Truncated sum-of-squares estimation of fractional time series models with generalized power law trend

Javier Hualde and Morten Ørregaard Nielsen

CREATES Research Paper 2022-07

Truncated sum-of-squares estimation of fractional time series models with generalized power law trend[*]

Javier Hualde[†]
Universidad Pública de Navarra

Morten Ørregaard Nielsen[‡]
Aarhus University

January 19, 2022

Abstract

We consider truncated (or conditional) sum-of-squares estimation of a parametric fractional time series model with an additive deterministic structure. The latter consists of both a drift term and a generalized power law trend. The memory parameter of the stochastic component and the power parameter of the deterministic trend component are both considered unknown real numbers to be estimated and belonging to arbitrarily large compact sets. Thus, our model captures different forms of nonstationarity and noninvertibility as well as a very flexible deterministic specification. 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 and due to the competition between stochastic and deterministic components. As expected, parameter estimates related to the deterministic component are shown to be consistent and asymptotically normal only for parts of the parameter space depending on the relative strength of the stochastic and deterministic components. In contrast, we establish consistency and asymptotic normality of parameter estimates related to the stochastic component for the entire parameter space. Furthermore, the asymptotic distribution of the latter estimates is unaffected by the presence of the deterministic component, even when this is not consistently estimable. We also include a small Monte Carlo simulation to illustrate our results.

MSC 2010 subject classifications: Primary 62M10; secondary 62F12.

JEL classification: C22.

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

[†]Email: `javier.hualde@unavarra.es`

[‡]Corresponding author. Email: `mon@econ.au.dk`

1

1 Introduction

A common approach to time series modeling is to assume an additive structure, where the observed process is a sum of latent stochastic and deterministic components. Regarding the former, the autoregressive moving average (ARMA) class, possibly including unit root nonstationary and noninvertible processes, is dominant. A general model that includes these processes as special cases, and also bridges the gap between stationary and invertible ARMA models and the unit root nonstationary or noninvertible models, is the fractionally integrated time series model. Specifically, a zero-mean fractional model for z_t is given by

$$\Delta_+^\delta z_t = u_t, \quad t \in \mathbb{Z}, \tag{1}$$

$$u_t = \omega(L; \boldsymbol{\varphi})\varepsilon_t, \quad t \in \mathbb{Z}, \tag{2}$$

where ε_t is a zero-mean and serially uncorrelated sequence, L is the lag operator, and, for any process ξ_t, $\Delta_+^\zeta \xi_t = \Delta^\zeta \xi_t \mathbb{I}(t \geq 1) = \sum_{i=0}^{t-1} \pi_i(-\zeta) \xi_{t-i} \mathbb{I}(t \geq 1)$ with $\mathbb{I}(\cdot)$ denoting the indicator function, $\pi_i(v) = 0$ for $i < 0$, $\pi_0(v) = 1$, and

$$\pi_i(v) = \frac{\Gamma(v+i)}{\Gamma(v)\Gamma(1+i)} = \frac{v(v+1)\dots(v+i-1)}{i!}, \quad i \geq 1, \tag{3}$$

denoting the coefficients in the binomial expansion of $(1-z)^{-v}$. The parameter δ is the "memory" of the process, which satisfies $\delta = 0$ for stationary and invertible ARMA models, $\delta = 1$ for unit root nonstationary models, and $\delta = -1$ for unit root noninvertible models.

The process z_t generated by (1) has been termed by the literature as a Type II fractionally integrated process of order δ. The Type II specification (1) assumes that the process is initialized at $t = 1$, but at the cost of more complicated proofs we conjecture that this could be generalized to any initialization under suitable conditions on the initial values. Johansen and Nielsen (2010, 2012a, 2016) developed maximum likelihood-based inference theory for fractional processes under more general assumptions on the initialization, where the inference is conditional on a finite number of initial values. Of course, finite sample behavior may depend on the initialization as investigated by Johansen and Nielsen (2016) using higher-order asymptotic expansions and numerical methods. To avoid further complications of the theory, we maintain the simpler Type II initialization in (1).

The function ω in (2) characterizes parametrically the short memory dependence present in u_t, and hence in z_t. It is given as

$$\omega(L; \boldsymbol{\varphi}) = \sum_{j=0}^{\infty} \omega_j(\boldsymbol{\varphi}) L^j, \tag{4}$$

where $\boldsymbol{\varphi}$ is an unknown $p \times 1$ vector collecting the short-memory parameters. For example, we could have

$$\omega(L; \boldsymbol{\varphi}) = \frac{\alpha_{MA}(L; \boldsymbol{\varphi})}{\alpha_{AR}(L; \boldsymbol{\varphi})}, \tag{5}$$

where α_{AR} and α_{MA} are polynomials of orders p_1 and p_2, respectively, with no common zeros and all roots outside the unit circle. Then (1),(2),(5) constitutes the fractionally integrated ARMA, or FARIMA(p_1, δ, p_2), model, where $\boldsymbol{\varphi}$ collects the AR and MA parameters. However, we will maintain the more general short-memory specification (4) instead of (5),

so that, under our conditions, z_t will not be restricted to belong to the FARIMA class. More precise conditions will be imposed on ω below.

An important implication of model (1),(2) is that $E(z_t) = 0$, so therefore, even if this setting has been employed in theoretical work (e.g., Hualde and Robinson, 2011 or Nielsen, 2015), it has limited empirical relevance. In practice, it would be natural to extend (1),(2) to allow for some deterministic structure such as, for example, a non-zero mean or drift. A very simple possibility would be to consider the observable process x_t generated as

$$\Delta_+^\delta (x_t - \mu) = u_t, \quad t \in \mathbb{Z}, , \tag{6}$$

which implies that

$$x_t = \mu \mathbb{I}(t \geq 1) + z_t = \mu \pi_{t-1}(1) + z_t, \quad t \in \mathbb{Z}, , \tag{7}$$

so x_t is a fractionally integrated process with drift if $\mu \neq 0$. This is a particular case of previous proposals in the literature, like Robinson (1994, 2005) or Robinson and Iacone (2005). An alternative specification for the deterministic component could be

$$\Delta_+^\delta x_t = \beta + u_t \text{ or } x_t = \Delta_+^{-\delta}\beta + \Delta_+^{-\delta}u_t = \beta\pi_{t-1}(\delta+1) + z_t, \tag{8}$$

see also Robinson (1994). By Stirling's approximation, $\pi_{t-1}(1+c)$ behaves like t^c, apart from a constant factor, and $\pi_{t-1}(1+c)$ is therefore denoted a *generalized power law trend* of order c or a *generalized polynomial trend* of order c. Thus, the presence of a non-zero β in (8) generates a deterministic trend component for x_t.

Several ideas arise from the previous discussion. First, the specification in (8) shows that, when dealing with fractional time series, the fractional coefficients π_{t-1} appear to be a more natural and elegant representation of deterministic terms compared with the usual powers of t. Of course, for the commonly applied cases $c = 0, 1$, $\pi_{t-1}(1+c)$ equals t^c for $t \geq 1$. More generally, considering π_{t-1} appears to be a natural approach to introducing a deterministic term that complements the Type II process z_t, due to the properties

$$\Delta_+^d \pi_{t-1}(c) = \pi_{t-1}(c-d) \text{ and } \sum_{j=0}^{t-1} \pi_j(c) = \pi_{t-1}(1+c). \tag{9}$$

Second, in view of (1), restricting the fractional order of the stochastic and deterministic components to be identical in (8) seems arbitrary. This motivates the more general model

$$x_t = \beta\pi_{t-1}(\gamma) + z_t, \tag{10}$$

where β and γ are both unknown real-valued parameters and z_t is generated by (1),(2), so, in particular, it can be either short or long memory. Under the model (1),(2),(10), the fractional order of the stochastic term is δ and that of the deterministic term is γ, thus allowing these orders to be modeled by two different parameters. In related work, Hualde and Nielsen (2020) analyze the model

$$x_t = \beta t_+^{\gamma-1} + z_t, \tag{11}$$

where $t_+^{\gamma-1} = t^{\gamma-1}\mathbb{I}(t \geq 1)$ (strictly speaking, they considered the power t_+^γ, but this makes no difference with the obvious re-labeling). Of course, since $\pi_{t-1}(\gamma) = t^{\gamma-1}\mathbb{I}(t \geq 1)$ for $\gamma = 1, 2,$

(10) and (11) are identical specifications for $\gamma = 1, 2$. Model (11) is a particular case of those of Robinson (2005) or Robinson and Iacone (2005), where more deterministic terms are included, but the power law parameters were assumed to be known. Also, (11) is embedded in Robinson's (2012) spatial model, where the power law parameters were unknown, but which required a short memory stochastic component. As will be seen below, apart from being more natural in a fractional setting, the specification (10) offers several crucial advantages over (11). Therefore, (11) is more accurately described as an approximation to (10) in a fractional setting with $\gamma \neq 1, 2$.

Third, the model (10) (or (11)) is, undoubtedly, quite restrictive. For example, it cannot accommodate the commonly applied constant plus linear trend specification. To illustrate the restrictive nature of (10) or (11), note that for both models $x_1 = \beta + z_1$, which implies that the β parameter is intimately linked to the initial observation. Thus, in practice, a large value of x_1 will lead to a large value of the estimated β, even in cases where the true slope of the deterministic component is small. This discussion just indicates that models (10) or (11) might be misspecified in most cases due to the omission of a drift component that determines the level of the observable time series. Thus, inspired by (7) and (10), we consider instead of (10) the more general structure

$$x_t = \mu \pi_{t-1}(1) + \beta \pi_{t-1}(\gamma) + z_t, \tag{12}$$

where for identification $\gamma \neq 1$. The deterministic specification in (12) covers the standard case of constant plus linear trend (when $\gamma = 2$). However, since γ is allowed to take any real value, (12) characterizes a wide range of deterministic behaviors. Compared to (10) (or (11)), the generalization in (12) appears to be particularly relevant for $\gamma < 1$, because in this case the deterministic structure would approach smoothly the drift μ as time increases. One interpretation of this case, which appears to be both realistic and coherent with e.g. economic time series, is that of a series moving around a deterministic structure that approaches an equilibrium value (given by μ). Furthermore, the parameter μ is the so-called "level parameter" in the terminology of Johansen and Nielsen (2016), the inclusion of which they argue can alleviate bias issues arising from non-zero initial conditions in the Type II context.

Fourth, model (12) is closed under fractional differencing in the sense that, for any $d \in \mathbb{R}$,

$$\Delta_+^d x_t = \mu \pi_{t-1}(1-d) + \beta \pi_{t-1}(\gamma - d) + \Delta_+^{d-\delta} u_t \tag{13}$$

has the exact same structure as (12). This property is an important advantage, because, as will be seen below, our proposed estimators depend crucially on fractional differences of the observables, and (13) will simplify the estimation procedure by avoiding the presence of "approximation errors" which routinely appear when taking fractional differences of polynomials of t. It is desirable, both from a mathematical and from a practical point of view, that fractional differences of a process belonging to a class of fractional time series belong to the same class of fractional time series. In other words, it would seem strange if fractional differences of a fractional time series process generated processes that were outside the class.

Unlike (10) and (12), the alternative model (11) given with $t_+^{\gamma-1}$ instead of $\pi_{t-1}(\gamma)$ is not closed under fractional differencing. Thus, modeling trends by means of π_{t-1} seems both more natural and more elegant than a power of t, and leads to several advantages, which are even more relevant because of the extra flexibility introduced by the drift term μ. In particular, we note that the properties (9) are only shared approximately by t_+^c, in the sense that $\Delta_+^d t_+^c$ is

only asymptotically equal to a constant times t_+^{c-d}, and only for some values of c and d. From a technical viewpoint, the "closedness" leads to simpler and more elegant proof arguments. From a practical viewpoint, for the model (11) discussed in Hualde and Nielsen (2020), only the case $\gamma > 0$ could be considered. Nicely, considering generalized trends (π_{t-1}) instead of powers of t permits consideration of any value of γ in an arbitrarily large compact set.

In our model (1),(2),(12), we allow both the stochastic and deterministic components to be of a fractional order, thereby placing them on an equal footing. Specifically, δ (which we permit to lie in an arbitrarily large compact interval) characterizes the behavior of $Var(x_t)$ and $Cov(x_t, x_{t-j})$ and γ characterizes the behaviour of $E(x_t)$. Thus, borrowing White and Granger's (2011) terminology, when $\delta > 1/2$, x_t has an increasing "stochastic trend in variance" because $Var(x_t)$ grows at rate $t^{2\delta-1}$. Similarly, x_t has a "stochastic trend in mean" because $E(x_t)$ grows at rate $t^{\gamma-1}$. In this sense, letting γ be real-valued appears as natural as letting δ be real-valued. Moreover, in the context of fractional models, letting γ be real-valued seems to be a natural alternative to the more standard linear trend model, which might suffer from severe specification problems with important implications for inference.

We note that if γ were known in (12), the estimation problem is simplified greatly. This case has been studied by Robinson (2005) who considered M-estimation of a model like (11) (although involving more deterministic terms) with *known* γ and allowing for fractional z_t. Other works that have considered a similar problem to ours, but assuming that z_t is at most a weakly dependent process, include Wu (1981), Phillips (2007), and Robinson (2012). The latter analyzes a more general spatial setting with more deterministic terms, but where the weakly dependent stochastic component is dominated by the deterministic structure. Related to the implications of the relative strength of deterministic and stochastic components for estimation, Johansen and Nielsen (2016) proved consistency and asymptotic normality of a truncated/conditional sum-of-squares estimator that ignores the deterministic components, whenever this component is dominated by the stochastic one. Other contributions that include power law trends, although in slightly different contexts, are Robinson and Marinucci (2000) and Robinson and Iacone (2005). Finally, as mentioned above, Hualde and Nielsen (2020) analyze the model given by (1),(2),(11), using the approximation given by the simple power $t_+^{\gamma-1}$ instead of the fractional coefficients $\pi_{t-1}(\gamma)$. Their work has two very important limitations with respect to our contribution in the present paper. First, they have no drift term, which, as justified before, limits substantially the empirical relevance of their model. Second, as a consequence of applying the approximation $t_+^{\gamma-1}$, they require a strong condition on the power law parameter, namely $\gamma > 0$, so cases where the deterministic structure tends very quickly to an equilibrium value (known to be 0 in their model) are not covered.

In the present paper we derive the limiting properties of a truncated/conditional sum-of-squares estimator of parameters in model (1),(2),(12). As in Hualde and Nielsen (2020), our setting is substantially more involved than in the related analysis of Hualde and Robinson (2011) (or Nielsen, 2015), which discussed parametric estimation in model (12) taking $\mu = \beta = 0$ as known. This avoids the huge complication of dealing with the competition between stochastic and deterministic components, especially related to the proof of consistency (prerequisite for proving limiting normality), which is very delicate because the loss function does not converge uniformly over a large admissible parameter space, even when knowledge of $\mu = \beta = 0$ is imposed. Additionally, the presence of a second deterministic term (the drift) over the simpler setting of Hualde and Nielsen (2020) introduces very sub-

stantial technical challenges.

As in Hualde and Nielsen (2020), we establish the limiting properties of our estimators, noting that results depend on the relative strength of the deterministic and stochastic components. We distinguish different cases depending on the true values of the parameters, which are denoted by subscript zero. When $\gamma_0 - 1/2 > \delta_0$ and $\delta_0 < 1/2$, we find that the estimators of all parameters are consistent and asymptotically normal. Next, when $\gamma_0 - 1/2 > \delta_0$ but $\delta_0 > 1/2$, consistency and asymptotic normality hold for all estimators except that of μ_0. Alternatively, when $\gamma_0 - 1/2 < \delta_0$, only the estimators of the parameters related to the stochastic component $(\delta_0, \boldsymbol{\varphi}_0)$ are consistent and asymptotically normal. In this case, the joint limiting distribution of the estimators of δ_0 and $\boldsymbol{\varphi}_0$ is unaffected by the presence of deterministic components that cannot be consistently estimated; a phenomenon that has been noted previously in, e.g., Heyde and Dai (1996), Abadir, Distaso, and Giraitis (2007), Iacone (2010), and Hualde and Nielsen (2020). If, in this case, $\delta_0 < 1/2$, we also provide a convergence rate for the estimator of μ_0.

The rest of the paper is organized as follows. First, in Section 2 we discuss the estimation problem in model (12) and compare with several alternatives. In Section 3 we present the main theoretical results of the paper. Next, a Monte Carlo experiment of finite sample performance is presented in Section 4, and we give some concluding remarks in Section 5. Finally, Section 6 collects the proofs of our main results, while all lemmas are stated in Sections 7 and 8. The proofs of all lemmas can be found in Sections 9 and 10.

2   The estimation problem

We introduce the following notation. Let prime denote transposition. The parameter for the stochastic component is $\boldsymbol{\tau} = (\delta, \boldsymbol{\varphi}')'$ with true value $\boldsymbol{\tau}_0 = (\delta_0, \boldsymbol{\varphi}_0')'$. We also define the parameters $\boldsymbol{\vartheta} = (\boldsymbol{\tau}', \gamma)'$ and $\boldsymbol{\phi} = (\mu, \beta)'$ with true values $\boldsymbol{\vartheta}_0 = (\boldsymbol{\tau}_0', \gamma_0)'$ and $\boldsymbol{\phi}_0 = (\mu_0, \beta_0)'$. Thus, the model is given by (12), which we repeat here for convenience,

$$x_t = \mu \pi_{t-1}(1) + \beta \pi_{t-1}(\gamma) + z_t,$$

where z_t is a stochastic term modeled by (1),(2). The data generating process (DGP) is given by (1),(2),(12) with parameters $\boldsymbol{\vartheta}_0$ and $\boldsymbol{\phi}_0$.

To illustrate the estimation problem, suppose for now that $\omega(L; \boldsymbol{\varphi}) = 1$, and hence $u_t = \varepsilon_t$, and ignore $\boldsymbol{\varphi}$, so that the parameters to be estimated are $\delta, \gamma, \mu, \beta$. Hualde and Nielsen's (2020) proof methods and results make evident the significant technical difficulties of enlarging their relatively restrictive model (11) with additional deterministic terms to capture a richer structure (e.g., including a drift). Thus, an interesting and tempting proposal is the "differencing and adding back" procedure, where the idea is to eliminate the drift by differencing to simplify the estimation problem. A similar idea has been considered for estimation of nonstationary processes by Velasco (1999a, 1999b) and Chen and Hurvich (2003) in combination with tapering of the periodogram in frequency domain methods. Specifically, consider $y_t = \Delta_+ x_t$ in (12); that is

$$y_t = \Delta_+ x_t = \mu \pi_{t-1}(0) + \beta \pi_{t-1}(\gamma - 1) + \Delta_+^{1-\delta} \varepsilon_t, \tag{14}$$

so that the observable process is $y_t = \beta_0 \pi_{t-1}(\gamma_0 - 1) + \Delta_+^{1-\delta_0} \varepsilon_t$ for $t \geq 2$ and $y_1 = \mu_0 + \beta_0 + \varepsilon_1$. We notice that μ_0 only affects one observation, y_1, so an apparently sensible approach could

be to "forget" about this effect and act as if the influence of μ_0 had been completely removed by differencing. In this setting, we will discuss several variations of differencing and adding back, and argue why these all fail for this model.

Supposed we observed

$$\widetilde{y}_t = \beta_0 \pi_{t-1}(\gamma_0 - 1) + \Delta_+^{1-\delta_0} \varepsilon_t, \quad t = 1, ..., T.$$

Then we could consider the loss function

$$\widetilde{Q}_T(\delta, \gamma, \beta) = T^{-1} \sum_{t=1}^{T} (\Delta_+^{\delta-1} \widetilde{y}_t - \beta \pi_{t-1}(\gamma - \delta))^2 \tag{15}$$

and derive the corresponding estimators. This approach would eliminate the drift parameter μ, and thus simplify the estimation problem. In particular, it would be relatively similar to that in Hualde and Nielsen (2020), with the only relevant difference of dealing with $\pi_{t-1}(\gamma-\delta)$ instead of $\Delta_+^\delta t^{\gamma-1}$ in the loss function. However, in practice we observe x_t, or equivalently y_t, for $t = 1, \ldots, T$, so that $\widetilde{Q}_T$ is infeasible because $\widetilde{y}_1$ is unobserved. Specifically, $\widetilde{y}_t = y_t$ for $t \geq 2$ but $\widetilde{y}_1 = \beta_0 + \varepsilon_1 \neq \mu_0 + \beta_0 + \varepsilon_1 = y_1$ if $\mu_0 \neq 0$. We now discuss three feasible alternatives.

First, inspired by $\widetilde{Q}_T$ in (15), we could ignore the presence of μ_0 in the single observation y_1 and set the loss function as

$$Q_{1T}(\delta, \gamma, \beta) = T^{-1} \sum_{t=1}^{T} (\Delta_+^{\delta-1} y_t - \beta \pi_{t-1}(\gamma - \delta))^2.$$

We note from (14) that $\Delta_+^{\delta-1} y_t = \mu_0 \pi_{t-1}(1 - \delta) + \beta_0 \pi_{t-1}(\gamma_0 - \delta) + \Delta_+^{\delta-\delta_0} \varepsilon_t$. Thus, comparing with the infeasible $\widetilde{Q}_T$, we find

$$Q_{1T}(\delta, \gamma, \beta) = T^{-1} \sum_{t=1}^{T} (\Delta_+^{\delta-1} \widetilde{y}_t - \beta \pi_{t-1}(\gamma - \delta) + \mu_0 \pi_{t-1}(1 - \delta))^2,$$

and the presence of the additional term $\mu_0 \pi_{t-1}(1 - \delta)$ in Q_{1T} (compared with $\widetilde{Q}_T$) is undesirable. For example, evaluated at the true values we find

$$Q_{1T}(\delta_0, \gamma_0, \beta_0) = T^{-1} \sum_{t=1}^{T} (\varepsilon_t + \mu_0 \pi_{t-1}(1 - \delta_0))^2,$$

where the contribution of the term $\mu_0 \pi_{t-1}(1 - \delta_0)$ is non-negligible, and in fact is dominant for $\delta_0 < 0$ if $\mu_0 \neq 0$.

Second, to eliminate the influence of the first observation, suppose we force a zero initial condition and consider

$$y_1^* = 0 \text{ and } y_t^* = \Delta x_t = \beta_0 \pi_{t-1}(\gamma_0 - 1) + \Delta_+^{1-\delta_0} \varepsilon_t \text{ for } t \geq 2.$$

We could then use the observed values $y_t^*, t = 1, \ldots, T$, in the estimation; that is, work with the loss function

$$Q_{2T}(\delta, \gamma, \beta) = T^{-1} \sum_{t=1}^{T} (\Delta_+^{\delta-1} y_t^* - \beta \pi_{t-1}(\gamma - \delta))^2.$$

Comparing again with the infeasible $\widetilde{Q}_T$, we find that

$$Q_{2T}(\delta, \gamma, \beta) = T^{-1} \sum_{t=1}^{T} (\Delta_+^{\delta-1} \widetilde{y}_t - \beta \pi_{t-1}(\gamma - \delta) - (\beta_0 + \varepsilon_1) \pi_{t-1}(1 - \delta))^2,$$

where the additional term $(\beta_0 + \varepsilon_1)\pi_{t-1}(1 - \delta)$ in Q_{2T} causes difficulties similar to the additional term in Q_{1T}. Furthermore, simply omitting the first observation and defining Q_{2T} by a summation over $t = 2, \ldots, T$ causes identical problems.

Third, instead of y_t in (14), suppose we consider the forward difference. That is, consider the observable

$$y_t^\dagger = x_{t+1} - x_t = \beta_0 \pi_t(\gamma_0 - 1) + \Delta_+^{1-\delta_0} \varepsilon_{t+1} \text{ for } t = 1, \ldots, T - 1$$

and the corresponding loss function

$$Q_{3T}(\delta, \gamma, \beta) = (T - 1)^{-1} \sum_{t=1}^{T-1} (\Delta_+^{\delta-1} y_t^\dagger - \beta \Delta_+^{\delta-1} \pi_t(\gamma - 1))^2.$$

This would appear to eliminate the influence of the first observation and μ_0. However, because $\Delta_+^{\delta-1} y_t^\dagger = \beta_0 \Delta_+^{\delta-1} \pi_t(\gamma_0 - 1) + \Delta_+^{\delta-\delta_0} \varepsilon_{t+1} - \pi_t(1 - \delta)\varepsilon_1$, the additional term $\pi_t(1 - \delta)\varepsilon_1$ again causes difficulties similar to the additional terms in Q_{1T} and Q_{2T}. Unreported Monte Carlo simulations confirm that estimators based on $Q_{iT}, i = 1, 2, 3$, are inconsistent for $\delta_0 < 0$.

The above discussion makes it clear that the "differencing and adding back" procedure cannot be used to simplify the estimation problem in our context. We therefore focus on the observed x_t for $t = 1, \ldots, T$. Motivated by the Gaussian log-likelihood and dealing with the general specification for ω given in (4), we consider the sum-of-squares loss function

$$L_T(\boldsymbol{\vartheta}, \boldsymbol{\phi}) = \frac{1}{T} \sum_{t=1}^{T} (\rho(L; \boldsymbol{\varphi})x_t(\delta) - \boldsymbol{\phi}' \boldsymbol{c}_{t-1}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}))^2. \tag{16}$$

Here we have defined $\rho(s; \boldsymbol{\varphi}) = \omega^{-1}(s; \boldsymbol{\varphi})$, see the discussion following Assumption A1, $\xi_t(d) = \Delta_+^d \xi_t$ for an arbitrary process ξ_t, and

$$\boldsymbol{c}_{t-1}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) = (c_{t-1}(1 - \delta, \boldsymbol{\varphi}), c_{t-1}(\gamma - \delta, \boldsymbol{\varphi}))'$$

with the convolution coefficient

$$c_{t-1}(d, \boldsymbol{\varphi}) = \rho(L; \boldsymbol{\varphi})\pi_{t-1}(d) = \sum_{j=0}^{t-1} \rho_j(\boldsymbol{\varphi})\pi_{t-j-1}(d). \tag{17}$$

Clearly, for a given $\boldsymbol{\vartheta}$, we can concentrate the loss function (16) with respect to $\boldsymbol{\phi}$, to obtain the concentrated loss function

$$R_T(\boldsymbol{\vartheta}) = \frac{1}{T} \sum_{t=1}^{T} \left(\rho(L; \boldsymbol{\varphi})x_t(\delta) - \widehat{\boldsymbol{\phi}}'(\boldsymbol{\vartheta})\boldsymbol{c}_{t-1}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \right)^2,$$

where

$$\widehat{\boldsymbol{\phi}}(\boldsymbol{\vartheta}) = \left(\sum_{t=1}^{T} \boldsymbol{c}_{t-1}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi})\boldsymbol{c}_{t-1}'(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \right)^{-1} \sum_{t=1}^{T} \boldsymbol{c}_{t-1}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi})\rho(L; \boldsymbol{\varphi})x_t(\delta). \tag{18}$$

Thus, letting the parameter space for $\boldsymbol{\vartheta}$ be denoted Ξ, which will be fully specified in Assumption A3 below, we propose the estimator

$$\widehat{\boldsymbol{\vartheta}} = \arg\min_{\boldsymbol{\vartheta} \in \Xi} R_T(\boldsymbol{\vartheta}), \tag{19}$$

along with $\widehat{\boldsymbol{\phi}} = \widehat{\boldsymbol{\phi}}(\widehat{\boldsymbol{\vartheta}})$ and $\widehat{\sigma}^2 = L_T(\widehat{\boldsymbol{\vartheta}}, \widehat{\boldsymbol{\phi}})$.

As proposed by Hualde and Robinson (2011), we call (19) the truncated sum-of-squares estimator, though it is also often called the nonlinear least squares or conditional sum-of-squares estimator. Because it is based on the Gaussian likelihood, it is expected to be asymptotically efficient under Gaussianity. For nonfractional ARMA models with a known integer-valued memory parameter, (19) was advocated by, e.g., Box and Jenkins (1971). For fractional time series, the estimator was suggested by Li and McLeod (1986) in stationary FARIMA models. The first rigorous asymptotic analyses of (19) with the memory parameter lying in an arbitrarily large compact interval were given by Hualde and Robinson (2011) and Nielsen (2015).

3 Main results

We first provide the assumptions needed for consistency of the estimator. Our conditions for the asymptotic analysis are nearly identical to those in Hualde and Robinson (2011) and Hualde and Nielsen (2020), with the only difference being in Assumption A3 below.

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 $\boldsymbol{\varphi} \in \Psi$, $\omega(e^{i\lambda}; \boldsymbol{\varphi})$ is differentiable in λ with derivative in Lip(ς) for $1/2 < \varsigma \leq 1$;

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

  (iv) for all $\boldsymbol{\varphi} \in \Psi$, $\omega_0(\boldsymbol{\varphi}) = 1$ and $|\omega(s; \boldsymbol{\varphi})| \neq 0, |s| \leq 1$.

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

A3. The parameter space for $\boldsymbol{\vartheta}$ is given by $\Xi = [\triangledown_1, \triangledown_2] \times \Psi \times \square$ and $\boldsymbol{\vartheta}_0 \in \Xi$. Here, $\triangledown_1$, $\triangledown_2$ are arbitrary real numbers such that $\triangledown_1 < \triangledown_2$, $\Psi \subseteq \mathbb{R}^p$ is compact and convex, and $\square = [\square_1, 1 - \kappa] \cup [1 + \kappa, \square_2]$ for arbitrary real numbers $\square_1, \square_2$, and an arbitrarily small real number $\kappa > 0$ such that $\square_1 < 1 - \kappa$ and $\square_2 > 1 + \kappa$. For $\boldsymbol{\phi}$ the parameter space is $\mathbb{R}^2$. Finally, if $\gamma_0 - 1/2 > \delta_0$, we also assume that $\beta_0 \neq 0$.

For a detailed discussion of A1 and A2 we refer to Hualde and Robinson (2011) and Hualde and Nielsen (2020). Note that writing $\omega^{-1}(s; \boldsymbol{\varphi}) = \rho(s; \boldsymbol{\varphi}) = \sum_{j=0}^{\infty} \rho_j(\boldsymbol{\varphi}) s^j$, Assumption A1 implies that $\rho_0(\boldsymbol{\varphi}) = 1$ for all $\boldsymbol{\varphi}$ and

$$\sup_{\boldsymbol{\varphi} \in \Psi} |\omega_j(\boldsymbol{\varphi})| = O(j^{-1-\varsigma}) \text{ as } j \to \infty, \tag{20}$$

$$\sup_{\boldsymbol{\varphi} \in \Psi} |\rho_j(\boldsymbol{\varphi})| = O(j^{-1-\varsigma}) \text{ as } j \to \infty, \tag{21}$$

$$\inf_{|s|=1, \boldsymbol{\varphi} \in \Psi} |\rho(s; \boldsymbol{\varphi})| > 0, \tag{22}$$

where $\varsigma > 1/2$ is given in A1(ii). A1 is easily satisfied for stationary and invertible ARMA models and for the exponential spectrum model of Bloomfield (1973). It is very similar to other conditions employed in asymptotic theory for the estimate $\widehat{\boldsymbol{\tau}}$, see 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 ω is a known parametric function. In fact ω satisfying A1 are invariably employed by practitioners.

Assumption A2 does not impose independence or identical distribution of ε_t, but requires conditional homoskedasticity. It is standard in the time series literature since Hannan (1973), although it may be quite strong for some empirical applications. We conjecture that this assumption could be relaxed to allow for both conditional and unconditional heteroskedasticiy following recent work by Cavaliere, Nielsen, and Taylor (2015, 2017, 2022). This would require replacing A2 by more complicated summability conditions on the cumulants of ε_t and would substantially complicate our proofs. Consequently, this seems beyond the scope of this paper.

Assumption A3 is very similar to A3 in Hualde and Nielsen (2020), but with two important differences. First, due to the inclusion of the extra drift term, we need $\gamma_0 \neq 1$ to guarantee identification of μ_0 and β_0 (see also Section 4.2). This condition, along with the need of dealing with compact parameter spaces in the consistency proof, leads to setting the parameter space for γ as $[\square_1, 1 - \kappa] \cup [1 + \kappa, \square_2]$. A similar requirement is imposed in the related setting of Robinson (2012). Second, translated to our notation, Hualde and Nielsen (2020) impose the condition that $\square_1 > 0$, which implies that just cases where $\gamma_0 > 0$ can be considered. This is due to the approximate nature of their model. Specifically, this additional condition helps to guarantee that their model is approximately closed under fractional differencing, i.e. to obtain that $\Delta_+^d t_+^c$ is approximately equal to a constant times t_+^{c-d}. In contrast, apart from the exclusion of the arbitrarily small open interval $(1 - \kappa, 1 + \kappa)$, we permit γ_0 to lie in an arbitrarily large set. Also, similarly to Hualde and Nielsen (2020), the parameter space for $\boldsymbol{\phi}$ is basically unrestricted, although we need the condition $\beta_0 \neq 0$ to guarantee the identification of γ_0 whenever this parameter can be consistently estimated.

As will be seen, when δ_0 is large, the stochastic signal dominates the deterministic trend. In particular, whenever $\delta_0 > \gamma_0 - 1/2$, γ_0 and β_0 cannot be consistently estimated, and if $\delta_0 > 1/2$ this problem also affects the estimation of the drift μ_0. On the other hand, when $\delta_0 < \gamma_0 - 1/2$ and/or $\delta_0 < 1/2$, at least part of the deterministic structure can be consistently estimated. Interestingly, for small values of δ_0 even very small and vanishing generalized trends (with small or negative γ_0) can be consistently estimated. In this sense, the value of δ_0 helps the identification of the deterministic trend.

Theorem 1. *Let x_t be generated by (1), (2), and (12) with true values $\boldsymbol{\vartheta}_0$ and $\boldsymbol{\phi}_0$, and let 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 $T \to \infty$.*

Because $\widehat{\boldsymbol{\phi}} = (\widehat{\mu}, \widehat{\beta})'$ has an explicit form, a prior consistency proof is not required, and therefore $\widehat{\boldsymbol{\phi}}$ is not included in Theorem 1. When applicable, we present directly the

asymptotic distribution of $\widehat{\phi}$ in Theorem 2(i) below. Specifically, it follows straightforwardly from Theorem 2 that $\widehat{\mu}$ is consistent if $\delta_0 < 1/2$ and $\widehat{\beta}$ is consistent when $\gamma_0 - 1/2 > \delta_0$.

Theorem 1(i) shows consistency of $\widehat{\boldsymbol{\vartheta}}$ when $\gamma_0 - 1/2 > \delta_0$, which we refer to as the strong deterministic trend case. On the other hand, Theorem 1(ii) shows consistency only of $\widehat{\boldsymbol{\tau}}$ when $\gamma_0 - 1/2 < \delta_0$, which is the weak deterministic trend case. In the latter case, γ_0 (and β_0) cannot possibly be consistently estimated. This is easily seen by considering for example $\delta_0 = 1$ (a random walk) in which case the parameters γ_0 and β_0 cannot be consistently estimated when $\gamma_0 < 3/2$ because the deterministic signal is drowned by the stochastic noise. This generalizes the well-known result that a level ($\gamma_0 = 1$) cannot be estimated consistently for a random walk, whereas a linear trend ($\gamma_0 = 2$) can be consistently estimated. Similarly, we note that, as will be discussed in the context of Theorem 2(i) below, the drift parameter μ_0 can be consistently estimated whenever $\delta_0 < 1/2$, i.e. when z_t is (asymptotically) stationary.

More generally, Theorem 1(ii) shows that, even in cases where the deterministic signal is not strong enough for consistent estimation of the deterministic structure, the parameter characterizing the stochastic component, $\boldsymbol{\tau}_0$, can nonetheless still be consistently estimated.

The proof of Theorem 1 is very challenging due to the non-uniform behaviour of the loss function over a large admissible parameter set. In a much simpler setting with absence of any deterministic component, this problem was acknowledged and solved by Hualde and Robinson (2011) and Nielsen (2015), where the difficulty arose due to the nonstationary/stationary behaviour of fractional differences of the observed process. Our proof strategy is similar to that in Hualde and Nielsen (2020) in that it takes advantage of the competition between the deterministic and stochastic components, although this is now more challenging because of the arbitrarily large parameter space for γ, and in particular because of the presence of a second deterministic term. Specifically, the latter complicates matters substantially due to the difficulties outlined in Section 2 and because dealing with the relative strengths of stochastic and deterministic terms is now more involved.

For the asymptotic distribution theory we define $\boldsymbol{b}_j(\boldsymbol{\varphi}_0) = \sum_{k=0}^{j-1} \omega_k(\boldsymbol{\varphi}_0) \partial \rho_{j-k}(\boldsymbol{\varphi}_0)/\partial \boldsymbol{\varphi}$ and the matrix

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

which is the Fisher information for the parameter $\boldsymbol{\tau}$ under Gaussianity; see Dahlhaus (1989) and Hualde and Robinson (2011). Also, we require an additional regularity condition.

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

(ii) for all λ, $\omega(e^{i\lambda}; \boldsymbol{\varphi})$ is thrice continuously differentiable in $\boldsymbol{\varphi}$ on a closed neighbourhood $\mathcal{N}_\epsilon(\boldsymbol{\varphi}_0)$ 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 λ with derivative in $\text{Lip}(\varsigma)$ for $1/2 < \varsigma \leq 1$;

(iii) the matrix $\boldsymbol{A}$ is nonsingular.

Assumption A4 is identical to A4 in Hualde and Nielsen (2020). It is slightly stronger than

A3 in Hualde and Robinson (2011). This strengthening seems necessary to obtain the bounds

$$\sup_{\varphi\in\mathcal{N}_\epsilon(\varphi_0)}\left|\frac{\partial\rho_j(\varphi)}{\partial\varphi_i}\right|=O(j^{-1-\varsigma}),\quad \sup_{\varphi\in\mathcal{N}_\epsilon(\varphi_0)}\left|\frac{\partial^2\rho_j(\varphi)}{\partial\varphi_i\partial\varphi_l}\right|=O(j^{-1-\varsigma}),$$

$$\sup_{\varphi\in\mathcal{N}_\epsilon(\varphi_0)}\left|\frac{\partial^3\rho_j(\varphi)}{\partial\varphi_i\partial\varphi_l\partial\varphi_k}\right|=O(j^{-1-\varsigma}), \tag{23}$$

where $\varsigma>1/2$ is given in A4(ii) and φ_i denotes the i-th element of φ; see Hualde and Nielsen (2020) for details. Again A4 is easily satisfied for ARMA models or the Bloomfield (1973) spectral model. For the latter model, the analytical formulas for $\boldsymbol{A}$, and hence for the asymptotic variance matrix, simplify neatly; see Robinson (1994). In practical implementations, though, numerical derivatives of the objective function will typically be used.

To describe the asymptotic distribution, we introduce additional notation. Let $\boldsymbol{I}_q$ and $\boldsymbol{0}_q$ denote the q-dimensional identity matrix and a q-vector of zeros, respectively, and define

$$\boldsymbol{P}_{\mu,\beta,T}=\mathrm{diag}(\boldsymbol{I}_{p+1},T^{\delta_0-\gamma_0+1},T^{\delta_0},T^{\delta_0-\gamma_0+1}\log T)\text{ and }\boldsymbol{W}_{\mu,\beta}=\begin{pmatrix}\sigma_0^2\boldsymbol{A} & \boldsymbol{0}_{p+1}\,\boldsymbol{0}_{p+1}\\\boldsymbol{0}'_{p+1} & \\\boldsymbol{0}'_{p+1} & \boldsymbol{V}_{\mu,\beta}\end{pmatrix},$$

$$\tag{24}$$

$$\boldsymbol{P}_{\beta,T}=\mathrm{diag}(\boldsymbol{I}_{p+1},T^{\delta_0-\gamma_0+1},T^{\delta_0-\gamma_0+1}\log T)\text{ and }\boldsymbol{W}_\beta=\begin{pmatrix}\sigma_0^2\boldsymbol{A} & \boldsymbol{0}_{p+1}\\\boldsymbol{0}'_{p+1} & v_\beta\end{pmatrix}, \tag{25}$$

where

$$\boldsymbol{V}_{\mu,\beta}=\begin{pmatrix}\frac{(\gamma_0-1)^2}{(\gamma_0-2\delta_0)^2}v_\beta & 0\\0 & \frac{\rho^2(1;\varphi_0)(\gamma_0-1)^2}{\Gamma^2(1-\delta_0)(1-2\delta_0)(\gamma_0-2\delta_0)^2}\end{pmatrix}\text{ and }v_\beta=\frac{\beta_0^2\rho^2(1;\varphi_0)}{\Gamma^2(\gamma_0-\delta_0)(2(\gamma_0-\delta_0)-1)^3}.$$

Theorem 2. *Let x_t be generated by (1), (2), and (12) with true values $\boldsymbol{\vartheta}_0$ and $\boldsymbol{\phi}_0$, and let Assumptions A1–A4 hold. Then, as $T\to\infty$:*

(i.a) If $\gamma_0-1/2>\delta_0$ and $\delta_0<1/2$,

$$T^{1/2}\boldsymbol{P}_{\mu,\beta,T}^{-1}\begin{pmatrix}\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\\\widehat{\mu}-\mu_0\\\widehat{\beta}-\beta_0\end{pmatrix}\to_d\begin{pmatrix}\boldsymbol{I}_{p+2} & \boldsymbol{0}_{p+2}\\\boldsymbol{0}'_{p+2} & 1\\\boldsymbol{0}'_{p+1} & -\beta_0 & 0\end{pmatrix}\boldsymbol{N}_{\mu,\beta}, \tag{26}$$

where $\boldsymbol{N}_{\mu,\beta}$ is a random variable distributed as $N(\boldsymbol{0}_{p+3},\sigma_0^2\boldsymbol{W}_{\mu,\beta}^{-1})$.

