boldsymbol{\varphi}_0) \tag{S.88}$$

$$+\frac{T^{\gamma_0-\delta_0+1/2}}{\log T}\sum_{t=1}^{T} (\varepsilon_t(\widehat{\boldsymbol{\tau}}) - \phi(L; \boldsymbol{\varphi}_0)\{u_t \mathbb{I}(t \geq 1)\}) k_{t,T}(\gamma_0, \delta_0, \boldsymbol{\varphi}_0) \tag{S.89}$$

$$+\frac{T^{\gamma_0-\delta_0+1/2}}{\log T}\sum_{t=1}^{T} \phi(L; \boldsymbol{\varphi}_0)\{u_t \mathbb{I}(t \geq 1)\}(k_{t,T}(\widehat{\gamma}, \widehat{\delta}, \widehat{\boldsymbol{\varphi}}) - k_{t,T}(\gamma_0, \delta_0, \boldsymbol{\varphi}_0)) \tag{S.90}$$

$$+\frac{T^{\gamma_0-\delta_0+1/2}}{\log T}\sum_{t=1}^{T} (\varepsilon_t(\widehat{\boldsymbol{\tau}}) - \phi(L; \boldsymbol{\varphi}_0)\{u_t \mathbb{I}(t \geq 1)\})(k_{t,T}(\widehat{\gamma}, \widehat{\delta}, \widehat{\boldsymbol{\varphi}}) - k_{t,T}(\gamma_0, \delta_0, \boldsymbol{\varphi}_0)). \tag{S.91}$$

Note that

$$\phi(L;\boldsymbol{\varphi}_0)\left\{u_t\mathbb{I}(t\geq 1)\right\}=\varepsilon_t-\sum_{j=t}^{\infty}\phi_j\left(\boldsymbol{\varphi}_0\right)u_{t-j}, \tag{S.92}$$

where, by Assumptions A1 and A2, it can be easily shown that

$$\sum_{j=t}^{\infty}\phi_j\left(\boldsymbol{\varphi}_0\right)u_{t-j}=O_p(t^{-1/2-\varsigma}). \tag{S.93}$$

Using summation by parts, (S.92), (S.93) and noting that, as in Lemmas S.14 and S.15,

$$k_{t,T}(\gamma_0,\delta_0,\boldsymbol{\varphi}_0)=O_p(t^{\gamma_0-\delta_0}T^{-1-2(\gamma_0-\delta_0)}), \tag{S.94}$$

it follows that (S.88) is $O_p(\log^{-1}T)$. Next, by (S.94) and Lemma S.7 with $\varkappa=1/2$ (because $\widehat{\boldsymbol{\tau}}$ is $T^{1/2}$-consistent), (S.89) is $O_p(T^{1/2-\varkappa}\log^{-1}T)=O_p(\log^{-1}T)$. Next, by summation by parts, (S.92), (S.93), the mean value theorem and Lemmas S.7 and S.16, it can be easily shown that (S.90) is $O_p(T^{\theta-1/2-(\gamma_0-\delta_0)})=o_p\left(1\right)$, setting $\theta<\gamma_0-\delta_0+1/2$. Finally, combining the arguments for (S.89) and (S.90), it is straightforward to show that (S.91) is $o_p\left(1\right)$, to conclude the proof of (S.85).

### S.2.4 Proof of Theorem 2(ii): the $\gamma_0+1/2<\delta_0$ case

First, noting (S.61), the loss function $R_T\left(\boldsymbol{\vartheta}\right)$ can be decomposed in the sum of two terms, $R_T\left(\boldsymbol{\vartheta}\right)=Q_T\left(\boldsymbol{\tau}\right)+S_T\left(\boldsymbol{\vartheta}\right)$, where $Q_T\left(\boldsymbol{\tau}\right)=T^{-1}\sum_{t=1}^{T}s_{1t}^2\left(\boldsymbol{\tau}\right)$ and

$$S_T\left(\boldsymbol{\vartheta}\right)=\frac{1}{T}\sum_{t=1}^{T}\left(d_t\left(\boldsymbol{\vartheta}\right)-s_{2t}\left(\boldsymbol{\vartheta}\right)\right)^2+\frac{2}{T}\sum_{t=1}^{T}s_{1t}\left(\boldsymbol{\tau}\right)\left(d_t\left(\boldsymbol{\vartheta}\right)-s_{2t}\left(\boldsymbol{\vartheta}\right)\right). \tag{S.95}$$

Thus, $Q_T\left(\boldsymbol{\tau}\right)$ is the loss function in Hualde and Robinson (2011). Now

$$\frac{\partial R_T(\widehat{\boldsymbol{\vartheta}})}{\partial\boldsymbol{\tau}}=0=\frac{\partial Q_T(\widehat{\boldsymbol{\tau}})}{\partial\boldsymbol{\tau}}+\frac{\partial S_T(\widehat{\boldsymbol{\vartheta}})}{\partial\boldsymbol{\tau}}, \tag{S.96}$$

and by the mean value theorem

$$\frac{\partial Q_T(\widehat{\boldsymbol{\tau}})}{\partial\boldsymbol{\tau}}=\frac{\partial Q_T(\boldsymbol{\tau}_0)}{\partial\boldsymbol{\tau}}+\frac{\partial^2 Q_T(\overline{\boldsymbol{\tau}})}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}'}\left(\widehat{\boldsymbol{\tau}}-\boldsymbol{\tau}_0\right), \tag{S.97}$$

where $\overline{\boldsymbol{\tau}}$ represents an intermediate point between $\widehat{\boldsymbol{\tau}}$ and $\boldsymbol{\tau}_0$ which is allowed to vary in the different rows of $\partial^2 Q_T(\cdot)/\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}'$. Inserting (S.97) in (S.96) we then find

$$T^{1/2}\left(\widehat{\boldsymbol{\tau}}-\boldsymbol{\tau}_0\right)=-\left(\frac{\partial^2 Q_T(\overline{\boldsymbol{\tau}})}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}'}\right)^{-1}T^{1/2}\frac{\partial Q_T(\boldsymbol{\tau}_0)}{\partial\boldsymbol{\tau}}-\left(\frac{\partial^2 Q_T(\overline{\boldsymbol{\tau}})}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}'}\right)^{-1}T^{1/2}\frac{\partial S_T(\widehat{\boldsymbol{\vartheta}})}{\partial\boldsymbol{\tau}}. \tag{S.98}$$

Now, by Hualde and Robinson (2011) (see the proof of their Theorem 2.2), the first term on the right-hand side of (S.98) has a $N\left(0,\boldsymbol{A}^{-1}\right)$ limiting distribution, and $\partial^2 Q_T(\overline{\boldsymbol{\tau}})/\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}'$ converges in probability to a nonsingular matrix. Thus, in view of (S.98), Theorem 2(ii) follows because $T^{1/2}\partial S_T(\widehat{\boldsymbol{\vartheta}})/\partial\boldsymbol{\tau}=o_p\left(1\right)$ by Lemma S.8.

### S.2.5 Proof of Corollary 1

First, by very similar steps to those in the proof of Theorem 1 it is possible to show that $\widehat{\boldsymbol{\tau}}_\gamma \to_p \boldsymbol{\tau}_0$ as $T \to \infty$, when $\gamma_0 + 1/2 > \delta_0$ or $\gamma_0 + 1/2 < \delta_0$. Given this result, the proof for case (ii) is basically identical to that of Theorem 2(ii), so we focus on case (i). By the mean value theorem,

$$\widehat{\boldsymbol{\tau}}_\gamma - \boldsymbol{\tau}_0 = -\left(\frac{\partial^2 R_T(\overline{\boldsymbol{\tau}}, \gamma_0)}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}'}\right)^{-1} \frac{\partial R_T(\boldsymbol{\tau}_0, \gamma_0)}{\partial \boldsymbol{\tau}},$$

where $\overline{\boldsymbol{\tau}}$ represents an intermediate point which is allowed to vary across the different rows of $\partial^2 R_T(\cdot)/\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}'$. Then, following the proof of Theorem 2(i), it is immediate to show that $T^{1/2}(\widehat{\boldsymbol{\tau}}_\gamma - \boldsymbol{\tau}_0) \to_d N\left(\mathbf{0}_{p+1}, \boldsymbol{A}^{-1}\right)$.

Next, noting (12), we have the mean value theorem expansion

$$\begin{aligned} T^{\gamma_0-\delta_0+1/2}(\hat{\mu}_\gamma - \mu_0) &= T^{\gamma_0-\delta_0+1/2} \sum_{t=1}^{T} \varepsilon_t(\widehat{\boldsymbol{\tau}}_\gamma) k_{t,T}(\gamma_0, \widehat{\delta}_\gamma, \widehat{\boldsymbol{\varphi}}_\gamma) \\ &= T^{\gamma_0-\delta_0+1/2} \sum_{t=1}^{T} \varepsilon_t(\boldsymbol{\tau}_0) k_{t,T}(\gamma_0, \delta_0, \boldsymbol{\varphi}_0) \\ &\quad + T^{\gamma_0-\delta_0+1/2} (\widehat{\boldsymbol{\tau}}_\gamma - \boldsymbol{\tau}_0)' \sum_{t=1}^{T} \frac{\partial}{\partial \boldsymbol{\tau}} \varepsilon_t(\boldsymbol{\tau}) k_{t,T}(\gamma_0, \delta, \boldsymbol{\varphi})\big|_{\boldsymbol{\tau}=\overline{\boldsymbol{\tau}}}, \end{aligned} \tag{S.99}$$

where $\overline{\boldsymbol{\tau}}$ represents again an intermediate value. Here, noting that $(\widehat{\boldsymbol{\tau}}_\gamma - \boldsymbol{\tau}_0)$ is $O_p(T^{-1/2})$, it can be easily shown that second term on the right-hand side of (S.99) is $o_p(1)$ by the same arguments as for (S.85). Also, by the same central limit theorem as applied in the proof of Theorem 2(i), the first term on the right-hand side of (S.99) is asymptotically normal with mean zero and variance given by

$$\begin{aligned} \lim_{T\to\infty} T^{2(\gamma_0-\delta_0+1/2)} \sigma_0^2 \sum_{t=1}^{T} k_{t,T}(\gamma_0, \delta_0, \boldsymbol{\varphi}_0)^2 &= \lim_{T\to\infty} \frac{\sigma_0^2}{T^{-2(\gamma_0-\delta_0+1/2)} \sum_{t=1}^{T} c_t(\gamma_0, \delta_0, \boldsymbol{\varphi}_0)^2} \\ &= \frac{\sigma_0^2 \Gamma^2(\gamma_0 - \delta_0 + 1)\, 2\,(\gamma_0 - \delta_0 + 1/2)}{\phi^2(1; \boldsymbol{\varphi}_0)\, \Gamma^2(\gamma_0 + 1)} \end{aligned}$$

using, in turn, Lemmas S.15, S.13, S.11, and S.10.

Finally, the joint convergence in (18) can be immediately established by direct application of the Cramer-Wold device, noting that

$$T^{1/2} \boldsymbol{P}_{\gamma,T}^{-1} \begin{pmatrix} \widehat{\boldsymbol{\tau}}_\gamma - \boldsymbol{\tau}_0 \\ \widehat{\mu}_\gamma - \mu_0 \end{pmatrix} = \begin{pmatrix} \frac{\partial^2 R_T(\boldsymbol{\tau}_0, \gamma_0)}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}'} & 0 \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} T^{-\frac{1}{2}} \sum_{t=1}^{T} \varepsilon_t \sum_{j=1}^{\infty} \boldsymbol{m}_j(\boldsymbol{\varphi}_0)\, \varepsilon_{t-j} \\ T^{\gamma_0-\delta_0+1/2} \sum_{t=1}^{T} \varepsilon_t k_{t,T}(\gamma_0, \delta_0, \boldsymbol{\varphi}_0) \end{pmatrix} + o_p(1),$$

and

$$T^{\gamma_0-\delta_0+1/2} \sum_{t=1}^{T} k_{t,T}(\gamma_0, \delta_0, \boldsymbol{\varphi}_0) \sum_{j=1}^{\infty} \boldsymbol{m}_j(\boldsymbol{\varphi}_0)\, \varepsilon_{t-j} = o_p(1),$$

as in the proof of (S.140).

## S.3 Auxiliary lemmas

**Lemma S.1** *Under Assumptions A1–A3,* $T^{-1}\sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}_0) \to_p \sigma_0^2$.

**Proof.** Clearly

$$\frac{1}{T}\sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}_0) = \frac{1}{T}\sum_{t=1}^{T} \left(\phi(L;\boldsymbol{\varphi}_0)\{u_t \mathbb{I}(t \geq 1)\}\right)^2 - \frac{1}{T}\left(\sum_{t=1}^{T} \phi(L;\boldsymbol{\varphi}_0)\{u_t \mathbb{I}(t \geq 1)\} h_{t,T}(\gamma_0,\delta_0,\boldsymbol{\varphi}_0)\right)^2. \quad \text{(S.100)}$$

In view of (S.92), (S.93), by Assumptions A1 and A2 and simple application of Lemma S.16, the second term on the right-hand side of (S.100) is $O_p\left(T^{2\theta-1}\right) = o_p(1)$ by choosing $\theta < 1/2$. Then the required result holds by (S.10). ■

**Lemma S.2** *Under Assumptions A1 and A3, for any* $g > 0$,

$$\lim_{T\to\infty} \inf_{\gamma_0-\delta\geq -1/2+g, |\gamma-\gamma_0|\geq g, \boldsymbol{\varphi}\in\Psi} \frac{1}{T^{2(\gamma_0-\delta)+1}} \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta}) > \epsilon.$$

**Proof.** Letting $\alpha > 0$ be arbitrarily small (in particular $\alpha < (\varsigma - 1/2)/3$, which implies $\alpha < 1/2$ and also $\alpha < g$) and defining $\Phi_1 = \{\boldsymbol{\vartheta} \in \Xi : \gamma - \delta \leq -1/2 - \alpha\}$, $\Phi_2 = \{\boldsymbol{\vartheta} \in \Xi : -1/2 - \alpha \leq \gamma - \delta \leq -1/2 + \alpha\}$, and $\Phi_3 = \{\boldsymbol{\vartheta} \in \Xi : \gamma - \delta \geq -1/2 + \alpha\}$, the result holds on showing

$$\lim_{T\to\infty} \inf_{\{\gamma_0-\delta\geq -1/2+g, |\gamma-\gamma_0|\geq g\}\cap\Phi_j} \frac{1}{T^{2(\gamma_0-\delta)+1}} \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta}) > \epsilon, \quad \text{(S.101)}$$

for $j = 1, 2, 3$. We first deal with $j = 1, 2$. Clearly

$$\frac{1}{T^{2(\gamma_0-\delta)+1}} \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta}) = \frac{\mu_0^2}{T^{2(\gamma_0-\delta)+1}} \sum_{t=1}^{T} c_t^2(\gamma_0,\delta,\boldsymbol{\varphi}) - \frac{\mu_0^2}{T^{2(\gamma_0-\delta)+1}} \left(\sum_{t=1}^{T} c_t(\gamma_0,\delta,\boldsymbol{\varphi}) h_{t,T}(\gamma,\delta,\boldsymbol{\varphi})\right)^2,$$

so because $|\gamma - \gamma_0| \geq g$ and by application of Lemma S.15, noting that $\gamma_0 - \delta \geq -1/2 + g > -1/2$, $\mu_0 \neq 0$, and (7), (S.101) for $j = 1, 2$ holds on showing

$$\lim_{T\to\infty} \inf_{\gamma_0-\delta\geq -1/2+g} \frac{1}{T^{2(\gamma_0-\delta)+1}} \sum_{t=1}^{T} b_t^2(\gamma_0,\delta) > \epsilon, \quad \text{(S.102)}$$

$$\sup_{\{\gamma_0-\delta\geq -1/2+g\}\cap\Phi_j} \frac{1}{T^{2(\gamma_0-\delta)+1}} \left(\sum_{t=1}^{T} c_t(\gamma_0,\delta,\boldsymbol{\varphi}) h_{t,T}(\gamma,\delta,\boldsymbol{\varphi})\right)^2 = o(1). \quad \text{(S.103)}$$

First, (S.102) follows by almost identical arguments to those in the proof of (S.198).

Next, we show (S.103) for $j = 1$. By (S.197) of Lemma S.15, the left-hand side of (S.103) is bounded by

$$\sup_{\{\gamma_0-\delta\geq-1/2+g\}\cap\Phi_1}\left(T^{-1/2-(\gamma_0-\delta)}\sum_{t=1}^{T}c_t\left(\gamma_0,\delta,\boldsymbol{\varphi}\right)c_t\left(\gamma,\delta,\boldsymbol{\varphi}\right)\right)^2. \tag{S.104}$$

By Lemma S.14, (S.104) is $O\left(T^{-2\alpha}\right)=o\left(1\right)$ to conclude the proof of (S.103), and therefore that of (S.101), for $j = 1$. Regarding $j = 2$, the left-hand side of (S.103) is bounded by

$$\frac{\left(\sup_{\{\gamma_0-\delta\geq-1/2+g\}\cap\Phi_2}T^{-(\gamma_0-2\delta+\gamma+1)}\sum_{t=1}^{T}c_t\left(\gamma_0,\delta,\boldsymbol{\varphi}\right)c_t\left(\gamma,\delta,\boldsymbol{\varphi}\right)\right)^2}{\inf_{\Phi_2}T^{-2(\gamma-\delta)-1}\sum_{t=1}^{T}c_t^2\left(\gamma,\delta,\boldsymbol{\varphi}\right)}, \tag{S.105}$$

where the denominator can be made arbitrarily large by setting $\alpha$ close enough to zero, see (S.198) of Lemma S.15. By (S.193) of Lemma S.14 the square-root of the numerator of (S.105) is $O(T^{-g+\alpha}\sum_{t=1}^{T}t^{g-1-\alpha})=O\left(1\right)$. This completes the proof of (S.103), and hence that of (S.101), for $j = 2$.