(i.b) If $\gamma_0-1/2>\delta_0$ and $\delta_0>1/2$,

$$T^{1/2}\boldsymbol{P}_{\beta,T}^{-1}\begin{pmatrix}\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\\\widehat{\beta}-\beta_0\end{pmatrix}\to_d\begin{pmatrix}\boldsymbol{I}_{p+2}\\\boldsymbol{0}'_{p+1} & -\beta_0\end{pmatrix}\boldsymbol{N}_\beta, \tag{27}$$

where $\boldsymbol{N}_\beta$ is a random variable distributed as $N(\boldsymbol{0}_{p+2},\sigma_0^2\boldsymbol{W}_\beta^{-1})$.

(ii) If $\gamma_0-1/2<\delta_0$,

$$T^{1/2}(\widehat{\boldsymbol{\tau}}-\boldsymbol{\tau}_0)\to_d N(\boldsymbol{0}_{p+1},\boldsymbol{A}^{-1}). \tag{28}$$

If, in addition, $\delta_0<1/2$, then for any $\epsilon>0$,

$$\widehat{\mu}-\mu_0=o_p\left(T^{\delta_0-1/2+\epsilon}\right). \tag{29}$$

The asymptotic distribution results in Theorem 2 are divided into two main cases. In Theorem 2(i) we first present the result for the strong deterministic trend case, $\gamma_0 - 1/2 > \delta_0$. Here, the deterministic signal is sufficiently strong, relative to that of the stochastic component, that we can prove joint asymptotic normality for the estimators $\widehat{\boldsymbol{\vartheta}}$ and $\widehat{\beta}$. Within this case, we distinguish between $\delta_0 < 1/2$, where we can also include $\widehat{\mu}$, and $\delta_0 > 1/2$ where we cannot include $\widehat{\mu}$. Whether $\widehat{\mu}$ can be included or not, $\widehat{\boldsymbol{\vartheta}}$ and $\widehat{\beta}$ retain identical properties.

In Theorem 2(ii) we present the result for the weak deterministic trend case, $\gamma_0 - 1/2 < \delta_0$. In this case, we can obtain the asymptotic distribution for the estimator of the stochastic component $\boldsymbol{\tau}_0$ only, but when $\delta_0 < 1/2$ we also prove consistency of $\widehat{\mu}$ with rate $T^{1/2 - \delta_0 + \epsilon}$ for any arbitrarily small $\epsilon > 0$. However, the estimator $\widehat{\mu}$ is a complicated function of $\widehat{\gamma}$, whose behavior is unknown in this case and indeed is not even consistent; see the discussion after Theorem 1. Thus, deriving an asymptotic distribution result for $\widehat{\mu}$ in case (ii) with $\delta_0 < 1/2$ does not seem to be possible.

Theorem 2(ii) shows that, even in cases where the deterministic signal is not strong enough for consistent estimation of γ_0 and β_0 (and possibly also μ_0), the estimator of the parameter characterizing the stochastic component, $\widehat{\boldsymbol{\tau}}$, has exactly the same limiting properties as in the strong deterministic trend case in Theorem 2(i.a). That is, the asymptotic distribution result for $\widehat{\boldsymbol{\tau}}$ in Theorem 2 is unaffected by the relative strengths of the stochastic and deterministic components. In particular, even when γ_0, β_0, and μ_0 cannot be consistently estimated, the asymptotic distribution of $\widehat{\boldsymbol{\tau}}$ is unaffected by their presence.

It is noteworthy that the asymptotic distribution of $\widehat{\boldsymbol{\tau}}$ is unaffected by the presence of the deterministic component in (12), and $\widehat{\boldsymbol{\tau}}$ has the same asymptotic distribution as in, e.g., Theorem 2.2 of Hualde and Robinson (2011). Specifically, the variance $\boldsymbol{A}^{-1}$ in the asymptotic distribution of $\widehat{\boldsymbol{\tau}}$ in (26), (27), and (28) is equal to the inverse Fisher information under Gaussianity; see Dahlhaus (1989). Because the estimate $\widehat{\boldsymbol{\tau}}$ is also asymptotically independent of the remaining parameter estimates, it therefore follows that $\widehat{\boldsymbol{\tau}}$ is asymptotically efficient under the additional assumption of Gaussianity, and this occurs regardless of the relative strength of the deterministic and stochastic components.

More generally, Theorem 2 makes it possible to conduct inference on the model parameters, with the caveat that the joint asymptotic distribution of $\widehat{\boldsymbol{\vartheta}}$, $\widehat{\mu}$, and $\widehat{\beta}$ given in (26) as well as that of $\widehat{\boldsymbol{\vartheta}}$ and $\widehat{\beta}$ given in (27) are both singular, which makes testing of joint hypotheses on $\boldsymbol{\vartheta}_0$ and β_0 impossible. However, separate inference can straightforwardly be conducted on $\boldsymbol{\vartheta}_0$ and β_0. For example, it is straightforward given (26) or (27) to construct confidence intervals and/or to test hypotheses such as $\gamma_0 = 2$ (deterministic trend in x_t is linear) or $\delta_0 = 1$ (stochastic component is of the random walk-type).

To conduct inference on the model parameters using Theorem 2 a consistent estimate of σ_0^2 is needed. To this end, consistency of the estimator $\widehat{\sigma}^2 = L_T(\widehat{\boldsymbol{\vartheta}}, \widehat{\boldsymbol{\phi}})$, see (16), is straightforwardly obtained using the methods in the proofs of Theorems 1 and 2.

Specifically, in view of the above comments, inference on the parameter $\boldsymbol{\tau}_0$ characterizing the stochastic component can be conducted by means of likelihood ratio tests or Wald/t tests. The former do not require estimation of the variance, whereas for the latter a consistent estimate of the variance of $\widehat{\boldsymbol{\tau}}$ can be obtained by numerical evaluation of the Hessian matrix. Because the (marginal) asymptotic distribution of $\widehat{\boldsymbol{\tau}}$ is the same across the different cases in Theorem 2, and because the Hessian matrix is asymptotically block-diagonal in each case, it

is a straightforward consequence of Theorem 2 that such likelihood ratio or Wald tests are asymptotically χ^2-distributed and t-tests are asymptotically standard normally distributed. Nonetheless, it may sometimes be of interest to determine which of the cases in Theorem 2 is relevant in a given situation. Of course there are many ways of doing so, and in Section 4.2 we consider a stepwise testing procedure.

We notice from (26) and (27) in Theorem 2(i) that $\widehat{\gamma}$ is $T^{\gamma_0-\delta_0-1/2}$-consistent whereas $\widehat{\beta}$ is only $T^{\gamma_0-\delta_0-1/2}/\log T$-consistent. In fact, it can be shown that if γ_0 were known, then the least squares regression estimator of β_0 would also be $T^{\gamma_0-\delta_0-1/2}$-consistent. Thus, there is a (small) rate-of-convergence loss from having to estimate the power law parameter γ_0.

Finally, we also notice from (26) that, in the case where they are both consistently estimable, $\widehat{\beta}$ and $\widehat{\mu}$ are asymptotically independent. This seemingly contradicts Robinson (2012), who considers the weakly dependent case ($\delta_0 = 0$ known) with unknown power law parameters. In contrast, we consider the power law parameter corresponding to β to be unknown (γ_0), while the power law parameter corresponding to μ is known (and equal to one). This fundamental difference ensures that, compared with a situation in which the power law parameter corresponding to μ were unknown, $\widehat{\mu}$ converges at a rate $\log T$ faster, and this in turn guarantees asymptotic independence from the coefficient estimate $\widehat{\beta}$. The technical justification for this asymptotic independence result is given in (138).

4 Monte Carlo simulations

4.1 Parameter estimation

We investigate the finite-sample performance of our estimators of γ_0 and δ_0 by means of a simple Monte Carlo experiment. We generate the observable series $x_t, t = 1, ..., T$, from (12) with $u_t = \varepsilon_t$ being an independent $N(0,1)$ sequence and $T = 64, 128, 256$. We fix $\mu_0 = 1$ (results are virtually unaffected by the value of μ_0, so there is no loss of generality here) and consider $\delta_0 = 0, 1, 2$, $\beta_0 = -5, 1, 10$, and 14 different values for γ_0 given by $\gamma_0 - \delta_0 = 1.2, 1.1, \ldots$, such that $\gamma_0 \neq 1$. Based on (12) and (13), it is tempting to think that results should be invariant to δ_0 as long as $\gamma_0 - \delta_0$ is kept fixed, but that is not the case. In fact, for fixed $\gamma_0 - \delta_0$, results depend on δ_0 in a subtle way, because whenever γ_0 is close to 1, results are worse due to near-multicollinearity of $\pi_t(1)$ and $\pi_t(\gamma_0)$; see (12). We computed $\widehat{\delta}, \widehat{\gamma}$ using the optimizing intervals $\delta \in [\delta_0 - 5, \delta_0 + 5]$ and $\gamma \in [-4, 0.99] \cup [1.01, 6]$, and we report Monte Carlo bias and standard deviation (SD) across 10,000 replications.

Results for Monte Carlo bias of $\widehat{\delta}$ are presented in Table 1. Here, the performance of $\widehat{\delta}$ reflects the asymptotic theory in Theorems 1 and 2. First, for all $(\delta_0, \gamma_0, \beta_0, T)$ combinations the bias is negative and it clearly decreases in absolute value as T increases, even for the boundary case $\gamma_0 - \delta_0 = 1/2$, which is not covered by our theory. Second, when the deterministic signal gets stronger (so $\gamma_0 - \delta_0$ is higher) results are slightly worse. Third, for fixed $\gamma_0 - \delta_0$, results are better for larger δ_0 because then γ_0 is further away from 1, so that the asymptotic multicollinearity problem is less noticeable. Finally, the results in this table are basically unaffected by the value of β_0.

Monte Carlo SD results for $\widehat{\delta}$ are given in Table 2. These again reflect the asymptotic theory and complement nicely the results in Table 1. As expected, the SD results improve as T increases and are largely unaffected by the values of γ_0, δ_0, and β_0.

Next, results for Monte Carlo bias of $\widehat{\gamma}$ are presented in Table 3. As expected from the

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

$\gamma_0 - \delta_0 \backslash T$	$\beta_0 = -5$			$\beta_0 = 1$			$\beta_0 = 10$		
	64	128	256	64	128	256	64	128	256
Panel A: $\delta_0 = 0$									
1.2	−0.159	−0.076	−0.039	−0.149	−0.076	−0.039	−0.160	−0.076	−0.039
1.1	−0.150	−0.073	−0.038	−0.129	−0.066	−0.035	−0.154	−0.074	−0.038
0.9	−0.151	−0.071	−0.037	−0.152	−0.073	−0.031	−0.151	−0.070	−0.036
0.8	−0.143	−0.068	−0.035	−0.147	−0.072	−0.037	−0.144	−0.068	−0.035
0.7	−0.143	−0.066	−0.025	−0.145	−0.070	−0.036	−0.138	−0.065	−0.029
0.6	−0.134	−0.063	−0.032	−0.142	−0.069	−0.035	−0.132	−0.063	−0.032
0.5	−0.128	−0.060	−0.031	−0.140	−0.068	−0.035	−0.127	−0.060	−0.031
0.4	−0.122	−0.057	−0.030	−0.138	−0.067	−0.034	−0.123	−0.057	−0.029
0.3	−0.116	−0.054	−0.028	−0.136	−0.066	−0.034	−0.115	−0.055	−0.028
0.2	−0.110	−0.052	−0.026	−0.134	−0.065	−0.034	−0.110	−0.052	−0.026
0.1	−0.104	−0.049	−0.025	−0.132	−0.064	−0.033	−0.104	−0.049	−0.024
0.0	−0.098	−0.046	−0.024	−0.131	−0.063	−0.033	−0.098	−0.046	−0.023
−0.1	−0.093	−0.044	−0.023	−0.129	−0.062	−0.032	−0.092	−0.043	−0.022
−0.2	−0.087	−0.042	−0.022	−0.128	−0.063	−0.032	−0.087	−0.041	−0.021
Panel B: $\delta_0 = 1$									
1.2	−0.129	−0.061	−0.032	−0.128	−0.062	−0.031	−0.130	−0.061	−0.031
1.1	−0.127	−0.060	−0.031	−0.125	−0.061	−0.031	−0.127	−0.060	−0.031
1.0	−0.124	−0.059	−0.030	−0.122	−0.060	−0.031	−0.125	−0.059	−0.030
0.9	−0.120	−0.057	−0.030	−0.116	−0.058	−0.029	−0.122	−0.057	−0.029
0.8	−0.117	−0.055	−0.029	−0.110	−0.055	−0.028	−0.118	−0.056	−0.028
0.7	−0.113	−0.053	−0.027	−0.101	−0.050	−0.026	−0.114	−0.053	−0.027
0.6	−0.108	−0.051	−0.026	−0.092	−0.046	−0.023	−0.110	−0.051	−0.026
0.5	−0.088	−0.042	−0.022	−0.098	−0.047	−0.021	−0.049	−0.036	−0.022
0.4	−0.087	−0.038	−0.018	−0.097	−0.047	−0.019	−0.055	−0.028	−0.015
0.3	−0.085	−0.039	−0.020	−0.075	−0.036	−0.018	−0.093	−0.043	−0.021
0.2	−0.074	−0.035	−0.017	−0.072	−0.034	−0.017	−0.083	−0.038	−0.019
0.1	−0.068	−0.032	−0.016	−0.068	−0.032	−0.016	−0.072	−0.033	−0.016
−0.1	−0.095	−0.046	−0.023	−0.097	−0.047	−0.025	−0.091	−0.042	−0.021
−0.2	−0.089	−0.042	−0.022	−0.096	−0.047	−0.024	−0.076	−0.034	−0.017
Panel C: $\delta_0 = 2$									
1.2	−0.114	−0.056	−0.030	−0.115	−0.057	−0.030	−0.115	−0.056	−0.030
1.1	−0.111	−0.056	−0.029	−0.112	−0.056	−0.029	−0.112	−0.055	−0.029
1.0	−0.108	−0.054	−0.028	−0.109	−0.054	−0.028	−0.109	−0.053	−0.028
0.9	−0.104	−0.052	−0.027	−0.103	−0.052	−0.027	−0.105	−0.051	−0.027
0.8	−0.099	−0.049	−0.026	−0.095	−0.049	−0.026	−0.100	−0.049	−0.026
0.7	−0.097	−0.046	−0.024	−0.087	−0.044	−0.023	−0.095	−0.047	−0.024
0.6	−0.091	−0.043	−0.023	−0.080	−0.039	−0.020	−0.089	−0.044	−0.023
0.5	−0.085	−0.040	−0.021	−0.076	−0.037	−0.018	−0.083	−0.041	−0.021
0.4	−0.079	−0.032	−0.015	−0.074	−0.036	−0.016	−0.078	−0.037	−0.019
0.3	−0.069	−0.033	−0.017	−0.060	−0.029	−0.014	−0.063	−0.034	−0.017
0.2	−0.063	−0.029	−0.015	−0.054	−0.026	−0.013	−0.066	−0.032	−0.015
0.1	−0.056	−0.025	−0.013	−0.049	−0.023	−0.012	−0.058	−0.028	−0.013
0.0	−0.050	−0.022	−0.011	−0.044	−0.021	−0.011	−0.051	−0.025	−0.011
−0.1	−0.044	−0.020	−0.010	−0.040	−0.019	−0.010	−0.045	−0.022	−0.010

Note: Based on 10,000 Monte Carlo replications.

Table 2: Monte Carlo standard deviation of $\widehat{\delta}$

$\gamma_0 - \delta_0 \backslash T$	$\beta_0 = -5$			$\beta_0 = 1$			$\beta_0 = 10$		
	64	128	256	64	128	256	64	128	256
Panel A: $\delta_0 = 0$									
1.2	0.158	0.092	0.059	0.162	0.092	0.059	0.158	0.092	0.058
1.1	0.157	0.091	0.059	0.160	0.094	0.060	0.158	0.092	0.058
0.9	0.157	0.091	0.059	0.153	0.091	0.059	0.156	0.092	0.058
0.8	0.157	0.091	0.058	0.156	0.091	0.059	0.158	0.091	0.058
0.7	0.157	0.090	0.059	0.156	0.092	0.059	0.155	0.090	0.061
0.6	0.154	0.089	0.058	0.156	0.092	0.059	0.154	0.090	0.058
0.5	0.153	0.089	0.058	0.156	0.092	0.059	0.153	0.090	0.057
0.4	0.152	0.088	0.058	0.156	0.092	0.059	0.150	0.090	0.057
0.3	0.150	0.088	0.057	0.156	0.091	0.059	0.150	0.088	0.057
0.2	0.149	0.087	0.057	0.156	0.092	0.059	0.149	0.088	0.056
0.1	0.148	0.087	0.056	0.156	0.092	0.059	0.147	0.087	0.056
0.0	0.146	0.086	0.056	0.157	0.092	0.059	0.145	0.086	0.056
−0.1	0.144	0.086	0.056	0.157	0.092	0.059	0.143	0.085	0.055
−0.2	0.142	0.085	0.055	0.157	0.092	0.059	0.140	0.085	0.055
Panel B: $\delta_0 = 1$									
1.2	0.149	0.089	0.058	0.151	0.088	0.057	0.150	0.089	0.057
1.1	0.149	0.088	0.058	0.151	0.088	0.057	0.150	0.089	0.057
1.0	0.149	0.088	0.058	0.153	0.088	0.056	0.149	0.089	0.057
0.9	0.150	0.089	0.057	0.156	0.089	0.057	0.149	0.089	0.057
0.8	0.147	0.088	0.057	0.159	0.091	0.057	0.148	0.089	0.057
0.7	0.147	0.088	0.057	0.160	0.092	0.058	0.147	0.089	0.057
0.6	0.147	0.088	0.057	0.158	0.091	0.059	0.147	0.088	0.057
0.5	0.175	0.099	0.061	0.157	0.091	0.058	0.215	0.113	0.061
0.4	0.157	0.093	0.060	0.149	0.088	0.058	0.183	0.107	0.066
0.3	0.149	0.089	0.057	0.153	0.089	0.057	0.146	0.088	0.056
0.2	0.149	0.088	0.057	0.153	0.089	0.057	0.146	0.088	0.056
0.1	0.150	0.089	0.057	0.151	0.088	0.057	0.147	0.089	0.056
−0.1	0.146	0.088	0.057	0.149	0.088	0.057	0.147	0.088	0.057
−0.2	0.145	0.088	0.057	0.149	0.088	0.057	0.143	0.086	0.055
Panel C: $\delta_0 = 2$									
1.2	0.142	0.088	0.056	0.144	0.088	0.056	0.144	0.088	0.057
1.1	0.143	0.088	0.057	0.145	0.088	0.056	0.145	0.088	0.056
1.0	0.142	0.088	0.057	0.147	0.088	0.057	0.145	0.088	0.056
0.9	0.142	0.088	0.057	0.149	0.089	0.057	0.145	0.088	0.056
0.8	0.142	0.087	0.056	0.152	0.089	0.057	0.145	0.087	0.057
0.7	0.144	0.087	0.056	0.153	0.092	0.058	0.144	0.086	0.056
0.6	0.143	0.087	0.056	0.150	0.092	0.058	0.143	0.085	0.056
0.5	0.142	0.086	0.056	0.146	0.090	0.057	0.141	0.085	0.056
0.4	0.140	0.092	0.060	0.144	0.089	0.057	0.140	0.084	0.056
0.3	0.138	0.084	0.055	0.146	0.088	0.056	0.152	0.084	0.055
0.2	0.138	0.083	0.055	0.141	0.087	0.056	0.137	0.083	0.055
0.1	0.135	0.082	0.054	0.139	0.085	0.055	0.134	0.082	0.054
0.0	0.133	0.081	0.054	0.139	0.084	0.055	0.131	0.081	0.054
−0.1	0.131	0.080	0.053	0.133	0.083	0.054	0.128	0.080	0.053

Note: Based on 10,000 Monte Carlo replications.

Table 3: Monte Carlo bias of $\widehat{\gamma}$

	$\beta_0 = -5$			$\beta_0 = 1$			$\beta_0 = 10$		
$\gamma_0 - \delta_0 \backslash T$	64	128	256	64	128	256	64	128	256
Panel A: $\delta_0 = 0$									
1.2	0.007	0.003	0.001	0.237	0.125	0.039	0.002	0.001	0.000
1.1	0.083	0.033	0.011	0.065	0.104	0.204	0.020	0.005	0.003
0.9	0.239	0.176	0.116	0.194	0.011	−0.065	0.075	0.053	0.042
0.8	0.069	0.026	0.011	0.280	0.078	0.082	0.010	0.006	0.002
0.7	0.031	0.013	0.087	0.340	0.161	0.138	0.005	0.003	0.016
0.6	0.022	0.008	0.005	0.398	0.211	0.157	0.003	0.002	0.001
0.5	0.016	0.005	0.002	0.454	0.311	0.178	0.002	0.001	0.000
0.4	0.011	0.001	0.001	0.533	0.358	0.225	0.002	0.001	0.000
0.3	0.005	−0.003	−0.004	0.608	0.409	0.300	0.000	0.000	−0.002
0.2	0.001	−0.007	−0.009	0.652	0.473	0.360	−0.001	−0.001	−0.003
0.1	−0.003	−0.009	−0.013	0.742	0.560	0.460	−0.002	−0.002	−0.004
0.0	−0.006	−0.011	−0.015	0.833	0.628	0.560	−0.003	−0.003	−0.005
−0.1	−0.008	−0.011	−0.014	0.904	0.731	0.630	−0.003	−0.003	−0.005
−0.2	−0.009	−0.012	−0.013	0.941	0.787	0.689	−0.004	−0.003	−0.005
Panel B: $\delta_0 = 1$									
1.2	0.000	0.000	0.000	0.005	0.001	0.001	0.000	0.000	0.000
1.1	0.000	0.000	0.000	0.006	0.003	0.001	0.000	0.000	0.000
1.0	0.000	0.000	0.000	0.001	0.006	0.005	0.000	0.000	0.000
0.9	0.000	0.000	0.000	−0.023	0.005	0.007	0.000	0.000	0.000
0.8	−0.001	0.000	0.001	−0.067	0.011	0.019	0.000	0.000	0.000
0.7	−0.014	0.001	0.002	−0.171	−0.079	0.008	−0.001	0.000	0.001
0.6	−0.051	−0.009	0.001	−0.263	−0.268	−0.196	−0.107	−0.012	0.001
0.5	−0.132	−0.056	−0.029	−0.294	−0.354	−0.370	−0.145	−0.041	−0.009
0.4	−0.213	−0.187	−0.168	−0.229	−0.334	−0.373	−0.185	−0.109	−0.060
0.3	−0.255	−0.276	−0.303	−0.134	−0.239	−0.316	−0.249	−0.201	−0.181
0.2	−0.136	−0.186	−0.272	−0.030	−0.139	−0.190	−0.328	−0.291	−0.283
0.1	0.031	−0.028	−0.113	0.094	−0.038	−0.099	−0.143	−0.167	−0.194
−0.1	0.196	0.146	0.036	0.280	0.123	0.119	0.067	−0.028	−0.027
−0.2	0.173	0.130	0.040	0.361	0.221	0.203	−0.097	−0.143	−0.109
Panel C: $\delta_0 = 2$									
1.2	0.000	0.000	0.000	0.002	0.002	0.001	0.000	0.000	0.000
1.1	0.001	0.000	0.000	0.002	0.004	0.001	0.000	0.000	0.000
1.0	0.001	0.000	0.000	−0.005	0.008	0.003	0.000	0.000	0.000
0.9	0.001	0.001	0.000	−0.032	0.009	0.008	0.000	0.000	0.000
0.8	0.002	0.001	0.000	−0.107	−0.015	0.010	0.001	0.000	0.000
0.7	0.002	0.001	0.001	−0.288	−0.138	−0.049	0.001	0.000	0.000
0.6	0.002	0.001	0.001	−0.510	−0.438	−0.324	0.001	0.000	0.000
0.5	0.000	0.000	0.000	−0.641	−0.692	−0.643	0.001	0.000	0.000
0.4	−0.006	−0.043	−0.054	−0.728	−0.802	−0.794	0.000	−0.001	−0.002
0.3	−0.024	−0.021	−0.019	−0.704	−0.811	−0.833	−0.034	−0.002	−0.002
0.2	−0.048	−0.046	−0.046	−0.667	−0.760	−0.792	−0.004	−0.006	−0.004
0.1	−0.087	−0.086	−0.080	−0.594	−0.695	−0.709	−0.007	−0.011	−0.008
0.0	−0.133	−0.141	−0.137	−0.506	−0.605	−0.614	−0.011	−0.016	−0.013
−0.1	−0.181	−0.195	−0.199	−0.403	−0.519	−0.528	−0.017	−0.022	−0.018

Note: Based on 10,000 Monte Carlo replications.

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

$\gamma_0 - \delta_0 \backslash T$	$\beta_0 = -5$			$\beta_0 = 1$			$\beta_0 = 10$		
	64	128	256	64	128	256	64	128	256
Panel A: $\delta_0 = 0$									
1.2	0.121	0.069	0.041	1.528	0.640	0.235	0.060	0.034	0.020
1.1	0.458	0.183	0.107	2.765	2.403	1.789	0.147	0.087	0.055
0.9	1.399	0.793	0.373	2.989	2.927	2.798	0.281	0.145	0.101
0.8	0.463	0.242	0.138	2.852	2.732	2.601	0.130	0.087	0.063
0.7	0.258	0.168	0.184	2.793	2.696	2.602	0.099	0.070	0.084
0.6	0.228	0.146	0.117	2.778	2.713	2.689	0.087	0.065	0.051
0.5	0.225	0.147	0.109	2.795	2.788	2.769	0.082	0.064	0.052
0.4	0.209	0.156	0.129	2.817	2.821	2.822	0.080	0.065	0.055
0.3	0.218	0.173	0.155	2.837	2.818	2.859	0.081	0.069	0.060
0.2	0.220	0.189	0.163	2.847	2.855	2.853	0.084	0.074	0.068
0.1	0.225	0.213	0.185	2.839	2.848	2.848	0.089	0.081	0.076
0.0	0.251	0.242	0.209	2.837	2.821	2.826	0.094	0.088	0.085
−0.1	0.270	0.262	0.227	2.832	2.807	2.802	0.100	0.095	0.093
−0.2	0.277	0.261	0.249	2.816	2.788	2.777	0.105	0.101	0.100
Panel B: $\delta_0 = 1$									
1.2	0.018	0.011	0.006	0.096	0.054	0.032	0.009	0.005	0.003
1.1	0.023	0.015	0.009	0.179	0.075	0.047	0.012	0.007	0.005
1.0	0.030	0.020	0.013	0.317	0.107	0.069	0.015	0.010	0.007
0.9	0.039	0.027	0.019	0.563	0.254	0.101	0.020	0.014	0.010
0.8	0.107	0.038	0.028	1.145	0.676	0.327	0.026	0.018	0.014
0.7	0.294	0.055	0.040	1.826	1.514	1.140	0.070	0.064	0.020
0.6	0.552	0.254	0.115	2.362	2.310	2.161	0.555	0.183	0.030
0.5	0.925	0.682	0.515	2.757	2.756	2.738	0.674	0.360	0.182
0.3	2.025	2.020	2.035	3.037	3.052	3.053	1.164	1.038	0.976
0.2	2.602	2.625	2.689	3.083	3.097	3.087	1.770	1.753	1.788
0.1	2.984	2.993	3.042	3.097	3.111	3.108	2.672	2.695	2.721
−0.1	3.002	3.017	3.050	3.098	3.112	3.118	2.725	2.753	2.745
−0.2	2.756	2.771	2.803	3.085	3.106	3.103	1.950	1.946	1.932
Panel C: $\delta_0 = 2$									
1.2	0.017	0.010	0.006	0.123	0.052	0.032	0.008	0.005	0.003
1.1	0.021	0.014	0.009	0.231	0.070	0.046	0.011	0.007	0.005
1.0	0.027	0.019	0.013	0.363	0.097	0.065	0.013	0.009	0.006
0.9	0.032	0.024	0.017	0.596	0.241	0.093	0.016	0.012	0.009
0.8	0.039	0.030	0.023	1.037	0.568	0.237	0.019	0.015	0.012
0.7	0.048	0.037	0.030	1.607	1.226	0.889	0.023	0.019	0.015
0.6	0.059	0.047	0.039	2.167	2.033	1.877	0.028	0.023	0.019
0.5	0.092	0.077	0.055	2.568	2.577	2.546	0.035	0.030	0.026
0.4	0.149	0.325	0.329	2.804	2.866	2.866	0.045	0.040	0.062
0.3	0.247	0.226	0.169	2.916	2.981	3.002	0.216	0.054	0.051
0.2	0.376	0.350	0.303	2.984	3.028	3.062	0.083	0.076	0.073
0.1	0.527	0.507	0.451	3.025	3.063	3.077	0.111	0.105	0.102
0.0	0.675	0.695	0.643	3.053	3.072	3.093	0.145	0.140	0.137
−0.1	0.806	0.842	0.822	3.060	3.084	3.096	0.186	0.185	0.181

Note: Based on 10,000 Monte Carlo replications.

asymptotic theory in Theorems 1 and 2, the behaviour of $\widehat{\gamma}$ is qualitatively different from that of $\widehat{\delta}$. When $\gamma_0 - \delta_0 \leq 1/2$, $\widehat{\gamma}$ is not consistent, and the bias is generally large (in absolute value) and does not decrease as T increases. This is, to some extent, mitigated for $\beta_0 = -5$ and $\beta_0 = 10$, where the coefficient on the deterministic trend is so large that the theoretically dominant stochastic component appears to be hidden in finite samples. On the other hand, when $\gamma_0 - \delta_0 > 1/2$ the bias is generally very small and decreases as $\gamma_0 - \delta_0$ increases, reflecting the fast convergence rates in those cases implied by Theorem 2. This effect is weaker when $\delta_0 = 0$, where $\gamma_0 - \delta_0 > 1/2$ implies that γ_0 is relatively close to 1, and the asymptotic multicollinearity clearly worsens the bias of the $\widehat{\gamma}$ estimates.

The Monte Carlo SD results for $\widehat{\gamma}$ are presented in Table 4. Again, these results are clearly in line with the predictions from asymptotic theory, and are qualitatively different from the SD results for $\widehat{\delta}$ in Table 2. For $\gamma_0 - \delta_0 \leq 1/2$, the SD is large and does not seem to decrease for larger values of T. The latter is seen regardless of the value of β_0. On the other hand, for $\gamma_0 - \delta_0 > 1/2$, the SD clearly decreases as either T or $\gamma_0 - \delta_0$ increases, although the results are relatively poor for $\beta_0 = 1, \delta_0 = 0$, due to the asymptotic multicollinearity problem. As was the case with the bias in Table 3, the SD is clearly smaller for $\beta_0 = -5$ or $\beta_0 = 10$ compared with $\beta_0 = 1$, reflecting the fact that the deterministic trend is easier to detect when its coefficient is larger in magnitude.

4.2   Testing procedure

Although parameter estimation and inference is straightforward using the results in Theorem 2, and for the parameters of the stochastic component it is not necessary to know which of the cases covered by Theorem 2 applies in any given situation. Nonetheless, it may sometimes be of interest to discern which case applies and, additionally, given that $\gamma_0 = 1$ leads to an identification problem, check whether the data support this possibility. Clearly, there are many possible ways of doing so, and we propose here a stepwise testing procedure. We do not pursue a formal analysis, which would involve very lengthy repetitions of techniques already developed in the proofs of Theorems 1 and 2, but instead illustrate the finite sample behavior by means of a small Monte Carlo simulation experiment. Our proposed testing procedure is as follows.

Step 1. Test $H_0^1 : \delta_0 = 1/2$ against $H_1^1 : \delta_0 < 1/2$ in model (12). Rejection of H_0^1 favors the possibility that μ_0 can be consistently estimated. This test is simple to implement because we conjecture that our proposed estimator $\widehat{\delta}$ derived from (19) has property (28) even in the case where $\gamma_0 = 1$. Assuming this conjecture is true, testing H_0^1 by means of a t-test is immediate. If H_0^1 is rejected, we proceed to Step 2, and if it is not then we proceed to Step 3.

Step 2. Test $H_0^2 : \beta_0 = 0$ against $H_1^2 : \beta_0 \neq 0$ in model (12). The result from Step 1 suggests that μ_0 can be consistently estimated, so, noting (12), it is crucial to determine whether $\gamma_0 = 1$ or $\gamma_0 \neq 1$. Nicely, the null $\gamma_0 = 1$ is equivalent to $\beta_0 = 0$ because both conditions lead to models that are observationally equivalent. Testing $H_0^2 : \beta_0 = 0$ against $H_1^2 : \beta_0 \neq 0$ is possible although γ_0 is not identified under the null. This is a classical problem in the hypothesis testing literature and several solutions have been provided (e.g. Hansen, 1996). We follow an LM approach. Let $(\widetilde{\boldsymbol{\tau}}, \widetilde{\mu})$ denote the restricted estimator that imposes H_0^2, and let $\gamma_F \in [\square_1, 1 - \kappa] \cup [1 + \kappa, \square_2]$ be a fixed number. Define the LM statistic

$$LM(\gamma_F) = \frac{T}{2\widetilde{\sigma}^2} \frac{\partial L_T(\widetilde{\boldsymbol{\tau}}, \gamma_F, \widetilde{\mu}, 0)}{\partial(\boldsymbol{\tau}', \mu, \beta)'} \left(\frac{\partial^2 L_T(\widetilde{\boldsymbol{\tau}}, \gamma_F, \widetilde{\mu}, 0)}{\partial(\boldsymbol{\tau}', \mu, \beta)\partial(\boldsymbol{\tau}', \mu, \beta)'} \right)^{-1} \frac{\partial L_T(\widetilde{\boldsymbol{\tau}}, \gamma_F, \widetilde{\mu}, 0)}{\partial(\boldsymbol{\tau}', \mu, \beta)},$$

where $\widetilde{\sigma}^2 = L_T(\widetilde{\boldsymbol{\tau}}, \gamma_F, \widetilde{\mu}, 0)$. We conjecture that, as $T \to \infty$,

$$LM(\gamma_F) \to_d \chi_1^2 \text{ under } H_0^2. \tag{30}$$

Another possibility is to consider the supremum over γ instead of a fixed γ_F, which would imply a nonstandard null limit distribution (e.g. Hansen, 1996). Our small Monte Carlo experiment will rely on (30) for a fixed γ_F. Additional, unreported Monte Carlo simulations suggest that results are relatively invariant to the choice of γ_F and also that $LM(\gamma_F)$ only has power if $\gamma_0 - 1/2 > \delta_0$. This makes sense because if if $\gamma_0 - 1/2 < \delta_0$ then the second term on the right-hand side of (12) is drowned by the stochastic component so it would be irrelevant whether $\beta_0 = 0$ or $\beta_0 \neq 0$. This is useful because rejection using $LM(\gamma_F)$ then favors the possibility that the generalized trend can be estimated consistently, whereas non-rejection suggests that the generalized trend is irrelevant (either because $\beta_0 = 0$ or because the trend is weak).

Step 3. Test $H_0^3 : \gamma_0 - 1/2 = \delta_0$ against $H_1^3 : \gamma_0 - 1/2 > \delta_0$ in model (12) under the restriction that $\mu = 0$. Because Step 1 suggests that μ_0 cannot be consistently estimated, we omit the drift from the analysis and in this step we check whether the generalized trend is strong. The test can be implemented as described in Hualde and Nielsen (2020).

Step 4. The outcomes of the tests in Steps 2 and 3 have the following implications.

M1: If H_0^2 is rejected then all model parameters can be estimated by (19) and $\widehat{\boldsymbol{\phi}}$.

M2: If H_0^2 is not rejected then we either employ (19) and $\widehat{\boldsymbol{\phi}}$ (ignoring the estimates of β_0 and γ_0), or we may use a simplified model that imposes $\beta = 0$.

M3: If H_0^3 is rejected then β_0, γ_0, and $\boldsymbol{\tau}_0$ can be consistently estimated. We either employ (19) and $\widehat{\boldsymbol{\phi}}$ (ignoring the estimate of μ_0), or we may use a simplified model that imposes $\mu = 0$.

M4: If H_0^3 is not rejected then only $\boldsymbol{\tau}_0$ can be consistently estimated. We either employ (19) and $\widehat{\boldsymbol{\phi}}$ (ignoring the estimates of μ_0, β_0, and γ_0), or we may use simplified versions of it in which we remove one or both deterministic components (and possibly ignoring estimates of the deterministic structure).

Summarizing, the outcome of the test procedure is one of M1,...,M4. As usual in model specification, we note that wrongly removing a component from the model renders all remaining estimates inconsistent. In the present context, Theorem 2 shows that retaining all deterministic components has no adverse effect asymptotically, in terms of efficiency or otherwise, on the estimation of the parameters associated with the stochastic component. This suggests that a conservative approach of always estimating the full model (12), including all deterministic components, is appropriate in most circumstances, while being careful in the interpretation of the parameters associated with the deterministic components. Nonetheless, we consider next the finite-sample properties of the above testing procedure.

Table 5 presents the proportion of cases (out of 10,000 replications) in which the stepwise testing procedure selects the different situations characterized by M1,...,M4. As before, the observable series $x_t, t = 1, \ldots, T$, was generated from (12) with $u_t = \varepsilon_t$ being an independent $N(0,1)$ sequence, and we present results for $T = 64, 256$, $\mu_0 = \beta_0 = 1$, $\delta_0 = 0, 0.4, 0.6, 1$,

Table 5: Monte Carlo selection frequencies of testing procedure

$\gamma_0 - \delta_0$	Model $\backslash T$	$\delta_0 = 0$		$\delta_0 = 0.4$		$\delta_0 = 0.6$		$\delta_0 = 1.0$	
		64	256	64	256	64	256	64	256
1.6	M1	**0.999**	**1.000**	**0.468**	**0.526**	0.090	0.002	0.000	0.000
1.6	M2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.6	M3	0.001	0.000	0.503	0.474	**0.862**	**0.998**	**0.942**	**1.000**
1.6	M4	0.000	0.000	0.029	0.000	0.048	0.000	0.058	0.000
1.4	M1	**0.998**	**1.000**	**0.454**	**0.528**	0.091	0.001	0.000	0.000
1.4	M2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.4	M3	0.001	0.000	0.480	0.472	**0.804**	**0.999**	**0.863**	**1.000**
1.4	M4	0.001	0.000	0.066	0.000	0.105	0.000	0.137	0.000
1.2	M1	**0.737**	**1.000**	**0.439**	**0.519**	0.085	0.001	0.000	0.000
1.2	M2	0.256	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.2	M3	0.003	0.000	0.418	0.481	**0.695**	**0.999**	**0.702**	**1.000**
1.2	M4	0.004	0.000	0.143	0.000	0.220	0.000	0.298	0.000
1.0	M1	0.104	0.077	**0.409**	**0.508**	0.080	0.001	0.000	0.000
1.0	M2	**0.893**	**0.923**	0.002	0.000	0.000	0.000	0.000	0.000
1.0	M3	0.000	0.000	0.283	0.486	**0.457**	**0.984**	**0.422**	**0.973**
1.0	M4	0.003	0.000	0.306	0.006	0.463	0.015	0.578	0.027
0.8	M1	**0.180**	**0.221**	**0.171**	**0.296**	0.067	0.001	0.000	0.000
0.8	M2	0.816	0.779	0.197	0.182	0.005	0.000	0.000	0.000
0.8	M3	0.000	0.000	0.105	0.287	**0.177**	**0.534**	**0.151**	**0.428**
0.8	M4	0.004	0.000	0.527	0.235	0.751	0.465	0.849	0.572
0.6	M1	**0.183**	**0.161**	0.076	0.055	0.033	0.001	0.000	0.000
0.6	M2	0.813	0.839	**0.274**	**0.406**	0.028	0.000	0.000	0.000
0.6	M3	0.000	0.000	0.028	0.036	**0.045**	**0.054**	**0.039**	**0.029**
0.6	M4	0.004	0.000	0.622	0.503	0.894	0.945	0.961	0.971
0.4	M1	0.147	0.100	0.086	0.066	0.026	0.000	0.000	0.000
0.4	M2	**0.850**	**0.900**	**0.273**	**0.392**	0.039	0.001	0.000	0.000
0.4	M3	0.000	0.000	0.010	0.008	0.015	0.025	0.030	0.056
0.4	M4	0.003	0.000	0.631	0.534	**0.920**	**0.974**	**0.970**	**0.944**
0.0	M1	0.107	0.076	0.081	0.059	0.029	0.000	0.000	0.000
0.0	M2	**0.888**	**0.924**	**0.271**	**0.397**	0.033	0.001	0.000	0.000
0.0	M3	0.000	0.000	0.002	0.003	0.023	0.051	0.050	0.098
0.0	M4	0.005	0.000	0.646	0.541	**0.915**	**0.948**	**0.950**	**0.902**
−0.4	M1	0.110	0.077	0.074	0.054	0.029	0.000	0.000	0.000
−0.4	M2	**0.887**	**0.923**	**0.261**	**0.400**	0.029	0.001	0.000	0.000
−0.4	M3	0.000	0.000	0.002	0.002	0.017	0.043	0.038	0.067
−0.4	M4	0.003	0.000	0.663	0.544	**0.925**	**0.956**	**0.962**	**0.933**

Note: Based on 10,000 Monte Carlo replications. Bold entries denote proportions of correct selections.

and 9 different values for γ_0 given by $\gamma_0 - \delta_0 = 1.6, 1.4, 1.2, \ldots, 0.4, 0.0, -0.4$. Numbers reported in bold correspond to proportions of correct choices. All tests were implemented with nominal size 0.05 and we fix $\gamma_F = 2$ in Step 2.

Overall, the behaviour of the testing procedure seems satisfactory. The results generally improve as T increases and correspond to what theory predicts. As expected, the correct identification of a strong trend (when $\gamma_0 - \delta_0 > 1/2$) is easier for larger values of $\gamma_0 - \delta_0$

and worsens substantially when $\gamma_0 - \delta_0 \leq 1$. This has to do with the relatively low power of Hualde and Nielsen's (2020) one-sided LM testing procedure for H_0^3 against H_1^3 when $\gamma_0 - \delta_0$ is close to $1/2$, and our results here are in line with theirs. Additionally, the test of H_0^1 against H_1^1 shows low power when $\delta_0 = 0.4$, which is certainly an adverse situation, though results improve as T increases.

5 Concluding remarks

We have proposed a parametric time series model that includes both a fractional stochastic component as well as a fractional deterministic component. The stochastic component is a fractionally integrated process driven by a memory parameter, δ, combined with a linear short-memory process. The deterministic component consists of the sum of a constant term and a flexible deterministic trend. The latter is fractional in the sense that it is defined using the same fractional coefficients as the fractional integration operator in the stochastic component, and similarly to the memory parameter δ, the deterministic trend is characterized by a power law parameter γ. Both the memory and power law parameters are assumed to lie in sets which can be arbitrarily large. Thus, our model may display many different behaviours, including various types of dependence (antipersistency, weak dependence, long memory) and very flexible deterministic trend functions.

Compared with our earlier work in Hualde and Nielsen (2020), there are three main differences in this paper. First, in this paper we apply the fractional coefficients directly to model the deterministic component instead of the approximation given by powers of t. As argued in the Introduction, the former are much more natural in a fractional context. Second, as a consequence of using the fractional coefficients to model the deterministic trend, we are able to relax the assumption $\gamma > 0$ on the power law parameter required by Hualde and Nielsen (2020). Instead, we allow the power law parameter to lie in any (arbitrarily large) compact interval. Third, we include an additional deterministic term such that we have both a constant/level and a trend. This complicates our analysis substantially, but clearly makes the model much more applicable in practice and also has the potential to alleviate bias arising from non-zero initial conditions in the Type II context (Johansen and Nielsen, 2016).

Our asymptotic results depend crucially on the relative strengths of the stochastic and deterministic components, as measured by the memory and power law parameters. Specifically, when the deterministic signal is sufficiently strong, that is if $\gamma_0 - \delta_0 > 1/2$ ($\delta_0 < 1/2$), the trend parameters γ_0 and β_0 (and level parameter μ_0) can be consistently estimated and their estimators are asymptotically normal. When the deterministic signal is weak, the parameters corresponding to the deterministic components (except μ_0 if $\delta_0 < 1/2$) cannot be consistently estimated. Remarkably, the asymptotic results for estimator corresponding to the stochastic part of the model (i.e., $\widehat{\boldsymbol{\tau}}$) are identical to those achieved in the simpler, purely stochastic, setting of Hualde and Robinson (2011), and are unaffected by the presence of the deterministic component, whether the parameters of the latter can be consistently estimated or not.

There are several interesting issues which have not been addressed in the present paper, but which will the object of future research. First, a semiparametric approach which focuses on estimating γ_0 and δ_0 (and possibly μ_0) without making parametric assumptions about the short-memory structure of z_t seems possible. Second, the fractional process which characterizes our model is Type II, and it would be of interest to determine whether our theory could also be developed for a Type I fractional process.

6 Proofs of theorems

Throughout, ϵ will denote a generic arbitrarily small positive constant, and K a generic arbitrarily large positive constant.

We note from the outset that many steps of the proofs are affected by the asymptotic behaviour of

$$\boldsymbol{\phi}_0'\boldsymbol{c}_{t-1}(1-\delta,\gamma_0-\delta,\boldsymbol{\varphi}) = \mu_0 c_{t-1}(1-\delta,\boldsymbol{\varphi}) + \beta_0 c_{t-1}(\gamma_0-\delta,\boldsymbol{\varphi}), \tag{31}$$

which, depending on the values of γ_0 and μ_0, is dominated by either the first or the second term on the right hand side of (31). Recalling that $\gamma_0 \in [\square_1, 1-\kappa] \cup [1+\kappa, \square_2]$ and also that if $\gamma_0 - 1/2 > \delta_0$ then $\beta_0 \neq 0$ (see Assumption A3), there are two cases. When $\mu_0 = 0$ or $\mu_0 \neq 0, \gamma_0 \geq 1+\kappa$, the second term dominates, whereas if $\mu_0 \neq 0, \gamma_0 \leq 1-\kappa$, the first term dominates. We will give the proof for the former case. The proof for the latter case ($\mu_0 \neq 0, \gamma_0 \leq 1-\kappa$) is very similar, just adapting many of the steps in the proof below and also some of the lemmas to the case where the first term in (31) is the dominant one, which mainly implies that "1" takes the role of "γ_0" in many parts of the proof.

6.1 Proof of Theorem 1(i): the $\gamma_0 - 1/2 > \delta_0$ case

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

$$\Pr(\widehat{\boldsymbol{\vartheta}} \in \overline{M}_\varepsilon) \to 0 \text{ as } T \to \infty, \tag{32}$$

$$\Pr(\widehat{\boldsymbol{\vartheta}} \in \overline{N}_\varepsilon \cap M_\varepsilon) \to 0 \text{ as } T \to \infty. \tag{33}$$

Strictly, ε should be $\varepsilon/\sqrt{2}$ in (32) and (33), but since ε is arbitrary this is irrelevant and we continue without the $\sqrt{2}$ factor. In (32) we prove consistency of $\widehat{\boldsymbol{\tau}}$, uniformly in γ. In (33) we prove consistency of $\widehat{\gamma}$, given that $\widehat{\boldsymbol{\tau}}$ is consistent and hence that $\boldsymbol{\tau}$ lies in a small neighborhood of $\boldsymbol{\tau}_0$.

Noting that from (13),

$$\rho(L;\boldsymbol{\varphi})x_t(\delta) = \boldsymbol{\phi}_0'\boldsymbol{c}_{t-1}(1-\delta,\gamma_0-\delta,\boldsymbol{\varphi}) + \rho(L;\boldsymbol{\varphi})u_t(\delta-\delta_0), \tag{34}$$

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

$$d_t(\boldsymbol{\vartheta}) = \boldsymbol{\phi}_0'\boldsymbol{c}_{t-1}(1-\delta,\gamma_0-\delta,\boldsymbol{\varphi})$$

$$- \boldsymbol{h}_{t-1,T}'(1-\delta,\gamma-\delta,\boldsymbol{\varphi})\sum_{j=1}^{T}\boldsymbol{h}_{j-1,T}(1-\delta,\gamma-\delta,\boldsymbol{\varphi})\boldsymbol{\phi}_0'\boldsymbol{c}_{j-1}(1-\delta,\gamma_0-\delta,\boldsymbol{\varphi}), \tag{35}$$

$$s_t(\boldsymbol{\vartheta}) = \rho(L;\boldsymbol{\varphi})u_t(\delta-\delta_0) - \boldsymbol{h}_{t-1,T}'(1-\delta,\gamma-\delta,\boldsymbol{\varphi})\sum_{j=1}^{T}\boldsymbol{h}_{j-1,T}(1-\delta,\gamma-\delta,\boldsymbol{\varphi})\rho(L;\boldsymbol{\varphi})u_j(\delta-\delta_0), \tag{36}$$

defining also the coefficient

$$\boldsymbol{h}_{t,T}(d_1,d_2,\boldsymbol{\varphi}) = \left(\sum_{j=1}^{T}\boldsymbol{c}_{j-1}(d_1,d_2,\boldsymbol{\varphi})\boldsymbol{c}_{j-1}'(d_1,d_2,\boldsymbol{\varphi})\right)^{-1/2}\boldsymbol{c}_t(d_1,d_2,\boldsymbol{\varphi}), \tag{37}$$

which clearly satisfies $\sum_{t=1}^{T} \boldsymbol{h}_{t-1,T}(d_1, d_2, \boldsymbol{\varphi}) \boldsymbol{h}'_{t-1,T}(d_1, d_2, \boldsymbol{\varphi}) = I_2$.

As in Hualde and Nielsen (2020) the strategy of proof exploits the different orders of magnitude of the stochastic term $s_t(\boldsymbol{\vartheta})$ and deterministic term $d_t(\boldsymbol{\vartheta})$ in $R_T(\boldsymbol{\vartheta})$ in different parts of the parameter space. Dealing with (32) and (33) allows us to consider separately the case where $\boldsymbol{\tau}$ is "far" from $\boldsymbol{\tau}_0$ (that is, $\overline{M}_\varepsilon$) and the case where γ is "far" from γ_0 (that is, $\overline{N}_\varepsilon$). Two important technical problems are that (a) when $\gamma = \gamma_0$ then $d_t(\boldsymbol{\vartheta}) = 0$ and (b) when δ is "far" from δ_0 then $u_t(\delta - \delta_0)$ may change from stationary to nonstationary. Both these features complicate greatly the treatment of (32) where $\gamma = \gamma_0$ is admissible. In (33) neither of these problems appear, and in that sense the proof of (33) is simpler than that of (32). Hence, we first give the proof of (33).

6.1.1 Proof of (33)

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 $s_{1t}(\boldsymbol{\tau}) = \rho(L; \boldsymbol{\varphi}) u_t(\delta - \delta_0)$ and

$$s_{2t}(\boldsymbol{\tau}, \gamma) = s_{2t}(\boldsymbol{\vartheta}) = \boldsymbol{h}'_{t-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \sum_{j=1}^{T} \boldsymbol{h}_{j-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) s_{1j}(\boldsymbol{\tau}); \qquad (38)$$

see (36). Thus, because $\sum_{t=1}^{T} \boldsymbol{h}_{t-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \boldsymbol{h}'_{t-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) = I_2$ by (37),

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

Because $\sum_{t=1}^{T} d_t(\boldsymbol{\vartheta}) \boldsymbol{c}_{t-1}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) = 0$, it holds that

$$\sum_{t=1}^{T} d_t(\boldsymbol{\vartheta}) s_t(\boldsymbol{\vartheta}) = \sum_{t=1}^{T} d_t(\boldsymbol{\vartheta}) \rho(L; \boldsymbol{\varphi}) u_t(\delta - \delta_0), \qquad (39)$$

where $\rho(L; \boldsymbol{\varphi}) u_t(\delta - \delta_0) = \sum_{j=0}^{t-1} c_j(\delta_0 - \delta, \boldsymbol{\varphi}) u_{t-j}$. Therefore,

$$R_T(\boldsymbol{\vartheta}) = \frac{1}{T} \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta}) + \frac{1}{T} \sum_{t=1}^{T} s_{1t}^2(\boldsymbol{\tau})$$

$$- \frac{1}{T} \left\| \sum_{j=1}^{T} s_{1j}(\boldsymbol{\tau}) \boldsymbol{h}_{j-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \right\|^2 + \frac{2}{T} \sum_{t=1}^{T} d_t(\boldsymbol{\vartheta}) s_{1t}(\boldsymbol{\tau}). \qquad (40)$$

Clearly, in view of (32) and for T sufficiently large, $\widehat{\boldsymbol{\vartheta}} \in \overline{N}_\varepsilon \cap M_\varepsilon$ is equivalent to $\inf_{\boldsymbol{\vartheta} \in \overline{N}_\varepsilon \cap M_\varepsilon} R_T(\boldsymbol{\vartheta}) \le \inf_{\boldsymbol{\vartheta} \in N_\varepsilon \cap M_\varepsilon} R_T(\boldsymbol{\vartheta})$. Because $R_T(\boldsymbol{\vartheta})$ is continuous in $\boldsymbol{\tau}$ and M_ε is compact, the infimum over $\overline{N}_\varepsilon \cap M_\varepsilon$ is attained for $\boldsymbol{\tau}$ and denoted $\widehat{\boldsymbol{\tau}}$, and it follows that $\inf_{\boldsymbol{\vartheta} \in \overline{N}_\varepsilon \cap M_\varepsilon} R_T(\boldsymbol{\vartheta}) = \inf_{\gamma \in \overline{N}_\varepsilon} R_T(\widehat{\boldsymbol{\tau}}, \gamma)$. Furthermore, it clearly holds that $\inf_{\boldsymbol{\vartheta} \in N_\varepsilon \cap M_\varepsilon} R_T(\boldsymbol{\vartheta}) \le \inf_{\gamma \in N_\varepsilon} R_T(\widehat{\boldsymbol{\tau}}, \gamma) \le R_T(\widehat{\boldsymbol{\tau}}, \gamma_0)$. Combining these inequalities we find that

$$\Pr(\widehat{\boldsymbol{\vartheta}} \in \overline{N}_\varepsilon \cap M_\varepsilon) = \Pr \left(\inf_{\boldsymbol{\vartheta} \in \overline{N}_\varepsilon \cap M_\varepsilon} R_T(\boldsymbol{\vartheta}) - \inf_{\boldsymbol{\vartheta} \in N_\varepsilon \cap M_\varepsilon} R_T(\boldsymbol{\vartheta}) \le 0 \right)$$

$$\le \Pr \left(\inf_{\gamma \in \overline{N}_\varepsilon} R_T(\widehat{\boldsymbol{\tau}}, \gamma) - R_T(\widehat{\boldsymbol{\tau}}, \gamma_0) \le 0 \right). \qquad (41)$$

Recalling that $d_t(\boldsymbol{\tau}, \gamma_0) = 0$, we obtain a useful cancellation of terms and find

$$R_T(\widehat{\boldsymbol{\tau}}, \gamma) - R_T(\widehat{\boldsymbol{\tau}}, \gamma_0) = \frac{1}{T} \sum_{t=1}^{T} d_t^2(\widehat{\boldsymbol{\tau}}, \gamma) + \frac{2}{T} \sum_{t=1}^{T} d_t(\widehat{\boldsymbol{\tau}}, \gamma) s_{1t}(\widehat{\boldsymbol{\tau}})$$

$$- \frac{1}{T} \left\| \sum_{j=1}^{T} s_{1j}(\widehat{\boldsymbol{\tau}}) \boldsymbol{h}_{j-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \right\|^2$$

$$+ \frac{1}{T} \left\| \sum_{j=1}^{T} s_{1j}(\widehat{\boldsymbol{\tau}}) \boldsymbol{h}_{j-1,T}(1 - \delta, \gamma_0 - \delta, \boldsymbol{\varphi}) \right\|^2 .$$

Thus, (33) holds if

$$\lim_{T \to \infty} \inf_{\boldsymbol{\vartheta} \in \overline{N}_\varepsilon \cap M_\varepsilon} \frac{1}{T^{2(\gamma_0 - \delta) - 1}} \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta}) > \epsilon, \tag{42}$$

$$\sup_{\boldsymbol{\vartheta} \in \overline{N}_\varepsilon \cap M_\varepsilon} \frac{1}{T^{2(\gamma_0 - \delta) - 1}} \left| \sum_{t=1}^{T} d_t(\boldsymbol{\vartheta}) s_{1t}(\boldsymbol{\tau}) \right| = o_p(1), \tag{43}$$

$$\sup_{\boldsymbol{\vartheta} \in \overline{N}_\varepsilon \cap M_\varepsilon} \frac{1}{T^{2(\gamma_0 - \delta) - 1}} \left\| \sum_{j=1}^{T} s_{1j}(\boldsymbol{\tau}) \boldsymbol{h}_{j-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \right\|^2 = o_p(1), \tag{44}$$

noting the change in the normalization compared with (41) ($T^{2(\gamma_0 - \delta_0) - 1}$ instead of T), which is justified because the right-hand side of the inequality inside the probability in (41) is 0, so multiplying the left- and right-hand sides of the inequality by a positive number does not alter the probability.

First, (42) follows from Lemma 2, noting that in $\overline{N}_\varepsilon \cap M_\varepsilon$, $\gamma_0 - \delta \geq \gamma_0 - \delta_0 - \varepsilon > 1/2$ setting ε small enough. Next, letting both ε and θ be sufficiently small and noting that in $\overline{N}_\varepsilon \cap M_\varepsilon$, $\delta_0 - \delta \geq -\varepsilon$, the left-hand side of (43) is $O_p(T^{1/2 + \delta_0 - \gamma_0 + 5\theta + \varepsilon}) = o_p(1)$ by (150) of Lemma 4. Finally, by (187) of Lemma 18 the left-hand side of (44) is $O_p(T^{-2(\gamma_0 - \delta_0 - 1/2 - \theta - \varepsilon)}) = o_p(1)$, to conclude the proof of (33).

6.1.2 Proof of (32)

To prove (32) we use

$$\Pr(\widehat{\boldsymbol{\vartheta}} \in \overline{M}_\varepsilon) = \Pr\left(\inf_{\boldsymbol{\vartheta} \in \overline{M}_\varepsilon} R_T(\boldsymbol{\vartheta}) \leq \inf_{\boldsymbol{\vartheta} \in M_\varepsilon} R_T(\boldsymbol{\vartheta}) \right) \leq \Pr\left(\inf_{\boldsymbol{\vartheta} \in \overline{M}_\varepsilon} S_T(\boldsymbol{\vartheta}) \leq 0 \right), \tag{45}$$

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 $\bigtriangledown_2 > \gamma_0 - 1 - \eta$. 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 $\mathcal{I}_1 = \{\delta : \bigtriangledown_1 \leq \delta \leq \delta_0 - 1/2 - \eta\}$, $\mathcal{I}_2 = \{\delta : \delta_0 - 1/2 - \eta \leq \delta \leq \delta_0 - 1/2\}$, $\mathcal{I}_3 = \{\delta : \delta_0 - 1/2 \leq \delta \leq \delta_0 - 1/2 + \eta\}$, $\mathcal{I}_4 = \{\delta : \delta_0 - 1/2 + \eta \leq \delta \leq \gamma_0 - 1 - \eta\}$, and $\mathcal{I}_5 = \{\delta : \gamma_0 - 1 - \eta \leq \delta \leq \bigtriangledown_2\}$, noting that the upper bound for η guarantees that $\mathcal{I}_4$ is non-empty, and correspondingly define $\mathcal{T}_i = \mathcal{I}_i \times \Psi$. Furthermore, fixing $\xi > 0$ and $\varrho > 0$ such that $\varrho < \min\{\eta/2, \kappa\}$, also define $\mathcal{H}_i = \{\boldsymbol{\vartheta} \in \overline{M}_\varepsilon : \boldsymbol{\tau} \in \mathcal{T}_i, |\gamma - \gamma_0| < \xi T^{-\varkappa_i}\}$,

$\overline{\mathcal{H}}_i = \{\boldsymbol{\vartheta} \in \overline{M}_\varepsilon : \boldsymbol{\tau} \in \mathcal{T}_i, \xi T^{-\varkappa_i} \leq |\gamma - \gamma_0| \leq \varrho\}$, and $\overline{\overline{\mathcal{H}}}_i = \{\boldsymbol{\vartheta} \in \overline{M}_\varepsilon : \boldsymbol{\tau} \in \mathcal{T}_i, |\gamma - \gamma_0| \geq \varrho\}$ for $i = 1, \ldots, 5$, where $\varkappa_i > 0$ will be defined subsequently, noting that $\overline{\mathcal{H}}_i$ is non-empty for any ξ, ϱ, for T large enough.

Now, by (45), (32) is justified by showing

$$\Pr\left(\inf_{\overline{\overline{\mathcal{H}}}_i} S_T(\boldsymbol{\vartheta}) \leq 0\right) \to 0 \text{ as } T \to \infty \text{ for } i = 1, \ldots, 5, \tag{46}$$

$$\Pr\left(\inf_{\overline{\mathcal{H}}_i} S_T(\boldsymbol{\vartheta}) \leq 0\right) \to 0 \text{ as } T \to \infty \text{ for } i = 1, \ldots, 5, \tag{47}$$

$$\Pr\left(\inf_{\mathcal{H}_i} S_T(\boldsymbol{\vartheta}) \leq 0\right) \to 0 \text{ as } T \to \infty \text{ for } i = 1, \ldots, 5. \tag{48}$$

The partitioning of the parameter space into $\mathcal{H}_i$, $\overline{\mathcal{H}}_i$, and $\overline{\overline{\mathcal{H}}}_i$ for $i = 1, \ldots, 5$ is useful because of the different behavior (orders of magnitude) of the stochastic term $s_t(\boldsymbol{\vartheta})$ and the deterministic term $d_t(\boldsymbol{\vartheta})$ on these sets. This motivates a separate analysis of (46), (47), and (48), at least for $i = 1, \ldots, 4$. We first give the main ideas and motivation for separate treatment of each of these sets, and then we give below the details of the proofs for each of these sets in separate subsections.

For $i = 5$ the stochastic term $s_t(\boldsymbol{\vartheta})$ dominates the deterministic term $d_t(\boldsymbol{\vartheta})$ in $S_T(\boldsymbol{\vartheta})$, and the contribution of $d_t(\boldsymbol{\vartheta})$ to $S_T(\boldsymbol{\vartheta})$ is negligible. Furthermore, because $u_t(\delta - \delta_0)$ is asymptotically stationary for $i = 5$, the proof for this case can be based on arguments from Hualde and Robinson (2011) for the purely stochastic term, $\rho(L; \boldsymbol{\varphi}) u_t(\delta - \delta_0)$.

For $i \leq 4$ the deterministic term $d_t(\boldsymbol{\vartheta})$ dominates the stochastic term $s_t(\boldsymbol{\vartheta})$ in $S_T(\boldsymbol{\vartheta})$, but only if $\gamma \neq \gamma_0$. Recall that $d_t(\boldsymbol{\vartheta}) = 0$ when $\gamma = \gamma_0$, which necessitates separate consideration of $\mathcal{H}_i$, $\overline{\mathcal{H}}_i$, and $\overline{\overline{\mathcal{H}}}_i$, at least for some i.

Specifically, on $\overline{\overline{\mathcal{H}}}_i$ we have $|\gamma - \gamma_0| \geq \varrho$, so that we can take advantage of $d_t(\boldsymbol{\vartheta})$ dominating $s_t(\boldsymbol{\vartheta})$ by an order of magnitude, so that the contribution of $s_t(\boldsymbol{\vartheta})$ to $S_T(\boldsymbol{\vartheta})$ is negligible. Thus, $\overline{\overline{\mathcal{H}}}_i$ can be dealt with for $i = 1, \ldots, 4$ with one proof that applies a uniform lower bound on, suitably normalized, $\sum_{t=1}^T d_t^2(\boldsymbol{\vartheta})$.

On $\mathcal{H}_i$ we have $|\gamma - \gamma_0| < \xi T^{-\varkappa_i}$, so that $\gamma = \gamma_0$ is admissible. Because $d_t(\boldsymbol{\vartheta}) = 0$ when $\gamma = \gamma_0$, we cannot exploit the lower bound on $\sum_{t=1}^T d_t^2(\boldsymbol{\vartheta})$. Thus, on $\mathcal{H}_i$ we need to deal carefully with the stochastic term $s_t(\boldsymbol{\vartheta})$ and we divide the parameter space and separately consider $i = 1, \ldots, 4$ as in Hualde and Robinson (2011), Johansen and Nielsen (2012a), and subsequent works. For $i = 4$, $u_t(\delta - \delta_0)$ is asymptotically stationary and we can apply the same proof as for $i = 5$ using the mean value theorem to show that the contribution of $d_t(\boldsymbol{\vartheta})$ is negligible. The cases $i = 2, 3$ deal with the discontinuity of the objective function, because $u_t(\delta - \delta_0)$ is near the border between stationarity and nonstationarity. The proof here uses the result of Hualde and Robinson (2011) that the contribution from the purely stochastic term, $\rho(L; \boldsymbol{\varphi}) u_t(\delta - \delta_0)$, can be made arbitrarily large, and we then show that the contribution from the deterministic term is bounded. Finally, for $i = 1$, $u_t(\delta - \delta_0)$ is nonstationary and we can use the method of Johansen and Nielsen (2019) and Hualde and Nielsen (2020), that avoids the strong moment condition (see Johansen and Nielsen, 2012b). The reason is that $u_t(\delta - \delta_0)$ has memory $\delta_0 - \delta \geq 1/2 + \eta$ for an arbitrarily small $\eta > 0$, and the most direct proof for $i = 1$ would involve justifying the weak convergence of the appropriately normalized sum

of squares of $u_t(\delta - \delta_0)$. The difficulty here is that for fixed δ convergence can be established under the condition that u_t has q finite moments, where $q \geq 2$ and $q > (\delta_0 - \delta - 1/2)^{-1}$. Thus, if $\delta_0 - \delta$ is close to $1/2$, the condition $q > (\delta_0 - \delta - 1/2)^{-1}$ is very strong, and, in fact, this is a serious technical problem because Johansen and Nielsen (2012b) showed that this moment condition is necessary and, moreover, in earlier parts of the proof η is required to be arbitrarily small. However, by using the lower bound in Lemma 22, we instead need to show the weak convergence of the appropriately normalized sum of squares of $u_t(\delta - \delta_0 - 1)$. This process has memory $\delta_0 - \delta + 1$, so this convergence does not require the strong moment condition because for $i = 1$, $\delta_0 - \delta + 1 \geq 3/2 + \eta$. In addition, note that Johansen and Nielsen (2019) require 8 moments, but this is not used to establish the bounds that we require. The cases $\overline{\mathcal{H}_i}$ for $i = 1, \ldots, 4$ also require joint treatment of the deterministic and stochastic terms in $S_T(\boldsymbol{\vartheta})$. In these case we show that, after suitable normalization, the contribution from the deterministic term $d_t(\boldsymbol{\vartheta})$ is "large" compared to that of the stochastic term $s_t(\boldsymbol{\vartheta})$. In that sense, the proofs here are opposite those of $\mathcal{H}_i$ for $i = 2, 3$. Because of the different behavior and normalization of $u_t(\delta - \delta_0)$, and hence of $s_t(\boldsymbol{\vartheta})$, for each of $i = 1, \ldots, 4$, slightly different proofs are needed for $\overline{\mathcal{H}_i}$ in each of these cases.

6.1.3   Proof of (46), (47), and (48) 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 $\delta_0 - \delta \leq 1 + \delta_0 - \gamma_0 + \eta < 1/2$, so $u_t(\delta - \delta_0)$ is asymptotically stationary. Let

$$S_T(\boldsymbol{\vartheta}) = U(\boldsymbol{\tau}) - r_T(\boldsymbol{\vartheta}), \tag{49}$$

where $U(\boldsymbol{\tau}) = E((\rho(L; \boldsymbol{\varphi})\Delta^{\delta-\delta_0}u_t)^2) - \sigma_0^2$ and, recalling $d_t(\boldsymbol{\vartheta}_0) = 0$,

$$r_T(\boldsymbol{\vartheta}) = \frac{1}{T}\sum_{t=1}^{T}(\rho(L; \boldsymbol{\varphi}_0)(u_t\mathbb{I}(t > 0)))^2 - \sigma_0^2$$

$$- \frac{1}{T}\sum_{t=1}^{T}\left((\rho(L; \boldsymbol{\varphi})u_t(\delta - \delta_0))^2 - E((\rho(L; \boldsymbol{\varphi})\Delta^{\delta-\delta_0}u_t)^2)\right)$$

$$- \frac{1}{T}\left\|\sum_{t=1}^{T}\rho(L; \boldsymbol{\varphi}_0)(u_t\mathbb{I}(t > 0))\boldsymbol{h}_{t-1,T}(1 - \delta_0, \gamma_0 - \delta_0, \boldsymbol{\varphi}_0)\right\|^2$$

$$+ \frac{1}{T}\left\|\sum_{t=1}^{T}\rho(L; \boldsymbol{\varphi})u_t(\delta - \delta_0)\boldsymbol{h}_{t-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi})\right\|^2$$

$$- \frac{2}{T}\sum_{t=1}^{T}d_t(\boldsymbol{\vartheta})s_t(\boldsymbol{\vartheta}) - \frac{1}{T}\sum_{t=1}^{T}d_t^2(\boldsymbol{\vartheta}).$$

It follows that (46), (47), and (48) for $i = 5$ hold if we show that

$$\inf_{\|\boldsymbol{\tau} - \boldsymbol{\tau}_0\| \geq \varepsilon, \boldsymbol{\tau} \in \mathcal{T}_5} U(\boldsymbol{\tau}) > \epsilon, \tag{50}$$

$$\frac{1}{T} \sum_{t=1}^{T} (\rho(L; \boldsymbol{\varphi}_0)(u_t \mathbb{I}(t > 0)))^2 - \sigma_0^2 = o_p(1), \tag{51}$$

$$\sup_{\|\boldsymbol{\tau} - \boldsymbol{\tau}_0\| \geq \varepsilon, \boldsymbol{\tau} \in \mathcal{T}_5} \frac{1}{T} \sum_{t=1}^{T} \left((\rho(L; \boldsymbol{\varphi}) u_t(\delta - \delta_0))^2 - E((\rho(L; \boldsymbol{\varphi}) \Delta^{\delta - \delta_0} u_t)^2) \right) = o_p(1), \tag{52}$$

$$\sup_{\mathcal{H}_5 \cup \overline{\mathcal{H}}_5 \cup \overline{\overline{\mathcal{H}}}_5} \frac{1}{T} \left| \sum_{t=1}^{T} d_t(\boldsymbol{\vartheta}) s_t(\boldsymbol{\vartheta}) \right| = o_p(1), \tag{53}$$

$$\sup_{\mathcal{H}_5 \cup \overline{\mathcal{H}}_5 \cup \overline{\overline{\mathcal{H}}}_5} \frac{1}{T} \left\| \sum_{t=1}^{T} \rho(L; \boldsymbol{\varphi}) u_t(\delta - \delta_0) \boldsymbol{h}_{t-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \right\|^2 = o_p(1). \tag{54}$$

First, (50), (51), and (52) follow by identical arguments to those in the proofs of (2.8) and (2.9) in Hualde and Robinson (2011). Next, for any $\theta > 0$, by (148) of Lemma 4 with $\gamma_0 - \delta \leq 1 + \eta$ and $\delta_0 - \delta \leq \delta_0 - \gamma_0 + 1 + \eta$, the left-hand side of (53) is $O_p(T^{\max\{\theta, 1 + \delta_0 - \gamma_0 + \eta\} + 4\theta - 1/2 + \eta})$ which is $o_p(1)$ for θ and η sufficiently small. In the same way, the left-hand side of (54) is $O_p(T^{2\max\{\theta, 1 + \delta_0 - \gamma_0 + \eta\} - 1})$ by (185) of Lemma 18, which is again $o_p(1)$ for θ and η small enough. This concludes the proof of (46), (47), and (48) for $i = 5$.

6.1.4 Proof of (46) for $i = 1, \ldots, 4$

In view of Lemma 1 and that $d_t(\boldsymbol{\vartheta}_0) = 0$, the result 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 $\gamma_0 - \delta \geq 1 + \eta$, so the probability above is bounded by

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

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

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

When $\delta \in \cup_{i=1}^{4} \mathcal{I}_i$ we have $\delta_0 - \delta \geq \delta_0 - \gamma_0 + 1 + \eta$, so by (150) of Lemma 4, for θ small enough,

$$\sup_{\overline{\overline{\mathcal{H}}}_i} \frac{2}{T^{2(\gamma_0 - \delta) - 1}} \left| \sum_{t=1}^{T} d_t(\boldsymbol{\vartheta}) s_t(\boldsymbol{\vartheta}) \right| = o_p(1),$$

and the proof follows by Lemma 2.

6.1.5 Proof of (47) and (48) for $i = 4$

Fix ζ such that $0 < \zeta < \eta$ and let $\varkappa_4 = \gamma_0 - \delta - 1 - \zeta$, noting that $\varkappa_4 \geq \eta - \zeta > 0$ when $\delta \in \mathcal{I}_4$. Then, because $d_t(\boldsymbol{\vartheta}_0) = 0$, (47) holds if

$$\Pr\left(\inf_{\overline{\mathcal{H}}_4} \frac{T^{2\varkappa_4}}{T^{2(\gamma_0-\delta)-1}} \left(\sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta}) - 2\left| \sum_{t=1}^{T} d_t(\boldsymbol{\vartheta})s_t(\boldsymbol{\vartheta}) \right| - \sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}_0) \right) \leq 0 \right) \to 0 \qquad (55)$$

as $T \to \infty$, noting the change in the normalization from (47) to (55), which is justified because the right-hand side of the inequality inside the probability in (47) is 0, so multiplying the left- and right-hand sides of the inequality by the same positive number does not alter the probability. By the Cauchy-Schwarz inequality and (39), the probability in (55) is bounded by

$$\Pr\left(\inf_{\overline{\mathcal{H}}_4} \frac{T^{2\varkappa_4}}{T^{2(\gamma_0-\delta)-1}} \left(\sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta})(1 - 2v_T(\boldsymbol{\vartheta})) - \sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}_0) \right) \leq 0 \right), \qquad (56)$$

where $v_T(\boldsymbol{\vartheta}) = (\sum_{t=1}^{T}(\rho(L;\boldsymbol{\varphi})u_t(\delta - \delta_0))^2 / \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta}))^{1/2}$. Then (55) holds if

$$\sup_{\overline{\mathcal{H}}_4} \frac{T^{2\varkappa_4}}{T^{2(\gamma_0-\delta)-1}} \sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}_0) = o_p(1), \qquad (57)$$

$$\Pr\left(\inf_{\overline{\mathcal{H}}_4} \frac{T^{2\varkappa_4}}{T^{2(\gamma_0-\delta)-1}} \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta})(1 - 2v_T(\boldsymbol{\vartheta})) \leq \epsilon \right) \to 0 \text{ as } T \to \infty. \qquad (58)$$

First, because $T^{2\varkappa_4 - 2(\gamma_0-\delta)+1} = T^{-1-2\zeta}$, (57) follows immediately by Lemma 1. Next, fixing c such that $0 < c < 1/2$, the probability in (58) equals

$$\Pr\left(\inf_{\overline{\mathcal{H}}_4} \frac{T^{2\varkappa_4}}{T^{2(\gamma_0-\delta)-1}} \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta})(1 - 2v_T(\boldsymbol{\vartheta})) \leq \epsilon, \sup_{\overline{\mathcal{H}}_4} v_T(\boldsymbol{\vartheta}) \leq c \right)$$

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

$$\leq \Pr\left(\inf_{\overline{\mathcal{H}}_4} \frac{T^{2\varkappa_4}}{T^{2(\gamma_0-\delta)-1}} \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta})(1 - 2c) \leq \epsilon \right) + \Pr\left(\sup_{\overline{\mathcal{H}}_4} v_T(\boldsymbol{\vartheta}) > c \right), \qquad (59)$$

so (58) holds on showing that

$$\lim_{T\to\infty} \inf_{\overline{\mathcal{H}}_4} \frac{T^{2\varkappa_4}}{T^{2(\gamma_0-\delta)-1}} \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta}) > \epsilon, \qquad (60)$$

$$\sup_{\overline{\mathcal{H}}_4} v_T(\boldsymbol{\vartheta}) = o_p(1). \qquad (61)$$

We first justify (60). Choose an arbitrarily small $\alpha > 0$. Then we initially show that

$$\lim_{T\to\infty} \inf_{\overline{\mathcal{H}}_4, 1-\delta \leq 1/2+\alpha} \frac{T^{2\varkappa_4}}{T^{2(\gamma_0-\delta)-1}} \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta}) > \epsilon. \qquad (62)$$

By the Cauchy-Schwarz inequality, $\sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta}) \geq T^{-1}\bar{d}_T^2(\boldsymbol{\vartheta})$, where $\bar{d}_t(\boldsymbol{\vartheta}) = \sum_{s=1}^{t} d_s(\boldsymbol{\vartheta})$, so that (62) holds by (146) of Lemma 3. Next we show

$$\lim_{T\to\infty} \inf_{\overline{\mathcal{H}}_{4},1-\delta\geq 1/2+\alpha} \frac{T^{2\varkappa_4}}{T^{2(\gamma_0-\delta)-1}} \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta}) > \epsilon. \tag{63}$$

By very similar arguments to those given in the proof of Lemma 2, using also Lemma 11, as well as very similar steps to those in the proof of Lemma 13, approximating sums by integrals (Lemma 10) and noting that the second term in (31) dominates because $\gamma_0 \geq 1 + \kappa$, it can be shown that

$$\frac{T^{2\varkappa_4}}{T^{2(\gamma_0-\delta)-1}} \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta}) = \frac{T^{2\varkappa_4}\beta_0^2 \rho^2(1;\boldsymbol{\varphi}_0)}{\Gamma^2(\gamma_0-\delta)} \left(\frac{1}{2(\gamma_0-\delta)-1} - \left(\begin{smallmatrix} \frac{1}{\gamma_0-2\delta} & \frac{1}{\gamma_0+\gamma-2\delta-1} \end{smallmatrix} \right) \right.$$
$$\left. \times \left(\begin{smallmatrix} \frac{1}{2(1-\delta)-1} & \frac{1}{\gamma-2\delta} \\ \frac{1}{\gamma-2\delta} & \frac{1}{2(\gamma-\delta)-1} \end{smallmatrix} \right)^{-1} \left(\begin{smallmatrix} \frac{1}{\gamma_0-2\delta} \\ \frac{1}{\gamma_0+\gamma-2\delta-1} \end{smallmatrix} \right) \right) + r_{4T}(\boldsymbol{\vartheta}), \tag{64}$$

where $\sup_{\overline{\mathcal{H}}_{4},1-\delta\geq 1/2+\alpha} |r_{4T}(\boldsymbol{\vartheta})| = o(1)$. It can be shown that the first term on the right-hand side of (64) equals

$$\frac{T^{2\varkappa_4}\beta_0^2 \rho^2(1;\boldsymbol{\varphi}_0)(\gamma_0-1)^2(\gamma_0-\gamma)^2}{\Gamma^2(\gamma_0-\delta)(2(\gamma_0-\delta)-1)(\gamma_0-2\delta)^2(\gamma_0+\gamma-2\delta-1)^2}, \tag{65}$$

so by (22), (64), and (65)

$$\inf_{\overline{\mathcal{H}}_{4},1-\delta\geq 1/2+\alpha} \frac{T^{2\varkappa_4}}{T^{2(\gamma_0-\delta)-1}} \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta}) \geq \frac{\kappa^2\epsilon}{K} \inf_{\overline{\mathcal{H}}_4} T^{2\varkappa_4}(\gamma_0-\gamma)^2 = \frac{\kappa^2\epsilon}{K}\xi^2 > \epsilon,$$

to conclude the proof of (60).

To show (61), note that

$$\sup_{\overline{\mathcal{H}}_4} v_T(\boldsymbol{\vartheta}) \leq \left(\frac{\sup_{\overline{\mathcal{H}}_4} T^{-1-2\zeta} \sum_{t=1}^{T}(\rho(L;\boldsymbol{\varphi})u_t(\delta-\delta_0))^2}{\inf_{\overline{\mathcal{H}}_4} T^{2\varkappa_4-(2(\gamma_0-\delta)-1)} \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta})} \right)^{1/2}, \tag{66}$$

recalling $\varkappa_4 = \gamma_0 - \delta - 1 - \zeta$. Here, $\sup_{\overline{\mathcal{H}}_4} T^{-1-2\zeta} \sum_{t=1}^{T}(\rho(L;\boldsymbol{\varphi})u_t(\delta-\delta_0))^2 = o_p(1)$ by Lemma 17 because $\delta_0 - \delta \leq 1/2 - \eta$. Then (66) is $o_p(1)$ by (60), which concludes the proof of (47) for $i = 4$.