Finally we show (S.101) for $j = 3$. By very similar steps to those in the proofs of (S.195), (S.196) in Lemma S.15, noting that $\gamma_0-\delta\geq-1/2+g$, it is straightforward to show that

$$\frac{1}{T^{2(\gamma_0-\delta)+1}}\sum_{t=1}^{T}d_t^2\left(\boldsymbol{\vartheta}\right)=\frac{\mu_0^2\phi^2(1;\boldsymbol{\varphi})}{T^{2(\gamma_0-\delta)+1}}\left(\sum_{t=1}^{T}b_t^2\left(\gamma_0,\delta\right)-\frac{\left(\sum_{t=1}^{T}b_t\left(\gamma_0,\delta\right)b_t\left(\gamma,\delta\right)\right)^2}{\sum_{t=1}^{T}b_t^2\left(\gamma,\delta\right)}\right)$$
$$+\,q_{1T}\left(\gamma_0,\gamma,\delta,\boldsymbol{\varphi}\right), \tag{S.106}$$

where $\sup_{\{\gamma_0-\delta\geq-1/2+g\}\cap\Phi_3}\left|q_{1T}\left(\gamma_0,\gamma,\delta,\boldsymbol{\varphi}\right)\right|=o\left(1\right)$. Next, by almost identical steps as in the proofs of (S.184) and (S.198), it can be easily shown that the first term on the right-hand side of (S.106) equals

$$\frac{\mu_0^2\phi^2(1;\boldsymbol{\varphi})\Gamma^2\left(\gamma_0+1\right)}{T^{2(\gamma_0-\delta)+1}}\left(\sum_{t=1}^{T}\pi_t^2\left(\gamma_0+1-\delta\right)-\frac{\left(\sum_{t=1}^{T}\pi_t\left(\gamma_0+1-\delta\right)\pi_t\left(\gamma+1-\delta\right)\right)^2}{\sum_{t=1}^{T}\pi_t^2\left(\gamma+1-\delta\right)}\right) \tag{S.107}$$
$$+\,q_{2T}\left(\gamma_0,\gamma,\delta,\boldsymbol{\varphi}\right),$$

where $\sup_{\{\gamma_0-\delta\geq-1/2+g\}\cap\Phi_3}\left|q_{2T}\left(\gamma_0,\gamma,\delta,\boldsymbol{\varphi}\right)\right|=o\left(1\right)$. Approximating sums by integrals, by (7) and given that $\Gamma^2\left(\gamma_0+1\right)>0$, the first term on (S.107) is bounded from below by

$$\epsilon\inf_{\{\gamma_0-\delta\geq-1/2+g\}\cap\Phi_3}\frac{1}{\Gamma^2\left(\gamma_0-\delta+1\right)}\left(\frac{1}{T^{2(\gamma_0-\delta)+1}}\sum_{t=1}^{T}t^{2(\gamma_0-\delta)}-\frac{\left(\frac{1}{T^{\gamma_0+\gamma-2\delta+1}}\sum_{t=1}^{T}t^{\gamma_0+\gamma-2\delta}\right)^2}{\frac{1}{T^{2(\gamma-\delta)+1}}\sum_{t=1}^{T}t^{2(\gamma-\delta)}}\right)$$
$$=\epsilon\inf_{\{\gamma_0-\delta\geq-1/2+g\}\cap\Phi_3}\frac{\left(\gamma_0-\gamma\right)^2}{\Gamma^2\left(\gamma_0-\delta+1\right)\left(2\left(\gamma_0-\delta\right)+1\right)\left(\gamma_0+\gamma-2\delta+1\right)^2}-o\left(1\right)$$
$$\geq\epsilon\inf_{\gamma_0-\delta\geq-1/2+g}\frac{g^2}{\Gamma^2\left(\gamma_0-\delta+1\right)2g\left(\alpha+g\right)^2}-o\left(1\right),$$

which is positive and bounded away from zero, to complete the proof of (S.101) for $j = 3$. ■

**Lemma S.3** *Under Assumptions A1 and A3, for $i = 1, \ldots, 4$,*

$$\frac{1}{T^{\gamma_0-\delta+1}}\frac{\partial \overline{d}_T(\boldsymbol{\vartheta})}{\partial \gamma} = \frac{\mu_0 \phi(1;\boldsymbol{\varphi})\Gamma(\gamma_0+1)}{\Gamma(\gamma_0-\delta+1)}\frac{2(\gamma-\delta)^2+2(\gamma-\delta)-(\gamma_0-\delta)}{(\gamma-\delta+1)^2(\gamma_0+\gamma-2\delta+1)^2} + g_T(\boldsymbol{\vartheta}), \quad \text{(S.108)}$$

*where $\sup_{\overline{\mathcal{H}}_i}|g_T(\boldsymbol{\vartheta})| = o(1)$, and for an arbitrarily large $K$ (setting $\xi$ large enough),*

$$\lim_{T\to\infty}\inf_{\overline{\mathcal{H}}_i}\frac{T^{2\varkappa_i}}{T^{2(\gamma_0-\delta)+2}}\overline{d}_T^2(\boldsymbol{\vartheta}) > K. \quad \text{(S.109)}$$

**Proof.** First, $\partial \overline{d}_T(\boldsymbol{\vartheta})/\partial\gamma$ equals

$$-\frac{\mu_0\sum_{t=1}^T c_t^{(1)}(\gamma,\delta,\boldsymbol{\varphi})\sum_{j=1}^T c_j(\gamma_0,\delta,\boldsymbol{\varphi})c_j(\gamma,\delta,\boldsymbol{\varphi})}{\sum_{j=1}^T c_j^2(\gamma,\delta,\boldsymbol{\varphi})}$$
$$-\frac{\mu_0\sum_{t=1}^T c_t(\gamma,\delta,\boldsymbol{\varphi})\sum_{j=1}^T c_j(\gamma_0,\delta,\boldsymbol{\varphi})c_j^{(1)}(\gamma,\delta,\boldsymbol{\varphi})}{\sum_{j=1}^T c_j^2(\gamma,\delta,\boldsymbol{\varphi})}$$
$$+\frac{2\mu_0\sum_{t=1}^T c_t(\gamma,\delta,\boldsymbol{\varphi})\sum_{j=1}^T c_j(\gamma_0,\delta,\boldsymbol{\varphi})c_j(\gamma,\delta,\boldsymbol{\varphi})\sum_{k=1}^T c_k(\gamma,\delta,\boldsymbol{\varphi})c_k^{(1)}(\gamma,\delta,\boldsymbol{\varphi})}{(\sum_{j=1}^T c_j^2(\gamma,\delta,\boldsymbol{\varphi}))^2}.$$

Noting that in $\cup_{i=1}^4\overline{\mathcal{H}}_i$, $\gamma_0-\delta\geq\eta$ and $\gamma-\delta\geq\eta-\varrho$, proceeding as in the proof of Lemma S.15,

$$\frac{1}{T^{\gamma_0-\delta+1}}\frac{\partial\overline{d}_T(\gamma,\boldsymbol{\tau})}{\partial\gamma} = -\frac{\mu_0\phi(1;\boldsymbol{\varphi})}{T^{\gamma_0-\delta+1}}\left(\frac{\sum_{t=1}^T b_t^{(1)}(\gamma,\delta)\sum_{j=1}^T b_j(\gamma_0,\delta)b_j(\gamma,\delta)}{\sum_{j=1}^T b_j^2(\gamma,\delta)}\right.$$
$$+\frac{\sum_{t=1}^T b_t(\gamma,\delta)\sum_{j=1}^T b_j(\gamma_0,\delta)b_j^{(1)}(\gamma,\delta)}{\sum_{j=1}^T b_j^2(\gamma,\delta)}$$
$$\left.-\frac{2\sum_{t=1}^T b_t(\gamma,\delta)\sum_{j=1}^T b_j(\gamma_0,\delta)b_j(\gamma,\delta)\sum_{k=1}^T b_k(\gamma,\delta)b_k^{(1)}(\gamma,\delta)}{(\sum_{j=1}^T b_j^2(\gamma,\delta))^2}\right)$$
$$+g_{1T}(\boldsymbol{\vartheta}), \quad \text{(S.110)}$$

where $\sup_{\overline{\mathcal{H}}_i}|g_{1T}(\boldsymbol{\vartheta})| = o(1)$. Now, substituting (S.185) into (S.110),

$$\frac{1}{T^{\gamma_0-\delta+1}}\frac{\partial\overline{d}_T(\gamma,\boldsymbol{\tau})}{\partial\gamma} = -\frac{\mu_0\phi(1;\boldsymbol{\varphi})}{T^{\gamma_0-\delta+1}}\left(\frac{\sum_{t=1}^T \log(t/T)b_t(\gamma,\delta)\sum_{j=1}^T b_j(\gamma_0,\delta)b_j(\gamma,\delta)}{\sum_{j=1}^T b_j^2(\gamma,\delta)}\right.$$
$$+\frac{\sum_{t=1}^T b_t(\gamma,\delta)\sum_{j=1}^T \log(j/T)b_j(\gamma_0,\delta)b_j(\gamma,\delta)}{\sum_{j=1}^T b_j^2(\gamma,\delta)}$$
$$\left.-\frac{2\sum_{t=1}^T b_t(\gamma,\delta)\sum_{j=1}^T b_j(\gamma_0,\delta)b_j(\gamma,\delta)\sum_{k=1}^T \log(k/T)b_k^2(\gamma,\delta)}{(\sum_{j=1}^T b_j^2(\gamma,\delta))^2}\right)+g_{2T}(\boldsymbol{\vartheta}),$$
$$\text{(S.111)}$$

where $\sup_{\overline{\mathcal{H}}_i}|g_{2T}(\boldsymbol{\vartheta})| = o(1)$, noting that the contribution of the second term on the right hand side of (S.185) and the $\log T$ terms cancel. Hence, using (S.184), it can be easily shown that

$$\frac{1}{T^{\gamma_0-\delta+1}}\frac{\partial\overline{d}_T(\gamma,\boldsymbol{\tau})}{\partial\gamma}=-\frac{\mu_0\phi(1;\boldsymbol{\varphi})\Gamma(\gamma_0+1)}{T^{\gamma_0-\delta+1}\Gamma(\gamma_0+1-\delta)}\left(\frac{\sum_{t=1}^{T}\log(t/T)t^{\gamma-\delta}\sum_{j=1}^{T}j^{\gamma_0+\gamma-2\delta}}{\sum_{j=1}^{T}j^{2\gamma-2\delta}}\right.$$
$$+\frac{\sum_{t=1}^{T}t^{\gamma-\delta}\sum_{j=1}^{T}\log(j/T)j^{\gamma_0+\gamma-2\delta}}{\sum_{j=1}^{T}j^{2\gamma-2\delta}}$$
$$\left.-\frac{2\sum_{t=1}^{T}t^{\gamma-\delta}\sum_{j=1}^{T}j^{\gamma_0+\gamma-2\delta}\sum_{k=1}^{T}\log(k/T)k^{2\gamma-2\delta}}{(\sum_{j=1}^{T}j^{2\gamma-2\delta})^2}\right)+g_{3T}(\boldsymbol{\vartheta}),$$

where $\sup_{\overline{\mathcal{H}}_i}|g_{3T}(\boldsymbol{\vartheta})|=o(1)$. Finally, (S.108) then follows by approximating sums by integrals, see Lemma S.10.

Next, because $\overline{d}_T(\gamma_0,\boldsymbol{\tau})=0$, the mean value theorem yields $\overline{d}_T(\gamma,\boldsymbol{\tau})=(\gamma-\gamma_0)\partial\overline{d}_T(\overline{\gamma},\boldsymbol{\tau})/\partial\gamma$, where $|\overline{\gamma}-\gamma_0|\le|\gamma-\gamma_0|$, so the left-hand side of (S.109) can be bounded from below by

$$\lim_{T\to\infty}\inf_{\overline{\mathcal{H}}_i}T^{2\varkappa_i}(\gamma-\gamma_0)^2\left(\frac{1}{T^{\gamma_0-\delta+1}}\frac{\partial\overline{d}_T(\overline{\gamma},\boldsymbol{\tau})}{\partial\gamma}\right)^2$$
$$\ge\lim_{T\to\infty}\inf_{\overline{\mathcal{H}}_i}T^{2\varkappa_i}(\gamma-\gamma_0)^2\inf_{\overline{\mathcal{H}}_i}\left(\frac{1}{T^{\gamma_0-\delta+1}}\frac{\partial\overline{d}_T(\gamma,\boldsymbol{\tau})}{\partial\gamma}\right)^2$$
$$=\xi^2\lim_{T\to\infty}\inf_{\overline{\mathcal{H}}_i}\left(\frac{1}{T^{\gamma_0-\delta+1}}\frac{\partial\overline{d}_T(\gamma,\boldsymbol{\tau})}{\partial\gamma}\right)^2.$$

Thus, setting $\xi$ large enough, (S.109) follows if

$$\lim_{T\to\infty}\inf_{\overline{\mathcal{H}}_i}\left(\frac{1}{T^{\gamma_0-\delta+1}}\frac{\partial\overline{d}_T(\boldsymbol{\tau},\gamma)}{\partial\gamma}\right)^2>\epsilon, \tag{S.112}$$

which, noting that $\gamma_0-\delta\ge\eta$, is a consequence of (7) and (S.108) because

$$\inf_{\overline{\mathcal{H}}_i}\left(2(\gamma-\delta)^2+2(\gamma-\delta)-(\gamma_0-\delta)\right)=\inf_{\overline{\mathcal{H}}_i}\left(2(\gamma-\delta)^2+(\gamma-\delta)-(\gamma_0-\gamma)\right)$$
$$\ge2(\eta-\varrho)^2+\eta-2\varrho>0.$$

■

**Lemma S.4** *Under the conditions of Theorem 1(ii) it holds that*

$$\Pr\left(\inf_{\boldsymbol{\vartheta}\in\overline{M}_\varepsilon}R_T(\boldsymbol{\vartheta})\le\sigma_0^2+\epsilon\right)\to0 \text{ as } T\to\infty.$$

**Proof.** Recall the intervals $\mathcal{I}_i$ and define $\mathcal{W}_i=\{\boldsymbol{\vartheta}\in\overline{M}_\varepsilon:\delta\in\mathcal{I}_i\}$ for $i=1,2,3$, and $\mathcal{W}_4=\{\boldsymbol{\vartheta}\in\overline{M}_\varepsilon:\delta\in\mathcal{I}_4\cup\mathcal{I}_5\}$. Then the result follows on showing

$$\Pr\left(\inf_{\mathcal{W}_i}R_T(\boldsymbol{\vartheta})\le\sigma_0^2+\epsilon\right)\to0\text{ as }T\to\infty \tag{S.113}$$

for $i = 1, \ldots, 4$, noting that

$$R_T(\boldsymbol{\vartheta}) = \frac{1}{T}\left(\sum_{t=1}^{T}\left(\phi(L;\boldsymbol{\varphi})\Delta_+^{\delta}x_t\right)^2 - \left(\sum_{t=1}^{T}\phi(L;\boldsymbol{\varphi})\Delta_+^{\delta}x_t h_{t,T}(\gamma,\delta,\boldsymbol{\varphi})\right)^2\right). \tag{S.114}$$

**Proof of (S.113) for** $i = 4$. Given (S.82), we first apply the bound

$$\frac{1}{T}\sum_{t=1}^{T}\left(\phi(L;\boldsymbol{\varphi})\Delta_+^{\delta}x_t\right)^2 \geq \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2(\boldsymbol{\tau}) - \frac{2|\mu_0|}{T}\left|\sum_{t=1}^{T}\varepsilon_t(\boldsymbol{\tau})c_t(\gamma_0,\delta,\boldsymbol{\varphi})\right| \tag{S.115}$$

and note that $\delta_0 - \delta \leq 1/2 - \eta$ when $\delta \in \mathcal{I}_4 \cup \mathcal{I}_5$, so $\Delta_+^{\delta-\delta_0}u_t$ is asymptotically stationary. In view of (S.114) and (S.115), the proof of (S.113) for $i = 4$ then follows by Hualde and Robinson (2011) (see the proof of their (2.7) for $i = 4$) by showing

$$\sup_{\mathcal{W}_4}\frac{1}{T}\left|\sum_{t=1}^{T}\varepsilon_t(\boldsymbol{\tau})c_t(\gamma_0,\delta,\boldsymbol{\varphi})\right| = o_p(1), \tag{S.116}$$

$$\sup_{\mathcal{W}_4}\frac{1}{T}\left(\sum_{t=1}^{T}\phi(L;\boldsymbol{\varphi})\Delta_+^{\delta}x_t h_{t,T}(\gamma,\delta,\boldsymbol{\varphi})\right)^2 = o_p(1). \tag{S.117}$$

First, noting that $\delta \in \mathcal{I}_4 \cup \mathcal{I}_5$ implies $\delta_0 - \delta \leq 1/2 - \eta$ and $\gamma_0 - \delta \leq 1/2 + \gamma_0 - \delta_0 - \eta$, (S.235) of Lemma S.21 implies that the left-hand side of (S.116) is $O_p(T^{\gamma_0-\delta_0+1/2-2\eta} + T^{-\varsigma-\eta} + T^{-1}\log T) = o_p(1)$.

Next, using (S.82), (S.117) follows by showing

$$\sup_{\mathcal{W}_4} T^{-1/2}\sum_{t=1}^{T}\varepsilon_t(\boldsymbol{\tau})h_{t,T}(\gamma,\delta,\boldsymbol{\varphi}) = o_p(1), \tag{S.118}$$

$$\sup_{\mathcal{W}_4} T^{-1/2}\sum_{t=1}^{T}c_t(\gamma_0,\delta,\boldsymbol{\varphi})h_{t,T}(\gamma,\delta,\boldsymbol{\varphi}) = o(1). \tag{S.119}$$

Here, (S.222) of Lemma S.18 shows that the left-hand side of (S.118) is $O_p(T^{\theta-1/2} + T^{-\eta}) = o_p(1)$ by choosing $\theta < 1/2$, while (S.232) of Lemma S.20 shows that the left-hand side of (S.119) $O(\left(T^{\theta-1/2} + T^{\gamma_0+1/2-\delta_0-\eta}\right)\log T) = o(1)$, to conclude the proof of (S.113) for $i = 4$.

**Proof of (S.113) for** $i = 3$. Noting (S.35), (S.114), and (S.115), the proof follows on showing

$$\sup_{\mathcal{W}_3}\frac{1}{T}\left|\sum_{t=1}^{T}\varepsilon_t(\boldsymbol{\tau})c_t(\gamma_0,\delta,\boldsymbol{\varphi})\right| = O_p(1), \tag{S.120}$$

$$\sup_{\mathcal{W}_3}\frac{1}{T}\left(\sum_{t=1}^{T}\phi(L;\boldsymbol{\varphi})\Delta_+^{\delta}x_t h_{t,T}(\gamma,\delta,\boldsymbol{\varphi})\right)^2 = O_p(1). \tag{S.121}$$

Both (S.120) and (S.121) follow straightforwardly by identical steps as those given in the proofs of (S.116) and (S.117) just replacing $\eta$ by 0.

**Proof of (S.113) for** $i = 2$. Clearly,

$$\Pr\left(\inf_{\mathcal{W}_2} R_T(\boldsymbol{\vartheta}) \le \sigma_0^2 + \epsilon\right) \le \Pr\left(\inf_{\mathcal{W}_2} \frac{T^{2(\delta_0-\delta)}}{T} \inf_{\mathcal{W}_2} \frac{T}{T^{2(\delta_0-\delta)}} R_T(\boldsymbol{\vartheta}) \le \sigma_0^2 + \epsilon\right)$$
$$= \Pr\left(\inf_{\mathcal{W}_2} \frac{T}{T^{2(\delta_0-\delta)}} R_T(\boldsymbol{\vartheta}) \le \sigma_0^2 + \epsilon\right). \quad \text{(S.122)}$$

Thus, in view of (S.43), (S.114), and (S.115), (S.113) for $i = 2$ follows on showing

$$\sup_{\mathcal{W}_2} \frac{1}{T^{2(\delta_0-\delta)}} \left| \sum_{t=1}^{T} \varepsilon_t(\boldsymbol{\tau}) c_t(\gamma_0, \delta, \boldsymbol{\varphi}) \right| = O_p(1), \quad \text{(S.123)}$$

$$\sup_{\mathcal{W}_2} \frac{1}{T^{2(\delta_0-\delta)}} \left( \sum_{t=1}^{T} \phi(L; \boldsymbol{\varphi}) \Delta_+^{\delta} x_t h_{t,T}(\gamma, \delta, \boldsymbol{\varphi}) \right)^2 = O_p(1). \quad \text{(S.124)}$$

The proofs of (S.123) and (S.124) are almost identical to those of (S.116) and (S.117), taking into account the different normalization, which implies using (S.233) instead of (S.232) in Lemma S.20, (S.236) instead of (S.235) in Lemma S.21, and (S.223) instead of (S.222) in Lemma S.18.