Next we show (48) 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 $\sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta})$ as in (60) because $d_t(\boldsymbol{\vartheta}) = 0$ 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 $\delta_0 - \delta \leq 1/2 - \eta$ and $\gamma_0 - \delta \geq 1 + \eta$, so that $u_t(\delta - \delta_0)$ is asymptotically stationary as in the proof for $i = 5$. Then, given (49), the result follows by (50), (51), (52) (whose proofs apply also for $\delta \in \mathcal{I}_4$), and showing also that

$$\sup_{\mathcal{H}_4} \frac{1}{T} \left\| \sum_{t=1}^{T} \rho(L;\boldsymbol{\varphi})u_t(\delta-\delta_0)\boldsymbol{h}_{t-1,T}(1-\delta,\gamma-\delta,\boldsymbol{\varphi}) \right\|^2 = o_p(1), \tag{67}$$

$$\sup_{\mathcal{H}_4} \frac{1}{T} \left| \sum_{t=1}^{T} d_t(\boldsymbol{\vartheta})s_t(\boldsymbol{\vartheta}) \right| = o_p(1). \tag{68}$$

From (185) of Lemma 18, the left-hand side of (67) is $O_p(T^{-2\eta}) = o_p(1)$ by choosing $\theta < 1/2 - \eta$. Next, because $|\gamma - \gamma_0| < \xi T^{-\varkappa_4}$ in $\mathcal{H}_4$, by (147) and (149) of Lemma 4 the left-hand side of (68) is $O_p(T^{\zeta - \eta + 4\theta}) = o_p(1)$ for θ small enough because $\zeta < \eta$.

6.1.6 Proof of (47) for $i = 1, 2, 3$

The proofs of (47) for $i = 1, 2, 3$ are nearly identical, but with different normalizations given by $q_1 = q_2 = 2(\delta_0 - \delta)$ and $q_3 = 1$. Fix $\chi_1 = \chi_2 = \gamma_0 - \delta_0 - 1/2 > 0$ and $\varkappa_3 = \gamma_0 - \delta - 1$, so that $\varkappa_3 \geq \gamma_0 - \delta_0 - 1/2 - \eta > 0$ because $\delta \in \mathcal{I}_3$, and finally let $\alpha > 0$ be an arbitrarily small number. To justify (47) for $i = 1, 2, 3$, we prove that

$$\Pr\left(\inf_{\overline{\mathcal{H}}_i, 1 - \delta \leq 1/2 + \alpha} S_T(\boldsymbol{\vartheta}) \leq 0\right) \to 0 \text{ as } T \to \infty, \tag{69}$$

$$\Pr\left(\inf_{\overline{\mathcal{H}}_i, 1 - \delta \geq 1/2 + \alpha} S_T(\boldsymbol{\vartheta}) \leq 0\right) \to 0 \text{ as } T \to \infty. \tag{70}$$

First, changing the normalization (T^{q_i} instead of T) and applying the Cauchy-Schwarz inequality, the probability in (69) is bounded by

$$\Pr\left(\inf_{\overline{\mathcal{H}}_i, 1 - \delta \leq 1/2 + \alpha} \frac{1}{T^{q_i + 1}}(\overline{d}_T(\boldsymbol{\vartheta}) + \overline{s}_T(\boldsymbol{\vartheta}))^2 - \frac{1}{T^{q_i}}\sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}_0) \leq 0\right), \tag{71}$$

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

$$\overline{s}_T(\boldsymbol{\vartheta}) = \sum_{t=1}^{T} s_t(\boldsymbol{\vartheta}) = \rho(L; \boldsymbol{\varphi})u_T(\delta - \delta_0 - 1)$$

$$- \sum_{t=1}^{T} \boldsymbol{h}'_{t-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi})\sum_{j=1}^{T} \boldsymbol{h}_{j-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi})\rho(L; \boldsymbol{\varphi})u_j(\delta - \delta_0). \tag{72}$$

The probability in (71) is bounded by

$$\Pr\left(\inf_{\overline{\mathcal{H}}_i, 1 - \delta \leq 1/2 + \alpha} \frac{1}{T^{q_i + 1}}\overline{d}_T^2(\boldsymbol{\vartheta})(1 - 2|\overline{v}_T(\boldsymbol{\vartheta})|) - \frac{1}{T^{q_i}}\sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}_0) \leq 0\right), \tag{73}$$

where $\overline{v}_T(\boldsymbol{\vartheta}) = \overline{s}_T(\boldsymbol{\vartheta})/\overline{d}_T(\boldsymbol{\vartheta})$. Applying Lemma 1, (69) for $i = 1, 2, 3$ then holds if

$$\Pr\left(\inf_{\overline{\mathcal{H}}_i, 1 - \delta \leq 1/2 + \alpha} \frac{1}{T^{q_i + 1}}\overline{d}_T^2(\boldsymbol{\vartheta})(1 - 2|\overline{v}_T(\boldsymbol{\vartheta})|) \leq K\right) \to 0 \text{ as } T \to \infty, \tag{74}$$

for an arbitrarily large K. As in (59), fixing c such that $0 < c < 1/2$, the probability in (74) is bounded by

$$\Pr\left(\inf_{\overline{\mathcal{H}}_i, 1 - \delta \leq 1/2 + \alpha} \frac{1}{T^{q_i + 1}}\overline{d}_T^2(\boldsymbol{\vartheta})(1 - 2c) \leq K\right) + \Pr\left(\sup_{\overline{\mathcal{H}}_i, 1 - \delta \leq 1/2 + \alpha} |\overline{v}_T(\boldsymbol{\vartheta})| > c\right), \tag{75}$$

so, as in (66), (74) holds if

$$\sup_{\overline{\mathcal{H}}_i, 1 - \delta \leq 1/2 + \alpha} \frac{1}{T^{q_i/2 + 1/2}}|\overline{s}_T(\boldsymbol{\vartheta})| = O_p(1), \tag{76}$$

$$\lim_{T \to \infty} \inf_{\overline{\mathcal{H}}_i, 1 - \delta \leq 1/2 + \alpha} \frac{1}{T^{q_i + 1}}\overline{d}_T^2(\boldsymbol{\vartheta}) > K. \tag{77}$$

For $\delta \in \mathcal{I}_3$ it holds that $\delta_0 - \delta \leq 1/2$, so in view of (72) the proof of (76) is immediate using (177) in Lemma 16 together with Lemmas 17 and 18 with $\theta < 1/2$. For $\delta \in \mathcal{I}_1 \cup \mathcal{I}_2$ it holds that $\delta_0 - \delta \geq 1/2$, so with the different normalization, (76) again follows by application of Lemmas 16, 17 and 18. Finally, because $T^{-(q_i+1)} = T^{2\varkappa_i - 2(\gamma_0 - \delta)}$, (77) follows by (146) of Lemma 3, to conclude the proof of (69) for $i = 1, 2, 3$.

Next, (70) holds if, for $i = 1, 2, 3$,

$$\Pr\left(\inf_{\overline{\mathcal{H}}_i, 1-\delta \geq 1/2 + \alpha} \frac{1}{T^{q_i}} \left(\sum_{t=1}^{T} (d_t(\boldsymbol{\vartheta}) + s_t(\boldsymbol{\vartheta}))^2 - \sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}_0) \right) \leq 0 \right) \to 0 \qquad (78)$$

as $T \to \infty$. Using Lemma 22 and proceeding as in the proof of (47) for $i = 4$, the probability in (78) is bounded by

$$\Pr\left(\inf_{\overline{\mathcal{H}}_i, 1-\delta \geq 1/2 + \alpha} \frac{1}{T^{q_i}} \left(\frac{\pi^2}{4T^2} \sum_{t=1}^{T} \overline{d}_t^2(\boldsymbol{\vartheta})(1 - \overline{w}_T(\boldsymbol{\vartheta})) - \sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}_0) \right) \leq \epsilon \right),$$

where

$$\overline{w}_T(\boldsymbol{\vartheta}) = \left(\frac{\sum_{t=1}^{T} \overline{s}_t^2(\boldsymbol{\vartheta})}{\sum_{t=1}^{T} \overline{d}_t^2(\boldsymbol{\vartheta})} \right)^{1/2}.$$

Then, because $T^{2\varkappa_i - (2(\gamma_0 - \delta) + 1)} = T^{-q_i - 2}$ for $i = 1, 2, 3$, by Lemma 1, (78) holds if

$$\lim_{T \to \infty} \inf_{\overline{\mathcal{H}}_i, 1-\delta \geq 1/2 + \alpha} \frac{T^{2\varkappa_i}}{T^{2(\gamma_0 - \delta) + 1}} \sum_{t=1}^{T} \overline{d}_t^2(\boldsymbol{\vartheta}) > K, \qquad (79)$$

$$\sup_{\overline{\mathcal{H}}_i, 1-\delta \geq 1/2 + \alpha} \frac{1}{T^{q_i + 2}} \sum_{t=1}^{T} \overline{s}_t^2(\boldsymbol{\vartheta}) = O_p(1). \qquad (80)$$

Again, by similar arguments to those given in the proofs of Lemmas 2 and 13, and using Lemmas 10 and 11 to approximate sums by integrals, it can be shown that

$$\frac{1}{T^{2(\gamma_0 - \delta) + 1}} \sum_{t=1}^{T} \overline{d}_t^2(\boldsymbol{\vartheta}) = \frac{\beta_0^2 \rho^2(1; \boldsymbol{\varphi}_0)}{\Gamma^2(\gamma_0 - \delta + 1)} \Lambda(\gamma, \delta) + p_T(\boldsymbol{\vartheta}),$$

where $\sup_{\overline{\mathcal{H}}_i, 1-\delta \geq 1/2 + \alpha} |p_T(\boldsymbol{\vartheta})| = o(1)$ for $i = 1, 2, 3$ and

$$\Lambda(\gamma, \delta) = \frac{1}{2(\gamma_0 - \delta) + 1} + \begin{pmatrix} \frac{(\gamma_0 - \delta)^2}{\gamma_0 - 2\delta} & \frac{(\gamma_0 - \delta)^2}{\gamma_0 + \gamma - 2\delta - 1} \end{pmatrix} \begin{pmatrix} \frac{1}{2(1-\delta)-1} & \frac{1}{\gamma - 2\delta} \\ \frac{1}{\gamma - 2\delta} & \frac{1}{2(\gamma - \delta) - 1} \end{pmatrix}^{-1}$$

$$\times \begin{pmatrix} \frac{1}{(2(1-\delta)+1)(1-\delta)^2} & \frac{1}{(\gamma - 2\delta + 2)(1-\delta)(\gamma - \delta)} \\ \frac{1}{(\gamma - 2\delta + 2)(1-\delta)(\gamma - \delta)} & \frac{1}{(2(\gamma - \delta) + 1)(\gamma - \delta)^2} \end{pmatrix} \begin{pmatrix} \frac{1}{2(1-\delta)-1} & \frac{1}{\gamma - 2\delta} \\ \frac{1}{\gamma - 2\delta} & \frac{1}{2(\gamma - \delta) - 1} \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{\gamma_0 - 2\delta} \\ \frac{1}{\gamma_0 + \gamma - 2\delta - 1} \end{pmatrix}$$

$$- 2 \begin{pmatrix} \frac{\gamma_0 - \delta}{\gamma_0 - 2\delta} & \frac{\gamma_0 - \delta}{\gamma_0 + \gamma - 2\delta - 1} \end{pmatrix} \begin{pmatrix} \frac{1}{2(1-\delta)-1} & \frac{1}{\gamma - 2\delta} \\ \frac{1}{\gamma - 2\delta} & \frac{1}{2(\gamma - \delta) - 1} \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{(\gamma_0 - 2\delta + 2)(1-\delta)} \\ \frac{1}{(\gamma_0 + \gamma - 2\delta + 1)(\gamma - \delta)} \end{pmatrix}.$$

By simple but very cumbersome calculations, it can be shown that

$$\Lambda(\gamma, \delta) = \frac{(\gamma_0 - 1)^2(\gamma_0 - \gamma)^2 \Lambda_N(\gamma, \delta)}{\Lambda_D(\gamma, \delta)},$$

where

$$\Lambda_D(\gamma, \delta) = (1 - \delta)^2 (\gamma - \delta)^2 (2(\gamma_0 - \delta) + 1)(2(1 - \delta) + 1)(2(\gamma - \delta) + 1)(\gamma_0 - 2\delta)^2$$
$$\times (\gamma_0 - 2\delta + 2)(\gamma_0 + \gamma - 2\delta - 1)^2 (\gamma_0 + \gamma - 2\delta + 1)(\gamma - 2\delta + 2),$$

and $\Lambda_N(\gamma, \delta)$ is a complicated function. Then, noting that $\gamma_0 \neq 1$ by Assumption A3 and that $|\gamma_0 - \gamma| \geq \xi T^{-\varkappa_i}$, setting ξ large enough, (79) holds by showing that

$$\inf_{\overline{\mathcal{H}}_{i}, 1-\delta \geq 1/2+\alpha} \Lambda_N(\gamma, \delta) > \epsilon. \tag{81}$$

First, given that $|\gamma_0 - \gamma| \leq \varrho$, it can be shown that

$$\inf_{\overline{\mathcal{H}}_{i}, 1-\delta \geq 1/2+\alpha} \Lambda_N(\gamma, \delta) \geq \inf_{\overline{\mathcal{H}}_{i}, 1-\delta \geq 1/2+\alpha} \Lambda_N(\gamma_0, \delta) - K\varrho, \tag{82}$$

where the second term on the right-hand side of (82) can be made arbitrarily small by setting ϱ small enough. Also, it can be shown that

$$\Lambda_N(\gamma_0, \delta) = (\gamma_0 - \delta) \sum_{k=0}^{6} (\gamma_0 - \delta - 1)^k g_k(\delta) + (2(1 - \delta) + 1)((1 - \delta)^2 - 1)^2,$$

where $g_k(\delta)$ are relatively complicated polynomials of δ, for which it can be shown that

$$\inf_{\overline{\mathcal{H}}_{i}, 1-\delta \geq 1/2+\alpha} g_k(\delta) > \epsilon$$

for $i = 1, 2, 3$, to justify (81) and conclude the proof of (79).

Next we show (80). Clearly,

$$\frac{1}{T^{q_i+2}} \sum_{t=1}^{T} \overline{s}_t^2(\boldsymbol{\vartheta})$$

$$\leq \frac{2}{T^{q_i+2}} \sum_{t=1}^{T} (\rho(L; \boldsymbol{\varphi}) u_t(\delta - \delta_0 - 1))^2$$

$$+ \frac{2}{T^{q_i+2}} \sum_{t=1}^{T} \left(\sum_{s=1}^{t} \sum_{j=1}^{T} \rho(L; \boldsymbol{\varphi}) u_j(\delta - \delta_0) \boldsymbol{h}'_{j-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \boldsymbol{h}_{s-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \right)^2. \tag{83}$$

First, we note that $\delta_0 - \delta + 1 \leq 3/2$ for $i = 3$ and $\delta_0 - \delta + 1 \geq 3/2$ for $i = 1, 2$. Then, by (183) and (184) of Lemma 17,

$$\sup_{\overline{\mathcal{H}}_{i}, 1-\delta \geq 1/2+\alpha} \frac{2}{T^{q_i+2}} \sum_{t=1}^{T} (\rho(L; \boldsymbol{\varphi}) u_t(\delta - \delta_0 - 1))^2 = O_p \left(\frac{1}{T^3} \sum_{t=1}^{T} t^2 \right) = O_p(1)$$

for $i = 1, 2, 3$. Similarly, by (176) of Lemma 16 and (185),(187) of Lemma 18,

$$\sup_{\overline{\mathcal{H}}_{i}, 1-\delta \geq 1/2+\alpha} \frac{2}{T^{q_i+2}} \sum_{t=1}^{T} \left(\sum_{s=1}^{t} \sum_{j=1}^{T} \rho(L; \boldsymbol{\varphi}) u_j(\delta - \delta_0) \boldsymbol{h}'_{j-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \boldsymbol{h}_{s-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \right)^2$$

$$= O_p \left(\frac{1}{T^3} \sum_{t=1}^{T} \left(T^{1/2} \sum_{s=1}^{t} s^{-1/2-\theta} T^\theta \right)^2 \right) = O_p(1)$$

for $i = 1, 2, 3$, to conclude the proof of (80), and therefore that of (70) and (47) for $i = 1, 2, 3$.

6.1.7 Proof of (48) for $i = 2, 3$

As with (47) for $i = 1, 2, 3$, the proofs of (48) for $i = 2, 3$ are nearly identical, but with different normalizations. Recall $q_2 = 2(\delta_0 - \delta)$ and $q_3 = 1$. Then (48) for $i = 2, 3$ holds if

$$\Pr\left(\inf_{\mathcal{H}_i} \frac{1}{T^{q_i}}\left(\sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}) - 2\left|\sum_{t=1}^{T} d_t(\boldsymbol{\vartheta})s_t(\boldsymbol{\vartheta})\right| - \sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}_0)\right) \leq 0\right) \to 0 \text{ as } T \to \infty, \quad (84)$$

where

$$\sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}) = \sum_{t=1}^{T}(\rho(L; \boldsymbol{\varphi})u_t(\delta - \delta_0))^2$$

$$- \left\|\sum_{t=1}^{T} \rho(L; \boldsymbol{\varphi})u_t(\delta - \delta_0)\boldsymbol{h}_{t-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi})\right\|^2. \quad (85)$$

In the proof of their (2.7) for $i = 2, 3$, Hualde and Robinson (2011) showed that

$$\Pr\left(\inf_{\|\boldsymbol{\tau} - \boldsymbol{\tau}_0\| \geq \varepsilon, \boldsymbol{\tau} \in \mathcal{T}_i} \frac{1}{T^{q_i}} \sum_{t=1}^{T}(\rho(L; \boldsymbol{\varphi})u_t(\delta - \delta_0))^2 > K\right) \to 1 \text{ as } T \to \infty \quad (86)$$

for any arbitrarily large fixed constant K (for small enough η). Thus, in view of (85), (48) for $i = 2, 3$ holds by (86) and Lemma 1 (noting that for $i = 2, 3$ we have $T^{-q_i} \leq T^{-1}$) on showing

$$\sup_{\mathcal{H}_i} \frac{1}{T^{q_i/2}} \left\|\sum_{t=1}^{T} \rho(L; \boldsymbol{\varphi})u_t(\delta - \delta_0)\boldsymbol{h}_{t-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi})\right\| = O_p(1), \quad (87)$$

$$\sup_{\mathcal{H}_i} \frac{1}{T^{q_i}} \left|\sum_{t=1}^{T} d_t(\boldsymbol{\vartheta})s_t(\boldsymbol{\vartheta})\right| = \xi O_p(1). \quad (88)$$

We note here that, even if ξ had to be set large enough in the proof of (77) (see the proof of Lemma 3), this can be dominated by the constant K fixed in (86), which can be chosen arbitrarily large by setting η small enough.

By (185) of Lemma 18, (87) holds for $i = 3$ because $\delta_0 - \delta \leq 1/2$ when $\delta \in \mathcal{I}_3$. Furthermore, noting that $\sup_{\mathcal{H}_3} |\gamma - \gamma_0| \leq \xi T^{-\varkappa_3}$ and that $\delta \in \mathcal{I}_3$ implies $\gamma_0 - \delta \geq \gamma_0 - \delta_0 + 1/2 - \eta > 1$ and $\delta_0 - \delta \leq 1/2$, it follows by (147) and (151) of Lemma 4 that the left-hand side of (88) with $i = 3$ is $\xi O_p(1)$ by choosing $\theta < 1/2$.

Next, (87) with $i = 2$ follows by (187) of Lemma 18 choosing $\theta < 1/2$, and noting that $\sup_{\mathcal{H}_2} |\gamma - \gamma_0| \leq \xi T^{-\varkappa_2} = \xi T^{-(\gamma_0 - \delta_0 - 1/2)}$ and that $\delta \in \mathcal{I}_2$ implies $\gamma_0 - \delta \geq \gamma_0 - \delta_0 + 1/2 > 1$ and $\delta_0 - \delta \geq 1/2$. Finally, (88) with $i = 2$ follows from (147) and (152) of Lemma 4 setting $\theta < 1/2$, to conclude the proof of (48) for $i = 2, 3$.

6.1.8 Proof of (48) for $i = 1$

Finally we show (48) for $i = 1$, which holds if

$$\Pr\left(\inf_{\mathcal{H}_1} \frac{1}{T^{2(\delta_0 - \delta)}}\left(\sum_{t=1}^{T}(d_t(\boldsymbol{\vartheta}) + s_t(\boldsymbol{\vartheta}))^2 - \sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}_0)\right) \leq 0\right) \to 0 \quad (89)$$

as $T \to \infty$. By Lemma 1 (noting that for $i = 1$ we have $2(\delta_0 - \delta) \leq -1 - 2\eta < -1$) and Lemma 22 with $Z_t = d_t(\boldsymbol{\vartheta}) + s_t(\boldsymbol{\vartheta})$, (89) follows if there exists an $\epsilon > 0$ such that

$$\Pr\left(\inf_{\mathcal{H}_1} \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{T^{\delta_0-\delta+1/2}}\overline{d}_t(\boldsymbol{\vartheta}) + \frac{1}{T^{\delta_0-\delta+1/2}}\overline{s}_t(\boldsymbol{\vartheta})\right)^2 > \epsilon\right) \to 1 \tag{90}$$

as $T \to \infty$. Note that in $\mathcal{H}_1$, $\gamma_0 - \delta \geq 1 + \eta$ and $|\gamma - \gamma_0| < \xi T^{-\varkappa_1}$, so there exists $\alpha > 0$ such that for T sufficiently large it holds that $\gamma - \delta \geq 1 + \alpha$. Defining the sets $\mathcal{G}_1 = \{\boldsymbol{\vartheta} : \boldsymbol{\tau} \in \mathcal{T}_1, \gamma - \delta \geq 1 + \alpha, 1 - \delta \leq 1/2 + \alpha, \gamma \in [\square_1, \square_2]\}$ and $\mathcal{G}_2 = \{\boldsymbol{\vartheta} : \boldsymbol{\tau} \in \mathcal{T}_1, \gamma - \delta \geq 1 + \alpha, 1 - \delta \geq 1/2 + \alpha, \gamma \in [\square_1, \square_2]\}$, (90) is justified by showing

$$\Pr\left(\inf_{\mathcal{G}_1} \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{T^{\delta_0-\delta+1/2}}\overline{d}_t(\boldsymbol{\vartheta}) + \frac{1}{T^{\delta_0-\delta+1/2}}\overline{s}_t(\boldsymbol{\vartheta})\right)^2 > \epsilon\right) \to 1, \tag{91}$$

$$\Pr\left(\inf_{\mathcal{G}_2} \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{T^{\delta_0-\delta+1/2}}\overline{d}_t(\boldsymbol{\vartheta}) + \frac{1}{T^{\delta_0-\delta+1/2}}\overline{s}_t(\boldsymbol{\vartheta})\right)^2 > \epsilon\right) \to 1, \tag{92}$$

as $T \to \infty$.

First we show (91). Initially we justify that

$$\frac{1}{T^{\delta_0-\delta+1/2}}\overline{s}_t(\boldsymbol{\vartheta}) = \frac{1}{T^{\delta_0-\delta+1/2}}(a_{1t}(\boldsymbol{\vartheta}) + a_{2t}(\boldsymbol{\vartheta})), \tag{93}$$

where

$$a_{1t}(\boldsymbol{\vartheta}) = \rho(L;\boldsymbol{\varphi})u_t(\delta-\delta_0-1) - \frac{c_{t-1}(\gamma-\delta+1,\boldsymbol{\varphi})\sum_{j=1}^{T}\rho(L;\boldsymbol{\varphi})u_j(\delta-\delta_0)c_{j-1}(\gamma-\delta,\boldsymbol{\varphi})}{\sum_{j=1}^{T}c_{j-1}^2(\gamma-\delta,\boldsymbol{\varphi})} \tag{94}$$

and $a_{2t}(\boldsymbol{\vartheta})$ is a remainder term such that for any arbitrarily small $\epsilon_1 > 0$, $\epsilon_2 > 0$, and for T sufficiently large,

$$\Pr\left(\sup_{\mathcal{G}_1,t} \frac{1}{T^{\delta_0-\delta+1/2}}|a_{2t,T}(\boldsymbol{\vartheta})| > \epsilon_1\right) < \epsilon_2. \tag{95}$$

As in (72),

$$\overline{s}_t(\boldsymbol{\vartheta}) = \rho(L;\boldsymbol{\varphi})u_t(\delta-\delta_0-1) - \sum_{j=1}^{T}\rho(L;\boldsymbol{\varphi})u_j(\delta-\delta_0)\boldsymbol{c}_{j-1}(1-\delta,\gamma-\delta,\boldsymbol{\varphi})$$

$$\times\left(\sum_{k=1}^{T}\boldsymbol{c}_{k-1}(1-\delta,\gamma-\delta,\boldsymbol{\varphi})\boldsymbol{c}'_{k-1}(1-\delta,\gamma-\delta,\boldsymbol{\varphi})\right)^{-1}\boldsymbol{c}_{t-1}(2-\delta,\gamma-\delta+1,\boldsymbol{\varphi})$$

$$= \rho(L;\boldsymbol{\varphi})u_t(\delta-\delta_0-1)$$

$$- \frac{a_{11t}(\boldsymbol{\vartheta}) + a_{12t}(\boldsymbol{\vartheta}) + a_{13t}(\boldsymbol{\vartheta}) + a_{14t}(\boldsymbol{\vartheta})}{\sum_{j=1}^{T}c_{j-1}^2(1-\delta,\boldsymbol{\varphi})\sum_{j=1}^{T}c_{j-1}^2(\gamma-\delta,\boldsymbol{\varphi}) - \left(\sum_{j=1}^{T}c_{j-1}(1-\delta,\boldsymbol{\varphi})c_{j-1}(\gamma-\delta,\boldsymbol{\varphi})\right)^2},$$

where

$$a_{11t}(\boldsymbol{\vartheta}) = c_{t-1}(2-\delta,\boldsymbol{\varphi}) \sum_{j=1}^{T} \rho(L;\boldsymbol{\varphi}) u_j(\delta-\delta_0) c_{j-1}(1-\delta,\boldsymbol{\varphi}) \sum_{k=1}^{T} c_{k-1}^2(\gamma-\delta,\boldsymbol{\varphi}),$$

$$a_{12t}(\boldsymbol{\vartheta}) = -c_{t-1}(\gamma-\delta+1,\boldsymbol{\varphi}) \sum_{j=1}^{T} \rho(L;\boldsymbol{\varphi}) u_j(\delta-\delta_0) c_{j-1}(1-\delta,\boldsymbol{\varphi}) \sum_{k=1}^{T} c_{k-1}(1-\delta,\boldsymbol{\varphi}) c_{k-1}(\gamma-\delta,\boldsymbol{\varphi}),$$

$$a_{13t}(\boldsymbol{\vartheta}) = c_{t-1}(\gamma-\delta+1,\boldsymbol{\varphi}) \sum_{j=1}^{T} \rho(L;\boldsymbol{\varphi}) u_j(\delta-\delta_0) c_{j-1}(\gamma-\delta,\boldsymbol{\varphi}) \sum_{k=1}^{T} c_{k-1}^2(1-\delta,\boldsymbol{\varphi}),$$

$$a_{14t}(\boldsymbol{\vartheta}) = -c_{t-1}(2-\delta,\boldsymbol{\varphi}) \sum_{j=1}^{T} \rho(L;\boldsymbol{\varphi}) u_j(\delta-\delta_0) c_{j-1}(\gamma-\delta,\boldsymbol{\varphi}) \sum_{k=1}^{T} c_{k-1}(1-\delta,\boldsymbol{\varphi}) c_{k-1}(\gamma-\delta,\boldsymbol{\varphi}).$$

Then, it can be straightforwardly shown that (93)-(95) follow by (159), (160), (162), (163), (164), (166), (170), (195), and (196). Next, let [·] denote the integer part of the argument and consider $S_{1T}(r,\boldsymbol{\vartheta}) = T^{\delta-\delta_0-1/2} a_{1[Tr]}(\boldsymbol{\vartheta})$ a process indexed by $(r,\boldsymbol{\vartheta})$ that is càdlàg in r and continuous in $\boldsymbol{\vartheta}$. By almost identical techniques to those of Hualde and Nielsen (2020) it can be shown that

$$S_{1T}(r,\boldsymbol{\vartheta}) \Rightarrow S_1(r,\boldsymbol{\vartheta}), \tag{96}$$

where

$$\begin{aligned} S_1(r,\boldsymbol{\vartheta}) = {} & \rho(1;\boldsymbol{\varphi})\omega(1;\boldsymbol{\varphi}_0) W(r;1+\delta_0-\delta) \\ & - \frac{\rho(1;\boldsymbol{\varphi})\omega(1;\boldsymbol{\varphi}_0)(2(\gamma-\delta)-1)}{\gamma-\delta} r^{\gamma-\delta} W(1;1+\delta_0-\delta) \\ & + \frac{\rho(1;\boldsymbol{\varphi})\omega(1;\boldsymbol{\varphi}_0)(2(\gamma-\delta)-1)(\gamma-\delta-1)}{\gamma-\delta} r^{\gamma-\delta} \int_0^1 u^{\gamma-\delta-2} W(u;1+\delta_0-\delta) du \end{aligned}$$

and $\Rightarrow$ 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 σ_0^2. Because $\bar{d}_t(\boldsymbol{\vartheta})$ is deterministic and $\bar{s}_t(\boldsymbol{\vartheta})$ is stochastic, and in view of the square in (91) and (93)–(95), (91) follows from (96) (also note Assumption A1(iv) and (22)). Note that the application of Lemma 22 going from (89) to (90) is essential to avoid a very strong necessary moment condition (Johansen and Nielsen, 2012b).

Next we show (92), which is similar to that of (91). Defining $S_{2T}(r,\boldsymbol{\vartheta}) = T^{\delta-\delta_0-1/2} \bar{s}_{[Tr]}(\boldsymbol{\vartheta})$, we show

$$S_{2T}(r,\boldsymbol{\vartheta}) \Rightarrow S_2(r,\boldsymbol{\vartheta}), \tag{97}$$

where

$$\begin{aligned} S_2(r,\boldsymbol{\vartheta}) = {} & \rho(1;\boldsymbol{\varphi})\omega(1;\boldsymbol{\varphi}_0) W(r;1+\delta_0-\delta) - \rho(1;\boldsymbol{\varphi})\omega(1;\boldsymbol{\varphi}_0) \\ & \times \left((1,1) W(1;1+\delta_0-\delta) - \int_0^1 (-\delta s^{-1-\delta}, (\gamma-\delta-1) s^{\gamma-\delta-2}) W(s;1+\delta_0-\delta) ds \right) \\ & \times \begin{pmatrix} \frac{1}{2(1-\delta)-1} & \frac{1}{\gamma-2\delta} \\ \frac{1}{\gamma-2\delta} & \frac{1}{2(\gamma-\delta)-1} \end{pmatrix}^{-1} \begin{pmatrix} \frac{r^{1-\delta}}{1-\delta} \\ \frac{r^{\gamma-\delta}}{\gamma-\delta} \end{pmatrix}, \end{aligned}$$

and $\Rightarrow$ now means weak convergence in the product space of functions that are càdlàg in $r \in [0,1]$ and continuous in $\boldsymbol{\vartheta} \in \mathcal{G}_2$ endowed with the Skorokhod topology in r and the uniform topology in $\boldsymbol{\vartheta}$. The proof of (97) is almost identical to that of (96), noting that in $\mathcal{G}_2$, $1 - \delta \geq 1/2 + \alpha$. Also, for $\boldsymbol{\vartheta} \in \mathcal{G}_2$ it holds that $\int_0^1 s^{-1-\delta} W(s; 1 + \delta_0 - \delta) ds$ is a well-defined random variable with zero mean and finite variance (e.g., for $\delta = \delta_0$ this variance is $2\sigma_0^2 ((1-\delta)(2(1-\delta)-1))^{-1})$. As in the proof of (91), (97) justifies (92), which completes the proof of (90), and hence that of (89). Thus the proof of (48) for $i = 1$ is completed and therefore that of (32).

6.2 Proof of Theorem 1(ii): the $\gamma_0 - 1/2 < \delta_0$ case

Clearly

$$\Pr\left(\|\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0\| \geq \varepsilon\right) = \Pr\left(\inf_{\boldsymbol{\vartheta} \in \overline{M}_\varepsilon} R_T(\boldsymbol{\vartheta}) \leq \inf_{\boldsymbol{\vartheta} \in M_\varepsilon} R_T(\boldsymbol{\vartheta})\right), \tag{98}$$

so, as in the proof of part (i), the result follows by showing that the right-hand side of (45) is $o(1)$, which, in view of Lemma 1, holds if

$$\Pr\left(\inf_{\boldsymbol{\vartheta} \in \overline{M}_\varepsilon} R_T(\boldsymbol{\vartheta}) \leq \sigma_0^2 + \epsilon\right) \to 0 \text{ as } T \to \infty. \tag{99}$$

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 (99) follows on showing that

$$\Pr\left(\inf_{\mathcal{W}_i} R_T(\boldsymbol{\vartheta}) \leq \sigma_0^2 + \epsilon\right) \to 0 \text{ as } T \to \infty \text{ for } i = 1, \ldots, 4. \tag{100}$$

The proofs of (100) for each of $i = 1, \ldots, 4$ follows in the next subsections. Each proof proceeds by showing that, because $\gamma_0 - 1/2 < \delta_0$, the deterministic term $d_t(\boldsymbol{\vartheta})$ in $R_T(\boldsymbol{\vartheta})$ is negligible with the appropriate normalization. We can therefore take advantage of lower bounds on the stochastic terms in $R_T(\boldsymbol{\vartheta})$ established in Hualde and Robinson (2011), Johansen and Nielsen (2019), and Hualde and Nielsen (2020). Note again that Johansen and Nielsen (2019) require 8 moments, but this is not used to establish the bounds that we require.

6.2.1 Proof of (100) for $i = 4$

We first note that

$$\begin{aligned}
R_T(\boldsymbol{\vartheta}) &= \frac{1}{T} \sum_{t=1}^T (\rho(L; \boldsymbol{\varphi}) x_t(\delta))^2 - \frac{1}{T} \left\| \sum_{t=1}^T \rho(L; \boldsymbol{\varphi}) x_t(\delta) \boldsymbol{h}_{t-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \right\|^2 \\
&\geq \frac{1}{T} \sum_{t=1}^T (\rho(L; \boldsymbol{\varphi}) u_t(\delta - \delta_0))^2 - \frac{2 \|\boldsymbol{\phi}_0\|}{T} \left\| \sum_{t=1}^T \rho(L; \boldsymbol{\varphi}) u_t(\delta - \delta_0) \boldsymbol{c}_{t-1}(1 - \delta, \gamma_0 - \delta, \boldsymbol{\varphi}) \right\| \\
&\quad - \frac{1}{T} \left\| \sum_{t=1}^T \rho(L; \boldsymbol{\varphi}) x_t(\delta) \boldsymbol{h}_{t-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \right\|^2,
\end{aligned} \tag{101}$$

where the inequality follows from (34). Because $\delta_0 - \delta \leq 1/2 - \eta$ when $\delta \in \mathcal{I}_4 \cup \mathcal{I}_5$, $u_t(\delta - \delta_0)$ is asymptotically stationary. In view of (101), the proof of (100) for $i = 4$ then follows by

Hualde and Robinson (2011) (see the proof of their (2.7) for $i = 4$) by showing

$$\sup_{\mathcal{W}_4} \frac{1}{T} \left| \sum_{t=1}^{T} \rho(L; \boldsymbol{\varphi}) u_t(\delta - \delta_0) \boldsymbol{c}_{t-1}(1 - \delta, \gamma_0 - \delta, \boldsymbol{\varphi}) \right| = o_p(1), \tag{102}$$

$$\sup_{\mathcal{W}_4} \frac{1}{T} \left\| \sum_{t=1}^{T} \rho(L; \boldsymbol{\varphi}) x_t(\delta) \boldsymbol{h}_{t-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \right\|^2 = o_p(1). \tag{103}$$

First, because $\delta \in \mathcal{I}_4 \cup \mathcal{I}_5$ implies $\delta_0 - \delta \leq 1/2 - \eta$ and $\gamma_0 - \delta \leq 1/2 + \gamma_0 - \delta_0 - \eta$, (193) of Lemma 20 implies that the left-hand side of (102) is $O_p(T^{\gamma_0 - \delta_0 - 1/2 - 2\eta} + T^{-\varsigma - \eta} + T^{-1} \log T) = o_p(1)$.

Next, using (34) and noting that the second term in (31) dominates because $\gamma_0 \geq 1 + \kappa$, (103) follows by showing

$$\sup_{\mathcal{W}_4} T^{-1/2} \sum_{t=1}^{T} \rho(L; \boldsymbol{\varphi}) u_t(\delta - \delta_0) \boldsymbol{h}_{t-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) = o_p(1), \tag{104}$$

$$\sup_{\mathcal{W}_4} T^{-1/2} \sum_{t=1}^{T} \boldsymbol{c}_{t-1}(\gamma_0 - \delta, \boldsymbol{\varphi}) \boldsymbol{h}_{t-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) = o(1). \tag{105}$$

Here, (185) of Lemma 18 shows that the left-hand side of (104) is $O_p(T^{\theta - 1/2} + T^{-\eta}) = o_p(1)$ by choosing $\theta < 1/2$, while (191) of Lemma 19 shows that the left-hand side of (105) $O(T^{\theta - 1/2} + T^{\gamma_0 - 1/2 - \delta_0 - \eta}) = o(1)$, to conclude the proof of (100) for $i = 4$.

6.2.2 Proof of (100) for $i = 2, 3$

Recall the different normalizations given by $q_3 = 1$ and $q_2 = 2(\delta_0 - \delta)$. Clearly,

$$\Pr\left(\inf_{\mathcal{W}_i} R_T(\boldsymbol{\vartheta}) \leq \sigma_0^2 + \epsilon\right) \leq \Pr\left(\inf_{\mathcal{W}_i} \frac{T^{q_i}}{T} \inf_{\mathcal{W}_i} \frac{T}{T^{q_i}} R_T(\boldsymbol{\vartheta}) \leq \sigma_0^2 + \epsilon\right)$$

$$= \Pr\left(\inf_{\mathcal{W}_i} \frac{T}{T^{q_i}} R_T(\boldsymbol{\vartheta}) \leq \sigma_0^2 + \epsilon\right). \tag{106}$$

Thus, in view of (86) and (101), the proof of (100) for $i = 2, 3$ follows on showing

$$\sup_{\mathcal{W}_i} \frac{1}{T^{q_i}} \left| \sum_{t=1}^{T} \rho(L; \boldsymbol{\varphi}) u_t(\delta - \delta_0) \boldsymbol{c}_{t-1}(1 - \delta, \gamma_0 - \delta, \boldsymbol{\varphi}) \right| = O_p(1), \tag{107}$$

$$\sup_{\mathcal{W}_i} \frac{1}{T^{q_i}} \left\| \sum_{t=1}^{T} \rho(L; \boldsymbol{\varphi}) x_t(\delta) \boldsymbol{h}_{t-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \right\|^2 = O_p(1). \tag{108}$$

For $i = 3$, both (107) and (108) follow straightforwardly by identical steps as those given in the proofs of (102) and (103) just replacing η by 0. For $i = 2$, we need to take into account the different normalization, which implies using (192) instead of (191) in Lemma 19, (194) instead of (193) in Lemma 20, and (187) instead of (185) in Lemma 18.

6.2.3 Proof of (100) for $i = 1$

Following identical steps to those given in (106),

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

so the required result follows on showing

$$\Pr\left(\inf_{\mathcal{W}_1} \frac{T}{T^{2(\delta_0-\delta)}} R_T(\boldsymbol{\vartheta}) > \epsilon\right) \to 1 \text{ as } T \to \infty. \tag{109}$$

First we show that

$$\frac{T}{T^{2(\delta_0-\delta)}} R_T(\boldsymbol{\vartheta}) \geq \frac{1}{T^{2(\delta_0-\delta)}} \sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}) + q_{1T}(\boldsymbol{\vartheta}), \tag{110}$$

where $\sup_{\mathcal{W}_1} |q_{1T}(\boldsymbol{\vartheta})| = o_p(1)$. Noting (35) and (39), the proof of (110) follows by showing

$$\sup_{\mathcal{W}_1} \frac{1}{T^{2(\delta_0-\delta)}} \left\| \sum_{t=1}^{T} \rho(L;\boldsymbol{\varphi}) u_t(\delta-\delta_0) c_{t-1}(\gamma_0-\delta,\boldsymbol{\varphi}) \right\| = o_p(1)$$

and

$$\sup_{\mathcal{W}_1} \frac{1}{T^{2(\delta_0-\delta)}} \left\| \sum_{j=1}^{T} c_{j-1}(\gamma_0-\delta,\boldsymbol{\varphi}) \boldsymbol{h}'_{j-1,T}(1-\delta,\gamma-\delta,\boldsymbol{\varphi}) \right\|$$
$$\times \left\| \sum_{t=1}^{T} \rho(L;\boldsymbol{\varphi}) u_t(\delta-\delta_0) \boldsymbol{h}_{t-1,T}(1-\delta,\gamma-\delta,\boldsymbol{\varphi}) \right\| = o_p(1).$$

Because $\delta_0 - \delta \geq 1/2 + \eta$ and $\gamma_0 - \delta \geq 1/2 + \eta + \gamma_0 - \delta_0$ on $\mathcal{W}_1$, these results follow straightforwardly by (187), (192), and (194), noting that also $\gamma_0 - 1/2 - \delta_0 < 0$.

Next, let $\alpha > 0$ be arbitrarily small (in particular $\alpha < (\varsigma - 1/2)/3$) and define the sets $\Phi_1, \ldots, \Phi_9$ as in the proof of Lemma 2. Then, in view of (110), (109) follows on showing

$$\Pr\left(\inf_{\mathcal{W}_1 \cap \Phi_j} \frac{1}{T^{2(\delta_0-\delta)}} \sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}) > \epsilon\right) \to 1 \text{ as } T \to \infty \tag{111}$$

for $j = 1, ..., 9$.

First we prove (111) for $j = 1, 2, 4, 5$. In their eqn. (S.126), Hualde and Nielsen (2020) showed that

$$\Pr\left(\inf_{\|\boldsymbol{\tau}-\boldsymbol{\tau}_0\|\geq\varepsilon, \boldsymbol{\tau}\in\mathcal{T}_1} \frac{1}{T^{2(\delta_0-\delta)}} \sum_{t=1}^{T} (\rho(L;\boldsymbol{\varphi}) u_t(\delta-\delta_0))^2 > \epsilon\right) \to 1 \text{ as } T \to \infty. \tag{112}$$

Then, in view of (85) and (112), (111) for $j = 1, 2, 4, 5$ holds if we show that

$$\sup_{\mathcal{W}_1 \cap \Phi_j} \frac{1}{T^{2(\delta_0-\delta)}} \left\| \sum_{t=1}^{T} \rho(L;\boldsymbol{\varphi}) u_t(\delta-\delta_0) \boldsymbol{h}_{t-1,T}(1-\delta,\gamma-\delta,\boldsymbol{\varphi}) \right\|^2 = o_p(1). \tag{113}$$

We cannot apply Lemma 18 to (113) because it only shows $O_p(1)$. Instead, we decompose as

$$\frac{1}{T^{2(\delta_0-\delta)}} \left\| \sum_{t=1}^{T} \rho(L;\boldsymbol{\varphi}) u_t(\delta-\delta_0) \boldsymbol{h}_{t-1,T}(1-\delta,\gamma-\delta,\boldsymbol{\varphi}) \right\|^2 = B_{1T} + B_{2T} + B_{3T},$$

where

$$B_{1T} = \frac{\left(\sum_{j=1}^T \rho(L;\boldsymbol{\varphi})u_j(\delta - \delta_0)c_{j-1}(\gamma - \delta, \boldsymbol{\varphi})\right)^2 \sum_{j=1}^T c_{j-1}^2(1 - \delta, \boldsymbol{\varphi})}{\sum_{j=1}^T c_{j-1}^2(1 - \delta, \boldsymbol{\varphi}) \sum_{j=1}^T c_{j-1}^2(\gamma - \delta, \boldsymbol{\varphi}) - \left(\sum_{j=1}^T c_{j-1}(1 - \delta, \boldsymbol{\varphi})c_{j-1}(\gamma - \delta, \boldsymbol{\varphi})\right)^2},$$

(114)

$$B_{2T} = \frac{\left(\sum_{j=1}^T \rho(L;\boldsymbol{\varphi})u_j(\delta - \delta_0)c_{j-1}(1 - \delta, \boldsymbol{\varphi})\right)^2 \sum_{j=1}^T c_{j-1}^2(\gamma - \delta, \boldsymbol{\varphi})}{\sum_{j=1}^T c_{j-1}^2(1 - \delta, \boldsymbol{\varphi}) \sum_{j=1}^T c_{j-1}^2(\gamma - \delta, \boldsymbol{\varphi}) - \left(\sum_{j=1}^T c_{j-1}(1 - \delta, \boldsymbol{\varphi})c_{j-1}(\gamma - \delta, \boldsymbol{\varphi})\right)^2},$$

(115)

$$B_{3T} = \frac{-2\sum_{j=1}^T \rho(L;\boldsymbol{\varphi})u_j(\delta - \delta_0)c_{j-1}(1 - \delta, \boldsymbol{\varphi}) \sum_{j=1}^T \rho(L;\boldsymbol{\varphi})u_j(\delta - \delta_0)c_{j-1}(\gamma - \delta, \boldsymbol{\varphi})}{\sum_{j=1}^T c_{j-1}^2(1 - \delta, \boldsymbol{\varphi}) \sum_{j=1}^T c_{j-1}^2(\gamma - \delta, \boldsymbol{\varphi}) - \left(\sum_{j=1}^T c_{j-1}(1 - \delta, \boldsymbol{\varphi})c_{j-1}(\gamma - \delta, \boldsymbol{\varphi})\right)^2}$$

$$\times \sum_{j=1}^T c_{j-1}(1 - \delta, \boldsymbol{\varphi})c_{j-1}(\gamma - \delta, \boldsymbol{\varphi}),$$

(116)

so (113) holds for $j = 1, 2, 4, 5$ if

$$\sup_{\mathcal{W}_1 \cap \Phi_j} |B_{iT}| = o_p(1) \quad \text{for } i = 1, 2, 3 \text{ and } j = 1, 2, 4, 5.$$

The proof of (113) for $j = 1$ follows straightforwardly by (165), (171), and (195). Next, we show (113) for $j = 2$. By (173) the normalized denominator can be made arbitrarily large by setting K large enough, so the result follows on showing that the supremum of the numerators in (114)–(116) is $O_p(1)$. This result in turn follows by (165) and (196) for B_{1T}, (168) and (195) for B_{2T}, and (167), (195), and (196) for B_{3T}. The proof of (113) for $j = 4$ is identical to that for $j = 2$ using symmetry of Φ_2 and Φ_4, and is therefore omitted. Regarding (113) for $j = 5$, we re-normalize as

$$B_{1T} = \frac{T^{1-2(\gamma-\delta)} \left(\sum_{j=1}^T \rho(L;\boldsymbol{\varphi})u_j(\delta - \delta_0)c_{j-1}(\gamma - \delta, \boldsymbol{\varphi})\right)^2}{T^{1-2(\gamma-\delta)} \left(\sum_{j=1}^T c_{j-1}^2(\gamma - \delta, \boldsymbol{\varphi}) - \frac{\left(\sum_{j=1}^T c_{j-1}(1-\delta,\boldsymbol{\varphi})c_{j-1}(\gamma-\delta,\boldsymbol{\varphi})\right)^2}{\sum_{j=1}^T c_{j-1}^2(1-\delta,\boldsymbol{\varphi})}\right)}.$$

(117)

By (174) the infimum of the denominator of (117) can be made arbitrarily large, whereas by (196) the supremum of the numerator is $O_p(1)$, so that $\sup_{\mathcal{W}_1 \cap \Phi_5} |B_{1T}| = o_p(1)$. The proof of $\sup_{\mathcal{W}_1 \cap \Phi_5} |B_{2T}| = o_p(1)$ is essentially identical and thus omitted, whereas that of $\sup_{\mathcal{W}_1 \cap \Phi_4} |B_{3T}| = o_p(1)$ uses (170) along with (174) and (196).

Finally, the proof of (111) for $j = 3, 6, 7, 8, 9$ is essentially identical to that of (90) with $\bar{d}_t(\boldsymbol{\vartheta}) = 0$. 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 $\gamma - \delta \geq 1 + \alpha$). 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 $\int_0^1 u^{\gamma - \delta - 2} W(u; 1 + \delta_0 - \delta) du$ is well defined. This completes the proof of (111) for $j = 3, 6, 7, 8, 9$, and therefore that of (100) for $i = 1$.

6.3 Proof of Theorem 2(i.a): the $\gamma_0 - 1/2 > \delta_0$ case with $\delta_0 < 1/2$

Defining $\boldsymbol{M}_T = \text{diag}(\boldsymbol{I}_{p+1}, T^{\delta_0 - \gamma_0 + 1})$ we initially show that

$$T^{1/2}\boldsymbol{M}_T^{-1}(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) \to_d N(\boldsymbol{0}_{p+2}, \sigma_0^2 \boldsymbol{V}_{\boldsymbol{\vartheta}}^{-1}) \text{ with } \boldsymbol{V}_{\boldsymbol{\vartheta}} = \begin{pmatrix} \sigma_0^2 \boldsymbol{A} & \boldsymbol{0}_{p+1} \\ \boldsymbol{0}'_{p+1} & \frac{(\gamma_0 - 1)^2}{(\gamma_0 - 2\delta_0)^2} v_\beta \end{pmatrix}. \tag{118}$$

By the mean value theorem,

$$\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0 = - \left(\frac{\partial^2 R_T(\overline{\boldsymbol{\vartheta}})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \right)^{-1} \frac{\partial R_T(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}}, \tag{119}$$

where $\overline{\boldsymbol{\vartheta}}$ represents an intermediate point which is allowed to vary across the different rows of $\partial^2 R_T(\cdot)/\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'$. We first analyze the score in (119). It can be straightforwardly 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 (40),

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

Then, by Lemma 5(a) it holds that

$$\frac{T^{1/2}}{2} \boldsymbol{M}_T \frac{\partial R_T(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} = \begin{pmatrix} T^{-1/2}\boldsymbol{I}_{p+1} & 0 \\ 0 & T^{1/2-(\gamma_0-\delta_0)} \end{pmatrix} \sum_{t=1}^T \varepsilon_t \begin{pmatrix} \frac{\partial s_{1t}(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\tau}} \\ \frac{\partial d_t(\boldsymbol{\vartheta}_0)}{\partial \gamma} \end{pmatrix} + o_p(1). \tag{120}$$

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

$$\frac{\partial d_t(\boldsymbol{\vartheta}_0)}{\partial \gamma} = -\beta_0 c_{t-1}^{(1)}(\gamma_0 - \delta_0, \boldsymbol{\varphi}_0)$$

$$+ \beta_0 \boldsymbol{h}'_{t-1,T}(1 - \delta_0, \gamma_0 - \delta_0, \boldsymbol{\varphi}_0) \sum_{j=1}^T \boldsymbol{h}_{t-1,T}(1 - \delta_0, \gamma_0 - \delta_0, \boldsymbol{\varphi}_0)c_{j-1}^{(1)}(\gamma_0 - \delta_0, \boldsymbol{\varphi}_0),$$

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

$$\frac{1}{T^{\gamma_0 - \delta_0 - 1/2}} \sum_{t=1}^T \varepsilon_t \frac{\partial d_t(\boldsymbol{\vartheta}_0)}{\partial \gamma} = \frac{\beta_0}{T^{\gamma_0 - \delta_0 - 1/2}} \sum_{t=1}^T \varepsilon_t \boldsymbol{h}'_{t-1,T}(1 - \delta_0, \gamma_0 - \delta_0, \boldsymbol{\varphi}_0)$$

$$\times \sum_{j=1}^T \boldsymbol{h}_{t-1,T}(1 - \delta_0, \gamma_0 - \delta_0, \boldsymbol{\varphi}_0)c_{j-1}^{(1)}(\gamma_0 - \delta_0, \boldsymbol{\varphi}_0)$$

$$- \frac{\beta_0}{T^{\gamma_0 - \delta_0 - 1/2}} \sum_{t=1}^T \varepsilon_t c_{t-1}^{(1)}(\gamma_0 - \delta_0, \boldsymbol{\varphi}_0). \tag{121}$$

By (17), $c_{t-1}^{(1)}(d, \boldsymbol{\varphi}) = \sum_{j=0}^{t-1} \rho_j(\boldsymbol{\varphi}) \pi_{t-j-1}^{(1)}(d)$, where $\pi_j^{(1)}(\cdot)$ is the first derivative of $\pi_j(\cdot)$ given by

$$\pi_j^{(1)}(d) = (\psi(d+j) - \psi(d))\pi_j(d), \tag{122}$$

with $\psi(\cdot)$ denoting the digamma function. Then, noting that $1 - \delta_0 > 1/2$ and $\gamma_0 - \delta_0 > 1/2$, by a similar analysis to that in the proof of Lemma 13, the right-hand side of (121) equals

$$\frac{\beta_0 \rho(1; \boldsymbol{\varphi}_0)}{T^{\gamma_0 - \delta_0 - 1/2}} \sum_{t=1}^T \varepsilon_t \left(\begin{array}{cc} \pi_{t-1}(1 - \delta_0) & \pi_{t-1}(\gamma_0 - \delta_0) \end{array} \right)$$

$$\times \left(\begin{array}{cc} \sum_{t=1}^T \pi_{t-1}^2(1-\delta_0) & \sum_{t=1}^T \pi_{t-1}(1-\delta_0)\pi_{t-1}(\gamma_0-\delta_0) \\ \sum_{t=1}^T \pi_{t-1}(1-\delta_0)\pi_{t-1}(\gamma_0-\delta_0) & \sum_{t=1}^T \pi_{t-1}^2(\gamma_0-\delta_0) \end{array} \right)^{-1}$$

$$\times \sum_{j=2}^T \left(\begin{array}{c} \pi_{j-1}(1-\delta_0) \\ \pi_{j-1}(\gamma_0-\delta_0) \end{array} \right) \pi_{j-1}^{(1)}(\gamma_0-\delta_0) - \frac{\beta_0\rho(1;\boldsymbol{\varphi}_0)}{T^{\gamma_0-\delta_0-1/2}} \sum_{t=2}^T \varepsilon_t \pi_{t-1}^{(1)}(\gamma_0-\delta_0) + o_p(1). \tag{123}$$

Using (122) and the approximation $\psi(z) = \log(z) + O(z^{-1})$, see Abramowitz and Stegun (1970, p. 259, eqn. 6.3.18), as well as careful cancellation of terms, it can be shown that the first two terms of (123) equal $\beta_0 \rho(1; \boldsymbol{\varphi}_0) T^{1/2+\delta_0-\gamma_0} \sum_{t=2}^T \varepsilon_t g_{t,T}(1-\delta_0, \gamma_0-\delta_0) + o_p(1)$, where

$$g_{t,T}(d_1, d_2) = \left(\begin{array}{cc} \pi_{t-1}(d_1) & \pi_{t-1}(d_2) \end{array} \right) \left(\begin{array}{cc} \sum_{t=1}^T \pi_{t-1}^2(d_1) & \sum_{t=1}^T \pi_{t-1}(d_1)\pi_{t-1}(d_2) \\ \sum_{t=1}^T \pi_{t-1}(d_1)\pi_{t-1}(d_2) & \sum_{t=1}^T \pi_{t-1}^2(d_2) \end{array} \right)^{-1}$$

$$\times \sum_{j=2}^T \log\left(\frac{d_2 + j - 1}{T} \right) \left(\begin{array}{c} \pi_{j-1}(d_1) \\ \pi_{j-1}(d_2) \end{array} \right) \pi_{j-1}(d_2) - \log\left(\frac{d_2 + t - 1}{T} \right) \pi_{t-1}(d_2).$$

Collecting these terms shows that

$$\frac{T^{1/2}}{2} \boldsymbol{M}_T \frac{\partial R_T(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} = \sum_{t=2}^T \varepsilon_t \boldsymbol{\eta}_{t,T} + o_p(1), \tag{124}$$

where

$$\boldsymbol{\eta}_{t,T} = \left(\begin{array}{c} T^{-1/2} \sum_{j=1}^\infty \boldsymbol{m}_j(\boldsymbol{\varphi}_0)\varepsilon_{t-j} \\ T^{1/2-(\gamma_0-\delta_0)}\beta_0\rho(1;\boldsymbol{\varphi}_0)g_{t,T}(1-\delta_0, \gamma_0-\delta_0) \end{array} \right).$$

Since $\varepsilon_t \boldsymbol{\eta}_{t,T}$ is a martingale difference sequence, we can apply the martingale central limit theorem (e.g., Corollary 3.1 of Hall and Heyde, 1980) to show that

$$\sum_{t=2}^T \varepsilon_t \boldsymbol{\eta}_{t,T} \to_d N(\boldsymbol{0}_{p+2}, \sigma_0^2 \boldsymbol{V}_{\boldsymbol{\vartheta}}). \tag{125}$$

To prove (125) we note in particular that $\sum_{j=1}^\infty \boldsymbol{m}_j(\boldsymbol{\varphi}_0)\boldsymbol{m}_j(\boldsymbol{\varphi}_0)' = \boldsymbol{A}$ by definition, see Assumption A4(iii), and from Lemma 10 and (158) of Lemma 11, we find that, as $T \to \infty$,

$$\frac{1}{T^{2(\gamma_0-\delta_0)-1}} \sum_{t=2}^T g_{t,T}^2(1-\delta_0, \gamma_0-\delta_0) \to \frac{1}{\Gamma^2(\gamma_0-\delta_0)} \left(\frac{2}{(2(\gamma_0-\delta_0)-1)^3} - \left(\begin{array}{cc} \frac{1}{(\gamma_0-2\delta_0)^2} & \frac{1}{(2(\gamma_0-\delta_0)-1)^2} \end{array} \right) \right.$$

$$\left. \times \left(\begin{array}{cc} \frac{1}{1-2\delta_0} & \frac{1}{\gamma_0-2\delta_0} \\ \frac{1}{\gamma_0-2\delta_0} & \frac{1}{2(\gamma_0-\delta_0)-1} \end{array} \right)^{-1} \left(\begin{array}{c} \frac{1}{(\gamma_0-2\delta_0)^2} \\ \frac{1}{(2(\gamma_0-\delta_0)-1)^2} \end{array} \right) \right)$$

$$= \frac{(\gamma_0-1)^2}{\Gamma^2(\gamma_0-\delta_0)(\gamma_0-2\delta_0)^2(2(\gamma_0-\delta_0)-1)^3}. \tag{126}$$

In addition,

$$\frac{1}{T^{\gamma_0 - \delta_0}} \sum_{t=2}^{T} g_{t,T}(1 - \delta_0, \gamma_0 - \delta_0) \sum_{j=1}^{\infty} \boldsymbol{m}_j(\boldsymbol{\varphi}_0)\varepsilon_{t-j} = o_p(1), \tag{127}$$

which, noting that

$$\sum_{k=1}^{t} \sum_{j=1}^{\infty} \boldsymbol{m}_j(\boldsymbol{\varphi}_0)\varepsilon_{k-j} = O_p(t^{1/2}), \tag{128}$$

follows by summation by parts using Lemmas 10 and 11, and shows block-diagonality of the variance matrix, $\boldsymbol{V}_{\boldsymbol{\vartheta}}$. Thus, $\sum_{t=2}^{T} \boldsymbol{\eta}_{t,T}\boldsymbol{\eta}'_{t,T} \to_p \boldsymbol{V}_{\boldsymbol{\vartheta}}$ as required in (125). Because the fourth moment of ε_t is finite, the (conditional) Lindeberg condition holds if, for example, $\sum_{t=2}^{T}(\boldsymbol{\xi}'\boldsymbol{\eta}_{t,T}\boldsymbol{\eta}'_{t,T}\boldsymbol{\xi})^2 \to_p 0$ for an arbitrary conforming vector $\boldsymbol{\xi}$, which follows straightforwardly by previous arguments.

In view of (119), (124), and (125), the proof of (118) is completed by showing that

$$\boldsymbol{M}_T \left(\frac{\partial^2 R_T(\overline{\boldsymbol{\vartheta}})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} - \frac{\partial^2 R_T(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \right) \boldsymbol{M}_T = o_p(1) \text{ and } \frac{1}{2}\boldsymbol{M}_T \frac{\partial^2 R_T(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \boldsymbol{M}_T \to_p \boldsymbol{V}_{\boldsymbol{\vartheta}}. \tag{129}$$

The first result in (129) is relatively straightforward due to Lemma 6. It is proven by a mean value expansion, noting that the derivatives add at most a multiplicative logarithmic factor, see Lemma 11, but that is more than compensated by the factor $\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0 = O_p(T^{-\chi})$, $\chi > 0$, from Lemma 6. The second result in (129) can be proven by identical methods to the corresponding proof in Hualde and Nielsen (2020, online supplement p. 16), and specifically follows by application of Lemma 5(b) and the same methods as applied in (126)–(127).

Given (118), the remaining part of (26) is justified as follows. From (18) and (34) we find that

$$\widehat{\boldsymbol{\phi}} = \widehat{\boldsymbol{\phi}}(\widehat{\boldsymbol{\vartheta}}) = \sum_{t=1}^{T} \boldsymbol{k}_{t-1,T}(1 - \widehat{\delta}, \widehat{\gamma} - \widehat{\delta}, \widehat{\boldsymbol{\varphi}})\boldsymbol{c}'_{t-1}(1 - \widehat{\delta}, \gamma_0 - \widehat{\delta}, \widehat{\boldsymbol{\varphi}})\boldsymbol{\phi}_0$$
$$+ \sum_{t=1}^{T} \boldsymbol{k}_{t-1,T}(1 - \widehat{\delta}, \widehat{\gamma} - \widehat{\delta}, \widehat{\boldsymbol{\varphi}})\rho(L; \widehat{\boldsymbol{\varphi}})u_t(\widehat{\delta} - \delta_0), \tag{130}$$

where

$$\boldsymbol{k}_{t,T}(d_1, d_2, \boldsymbol{\varphi}) = \left(\sum_{j=1}^{T} \boldsymbol{c}_{j-1}(d_1, d_2, \boldsymbol{\varphi})\boldsymbol{c}'_{j-1}(d_1, d_2, \boldsymbol{\varphi}) \right)^{-1} \boldsymbol{c}_t(d_1, d_2, \boldsymbol{\varphi}).$$

By the mean value theorem,

$$\boldsymbol{k}_{t-1,T}(1 - \widehat{\delta}, \widehat{\gamma} - \widehat{\delta}, \widehat{\boldsymbol{\varphi}}) = \boldsymbol{k}_{t-1,T}(1 - \widehat{\delta}, \gamma_0 - \widehat{\delta}, \widehat{\boldsymbol{\varphi}}) + \boldsymbol{k}^{(2)}_{t-1,T}(1 - \widehat{\delta}, \overline{\gamma} - \widehat{\delta}, \widehat{\boldsymbol{\varphi}})(\widehat{\gamma} - \gamma_0),$$

where $\boldsymbol{k}^{(2)}_{t,T}(\cdot, \cdot, \cdot)$ is the derivative of $\boldsymbol{k}_{t,T}(\cdot, \cdot, \cdot)$ with respect to the second argument and $|\overline{\gamma} - \gamma_0| \leq |\widehat{\gamma} - \gamma_0|$. Thus, defining $N_T = \mathrm{diag}(T^{1/2-\delta_0}, T^{\gamma_0-\delta_0-1/2}/\log T)$,

$$N_T(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0) = (\widehat{\gamma} - \gamma_0)N_T \sum_{t=1}^{T} \boldsymbol{k}^{(2)}_{t-1,T}(1 - \widehat{\delta}, \overline{\gamma} - \widehat{\delta}, \widehat{\boldsymbol{\varphi}})\boldsymbol{c}'_{t-1}(1 - \widehat{\delta}, \gamma_0 - \widehat{\delta}, \widehat{\boldsymbol{\varphi}})\boldsymbol{\phi}_0$$
$$+ N_T \sum_{t=1}^{T} \boldsymbol{k}_{t-1,T}(1 - \widehat{\delta}, \widehat{\gamma} - \widehat{\delta}, \widehat{\boldsymbol{\varphi}})\rho(L; \widehat{\boldsymbol{\varphi}})u_t(\widehat{\delta} - \delta_0). \tag{131}$$

First, by application of Lemmas 6 and 7 it follows straightforwardly that the first term on the right-hand side of (131) equals

$$
(\widehat{\gamma} - \gamma_0)N_T \sum_{t=1}^{T} \boldsymbol{k}^{(2)}_{t-1,T}(1 - \delta_0, \gamma_0 - \delta_0, \boldsymbol{\varphi}_0)\boldsymbol{c}'_{t-1}(1 - \delta_0, \gamma_0 - \delta_0, \boldsymbol{\varphi}_0)\boldsymbol{\phi}_0 + o_p(1). \tag{132}
$$

Then, noting that

$$
\frac{\partial \boldsymbol{k}_{t,T}(d_1, d_2, \boldsymbol{\varphi})}{\partial d_2} = -\left(\sum_{j=1}^{T} \boldsymbol{c}_{j-1}(d_1, d_2, \boldsymbol{\varphi})\boldsymbol{c}'_{j-1}(d_1, d_2, \boldsymbol{\varphi}) \right)^{-1} \left(\frac{\partial}{\partial d_2} \sum_{j=1}^{T} \boldsymbol{c}_{j-1}(d_1, d_2, \boldsymbol{\varphi})\boldsymbol{c}'_{j-1}(d_1, d_2, \boldsymbol{\varphi}) \right)
$$

$$
\times \left(\sum_{j=1}^{T} \boldsymbol{c}_{j-1}(d_1, d_2, \boldsymbol{\varphi})\boldsymbol{c}'_{j-1}(d_1, d_2, \boldsymbol{\varphi}) \right)^{-1} \boldsymbol{c}_t(d_1, d_2, \boldsymbol{\varphi})
$$

$$
+ \left(\sum_{j=1}^{T} \boldsymbol{c}_{j-1}(d_1, d_2, \boldsymbol{\varphi})\boldsymbol{c}'_{j-1}(d_1, d_2, \boldsymbol{\varphi}) \right)^{-1} \left(\begin{array}{c} 0 \\ c_t^{(1)}(d_2, \boldsymbol{\varphi}) \end{array} \right),
$$

(132) is

$$
- (\widehat{\gamma} - \gamma_0)N_T \left(\sum_{j=1}^{T} \boldsymbol{c}_{j-1}(1 - \delta_0, \gamma_0 - \delta_0, \boldsymbol{\varphi}_0)\boldsymbol{c}'_{j-1}(1 - \delta_0, \gamma_0 - \delta_0, \boldsymbol{\varphi}_0) \right)^{-1}
$$

$$
\times \sum_{j=1}^{T} \boldsymbol{c}_{j-1}(1 - \delta_0, \gamma_0 - \delta_0, \boldsymbol{\varphi}_0) \left(\begin{array}{cc} 0 & c_t^{(1)}(\gamma_0 - \delta_0, \boldsymbol{\varphi}_0) \end{array} \right) \boldsymbol{\phi}_0 + o_p(1)
$$

$$
= -(\widehat{\gamma} - \gamma_0)N_T \left(\sum_{j=1}^{T} \boldsymbol{c}_{j-1}(1 - \delta_0, \gamma_0 - \delta_0, \boldsymbol{\varphi}_0)\boldsymbol{c}'_{j-1}(1 - \delta_0, \gamma_0 - \delta_0, \boldsymbol{\varphi}_0) \right)^{-1}
$$

$$
\times \sum_{j=1}^{T} \boldsymbol{c}_{j-1}(1 - \delta_0, \gamma_0 - \delta_0, \boldsymbol{\varphi}_0) \left(\begin{array}{cc} c_{j-1}(1 - \delta_0, \boldsymbol{\varphi}_0) & c_{j-1}^{(1)}(\gamma_0 - \delta_0, \boldsymbol{\varphi}_0) \end{array} \right) \left(\begin{array}{cc} 0 & 0 \\ 0 & 1 \end{array} \right) \boldsymbol{\phi}_0 + o_p(1). \tag{133}
$$

Now we note that by (122) it can be straightforwardly shown that $c_{j-1}^{(1)}(d, \boldsymbol{\varphi}) = c_{j-1}(d, \boldsymbol{\varphi})(\log j)(1 + o(1))$ and apply summation by parts,

$$
\sum_{j=1}^{T} c_{j-1}(d_1, \boldsymbol{\varphi})c_{j-1}^{(1)}(d_2, \boldsymbol{\varphi}) = \sum_{j=1}^{T} c_{j-1}(d_1, \boldsymbol{\varphi})c_{j-1}(d_2, \boldsymbol{\varphi})(\log j)(1 + o(1))
$$

$$
= (\log T) \sum_{j=1}^{T} c_{j-1}(d_1, \boldsymbol{\varphi})c_{j-1}(d_2, \boldsymbol{\varphi})(1 + o(1))
$$

$$
- \sum_{k=1}^{T-1} (\log(k + 1) - \log k) \sum_{j=1}^{k} c_{j-1}(d_1, \boldsymbol{\varphi})c_{j-1}(d_2, \boldsymbol{\varphi})(1 + o(1)).
$$

Because $\log(k+1) - \log k = \log(1 + k^{-1}) = k^{-1} + O(k^{-2})$, see Abramowitz and Stegun (1970, p. 68, eqn. 4.1.24), we find that $\sum_{j=1}^{T} c_{j-1}(d_1, \boldsymbol{\varphi})c_{j-1}^{(1)}(d_2, \boldsymbol{\varphi}) = (\log T) \sum_{j=1}^{T} c_{j-1}(d_1, \boldsymbol{\varphi})c_{j-1}(d_2, \boldsymbol{\varphi})(1+$

$o(1)$), see also Lemma 13. Inserting this result into (133), we find that (133) is

$$-(\widehat{\gamma}-\gamma_0)\begin{pmatrix} T^{1/2-\delta_0} & 0 \\ 0 & T^{\gamma_0-\delta_0-1/2} \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\phi_0+o_p(1) = -\begin{pmatrix} 0 \\ T^{\gamma_0-\delta_0-1/2}(\widehat{\gamma}-\gamma_0)\beta_0 \end{pmatrix}+o_p(1).$$
(134)

Next, proceeding as in the proof of (123), using Lemmas 6, 7, 10, 11, and 13, it can be straightforwardly shown that the second term on the right-hand side of (131) is

$$\begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\log T} \end{pmatrix}\sum_{t=1}^{T}\varepsilon_t\boldsymbol{\theta}_{t,T}+o_p(1) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\sum_{t=1}^{T}\varepsilon_t\boldsymbol{\theta}_{t,T}+o_p(1),$$
(135)

where

$$\boldsymbol{\theta}_{t,T} = \frac{1}{\rho(1;\boldsymbol{\varphi}_0)}\begin{pmatrix} \frac{1}{\Gamma^2(1-\delta_0)(1-2\delta_0)} & \frac{1}{\Gamma(1-\delta_0)\Gamma(\gamma_0-\delta_0)(\gamma_0-2\delta_0)} \\ \frac{1}{\Gamma(1-\delta_0)\Gamma(\gamma_0-\delta_0)(\gamma_0-2\delta_0)} & \frac{1}{\Gamma^2(\gamma_0-\delta_0)(2(\gamma_0-\delta_0)-1)} \end{pmatrix}^{-1}\begin{pmatrix} \frac{\pi_{t-1}(1-\delta_0)}{T^{1/2-\delta_0}} \\ \frac{\pi_{t-1}(\gamma_0-\delta_0)}{T^{\gamma_0-\delta_0-1/2}} \end{pmatrix}.$$

Therefore, collecting (131)–(135),

$$N_T(\widehat{\boldsymbol{\phi}}-\boldsymbol{\phi}_0) = \begin{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix}\sum_{t=1}^{T}\varepsilon_t\boldsymbol{\theta}_{t,T} \\ -T^{\gamma_0-\delta_0-1/2}(\widehat{\gamma}-\gamma_0)\beta_0 \end{pmatrix}+o_p(1).$$
(136)

Using summation by parts and (128) it can be shown that

$$\frac{1}{T^{1/2}}\sum_{t=1}^{T}\begin{pmatrix} 1 & 0 \end{pmatrix}\boldsymbol{\theta}_{t,T}\sum_{j=1}^{\infty}m_j(\boldsymbol{\varphi}_0)\varepsilon_{t-j} = o_p(1).$$
(137)

Additionally, as in the proof of (126), and noting that the two main terms corresponding to the two terms in $g_{t,T}(1-\delta_0,\gamma_0-\delta_0)$ cancel, it can be shown that

$$\frac{1}{T^{\gamma_0-\delta_0-1/2}}\sum_{t=1}^{T}\begin{pmatrix} 1 & 0 \end{pmatrix}\boldsymbol{\theta}_{t,T}g_{t,T}(1-\delta_0,\gamma_0-\delta_0) = o(1),$$
(138)

and

$$\begin{pmatrix} 1 & 0 \end{pmatrix}\sum_{t=1}^{T}\boldsymbol{\theta}_{t,T}\boldsymbol{\theta}'_{t,T}\begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$= \frac{1}{\rho^2(1;\boldsymbol{\varphi}_0)}\begin{pmatrix} 1 & 0 \end{pmatrix}\begin{pmatrix} \frac{1}{\Gamma^2(1-\delta_0)(1-2\delta_0)} & \frac{1}{\Gamma(1-\delta_0)\Gamma(\gamma_0-\delta_0)(\gamma_0-2\delta_0)} \\ \frac{1}{\Gamma(1-\delta_0)\Gamma(\gamma_0-\delta_0)(\gamma_0-2\delta_0)} & \frac{1}{\Gamma^2(\gamma_0-\delta_0)(2(\gamma_0-\delta_0)-1)} \end{pmatrix}^{-1}\begin{pmatrix} 1 \\ 0 \end{pmatrix}+o(1)$$

$$= \frac{\Gamma^2(1-\delta_0)(1-2\delta_0)(\gamma_0-2\delta_0)^2}{\rho^2(1;\boldsymbol{\varphi}_0)(\gamma_0-1)^2}+o(1).$$
(139)

Then (26) holds by (136)–(139) using the Cramér-Wold device.

6.4 Proof of Theorem 2(i.b): the $\gamma_0 - 1/2 > \delta_0$ case with $\delta_0 > 1/2$

The proof for this case (with $\delta_0 > 1/2$) is almost identical to that for case (i.a), but simpler. The main difference is that now, by (169) of Lemma 14, it can be straightforwardly shown

that

$$\frac{1}{T^{\gamma_0 - \delta_0 - 1/2}} \sum_{t=1}^{T} c_{t-1}(1 - \delta_0, \boldsymbol{\varphi}_0) c_{t-1}^{(1)}(\gamma_0 - \delta_0, \boldsymbol{\varphi}_0) = o(1),$$

$$\frac{1}{T^{\gamma_0 - \delta_0 - 1/2}} \sum_{t=1}^{T} c_{t-1}(1 - \delta_0, \boldsymbol{\varphi}_0) \varepsilon_t = o_p(1). \tag{140}$$

Using also (162), the consequence is that first term on the right-hand side of (121) is

$$\frac{\beta_0}{T^{\gamma_0 - \delta_0 - 1/2}} \sum_{t=1}^{T} \varepsilon_t c_{t-1}(\gamma_0 - \delta_0, \boldsymbol{\varphi}_0) \frac{\sum_{j=1}^{T} c_{j-1}(\gamma_0 - \delta_0, \boldsymbol{\varphi}_0) c_{j-1}^{(1)}(\gamma_0 - \delta_0, \boldsymbol{\varphi}_0)}{\sum_{j=1}^{T} c_{j-1}^2(\gamma_0 - \delta_0, \boldsymbol{\varphi}_0)} + o_p(1),$$

while the second term on the right-hand side of (121) is unchanged. Therefore, instead of $g_{t,T}(1 - \delta_0, \gamma_0 - \delta_0)$ we now have $g_{t,T}(\gamma_0 - \delta_0)$ with

$$g_{t,T}(d) = \frac{\pi_{t-1}(d) \sum_{j=1}^{T} \log\left(\frac{d+j-1}{T}\right) \pi_{j-1}^2(d)}{\sum_{j=1}^{T} \pi_{j-1}^2(d)} - \log\left(\frac{d+t-1}{T}\right) \pi_{t-1}(d),$$

which satisfies

$$\frac{1}{T^{2(\gamma_0 - \delta_0) - 1}} \sum_{t=1}^{T} g_{t,T}^2(\gamma_0 - \delta_0) \to \frac{1}{\Gamma^2(\gamma_0 - \delta_0)(2(\gamma_0 - \delta_0) - 1)^3}.$$

Hence, $T^{1-2(\gamma_0 - \delta_0)} \beta_0^2 \rho^2(1; \boldsymbol{\varphi}_0) \sum_{t=1}^{T} g_{t,T}^2(\gamma_0 - \delta_0) \to v_\beta$, which proves the required result for $\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0$ by an identical proof to that of (118), see also (125) and (129). To complete the proof we note that (134) still holds for (the second element of) the first term on the right-hand side of (131), whereas (the second element of) the second term on the right-hand side of (131) is now $o_p(1)$ using Lemmas 6, 7, 10, and 11 together with (140) and (162).

6.5 Proof of Theorem 2(ii): the $\gamma_0 - 1/2 < \delta_0$ case

Similarly to (40), we decompose the loss function $R_T(\boldsymbol{\vartheta})$ into the sum of two terms, $R_T(\boldsymbol{\vartheta}) = Q_T(\boldsymbol{\tau}) + J_T(\boldsymbol{\vartheta})$, where $Q_T(\boldsymbol{\tau}) = T^{-1} \sum_{t=1}^{T} s_{1t}^2(\boldsymbol{\tau})$ is the loss function in Hualde and Robinson (2011) and

$$J_T(\boldsymbol{\vartheta}) = \frac{1}{T} \sum_{t=1}^{T} (d_t(\boldsymbol{\vartheta}) - s_{2t}(\boldsymbol{\vartheta}))^2 + \frac{2}{T} \sum_{t=1}^{T} s_{1t}(\boldsymbol{\tau})(d_t(\boldsymbol{\vartheta}) - s_{2t}(\boldsymbol{\vartheta})). \tag{141}$$

Then

$$\frac{\partial R_T(\widehat{\boldsymbol{\vartheta}})}{\partial \boldsymbol{\tau}} = \mathbf{0}_{p+1} = \frac{\partial Q_T(\widehat{\boldsymbol{\tau}})}{\partial \boldsymbol{\tau}} + \frac{\partial J_T(\widehat{\boldsymbol{\vartheta}})}{\partial \boldsymbol{\tau}}, \tag{142}$$

and by the mean value theorem

$$\frac{\partial Q_T(\widehat{\boldsymbol{\tau}})}{\partial \boldsymbol{\tau}} = \frac{\partial Q_T(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\tau}} + \frac{\partial^2 Q_T(\overline{\boldsymbol{\tau}})}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}'} (\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0), \tag{143}$$

where $\overline{\boldsymbol{\tau}}$ is an intermediate point between $\widehat{\boldsymbol{\tau}}$ and $\boldsymbol{\tau}_0$ which is allowed to vary in different rows of $\partial^2 Q_T(\cdot)/\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}'$. Inserting (143) into (142) we then find

$$T^{1/2}(\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) = -\left(\frac{\partial^2 Q_T(\overline{\boldsymbol{\tau}})}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}'}\right)^{-1} T^{1/2} \frac{\partial Q_T(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\tau}} - \left(\frac{\partial^2 Q_T(\overline{\boldsymbol{\tau}})}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}'}\right)^{-1} T^{1/2} \frac{\partial J_T(\widehat{\boldsymbol{\vartheta}})}{\partial \boldsymbol{\tau}}. \tag{144}$$

From the proof of Hualde and Robinson (2011, Theorem 2.2), the first term on the right-hand side of (144) has a $N\left(\mathbf{0}_{p+1}, \boldsymbol{A}^{-1}\right)$ 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 (144), the required result follows because $T^{1/2} \partial J_T(\widehat{\boldsymbol{\vartheta}})/\partial \boldsymbol{\tau} = o_p(1)$ by Lemma 8.

Finally we prove (29). By (130), noting (28) and using Lemma 7 repeatedly, we find that

$$T^{1/2-\delta_0}(\widehat{\mu} - \mu_0) = \frac{a_{1T} + a_{2T} + a_{3T} + a_{4T}}{b_T} + o_p(1),$$

where

$$a_{1T} = \frac{1}{T^{1/2-\delta_0}} \beta_0 \sum_{j=1}^{T} c_{j-1}(1 - \delta_0, \boldsymbol{\varphi}_0) c_{j-1}(\gamma_0 - \delta_0, \boldsymbol{\varphi}_0),$$

$$a_{2T} = -\frac{1}{T^{1/2-\delta_0}} \beta_0 \sum_{j=1}^{T} c_{j-1}(1 - \delta_0, \boldsymbol{\varphi}_0) h_{j-1,T}(\widehat{\gamma} - \delta_0, \boldsymbol{\varphi}_0) \sum_{k=1}^{T} h_{k-1,T}(\widehat{\gamma} - \delta_0, \boldsymbol{\varphi}_0) c_{k-1}(\gamma_0 - \delta_0, \boldsymbol{\varphi}_0),$$

$$a_{3T} = \frac{1}{T^{1/2-\delta_0}} \sum_{j=1}^{T} c_{j-1}(1 - \delta_0, \boldsymbol{\varphi}_0) \varepsilon_j,$$

$$a_{4T} = -\frac{1}{T^{1/2-\delta_0}} \sum_{j=1}^{T} c_{j-1}(1 - \delta_0, \boldsymbol{\varphi}_0) h_{j-1,T}(\widehat{\gamma} - \delta_0, \boldsymbol{\varphi}_0) \sum_{k=1}^{T} h_{k-1,T}(\widehat{\gamma} - \delta_0, \boldsymbol{\varphi}_0) \varepsilon_k,$$

$$b_T = \frac{1}{T^{1-2\delta_0}} \left(\sum_{j=1}^{T} c_{j-1}^2(1 - \delta_0, \boldsymbol{\varphi}_0) - \frac{\left(\sum_{j=1}^{T} c_{j-1}(1 - \delta_0, \boldsymbol{\varphi}_0) c_{j-1}(\widehat{\gamma} - \delta_0, \boldsymbol{\varphi}_0) \right)^2}{\sum_{j=1}^{T} c_{j-1}^2(\widehat{\gamma} - \delta_0, \boldsymbol{\varphi}_0)} \right),$$

and $h_{t,T}(d, \boldsymbol{\varphi}) = c_t(d, \boldsymbol{\varphi})(\sum_{j=1}^{T} c_{j-1}^2(d, \boldsymbol{\varphi}))^{-1/2}$. For T sufficiently large, b_T is bounded away from zero almost surely by (172) of Lemma 15. Moreover, by direct application of (159), (176), and (185) (noting that the latter two also hold for $h_{t,T}(d, \boldsymbol{\varphi})$, see (253)), it holds that $a_{1T} = o(1)$, $a_{2T} = o_p(1)$, and $a_{4T} = O_p(T^\theta)$, while Lemmas 11 and 13 imply that $a_{3T} \to_d N(0, \rho^2(1; \boldsymbol{\varphi}_0)/((1 - 2\delta_0)\Gamma^2(1 - \delta_0)))$ by the martingale central limit theorem. This proves (29) by choosing $\theta < \epsilon$.