**Proof of (S.113) for** $i = 1$. Following identical steps to those given in (S.122),

$$\Pr\left(\inf_{\mathcal{W}_1} R_T(\boldsymbol{\vartheta}) \le \sigma_0^2 + \epsilon\right) \le \Pr\left(\inf_{\mathcal{W}_1} \frac{T}{T^{2(\delta_0-\delta)}} R_T(\boldsymbol{\vartheta}) \le \frac{\sigma_0^2 + \epsilon}{T^{2\eta}}\right).$$

Letting $\alpha > 0$ be arbitrarily small (in particular $\alpha < (\varsigma - 1/2)/3$) and defining the sets

$$\Phi_1 = \{\boldsymbol{\vartheta} \in \Xi : \gamma - \delta \le -1/2 - \alpha\}, \quad \Phi_2 = \{\boldsymbol{\vartheta} \in \Xi : -1/2 - \alpha \le \gamma - \delta \le -1/2 + \alpha\},$$
$$\Phi_3 = \{\boldsymbol{\vartheta} \in \Xi : \gamma - \delta \ge -1/2 + \alpha\},$$

the required result follows on showing

$$\Pr\left(\inf_{\mathcal{W}_1 \cap \Phi_j} \frac{T}{T^{2(\delta_0-\delta)}} R_T(\boldsymbol{\vartheta}) > \epsilon\right) \to 1 \text{ as } T \to \infty \quad \text{(S.125)}$$

for $j = 1, 2, 3$ and $\epsilon > 0$ arbitrarily small. In the proof of their (2.7) for $i = 1$, Hualde and Robinson (2011) showed that

$$\Pr\left(\inf_{\|\boldsymbol{\tau}-\boldsymbol{\tau}_0\| \ge \varepsilon, \boldsymbol{\tau} \in \mathcal{T}_1} \frac{1}{T^{2(\delta_0-\delta)}} \sum_{t=1}^{T} \varepsilon_t^2(\boldsymbol{\tau}) > \epsilon\right) \to 1 \text{ as } T \to \infty, \quad \text{(S.126)}$$

although their proof based on the Cauchy-Schwarz inequality does not appear sufficient; see the discussion just below (S.52). We shortly prove (S.126), so in view of (S.82), (S.114), and (S.115), (S.125) for $j = 1, 2$ holds if we also show that

$$\sup_{\mathcal{W}_1} \frac{1}{T^{2(\delta_0-\delta)}} \sum_{t=1}^{T} \varepsilon_t(\boldsymbol{\tau}) c_t(\gamma_0, \delta, \boldsymbol{\varphi}) = o_p(1), \quad \text{(S.127)}$$

$$\sup_{\mathcal{W}_1} \frac{1}{T^{\delta_0-\delta}} \sum_{t=1}^{T} c_t(\gamma_0, \delta, \boldsymbol{\varphi}) h_{t,T}(\gamma, \delta, \boldsymbol{\varphi}) = o(1), \quad \text{(S.128)}$$

$$\sup_{\mathcal{W}_1 \cap \Phi_j} \frac{1}{T^{2(\delta_0-\delta)}} \left( \sum_{t=1}^{T} \varepsilon_t(\boldsymbol{\tau}) h_{t,T}(\gamma, \delta, \boldsymbol{\varphi}) \right)^2 = o_p(1). \quad \text{(S.129)}$$

First, noting that $\delta_0 - \delta \geq 1/2 + \eta$ and $\gamma_0 - \delta \geq 1/2 + \eta + \gamma_0 - \delta_0$, by (S.236) of Lemma S.21 and (S.233) of Lemma S.20 with $\theta < 1/2 + \eta$, the left-hand sides of (S.127) and (S.128) are $O_p(T^{\max\{\gamma_0-\delta_0+1/2, -\varsigma-\eta\}} + T^{-1-2\eta}\log T) = o_p(1)$ and $O(T^{\gamma_0-\delta_0+1/2} + T^{-1/2-\eta+\theta}\log T) = o(1)$, respectively. Next, for $j = 1$, by (S.197) of Lemma S.15, the left-hand side of (S.129) is

$$\sup_{\mathcal{W}_1\cap\Phi_1} T^{2(\delta-\delta_0)}\left(\sum_{t=1}^{T}\varepsilon_t(\boldsymbol{\tau})\,c_t(\gamma,\delta,\boldsymbol{\varphi})\right)^2,$$

which is easily shown to be $O_p(T^{-2\alpha}) = o_p(1)$ by (S.192) of Lemma S.14 and (S.219) of Lemma S.17. For $j = 2$, we use (S.3) to bound the left-hand side of (S.129) by

$$\frac{\sup_{\mathcal{W}_1\cap\Phi_2} T^{-1}\left(\sum_{t=1}^{T} T^{-(\delta_0-\delta)}\varepsilon_t(\boldsymbol{\tau})\,T^{-(\gamma-\delta)}c_t(\gamma,\delta,\boldsymbol{\varphi})\right)^2}{\inf_{\mathcal{W}_1\cap\Phi_2} T^{-2(\gamma-\delta)-1}\sum_{j=1}^{T} c_j^2(\gamma,\delta,\boldsymbol{\varphi})},$$

where the denominator can be made arbitrarily large by (S.198) of Lemma S.15 and the numerator is easily seen to be $O_p(1)$ by direct application of (S.193) of Lemma S.14 and (S.219) of Lemma S.17.

Thus, to prove (S.125) for $j = 1, 2$ it only remains to prove (S.126). By application of the bound (S.181) in Lemma S.12,

$$\frac{1}{T^{2(\delta_0-\delta)}}\sum_{t=1}^{T}\varepsilon_t^2(\boldsymbol{\tau}) \geq \left(1 + O(T^{-1})\right)\frac{\pi^2/4}{T^{2(\delta_0-\delta)+2}}\sum_{t=1}^{T}\varepsilon_t^2(\delta_0-\delta+1,\boldsymbol{\varphi}),$$

where $\varepsilon_t(\delta_0-\delta+1,\boldsymbol{\varphi}) = \sum_{j=1}^{t}\varepsilon_t(\boldsymbol{\tau})$ is defined in (S.27) and the $O(T^{-1})$ term does not depend on the parameters. Proceeding as in the proof of (S.52), we obtain the weak convergence

$$T^{\delta-\delta_0-1/2}\varepsilon_{[Tr]}(\delta-1,\boldsymbol{\varphi}) \Rightarrow \phi(1;\boldsymbol{\varphi})\,\omega(1;\boldsymbol{\varphi}_0)\,W(r;1+\delta_0-\delta), \tag{S.130}$$

so by the continuous mapping theorem,

$$\frac{\pi^2/4}{T^{2(\delta_0-\delta)+2}}\sum_{t=1}^{T}\varepsilon_t^2(\delta-1,\boldsymbol{\varphi}) \Rightarrow \frac{\pi^2}{4}\phi^2(1;\boldsymbol{\varphi})\,\omega^2(1;\boldsymbol{\varphi}_0)\int_0^1 W(r;1+\delta_0-\delta)^2\,dr.$$

It follows that the probability in (S.126) is bounded from below by

$$\begin{aligned}&\Pr\left(\left(1+O(T^{-1})\right)\inf_{\|\boldsymbol{\tau}-\boldsymbol{\tau}_0\|\geq\varepsilon,\boldsymbol{\tau}\in\mathcal{T}_1}\frac{\pi^2/4}{T^{2(\delta_0-\delta)+2}}\sum_{t=1}^{T}\varepsilon_t^2(\delta-1,\boldsymbol{\varphi}) > \epsilon\right)\\ &\quad\to \Pr\left(\inf_{\|\boldsymbol{\tau}-\boldsymbol{\tau}_0\|\geq\varepsilon,\boldsymbol{\tau}\in\mathcal{T}_1}\frac{\pi^2}{4}\phi^2(1;\boldsymbol{\varphi})\,\omega^2(1;\boldsymbol{\varphi}_0)\int_0^1 W(r;1+\delta_0-\delta)^2\,dr > \epsilon\right),\end{aligned}$$

and (S.126) follows because $\epsilon > 0$ is arbitrarily small. For additional details, see Lemma 3 of Johansen and Nielsen (2018) for the same argument.

We finally show (S.125) for $j = 3$. Using (S.15), applying (S.127) and (S.128) together with Lemmas S.16, S.17, we have

$$T^{1-2(\delta_0-\delta)}R_T(\boldsymbol{\vartheta}) \geq T^{-2(\delta_0-\delta)}\sum_{t=1}^{T}s_t^2(\boldsymbol{\vartheta}) + q_{1T}(\boldsymbol{\vartheta}),$$

where $\sup_{\mathcal{W}_1\cap\Phi_3}|q_{1T}(\boldsymbol{\vartheta})| = o_p(1)$. Thus, (S.125) for $j=3$ holds on showing

$$\Pr\left(\inf_{\mathcal{W}_1\cap\Phi_3}\frac{1}{T^{2(\delta_0-\delta)}}\sum_{t=1}^{T}s_t^2\left(\boldsymbol{\vartheta}\right)>\epsilon\right)\to 1 \text{ as } T\to\infty. \tag{S.131}$$

The proof of (S.131) is essentially identical to the main steps of the proof of (S.51). The only relevant difference is that we now need to establish a convergence result on a larger set where $\gamma-\delta\geq -1/2+\alpha$ (instead of $\gamma-\delta\geq\alpha$). However, this does not lead to any relevant changes in the proof because for fixed $\boldsymbol{\vartheta}$ such that $\gamma-\delta\geq-1/2+\alpha$ and $1+\delta_0-\delta\geq 3/2+\eta$ with $\alpha>0,\eta>0$, the integral $\int_0^1 u^{\gamma-\delta-1}W(u;1+\delta_0-\delta)du$ is well defined. This completes the proof of (S.125) for $j=3$ and therefore that of (S.113) for $i=1$. ■

**Lemma S.5** *Under the conditions of Theorem 2(i) it holds that:*

*(a) The first-order derivatives satisfy*

$$\frac{1}{T^{1/2}}\sum_{t=1}^{T}\left(s_t\left(\boldsymbol{\vartheta}_0\right)-\varepsilon_t\right)\frac{\partial s_{1t}\left(\boldsymbol{\tau}_0\right)}{\partial\boldsymbol{\tau}}=o_p\left(1\right),\ \frac{1}{T^{\gamma_0-\delta_0+1/2}}\sum_{t=1}^{T}\left(s_t\left(\boldsymbol{\vartheta}_0\right)-\varepsilon_t\right)\frac{\partial d_t\left(\boldsymbol{\vartheta}_0\right)}{\partial\gamma}=o_p\left(1\right), \tag{S.132}$$

$$\frac{1}{T^{1/2}}\sum_{t=1}^{T}s_t\left(\boldsymbol{\vartheta}_0\right)\frac{\partial s_{2t}\left(\boldsymbol{\vartheta}_0\right)}{\partial\boldsymbol{\tau}}=o_p\left(1\right),\ \frac{1}{T^{\gamma_0-\delta_0+1/2}}\sum_{t=1}^{T}s_t\left(\boldsymbol{\vartheta}_0\right)\frac{\partial s_{2t}\left(\boldsymbol{\vartheta}_0\right)}{\partial\gamma}=o_p\left(1\right). \tag{S.133}$$

*(b) The second-order derivatives satisfy*

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial s_{1t}\left(\boldsymbol{\tau}_0\right)}{\partial\boldsymbol{\tau}}\frac{\partial s_{2t}\left(\boldsymbol{\vartheta}_0\right)}{\partial\boldsymbol{\tau}'}=o_p\left(1\right),\ \frac{1}{T^{\gamma_0-\delta_0+1}}\sum_{t=1}^{T}\frac{\partial s_{1t}\left(\boldsymbol{\tau}_0\right)}{\partial\boldsymbol{\tau}}\frac{\partial s_{2t}\left(\boldsymbol{\vartheta}_0\right)}{\partial\gamma}=o_p\left(1\right),$$

$$\frac{1}{T^{\gamma_0-\delta_0+1}}\sum_{t=1}^{T}\frac{\partial s_{1t}\left(\boldsymbol{\tau}_0\right)}{\partial\boldsymbol{\tau}}\frac{\partial d_t\left(\boldsymbol{\vartheta}_0\right)}{\partial\gamma}=o_p\left(1\right),\ \frac{1}{T}\sum_{t=1}^{T}\frac{\partial s_{2t}\left(\boldsymbol{\vartheta}_0\right)}{\partial\boldsymbol{\tau}}\frac{\partial s_{2t}\left(\boldsymbol{\vartheta}_0\right)}{\partial\boldsymbol{\tau}'}=o_p\left(1\right),$$

$$\frac{1}{T^{\gamma_0-\delta_0+1}}\sum_{t=1}^{T}\frac{\partial s_{2t}\left(\boldsymbol{\vartheta}_0\right)}{\partial\boldsymbol{\tau}}\frac{\partial s_{2t}\left(\boldsymbol{\vartheta}_0\right)}{\partial\gamma}=o_p\left(1\right),\ \frac{1}{T^{\gamma_0-\delta_0+1}}\sum_{t=1}^{T}\frac{\partial s_{2t}\left(\boldsymbol{\vartheta}_0\right)}{\partial\boldsymbol{\tau}}\frac{\partial d_t\left(\boldsymbol{\vartheta}_0\right)}{\partial\gamma}=o_p\left(1\right),$$

$$\frac{1}{T^{2(\gamma_0-\delta_0)+1}}\sum_{t=1}^{T}\left(\frac{\partial s_{2t}\left(\boldsymbol{\vartheta}_0\right)}{\partial\gamma}\right)^2=o_p\left(1\right),\ \frac{1}{T^{2(\gamma_0-\delta_0)+1}}\sum_{t=1}^{T}\frac{\partial s_{2t}\left(\boldsymbol{\vartheta}_0\right)}{\partial\gamma}\frac{\partial d_t\left(\boldsymbol{\vartheta}_0\right)}{\partial\gamma}=o_p\left(1\right),$$

$$\frac{1}{T}\sum_{t=1}^{T}s_t(\boldsymbol{\vartheta}_0)\frac{\partial^2 s_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}'}=o_p\left(1\right),\ \frac{1}{T^{\gamma_0-\delta_0+1}}\sum_{t=1}^{T}s_t(\boldsymbol{\vartheta}_0)\frac{\partial^2 s_{2t}(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\tau}\partial\gamma}=o_p\left(1\right),$$

$$\frac{1}{T^{2(\gamma_0-\delta_0)+1}}\sum_{t=1}^{T}s_t(\boldsymbol{\vartheta}_0)\frac{\partial^2 s_{2t}(\boldsymbol{\vartheta}_0)}{\partial\gamma^2}=o_p\left(1\right),\ \frac{1}{T^{\gamma_0-\delta_0+1}}\sum_{t=1}^{T}s_t(\boldsymbol{\vartheta}_0)\frac{\partial^2 d_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\tau}\partial\gamma}=o_p\left(1\right),$$

$$\frac{1}{T^{2(\gamma_0-\delta_0)+1}}\sum_{t=1}^{T}s_t(\boldsymbol{\vartheta}_0)\frac{\partial^2 d_t(\boldsymbol{\vartheta}_0)}{\partial\gamma^2}=o_p\left(1\right).$$

**Proof.** First we show the first equality in (S.132). By definition of $s_t\left(\boldsymbol{\vartheta}\right)$,

$$s_t\left(\boldsymbol{\vartheta}_0\right)-\varepsilon_t=-\sum_{j=t}^{\infty}\phi_j(\boldsymbol{\varphi}_0)u_{t-j}-h_{t,T}\left(\gamma_0,\delta_0,\boldsymbol{\varphi}_0\right)\sum_{j=1}^{T}s_{1j}\left(\boldsymbol{\tau}_0\right)h_{j,T}\left(\gamma_0,\delta_0,\boldsymbol{\varphi}_0\right), \tag{S.134}$$

where $s_{1j}(\boldsymbol{\tau}_0) = \varepsilon_j - \sum_{k=j}^{\infty} \phi_k(\boldsymbol{\varphi}_0) u_{j-k}$, so the result holds if

$$\frac{1}{T^{1/2}} \sum_{t=1}^{T} \sum_{j=t}^{\infty} \phi_j(\boldsymbol{\varphi}_0) u_{t-j} \frac{\partial s_{1t}(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\tau}} = o_p(1), \tag{S.135}$$

$$\frac{1}{T^{1/2}} \sum_{t=1}^{T} h_{t,T}(\gamma_0, \delta_0, \boldsymbol{\varphi}_0) \frac{\partial s_{1t}(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\tau}} \sum_{j=1}^{T} s_{1j}(\boldsymbol{\tau}_0) h_{j,T}(\gamma_0, \delta_0, \boldsymbol{\varphi}_0) = o_p(1). \tag{S.136}$$

First, for $t \geq 2$,

$$\frac{\partial s_{1t}(\boldsymbol{\tau}_0)}{\partial \delta} = -\sum_{j=1}^{t-1} \sum_{k=0}^{j-1} \phi_k(\boldsymbol{\varphi}_0)(j-k)^{-1} u_{t-j}, \tag{S.137}$$

so, noting (6) and applying Lemma S.9,

$$E\left|\frac{\partial s_{1t}(\boldsymbol{\tau}_0)}{\partial \delta}\right| \leq K \sum_{j=1}^{t-1} \sum_{k=1}^{j-1} k^{-1-\varsigma}(j-k)^{-1} \leq K \sum_{j=1}^{t-1} j^{-1} \log j \leq K \log^2 t.$$

Similarly, by (14),

$$E\left\|\frac{\partial s_{1t}(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\varphi}}\right\| = E\left\|\sum_{j=1}^{t-1} \frac{\partial \phi_j(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}} u_{t-j}\right\| = O(1), \tag{S.138}$$

so that

$$E\left\|\frac{\partial s_{1t}(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\tau}}\right\| = O\left(\log^2 t\right).$$

Thus, noting (S.93),

$$E\left\|\frac{1}{T^{1/2}} \sum_{t=1}^{T} \sum_{j=t}^{\infty} \phi_j(\boldsymbol{\varphi}_0) u_{t-j} \frac{\partial s_{1t}(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\tau}}\right\| \leq \frac{K}{T^{1/2}} \sum_{t=1}^{T} t^{-1/2-\varsigma} \log^2 t \leq K T^{-1/2} = o(1)$$

because $\varsigma > 1/2$, which proves (S.135). Next, by (S.93) and (S.209) of Lemma S.16, it is straightforward to show that

$$\sum_{j=1}^{T} s_{1j}(\boldsymbol{\tau}_0) h_{j,T}(\gamma_0, \delta_0, \boldsymbol{\varphi}_0) = \sum_{j=1}^{T} (\varepsilon_j - \sum_{k=j}^{\infty} \phi_k(\boldsymbol{\varphi}_0) u_{j-k}) h_{j,T}(\gamma_0, \delta_0, \boldsymbol{\varphi}_0) = O_p(T^{\theta}), \tag{S.139}$$

so (S.136) holds on showing that

$$\frac{1}{T^{1/2}} \sum_{t=1}^{T} h_{t,T}(\gamma_0, \delta_0, \boldsymbol{\varphi}_0) \frac{\partial s_{1t}(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\tau}} = O_p\left(T^{\theta-1/2} \log T\right) \tag{S.140}$$

and setting $\theta < 1/4$. Noting (S.137), by simple calculations,

$$\frac{\partial s_{1t}(\boldsymbol{\tau}_0)}{\partial \delta} = -\sum_{j=1}^{t-1} \frac{1}{j} \varepsilon_{t-j} + \sum_{j=1}^{t-1} \frac{1}{j} \sum_{k=t-j}^{\infty} \phi_k(\boldsymbol{\varphi}_0) u_{t-j-k}. \tag{S.141}$$

The contribution of the first term on the right-hand side of (S.141) to the left-hand side of (S.140) is, by (S.209) of Lemma S.16,

$$\begin{aligned}-\frac{1}{T^{1/2}}\sum_{t=2}^{T}\sum_{j=1}^{t-1}j^{-1}\varepsilon_{t-j}h_{t,T}\left(\gamma_0,\delta_0,\boldsymbol{\varphi}_0\right) &= -\frac{1}{T^{1/2}}\sum_{t=1}^{T}\varepsilon_t\sum_{j=1}^{T-t}j^{-1}h_{t+j,T}\left(\gamma_0,\delta_0,\boldsymbol{\varphi}_0\right)\\ &= O_p\left(\frac{T^{\theta}}{T^{1/2}}\left(\sum_{t=1}^{T}\left(\sum_{j=1}^{T-t}\frac{1}{j}\left(t+j\right)^{-1/2-\theta}\right)^2\right)^{1/2}\right)\\ &= O_p(T^{\theta-1/2}\log T) \qquad \text{(S.142)}\end{aligned}$$

because

$$\sum_{j=1}^{T-t}j^{-1}\left(t+j\right)^{-1/2-\theta}\leq t^{-1/2-\theta}\sum_{j=1}^{T}j^{-1}\leq Kt^{-1/2-\theta}\log T.$$

Similarly, the contribution of the second term on the right-hand side of (S.141) to the left-hand side of (S.140) is

$$\frac{1}{T^{1/2}}\sum_{t=1}^{T}h_{t,T}\left(\gamma_0,\delta_0,\boldsymbol{\varphi}_0\right)\sum_{j=1}^{t-1}j^{-1}\sum_{k=t-j}^{\infty}\phi_k(\boldsymbol{\varphi}_0)u_{t-j-k},$$

which can be easily shown to be $O_p\left(T^{\theta-1/2}\log T\right)$ by (S.93), Lemma S.9 and (S.209) of Lemma S.16. Next, the contribution of $\partial s_{1t}\left(\boldsymbol{\tau}_0\right)/\partial\boldsymbol{\varphi}$ to the left-hand side of (S.140) is

$$\frac{1}{T^{1/2}}\sum_{t=1}^{T}u_t\sum_{j=1}^{T-t}\frac{\partial\phi_j(\boldsymbol{\varphi}_0)}{\partial\boldsymbol{\varphi}}h_{t+j,T}\left(\gamma_0,\delta_0,\boldsymbol{\varphi}_0\right),$$

which, by very similar arguments to (S.142), can easily be shown to be $O_p\left(T^{\theta-1/2}\right)$ by (14) and (S.209) of Lemma S.16, to conclude the proof of (S.140) and hence of the first equality in (S.132).