7 Auxiliary lemmas

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

Lemma 2. *Under Assumptions A1 and A3, for any $g > 0$ and any $\kappa > 0$, $\beta_0 \neq 0$, and either $\mu_0 = 0$ or $\mu_0 \neq 0, \gamma_0 \geq 1 + \kappa$,*

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

Lemma 3. *Under Assumptions A1 and A3, for any $\kappa > 0$, $\beta_0 \neq 0$, and either $\mu_0 = 0$ or $\mu_0 \neq 0, \gamma_0 \geq 1 + \kappa$,*

$$\frac{1}{T^{\gamma_0-\delta}} \frac{\partial \overline{d}_T(\boldsymbol{\vartheta})}{\partial \gamma} = \frac{\beta_0 \rho(1; \boldsymbol{\varphi})}{\Gamma(\gamma_0-\delta)} \frac{2(\gamma-\delta)^2 - 2(\gamma-\delta) + 1 - (\gamma_0-\delta)}{(\gamma-\delta)^2(\gamma_0+\gamma-2\delta-1)^2} + g_T(\boldsymbol{\vartheta}), \tag{145}$$

where for any arbitrarily small $\alpha > 0$, $\sup_{\overline{\mathcal{H}}_i, 1-\delta \leq 1/2+\alpha} |g_T(\boldsymbol{\vartheta})| = o(1)$. Additionally, for some fixed $\epsilon > 0$, which does not depend on ξ or T,

$$\lim_{T \to \infty} \inf_{\overline{\mathcal{H}}_i, 1-\delta \leq 1/2+\alpha} \frac{T^{2\varkappa_i}}{T^{2(\gamma_0 - \delta)}} \overline{d}_T^2(\boldsymbol{\vartheta}) > \epsilon \xi^2 \quad \text{for } i = 1, \ldots, 4. \tag{146}$$

Lemma 4. *Under Assumptions A1–A3, for any $\alpha > 0$, $\kappa > 0$, and θ such that $0 < \theta < \min\{\varsigma - 1/2, \alpha\}$, for $\beta_0 \neq 0$, and for either $\mu_0 = 0$ or $\mu_0 \neq 0, \gamma_0 \geq 1 + \kappa$,*

$$\left| \sum_{t=1}^{T} d_t(\boldsymbol{\vartheta}) s_t(\boldsymbol{\vartheta}) \right| \leq |\gamma - \gamma_0| |M_T(\boldsymbol{\vartheta})|, \tag{147}$$

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

$$\sup_{\delta_0 - \delta \leq g, \gamma_0 - \delta \leq 1/2+\alpha} |M_T(\boldsymbol{\vartheta})| = O_p(T^{\max\{\theta, g\} + 4\theta + \alpha}), \tag{148}$$

$$\sup_{\delta_0 - \delta \leq g, \gamma_0 - \delta \geq 1/2+\alpha} T^{\delta - \gamma_0} |M_T(\boldsymbol{\vartheta})| = O_p(T^{\max\{\theta, g\} + 4\theta - 1/2}), \tag{149}$$

$$\sup_{\delta_0 - \delta \geq g, \gamma_0 - \delta \geq 1/2+\alpha} T^{2\delta - \delta_0 - \gamma_0} |M_T(\boldsymbol{\vartheta})| = O_p(T^{\max\{\theta, g\} + 4\theta - g - 1/2}), \tag{150}$$

$$\sup_{\delta_0 - \delta \leq g, \gamma_0 - \delta \geq 1/2+\alpha, \gamma - \delta \geq 1/2+\theta} T^{\delta - \gamma_0} |M_T(\boldsymbol{\vartheta})| = O_p(T^{\max\{\theta, g\} - 1/2}), \tag{151}$$

$$\sup_{\delta_0 - \delta \geq g, \gamma_0 - \delta \geq 1/2+\alpha, \gamma - \delta \geq 1/2+\theta} T^{2\delta - \delta_0 - \gamma_0} |M_T(\boldsymbol{\vartheta})| = O_p(T^{\max\{\theta, g\} - g - 1/2}). \tag{152}$$

Lemma 5. *Under the conditions of Theorem 2(i) it holds that:*
(a) The first-order derivatives satisfy

$$\frac{1}{T^{1/2}} \sum_{t=1}^{T} (s_t(\boldsymbol{\vartheta}_0) - \varepsilon_t) \frac{\partial s_{1t}(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\tau}} = o_p(1), \quad \frac{1}{T^{\gamma_0 - \delta_0 - 1/2}} \sum_{t=1}^{T} (s_t(\boldsymbol{\vartheta}_0) - \varepsilon_t) \frac{\partial d_t(\boldsymbol{\vartheta}_0)}{\partial \gamma} = o_p(1), \tag{153}$$

$$\frac{1}{T^{1/2}} \sum_{t=1}^{T} s_t(\boldsymbol{\vartheta}_0) \frac{\partial s_{2t}(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\tau}} = o_p(1), \quad \frac{1}{T^{\gamma_0 - \delta_0 - 1/2}} \sum_{t=1}^{T} s_t(\boldsymbol{\vartheta}_0) \frac{\partial s_{2t}(\boldsymbol{\vartheta}_0)}{\partial \gamma} = o_p(1). \tag{154}$$

(b) The second-order derivatives satisfy

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

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

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

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

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

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

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

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

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

$$c_t(\widehat{\delta}, \widehat{\boldsymbol{\varphi}}) = c_t(\delta_0, \boldsymbol{\varphi}_0) + O_p(T^{-\varkappa}t^{\max\{\delta_0-1,-1-\varsigma\}}(\log t)^2), \tag{155}$$

$$c_t^{(1)}(\widehat{\delta}, \widehat{\boldsymbol{\varphi}}) = c_t^{(1)}(\delta_0, \boldsymbol{\varphi}_0) + O_p(T^{-\varkappa}t^{\max\{\delta_0-1,-1-\varsigma\}}(\log t)^3), \tag{156}$$

and, uniformly in $t = 1, ..., T$,

$$\rho(L; \widehat{\boldsymbol{\varphi}})u_t(\widehat{\delta} - \delta_0) = \sum_{j=0}^{t-1}\rho_j(\widehat{\boldsymbol{\varphi}})u_{t-j}(\widehat{\delta} - \delta_0) = \sum_{j=0}^{t-1}\rho_j(\boldsymbol{\varphi}_0)u_{t-j} + O_p(T^{-\varkappa}). \tag{157}$$

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

8 Technical lemmas

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

Lemma 10. *For any $d > 0$ and any fixed $a \ge 0$, as $T \to \infty$,*

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

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

$$\pi_j(-v) = \frac{1}{\Gamma(-v)}j^{-v-1}(1 + \epsilon_j(v)), \tag{158}$$

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

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

Lemma 12. *Under Assumptions A1 and A3, uniformly in* $t = 1, \dots, T$ *and* $T \geq 1$, *for* $m \geq 0$,

$$\sup_{d \leq g, \varphi \in \Psi} \left| \frac{\partial^m c_t(d, \varphi)}{\partial d^m} \right| = O(t^{\max\{g-1, -1-\varsigma\}}(\log t)^m), \tag{159}$$

$$\sup_{d \geq g, \varphi \in \Psi} \left| \frac{\partial^m}{\partial d^m} T^{-d} c_t(d, \varphi) \right| = O(T^{-g} t^{\max\{g-1, -1-\varsigma\}}(1 + |\log(t/T)|)^m). \tag{160}$$

Lemma 13. *Under Assumptions A1 and A3,*

$$\frac{1}{T^{2d-1}} \sum_{t=1}^{T} c_{t-1}^2(d, \varphi) = \frac{\rho^2(1; \varphi)}{T^{2d-1}} \sum_{t=1}^{T} \pi_{t-1}^2(d) + |r_T(d, \varphi)|, \tag{161}$$

where, for any $\eta > 0$, $\sup_{d \geq 1/2+\eta, \varphi \in \Psi} |r_T(d, \varphi)| = o(1)$.

Lemma 14. *Under Assumptions A1 and A3, for any* $\underline{d} \leq \overline{d}$ *and* α *such that* $0 < \alpha < (\varsigma - 1/2)/3$,

$$\inf_{\underline{d} \leq d \leq \overline{d}, \varphi \in \Psi} \sum_{t=1}^{T} c_{t-1}^2(d, \varphi) \geq 1, \tag{162}$$

$$\inf_{1/2-\alpha \leq d \leq 1/2+\alpha, \varphi \in \Psi} \frac{1}{T^{2d-1}} \sum_{t=1}^{T} c_{t-1}^2(d, \varphi) \geq \frac{\epsilon}{\alpha} + o(1), \tag{163}$$

$$\inf_{1/2+\alpha \leq d \leq \overline{d}, \varphi \in \Psi} \frac{1}{T^{2d-1}} \sum_{t=1}^{T} c_{t-1}^2(d, \varphi) \geq \epsilon + o(1), \tag{164}$$

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

Additionally, for any $\alpha > 0$ *and* $g > \alpha$, *and for* $m = 0, 1$,

$$\sup_{d_1 \leq 1/2-\alpha, d_2 \leq 1/2-\alpha, \varphi \in \Psi} \sum_{t=1}^{T} c_{t-1}(d_1, \varphi) c_{t-1}(d_2, \varphi) = O(1), \tag{165}$$

$$\sup_{d_1 \geq 1/2+g, d_2 \leq 1/2-\alpha, \varphi \in \Psi} \frac{1}{T^{d_1-1/2}} \sum_{t=1}^{T} \frac{\partial^m}{\partial d_1^m} c_{t-1}(d_1, \varphi) c_{t-1}(d_2, \varphi) = O(T^{-\alpha}(\log T)^m), \tag{166}$$

$$\sup_{d_1 \geq 1/2-\alpha, d_2 \leq 1/2-\alpha, \varphi \in \Psi} \frac{1}{T^{d_1-1/2}} \sum_{t=1}^{T} c_{t-1}(d_1, \varphi) c_{t-1}(d_2, \varphi) = O(T^{\alpha}), \tag{167}$$

$$\sup_{d_1 \geq 1/2-\alpha, \varphi \in \Psi} \frac{1}{T^{2d_1-1}} \sum_{t=1}^{T} c_{t-1}^2(d_1, \varphi) = O(T^{2\alpha}), \tag{168}$$

$$\sup_{d_1 \geq 1/2+\alpha, d_2 \leq 1/2-\alpha, \varphi \in \Psi} \frac{1}{T^{d_1-1/2}} \sum_{t=1}^{T} c_{t-1}(d_1, \varphi) c_{t-1}(d_2, \varphi) = O(T^{-\alpha} \log T). \tag{169}$$

Finally, for any $\eta > 0$ and $\theta \geq -1/2 - \varrho$, and for $m = 0, 1$,

$$\sup_{\substack{d_1 \geq 1/2+\theta, d_2 \geq 1/2+\theta \\ d_1+d_2 \geq 1+\eta, \boldsymbol{\varphi} \in \Psi}} \frac{1}{T^{d_1+d_2-1}} \sum_{t=1}^{T} \frac{\partial^m}{\partial d_1^m} c_{t-1}(d_1, \boldsymbol{\varphi}) c_{t-1}(d_2, \boldsymbol{\varphi}) = O((\log T)^m). \tag{170}$$

Lemma 15. *Under Assumptions A1 and A3, for any $\alpha > 0$ and $\kappa > 2\alpha$, there exists a fixed $\epsilon > 0$, which does not depend on α or T, such that, uniformly in $|d_1 - d_2| \geq \kappa$,*

$$\inf_{\substack{d_1 \leq \frac{1}{2}-\alpha \\ d_2 \leq \frac{1}{2}-\alpha}} \left(\sum_{t=1}^{T} c_{t-1}^2(d_1, \boldsymbol{\varphi}) \sum_{j=1}^{T} c_{j-1}^2(d_2, \boldsymbol{\varphi}) - \left(\sum_{t=1}^{T} c_{t-1}(d_1, \boldsymbol{\varphi}) c_{t-1}(d_2, \boldsymbol{\varphi}) \right)^2 \right) \geq \kappa^2, \tag{171}$$

$$\inf_{\frac{1}{2}+\alpha \leq d_1} \frac{1}{T^{2d_1-1}} \left(\sum_{t=1}^{T} c_{t-1}^2(d_1, \boldsymbol{\varphi}) - \frac{\left(\sum_{t=1}^{T} c_{t-1}(d_1, \boldsymbol{\varphi}) c_{t-1}(d_2, \boldsymbol{\varphi}) \right)^2}{\sum_{t=1}^{T} c_{t-1}^2(d_2, \boldsymbol{\varphi})} \right) \geq \epsilon, \tag{172}$$

$$\inf_{\substack{\frac{1}{2}-\alpha \leq d_1 \leq \frac{1}{2}+\alpha \\ d_2 \leq \frac{1}{2}-\alpha}} \frac{1}{T^{2d_1-1}} \left(\sum_{t=1}^{T} c_{t-1}^2(d_1, \boldsymbol{\varphi}) \sum_{j=1}^{T} c_{j-1}^2(d_2, \boldsymbol{\varphi}) - \left(\sum_{t=1}^{T} c_{t-1}(d_1, \boldsymbol{\varphi}) c_{t-1}(d_2, \boldsymbol{\varphi}) \right)^2 \right) \geq \frac{\epsilon}{\alpha} + o(1), \tag{173}$$

$$\inf_{\substack{\frac{1}{2}-\alpha \leq d_1 \leq \frac{1}{2}+\alpha \\ \frac{1}{2}-\alpha \leq d_2 \leq \frac{1}{2}+\alpha}} \frac{1}{T^{2d_1-1}} \left(\sum_{t=1}^{T} c_{t-1}^2(d_1, \boldsymbol{\varphi}) - \frac{\left(\sum_{t=1}^{T} c_{t-1}(d_1, \boldsymbol{\varphi}) c_{t-1}(d_2, \boldsymbol{\varphi}) \right)^2}{\sum_{t=1}^{T} c_{t-1}^2(d_2, \boldsymbol{\varphi})} \right) \geq \frac{\epsilon}{\alpha} + O(1), \tag{174}$$

$$\inf_{\substack{\frac{1}{2}-\alpha \leq d_1 \leq \frac{1}{2}+\alpha \\ \frac{1}{2}-\alpha \leq d_2 \leq \frac{1}{2}+\alpha}} \frac{1}{T^{2d_1+2d_2-2}} \left(\sum_{t=1}^{T} c_{t-1}^2(d_1, \boldsymbol{\varphi}) \sum_{j=1}^{T} c_{j-1}^2(d_2, \boldsymbol{\varphi}) - \left(\sum_{t=1}^{T} c_{t-1}(d_1, \boldsymbol{\varphi}) c_{t-1}(d_2, \boldsymbol{\varphi}) \right)^2 \right) \geq \frac{\epsilon}{\alpha} + O(1). \tag{175}$$

Lemma 16. *Let θ and κ be arbitrary positive numbers such that $0 < \theta < \min\{\varsigma - 1/2, \kappa/2\}$. Then, under Assumptions A1 and A3, for any real numbers $\underline{d}_1, \underline{d}_2 \leq 1/2 - \theta$ and $\overline{d}_1, \overline{d}_2 \geq 1/2 + \theta$, uniformly in $|d_1 - d_2| \geq \kappa$ and in $t, k = 1, \ldots, T$ and $T \geq 1$,*

$$\sup_{d_1 \in [\underline{d}_1, \overline{d}_1], d_2 \in [\underline{d}_2, \overline{d}_2], \boldsymbol{\varphi} \in \Psi} \|\boldsymbol{h}_{t-1,T}(d_1, d_2, \boldsymbol{\varphi})\| = O(t^{-1/2-\theta} T^\theta), \tag{176}$$

$$\sup_{d_1 \in [\underline{d}_1, \overline{d}_1], , \boldsymbol{\varphi} \in \Psi} \left| \sum_{t=1}^{T} h_{t-1,T}(d_1, \boldsymbol{\varphi}) \right| = O(T^{1/2}), \tag{177}$$

$$\sup_{d_1 \in [\underline{d}_1, \overline{d}_1], d_2 \in [\underline{d}_2, \overline{d}_2], \boldsymbol{\varphi} \in \Psi} \|\boldsymbol{h}_{t,T}(d_1, d_2, \boldsymbol{\varphi}) - \boldsymbol{h}_{t-1,T}(d_1, d_2, \boldsymbol{\varphi})\| = O(t^{-3/2-\theta} T^\theta), \tag{178}$$

$$\sup_{d_1 \in [\underline{d}_1, \overline{d}_1], d_2 \in [\underline{d}_2, \overline{d}_2], \boldsymbol{\varphi} \in \Psi} \left| \frac{\partial}{\partial d_2} \left(\boldsymbol{h}_{t-1,T}'(d_1, d_2, \boldsymbol{\varphi}) \boldsymbol{h}_{k-1,T}(d_1, d_2, \boldsymbol{\varphi}) \right) \right| = O(t^{-1/2-\theta} k^{-1/2-\theta} T^{5\theta}), \tag{179}$$

$$\sup_{d_1\in[\underline{d}_1,\bar{d}_1],d_2\in[\underline{d}_2,\bar{d}_2],\varphi\in\Psi}\left|\frac{\partial}{\partial d_2}((\boldsymbol{h}_{t,T}(d_1,d_2,\varphi)-\boldsymbol{h}_{t-1,T}(d_1,d_2,\varphi))'\boldsymbol{h}_{k-1,T}(d_1,d_2,\varphi))\right|$$

$$=O(t^{-3/2-\theta}k^{-1/2-\theta}T^{5\theta}),\tag{180}$$

Additionally, for any α such that $0<\alpha<\theta$,

$$\sup_{d_1\in[\underline{d}_1,\bar{d}_1],d_2\in[1/2+\theta,\bar{d}_2],\varphi\in\Psi}\left|\frac{\partial}{\partial d_2}\left(\boldsymbol{h}'_{t-1,T}(d_1,d_2,\varphi)\boldsymbol{h}_{k-1,T}(d_1,d_2,\varphi)\right)\right|=O(t^{-1/2-\alpha}k^{-1/2-\alpha}T^{2\alpha}),\tag{181}$$

$$\sup_{d_1\in[\underline{d}_1,\bar{d}_1],d_2\in[1/2+\theta,\bar{d}_2],\varphi\in\Psi}\left|\frac{\partial}{\partial d_2}((\boldsymbol{h}_{t,T}(d_1,d_2,\varphi)-\boldsymbol{h}_{t-1,T}(d_1,d_2,\varphi))'\boldsymbol{h}_{k-1,T}(d_1,d_2,\varphi))\right|$$

$$=O(t^{-3/2-\alpha}k^{-1/2-\alpha}T^{2\alpha}),\tag{182}$$

Lemma 17. *Under Assumptions A1–A3, uniformly in $t=1,\dots,T$, $T\geq 1$, and $\varphi\in\Psi$,*

$$\sup_{d\leq g}|\rho(L;\varphi)u_t(-d)|=O_p(t^{g-1/2}+\log t\mathbb{I}(g=1/2)+\mathbb{I}(g<1/2)),\tag{183}$$

$$\sup_{d\geq g}|T^{-d}\rho(L;\varphi)u_t(-d)|=O_p(T^{-g}(t^{g-1/2}+\log t\mathbb{I}(g=1/2)+\mathbb{I}(g<1/2))).\tag{184}$$

Lemma 18. *Let θ and κ be arbitrary positive numbers such that $0<\theta<\min\{\varsigma-1/2,\kappa/2\}$. Then, under Assumptions A1–A3, uniformly in $\boldsymbol{\vartheta}\in\Xi$ and in $t,k=1,\dots,T$ and $T\geq 1$,*

$$\sup_{\delta_0-\delta\leq g}\left\|\sum_{t=1}^T\rho(L;\varphi)u_t(\delta-\delta_0)\boldsymbol{h}_{t-1,T}(1-\delta,\gamma-\delta,\varphi)\right\|=O_p(T^{\max\{\theta,g\}}),\tag{185}$$

$$\sup_{\delta_0-\delta\leq g}\left|\sum_{t=1}^T\rho(L;\varphi)u_t(\delta-\delta_0)\frac{\partial}{\partial\gamma}\left(\boldsymbol{h}'_{t-1,T}(1-\delta,\gamma-\delta,\varphi)\boldsymbol{h}_{k-1,T}(1-\delta,\gamma-\delta,\varphi)\right)\right|$$

$$=O_p(k^{-1/2-\theta}T^{\max\{\theta,g\}+5\theta}),\tag{186}$$

$$\sup_{\delta_0-\delta\geq g}\frac{1}{T^{\delta_0-\delta}}\left\|\sum_{t=1}^T\rho(L;\varphi)u_t(\delta-\delta_0)\boldsymbol{h}_{t-1,T}(1-\delta,\gamma-\delta,\varphi)\right\|=O_p(T^{\max\{\theta,g\}-g}),\tag{187}$$

$$\sup_{\delta_0-\delta\geq g}\frac{1}{T^{\delta_0-\delta}}\left|\sum_{t=1}^T\rho(L;\varphi)u_t(\delta-\delta_0)\frac{\partial}{\partial\gamma}\left(\boldsymbol{h}'_{t-1,T}(1-\delta,\gamma-\delta,\varphi)\boldsymbol{h}_{k-1,T}(1-\delta,\gamma-\delta,\varphi)\right)\right|$$

$$=O_p(k^{-1/2-\theta}T^{\max\{\theta,g\}-g+5\theta}),\tag{188}$$

$$\sup_{\delta_0-\delta\leq g,\gamma-\delta\geq 1/2+\theta}\left|\sum_{t=1}^T\rho(L;\varphi)u_t(\delta-\delta_0)\frac{\partial}{\partial\gamma}\left(\boldsymbol{h}'_{t-1,T}(1-\delta,\gamma-\delta,\varphi)\boldsymbol{h}_{k-1,T}(1-\delta,\gamma-\delta,\varphi)\right)\right|$$

$$=O_p(k^{-1/2-\theta}T^{\max\{\theta\alpha,g\}+\theta}),\tag{189}$$

$$\sup_{\delta_0-\delta\geq g,\gamma-\delta\geq 1/2+\theta}\frac{1}{T^{\delta_0-\delta}}\left|\sum_{t=1}^T\rho(L;\varphi)u_t(\delta-\delta_0)\frac{\partial}{\partial\gamma}\left(\boldsymbol{h}'_{t-1,T}(1-\delta,\gamma-\delta,\varphi)\boldsymbol{h}_{k-1,T}(1-\delta,\gamma-\delta,\varphi)\right)\right|$$

$$=O_p(k^{-1/2-\theta}T^{\max\{\theta,g\}+\theta-g}).\tag{190}$$

Lemma 19. *Let $\theta > 0$ be an arbitrary number. Then, under Assumptions A1 and A3, uniformly in $\boldsymbol{\vartheta} \in \Xi$,*

$$\sup_{\gamma_0 - \delta \leq g} \left\| \sum_{t=1}^{T} c_{t-1}(\gamma_0 - \delta, \boldsymbol{\varphi}) \boldsymbol{h}_{t-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \right\| = O(T^{\max\{\theta, g-1/2\}}), \qquad (191)$$

$$\sup_{\gamma_0 - \delta \geq g} \frac{1}{T^{\gamma_0 - \delta}} \left\| \sum_{t=1}^{T} c_{t-1}(\gamma_0 - \delta, \boldsymbol{\varphi}) \boldsymbol{h}_{t-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \right\| = O(T^{\max\{\theta, g-1/2\}-g}). \qquad (192)$$

Lemma 20. *Under Assumptions A1–A3, uniformly in $\boldsymbol{\vartheta} \in \Xi$,*

$$\sup_{\delta_0 - \delta \leq g_1, \gamma_0 - \delta \leq g_2} \frac{1}{T} \left| \sum_{t=1}^{T} \rho(L; \boldsymbol{\varphi}) u_t(\delta - \delta_0) c_{t-1}(\gamma_0 - \delta, \boldsymbol{\varphi}) \right|$$
$$= O_p(T^{g_1+g_2-3/2} + T^{-1}(\log T) + T^{g_1-1/2-\varsigma} + T^{\max\{g_2-2,0\}}(\log T)^2 \mathbb{I}(g_1 \leq -1/2)), \qquad (193)$$

$$\sup_{\delta_0 - \delta \geq g_1, \gamma_0 - \delta \geq g_2} \frac{1}{T^{\gamma_0 + \delta_0 - 2\delta}} \left| \sum_{t=1}^{T} \rho(L; \boldsymbol{\varphi}) u_t(\delta - \delta_0) c_{t-1}(\gamma_0 - \delta, \boldsymbol{\varphi}) \right|$$
$$= O_p(T^{-1/2} + T^{1/2-g_2-\varsigma} + T^{-g_1-g_2}(\log T) + T^{-g_1-1}(\log T)^2 \mathbb{I}(g_1 \leq -1/2)). \qquad (194)$$

Lemma 21. *Let η, α be arbitrary positive numbers such that $\alpha < \eta + 1/2$. Then, under Assumptions A1–A3, uniformly in $\boldsymbol{\vartheta} \in \Xi$,*

$$\sup_{\delta_0 - \delta \geq 1/2+\eta, d \leq 1/2-\alpha} \frac{1}{T^{\delta_0 - \delta}} \left| \sum_{t=1}^{T} \rho(L; \boldsymbol{\varphi}) u_t(\delta - \delta_0) c_{t-1}(d, \boldsymbol{\varphi}) \right| = O_p(T^{-\alpha}), \qquad (195)$$

$$\sup_{\delta_0 - \delta \geq 1/2+\eta, d \geq 1/2-\alpha} \frac{1}{T^{\delta_0 - \delta + d - 1/2}} \left| \sum_{t=1}^{T} \rho(L; \boldsymbol{\varphi}) u_t(\delta - \delta_0) c_{t-1}(d, \boldsymbol{\varphi}) \right| = O_p(1). \qquad (196)$$

Lemma 22. *Let $Z_t, t = 1, \ldots, T$, be arbitrary. Then*

$$\sum_{t=1}^{T} Z_t^2 \geq \left(\frac{\pi^2}{4} T^{-2} + O(T^{-3}) \right) \sum_{t=1}^{T} (\Delta_+^{-1} Z_t)^2, \qquad (197)$$

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

9 Proofs of auxiliary lemmas

9.1 Proof of Lemma 1

We first find that

$$\frac{1}{T} \sum_{t=1}^{T} s_t^2(\boldsymbol{\vartheta}_0) = \frac{1}{T} \sum_{t=1}^{T} (\rho(L; \boldsymbol{\varphi}_0)(u_t \mathbb{I}(t > 0)))^2$$
$$- \frac{1}{T} \left\| \sum_{t=1}^{T} \rho(L; \boldsymbol{\varphi}_0)(u_t \mathbb{I}(t > 0)) \boldsymbol{h}_{t-1,T}(1 - \delta_0, \gamma_0 - \delta_0, \boldsymbol{\varphi}_0) \right\|^2, \qquad (198)$$

where

$$\rho(L; \boldsymbol{\varphi}_0)(u_t \mathbb{I}(t > 0)) = \varepsilon_t - \sum_{j=t}^{\infty} \rho_j(\boldsymbol{\varphi}_0) u_{t-j}.$$

It can be straightforwardly shown that

$$\sum_{j=t}^{\infty} \rho_j(\boldsymbol{\varphi}_0) u_{t-j} = O_p(t^{-1/2-\varsigma})$$

by Assumptions A1 and A2. By simple application of Lemma 16, the second term on the right-hand side of (198) is $O_p\left(T^{2\theta-1}\right) = o_p(1)$ by choosing $\theta < 1/2$. Then the required result holds because $T^{-1}\sum_{t=1}^{T} \varepsilon_t^2 \to_p \sigma_0^2$ by a law of large numbers.

9.2 Proof of Lemma 2

Letting $\alpha > 0$ be arbitrarily small (in particular $\alpha < \min\{(\varsigma - 1/2)/3, g, \kappa/2\}$, which implies $\alpha < 1/2$) and defining the sets

$$\begin{aligned}
\Phi_1 &= \{\boldsymbol{\vartheta} \in \Xi : 1 - \delta \leq 1/2 - \alpha, \gamma - \delta \leq 1/2 - \alpha\}, \\
\Phi_2 &= \{\boldsymbol{\vartheta} \in \Xi : 1 - \delta \leq 1/2 - \alpha, 1/2 - \alpha \leq \gamma - \delta \leq 1/2 + \alpha\}, \\
\Phi_3 &= \{\boldsymbol{\vartheta} \in \Xi : 1 - \delta \leq 1/2 - \alpha, \gamma - \delta \geq 1/2 + \alpha\}, \\
\Phi_4 &= \{\boldsymbol{\vartheta} \in \Xi : 1/2 - \alpha \leq 1 - \delta \leq 1/2 + \alpha, \gamma - \delta \leq 1/2 - \alpha\}, \\
\Phi_5 &= \{\boldsymbol{\vartheta} \in \Xi : 1/2 - \alpha \leq 1 - \delta \leq 1/2 + \alpha, 1/2 - \alpha \leq \gamma - \delta \leq 1/2 + \alpha\}, \\
\Phi_6 &= \{\boldsymbol{\vartheta} \in \Xi : 1/2 - \alpha \leq 1 - \delta \leq 1/2 + \alpha, \gamma - \delta \geq 1/2 + \alpha\}, \\
\Phi_7 &= \{\boldsymbol{\vartheta} \in \Xi : 1 - \delta \geq 1/2 + \alpha, \gamma - \delta \leq 1/2 - \alpha\}, \\
\Phi_8 &= \{\boldsymbol{\vartheta} \in \Xi : 1 - \delta \geq 1/2 + \alpha, 1/2 - \alpha \leq \gamma - \delta \leq 1/2 + \alpha\}, \\
\Phi_9 &= \{\boldsymbol{\vartheta} \in \Xi : 1 - \delta \geq 1/2 + \alpha, \gamma - \delta \geq 1/2 + \alpha\},
\end{aligned}$$

the result holds on showing that

$$\lim_{T \to \infty} \inf_{\{\gamma_0 - \delta \geq 1/2 + g, |\gamma - \gamma_0| \geq g\} \cap \Phi_j} \frac{1}{T^{2(\gamma_0 - \delta) - 1}} \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta}) > \epsilon \tag{199}$$

for $j = 1, ..., 9$. Clearly,

$$\begin{aligned}
\frac{1}{T^{2(\gamma_0-\delta)-1}} \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta}) &= \frac{1}{T^{2(\gamma_0-\delta)-1}} \boldsymbol{\phi}_0' \sum_{t=1}^{T} \boldsymbol{c}_{t-1}(1-\delta, \gamma_0 - \delta, \boldsymbol{\varphi}) \boldsymbol{c}_{t-1}'(1-\delta, \gamma_0 - \delta, \boldsymbol{\varphi}) \boldsymbol{\phi}_0 \\
&\quad - \frac{1}{T^{2(\gamma_0-\delta)-1}} \boldsymbol{\phi}_0' \sum_{t=1}^{T} \boldsymbol{c}_{t-1}(1-\delta, \gamma_0 - \delta, \boldsymbol{\varphi}) \boldsymbol{h}_{t-1}'(1-\delta, \gamma - \delta, \boldsymbol{\varphi}) \\
&\quad \times \sum_{j=1}^{T} \boldsymbol{h}_{j-1}(1-\delta, \gamma - \delta, \boldsymbol{\varphi}) \boldsymbol{c}_{j-1}'(1-\delta, \gamma_0 - \delta, \boldsymbol{\varphi}) \boldsymbol{\phi}_0. \tag{200}
\end{aligned}$$

The first term on the right-hand side of (200) is

$$\frac{1}{T^{2(\gamma_0-\delta)-1}} \sum_{t=1}^{T} (\mu_0 c_{t-1}(1-\delta, \boldsymbol{\varphi}) + \beta_0 c_{t-1}(\gamma_0 - \delta, \boldsymbol{\varphi}))^2, \tag{201}$$

so, noting that $\beta_0 \neq 0$ and either $\mu_0 = 0$ or $\mu_0 \neq 0$, $\gamma_0 \geq 1+\kappa$, and also $\gamma_0 - \delta \geq 1/2+g > 1/2$, the second term within the parenthesis in (201) dominates. As will be seen, the proof of (199) changes substantially depending on whether none of $1 - \delta$ and $\gamma - \delta$ are larger than $1/2 + \alpha$ or not. We deal first with the former situation, which corresponds to cases $j = 1, 2, 4, 5$.

Specifically, noting (200), (201), and $\beta_0 \neq 0$, $|\gamma - \gamma_0| \geq g$, by application of Lemma 13, (199) for $j = 1, 2, 4, 5$ holds on showing that

$$\lim_{T \to \infty} \inf_{\gamma_0 - \delta \geq 1/2+g} \frac{1}{T^{2(\gamma_0-\delta)-1}} \sum_{t=1}^{T} \pi_{t-1}^2(\gamma_0 - \delta) > \epsilon, \qquad (202)$$

$$\sup_{\{\gamma_0 - \delta \geq 1/2+g, |\gamma - \gamma_0| \geq g\} \cap \Phi_j} \frac{1}{T^{2(\gamma_0-\delta)-1}} \left\| \sum_{t=1}^{T} c_{t-1}(\gamma_0 - \delta, \boldsymbol{\varphi}) \boldsymbol{h}_{t-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \right\|^2 = o(1). \quad (203)$$

First, we show (202). By Lemma 10 and (158) in Lemma 11,

$$\inf_{\gamma_0 - \delta \geq 1/2+g} \frac{1}{T^{2(\gamma_0-\delta)-1}} \sum_{t=1}^{T} \pi_{t-1}^2(\gamma_0 - \delta) \geq \epsilon \inf_{\gamma_0 - \delta \geq 1/2+g} \frac{1}{T} \sum_{t=1}^{T} \left(\frac{t}{T} \right)^{2(\gamma_0-\delta)-2} = \epsilon \frac{1}{2g} + o(1), \quad (204)$$

so that (202) holds by taking limits as $T \to \infty$.

Next, we show (203) for $j = 1$. Clearly,

$$\frac{1}{T^{2(\gamma_0-\delta)-1}} \left\| \sum_{t=1}^{T} c_{t-1}(\gamma_0 - \delta, \boldsymbol{\varphi}) \boldsymbol{h}_{t-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \right\|^2 = \frac{N_{1T} + N_{2T} + N_{3T}}{T^{2(\gamma_0-\delta)-1} D_T}, \qquad (205)$$

where

$$N_{1T} = \left(\sum_{t=1}^{T} c_{t-1}(\gamma_0 - \delta, \boldsymbol{\varphi}) c_{t-1}(1 - \delta, \boldsymbol{\varphi}) \right)^2 \sum_{t=1}^{T} c_{t-1}^2(\gamma - \delta, \boldsymbol{\varphi}),$$

$$N_{2T} = \left(\sum_{t=1}^{T} c_{t-1}(\gamma_0 - \delta, \boldsymbol{\varphi}) c_{t-1}(\gamma - \delta, \boldsymbol{\varphi}) \right)^2 \sum_{t=1}^{T} c_{t-1}^2(1 - \delta, \boldsymbol{\varphi}),$$

$$N_{3T} = -2 \sum_{t=1}^{T} c_{t-1}(\gamma_0 - \delta, \boldsymbol{\varphi}) c_{t-1}(1 - \delta, \boldsymbol{\varphi}) \sum_{t=1}^{T} c_{t-1}(\gamma_0 - \delta, \boldsymbol{\varphi}) c_{t-1}(\gamma - \delta, \boldsymbol{\varphi})$$

$$\times \sum_{t=1}^{T} c_{t-1}(1 - \delta, \boldsymbol{\varphi}) c_{t-1}(\gamma - \delta, \boldsymbol{\varphi}),$$

$$D_T = \sum_{t=1}^{T} c_{t-1}^2(1 - \delta, \boldsymbol{\varphi}) \sum_{t=1}^{T} c_{t-1}^2(\gamma - \delta, \boldsymbol{\varphi}) - \left(\sum_{t=1}^{T} c_{t-1}(1 - \delta, \boldsymbol{\varphi}) c_{t-1}(\gamma - \delta, \boldsymbol{\varphi}) \right)^2.$$

Thus, (203) for $j = 1$ follows by (165), (166) of Lemma 14 and (171) of Lemma 15. Next we show (203) for $j = 2$. Noting (205), the proof consists on showing that the infimum of $T^{1-2(\gamma-\delta)} D_T$ can be made arbitrarily large by setting α sufficiently close to zero, and also that the supremum of $T^{2-2(\gamma-\delta)-2(\gamma_0-\delta)}(N_{1T} + N_{2T} + N_{3T}) = O(1)$. These results follow by (173) of Lemma 15 and (165), (166), (170), (167), (168) of Lemma 14, respectively. Next, the proof of (203) for $j = 4$ is basically identical to that for $j = 2$, so it is omitted.

To prove (203) for $j = 5$ we now write

$$\frac{1}{T^{2(\gamma_0-\delta)-1}} \left\| \sum_{t=1}^{T} c_{t-1}(\gamma_0 - \delta, \boldsymbol{\varphi}) \boldsymbol{h}_{t-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \right\|^2 = A_{1T} + A_{2T} + A_{3T},$$

where

$$A_{1T} = \frac{\frac{1}{T^{2(\gamma_0-\delta)+2(1-\delta)-2}} \left(\sum_{t=1}^{T} c_{t-1}(\gamma_0 - \delta, \boldsymbol{\varphi}) c_{t-1}(1 - \delta, \boldsymbol{\varphi}) \right)^2}{\frac{1}{T^{2(1-\delta)-1}} \left(\sum_{t=1}^{T} c_{t-1}^2(1 - \delta, \boldsymbol{\varphi}) - \frac{\left(\sum_{t=1}^{T} c_{t-1}(1-\delta,\boldsymbol{\varphi}) c_{t-1}(\gamma-\delta,\boldsymbol{\varphi}) \right)^2}{\sum_{t=1}^{T} c_{t-1}^2(\gamma-\delta,\boldsymbol{\varphi})} \right)},$$

$$A_{2T} = \frac{\frac{1}{T^{2(\gamma_0-\delta)+2(\gamma-\delta)-2}} \left(\sum_{t=1}^{T} c_{t-1}(\gamma_0 - \delta, \boldsymbol{\varphi}) c_{t-1}(\gamma - \delta, \boldsymbol{\varphi}) \right)^2}{\frac{1}{T^{2(\gamma-\delta)-1}} \left(\sum_{t=1}^{T} c_{t-1}^2(\gamma - \delta, \boldsymbol{\varphi}) - \frac{\left(\sum_{t=1}^{T} c_{t-1}(1-\delta,\boldsymbol{\varphi}) c_{t-1}(\gamma-\delta,\boldsymbol{\varphi}) \right)^2}{\sum_{t=1}^{T} c_{t-1}^2(1-\delta,\boldsymbol{\varphi})} \right)},$$

$$A_{3T} = -2 \frac{\frac{1}{T^{2(\gamma_0-\delta)+2(\gamma-\delta)+2(1-\delta)-3}} \sum_{t=1}^{T} c_{t-1}(\gamma_0 - \delta, \boldsymbol{\varphi}) c_{t-1}(1 - \delta, \boldsymbol{\varphi})}{\left(\sum_{t=1}^{T} c_{t-1}^2(1 - \delta, \boldsymbol{\varphi}) \sum_{t=1}^{T} c_{t-1}^2(\gamma - \delta, \boldsymbol{\varphi}) - \left(\sum_{t=1}^{T} c_{t-1}(1 - \delta, \boldsymbol{\varphi}) c_{t-1}(\gamma - \delta, \boldsymbol{\varphi}) \right)^2 \right)}$$

$$\times \frac{\sum_{t=1}^{T} c_{t-1}(\gamma_0 - \delta, \boldsymbol{\varphi}) c_{t-1}(\gamma - \delta, \boldsymbol{\varphi}) \sum_{t=1}^{T} c_{t-1}(1 - \delta, \boldsymbol{\varphi}) c_{t-1}(\gamma - \delta, \boldsymbol{\varphi})}{\frac{1}{T^{2(\gamma-\delta)+2(1-\delta)-2}}}.$$

The proof now follows straightforwardly by application of (170) of Lemma 14 to the numerators and (174), (175) of Lemma 15 to the denominators.

Next we show (199) for $j = 3$. We use the decomposition (205), and first note that, by (162), (164), (166), and (169),

$$\sup_{\{\gamma_0-\delta \geq 1/2+g, |\gamma-\gamma_0| \geq g\} \cap \Phi_3} T^{1-2(\gamma_0-\delta)} N_{1T}/D_T = o(1).$$

Similarly, by (162) and (169),

$$\frac{N_{2T}}{T^{2(\gamma_0-\delta)-1} D_T} = \frac{1}{T^{2(\gamma_0-\delta)-1}} \frac{\left(\sum_{t=1}^{T} c_{t-1}(\gamma_0 - \delta, \boldsymbol{\varphi}) c_{t-1}(\gamma - \delta, \boldsymbol{\varphi}) \right)^2}{\sum_{t=1}^{T} c_{t-1}^2(\gamma - \delta, \boldsymbol{\varphi})} + q_{1T}(\boldsymbol{\vartheta}),$$

where $\sup_{\{\gamma_0-\delta \geq 1/2+g, |\gamma-\gamma_0| \geq g\} \cap \Phi_3} |q_{1T}(\boldsymbol{\vartheta})| = o(1)$. Additionally, by (162), (164), (166), (170), and (169),

$$\sup_{\{\gamma_0-\delta \geq 1/2+g, |\gamma-\gamma_0| \geq g\} \cap \Phi_3} T^{1-2(\gamma_0-\delta)} N_{3T}/D_T = o(1).$$

Collecting all these results, by very similar steps to those in the proof of Lemma 13, noting that $\gamma_0 - \delta \geq 1/2 + g$, it is straightforward to show that

$$\frac{1}{T^{2(\gamma_0-\delta)-1}} \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta}) = \frac{\beta_0^2 \rho^2(1; \boldsymbol{\varphi})}{T^{2(\gamma_0-\delta)-1}} \left(\sum_{t=1}^{T} \pi_{t-1}^2(\gamma_0 - \delta) - \frac{\left(\sum_{t=1}^{T} \pi_{t-1}(\gamma_0 - \delta) \pi_{t-1}(\gamma - \delta) \right)^2}{\sum_{t=1}^{T} \pi_{t-1}^2(\gamma - \delta)} \right)$$

$$+ q_{2T}(\boldsymbol{\vartheta}), \tag{206}$$

where $\sup_{\{\gamma_0-\delta\geq 1/2+g\}\cap\Phi_3}|q_{2T}(\boldsymbol{\vartheta})| = o(1)$. Next, using (158) of Lemma 11 and approximating sums by integrals, see Lemma 10,

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

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

$$= \epsilon \inf_{\{\gamma_0-\delta\geq 1/2+g, |\gamma-\gamma_0|\geq g\}\cap\Phi_3} \frac{(\gamma_0-\gamma)^2}{\Gamma^2(\gamma_0-\delta)(2(\gamma_0-\delta)-1)(\gamma_0+\gamma-2\delta-1)^2} - o(1)$$

$$\geq \epsilon \inf_{\gamma_0-\delta\geq 1/2+g} \frac{g^2}{\Gamma^2(\gamma_0-\delta)2g(\alpha+g)^2} - o(1),$$

which is positive and bounded away from zero, to complete the proof of (199) for $j = 3$.

Next we show (203) for $j = 6$. Similar to above,

$$\frac{1}{T^{2(\gamma_0-\delta)-1}}\sum_{t=1}^{T}d_t^2(\boldsymbol{\vartheta}) = \beta_0^2\left(\frac{1}{T^{2(\gamma_0-\delta)-1}}\sum_{t=1}^{T}c_{t-1}^2(\gamma_0-\delta,\boldsymbol{\varphi}) - \frac{N_{1T}+N_{2T}+N_{3T}}{T^{2(\gamma_0-\delta)-1}D_T}\right) + q_{3T}(\boldsymbol{\vartheta}),$$

where $\sup_{\{\gamma_0-\delta\geq 1/2+g, |\gamma-\gamma_0|\geq g\}\cap\Phi_6}|q_{3T}(\boldsymbol{\vartheta})| = o(1)$. Noting that

$$\frac{N_{1T}}{T^{2(\gamma_0-\delta)-1}D_T} = \frac{\left(\frac{1}{T^{\gamma_0-\delta+1-\delta-1}}\sum_{t=1}^{T}c_{t-1}(\gamma_0-\delta,\boldsymbol{\varphi})c_{t-1}(1-\delta,\boldsymbol{\varphi})\right)^2}{\frac{1}{T^{2(\gamma-\delta)-1}}\sum_{t=1}^{T}c_{t-1}^2(\gamma-\delta,\boldsymbol{\varphi}) - \frac{\left(\frac{1}{T^{\gamma-\delta+1-\delta-1}}\sum_{t=1}^{T}c_{t-1}(1-\delta,\boldsymbol{\varphi})c_{t-1}(\gamma-\delta,\boldsymbol{\varphi})\right)^2}{\frac{1}{T^{2(1-\delta)-1}}\sum_{t=1}^{T}c_{t-1}^2(1-\delta,\boldsymbol{\varphi})}}$$

$$\times \frac{\frac{1}{T^{2(\gamma-\delta)-1}}\sum_{t=1}^{T}c_{t-1}^2(\gamma-\delta,\boldsymbol{\varphi})}{\frac{1}{T^{2(1-\delta)-1}}\sum_{t=1}^{T}c_{t-1}^2(1-\delta,\boldsymbol{\varphi})},$$

$$\frac{N_{2T}}{T^{2(\gamma_0-\delta)-1}D_T} = \frac{\left(\frac{1}{T^{\gamma_0-\delta+\gamma-\delta-1}}\sum_{t=1}^{T}c_{t-1}(\gamma_0-\delta,\boldsymbol{\varphi})c_{t-1}(\gamma-\delta,\boldsymbol{\varphi})\right)^2}{\frac{1}{T^{2(\gamma-\delta)-1}}\sum_{t=1}^{T}c_{t-1}^2(\gamma-\delta,\boldsymbol{\varphi}) - \frac{\left(\frac{1}{T^{\gamma-\delta+1-\delta-1}}\sum_{t=1}^{T}c_{t-1}(1-\delta,\boldsymbol{\varphi})c_{t-1}(\gamma-\delta,\boldsymbol{\varphi})\right)^2}{\frac{1}{T^{2(1-\delta)-1}}\sum_{t=1}^{T}c_{t-1}^2(1-\delta,\boldsymbol{\varphi})}},$$

$$\frac{N_{3T}}{T^{2(\gamma_0-\delta)-1}D_T} = \frac{-2\frac{1}{T^{\gamma_0-\delta+1-\delta-1}}\sum_{t=1}^{T}c_{t-1}(\gamma_0-\delta,\boldsymbol{\varphi})c_{t-1}(1-\delta,\boldsymbol{\varphi})}{\frac{1}{T^{2(\gamma-\delta)-1}}\sum_{t=1}^{T}c_{t-1}^2(\gamma-\delta,\boldsymbol{\varphi}) - \frac{\left(\frac{1}{T^{\gamma-\delta+1-\delta-1}}\sum_{t=1}^{T}c_{t-1}(1-\delta,\boldsymbol{\varphi})c_{t-1}(\gamma-\delta,\boldsymbol{\varphi})\right)^2}{\frac{1}{T^{2(1-\delta)-1}}\sum_{t=1}^{T}c_{t-1}^2(1-\delta,\boldsymbol{\varphi})}}$$

$$\times \frac{1}{T^{\gamma_0-\delta+\gamma-\delta-1}}\sum_{t=1}^{T}c_{t-1}(\gamma_0-\delta,\boldsymbol{\varphi})c_{t-1}(\gamma-\delta,\boldsymbol{\varphi})$$

$$\times \frac{\frac{1}{T^{1-\delta+\gamma-\delta-1}}\sum_{t=1}^{T}c_{t-1}(1-\delta,\boldsymbol{\varphi})c_{t-1}(\gamma-\delta,\boldsymbol{\varphi})}{\frac{1}{T^{2(1-\delta)-1}}\sum_{t=1}^{T}c_{t-1}^2(1-\delta,\boldsymbol{\varphi})},$$

and using (163) and (170), the proof follows by identical steps as those given in the treatment of (206). Next, the proofs of (203) for $j = 7, 8$ are omitted because they is identical to those

for $j = 3, 6$, respectively. Finally, we show (203) for $j = 9$. The proof is relatively similar to that for $j = 3$. In particular, by the same methods as applied in the proof for $j = 3$, it can be shown that

$$
\frac{1}{T^{2(\gamma_0-\delta)-1}} \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta}) = \beta_0^2 \rho^2(1;\boldsymbol{\varphi}) \frac{1}{T^{2(\gamma_0-\delta)-1}} \sum_{t=1}^{T} t^{2(\gamma_0-\delta-1)} - \beta_0^2 \rho^2(1;\boldsymbol{\varphi})
$$

$$
\times \left(\begin{array}{cc} \frac{1}{T^{\gamma_0+1-2\delta-1}} \sum_{t=1}^{T} t^{\gamma_0+1-2\delta-2} & \frac{1}{T^{\gamma_0+\gamma-2\delta-1}} \sum_{t=1}^{T} t^{\gamma_0+\gamma-2\delta-2} \end{array} \right)
$$

$$
\times \left(\begin{array}{cc} \frac{1}{T^{2(1-\delta)-1}} \sum_{t=1}^{T} t^{2(1-\delta-1)} & \frac{1}{T^{1+\gamma-2\delta-1}} \sum_{t=1}^{T} t^{1+\gamma-2\delta-2} \\ \frac{1}{T^{1+\gamma-2\delta-1}} \sum_{t=1}^{T} t^{1+\gamma-2\delta-2} & \frac{1}{T^{2(\gamma-\delta)-1}} \sum_{t=1}^{T} t^{2(\gamma-\delta-1)} \end{array} \right)^{-1}
$$

$$
\times \left(\begin{array}{c} \frac{1}{T^{\gamma_0+1-2\delta-1}} \sum_{t=1}^{T} t^{\gamma_0+1-2\delta-2} \\ \frac{1}{T^{\gamma_0+\gamma-2\delta-1}} \sum_{t=1}^{T} t^{\gamma_0+\gamma-2\delta-2} \end{array} \right) + q_{3T}(\boldsymbol{\vartheta}),
$$

where $\sup_{\{\gamma_0-\delta \geq 1/2+g\} \cap \Phi_9} |q_{3T}(\boldsymbol{\vartheta})| = o(1)$. Then, using (158) of Lemma 11 and approximating sums by integrals, see Lemma 10,

$$
\inf_{\{\gamma_0-\delta \geq 1/2+g, |\gamma-\gamma_0| \geq g\} \cap \Phi_9} \frac{1}{T^{2(\gamma_0-\delta)-1}} \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta})
$$

$$
\geq \epsilon \inf_{\{\gamma_0-\delta \geq 1/2+g, |\gamma-\gamma_0| \geq g\} \cap \Phi_9} \frac{(\gamma_0-1)^{2\cdot}(\gamma_0-\gamma)^2}{\Gamma^2(\gamma_0-\delta)(2(\gamma_0-\delta)-1)(\gamma_0-2\delta)^2(\gamma_0+\gamma-2\delta-1)^2}
$$

$$
\geq \epsilon \frac{\kappa^2 g^2}{2g(\alpha+g)^4},
$$

which is positive and bounded away from zero, to complete the proof (203) for $j = 9$, and hence that of Lemma 2.

9.3   Proof of Lemma 3

To prove (145), first note that $\partial \bar{d}_T(\boldsymbol{\vartheta})/\partial\gamma$ equals

$$
-\boldsymbol{\phi}_0' \sum_{t=1}^{T} \boldsymbol{c}_{t-1}(1-\delta,\gamma_0-\delta,\boldsymbol{\varphi}) \left(\frac{\partial}{\partial\gamma} \boldsymbol{c}_{t-1}'(1-\delta,\gamma-\delta,\boldsymbol{\varphi}) \right)
$$

$$
\times \left(\sum_{j=1}^{T} \boldsymbol{c}_{j-1}(1-\delta,\gamma-\delta,\boldsymbol{\varphi}) \boldsymbol{c}_{j-1}'(1-\delta,\gamma-\delta,\boldsymbol{\varphi}) \right)^{-1} \sum_{s=1}^{T} \boldsymbol{c}_{s-1}(1-\delta,\gamma-\delta,\boldsymbol{\varphi})
$$

$$
-\boldsymbol{\phi}_0' \sum_{t=1}^{T} \boldsymbol{c}_{t-1}(1-\delta,\gamma_0-\delta,\boldsymbol{\varphi}) \boldsymbol{c}_{t-1}'(1-\delta,\gamma-\delta,\boldsymbol{\varphi})
$$

$$
\times \left(\sum_{j=1}^{T} \boldsymbol{c}_{j-1}(1-\delta,\gamma-\delta,\boldsymbol{\varphi}) \boldsymbol{c}_{j-1}'(1-\delta,\gamma-\delta,\boldsymbol{\varphi}) \right)^{-1} \sum_{s=1}^{T} \frac{\partial}{\partial\gamma} \boldsymbol{c}_{s-1}(1-\delta,\gamma-\delta,\boldsymbol{\varphi})
$$

$$
+\boldsymbol{\phi}_0' \sum_{t=1}^{T} \boldsymbol{c}_{t-1}(1-\delta,\gamma_0-\delta,\boldsymbol{\varphi}) \boldsymbol{c}_{t-1}'(1-\delta,\gamma-\delta,\boldsymbol{\varphi})
$$

$$
\times \left(\sum_{j=1}^{T} \boldsymbol{c}_{j-1}(1-\delta,\gamma-\delta,\boldsymbol{\varphi}) \boldsymbol{c}_{j-1}'(1-\delta,\gamma-\delta,\boldsymbol{\varphi}) \right)^{-1}
$$

$$
\times \left(\frac{\partial}{\partial\gamma} \sum_{j=1}^{T} \boldsymbol{c}_{j-1}(1-\delta,\gamma-\delta,\boldsymbol{\varphi}) \boldsymbol{c}_{j-1}'(1-\delta,\gamma-\delta,\boldsymbol{\varphi}) \right)
$$

$$
\times \left(\sum_{j=1}^{T} \boldsymbol{c}_{j-1}(1-\delta,\gamma-\delta,\boldsymbol{\varphi}) \boldsymbol{c}_{j-1}'(1-\delta,\gamma-\delta,\boldsymbol{\varphi}) \right)^{-1} \sum_{s=1}^{T} \boldsymbol{c}_{s-1}(1-\delta,\gamma-\delta,\boldsymbol{\varphi}).
$$

Note that in $\cup_{i=1}^{4} \overline{\mathcal{H}}_i$ we have $\gamma_0 - \delta \geq 1+\eta$ and $\gamma - \delta \geq 1+\eta-\varrho$. Setting $\Upsilon_1 = \{\boldsymbol{\vartheta} \in \Xi : 1-\delta \leq 1/2-\alpha\}$ and $\Upsilon_2 = \{\boldsymbol{\vartheta} \in \Xi : 1/2-\alpha \leq 1-\delta \leq 1/2+\alpha\}$, we show (145) with

$$
\sup_{\overline{\mathcal{H}}_i, \Upsilon_j} |g_T(\boldsymbol{\vartheta})| = o(1) \text{ for } j = 1, 2.
$$

We first give the proof for $j = 1$. By (159), (160) of Lemma 12 it follows immediately that $\sup_{\overline{\mathcal{H}}_i, \Upsilon_1} \left| \sum_{t=1}^{T} c_{t-1}(1-\delta,\boldsymbol{\varphi}) \right| = O(T^{1/2-\alpha})$ and $\sup_{\overline{\mathcal{H}}_i, \Upsilon_1} T^{\delta-\gamma} \left| \sum_{t=1}^{T} c_{t-1}(\gamma-\delta,\boldsymbol{\varphi}) \right| = O(1)$. Using Lemma 13 and (165), (166), (170) of Lemma 14, it can then straightforwardly be

shown that

$$\frac{1}{T^{\gamma_0-\delta}}\frac{\partial \bar{d}_T(\boldsymbol{\vartheta})}{\partial \gamma} =$$

$$-\frac{\beta_0\rho(1;\boldsymbol{\varphi})}{T^{\gamma_0-\delta}}\left(\frac{\sum_{j=1}^T \pi_{j-1}^{(1)}(\gamma-\delta)\sum_{j=1}^T \pi_{j-1}(\gamma_0-\delta)\pi_{j-1}(\gamma-\delta)}{\sum_{j=1}^T \pi_{j-1}^2(\gamma-\delta)}\right.$$

$$-\frac{2\sum_{j=1}^T \pi_{j-1}(\gamma-\delta)\sum_{j=1}^T \pi_{j-1}(\gamma_0-\delta)\pi_{j-1}(\gamma-\delta)\sum_{j=1}^T \pi_{j-1}(\gamma-\delta)\pi_{j-1}^{(1)}(\gamma-\delta)}{(\sum_{j=1}^T \pi_{j-1}^2(\gamma-\delta))^2}$$

$$\left.+\frac{\sum_{j=1}^T \pi_{j-1}(\gamma-\delta)\sum_{j=1}^T \pi_{j-1}(\gamma_0-\delta)\pi_{j-1}^{(1)}(\gamma-\delta)}{\sum_{j=1}^T \pi_{j-1}^2(\gamma-\delta)}\right) + g_{1,T}(\boldsymbol{\vartheta}), \tag{207}$$

where $\sup_{\overline{\mathcal{H}}_i,\Upsilon_1}|g_{1,T}(\boldsymbol{\vartheta})| = o(1)$. Now, substituting (122) into (207) (noting that the contribution of $\psi(d)$ cancels), approximating $\psi(d+j)$ by $\log j$, introducing $\log T$ terms (which cancel) and using (158) in Lemma 11, it can be shown that

$$\frac{1}{T^{\gamma_0-\delta}}\frac{\partial \bar{d}_T(\boldsymbol{\vartheta})}{\partial \gamma} = -\frac{\beta_0\rho(1;\boldsymbol{\varphi})}{\Gamma(\gamma_0-\delta)T^{\gamma_0-\delta}}\left(\frac{\sum_{j=1}^T \log(j/T)j^{\gamma-\delta-1}\sum_{j=1}^T j^{\gamma_0+\gamma-2\delta-2}}{\sum_{j=1}^T j^{2\gamma-2\delta-2}}\right.$$

$$-\frac{2\sum_{j=1}^T j^{\gamma-\delta-1}\sum_{j=1}^T j^{\gamma_0+\gamma-2\delta-2}\sum_{j=1}^T \log(j/T)j^{2\gamma-2\delta-2}}{(\sum_{j=1}^T j^{2\gamma-2\delta-2})^2}$$

$$\left.+\frac{\sum_{j=1}^T j^{\gamma-\delta-1}\sum_{j=1}^T \log(j/T)j^{\gamma_0+\gamma-2\delta-2}}{\sum_{j=1}^T j^{2\gamma-2\delta-2}}\right) + g_{2,T}(\boldsymbol{\vartheta}), \tag{208}$$

where $\sup_{\overline{\mathcal{H}}_i,\Upsilon_1}|g_{2,T}(\boldsymbol{\vartheta})| = o(1)$. Finally, (145) then follows by approximating sums by integrals, see Lemma 10. The proof for $j = 2$ is almost identical to that for $j = 1$, just noting that, by (160) of Lemma 12, $\sup_{\overline{\mathcal{H}}_i,\Upsilon_2} T^{\delta-1}\left|\sum_{t=1}^T c_{t-1}(1-\delta,\boldsymbol{\varphi})\right| = O(1)$, and also using (163) (for α small enough) and (170) of Lemma 14.

Next we show (146). Define $\bar{d}_T(\boldsymbol{\tau},\gamma) = \bar{d}_T(\boldsymbol{\vartheta})$. Because $\bar{d}_T(\boldsymbol{\tau},\gamma_0) = 0$, the mean value theorem yields $\bar{d}_T(\boldsymbol{\tau},\gamma) = (\gamma-\gamma_0)\partial \bar{d}_T(\boldsymbol{\tau},\overline{\gamma})/\partial \gamma$, where $|\overline{\gamma}-\gamma_0| \leq |\gamma-\gamma_0|$, so the left-hand side of (146) can be bounded from below by

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

Thus, (146) follows if

$$\lim_{T\to\infty}\inf_{\overline{\mathcal{H}}_i,1-\delta\leq\frac{1}{2}+\alpha}\left(\frac{1}{T^{\gamma_0-\delta}}\frac{\partial \bar{d}_T(\boldsymbol{\tau},\gamma)}{\partial \gamma}\right)^2 > \epsilon, \tag{209}$$

which, noting that $\gamma_0 - \delta \geq 1 + \eta$ and $\gamma - \delta \geq 1 + \eta - \varrho$, is a consequence of (22) and (145) because

$$\inf_{\overline{\mathcal{H}}_i}\left(2(\gamma-\delta)^2 - 2(\gamma-\delta) + 1 - (\gamma_0-\delta)\right) = \inf_{\overline{\mathcal{H}}_i}\left(2(\gamma-\delta)^2 - 3(\gamma-\delta) + 1 - (\gamma_0-\gamma)\right)$$

$$\geq 2(\eta-\varrho)^2 + \eta - 2\varrho > 0.$$

9.4 Proof of Lemma 4

Letting $d_t(\boldsymbol{\tau}, \gamma) = d_t(\boldsymbol{\vartheta})$, noting (39) and that $d_t(\boldsymbol{\tau}, \gamma_0) = 0$, by the mean value theorem,

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

where $|\overline{\gamma} - \gamma_0| \leq |\gamma - \gamma_0|$ and

$$M_T(\boldsymbol{\vartheta}) = \frac{\partial}{\partial \gamma} \sum_{t=1}^T \rho(L; \boldsymbol{\varphi}) u_t (\delta - \delta_0) \boldsymbol{h}'_{t-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi})$$

$$\times \sum_{j=1}^T \boldsymbol{h}_{j-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \boldsymbol{c}'_{j-1}(1 - \delta, \gamma_0 - \delta, \boldsymbol{\varphi}) \phi_0.$$

Thus,

$$|M_T(\boldsymbol{\vartheta})| \leq K \sum_{j=1}^T \left| \sum_{t=1}^T \rho(L; \boldsymbol{\varphi}) u_t (\delta - \delta_0) \frac{\partial}{\partial \gamma} (\boldsymbol{h}'_{t-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \boldsymbol{h}_{j-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi})) \right|$$

$$\times \|\boldsymbol{c}_{j-1}(1 - \delta, \gamma_0 - \delta, \boldsymbol{\varphi})\| \,.$$

Then (148) follows by (159) of Lemma 12, noting that for $\gamma_0 \geq 1 + \kappa$ the second component of $\boldsymbol{c}_{j-1}(1-\delta, \gamma_0-\delta, \boldsymbol{\varphi})$ dominates, and (186) of Lemma 18. Similarly, (149)–(152) all follow using (160) of Lemma 12 together with (186), (188), (189), and (190) of Lemma 18, respectively.

9.5 Proof of Lemma 5

The proof is omitted as it is very similar to that of Lemma S.5 of Hualde and Nielsen (2020).

9.6 Proof of Lemma 6

As in the proof of Theorem 1(i), noting (32), (33), (41), (45), the result holds on establishing that

$$\Pr\left(\inf_{\boldsymbol{\vartheta} \in \overline{M}^*_\varepsilon} S_T(\boldsymbol{\vartheta}) \leq 0 \right) \to 0 \text{ as } T \to \infty, \qquad (210)$$

$$\Pr\left(\widehat{\boldsymbol{\vartheta}} \in \overline{N}^*_\varepsilon \cap M^*_\varepsilon, \inf_{\overline{N}^*_\varepsilon \cap M^*_\varepsilon} R_T(\widehat{\boldsymbol{\tau}}, \gamma) - R_T(\widehat{\boldsymbol{\tau}}, \gamma_0) \leq 0 \right) \to 0 \text{ as } T \to \infty, \qquad (211)$$

where

$$M^*_\varepsilon = \left\{ \boldsymbol{\vartheta} \in \Xi : \|\boldsymbol{\tau} - \boldsymbol{\tau}_0\| < \varepsilon T^{-\varkappa} \right\}, \quad \overline{M}^*_\varepsilon = \left\{ \boldsymbol{\vartheta} \in \Xi : \varepsilon T^{-\varkappa} \leq \|\boldsymbol{\tau} - \boldsymbol{\tau}_0\| < \varepsilon \right\},$$
$$N^*_\varepsilon = \left\{ \boldsymbol{\vartheta} \in \Xi : |\gamma - \gamma_0| < \varepsilon T^{-\varkappa} \right\}, \quad \overline{N}^*_\varepsilon = \left\{ \boldsymbol{\vartheta} \in \Xi : \varepsilon T^{-\varkappa} \leq |\gamma - \gamma_0| < \varepsilon \right\}.$$

We first prove (210), which, defining $\mathcal{J}_i = \{ \boldsymbol{\vartheta} \in \overline{M}^*_\varepsilon : \delta \in \mathcal{I}_i \}$ for $i = 4, 5$, holds if

$$\Pr\left(\inf_{\mathcal{J}_i} S_T(\boldsymbol{\vartheta}) \leq 0 \right) \to 0 \text{ as } T \to \infty \qquad (212)$$

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, (212) for $i = 5$ would hold if

$$\Pr\left(\inf_{\mathcal{J}_5} T^{2\varkappa} S_T(\boldsymbol{\vartheta}) \leq 0 \right) \to 0 \text{ as } T \to \infty. \qquad (213)$$

Proceeding as in the proof of (46)–(48) for $i = 5$, (213) holds if

$$\inf_{\mathcal{J}_5} T^{2\varkappa} U(\boldsymbol{\tau}) > \epsilon, \tag{214}$$

$$\frac{1}{T^{1-2\varkappa}} \sum_{t=1}^{T} \left((\rho(L; \boldsymbol{\varphi}_0)(u_t \mathbb{I}(t > 0)))^2 - \sigma_0^2 \right) = o_p(1), \tag{215}$$

$$\sup_{\mathcal{J}_5} \frac{1}{T^{1-2\varkappa}} \sum_{t=1}^{T} \left((\rho(L; \boldsymbol{\varphi}) u_t (\delta - \delta_0))^2 - E((\rho(L; \boldsymbol{\varphi}) \Delta^{\delta - \delta_0} u_t)^2) \right) = o_p(1), \tag{216}$$

$$\sup_{\mathcal{J}_5} \frac{1}{T^{1-2\varkappa}} \left| \sum_{t=1}^{T} d_t(\boldsymbol{\vartheta}) s_t(\boldsymbol{\vartheta}) \right| = o_p(1), \tag{217}$$

$$\sup_{\mathcal{J}_5} \frac{1}{T^{1-2\varkappa}} \left(\sum_{t=1}^{T} \rho(L; \boldsymbol{\varphi}) u_t (\delta - \delta_0) \boldsymbol{h}_{t-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \right)^2 = o_p(1). \tag{218}$$

Clearly (214)-(216) hold as in (S.152)-(S.154) of Hualde and Nielsen (2020) and the proofs of (217), (218) are omitted as, for small enough $\varkappa$, they follow by almost identical arguments to those of (53), (54), respectively.

Next, the proof of (212) for $i = 4$ is omitted because it is implied by almost identical results to (46)–(48) 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 (210).

Finally, we prove (211). For the same reason as in the proof of (41), we need to prove that

$$\inf_{\boldsymbol{\vartheta} \in \overline{N}_\varepsilon^* \cap M_\varepsilon^*} \frac{T^{2\varkappa}}{T^{2(\gamma_0 - \delta) - 1}} \sum_{t=1}^{T} d_t^2(\boldsymbol{\vartheta}) > \epsilon, \tag{219}$$

$$\sup_{\boldsymbol{\vartheta} \in \overline{N}_\varepsilon^* \cap M_\varepsilon^*} \frac{T^{2\varkappa}}{T^{2(\gamma_0 - \delta) - 1}} \left| \sum_{t=1}^{T} d_t(\boldsymbol{\vartheta}) s_{1t}(\boldsymbol{\tau}) \right| = o_p(1), \tag{220}$$

$$\sup_{\boldsymbol{\vartheta} \in \overline{N}_\varepsilon^* \cap M_\varepsilon^*} \frac{T^{2\varkappa}}{T^{2(\gamma_0 - \delta) - 1}} \left(\sum_{j=1}^{T} s_{1j}(\boldsymbol{\tau}) \boldsymbol{h}_{j-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \right)^2 = o_p(1). \tag{221}$$

As in (60), the proof of (219) follows by Lemma 3, whereas the proofs of (220) and (221) hold as in (43) and (44) for $\varkappa > 0$ sufficiently small.

9.7 Proof of Lemma 7

First we show (155). Clearly,

$$c_t(\widehat{\delta}, \widehat{\boldsymbol{\varphi}}) - c_t(\delta_0, \boldsymbol{\varphi}_0) = \sum_{j=0}^{t} (\rho_j(\widehat{\boldsymbol{\varphi}}) - \rho_j(\boldsymbol{\varphi}_0)) \pi_{t-j}(\delta_0) + \sum_{j=0}^{t} (\pi_{t-j}(\widehat{\delta}) - \pi_{t-j}(\delta_0)) \rho_j(\boldsymbol{\varphi}_0)$$

$$+ \sum_{j=0}^{t} (\rho_j(\widehat{\boldsymbol{\varphi}}) - \rho_j(\boldsymbol{\varphi}_0))(\pi_{t-j}(\widehat{\delta}) - \pi_{t-j}(\delta_0)). \tag{222}$$

Fix $\epsilon < 1/2$. Then

$$\rho_j(\widehat{\boldsymbol{\varphi}}) - \rho_j(\boldsymbol{\varphi}_0) = (\rho_j(\widehat{\boldsymbol{\varphi}}) - \rho_j(\boldsymbol{\varphi}_0))(\mathbb{I}(\|\widehat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0\| < \epsilon) + \mathbb{I}(\|\widehat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0\| \geq \epsilon)), \tag{223}$$

so by the mean value theorem the left-hand side of (223) is bounded by

$$\sup_{\|\boldsymbol{\varphi}-\boldsymbol{\varphi}_0\|<\epsilon} \left\| \frac{\partial \rho_j(\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}} \right\| \|\widehat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0\| + K \sup_{\boldsymbol{\varphi}\in\Psi} |\rho_j(\boldsymbol{\varphi})| \frac{\|\widehat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0\|^N}{\epsilon^N}, \tag{224}$$

for any arbitrarily large fixed number N. By (21) and the $T^{-\varkappa}$-consistency of $\widehat{\boldsymbol{\tau}}$, the second term in (224) is of smaller order, whereas by (23) the first term is $O_p(T^{-\varkappa}j^{-1-\varsigma})$. This implies that the first term on the right-hand side of (222) is $O_p(T^{-\varkappa}t^{\max\{\delta_0-1,-1-\varsigma\}}\log t)$ by Lemmas 9 and 11. Next, by straightforward application of Lemma 11 and a second-order Taylor expansion, $\pi_{t-j}(\widehat{\delta}) - \pi_{t-j}(\delta_0) = O_p((t-j)^{\delta_0-1}(\log t)T^{-\varkappa})$, so by (21) and Lemma 9 the second term on the right-hand side of (222) is $O_p(T^{-\varkappa}t^{\max\{\delta_0-1,-1-\varsigma\}}(\log t)^2)$. Finally, combining the arguments for the first two terms, the third term on the right-hand side of (222) is of smaller order, to conclude the proof of (155).

The proof of (156) is omitted because it is almost identical to that of (155) with the only difference that the coefficients $\pi_{t-j}^{(1)}(\cdot)$ instead of $\pi_{t-j}(\cdot)$ lead to an extra $(\log t)$-factor; see Lemma 11. Finally, we show (157). Clearly, the left-hand side of (157) is

$$\sum_{j=0}^{t-1} \rho_j(\boldsymbol{\varphi}_0)u_{t-j} + \sum_{j=0}^{t-1} (\rho_j(\widehat{\boldsymbol{\varphi}}) - \rho_j(\boldsymbol{\varphi}_0))u_{t-j} + \sum_{j=0}^{t-1} \rho_j(\widehat{\boldsymbol{\varphi}})(u_{t-j}(\widehat{\delta} - \delta_0) - u_{t-j}). \tag{225}$$

Using the mean value theorem as in (223) and (224) and summation by parts, it can be shown that the second term in (225) is $O_p(T^{-\varkappa})$. Similarly, by Lemma C.5 of Robinson and Hualde (2003) and (21), the third term in (225) is also $O_p(T^{-\varkappa})$, to conclude the proof of (157).

9.8 Proof of Lemma 8

First, for any $\epsilon > 0$, clearly

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

Then, in view of Theorem 1(ii) and (141), the result holds on showing

$$\sup_{\boldsymbol{\vartheta}\in M_\varepsilon} \frac{1}{T^{1/2}} \sum_{t=1}^{T} (d_t(\boldsymbol{\vartheta}) - s_{2t}(\boldsymbol{\vartheta})) \left(\frac{\partial d_t(\boldsymbol{\vartheta})}{\partial\boldsymbol{\tau}} - \frac{\partial s_{2t}(\boldsymbol{\vartheta})}{\partial\boldsymbol{\tau}} \right) = o_p(1), \tag{226}$$

$$\sup_{\boldsymbol{\vartheta}\in M_\varepsilon} \frac{1}{T^{1/2}} \sum_{t=1}^{T} s_{1t}(\boldsymbol{\tau}) \left(\frac{\partial d_t(\boldsymbol{\vartheta})}{\partial\boldsymbol{\tau}} - \frac{\partial s_{2t}(\boldsymbol{\vartheta})}{\partial\boldsymbol{\tau}} \right) = o_p(1), \tag{227}$$

$$\sup_{\boldsymbol{\vartheta}\in M_\varepsilon} \frac{1}{T^{1/2}} \sum_{t=1}^{T} \frac{\partial s_{1t}(\boldsymbol{\tau})}{\partial\boldsymbol{\tau}}(d_t(\boldsymbol{\vartheta}) - s_{2t}(\boldsymbol{\vartheta})) = o_p(1), \tag{228}$$

recalling $M_\varepsilon = \{\boldsymbol{\vartheta} \in \Xi : \|\boldsymbol{\tau} - \boldsymbol{\tau}_0\| < \varepsilon\}$.

The proof of (226) follows upon showing that, for any $\theta > 0$ and ε such that $0 < \varepsilon < \theta$,

$$\sup_{\boldsymbol{\vartheta} \in M_\varepsilon} |d_t(\boldsymbol{\vartheta})| = O(t^{\max\{\gamma_0 - \delta_0 + \varepsilon - 1, -1 - \varsigma\}} + T^{2\theta} t^{-1/2 - \theta}), \tag{229}$$

$$\sup_{\boldsymbol{\vartheta} \in M_\varepsilon} |s_{2t}(\boldsymbol{\vartheta})| = O_p(T^{2\theta} t^{-1/2 - \theta}), \tag{230}$$

$$\sup_{\boldsymbol{\vartheta} \in M_\varepsilon} \left\| \frac{\partial d_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\tau}} \right\| = O(t^{\max\{\gamma_0 - \delta_0 + \varepsilon - 1, -1 - \varsigma\}} \log t + T^{5\theta} t^{-1/2 - \theta}), \tag{231}$$

$$\sup_{\boldsymbol{\vartheta} \in M_\varepsilon} \left\| \frac{\partial s_{2t}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\tau}} \right\| = O_p(T^{6\theta} t^{-1/2 - \theta}), \tag{232}$$

and then letting θ be sufficiently small. We only show (231) and (232) because the proofs for (229) and (230) are very similar but simpler. First, by (159) of Lemma 12,

$$\sup_{\boldsymbol{\vartheta} \in M_\varepsilon} |c_t(\gamma_0 - \delta, \boldsymbol{\varphi})| = O(t^{\max\{\gamma_0 - \delta_0 + \varepsilon - 1, -1 - \varsigma\}}), \tag{233}$$

and by a simple modification of that result,

$$\sup_{\boldsymbol{\vartheta} \in M_\varepsilon} \left\| \frac{\partial c_t(\gamma_0 - \delta, \boldsymbol{\varphi})}{\partial \boldsymbol{\tau}} \right\| = O(t^{\max\{\gamma_0 - \delta_0 + \varepsilon - 1, -1 - \varsigma\}} \log t). \tag{234}$$

Then (231) follows by direct application of (179) of Lemma 16, (233) and (234), noting that the bound in (179) also applies if the derivative is taken with respect to d_1 or $\boldsymbol{\varphi}$.

To prove (232), note from (38) that

$$\frac{\partial s_{2t}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\tau}} = \frac{\partial}{\partial \boldsymbol{\tau}} \left(\boldsymbol{h}'_{t-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \sum_{j=1}^{T} \boldsymbol{h}_{j-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \right) s_{1j}(\boldsymbol{\tau})$$

$$+ \boldsymbol{h}'_{t-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \sum_{j=1}^{T} \boldsymbol{h}_{j-1,T}(1 - \delta, \gamma - \delta, \boldsymbol{\varphi}) \frac{\partial s_{1j}(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}}. \tag{235}$$

First, the $\sup_{\boldsymbol{\vartheta} \in M_\varepsilon}$ of the absolute values of the first term on the right-hand side of (235) is $O_p(T^{6\theta} t^{-1/2 - \theta})$ by (186) of Lemma 18, noting that $\delta_0 - \delta \leq \varepsilon$ and that this bound also applies if the derivatives are taken with respect to $\boldsymbol{\tau}$. For the second term on the right-hand side of (235), noting that $\sum_{l=1}^{t} s_{1l}(\boldsymbol{\tau}) = \sum_{j=0}^{t-1} c_j(\delta_0 - \delta, \boldsymbol{\varphi}) \sum_{l=1}^{t-j} u_l$, it is straightforward to show that, by (159) of Lemma 12,

$$\sup_{\boldsymbol{\vartheta} \in M_\varepsilon} \left\| \sum_{j=1}^{t} \frac{\partial s_{1j}(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}} \right\| = O_p(t^{1/2 + \varepsilon} \log t). \tag{236}$$

Therefore, using summation by parts as in the proof of Lemma 18 and by (176) of Lemma 16, the $\sup_{\boldsymbol{\vartheta} \in M_\varepsilon}$ of the absolute value of the second term on the right-hand side of (235) is $O_p(T^{2\theta} t^{-1/2 - \theta})$, to justify (232) and hence (226).

Finally, (227) and (228) can be established by using summation by parts followed by direct application of the results in (229), (231), (236), and Lemma 17, noting also that by

previous arguments it can be straightforwardly shown that

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

$$\sup_{\boldsymbol{\vartheta} \in M_\varepsilon} \left\| \frac{\partial d_{t+1}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\tau}} - \frac{\partial d_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\tau}} \right\| = O(t^{\max\{\gamma_0 - \delta_0 + \varepsilon - 2, -1 - \varsigma\}} \log t + T^{5\theta} t^{-3/2 - \theta}),$$

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

$$\sup_{\boldsymbol{\vartheta} \in M_\varepsilon} \left\| \frac{\partial s_{2t+1}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\tau}} - \frac{\partial s_{2t}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\tau}} \right\| = O_p(T^{6\theta} t^{-3/2 - \theta}).$$

10 Proofs of technical lemmas

10.1 Proof of Lemma 9

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

10.2 Proof of Lemma 10

The proof of Lemma 10 is given in Lemma S.10 of Hualde and Nielsen (2020).

10.3 Proof of Lemma 11

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

10.4 Proof of Lemma 12

The proof of Lemma 12 is almost identical to that of Lemma 1 of Hualde and Robinson (2011) and is therefore omitted.