Next, because $\sum_{t=1}^{T}c_t\left(\gamma_0,\delta_0,\boldsymbol{\varphi}_0\right)\partial d_t\left(\boldsymbol{\vartheta}_0\right)/\partial\gamma=0$, the proof of the second equality in (S.132) follows by showing that

$$\frac{1}{T^{\gamma_0-\delta_0+1/2}}\sum_{t=1}^{T}\sum_{j=t}^{\infty}\phi_j(\boldsymbol{\varphi}_0)u_{t-j}\frac{\partial d_t\left(\boldsymbol{\vartheta}_0\right)}{\partial\gamma}=o_p\left(1\right),$$

which, noting the proof of (S.69) and (S.93), follows easily by previous arguments.

The proofs of the two equalities in (S.133) are almost identical, but the second is simpler, so we show only the first. By (S.134), the first equality in (S.133) holds if

$$\frac{1}{T^{1/2}}\sum_{t=1}^{T}s_{1t}\left(\boldsymbol{\tau}_0\right)\frac{\partial s_{2t}\left(\boldsymbol{\vartheta}_0\right)}{\partial\boldsymbol{\tau}}=o_p\left(1\right), \qquad \text{(S.143)}$$

$$\frac{1}{T^{1/2}}\sum_{t=1}^{T}h_{t,T}\left(\gamma_0,\delta_0,\boldsymbol{\varphi}_0\right)\frac{\partial s_{2t}\left(\boldsymbol{\vartheta}_0\right)}{\partial\boldsymbol{\tau}}\sum_{j=1}^{T}s_{1j}\left(\boldsymbol{\tau}_0\right)h_{j,T}\left(\gamma_0,\delta_0,\boldsymbol{\varphi}_0\right)=o_p\left(1\right). \qquad \text{(S.144)}$$

As defined before, $s_{2t}(\boldsymbol{\vartheta}) = h_{t,T}(\gamma,\delta,\boldsymbol{\varphi})\sum_{j=1}^{T} s_{1j}(\boldsymbol{\tau})h_{j,T}(\gamma,\delta,\boldsymbol{\varphi})$ so that

$$\begin{aligned}\frac{\partial s_{2t}(\boldsymbol{\vartheta})}{\partial\boldsymbol{\tau}} =& \frac{\partial h_{t,T}(\gamma,\delta,\boldsymbol{\varphi})}{\partial\boldsymbol{\tau}}\sum_{j=1}^{T} s_{1j}(\boldsymbol{\tau})\, h_{j,T}(\gamma,\delta,\boldsymbol{\varphi}) \\ &+ h_{t,T}(\gamma,\delta,\boldsymbol{\varphi})\sum_{j=1}^{T}\frac{\partial s_{1j}(\boldsymbol{\tau})}{\partial\boldsymbol{\tau}} h_{j,T}(\gamma,\delta,\boldsymbol{\varphi}) \\ &+ h_{t,T}(\gamma,\delta,\boldsymbol{\varphi})\sum_{j=1}^{T} s_{1j}(\boldsymbol{\tau})\,\frac{\partial h_{j,T}(\gamma,\delta,\boldsymbol{\varphi})}{\partial\boldsymbol{\tau}}. \end{aligned} \tag{S.145}$$

First, given that $\gamma_0 + 1/2 > \delta_0$, setting $\theta < \gamma_0 - \delta_0 + 1/2$, by a simple modification of the proof of (S.210) of Lemma S.16,

$$\left\|\frac{\partial h_{t,T}(\gamma_0,\delta_0,\boldsymbol{\varphi}_0)}{\partial\boldsymbol{\tau}}\right\| = O\left(t^{-1/2}\left(\frac{T}{t}\right)^{\theta}\log T\right). \tag{S.146}$$

Then noting (S.139), (S.140), and by application of Lemma S.16, it follows that

$$\frac{\partial s_{2t}(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\tau}} = O_p(t^{-1/2-\theta}T^{2\theta}\log T). \tag{S.147}$$

By (S.93) and (S.147), it follows that the left-hand side of (S.143) is $O_p(T^{2\theta-1/2}\log T) = o_p(1)$ by setting $\theta < 1/4$. Similarly, by (S.139), (S.147) and (S.209) of Lemma S.16, the left-hand side of (S.144) is $O_p(T^{4\theta-1/2}\log T) = o_p(1)$ by setting $\theta < 1/8$. This concludes the proof of the first equality in (S.133).

Finally, the proofs for the results in part (b) are heavily based on the arguments employed in the proofs of (S.132) and (S.133), and are therefore omitted. ■

**Lemma S.6** *Under the conditions of Theorem 2(i), for some fixed* $\varkappa > 0$, $T^{\varkappa}(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) \to_p 0$.

**Proof.** As in the proof of Theorem 1(i), noting (S.1), (S.2), (S.4), (S.62), the result holds on establishing that

$$\Pr\left(\inf_{\boldsymbol{\vartheta}\in\overline{M}_{\varepsilon}^{*}} S_T(\boldsymbol{\vartheta}) \le 0\right) \to 0 \text{ as } T\to\infty, \tag{S.148}$$

$$\Pr\left(\widehat{\boldsymbol{\vartheta}} \in \overline{N}_{\varepsilon}^{*}\cap M_{\varepsilon}^{*}, \inf_{\overline{N}_{\varepsilon}^{*}\cap M_{\varepsilon}^{*}} R_T(\widehat{\boldsymbol{\tau}},\gamma) - R_T(\widehat{\boldsymbol{\tau}},\gamma_0) \le 0\right) \to 0 \text{ as } T\to\infty, \tag{S.149}$$

where

$$\begin{aligned} M_{\varepsilon}^{*} &= \left\{\boldsymbol{\vartheta}\in\Xi : \|\boldsymbol{\tau}-\boldsymbol{\tau}_0\| < \varepsilon T^{-\varkappa}\right\}, \quad \overline{M}_{\varepsilon}^{*} = \{\boldsymbol{\vartheta}\in\Xi : \varepsilon T^{-\varkappa} \le \|\boldsymbol{\tau}-\boldsymbol{\tau}_0\| < \varepsilon\}, \\ N_{\varepsilon}^{*} &= \left\{\boldsymbol{\vartheta}\in\Xi : |\gamma-\gamma_0| < \varepsilon T^{-\varkappa}\right\}, \quad \overline{N}_{\varepsilon}^{*} = \{\boldsymbol{\vartheta}\in\Xi : \varepsilon T^{-\varkappa} \le |\gamma-\gamma_0| < \varepsilon\}. \end{aligned}$$

We first prove (S.148), which, defining $\mathcal{J}_i = \{\boldsymbol{\vartheta}\in\overline{M}_{\varepsilon}^{*} : \delta\in\mathcal{I}_i\}$, $i = 4, 5$, holds if

$$\Pr\left(\inf_{\mathcal{J}_i} S_T(\boldsymbol{\vartheta}) \le 0\right) \to 0 \text{ as } T\to\infty \tag{S.150}$$

for $i = 4, 5$. Note here that $\boldsymbol{\vartheta} \in \overline{M}_{\varepsilon}^{*}$ implies $\|\boldsymbol{\tau} - \boldsymbol{\tau}_0\| < \varepsilon$, so necessarily $\delta \in \mathcal{I}_4 \cup \mathcal{I}_5$ and there is no need to consider the intervals $\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3$. Clearly, (S.150) for $i = 5$ would hold if

$$\Pr\left(\inf_{\mathcal{J}_5} T^{2\varkappa} S_T(\boldsymbol{\vartheta}) \leq 0\right) \to 0 \text{ as } T \to \infty. \tag{S.151}$$

Proceeding as in the proof of (S.5)–(S.7) for $i = 5$, (S.151) holds if

$$\inf_{\mathcal{J}_5} T^{2\varkappa} U(\boldsymbol{\tau}) > \epsilon, \tag{S.152}$$

$$\frac{1}{T^{1-2\varkappa}} \sum_{t=1}^{T} \left( (\phi(L; \boldsymbol{\varphi}_0) \{u_t \mathbb{I}(t \geq 1)\})^2 - \sigma_0^2 \right) = o_p(1), \tag{S.153}$$

$$\sup_{\mathcal{J}_5} \frac{1}{T^{1-2\varkappa}} \sum_{t=1}^{T} \left( \varepsilon_t^2(\boldsymbol{\tau}) - E\left( \left( \phi(L; \boldsymbol{\varphi}) \Delta^{\delta - \delta_0} u_t \right)^2 \right) \right) = o_p(1), \tag{S.154}$$

$$\sup_{\mathcal{J}_5} \frac{1}{T^{1-2\varkappa}} \left| \sum_{t=1}^{T} d_t(\boldsymbol{\vartheta}) s_t(\boldsymbol{\vartheta}) \right| = o_p(1), \tag{S.155}$$

$$\sup_{\mathcal{J}_5} \frac{1}{T^{1-2\varkappa}} \left( \sum_{t=1}^{T} \varepsilon_t(\boldsymbol{\tau}) h_{t,T}(\gamma, \delta, \boldsymbol{\varphi}) \right)^2 = o_p(1). \tag{S.156}$$

First, we justify (S.152). Clearly

$$U(\boldsymbol{\tau}) = \sigma_0^2 \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\left|\phi\left(e^{i\lambda}; \boldsymbol{\varphi}\right)\right|^2}{\left|\phi\left(e^{i\lambda}; \boldsymbol{\varphi}_0\right)\right|^2} \left|1 - e^{i\lambda}\right|^{2(\delta - \delta_0)} d\lambda - 1 \right),$$

and we show that $U(\boldsymbol{\tau})$ is a strictly convex function at $\boldsymbol{\tau}_0$ with a strict local minimum at $\boldsymbol{\tau} = \boldsymbol{\tau}_0$. Noting that

$$\int_{-\pi}^{\pi} \frac{e^{iq\lambda}}{\phi\left(e^{i\lambda}; \boldsymbol{\varphi}_0\right)} d\lambda = 0 \text{ for any } q = \pm 1, \pm 2, \ldots, \tag{S.157}$$

and $\int_{-\pi}^{\pi} \log(2 - 2\cos\lambda) \, d\lambda = 0$, it is straightforward to show that $\partial U(\boldsymbol{\tau}_0) / \partial \boldsymbol{\tau} = 0$. Similarly, using again (S.157),

$$\begin{aligned}
\frac{\partial^2 U(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}'} &= \begin{pmatrix} \int_{-\pi}^{\pi} \log^2(2 - 2\cos\lambda) \, d\lambda & 2\int_{-\pi}^{\pi} \frac{\partial \phi(e^{i\lambda}; \boldsymbol{\varphi}_0)/\partial \boldsymbol{\varphi}'}{\phi(e^{i\lambda}; \boldsymbol{\varphi}_0)} \log(2 - 2\cos\lambda) \, d\lambda \\ 2\int_{-\pi}^{\pi} \frac{\partial \phi(e^{i\lambda}; \boldsymbol{\varphi}_0)/\partial \boldsymbol{\varphi}}{\phi(e^{i\lambda}; \boldsymbol{\varphi}_0)} \log(2 - 2\cos\lambda) \, d\lambda & 2\int_{-\pi}^{\pi} \frac{\partial \phi(e^{i\lambda}; \boldsymbol{\varphi}_0)/\partial \boldsymbol{\varphi} \partial \phi(e^{i\lambda}; \boldsymbol{\varphi}_0)/\partial \boldsymbol{\varphi}'}{\left|\phi(e^{i\lambda}; \boldsymbol{\varphi}_0)\right|^2} d\lambda \end{pmatrix} \\
&= \begin{pmatrix} 2\pi^3/3 & -4\pi \sum_{j=1}^{\infty} \boldsymbol{b}_j'(\boldsymbol{\varphi}_0)/j \\ -4\pi \sum_{j=1}^{\infty} \boldsymbol{b}_j(\boldsymbol{\varphi}_0)/j & 4\pi \sum_{j=1}^{\infty} \boldsymbol{b}_j(\boldsymbol{\varphi}_0) \boldsymbol{b}_j'(\boldsymbol{\varphi}_0) \end{pmatrix},
\end{aligned}$$

which by A4(iii) is positive definite, to complete the proof of strict convexity of $U(\boldsymbol{\tau})$ at $\boldsymbol{\tau}_0$. Thus, by continuity there exists a point $\boldsymbol{\tau}^*$ such that $\|\boldsymbol{\tau}_0 - \boldsymbol{\tau}^*\| = \varepsilon T^{-\varkappa}$ and $\inf_{\mathcal{J}_5} U(\boldsymbol{\tau}) = U(\boldsymbol{\tau}^*)$. Then, noting that $U(\boldsymbol{\tau}_0) = 0$ and $\partial U(\boldsymbol{\tau}_0) / \partial \boldsymbol{\tau} = 0$, by Taylor's expansion,

$$U(\boldsymbol{\tau}^*) \geq \frac{1}{2} (\boldsymbol{\tau}^* - \boldsymbol{\tau}_0)' \frac{\partial^2 U(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}'} (\boldsymbol{\tau}^* - \boldsymbol{\tau}_0) - |w_T|, \tag{S.158}$$

where it can be shown that $w_T = O\left(T^{-3\varkappa}\right)$. Here, the main issue is to justify that the third derivative of $U(\boldsymbol{\tau})$ evaluated at an arbitrarily small neighborhood of $\boldsymbol{\tau}_0$ is bounded, but this follows straightforwardly from A4(ii). Additionally,

$$(\boldsymbol{\tau}^* - \boldsymbol{\tau}_0)' \frac{\partial^2 U(\boldsymbol{\tau}_0)}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}'} (\boldsymbol{\tau}^* - \boldsymbol{\tau}_0) \geq \underline{\lambda} \left\|\boldsymbol{\tau}^* - \boldsymbol{\tau}_0\right\|^2,$$

where $\underline{\lambda}$ denotes the minimum eigenvalue of the matrix $\partial^2 U(\boldsymbol{\tau}_0)/\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}'$, so by (S.158) and noting that $\underline{\lambda}$ is strictly positive, for a sufficiently small $\epsilon > 0$,

$$U(\boldsymbol{\tau}^*) > \frac{\epsilon}{\varepsilon^2} \left\|\boldsymbol{\tau}^* - \boldsymbol{\tau}_0\right\|^2,$$

which justifies (S.152). The proofs of (S.153)–(S.156) are omitted as, for small enough $\varkappa$, they follow by almost identical arguments to those of (S.10)–(S.13).

Next, the proof of (S.150) for $i = 4$ is omitted because it is basically identical to those of (S.5)–(S.7) for $i = 4$. The only difference is that now $\varepsilon T^{-\varkappa} \leq \|\boldsymbol{\tau} - \boldsymbol{\tau}_0\| \leq \varepsilon$ instead of $\|\boldsymbol{\tau} - \boldsymbol{\tau}_0\| \geq \varepsilon$, but this does not make any difference. This completes the justification of (S.148).

Finally, we prove (S.149). For the same reason as in the proof of (S.62), we need to prove that

$$\lim_{T\to\infty} \inf_{\boldsymbol{\vartheta}\in\overline{N}_\varepsilon^*\cap M_\varepsilon^*} \frac{T^{2\varkappa}}{T^{2(\gamma_0-\delta)+1}} \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta}) > \epsilon, \tag{S.159}$$

$$\sup_{\boldsymbol{\vartheta}\in\overline{N}_\varepsilon^*\cap M_\varepsilon^*} \frac{T^{2\varkappa}}{T^{2(\gamma_0-\delta)+1}} \left|\sum_{t=1}^{T} d_t(\boldsymbol{\vartheta})\, s_{1t}(\boldsymbol{\tau})\right| = o_p(1), \tag{S.160}$$

$$\sup_{\boldsymbol{\vartheta}\in\overline{N}_\varepsilon^*\cap M_\varepsilon^*} \frac{T^{2\varkappa}}{T^{2(\gamma_0-\delta)+1}} \left(\sum_{j=1}^{T} s_{1j}(\boldsymbol{\tau})\, h_{j,T}(\gamma,\delta,\boldsymbol{\varphi})\right)^2 = o_p(1). \tag{S.161}$$

As in (S.20), the proof of (S.159) follows by Lemma S.3, whereas the proofs of (S.160) and (S.161) hold as in (S.64) and (S.65) for $\varkappa > 0$ sufficiently small. ■

**Lemma S.7** *Let* $\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0 = O_p\left(T^{-\varkappa}\right)$, $\widehat{\gamma} - \gamma_0 = O_p\left(T^{-\varkappa}\right)$ *for* $\varkappa > 0$. *Then, under Assumptions A1–A4,*

$$c_t(\widehat{\gamma},\widehat{\delta},\widehat{\boldsymbol{\varphi}}) = c_t(\gamma_0,\delta_0,\boldsymbol{\varphi}_0) + O_p\left(T^{-\varkappa} t^{\max\{\gamma_0-\delta_0,-1-\varsigma\}} \log^2 t\right), \tag{S.162}$$

$$c_t^{(1)}(\widehat{\gamma},\widehat{\delta},\widehat{\boldsymbol{\varphi}}) = c_t^{(1)}(\gamma_0,\delta_0,\boldsymbol{\varphi}_0) + O_p\left(T^{-\varkappa} t^{\max\{\gamma_0-\delta_0,-1-\varsigma\}} \log^3 t\right), \tag{S.163}$$

*and, uniformly in* $t = 1,\dots,T$,

$$\varepsilon_t(\widehat{\boldsymbol{\tau}}) = \sum_{j=0}^{t-1} a_j\left(\delta_0 - \widehat{\delta}, \widehat{\boldsymbol{\varphi}}\right) u_{t-j} = \sum_{j=0}^{t-1} \phi_j(\boldsymbol{\varphi}_0)\, u_{t-j} + O_p\left(T^{-\varkappa}\right). \tag{S.164}$$

**Proof.** First we show (S.162). Clearly

$$
\begin{aligned}
c_t(\widehat{\gamma},\widehat{\delta},\widehat{\boldsymbol{\varphi}})-c_t(\gamma_0,\delta_0,\boldsymbol{\varphi}_0) &= \sum_{j=1}^{t} b_j\left(\gamma_0,\delta_0\right)\left(\phi_{t-j}\left(\widehat{\boldsymbol{\varphi}}\right)-\phi_{t-j}\left(\boldsymbol{\varphi}_0\right)\right) \\
&\quad + \sum_{j=1}^{t}(b_j\left(\widehat{\gamma},\widehat{\delta}\right)-b_j\left(\gamma_0,\delta_0\right))\phi_{t-j}\left(\boldsymbol{\varphi}_0\right) \\
&\quad + \sum_{j=1}^{t}(b_j\left(\widehat{\gamma},\widehat{\delta}\right)-b_j\left(\gamma_0,\delta_0\right))\left(\phi_{t-j}\left(\widehat{\boldsymbol{\varphi}}\right)-\phi_{t-j}\left(\boldsymbol{\varphi}_0\right)\right). \qquad \text{(S.165)}
\end{aligned}
$$

Fix $\epsilon < 1/2$. Then

$$
\phi_j\left(\widehat{\boldsymbol{\varphi}}\right)-\phi_j\left(\boldsymbol{\varphi}_0\right)=\left(\phi_j\left(\widehat{\boldsymbol{\varphi}}\right)-\phi_j\left(\boldsymbol{\varphi}_0\right)\right)\left(\mathbb{I}\left(\|\widehat{\boldsymbol{\varphi}}-\boldsymbol{\varphi}_0\|<\epsilon\right)+\mathbb{I}\left(\|\widehat{\boldsymbol{\varphi}}-\boldsymbol{\varphi}_0\|\geq\epsilon\right)\right), \qquad \text{(S.166)}
$$

so by the mean value theorem the left-hand side of (S.166) is bounded by

$$
\sup_{\|\boldsymbol{\varphi}-\boldsymbol{\varphi}_0\|<\epsilon}\left\|\frac{\partial\phi_j\left(\boldsymbol{\varphi}\right)}{\partial\boldsymbol{\varphi}}\right\|\|\widehat{\boldsymbol{\varphi}}-\boldsymbol{\varphi}_0\|+K\sup_{\boldsymbol{\varphi}\in\Psi}\left|\phi_j\left(\boldsymbol{\varphi}\right)\right|\frac{\|\widehat{\boldsymbol{\varphi}}-\boldsymbol{\varphi}_0\|^N}{\epsilon^N}, \qquad \text{(S.167)}
$$

for any arbitrarily large fixed number $N$. Then by (6) and the $T^{-\varkappa}$-consistency of $\widehat{\boldsymbol{\tau}}$, the second term in (S.167) is of smaller order, whereas by (14), the first one is $O_p(T^{-\varkappa}j^{-1-\varsigma})$. This implies that the first term on the right-hand side of (S.165) is $O_p(T^{-\varkappa}t^{\max\{\gamma_0-\delta_0,-1-\varsigma\}}\log t)$ by (S.184) of Lemma S.13, using also Lemmas S.9 and S.11.

Next, we show that

$$
b_j(\widehat{\gamma},\widehat{\delta})=b_j\left(\gamma_0,\delta_0\right)+O_p(T^{-\varkappa}j^{\gamma_0-\delta_0}\log j). \qquad \text{(S.168)}
$$

Using a $N$'th-order Taylor expansion,

$$
b_j(\widehat{\gamma},\widehat{\delta})=b_j\left(\gamma_0,\delta_0\right)+(\widehat{\gamma}-\gamma_0,\widehat{\delta}-\delta_0)(b_j^{(1)}\left(\gamma_0,\delta_0\right),b_j^{(2)}\left(\gamma_0,\delta_0\right))'+p_{j,N}(\widehat{\gamma},\widehat{\delta},\gamma_0,\delta_0),
$$

where $b_j^{(i)}\left(\cdot,\cdot\right)$ denotes derivative of $b_j\left(\cdot,\cdot\right)$ with respect to the $i$'th argument and $p_{j,N}(\widehat{\gamma},\widehat{\delta},\gamma_0,\delta_0)$ collects derivatives of $b_j\left(\cdot,\cdot\right)$ evaluated at $\left(\gamma_0,\delta_0\right)$ (whose order of magnitude is given by (S.186)), powers of $(\widehat{\gamma}-\gamma_0)$, $(\widehat{\delta}-\delta_0)$, and a last term which involves, $(\widehat{\gamma}-\gamma_0)^N$, $(\widehat{\delta}-\delta_0)^N$ and $N$'th derivatives of $b_j\left(\cdot,\cdot\right)$ evaluated at intermediate points. This last term is bounded by $KT^{-N\varkappa}\log^N T\sup_{\gamma\in[\Box_1,\Box_2],\delta\in[\nabla_1,\nabla_2]}\sum_{k=1}^{j-1}j^{-\delta-1}\left(j-k\right)^{\gamma}$, which can be easily shown to be $o_p(T^{-\varkappa}j^{\gamma_0-\delta_0}\log j)$ for $N$ large enough. Then, (S.168) follows by (S.186), and using also (6), Lemma S.9 and (S.184), the second term on the right-hand side of (S.165) is

$$
O_p(T^{-\varkappa}t^{\max\{\gamma_0-\delta_0,-1-\varsigma\}}\log^2 t).
$$

Finally, combining the arguments for the first two terms, the third term on the right-hand side of (S.165) is of smaller order, to conclude the proof of (S.162).

The proof of (S.163) is omitted because it is almost identical to that of (S.162) with the only difference that the coefficients $b_j^{(1)}(\cdot,\cdot)$ instead of $b_j(\cdot,\cdot)$ lead to an extra $\log t$ factor, see Lemma S.13.

Finally, we show (S.164). Clearly

$$a_j(\delta_0-\widehat{\delta},\widehat{\boldsymbol{\varphi}}) = \phi_j(\boldsymbol{\varphi}_0)+\sum_{k=0}^{j}(\phi_k(\widehat{\boldsymbol{\varphi}})-\phi_k(\boldsymbol{\varphi}_0))\pi_{j-k}(0)+\sum_{k=0}^{j}\phi_k(\boldsymbol{\varphi}_0)(\pi_{j-k}(\delta_0-\widehat{\delta})-\pi_{j-k}(0))$$
$$+\sum_{k=0}^{j}(\phi_k(\widehat{\boldsymbol{\varphi}})-\phi_k(\boldsymbol{\varphi}_0))(\pi_{j-k}(\delta_0-\widehat{\delta})-\pi_{j-k}(0)),$$

so that

$$\begin{aligned}\varepsilon_t(\widehat{\boldsymbol{\tau}}) =& \sum_{j=0}^{t-1}\phi_j(\boldsymbol{\varphi}_0)u_{t-j}+\sum_{j=0}^{t-1}(\phi_j(\widehat{\boldsymbol{\varphi}})-\phi_j(\boldsymbol{\varphi}_0))u_{t-j}\\ &+\sum_{j=0}^{t-1}\sum_{k=0}^{j}\phi_k(\boldsymbol{\varphi}_0)(\pi_{j-k}(\delta_0-\widehat{\delta})-\pi_{j-k}(0))u_{t-j}\\ &+\sum_{j=0}^{t-1}\sum_{k=0}^{j}(\phi_k(\widehat{\boldsymbol{\varphi}})-\phi_k(\boldsymbol{\varphi}_0))(\pi_{j-k}(\delta_0-\widehat{\delta})-\pi_{j-k}(0))u_{t-j}. \end{aligned} \tag{S.169}$$

Using the mean value theorem as in (S.166) and (S.167) and summation by parts, it can be shown that the second term on the right-hand side of (S.169) is $O_p(T^{-\varkappa})$. Similarly, by Lemma C.5 of Robinson and Hualde (2003) and (6), the third and fourth terms on the right-hand side of (S.169) are also $O_p(T^{-\varkappa})$, to conclude the proof of (S.164). ■

**Lemma S.8** *Under the conditions of Theorem 2(ii), $T^{1/2}\partial S_T(\widehat{\boldsymbol{\vartheta}})/\partial\boldsymbol{\tau}=o_p(1)$.*

**Proof.** First, for any $\epsilon>0$, clearly

$$\begin{aligned}\Pr\left(\left\|T^{1/2}\partial S_T(\widehat{\boldsymbol{\vartheta}})/\partial\boldsymbol{\tau}\right\|\geq\epsilon\right) =& \Pr\left(\left\|T^{1/2}\partial S_T(\widehat{\boldsymbol{\vartheta}})/\partial\boldsymbol{\tau}\right\|\geq\epsilon,\|\widehat{\boldsymbol{\tau}}-\boldsymbol{\tau}_0\|<\varepsilon\right)\\ &+\Pr\left(\left\|T^{1/2}\partial S_T(\widehat{\boldsymbol{\vartheta}})/\partial\boldsymbol{\tau}\right\|\geq\epsilon,\|\widehat{\boldsymbol{\tau}}-\boldsymbol{\tau}_0\|\geq\varepsilon\right)\\ \leq& \Pr\left(\left\|T^{1/2}\partial S_T(\widehat{\boldsymbol{\vartheta}})/\partial\boldsymbol{\tau}\right\|\geq\epsilon,\|\widehat{\boldsymbol{\tau}}-\boldsymbol{\tau}_0\|<\varepsilon\right)\\ &+\Pr\left(\|\widehat{\boldsymbol{\tau}}-\boldsymbol{\tau}_0\|\geq\varepsilon\right),\end{aligned}$$

so, in view of Theorem 1(ii) and (S.95), the result holds on showing

$$\sup_{\boldsymbol{\vartheta}\in M_\varepsilon}\frac{1}{T^{1/2}}\sum_{t=1}^{T}(d_t(\boldsymbol{\vartheta})-s_{2t}(\boldsymbol{\vartheta}))\left(\frac{\partial d_t(\boldsymbol{\vartheta})}{\partial\boldsymbol{\tau}}-\frac{\partial s_{2t}(\boldsymbol{\vartheta})}{\partial\boldsymbol{\tau}}\right)=o_p(1), \tag{S.170}$$

$$\sup_{\boldsymbol{\vartheta}\in M_\varepsilon}\frac{1}{T^{1/2}}\sum_{t=1}^{T}s_{1t}(\boldsymbol{\tau})\left(\frac{\partial d_t(\boldsymbol{\vartheta})}{\partial\boldsymbol{\tau}}-\frac{\partial s_{2t}(\boldsymbol{\vartheta})}{\partial\boldsymbol{\tau}}\right)=o_p(1), \tag{S.171}$$

$$\sup_{\boldsymbol{\vartheta}\in M_\varepsilon}\frac{1}{T^{1/2}}\sum_{t=1}^{T}\frac{\partial s_{1t}(\boldsymbol{\tau})}{\partial\boldsymbol{\tau}}(d_t(\boldsymbol{\vartheta})-s_{2t}(\boldsymbol{\vartheta}))=o_p(1). \tag{S.172}$$

The proof of (S.170) follows upon showing that, for any $\theta > 0$ and $\varepsilon$ such that $0 < \varepsilon < \theta$,

$$\sup_{\boldsymbol{\vartheta}\in M_\varepsilon} |d_t(\boldsymbol{\vartheta})| = O(t^{\max\{\gamma_0-\delta_0+\varepsilon,-1-\varsigma\}} + T^{2\theta}t^{-1/2-\theta}), \tag{S.173}$$

$$\sup_{\boldsymbol{\vartheta}\in M_\varepsilon} |s_{2t}(\boldsymbol{\vartheta})| = O_p(T^{2\theta}t^{-1/2-\theta}), \tag{S.174}$$

$$\sup_{\boldsymbol{\vartheta}\in M_\varepsilon} \left\| \frac{\partial d_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\tau}} \right\| = O(t^{\max\{\gamma_0-\delta_0+\varepsilon,-1-\varsigma\}} \log t + T^{4\theta}t^{-1/2-\theta}), \tag{S.175}$$

$$\sup_{\boldsymbol{\vartheta}\in M_\varepsilon} \left\| \frac{\partial s_{2t}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\tau}} \right\| = O_p(T^{4\theta}t^{-1/2-\theta}), \tag{S.176}$$

and then letting $\theta$ be sufficiently small. We only show (S.175) and (S.176) because the proofs for (S.173) and (S.174) are very similar but simpler. First, by (S.192) of Lemma S.14,

$$\sup_{\boldsymbol{\vartheta}\in M_\varepsilon} |c_t(\gamma_0, \delta, \boldsymbol{\varphi})| = O(t^{\max\{\gamma_0-\delta_0+\varepsilon,-1-\varsigma\}}), \tag{S.177}$$

and by a simple modification of that result,

$$\sup_{\boldsymbol{\vartheta}\in M_\varepsilon} \left\| \frac{\partial c_t(\gamma_0, \delta, \boldsymbol{\varphi})}{\partial \boldsymbol{\tau}} \right\| = O(t^{\max\{\gamma_0-\delta_0+\varepsilon,-1-\varsigma\}} \log t). \tag{S.178}$$

Then (S.175) follows by direct application of (S.177) and (S.178) and (S.209) of Lemma S.16, noting that the bound in (S.209) also applies if the derivative is taken with respect to $\boldsymbol{\tau}$.

To prove (S.176) we apply (S.145), where the $\sup_{\boldsymbol{\vartheta}\in M_\varepsilon}$ of the absolute values of the first and third terms on the right-hand side are $O_p(T^{4\theta}t^{-1/2-\theta})$ by direct application of (S.209) of Lemma S.16 and (S.222) of Lemma S.18, noting that $\delta_0 - \delta \leq \varepsilon$ and that these bounds also apply if the derivatives are taken with respect to $\boldsymbol{\tau}$. For the second term on the right-hand side of (S.145), noting that $\sum_{l=1}^{t} s_{1l}(\boldsymbol{\tau}) = \sum_{j=0}^{t-1} a_j(\delta_0 - \delta, \boldsymbol{\varphi}) \sum_{l=1}^{t-j} u_l$, it is straightforward to show that, by (S.192) of Lemma S.14,

$$\sup_{\boldsymbol{\vartheta}\in M_\varepsilon} \left\| \sum_{j=1}^{t} \frac{\partial s_{1j}(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}} \right\| = O_p(t^{1/2+\varepsilon} \log t). \tag{S.179}$$

Therefore, by (S.209) of Lemma S.16 and using summation by parts as in the proof of Lemma S.18, the $\sup_{\boldsymbol{\vartheta}\in M_\varepsilon}$ of the absolute value of the second term on the right-hand side of (S.145) is $O_p(T^{2\theta}t^{-1/2-\theta})$, to justify (S.176) and hence (S.170).

Finally, (S.171) and (S.172) can be established by using summation by parts followed by direct application of the results in (S.173), (S.175), (S.179), and Lemma S.17, noting also that by previous arguments it can be easily shown that

$$\sup_{\boldsymbol{\vartheta}\in M_\varepsilon} |d_{t+1}(\boldsymbol{\vartheta}) - d_t(\boldsymbol{\vartheta})| = O(t^{\max\{\gamma_0-\delta_0+\varepsilon-1,-1-\varsigma\}} + T^{2\theta}t^{-3/2-\theta}),$$

$$\sup_{\boldsymbol{\vartheta}\in M_\varepsilon} \left\| \frac{\partial d_{t+1}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\tau}} - \frac{\partial d_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\tau}} \right\| = O(t^{\max\{\gamma_0-\delta_0+\varepsilon-1,-1-\varsigma\}} \log t + T^{4\theta}t^{-3/2-\theta}),$$

$$\sup_{\boldsymbol{\vartheta}\in M_\varepsilon} |s_{2t+1}(\boldsymbol{\vartheta}) - s_{2t}(\boldsymbol{\vartheta})| = O_p(T^{2\theta}t^{-3/2-\theta}),$$

$$\sup_{\boldsymbol{\vartheta}\in M_\varepsilon} \left\| \frac{\partial s_{2t+1}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\tau}} - \frac{\partial s_{2t}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\tau}} \right\| = O_p(T^{4\theta}t^{-3/2-\theta}).$$

■

## S.4 Technical lemmas

**Lemma S.9** *Uniformly for* $\max\{|\alpha|, |\beta|\} \leq a_0$, $\sum_{j=1}^{t-1} j^{\alpha-1}(t-j)^{\beta-1} \leq K(\log t)t^{\max\{\alpha+\beta-1,\alpha-1,\beta-1\}}$.

**Proof.** The proof of Lemma S.9 is given in Lemma B.4 of Johansen and Nielsen (2010). ■

**Lemma S.10** *For any* $d > -1$ *and any fixed* $a \geq 0$, *as* $T \to \infty$,

$$\frac{1}{T^{d+1}} \sum_{t=1}^{T} t^d \to \frac{1}{d+1}, \quad \frac{1}{T^{d+1}} \sum_{t=1}^{T} \log\left(\frac{t+a}{T}\right) t^d \to -\frac{1}{(d+1)^2},$$

$$\frac{1}{T^{d+1}} \sum_{t=1}^{T} \log^2\left(\frac{t+a}{T}\right) t^d \to \frac{2}{(d+1)^3}.$$

**Proof.** The proof of the first result is straightforward by approximating the sum by an integral. Next, by the mean value theorem, it is simple to show that

$$\frac{1}{T^{d+1}} \sum_{t=1}^{T} \log\left(\frac{t+a}{T}\right) t^d = \frac{1}{T^{d+1}} \sum_{t=1}^{T} \log\left(\frac{t}{T}\right) t^d + o(1).$$

Approximating the sum by an integral we find

$$\frac{1}{T} \sum_{t=1}^{T} \log\left(\frac{t}{T}\right)\left(\frac{t}{T}\right)^d \sim \int_0^1 \log(x)\, x^d dx = B(d+1,1)\left(\psi(d+1) - \psi(d+2)\right),$$

see p. 535 of Gradshteyn and Ryzhik (2000), where $B(x,y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$ is the Beta function and $\psi(\cdot)$ is the digamma function. Thus, the second result follows by the recurrence formulae for the gamma and digamma functions, see pp. 256 and 258 of Abramowitz and Stegun (1970). Similarly, $T^{-d-1}\sum_{t=1}^{T} \log^2((t+a)/T)t^d$ can be approximated by

$$\frac{1}{T} \sum_{t=1}^{T} \log^2\left(\frac{t}{T}\right)\left(\frac{t}{T}\right)^d \sim \int_0^1 \log^2(x)\, x^d dx$$
$$= B(d+1,1)\left(\left(\psi(d+1) - \psi(d+2)\right)^2 + \psi'(d+1) - \psi'(d+2)\right),$$

see p. 538 of Gradshteyn and Ryzhik (2000), where $\psi'(\cdot)$ is the trigamma function. Then the third result follows by the recurrence formulae for the gamma, digamma, and trigamma functions, see pp. 256, 258, and 260 of Abramowitz and Stegun (1970). ■

**Lemma S.11** *Let* $j \geq 1$ *and* $\mathbb{K}$ *denote any compact subset of* $\mathbb{R}\backslash\mathbb{N}_0$. *Then*

$$\pi_j(-v) = \frac{1}{\Gamma(-v)} j^{-v-1}(1 + \epsilon_j(v)), \tag{S.180}$$

*where* $\max_{v\in\mathbb{K}} |\epsilon_j(v)| \to 0$ *as* $j \to \infty$. *Thus, uniformly in* $j \geq 1, m \geq 0$,

*(i)* $\pi_j(-v) \geq Kj^{-v-1}$ *uniformly in* $v \in \mathbb{K}$,

*(ii)* $|\frac{\partial^m}{\partial u^m}\pi_j(u)| \leq K(1+\log j)^m j^{u-1}$ *uniformly in* $|u| \leq u_0$,

*(iii)* $|\frac{\partial^m}{\partial u^m} T^{-u}\pi_j(u)| \leq KT^{-u}(1+|\log(j/T)|)^m j^{u-1}$ *uniformly in* $|u| \leq u_0$.

**Proof.** The proof of Lemma S.11 is given in Lemma B.3 of Johansen and Nielsen (2010) and Lemma A.5 of Johansen and Nielsen (2012). ■

**Lemma S.12** *Let* $Z_t, t = 1, \ldots, T$, *be arbitrary. Then*

$$\sum_{t=1}^{T} Z_t^2 \geq \left(\frac{\pi^2}{4} T^{-2} + O(T^{-3})\right) \sum_{t=1}^{T} \left(\Delta_+^{-1} Z_t\right)^2, \tag{S.181}$$

*where the* $O(T^{-3})$ *term does not depend on any parameters.*

**Proof.** Following the proof of Lemma 2 of Johansen and Nielsen (2018), we let $Z = (Z_1, \ldots, Z_T)'$ and define the cumulation matrix

$$C = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 1 & \cdots & 1 & 1 \end{pmatrix}$$

such that $(\Delta_+^{-1} Z_1, \ldots, \Delta_+^{-1} Z_T)' = CZ$. Then, using $X = CZ$,

$$\frac{\sum_{t=1}^{T} Z_t^2}{\sum_{t=1}^{T} \left(\Delta_+^{-1} Z_t\right)^2} = \frac{Z'Z}{Z'C'CZ} = \frac{X'C'^{-1}C^{-1}X}{X'X} \geq \lambda_{\min}\left(C'^{-1}C^{-1}\right), \tag{S.182}$$

where the inequality follows from, e.g., Horn and Johnson (2013, p. 258). From Rutherford (1948), see also Tanaka (1996, eqn. (1.4)), we find the eigenvalues

$$\lambda_t\left(C'^{-1}C^{-1}\right) = 4\sin^2\left(\frac{\pi}{2}\frac{2t-1}{2T+1}\right), \quad t = 1, \ldots, T,$$

such that, in particular,

$$\lambda_{\min}\left(C'^{-1}C^{-1}\right) = 4\sin^2\left(\frac{\pi}{2}\frac{1}{2T+1}\right) = \frac{\pi^2}{4} T^{-2} + O(T^{-3}). \tag{S.183}$$

The bound (S.181) follows by combining (S.182) and (S.183). ■

**Lemma S.13** *For any real numbers* $\diamondsuit > -1$, $\overline{d}_1$, $\underline{d}_2$, $\overline{d}_2$, *and* $\diamondsuit \leq d_1 \leq \overline{d}_1$, $\underline{d}_2 \leq d_2 \leq \overline{d}_2$, $m \geq 1$, $0 \leq p \leq m$, *denoting by* $\psi(\cdot)$ *the digamma function,*

$$b_t(d_1, d_2) = \Gamma(d_1+1)\pi_t(d_1+1-d_2) + s_{1t}(d_1, d_2), \tag{S.184}$$

$$\frac{\partial b_t(d_1, d_2)}{\partial d_1} = \log(t) b_t(d_1, d_2) + (\psi(d_1+1) - \psi(d_1+1-d_2))\Gamma(d_1+1)\pi_t(d_1+1-d_2) + s_{2t}(d_1, d_2), \tag{S.185}$$

$$\frac{\partial^m b_t(d_1, d_2)}{\partial d_1^p \partial d_2^{m-p}} = O\left(t^{d_1-d_2}(\log t)^m\right), \tag{S.186}$$

*where* $|s_{1t}(d_1, d_2)| \leq K(t^{d_1-d_2-1} + t^{-d_2-1})$ *and* $|s_{2t}(d_1, d_2)| \leq K((t^{d_1-d_2-1} + t^{-d_2-1})\log t)$.

**Proof.** First we show (S.184). For integer $d_2$, it can be easily shown that (S.184) holds with $|r_{1t}(d_1, d_2)| \leq Kt^{d_1-d_2-1}$ using standard results for summations and the Taylor's theorem. Next, when $d_1$ is integer and $d_2$ is noninteger, by (3) and standard properties of the gamma function, it is simple to show that for integer $k \geq 1$,

$$j^k \pi_j(-d_2) = \frac{\Gamma(k-d_2)}{\Gamma(-d_2)} \pi_{j-1}(k-d_2) + \sum_{l=1}^{k-1} g_l(d_2) \pi_{j-1}(l-d_2), \tag{S.187}$$

where $g_l(\cdot)$ are polynomials and the second term on the right-hand side of (S.187) is 0 when $k=1$. Then

$$\sum_{j=0}^{t-1} j^k \pi_j(-d_2) = \frac{\Gamma(k-d_2)}{\Gamma(-d_2)} \pi_{t-2}(k+1-d_2) + O\left(t^{k-d_2-1}\right)$$

using Lemma S.11. Then, by (11) it is simple, but tedious, to show that (S.184) holds with $|r_{1t}(d_1, d_2)| \leq Kt^{d_1-d_2-1}$.

Next, we deal with the case where neither $d_1$ nor $d_2$ are integers. From Abramowitz and Stegun (1970, p. 257, eqn. 6.1.47), for a fixed integer $p \geq 1$ and any $k = -1, 0, 1, \ldots, p-1$,

$$\frac{\Gamma(t+d_1-k)}{\Gamma(t+1)} = t^{d_1-k-1} + \lambda_1(d_1-k) t^{d_1-k-2} + \cdots + \lambda_{p-k-1}(d_1-k) t^{d_1-p} + r_{p-k-1,t}(d_1-k),$$

for $t \neq k-d_1, k-d_1-1, k-d_1-2, \ldots,$ (which does not hold because $d_1$ is not integer), where for any $l$, $k$, $\sup_{\lozenge \leq d_1 \leq \bar{d}_1} |\lambda_l(d_1-k)| \leq K$ and $|r_{p-k-1,t}(d_1-k)| \leq Kt^{d_1-p-1}$. Then by recursive substitution it can be easily shown that

$$t^{d_1} = \frac{\Gamma(t+d_1+1)}{\Gamma(t+1)} + \nu_1(d_1) \frac{\Gamma(t+d_1)}{\Gamma(t+1)} + \cdots + \nu_p(d_1) \frac{\Gamma(t+d_1-p+1)}{\Gamma(t+1)} + q_{p,t}(d_1),$$

where the $\nu_l(d_1)$'s are complicated combinations of the $\lambda_l(\cdot)$'s, and $q_{p,t}(d_1)$ is a weighted sum of the $r_{p-k-1,t}(\cdot)$'s with coefficients which depend on a complicated manner on the $\lambda_l(\cdot)$'s. Given that $p$ is fixed it can be shown that for any $l$, $\sup_{\lozenge \leq d_1 \leq \bar{d}_1} |\nu_l(d_1)| \leq K$ and $|q_{p,t}(d_1)| \leq Kt^{d_1-p-1}$. Therefore, noting (3), it is immediate to show that

$$t^{d_1} = \Gamma(d_1+1) \pi_t(d_1+1) + \bar{\nu}_1(d_1) \pi_t(d_1) + \cdots + \bar{\nu}_p(d_1) \pi_t(d_1-p+1) + \bar{q}_{p,t}(d_1),$$

where $\bar{\nu}_k(d_1) = \Gamma(d_1+1)\Gamma(d_1-k+1)\nu_k(d_1)$, $\bar{q}_{p,t}(d_1) = \Gamma(d_1+1) q_{p,t}(d_1)$. Then, given that $\Delta^d \pi_t(c) = \pi_t(c-d)$,

$$\begin{aligned} b_t(d_1, d_2) =& \Gamma(d_1+1) \pi_t(d_1+1-d_2) + \bar{\nu}_1(d_1) \pi_t(d_1-d_2) + \cdots + \bar{\nu}_p(d_1) \pi_t(d_1-p+1-d_2) \\ &+ \sum_{j=0}^{t-1} \pi_j(-d_2) \bar{q}_{p,t-j}(d_1). \end{aligned} \tag{S.188}$$

Clearly, by Lemma S.11,

$$\left| \sum_{j=0}^{t-1} \pi_j(-d_2) \bar{q}_{p,t-j}(d_1) \right| \leq K \sum_{j=1}^{t-1} j^{-d_2-1} (t-j)^{d_1-p-1} \leq Kt^{-d_2-1},$$

for $p$ large enough, so (S.184) immediately follows by Lemma S.11.

Next we show (S.185). When $d_1$ and/or $d_2$ are integers the proof follows from relatively simple (but cumbersome) arguments very similar to those employed in the proof of (S.184). Thus we focus on the case where $d_1$ and $d_2$ are not integers. By (S.188),

$$\begin{aligned}\frac{\partial b_t(d_1,d_2)}{\partial d_1} =&\Gamma^{(1)}(d_1+1)\,\pi_t(d_1+1-d_2)+\Gamma(d_1+1)\,\pi_t^{(1)}(d_1+1-d_2)\\ &+\sum_{k=1}^{p}\left(\overline{\nu}_k^{(1)}(d_1)\,\pi_t(d_1-d_2-k+1)+\overline{\nu}_k(d_1)\,\pi_t^{(1)}(d_1-d_2-k+1)\right)\\ &+\sum_{j=0}^{t-1}\pi_j(-d_2)\,\overline{q}_{p,t-j}^{(1)}(d_1),\end{aligned} \tag{S.189}$$

where superscript (1) denotes first derivative. Noting that

$$\pi_j^{(1)}(d)=(\psi(d+j)-\psi(d))\,\pi_j(d) \tag{S.190}$$

and that for a fixed $a$,

$$\psi(t+a)=\log t+O\left(t^{-1}\right), \tag{S.191}$$

see Abramowitz and Stegun (1970, p. 259, eqn. 6.3.18), it can be easily shown that $|\overline{q}_{p,j}^{(1)}(d_1)|\leq Kj^{d_1-p-1}\log j$, so, for $p$ large enough, the last term on the right-hand side of (S.189) is $O\left(t^{-d_2-1}\right)$. Then by (S.184), (S.189), and noting that $\Gamma^{(1)}(d_1+1)=\Gamma(d_1+1)\,\psi(d_1+1)$,

$$\begin{aligned}&\frac{\partial b_t(d_1,d_2)}{\partial d_1}\\ &=\log(t)\,\Gamma(d_1+1)\,\pi_t(d_1+1-d_2)\\ &\quad+\Gamma(d_1+1)\,(\psi(t+d_1+1-d_2)-\psi(d_1+1-d_2)+\psi(d_1+1)-\log t)\,\pi_t(d_1+1-d_2)\\ &\quad+O\left(t^{d_1-d_2-1}\log t+t^{-d_2-1}\right),\end{aligned}$$

so (S.185) is justified by (S.184) and (S.191).

Finally, the proof of (S.186) follows by taking appropriate derivatives on $b_t(d_1,d_2)$, as in (S.189), and using standard bounds for the derivatives of the digamma function given by

$$\left|\frac{\partial^l\psi(z)}{\partial z^l}\right|\leq Kz^{-l},\quad l=1,2,3,\ldots$$

■

**Lemma S.14** *Under Assumptions A1, A3, uniformly in* $t=1,\ldots,T$ *and* $T\geq 1$*, for any* *real numbers* $\Diamond > -1$, $\overline{d}_1$, $\underline{d}_2$, $\overline{d}_2$, *and* $\Diamond \leq d_1 \leq \overline{d}_1$, $\underline{d}_2 \leq d_2 \leq \overline{d}_2$, $m \geq 0$, $0 \leq p \leq m$,

$$\sup_{d_1 \geq \Diamond, d_1 - d_2 \leq g, \boldsymbol{\varphi} \in \Psi} \left| \frac{\partial^m c_t(d_1, d_2, \boldsymbol{\varphi})}{\partial d_1^p \partial d_2^{m-p}} \right| = O(t^{\max\{g,-1-\varsigma\}} (\log t)^m), \tag{S.192}$$

$$\sup_{d_1 \geq \Diamond, d_1 - d_2 \geq g, \boldsymbol{\varphi} \in \Psi} \left| \frac{\partial^m}{\partial d_1^p \partial d_2^{m-p}} T^{-(d_1-d_2)} c_t(d_1, d_2, \boldsymbol{\varphi}) \right| = O(T^{-g} t^{\max\{g,-1-\varsigma\}} (1 + |\log(t/T)|)^m). \tag{S.193}$$

$$\sup_{d \leq g, \boldsymbol{\varphi} \in \Psi} \left| \frac{\partial^m a_t(d, \boldsymbol{\varphi})}{\partial d^m} \right| = O(t^{\max\{g-1,-1-\varsigma\}} (\log t)^m), \tag{S.194}$$

$$\sup_{d \geq g, \boldsymbol{\varphi} \in \Psi} \left| \frac{\partial^m}{\partial d^m} T^{-d} a_t(d, \boldsymbol{\varphi}) \right| = O(T^{-g} t^{\max\{g-1,-1-\varsigma\}} (1 + |\log(t/T)|)^m).$$

**Proof.** Using Lemma S.13, the proof of Lemma S.14 is almost identical to that of Lemma 1 of Hualde and Robinson (2011) and is therefore omitted. ■

**Lemma S.15** *Under Assumptions A1 and A3, for any real numbers* $\Diamond > -1$, $\overline{d}_1$, $\underline{d}_2$, $\overline{d}_2$, *and* $\Diamond \leq d_1 \leq \overline{d}_1$, $\underline{d}_2 \leq d_2 \leq \overline{d}_2$,

$$\frac{1}{T^{2(d_1-d_2)+1}} \sum_{t=1}^{T} c_t^2(d_1, d_2, \boldsymbol{\varphi}) \geq \frac{\phi^2(1; \boldsymbol{\varphi})}{T^{2(d_1-d_2)+1}} \sum_{t=1}^{T} b_t^2(d_1, d_2) - |r_{1T}(d_1, d_2, \boldsymbol{\varphi})|, \tag{S.195}$$

$$\frac{1}{T^{2(d_1-d_2)+1}} \sum_{t=1}^{T} c_t^2(d_1, d_2, \boldsymbol{\varphi}) \leq \frac{\phi^2(1; \boldsymbol{\varphi})}{T^{2(d_1-d_2)+1}} \sum_{t=1}^{T} b_t^2(d_1, d_2) + |r_{2T}(d_1, d_2, \boldsymbol{\varphi})|, \tag{S.196}$$

*where, for any* $\eta > 0$, $\sup_{d_1 \geq \Diamond, d_1 - d_2 \geq -1/2+\eta, \boldsymbol{\varphi} \in \Psi} |r_{iT}(d_1, d_2, \boldsymbol{\varphi})| = o(1)$, $i = 1, 2$. *Furthermore, for any* $\alpha$ *such that* $0 < \alpha < \min\{(\varsigma - 1/2)/3, (1 + \Diamond)/2\}$,

$$\inf_{d_1, d_2, \boldsymbol{\varphi} \in \Psi} \sum_{t=1}^{T} c_t^2(d_1, d_2, \boldsymbol{\varphi}) \geq 1, \tag{S.197}$$

$$\inf_{d_1 \geq \Diamond, -1/2-\alpha \leq d_1 - d_2 \leq -1/2+\alpha, \boldsymbol{\varphi} \in \Psi} \frac{1}{T^{2(d_1-d_2)+1}} \sum_{t=1}^{T} c_t^2(d_1, d_2, \boldsymbol{\varphi}) \geq \frac{\epsilon}{\alpha} + o(1), \tag{S.198}$$

$$\inf_{d_1 \geq \Diamond, d_1 - d_2 \geq -1/2-\alpha, \boldsymbol{\varphi} \in \Psi} \frac{1}{T^{2(d_1-d_2)+1}} \sum_{t=1}^{T} c_t^2(d_1, d_2, \boldsymbol{\varphi}) \geq \epsilon, \tag{S.199}$$

*for some* $\epsilon > 0$, *which does not depend on* $\alpha$ *or* $T$.

**Proof.** First we show (S.195). By summation by parts

$$c_t(d_1, d_2, \boldsymbol{\varphi}) = \sum_{j=1}^{t} b_j(d_1, d_2) \phi_{t-j}(\boldsymbol{\varphi}) = b_t(d_1, d_2) \sum_{j=0}^{t-1} \phi_j(\boldsymbol{\varphi}) - \sum_{j=1}^{t-1} (b_{j+1}(d_1, d_2) - b_j(d_1, d_2)) \sum_{l=1}^{j} \phi_{t-l}(\boldsymbol{\varphi}).$$

Using (S.28),

$$b_{j+1}(d_1,d_2) - b_j(d_1,d_2) = \sum_{k=0}^{j} \pi_k(-d_2-1)(j+1-k)^{d_1} = b_{j+1}(d_1,d_2+1)$$

and

$$\begin{aligned} c_t(d_1,d_2,\boldsymbol{\varphi}) =& \phi(1;\boldsymbol{\varphi}) b_t(d_1,d_2) - b_t(d_1,d_2)\sum_{k=t}^{\infty}\phi_k(\boldsymbol{\varphi}) \\ & - \sum_{j=1}^{t-1} b_{j+1}(d_1,d_2+1)\sum_{l=1}^{j}\phi_{t-l}(\boldsymbol{\varphi}). \end{aligned} \tag{S.200}$$

Then

$$\begin{aligned} \sum_{t=1}^{T} c_t^2(d_1,d_2,\boldsymbol{\varphi}) \geq& \phi^2(1;\boldsymbol{\varphi})\sum_{t=1}^{T} b_t^2(d_1,d_2) - 2\phi(1;\boldsymbol{\varphi})\sum_{t=1}^{T} b_t^2(d_1,d_2)\sum_{j=t}^{\infty}\phi_j(\boldsymbol{\varphi}) \\ & - 2\phi(1;\boldsymbol{\varphi})\sum_{t=1}^{T} b_t(d_1,d_2)\sum_{j=1}^{t-1} b_{j+1}(d_1,d_2+1)\sum_{l=1}^{j}\phi_{t-l}(\boldsymbol{\varphi}) \\ & + 2\sum_{t=1}^{T} b_t(d_1,d_2)\sum_{j=t}^{\infty}\phi_j(\boldsymbol{\varphi})\sum_{k=1}^{t-1} b_{k+1}(d_1,d_2+1)\sum_{l=1}^{k}\phi_{t-l}(\boldsymbol{\varphi}). \end{aligned} \tag{S.201}$$

Noting (6), the fourth term on the right-hand side of (S.201) is of smaller order than the third term. Then the proof of (S.195) follows on showing

$$\sup_{d_1\geq\lozenge, d_1-d_2\geq -1/2+\eta, \boldsymbol{\varphi}\in\Psi} \frac{1}{T^{2(d_1-d_2)+1}}\left|\sum_{t=1}^{T} b_t^2(d_1,d_2)\sum_{j=t}^{\infty}\phi_j(\boldsymbol{\varphi})\right| = o(1), \tag{S.202}$$

$$\sup_{d_1\geq\lozenge, d_1-d_2\geq -1/2+\eta, \boldsymbol{\varphi}\in\Psi} \frac{1}{T^{2(d_1-d_2)+1}}\left|\sum_{t=1}^{T} b_t(d_1,d_2)\sum_{j=1}^{t-1} b_{j+1}(d_1,d_2+1)\sum_{l=1}^{j}\phi_{t-l}(\boldsymbol{\varphi})\right| = o(1). \tag{S.203}$$

First, by (6), Lemma S.9 and (S.184) of Lemma S.13, the left-hand side of (S.202) is bounded by

$$\begin{aligned} K\sup_{d_1-d_2\geq -1/2+\eta}\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{2(d_1-d_2)} t^{-\varsigma} \leq& K\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{-1+2\eta} t^{-\varsigma} \\ \leq& K\frac{1}{T^{2\eta}}\sum_{t=1}^{T} t^{-1+2\eta-\varsigma} = o(1), \end{aligned}$$

so (S.202) holds. Similarly, the left-hand side of (S.203) is bounded by

$$
\begin{aligned}
K \sup_{d_1-d_2\geq -1/2+\eta} \frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{d_1-d_2}\sum_{j=1}^{t-1}\left(\frac{j}{T}\right)^{d_1-d_2} j^{-1}(t-j)^{-\varsigma} \\
&\leq K\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{-1/2+\eta}\sum_{j=1}^{t-1}\left(\frac{j}{T}\right)^{-1/2+\eta} j^{-1}(t-j)^{-\varsigma} \\
&\leq K\frac{1}{T^{2\eta}}\sum_{t=1}^{T} t^{-1/2+\eta}\sum_{j=1}^{t-1} j^{-3/2+\eta}(t-j)^{-\varsigma} \\
&\leq K\frac{1}{T^{2\eta}}\sum_{t=1}^{T}\left(t^{-1/2+\eta-\varsigma}+t^{-1+2\eta-\varsigma}\right)(1+\log t) = o(1),
\end{aligned}
$$

noting that $\varsigma > 1/2$, where the third inequality is due to Lemma S.9, to conclude the proof of (S.195).

Next, in view of (S.200), the proof of (S.196) follows by (S.202), (S.203) and

$$
\sup_{d_1\geq\lozenge, d_1-d_2\geq -1/2+\eta, \boldsymbol{\varphi}\in\Psi} \frac{1}{T^{2(d_1-d_2)+1}}\sum_{t=1}^{T} b_t^2(d_1,d_2)\left(\sum_{j=t}^{\infty}\phi_j(\boldsymbol{\varphi})\right)^2 = o(1),
$$

$$
\sup_{d_1\geq\lozenge, d_1-d_2\geq -1/2+\eta, \boldsymbol{\varphi}\in\Psi} \frac{1}{T^{2(d_1-d_2)+1}}\sum_{t=1}^{T}\left(\sum_{j=1}^{t-1} b_{j+1}(d_1,d_2+1)\sum_{l=1}^{j}\phi_{t-l}(\boldsymbol{\varphi})\right)^2 = o(1),
$$

which follow by straightforward arguments using (6), Lemma S.9 and (S.184).

The proof of (S.197) is immediate because

$$
\inf_{d_1,d_2,\boldsymbol{\varphi}\in\Psi}\sum_{t=1}^{T} c_t^2(d_1,d_2,\boldsymbol{\varphi}) \geq \inf_{d_1,d_2,\boldsymbol{\varphi}\in\Psi} c_1^2(d_1,d_2,\boldsymbol{\varphi}) = 1.
$$

For the proof of (S.198), denoting by $[\cdot]$ the integer part of the argument and given that

$$
\frac{1}{T^{2(d_1-d_2)+1}}\sum_{t=1}^{T} c_t^2(d_1,d_2,\boldsymbol{\varphi}) \geq \frac{1}{T^{2(d_1-d_2)+1}}\sum_{t=[T^{1/2}]}^{T} c_t^2(d_1,d_2,\boldsymbol{\varphi}), \tag{S.204}
$$

as in the proof of (S.195), the right-hand side of (S.204) is bounded from below by

$$
\epsilon \inf_{d_1\geq\lozenge, -1/2-\alpha\leq d_1-d_2\leq -1/2+\alpha} \frac{1}{T^{2(d_1-d_2)+1}}\sum_{t=[T^{1/2}]}^{T} b_t^2(d_1,d_2) \tag{S.205}
$$

$$
-\sup_{d_1\geq\lozenge, -1/2-\alpha\leq d_1-d_2\leq -1/2+\alpha, \boldsymbol{\varphi}\in\Psi} \frac{K}{T^{2(d_1-d_2)+1}}\left|\sum_{t=[T^{1/2}]}^{T} b_t^2(d_1,d_2)\sum_{j=t}^{\infty}\phi_j(\boldsymbol{\varphi})\right| \tag{S.206}
$$

$$
-\sup_{d_1\geq\lozenge, -1/2-\alpha\leq d_1-d_2\leq -1/2+\alpha, \boldsymbol{\varphi}\in\Psi} \frac{K}{T^{2(d_1-d_2)+1}}\left|\sum_{t=[T^{1/2}]}^{T} b_t(d_1,d_2)\sum_{j=1}^{t-1} b_{j+1}(d_1,d_2+1)\sum_{l=1}^{j}\phi_{t-l}(\boldsymbol{\varphi})\right|. \tag{S.207}
$$

First, by (6) and (S.184), (S.206) is bounded by

$$\sup_{-1/2-\alpha\leq d_1-d_2\leq -1/2+\alpha}\frac{K}{T}\sum_{t=[T^{1/2}]}^{T}\left(\frac{t}{T}\right)^{2(d_1-d_2)}t^{-\varsigma}\leq \frac{K}{T}\sum_{t=[T^{1/2}]}^{T}\left(\frac{t}{T}\right)^{-1-2\alpha}t^{-\varsigma}=O\left(T^{\alpha-\varsigma/2}\right)=o\left(1\right),$$

because $\alpha<(\varsigma-1/2)/3<\varsigma/2$. Similarly, (S.207) is bounded by

$$\begin{aligned}KT^{2\alpha}\sum_{t=[T^{1/2}]}^{T}t^{-1/2-\alpha}\sum_{j=1}^{t-1}j^{-3/2-\alpha}\left(t-j\right)^{-\varsigma}&\leq KT^{2\alpha}\log T\sum_{t=[T^{1/2}]}^{T}t^{-1/2-\alpha-\varsigma}\\&\leq KT^{3\alpha/2+1/4-\varsigma/2}\log T,\end{aligned}$$

which is $o(1)$ because $\alpha<(\varsigma-1/2)/3$.

Finally, by (S.184), (S.205) is bounded from below by

$$\begin{aligned}&\epsilon\inf_{d_1\geq\lozenge,-1/2-\alpha\leq d_1-d_2\leq -1/2+\alpha}\frac{\Gamma^2\left(d_1+1\right)}{T^{2(d_1-d_2)+1}}\sum_{t=\left[T^{1/2}\right]}^{T}\pi_t^2\left(d_1+1-d_2\right)\\&-K\sup_{d_1\geq\lozenge,-1/2-\alpha\leq d_1-d_2\leq -1/2+\alpha}\frac{1}{T^{2(d_1-d_2)+1}}\sum_{t=\left[T^{1/2}\right]}^{T}\left|\pi_t\left(d_1+1-d_2\right)\right|\left|s_{1t}\left(d_1,d_2\right)\right|.\end{aligned}\qquad\text{(S.208)}$$

By (S.184) and Lemma S.11 the second term on (S.208) is bounded by

$$\begin{aligned}&K\sup_{d_1\geq\lozenge,-1/2-\alpha\leq d_1-d_2\leq -1/2+\alpha}\frac{1}{T}\sum_{t=\left[T^{1/2}\right]}^{T}\left(\frac{t}{T}\right)^{2(d_1-d_2)}\left(t^{-1}+t^{-d_1-1}\right)\\&\leq KT^{2\alpha}\sum_{t=\left[T^{1/2}\right]}^{T}t^{-1-2\alpha}\left(t^{-1}+t^{-\lozenge-1}\right)\leq K\left(T^{\alpha-1/2}+T^{\alpha-1/2-\lozenge/2}\right)=o\left(1\right),\end{aligned}$$

because $\alpha<\min\left\{(\varsigma-1/2)/3,(1+\lozenge)/2\right\}$. Next using Lemma S.11 the first term on (S.208) is bounded from below by

$$\begin{aligned}\epsilon\inf_{-1/2-\alpha\leq d_1-d_2\leq -1/2+\alpha}\frac{1}{T}\sum_{t=\left[T^{1/2}\right]}^{T}\left(\frac{t}{T}\right)^{2(d_1-d_2)}&\geq\epsilon\frac{1}{T}\sum_{t=\left[T^{1/2}\right]}^{T}\left(\frac{t}{T}\right)^{-1+2\alpha}\geq\epsilon\int_{[T^{1/2}]/T}^{1}x^{2\alpha-1}dx\\&=\epsilon\frac{1-([T^{1/2}]/T)^{2\alpha}}{2\alpha}=\frac{\epsilon}{2\alpha}-O\left(T^{-\alpha}\right)\end{aligned}$$

In view of (S.204), (S.206), and (S.207), this proves (S.198).

Finally, the proof for (S.199) is almost identical to that for (S.198) with the only difference of the treatment of the first term on (S.208). Here, noting that $d_1-d_2\leq\overline{d}_1-\underline{d}_2$, defining $g_T=T^{-\frac{1}{2}\left(2\left(\overline{d}_1-\underline{d}_2\right)+1\right)}\mathbb{I}\left(\overline{d}_1-\underline{d}_2<-1/2\right)+\log T\mathbb{I}\left(\overline{d}_1-\underline{d}_2=-1/2\right)+\mathbb{I}\left(\overline{d}_1-\underline{d}_2>-1/2\right)$,

$$\epsilon\inf_{d_1-d_2\geq -1/2-\alpha}\frac{1}{T}\sum_{t=\left[T^{1/2}\right]}^{T}\left(\frac{t}{T}\right)^{2(d_1-d_2)}\geq\epsilon\frac{1}{T}\sum_{t=\left[T^{1/2}\right]}^{T}\left(\frac{t}{T}\right)^{2\left(\overline{d}_1-\underline{d}_2\right)}\geq\epsilon g_T\geq\epsilon,$$

to conclude the proof of Lemma S.15. ■

**Lemma S.16** *Let $\theta$ be an arbitrary number such that $0 < \theta < \varsigma - 1/2$. Then, under Assumptions A1 and A3, for any real numbers $\Diamond > -1$, $D_1 < -1/2 - \theta$ and $D_2 > -1/2 + \theta$, $m = 0, 1$, $0 \leq p \leq m$, uniformly in $t = 1, \ldots, T$ and $T \geq 1$,*

$$\sup_{d_1 \geq \Diamond, d_1 - d_2 \in [D_1, D_2], \boldsymbol{\varphi} \in \Psi} \left| \frac{\partial^m h_{t,T}(d_1, d_2, \boldsymbol{\varphi})}{\partial d_1^p \partial d_2^{m-p}} \right| = O(t^{-1/2-\theta} T^{\theta + 2\theta m}), \quad \text{(S.209)}$$

$$\sup_{d_1 \geq \Diamond, d_1 - d_2 \geq -1/2+\theta, \boldsymbol{\varphi} \in \Psi} \left| \frac{\partial h_{t,T}(d_1, d_2, \boldsymbol{\varphi})}{\partial d_1} \right| = O\left( t^{-1/2+\theta} T^{-\theta} \left(1 + |\log(t/T)|\right)\right), \quad \text{(S.210)}$$

$$\sup_{d_1 \geq \Diamond, d_1 - d_2 \in [D_1, D_2], \boldsymbol{\varphi} \in \Psi} \left| \frac{\partial^m (h_{t+1,T}(d_1, d_2, \boldsymbol{\varphi}) - h_{t,T}(d_1, d_2, \boldsymbol{\varphi}))}{\partial d_1^p \partial d_2^{m-p}} \right| = O(t^{-3/2-\theta} T^{\theta + 2\theta m}), \quad \text{(S.211)}$$

$$\sup_{d_1 \geq \Diamond, d_1 - d_2 \geq -1/2+\theta, \boldsymbol{\varphi} \in \Psi} \left| \frac{\partial (h_{t+1,T}(d_1, d_2, \boldsymbol{\varphi}) - h_{t,T}(d_1, d_2, \boldsymbol{\varphi}))}{\partial d_1} \right| = O\left( t^{-3/2+\theta} T^{-\theta} \left(1 + |\log(t/T)|\right)\right), \quad \text{(S.212)}$$

$$\sup_{d_1 \geq \Diamond, d_1 - d_2 \in [D_1, D_2], \boldsymbol{\varphi} \in \Psi} \left| \sum_{t=1}^{T} h_{t,T}(d_1, d_2, \boldsymbol{\varphi}) \right| = O(T^{1/2}). \quad \text{(S.213)}$$

**Proof.** The left-hand side of (S.209) is bounded by

$$\sup_{d_1 \geq \Diamond, D_1 \leq d_1 - d_2 \leq -1/2-\theta, \boldsymbol{\varphi} \in \Psi} \left| \frac{\partial^m}{\partial d_1^p \partial d_2^{m-p}} h_{t,T}(d_1, d_2, \boldsymbol{\varphi}) \right| + \sup_{d_1 \geq \Diamond, -1/2-\theta \leq d_1 - d_2 \leq D_2, \boldsymbol{\varphi} \in \Psi} \left| \frac{\partial^m}{\partial d_1^p \partial d_2^{m-p}} h_{t,T}(d_1, d_2, \boldsymbol{\varphi}) \right|. \quad \text{(S.214)}$$

Suppose first that $m = 0$. Using the definition (S.3) and applying (S.192) of Lemma S.14 and (S.197) of Lemma S.15, the first term of (S.214) is bounded by

$$\frac{\sup_{d_1 \geq \Diamond, D_1 \leq d_1 - d_2 \leq -1/2-\theta, \boldsymbol{\varphi} \in \Psi} |c_t(d_1, d_2, \boldsymbol{\varphi})|}{\inf_{d_1 \geq \Diamond, D_1 \leq d_1 - d_2 \leq -1/2-\theta, \boldsymbol{\varphi} \in \Psi} \left( \sum_{j=1}^{T} c_j^2(d_1, d_2, \boldsymbol{\varphi}) \right)^{1/2}} \leq \sup_{d_1 \geq \Diamond, D_1 \leq d_1 - d_2 \leq -1/2-\theta, \boldsymbol{\varphi} \in \Psi} |c_t(d_1, d_2, \boldsymbol{\varphi})| = O(t^{-1/2-\theta}),$$

so the bound in (S.209) applies to the first term of (S.214) (although it is not tight). Next, the second term of (S.214) is bounded by

$$\frac{\sup_{d_1 \geq \Diamond, -1/2-\theta \leq d_1 - d_2 \leq D_2, \boldsymbol{\varphi} \in \Psi} T^{-(d_1 - d_2)} |c_t(d_1, d_2, \boldsymbol{\varphi})|}{\inf_{d_1 \geq \Diamond, -1/2-\theta \leq d_1 - d_2 \leq D_2, \boldsymbol{\varphi} \in \Psi} \left( T^{-2(d_1 - d_2)} \sum_{j=1}^{T} c_j^2(d_1, d_2, \boldsymbol{\varphi}) \right)^{1/2}}.$$

By (S.193) of Lemma S.14 the numerator is $O(t^{-1/2-\theta} T^{1/2+\theta})$ and by (S.199) of Lemma S.15 the denominator is bounded from below by $\epsilon T^{1/2}$. Thus (S.209) for $m = 0$ follows.

Next, for the derivative we find

$$\frac{\partial}{\partial d_1^p \partial d_2^{1-p}} h_{t,T}(d_1, d_2, \boldsymbol{\varphi}) = \frac{\partial c_t(d_1, d_2, \boldsymbol{\varphi}) / \partial d_1^p \partial d_2^{1-p}}{\left(\sum_{j=1}^T c_j^2(d_1, d_2, \boldsymbol{\varphi})\right)^{1/2}} - \frac{h_{t,T}(d_1, d_2, \boldsymbol{\varphi}) \sum_{j=1}^T c_j(d_1, d_2, \boldsymbol{\varphi}) \partial c_j(d_1, d_2, \boldsymbol{\varphi}) / \partial d_1^p \partial d_2^{1-p}}{\sum_{j=1}^T c_j^2(d_1, d_2, \boldsymbol{\varphi})}. \tag{S.215}$$

First we show (S.209). Proceeding as in the proof for $m = 0$, taking into account the extra log-term arising from (S.192) in Lemma S.14, the first term of (S.215) is $O(t^{-1/2}(T/t)^\theta \log T)$, so the bound in (S.209) applies. Next, using again (S.192) in Lemma S.14 and also (S.209) for $m = 0$, the second term of (S.215) is $O(t^{-1/2}(T/t)^\theta T^{2\theta} \sum_{j=1}^T j^{-1-2\theta} \log j)$, so the bound in (S.209) applies for $m = 1$.

Next we show (S.210). Clearly

$$\frac{\partial h_{t,T}(d_1, d_2, \boldsymbol{\varphi})}{\partial d_1} = \frac{\partial T^{-(d_1-d_2)} c_t(d_1, d_2, \boldsymbol{\varphi}) / \partial d_1}{\left(\sum_{j=1}^T T^{-2(d_1-d_2)} c_j^2(d_1, d_2, \boldsymbol{\varphi})\right)^{1/2}} - \frac{T^{-(d_1-d_2)} c_t(d_1, d_2, \boldsymbol{\varphi}) \sum_{j=1}^T T^{-(d_1-d_2)} c_j(d_1, d_2, \boldsymbol{\varphi}) \partial T^{-(d_1-d_2)} c_j(d_1, d_2, \boldsymbol{\varphi}) / \partial d_1}{\left(\sum_{j=1}^T T^{-2(d_1-d_2)} c_j^2(d_1, d_2, \boldsymbol{\varphi})\right)^{3/2}}.$$

First,

$$\begin{aligned}
&\sup_{d_1 \geq \Diamond, d_1-d_2 \geq -1/2+\theta, \boldsymbol{\varphi} \in \Psi} \left| \frac{\partial T^{-(d_1-d_2)} c_t(d_1, d_2, \boldsymbol{\varphi}) / \partial d_1}{\left(\sum_{j=1}^T T^{-2(d_1-d_2)} c_j^2(d_1, d_2, \boldsymbol{\varphi})\right)^{1/2}} \right| \\
&\leq \frac{\sup_{d_1 \geq \Diamond, d_1-d_2 \geq -1/2+\theta, \boldsymbol{\varphi} \in \Psi} \left| \partial T^{-(d_1-d_2)} c_t(d_1, d_2, \boldsymbol{\varphi}) / \partial d_1 \right|}{\left(\inf_{d_1 \geq \Diamond, d_1-d_2 \geq -1/2+\theta, \boldsymbol{\varphi} \in \Psi} \sum_{j=1}^T T^{-2(d_1-d_2)} c_j^2(d_1, d_2, \boldsymbol{\varphi})\right)^{1/2}} \\
&= O\left(t^{-1/2}(t/T)^\theta (1 + |\log(t/T)|)\right)
\end{aligned} \tag{S.216}$$

by (S.193) of Lemma S.14 and (S.199) of Lemma S.15. Similarly, like in (S.216),

$$\sup_{d_1 \geq \Diamond, d_1-d_2 \geq -1/2+\theta, \boldsymbol{\varphi} \in \Psi} \left| \frac{T^{-(d_1-d_2)} c_t(d_1, d_2, \boldsymbol{\varphi})}{\left(\sum_{j=1}^T T^{-2(d_1-d_2)} c_j^2(d_1, d_2, \boldsymbol{\varphi})\right)^{1/2}} \right| = O\left(t^{-1/2}(t/T)^\theta\right), \tag{S.217}$$

so, by (S.216) and (S.217), it is straightforward to show that

$$\sup_{d_1 \geq \Diamond, d_1-d_2 \geq -1/2+\theta, \boldsymbol{\varphi} \in \Psi} \left| \frac{\frac{1}{T^{d_1-d_2}} c_t(d_1, d_2, \boldsymbol{\varphi}) \sum_{j=1}^T \frac{1}{T^{d_1-d_2}} c_j(d_1, d_2, \boldsymbol{\varphi}) \partial \frac{1}{T^{d_1-d_2}} c_j(d_1, d_2, \boldsymbol{\varphi}) / \partial d_1}{\left(\sum_{j=1}^T T^{-2(d_1-d_2)} c_j^2(d_1, d_2, \boldsymbol{\varphi})\right)^{3/2}} \right|$$

$$= O\left(t^{-1/2}(t/T)^\theta\right),$$

to conclude the proof of (S.210).

The proofs of (S.211)–(S.213) are omitted because they follow by identical arguments, noting that

$$h_{t,T}(d_1,d_2,\boldsymbol{\varphi}) - h_{t-1,T}(d_1,d_2,\boldsymbol{\varphi}) = \frac{c_t(d_1,d_2+1,\boldsymbol{\varphi})}{\left(\sum_{j=1}^{T} c_j^2(d_1,d_2,\boldsymbol{\varphi})\right)^{1/2}}.$$

■

**Lemma S.17** *Under Assumptions A1–A3, uniformly in $t=1,\ldots,T$, $T\geq 1$, and $\boldsymbol{\varphi}\in\Psi$,*

$$\sup_{d\leq g}\left|\phi(L;\boldsymbol{\varphi})\Delta^{-d}\{u_t\mathbb{I}(t\geq 1)\}\right| = O_p(t^{g-1/2}+\log t\,\mathbb{I}(g=1/2)+\mathbb{I}(g<1/2)), \quad \text{(S.218)}$$

$$\sup_{d\geq g}\left|T^{-d}\phi(L;\boldsymbol{\varphi})\Delta^{-d}\{u_t\mathbb{I}(t\geq 1)\}\right| = O_p(T^{-g}(t^{g-1/2}+\log t\,\mathbb{I}(g=1/2)+\mathbb{I}(g<1/2))). \quad \text{(S.219)}$$

**Proof.** First we show (S.218). Write $\phi(L;\boldsymbol{\varphi})\Delta^{-d}\{u_t\mathbb{I}(t\geq 1)\} = \sum_{j=0}^{t-1} a_j(d,\boldsymbol{\varphi})u_{t-j}$ and apply summation by parts,

$$\sum_{j=0}^{t-1} a_j(d,\boldsymbol{\varphi})u_{t-j} = a_{t-1}(d,\boldsymbol{\varphi})\sum_{j=0}^{t-1}u_{t-j} - \sum_{j=0}^{t-2}(a_{j+1}(d,\boldsymbol{\varphi})-a_j(d,\boldsymbol{\varphi}))\sum_{l=0}^{j}u_{t-l}. \quad \text{(S.220)}$$

Noting that $a_{j+1}(d,\boldsymbol{\varphi})-a_j(d,\boldsymbol{\varphi}) = a_{j+1}(d-1,\boldsymbol{\varphi})$, the right-hand side of (S.220) is bounded by

$$|a_{t-1}(d,\boldsymbol{\varphi})|\left|\sum_{j=0}^{t-1}u_{t-j}\right| + \sum_{j=0}^{t-2}|a_{j+1}(d-1,\boldsymbol{\varphi})|\left|\sum_{l=0}^{j}u_{t-l}\right|. \quad \text{(S.221)}$$

Under our conditions, $E\left|\sum_{l=1}^{t}u_l\right| = O\left(t^{1/2}\right)$, so, in view of (S.194) of Lemma S.14, the expectation of the left-hand side of (S.218) is bounded by

$$Kt^{\max\{g-1/2,-1/2-\varsigma\}} + K\sum_{j=1}^{t} j^{\max\{g-3/2,-1/2-\varsigma\}} \leq K(t^{g-1/2}+\log t\,\mathbb{I}(g=1/2)+\mathbb{I}(g<1/2))$$

to conclude the proof of (S.218). The proof of (S.219) is omitted because it is almost identical to that for (S.218). ■

**Lemma S.18** *Let $\theta$ be an arbitrary number such that $0<\theta<\varsigma-1/2$. Then, under Assumptions A1–A3, for $m=0,1$, and uniformly in $\boldsymbol{\vartheta}\in\Xi$,*

$$\sup_{\boldsymbol{\vartheta}\in\Xi,\delta_0-\delta\leq g}\left|\sum_{t=1}^{T}\varepsilon_t(\boldsymbol{\tau})\frac{\partial^m}{\partial\gamma^m}h_{t,T}(\gamma,\delta,\boldsymbol{\varphi})\right| = O_p(T^{\max\{\theta,g\}+2\theta m}), \quad \text{(S.222)}$$

$$\sup_{\boldsymbol{\vartheta}\in\Xi,\delta_0-\delta\geq g}\frac{1}{T^{\delta_0-\delta}}\left|\sum_{t=1}^{T}\varepsilon_t(\boldsymbol{\tau})\frac{\partial^m}{\partial\gamma^m}h_{t,T}(\gamma,\delta,\boldsymbol{\varphi})\right| = O_p(T^{\max\{\theta,g\}-g+2\theta m}), \quad \text{(S.223)}$$

$$\sup_{\boldsymbol{\vartheta}\in\Xi,\delta_0-\delta\leq g,\gamma-\delta\geq -1/2+\theta}\left|\sum_{t=1}^{T}\varepsilon_t(\boldsymbol{\tau})\frac{\partial}{\partial\gamma}h_{t,T}(\gamma,\delta,\boldsymbol{\varphi})\right| = O_p(T^{\max\{\theta,g\}}), \quad \text{(S.224)}$$

$$\sup_{\boldsymbol{\vartheta}\in\Xi,\delta_0-\delta\geq g,\gamma-\delta\geq -1/2+\theta}\frac{1}{T^{\delta_0-\delta}}\left|\sum_{t=1}^{T}\varepsilon_t(\boldsymbol{\tau})\frac{\partial}{\partial\gamma}h_{t,T}(\gamma,\delta,\boldsymbol{\varphi})\right| = O_p(T^{\max\{\theta,g\}-g}). \quad \text{(S.225)}$$

**Proof.** By summation by parts as in (S.237), we find

$$\left|\sum_{t=1}^{T}\varepsilon_t(\boldsymbol{\tau})h_{t,T}\left(\gamma,\delta,\boldsymbol{\varphi}\right)\right| \leq \left|h_{T,T}\left(\gamma,\delta,\boldsymbol{\varphi}\right)\right|\left|\varepsilon_T(\delta_0-\delta+1,\boldsymbol{\varphi})\right| + \sum_{t=1}^{T-1}\left|h_{t+1,T}\left(\gamma,\delta,\boldsymbol{\varphi}\right)-h_{t,T}\left(\gamma,\delta,\boldsymbol{\varphi}\right)\right|\left|\varepsilon_t(\delta_0-\delta+1,\boldsymbol{\varphi})\right|,$$

noting (S.27). First, application of (S.209), (S.211) of Lemma S.16 together with (S.218), (S.219) of Lemma S.17 implies (S.222) and (S.223). Next, (S.224) and (S.225) follow from (S.210), (S.212) of Lemma S.16 and (S.218), (S.219) of Lemma S.17. ■

**Lemma S.19** *Under Assumptions A1–A3, for any* $g_2 > -1/2$ *and for any arbitrary* $\theta$ *such that* $0 < \theta < \min\{\varsigma - 1/2, g_2 + 1/2\}$,

$$\left|\sum_{t=1}^{T} d_t\left(\boldsymbol{\vartheta}\right)s_t\left(\boldsymbol{\vartheta}\right)\right| \leq \left|\gamma-\gamma_0\right|\left|M_T\left(\boldsymbol{\vartheta}\right)\right|, \tag{S.226}$$

*where, uniformly in* $\boldsymbol{\vartheta} \in \Xi$,

$$\sup_{\boldsymbol{\vartheta}\in\Xi,\delta_0-\delta\leq g_1,\gamma_0-\delta\leq g_2}\left|M_T\left(\boldsymbol{\vartheta}\right)\right| = O_p(T^{\max\{\theta,g_1\}+2\theta+g_2+1/2}), \tag{S.227}$$

$$\sup_{\boldsymbol{\vartheta}\in\Xi,\delta_0-\delta\leq g_1,\gamma_0-\delta\geq g_2} T^{\delta-\gamma_0}\left|M_T\left(\boldsymbol{\vartheta}\right)\right| = O_p(T^{\max\{\theta,g_1\}+2\theta+1/2}), \tag{S.228}$$

$$\sup_{\boldsymbol{\vartheta}\in\Xi,\delta_0-\delta\geq g_1,\gamma_0-\delta\geq g_2} T^{2\delta-\delta_0-\gamma_0}\left|M_T\left(\boldsymbol{\vartheta}\right)\right| = O_p(T^{\max\{\theta,g_1\}+2\theta-g_1+1/2}), \tag{S.229}$$

$$\sup_{\boldsymbol{\vartheta}\in\Xi,\delta_0-\delta\leq g_1,\gamma_0-\delta\geq g_2,\gamma-\delta\geq-1/2+\theta} T^{\delta-\gamma_0}\left|M_T\left(\boldsymbol{\vartheta}\right)\right| = O_p(T^{\max\{\theta,g_1\}+1/2}), \tag{S.230}$$

$$\sup_{\boldsymbol{\vartheta}\in\Xi,\delta_0-\delta\geq g_1,\gamma_0-\delta\geq g_2,\gamma-\delta\geq-1/2+\theta} T^{2\delta-\delta_0-\gamma_0}\left|M_T\left(\boldsymbol{\vartheta}\right)\right| = O_p(T^{\max\{\theta,g_1\}-g_1+1/2}). \tag{S.231}$$

**Proof.** Letting $d_t\left(\boldsymbol{\tau},\gamma\right) = d_t\left(\boldsymbol{\vartheta}\right)$, noting (S.15) and that $d_t\left(\boldsymbol{\tau},\gamma_0\right) = 0$, by the mean value theorem,

$$\left|\sum_{t=1}^{T} d_t\left(\boldsymbol{\vartheta}\right)s_t\left(\boldsymbol{\vartheta}\right)\right| \leq \left|\gamma-\gamma_0\right|\left|\frac{\partial}{\partial\gamma}\sum_{t=1}^{T} d_t\left(\boldsymbol{\tau},\overline{\gamma}\right)\varepsilon_t(\boldsymbol{\tau})\right|,$$

where $\left|\overline{\gamma}-\gamma_0\right| \leq \left|\gamma-\gamma_0\right|$. Then we find the bound

$$\left|\frac{\partial}{\partial\gamma}\sum_{t=1}^{T} d_t\left(\boldsymbol{\tau},\gamma\right)\varepsilon_t(\boldsymbol{\tau})\right| \leq \left|\sum_{t=1}^{T}\varepsilon_t(\boldsymbol{\tau})\frac{\partial h_{t,T}\left(\gamma,\delta,\boldsymbol{\varphi}\right)}{\partial\gamma}\sum_{j=1}^{T}c_j\left(\gamma_0,\delta,\boldsymbol{\varphi}\right)h_{j,T}\left(\gamma,\delta,\boldsymbol{\varphi}\right)\right| + \left|\sum_{t=1}^{T}\varepsilon_t(\boldsymbol{\tau})h_{t,T}\left(\gamma,\delta,\boldsymbol{\varphi}\right)\sum_{j=1}^{T}c_j\left(\gamma_0,\delta,\boldsymbol{\varphi}\right)\frac{\partial h_{j,T}\left(\gamma,\delta,\boldsymbol{\varphi}\right)}{\partial\gamma}\right|.$$

The results (S.227)–(S.229) now all follow by direct application of (S.232), (S.233) of Lemma S.20 with $\theta < g + 1/2$ and (S.222), (S.223) of Lemma S.18. Results (S.230) and (S.231) are derived straightforwardly from (S.224), (S.225) of Lemma S.18 and (S.234) of Lemma S.20. ■

**Lemma S.20** *Let $\theta$ be an arbitrary number such that $0 < \theta < \varsigma - 1/2$. Then, under Assumptions A1 and A3, for $m = 0, 1$ and uniformly in $\boldsymbol{\vartheta} \in \Xi$,*

$$\sup_{\boldsymbol{\vartheta}\in\Xi,\gamma_0-\delta\leq g} \left| \sum_{t=1}^{T} c_t\left(\gamma_0,\delta,\boldsymbol{\varphi}\right) \frac{\partial^m}{\partial\gamma^m} h_{t,T}\left(\gamma,\delta,\boldsymbol{\varphi}\right) \right| = O(T^{\max\{\theta,g+1/2\}+2\theta m}\log T), \tag{S.232}$$

$$\sup_{\boldsymbol{\vartheta}\in\Xi,\gamma_0-\delta\geq g} \frac{1}{T^{\gamma_0-\delta}} \left| \sum_{t=1}^{T} c_t\left(\gamma_0,\delta,\boldsymbol{\varphi}\right) \frac{\partial^m}{\partial\gamma^m} h_{t,T}\left(\gamma,\delta,\boldsymbol{\varphi}\right) \right| = O(T^{\max\{\theta,g+1/2\}-g+2\theta m}\log T), \tag{S.233}$$

*and for $g > -1/2$, uniformly in $\boldsymbol{\vartheta} \in \Xi$,*

$$\sup_{\boldsymbol{\vartheta}\in\Xi,\gamma_0-\delta\geq g,\gamma-\delta\geq -1/2+\theta} \frac{1}{T^{\gamma_0-\delta}} \left| \sum_{t=1}^{T} c_t\left(\gamma_0,\delta,\boldsymbol{\varphi}\right) \frac{\partial}{\partial\gamma} h_{t,T}\left(\gamma,\delta,\boldsymbol{\varphi}\right) \right| = O(T^{1/2}). \tag{S.234}$$

**Proof.** The results follow by direct application of (S.209), (S.210) in Lemma S.16 and (S.192), (S.193) of Lemma S.14. ■

**Lemma S.21** *Under Assumptions A1–A3, uniformly in $\boldsymbol{\vartheta} \in \Xi$,*

$$\sup_{\boldsymbol{\vartheta}\in\Xi,\delta_0-\delta\leq g_1,\gamma_0-\delta\leq g_2} \frac{1}{T} \left| \sum_{t=1}^{T} \varepsilon_t(\boldsymbol{\tau}) c_t\left(\gamma_0,\delta,\boldsymbol{\varphi}\right) \right|$$
$$= O_p(T^{g_1+g_2-1/2} + T^{-1}\log T + T^{g_1-1/2-\varsigma} + T^{g_2-1}\log^2 T\mathbb{I}\left(g_1 \leq -1/2\right)), \tag{S.235}$$

$$\sup_{\boldsymbol{\vartheta}\in\Xi,\delta_0-\delta\geq g_1,\gamma_0-\delta\geq g_2} \frac{1}{T^{\gamma_0+\delta_0-2\delta}} \left| \sum_{t=1}^{T} \varepsilon_t(\boldsymbol{\tau}) c_t\left(\gamma_0,\delta,\boldsymbol{\varphi}\right) \right|$$
$$= O_p(T^{1/2} + T^{1/2-g_2-\varsigma} + T^{-g_1-g_2}\log T + T^{-g_1}\log^2 T\mathbb{I}\left(g_1 \leq -1/2\right)). \tag{S.236}$$

**Proof.** By summation by parts and (S.53) we find

$$\left| \sum_{t=1}^{T} \varepsilon_t(\boldsymbol{\tau}) c_t\left(\gamma_0,\delta,\boldsymbol{\varphi}\right) \right| \leq \left| c_T\left(\gamma_0,\delta,\boldsymbol{\varphi}\right) \right| \left| \varepsilon_T(\delta_0-\delta+1,\boldsymbol{\varphi}) \right|$$
$$+ \left| \sum_{t=1}^{T-1} c_{t+1}\left(\gamma_0,\delta+1,\boldsymbol{\varphi}\right) \varepsilon_t(\delta_0-\delta+1,\boldsymbol{\varphi}) \right|, \tag{S.237}$$

see (S.27). The result (S.235) then follows by application of (S.218) of Lemma S.17 and (S.192) of Lemma S.14, while the result (S.236) follows by application of (S.219) of Lemma S.17 and (S.193) of Lemma S.14. ■

## References


1. Abramowitz, M. and Stegun, I.A. (1970). *Handbook of Mathematical Functions*, National Bureau of Standards, Washington, D.C.
2. Gradshteyn, I.S. and Ryzhik, I.M. (2000). *Tables of Integrals, Series and Products*, Academic Press, New York.

3. Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and Its Application*, Academic Press, New York.
4. Horn, R.A., and Johnson, C.R. (2013). *Matrix Analysis*, 2nd ed., Cambridge University Press, Cambridge.
5. Hosoya, Y. (2005). Fractional invariance principle. *Journal of Time Series Analysis* **26**, 463–486.
6. Hualde, J. and Nielsen, M.Ø. (2019). Truncated sum of squares estimation of fractional time series models with deterministic trends. QED working paper 1376, Queen's University.
7. Hualde, J. and Robinson, P.M. (2011). Gaussian pseudo-maximum likelihood estimation of fractional time series models. *Annals of Statistics* **39**, 3152–3181.
8. Johansen, S. and Nielsen, M.Ø. (2010). Likelihood inference for a nonstationary fractional autoregressive model. *Journal of Econometrics* **158**, 51–66.
9. Johansen, S. and Nielsen, M.Ø. (2012). Likelihood inference for a fractionally cointegrated vector autoregressive model. *Econometrica* **80**, 2667–2732.
10. Johansen, S. and Nielsen, M.Ø. (2018). Nonstationary cointegration in the fractionally cointegrated VAR model. Forthcoming in *Journal of Time Series Analysis*.
11. Robinson, P.M. and Hualde, J. (2003). Cointegration in fractional systems with unknown integration orders. *Econometrica* **71**, 1727–1766.
12. Robinson, P.M. and Marinucci, D. (2000). The averaged periodogram for nonstationary vector time series. *Statistical Inference for Stochastic Processes* **3**, 149–160.
13. Rutherford, D.E. (1948). Some continuant determinants arising in physics and chemistry. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics* **62**, 229–236.
14. Tanaka, K. (1996). *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*, John Wiley and Sons, New York.