10.5 Proof of Lemma 13

First we show that

$$\frac{1}{T^{2d-1}} \sum_{t=1}^T c_{t-1}^2(d, \boldsymbol{\varphi}) \geq \frac{\rho^2(1; \boldsymbol{\varphi})}{T^{2d-1}} \sum_{t=1}^T \pi_{t-1}^2(d) - |r_{1,T}(d, \boldsymbol{\varphi})|. \tag{237}$$

By (17),

$$\sum_{t=1}^T c_{t-1}^2(d, \boldsymbol{\varphi}) \geq \rho^2(1; \boldsymbol{\varphi}) \sum_{t=1}^T \pi_{t-1}^2(d) - 2\rho(1; \boldsymbol{\varphi}) \sum_{t=1}^T \pi_{t-1}^2(d) \sum_{k=t}^\infty \rho_k(\boldsymbol{\varphi})$$

$$- 2\rho(1; \boldsymbol{\varphi}) \sum_{t=2}^T \pi_{t-1}(d) \sum_{k=0}^{t-2} \pi_{k+1}(d-1) \sum_{l=0}^k \rho_{t-1-l}(\boldsymbol{\varphi})$$

$$+ 2 \sum_{t=1}^T \pi_{t-1}(d) \sum_{j=t}^\infty \rho_j(\boldsymbol{\varphi}) \sum_{k=0}^{t-2} \pi_{k+1}(d-1) \sum_{l=0}^k \rho_{t-1-l}(\boldsymbol{\varphi}). \tag{238}$$

Because of (21), the fourth term on the right-hand side of (238) is of smaller order than the third term. Then the proof of (237) follows on showing

$$\sup_{d \geq 1/2+\eta, \varphi \in \Psi} \frac{1}{T^{2d-1}} \left| \sum_{t=1}^{T} \pi_{t-1}^2(d) \sum_{k=t}^{\infty} \rho_k(\varphi) \right| = o(1), \tag{239}$$

$$\sup_{d \geq 1/2+\eta, \varphi \in \Psi} \frac{1}{T^{2d-1}} \left| \sum_{t=2}^{T} \pi_{t-1}(d) \sum_{k=0}^{t-2} \pi_{k+1}(d-1) \sum_{l=0}^{k} \rho_{t-1-l}(\varphi) \right| = o(1). \tag{240}$$

First, by (21) and Lemma 11, the left-hand side of (239) is bounded by

$$K \sup_{d \geq 1/2+\eta} T \sum_{t=1}^{T} \left(\frac{t}{T} \right)^{2d} t^{-2-\varsigma} \leq KT \sum_{t=1}^{T} \left(\frac{t}{T} \right)^{1+2\eta} t^{-2-\varsigma}$$

$$\leq KT^{-2\eta} \sum_{t=1}^{T} t^{-1+2\eta-\varsigma} = o(1),$$

so (239) holds. Similarly, the left-hand side of (240) is bounded by

$$K \sup_{d \geq 1/2+\eta} T \sum_{t=1}^{T} \left(\frac{t}{T} \right)^{d} t^{-1} \sum_{k=1}^{t-1} \left(\frac{k}{T} \right)^{d} k^{-2} (t-k)^{-\varsigma}$$

$$\leq KT^{-2\eta} \sum_{t=1}^{T} t^{-1/2+\eta} \sum_{k=1}^{t-1} k^{-3/2+\eta} (t-k)^{-\varsigma}$$

$$\leq KT^{-2\eta} \sum_{t=1}^{T} t^{-1/2+\eta+\max\{\eta-3/2, -\varsigma, \eta-\varsigma-1/2\}} (1 + \log t) = o(1)$$

for any $\eta > 0$ because $\varsigma > 1/2$, where the second inequality is due to Lemma 9. This concludes the proof of (237). Next, we show that

$$\frac{1}{T^{2d-1}} \sum_{t=1}^{T} c_{t-1}^2(d, \varphi) \leq \frac{\rho^2(1; \varphi)}{T^{2d-1}} \sum_{t=1}^{T} \pi_{t-1}^2(d) + |r_{2,T}(d, \varphi)|, \tag{241}$$

which combined with (237) is sufficient to prove (161). In view of (17), the proof of (241) follows by (239), (240) and

$$\sup_{d \geq 1/2+\eta, \varphi \in \Psi} \frac{1}{T^{2d-1}} \sum_{t=1}^{T} \pi_{t-1}^2(d) \left(\sum_{k=t}^{\infty} \rho_k(\varphi) \right)^2 = o(1),$$

$$\sup_{d \geq 1/2+\eta, \varphi \in \Psi} \frac{1}{T^{2d-1}} \sum_{t=2}^{T} \left(\sum_{k=0}^{t-2} \pi_{k+1}(d-1) \sum_{l=0}^{k} \rho_{t-1-l}(\varphi) \right)^2 = o(1),$$

which follow by straightforward arguments using (21) and Lemma 9.

10.6 Proof of Lemma 14

The proof of (162) is immediate because

$$\inf_{\underline{d} \leq d \leq \overline{d}, \varphi \in \Psi} \sum_{t=1}^{T} c_{t-1}^2(d, \varphi) \geq \inf_{\underline{d} \leq d \leq \overline{d}, \varphi \in \Psi} c_0^2(d, \varphi) = 1.$$

For the proof of (163), we first truncate the summation,

$$\inf_{1/2-\alpha\leq d\leq 1/2+\alpha,\boldsymbol{\varphi}\in\Psi} T^{1-2d}\sum_{t=1}^{T}c_{t-1}^{2}(d,\boldsymbol{\varphi}) \geq \inf_{1/2-\alpha\leq d\leq 1/2+\alpha,\boldsymbol{\varphi}\in\Psi} T^{1-2d}\sum_{t=[T^{1/2}]}^{T}c_{t-1}^{2}(d,\boldsymbol{\varphi}). \quad (242)$$

Then, noting (238), the right-hand side of (242) is bounded from below by

$$\epsilon \inf_{1/2-\alpha\leq d\leq 1/2+\alpha}\frac{1}{T^{2d-1}}\sum_{t=[T^{1/2}]}^{T}\pi_{t-1}^{2}(d) \quad (243)$$

$$- \sup_{1/2-\alpha\leq d\leq 1/2+\alpha,\boldsymbol{\varphi}\in\Psi}\frac{K}{T^{2d-1}}\left|\sum_{t=[T^{1/2}]}^{T}\pi_{t-1}^{2}(d)\sum_{l=t}^{\infty}\rho_{l}(\boldsymbol{\varphi})\right| \quad (244)$$

$$- \sup_{1/2-\alpha\leq d\leq 1/2+\alpha,\boldsymbol{\varphi}\in\Psi}\frac{K}{T^{2d-1}}\left|\sum_{t=[T^{1/2}]}^{T}\pi_{t-1}(d)\sum_{k=0}^{t-2}\pi_{k+1}(d-1)\sum_{l=0}^{k}\rho_{t-1-l}(\boldsymbol{\varphi})\right|. \quad (245)$$

First, noting that $\sup_{d\geq 1/2-\alpha}(t/T)^{2d} = (t/T)^{1-2\alpha}$, by (21) and Lemma 11, it can be readily shown that (244) is $O(T^{\alpha-\varsigma/2}) = o(1)$ because $\alpha < (\varsigma-1/2)/3 < \varsigma/2$. Similarly, by identical arguments, (245) is bounded by

$$KT^{2\alpha}\sum_{t=[T^{1/2}]}^{T}t^{-1/2-\alpha}\sum_{k=1}^{t-1}k^{-3/2-\alpha}(t-k)^{-\varsigma} \leq KT^{2\alpha}\log T\sum_{t=[T^{1/2}]}^{T}t^{-1/2-\alpha}t^{-\varsigma}$$

$$\leq KT^{3\alpha/2+1/4-\varsigma/2}\log T$$

by Lemma 9. Then (245) is $o(1)$ because $\alpha < (\varsigma-1/2)/3$. Finally, by (158) in Lemma 11, (243) is bounded from below by

$$\epsilon \inf_{1/2-\alpha\leq d\leq 1/2+\alpha}\frac{1}{T^{2d-1}}\sum_{t=[T^{1/2}]}^{T}t^{2d-2} \geq \epsilon\frac{1}{T}\sum_{t=[T^{1/2}]}^{T}\left(\frac{t}{T}\right)^{2\alpha-1} \geq \epsilon\int_{[T^{1/2}]/T}^{1}x^{2\alpha-1}dx$$

$$= \epsilon\frac{1-([T^{1/2}]/T)^{2\alpha}}{2\alpha} = \frac{\epsilon}{2\alpha}-O\left(T^{-\alpha}\right).$$

In view of (242), (244), and (245), this proves (163).

Next, we show (164). By (161),

$$\inf_{1/2+\alpha\leq d\leq\bar{d},\boldsymbol{\varphi}\in\Psi}\frac{1}{T^{2d-1}}\sum_{t=1}^{T}c_{t-1}^{2}(d,\boldsymbol{\varphi}) \geq \inf_{1/2+\alpha\leq d\leq\bar{d},\boldsymbol{\varphi}\in\Psi}\frac{\rho^{2}(1;\boldsymbol{\varphi})}{T^{2d-1}}\sum_{t=1}^{T}\pi_{t-1}^{2}(d)-o(1),$$

so that (164) follows by (158) and approximating sums by integrals (Lemma 10), noting that $d\leq\bar{d}$.

Regarding the rest of the results, we omit the proofs of (165), (166), (167), (168), and (169) because they follow almost immediately from (159) and (160) of Lemma 12. Next, by

Lemma 1 of Hualde and Robinson (2011), see also (160) and Lemma 11, for T sufficiently large the left-hand side of (170) is bounded by

$$\sup_{\substack{d_1 \geq 1/2+\theta, d_2 \geq 1/2+\theta \\ d_1+d_2 \geq 1+\eta}} \frac{K}{T^{d_1+d_2-1}} \sum_{t=1}^{T} t^{\max\{d_1-1,-1-\varrho\}} t^{\max\{d_2-1,-1-\varrho\}} (1+\log t)^m$$

$$\leq \sup_{\substack{d_1 \geq 1/2+\theta, d_2 \geq 1/2+\theta \\ d_1+d_2 \geq 1+\eta}} \frac{K(\log T)^m}{T^{d_1+d_2-1}} \sum_{t=1}^{T} t^{d_1+d_2-2} \leq K(\log T)^m,$$

noting that the supremum occurs at $d_1 + d_2 = 1 + \eta > 1$ and applying Lemma 10.

10.7 Proof of Lemma 15

First, we show (171). It holds that

$$\sum_{t=1}^{T} c_{t-1}^2(d_1, \boldsymbol{\varphi}) \sum_{t=1}^{T} c_{t-1}^2(d_2, \boldsymbol{\varphi}) - \left(\sum_{t=1}^{T} c_{t-1}(d_1, \boldsymbol{\varphi}) c_{t-1}(d_2, \boldsymbol{\varphi}) \right)^2$$

$$= \sum_{t=1}^{T-1} \sum_{j=1, j \neq t}^{T} \left(c_{t-1}(d_1, \boldsymbol{\varphi}) c_{j-1}(d_2, \boldsymbol{\varphi}) - c_{t-1}(d_2, \boldsymbol{\varphi}) c_{j-1}(d_1, \boldsymbol{\varphi}) \right)^2, \tag{246}$$

so the left-hand side of (171) is bounded from below by

$$\inf_{d_1 \leq \frac{1}{2}-\alpha, d_2 \leq \frac{1}{2}-\alpha, |d_1-d_2| \geq \kappa} (c_1(d_1, \boldsymbol{\varphi}) - c_1(d_2, \boldsymbol{\varphi}))^2 = \inf_{d_1 \leq \frac{1}{2}-\alpha, d_2 \leq \frac{1}{2}-\alpha, |d_1-d_2| \geq \kappa} (\pi_1(d_1) - \pi_1(d_2))^2$$

$$= \inf_{d_1 \leq \frac{1}{2}-\alpha, d_2 \leq \frac{1}{2}-\alpha, |d_1-d_2| \geq \kappa} (d_1 - d_2)^2 = \kappa^2.$$

Next, (172) follows by showing that, uniformly in $|d_1 - d_2| \geq \kappa$,

$$\inf_{\frac{1}{2}+\alpha \leq d_1, d_2 \leq \frac{1}{2}-\alpha} \frac{1}{T^{2d_1-1}} \left(\sum_{t=1}^{T} c_{t-1}^2(d_1, \boldsymbol{\varphi}) - \frac{\left(\sum_{t=1}^{T} c_{t-1}(d_1, \boldsymbol{\varphi}) c_{t-1}(d_2, \boldsymbol{\varphi}) \right)^2}{\sum_{t=1}^{T} c_{t-1}^2(d_2, \boldsymbol{\varphi})} \right) \geq \epsilon, \tag{247}$$

$$\inf_{\frac{1}{2}+\alpha \leq d_1, \frac{1}{2}-\alpha \leq d_2 \leq \frac{1}{2}+\alpha} \frac{1}{T^{2d_1-1}} \left(\sum_{t=1}^{T} c_{t-1}^2(d_1, \boldsymbol{\varphi}) - \frac{\left(\sum_{t=1}^{T} c_{t-1}(d_1, \boldsymbol{\varphi}) c_{t-1}(d_2, \boldsymbol{\varphi}) \right)^2}{\sum_{t=1}^{T} c_{t-1}^2(d_2, \boldsymbol{\varphi})} \right) \geq \epsilon, \tag{248}$$

$$\inf_{\frac{1}{2}+\alpha \leq d_1, \frac{1}{2}+\alpha \leq d_2} \frac{1}{T^{2d_1-1}} \left(\sum_{t=1}^{T} c_{t-1}^2(d_1, \boldsymbol{\varphi}) - \frac{\left(\sum_{t=1}^{T} c_{t-1}(d_1, \boldsymbol{\varphi}) c_{t-1}(d_2, \boldsymbol{\varphi}) \right)^2}{\sum_{t=1}^{T} c_{t-1}^2(d_2, \boldsymbol{\varphi})} \right) \geq \epsilon. \tag{249}$$

First, (247) follows by (162), (164), and (166) (noting that replacing α by g in (166) just adds a $\log T$ factor to the bound). Next, (248) holds by (163), (164) and (170). Finally, (249) follows by very similar arguments to those in (199) for $j = 3$ in the proof of Lemma 2 and approximating sums by integrals (Lemma 10).

To prove (173), we note that the left-hand side is bounded from below, using (246), by

$$\inf_{\frac{1}{2}-\alpha\leq d_1\leq\frac{1}{2}+\alpha, d_2\leq\frac{1}{2}-\alpha, |d_1-d_2|\geq\kappa} \frac{1}{T^{2d_1-1}} \sum_{t=1}^{T} (c_{t-1}(d_1,\boldsymbol{\varphi}) - c_{t-1}(d_2,\boldsymbol{\varphi}))^2$$

$$\geq \inf_{\frac{1}{2}-\alpha\leq d_1\leq\frac{1}{2}+\alpha, d_2\leq\frac{1}{2}-\alpha, |d_1-d_2|\geq\kappa} \frac{1}{T^{2d_1-1}} \sum_{t=[T^{1/2}]}^{T} (c_{t-1}(d_1,\boldsymbol{\varphi}) - c_{t-1}(d_2,\boldsymbol{\varphi}))^2$$

$$\geq \inf_{\frac{1}{2}-\alpha\leq d_1\leq\frac{1}{2}+\alpha} \frac{1}{T^{2d_1-1}} \sum_{t=[T^{1/2}]}^{T} c_{t-1}^2(d_1,\boldsymbol{\varphi})$$

$$- \sup_{\frac{1}{2}-\alpha\leq d_1\leq\frac{1}{2}+\alpha, d_2\leq\frac{1}{2}-\alpha, |d_1-d_2|\geq\kappa} \frac{2}{T^{2d_1-1}} \sum_{t=[T^{1/2}]}^{T} c_{t-1}(d_1,\boldsymbol{\varphi})c_{t-1}(d_2,\boldsymbol{\varphi}). \tag{250}$$

The first term on the right-hand side of (250) is bounded from below by $\epsilon/\alpha + o(1)$ as in the proof of (163) of Lemma 14. Using (159) and taking into account the range of values that d_1 and d_2 could take, the second term on the right-hand side of (250) is bounded by

$$\sup_{\frac{1}{2}-\alpha\leq d_1\leq\frac{1}{2}+\alpha, d_2\leq\frac{1}{2}-\alpha, |d_1-d_2|\geq\kappa} KT \sum_{t=[T^{1/2}]}^{T} \frac{1}{T^{d_1}} \left(\frac{t}{T}\right)^{d_1} t^{d_2-2}. \tag{251}$$

Clearly the supremum is attained at the minimum value that d_1 could take and at the maximum value that d_2 could take. Noting that $|d_1 - d_2| \geq \kappa$, this necessarily occurs at $d_1 = 1/2 - \alpha + \theta\kappa$, $d_2 = 1/2 - \alpha - (1-\theta)\kappa$ for some $\theta \in [0, 2\alpha/\kappa]$. Thus, (251) is bounded by

$$KT \sum_{t=[T^{1/2}]}^{T} \frac{1}{T^{1/2-\alpha+\theta\kappa}} \left(\frac{t}{T}\right)^{1/2-\alpha+\theta\kappa} t^{1/2-\alpha-(1-\theta)\kappa-2} = O\left(T^{\alpha-\theta\kappa-\kappa/2} + T^{-\kappa}\log T\right) = o(1),$$

to conclude the proof of (173). Finally, the proofs of (174) and (175) follow straightforwardly by (163) and (170).

10.8   Proof of Lemma 16

Without loss of generality we assume that $d_1 > d_2$ in the proofs of (176)–(180). Letting

$$h_{t,T}(d,\boldsymbol{\varphi}) = \frac{c_t(d,\boldsymbol{\varphi})}{\left(\sum_{j=1}^{T} c_{j-1}^2(d,\boldsymbol{\varphi})\right)^{1/2}}, \tag{252}$$

we initially show that

$$\sup_{d\in[\underline{d}_1,\overline{d}_1],\boldsymbol{\varphi}\in\Psi} |h_{t-1,T}(d_1,\boldsymbol{\varphi})| = O(t^{-1/2-\theta}T^{\theta}). \tag{253}$$

The left-hand side of (253) is bounded by

$$\sup_{\underline{d}_1\leq d_1\leq 1/2-\theta,\boldsymbol{\varphi}\in\Psi} |h_{t-1,T}(d_1,\boldsymbol{\varphi})| + \sup_{1/2-\theta\leq d_1\leq\overline{d}_1,\boldsymbol{\varphi}\in\Psi} |h_{t-1,T}(d_1,\boldsymbol{\varphi})|. \tag{254}$$

Using the definition (252) and applying (159) of Lemma 12 and (162) of Lemma 14, the first term of (254) is bounded by

$$
\sup_{\underline{d}_1 \leq d_1 \leq 1/2-\theta, \boldsymbol{\varphi} \in \Psi} |h_{t-1,T}(d_1, \boldsymbol{\varphi})| \leq \frac{\sup_{\underline{d}_1 \leq d_1 \leq 1/2-\theta, \boldsymbol{\varphi} \in \Psi} |c_{t-1}(d_1, \boldsymbol{\varphi})|}{\inf_{\underline{d}_1 \leq d_1 \leq 1/2-\theta, \boldsymbol{\varphi} \in \Psi} \left(\sum_{j=1}^{T} c_{j-1}^2(d_1, \boldsymbol{\varphi}) \right)^{1/2}}
$$

$$
\leq \sup_{\underline{d}_1 \leq d_1 \leq 1/2-\theta, \boldsymbol{\varphi} \in \Psi} |c_{t-1}(d_1, \boldsymbol{\varphi})| = O(t^{-1/2-\theta}), \qquad (255)
$$

so the bound in (253) applies to the first term of (254) (although it is not tight). Next, the second term of (254) is bounded by

$$
\frac{\sup_{1/2-\theta \leq d_1 \leq \bar{d}_1, \boldsymbol{\varphi} \in \Psi} T^{-d_1} |c_{t-1}(d_1, \boldsymbol{\varphi})|}{\inf_{1/2-\theta \leq d_1 \leq \bar{d}_1, \boldsymbol{\varphi} \in \Psi} \left(\sum_{j=1}^{T} T^{-2d_1} c_{j-1}^2(d_1, \boldsymbol{\varphi}) \right)^{1/2}}.
$$

By (160) of Lemma 12 the numerator is $O(t^{-1}(t/T)^{1/2-\theta})$ and by Lemma 13 the denominator is bounded from below by $\epsilon \theta^{-1/2} T^{-1/2}$. Thus (253) follows.

We now prove (176). Clearly,

$$
\|\boldsymbol{h}_{t-1,T}(d_1, d_2, \boldsymbol{\varphi})\|^2 = \frac{N_{1t} + N_{2t} + N_{3t}}{D_T}
$$

where

$$
N_{1t} = c_{t-1}^2(d_1, \boldsymbol{\varphi}) \sum_{j=1}^{T} c_{j-1}^2(d_2, \boldsymbol{\varphi}),
$$

$$
N_{2t} = c_{t-1}^2(d_2, \boldsymbol{\varphi}) \sum_{j=1}^{T} c_{j-1}^2(d_1, \boldsymbol{\varphi}),
$$

$$
N_{3t} = -2 c_{t-1}(d_1, \boldsymbol{\varphi}) c_{t-1}(d_2, \boldsymbol{\varphi}) \sum_{j=1}^{T} c_{j-1}(d_1, \boldsymbol{\varphi}) c_{j-1}(d_2, \boldsymbol{\varphi}),
$$

$$
D_T = \sum_{j=1}^{T} c_{j-1}^2(d_1, \boldsymbol{\varphi}) \sum_{j=1}^{T} c_{j-1}^2(d_2, \boldsymbol{\varphi}) - \left(\sum_{j=1}^{T} c_{j-1}(d_1, \boldsymbol{\varphi}) c_{j-1}(d_2, \boldsymbol{\varphi}) \right)^2. \qquad (256)
$$

By application of the Cauchy-Schwarz inequality to the numerator and (246) to the denominator,

$$
\|\boldsymbol{h}_{t-1,T}(d_1, d_2, \boldsymbol{\varphi})\|^2 \leq \frac{2(N_{1t} + N_{2t})}{D_T} \qquad (257)
$$

$$
= \frac{2(N_{1t} + N_{2t})}{\sum_{k=1}^{T-1} \sum_{j=1, j\neq k}^{T} \left(c_{k-1}(d_1, \boldsymbol{\varphi}) c_{j-1}(d_2, \boldsymbol{\varphi}) - c_{k-1}(d_2, \boldsymbol{\varphi}) c_{j-1}(d_1, \boldsymbol{\varphi}) \right)^2}
$$

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

$$
\qquad (258)
$$

We will prove (176) by considering several parts of the parameter space in turn. First, we show that

$$\sup_{\substack{\underline{d}_1 \leq d_1 \leq 1/2-\theta, \underline{d}_2 \leq d_2 \leq 1/2-\theta \\ \varphi \in \Psi, |d_1-d_2| \geq \kappa}} \|\boldsymbol{h}_{t-1,T}(d_1, d_2, \boldsymbol{\varphi})\|^2 = O(t^{-1-2\theta}T^{2\theta}), \tag{259}$$

noting that in this region necessarily $d_2 \leq 1/2 - \theta - \kappa$. Using the bounds (253) and (258), (259) follows by showing that

$$\inf_{\substack{\underline{d}_1 \leq d_1 \leq 1/2-\theta, \underline{d}_2 \leq d_2 \leq 1/2-\theta \\ \varphi \in \Psi, |d_1-d_2| \geq \kappa}} \sum_{k=1}^{T-1} \sum_{j \neq t=1}^{T} (h_{k-1,T}(d_1, \boldsymbol{\varphi})h_{j-1,T}(d_2, \boldsymbol{\varphi}) - h_{k-1,T}(d_2, \boldsymbol{\varphi})h_{j-1,T}(d_1, \boldsymbol{\varphi}))^2$$

$$= \inf_{\substack{\underline{d}_1 \leq d_1 \leq 1/2-\theta, \underline{d}_2 \leq d_2 \leq 1/2-\theta \\ \varphi \in \Psi, |d_1-d_2| \geq \kappa}} \frac{\sum_{k=1}^{T-1} \sum_{j \neq t=1}^{T} (c_{k-1}(d_1, \boldsymbol{\varphi})c_{j-1}(d_2, \boldsymbol{\varphi}) - c_{k-1}(d_2, \boldsymbol{\varphi})c_{j-1}(d_1, \boldsymbol{\varphi}))^2}{\sum_{t=1}^{T} c_{t-1}^2(d_1, \boldsymbol{\varphi}) \sum_{j=1}^{T} c_{j-1}^2(d_2, \boldsymbol{\varphi})} > \epsilon. \tag{260}$$

The supremum of the denominator in (260) is $O(1)$ by (165), while the numerator is bounded from below by

$$\inf_{\substack{\underline{d}_1 \leq d_1 \leq 1/2-\theta, \underline{d}_2 \leq d_2 \leq 1/2-\theta \\ \varphi \in \Psi, |d_1-d_2| \geq \kappa}} \sum_{k=1}^{T-1} \sum_{j \neq t=1}^{T} (c_{k-1}(d_1, \boldsymbol{\varphi})c_{j-1}(d_2, \boldsymbol{\varphi}) - c_{k-1}(d_2, \boldsymbol{\varphi})c_{j-1}(d_1, \boldsymbol{\varphi}))^2$$

$$\geq \inf_{\substack{\underline{d}_1 \leq d_1 \leq 1/2-\theta, \underline{d}_2 \leq d_2 \leq 1/2-\theta \\ \varphi \in \Psi, |d_1-d_2| \geq \kappa}} (c_1(d_1, \boldsymbol{\varphi}) - c_1(d_2, \boldsymbol{\varphi}))^2 = \inf_{|d_1-d_2| \geq \kappa} (d_1 - d_2)^2 = \kappa^2$$

as in the proof of (171).

We next show that

$$\sup_{1/2-\theta \leq d_1 \leq 1/2+\theta, \varphi \in \Psi, |d_1-d_2| \geq \kappa} \|\boldsymbol{h}_{t-1,T}(d_1, d_2, \boldsymbol{\varphi})\|^2 = O(t^{-1-2\theta}T^{2\theta}), \tag{261}$$

noting that in this region necessarily $d_2 \leq 1/2 + \theta - \kappa < 1/2$. By (246), (252), and (258), the left-hand side of (261) is bounded by

$$\sup_{\substack{1/2-\theta \leq d_1 \leq 1/2+\theta \\ \varphi \in \Psi, |d_1-d_2| \geq \kappa}} 2 \frac{T^{1-2d_1} c_{t-1}^2(d_1, \boldsymbol{\varphi}) \sum_{j=1}^{T} c_{j-1}^2(d_2, \boldsymbol{\varphi}) + c_{t-1}^2(d_2, \boldsymbol{\varphi}) T^{1-2d_1} \sum_{j=1}^{T} c_{j-1}^2(d_2, \boldsymbol{\varphi})}{T^{1-2d_1} \sum_{k=[T^{1/2}]}^{T} (c_{k-1}(d_1, \boldsymbol{\varphi}) - c_{k-1}(d_2, \boldsymbol{\varphi}))^2}.$$

By simple application of (160) of Lemma 12 and (165) of Lemma 14,

$$\sup_{1/2-\theta \leq d_1 \leq 1/2+\theta, \varphi \in \Psi, |d_1-d_2| \geq \kappa} \frac{1}{T^{2d_1-1}} c_{t-1}^2(d_1, \boldsymbol{\varphi}) \sum_{j=1}^{T} c_{j-1}^2(d_2, \boldsymbol{\varphi}) = O\left(t^{-1-2\theta}T^{2\theta}\right). \tag{262}$$

Furthermore,

$$\sup_{\substack{1/2-\theta \leq d_1 \leq 1/2+\theta \\ \varphi \in \Psi, |d_1-d_2| \geq \kappa}} c_{t-1}^2(d_2, \boldsymbol{\varphi}) \frac{1}{T^{2d_1-1}} \sum_{j=1}^{T} c_{j-1}^2(d_1, \boldsymbol{\varphi}) = O(t^{-1+2\theta-2\kappa}T^{2\theta}) = O(t^{-1-2\theta}T^{2\theta}), \tag{263}$$

where the first equality in (263) follows by (159) of Lemma 12 and (168) of Lemma 14, whereas the second holds because $\theta \leq \kappa/2$. Then, in view of (262) and (263), the proof of (261) follows by showing that

$$\inf_{1/2-\theta \leq d_1 \leq 1/2+\theta, \boldsymbol{\varphi} \in \Psi} \frac{1}{T^{2d_1-1}} \sum_{k=[T^{1/2}]}^{T} c_{k-1}^2(d_1, \boldsymbol{\varphi}) > \epsilon, \tag{264}$$

$$\sup_{1/2-\theta \leq d_1 \leq 1/2+\theta, \boldsymbol{\varphi} \in \Psi, |d_1-d_2| \geq \kappa} \frac{1}{T^{2d_1-1}} \sum_{k=[T^{1/2}]}^{T} c_{k-1}(d_1, \boldsymbol{\varphi}) c_{k-1}(d_2, \boldsymbol{\varphi}) = o(1). \tag{265}$$

First, (264) follows as in the proof of (163). Noting (159) and (160), the supremum in (265) occurs at the smallest possible value for d_1 and the largest one for d_2, which is $d_2 \leq d_1 - \kappa$. Thus, by (159) and (160), the left-hand side of (265) is bounded by

$$KT \sum_{k=[T^{1/2}]}^{T} \left(\frac{k}{T}\right)^{d_1} k^{-1} \left(\frac{k}{T}\right)^{d_1} k^{-\kappa-1} \leq KT^{2\theta} \sum_{k=[T^{1/2}]}^{T} k^{-1-2\theta-\kappa} = O\left(T^{\theta-\kappa/2}\right),$$

so that (265) is justified and the proof of (261) concluded.

Next, we show that

$$\sup_{\substack{1/2+\theta \leq d_1 \leq \bar{d}_1, \underline{d}_2 \leq d_2 \leq 1/2-\theta \\ \boldsymbol{\varphi} \in \Psi, |d_1-d_2| \geq \kappa}} \|\boldsymbol{h}_{t-1,T}(d_1, d_2, \boldsymbol{\varphi})\|^2 = O(t^{-1-2\theta}T^{2\theta}), \tag{266}$$

noting that in this region $d_2 \leq \min\{1/2 - \theta, d_1 - \kappa\} < 1/2$. The proof of (266) is nearly identical to that of (259). We normalize by T^{2d_1-1} in both the numerator and denominator of (260), and the denominator is then $O(1)$ by application of (165) and (170) of Lemma 14. The (normalized) numerator of (260) is bounded from below by

$$\inf_{\substack{1/2+\theta \leq d_1 \leq \bar{d}_1, \underline{d}_2 \leq d_2 \leq 1/2-\theta \\ \boldsymbol{\varphi} \in \Psi, |d_1-d_2| \geq \kappa}} T^{2d_1-1} \sum_{k=1}^{T} \left(c_{k-1}(d_1, \boldsymbol{\varphi}) - c_{k-1}(d_2, \boldsymbol{\varphi})\right)^2 > \epsilon$$

by application of (162) and (164) of Lemma 14.

Next, we show that

$$\sup_{\substack{1/2+\theta \leq d_1 \leq \bar{d}_1, 1/2-\theta \leq d_2 \leq 1/2+\theta \\ \boldsymbol{\varphi} \in \Psi, |d_1-d_2| \geq \kappa}} \|\boldsymbol{h}_{t-1,T}(d_1, d_2, \boldsymbol{\varphi})\|^2 = O(t^{-1-2\theta}T^{2\theta}). \tag{267}$$

In view of (257), (267) follows by showing that

$$\sup_{\substack{1/2+\theta \leq d_1 \leq \bar{d}_1 \\ 1/2-\theta \leq d_2 \leq 1/2+\theta \\ \boldsymbol{\varphi} \in \Psi, |d_1-d_2| \geq \kappa}} \frac{T^{2d_1-1} c_{t-1}^2(d_1, \boldsymbol{\varphi}) + h_{t-1,T}^2(d_2, \boldsymbol{\varphi}) T^{2d_1-1} \sum_{j=1}^{T} c_{j-1}^2(d_1, \boldsymbol{\varphi})}{T^{2d_1-1} \left(\sum_{j=1}^{T} c_{j-1}^2(d_1, \boldsymbol{\varphi}) - \frac{\left(\sum_{j=1}^{T} c_{j-1}(d_1, \boldsymbol{\varphi}) c_{j-1}(d_2, \boldsymbol{\varphi})\right)^2}{\sum_{j=1}^{T} c_{j-1}^2(d_2, \boldsymbol{\varphi})}\right)} = O(t^{-1-2\theta}T^{2\theta}),$$

and this holds by (170), (172), and (253), noting also that by (160),

$$\sup_{1/2+\theta \leq d_1 \leq \bar{d}_1, \boldsymbol{\varphi} \in \Psi} T^{2d_1-1} c_{t-1}^2(d_1, \boldsymbol{\varphi}) = O(t^{-1+2\theta}T^{-2\theta}) = O(t^{-1-2\theta}T^{2\theta})$$

because $t \leq T$.

Finally, we show that

$$\sup_{\substack{1/2+\theta \leq d_1 \leq \bar{d}_1, 1/2+\theta \leq d_2 \leq \bar{d}_2 \\ \varphi \in \Psi, |d_1-d_2| \geq \kappa}} \|\boldsymbol{h}_{t-1,T}(d_1, d_2, \varphi)\|^2 = O(t^{-1-2\theta}T^{2\theta}). \tag{268}$$

In view of (258), (268) holds by (253) and

$$\inf_{\substack{1/2+\theta \leq d_1 \leq \bar{d}_1, 1/2+\theta \leq d_2 \leq \bar{d}_2 \\ \varphi \in \Psi, |d_1-d_2| \geq \kappa}} \sum_{k=1}^{T-1} \sum_{j \neq t=1}^{T} (h_{k-1,T}(d_1, \varphi)h_{j-1,T}(d_2, \varphi) - h_{k-1,T}(d_2, \varphi)h_{j-1,T}(d_1, \varphi))^2 > \epsilon. \tag{269}$$

By very similar steps to those in the proof of Lemma 13, it can be shown that

$$\sum_{k=1}^{T-1} \sum_{j \neq t=1}^{T} (h_{k-1,T}(d_1, \varphi)h_{j-1,T}(d_2, \varphi) - h_{k-1,T}(d_2, \varphi)h_{j-1,T}(d_1, \varphi))^2$$

$$= \frac{\sum_{k=1}^{T-1} \sum_{j \neq t=1}^{T} \left(T^{1/2-d_1}\pi_{k-1}(d_1)T^{1/2-d_2}\pi_{j-1}(d_2) - T^{1/2-d_2}\pi_{k-1}(d_2)T^{1/2-d_1}\pi_{j-1}(d_1)\right)^2}{T^{1-2d_1}\sum_{j=1}^{T}\pi_{j-1}^2(d_1)T^{1-2d_2}\sum_{k=1}^{T}\pi_{k-1}^2(d_2)}$$

$$+ g(d_1, d_2, \varphi),$$

where $\sup_{1/2+\theta \leq d_1 \leq \bar{d}_1, 1/2+\theta \leq d_2 \leq \bar{d}_2, \varphi \in \Psi} |g(d_1, d_2, \varphi)| = o(1)$. Next, using this result together with Lemma 11 and approximating sums by integrals, see Lemma 10, the left-hand side of (269) is bounded from below by

$$\epsilon \frac{\inf_{1/2+\theta \leq d_1 \leq \bar{d}_1, 1/2+\theta \leq d_2 \leq \bar{d}_2} \left(\int_0^1 \int_0^1 (x^{d_1-1}y^{d_2-1} - x^{d_2-1}y^{d_1-1})^2 dxdy\right)}{\sup_{1/2+\theta \leq d_1 \leq \bar{d}_1, 1/2+\theta \leq d_2 \leq \bar{d}_2}(2d_1-1)^{-1}(2d_2-1)^{-1}} - o(1) > \epsilon,$$

so (268) is justified, to conclude the proof of (176).

Next, (177) follows trivially from (176), while the proof of (178) is omitted because it is basically identical to that of (176). We only need to note that instead of (253) we have

$$\sup_{d_1 \in [\underline{d}_1, \bar{d}_1], \varphi \in \Psi} |h_{t,T}(d_1, \varphi) - h_{t-1,T}(d_1, \varphi)| = O(t^{-3/2-\theta}T^{\theta}),$$

which can be justified exactly like (253) because

$$h_{t,T}(d, \varphi) - h_{t-1,T}(d, \varphi) = \frac{c_t(d-1, \varphi)}{\left(\sum_{j=1}^{T} c_{j-1}^2(d, \varphi)\right)^{1/2}}. \tag{270}$$

We next show (179), for which we first present a preliminary result. By the same proof as that of (253), taking into account the extra log-term arising from (159) and (160) of Lemma 12, it holds that

$$\sup_{d_1 \in [\underline{d}_1, \bar{d}_1], \varphi \in \Psi} \left| \frac{\frac{\partial}{\partial d_1} c_{t-1}(d_1, \varphi)}{\left(\sum_{j=1}^{T} c_{j-1}^2(d_1, \varphi)\right)^{1/2}} \right| = O(t^{-1/2-\theta}T^{\theta} \log T). \tag{271}$$

Clearly,

$$
\boldsymbol{h}'_{t-1,T}(d_1, d_2, \boldsymbol{\varphi})\boldsymbol{h}_{k-1,T}(d_1, d_2, \boldsymbol{\varphi})
$$

$$
= \boldsymbol{c}'_{t-1}(d_1, d_2, \boldsymbol{\varphi}) \left(\sum_{j=1}^{T} \boldsymbol{c}_{j-1}(d_1, d_2, \boldsymbol{\varphi})\boldsymbol{c}'_{j-1}(d_1, d_2, \boldsymbol{\varphi}) \right)^{-1} \boldsymbol{c}_{k-1}(d_1, d_2, \boldsymbol{\varphi}),
$$

so that

$$
\frac{\partial}{\partial d_2} \left(\boldsymbol{h}'_{t-1,T}(d_1, d_2, \boldsymbol{\varphi})\boldsymbol{h}_{k-1,T}(d_1, d_2, \boldsymbol{\varphi}) \right) = A_{1tk} + A_{2tk} + A_{3tk}, \tag{272}
$$

where

$$
A_{1tk} = \left(\frac{\partial}{\partial d_2} \boldsymbol{c}'_{t-1}(d_1, d_2, \boldsymbol{\varphi}) \right) \left(\sum_{j=1}^{T} \boldsymbol{c}_{j-1}(d_1, d_2, \boldsymbol{\varphi})\boldsymbol{c}'_{j-1}(d_1, d_2, \boldsymbol{\varphi}) \right)^{-1} \boldsymbol{c}_{k-1}(d_1, d_2, \boldsymbol{\varphi}),
$$

$$
A_{2tk} = \boldsymbol{c}'_{t-1}(d_1, d_2, \boldsymbol{\varphi}) \left(\sum_{j=1}^{T} \boldsymbol{c}_{j-1}(d_1, d_2, \boldsymbol{\varphi})\boldsymbol{c}'_{j-1}(d_1, d_2, \boldsymbol{\varphi}) \right)^{-1} \left(\frac{\partial}{\partial d_2} \boldsymbol{c}_{k-1}(d_1, d_2, \boldsymbol{\varphi}) \right),
$$

$$
A_{3tk} = -\boldsymbol{c}'_{t-1}(d_1, d_2, \boldsymbol{\varphi}) \left(\sum_{j=1}^{T} \boldsymbol{c}_{j-1}(d_1, d_2, \boldsymbol{\varphi})\boldsymbol{c}'_{j-1}(d_1, d_2, \boldsymbol{\varphi}) \right)^{-1}
$$

$$
\times \left(\frac{\partial}{\partial d_2} \sum_{j=1}^{T} \boldsymbol{c}_{j-1}(d_1, d_2, \boldsymbol{\varphi})\boldsymbol{c}'_{j-1}(d_1, d_2, \boldsymbol{\varphi}) \right) \left(\sum_{j=1}^{T} \boldsymbol{c}_{j-1}(d_1, d_2, \boldsymbol{\varphi})\boldsymbol{c}'_{j-1}(d_1, d_2, \boldsymbol{\varphi}) \right)^{-1}
$$

$$
\times \boldsymbol{c}_{k-1}(d_1, d_2, \boldsymbol{\varphi}).
$$

Clearly, recalling the definition of D_T in (256),

$$
A_{1tk} = \left(\frac{\partial c_{t-1}(d_2, \boldsymbol{\varphi})}{\partial d_2} \right) \frac{c_{k-1}(d_2, \boldsymbol{\varphi}) \sum_{j=1}^{T} c_{j-1}^2(d_1, \boldsymbol{\varphi})}{D_T}
$$

$$
- \left(\frac{\partial c_{t-1}(d_2, \boldsymbol{\varphi})}{\partial d_2} \right) \frac{c_{k-1}(d_1, \boldsymbol{\varphi}) \sum_{j=1}^{T} c_{j-1}(d_1, \boldsymbol{\varphi})c_{j-1}(d_2, \boldsymbol{\varphi})}{D_T}.
$$

Then, proceeding as in the proof of (176), it can be straightforwardly shown that

$$
\sup_{d_1 \in [\underline{d}_1, \overline{d}_1], d_2 \in [\underline{d}_2, \overline{d}_2], \boldsymbol{\varphi} \in \Psi} |A_{1tk}| = O(t^{-1/2-\theta}k^{-1/2-\theta}T^{3\theta}), \tag{273}
$$

where the extra T^θ-factor (compared with (176)) captures a $(\log T)$-factor due to having $\partial c_{t-1}(d_2, \boldsymbol{\varphi})/\partial d_2$ instead of $c_{t-1}(d_2, \boldsymbol{\varphi})$; see Lemma 12. Similarly, noting the symmetry of A_{1tk} and A_{2tk},

$$
\sup_{d_1 \in [\underline{d}_1, \overline{d}_1], d_2 \in [\underline{d}_2, \overline{d}_2], \boldsymbol{\varphi} \in \Psi} |A_{2tk}| = O(t^{-1/2-\theta}k^{-1/2-\theta}T^{3\theta}). \tag{274}
$$

Finally,

$$
\begin{aligned}
A_{3tk} = & -\sum_{l=1}^{T} \boldsymbol{c}'_{t-1}(d_1, d_2, \boldsymbol{\varphi}) \left(\sum_{j=1}^{T} \boldsymbol{c}_{j-1}(d_1, d_2, \boldsymbol{\varphi}) \boldsymbol{c}'_{j-1}(d_1, d_2, \boldsymbol{\varphi}) \right)^{-1} \begin{pmatrix} 0 \\ \frac{\partial c_{l-1}(d_2, \boldsymbol{\varphi})}{\partial d_2} \end{pmatrix} \\
& \times \boldsymbol{c}'_{l-1}(d_1, d_2, \boldsymbol{\varphi}) \left(\sum_{j=1}^{T} \boldsymbol{c}_{j-1}(d_1, d_2, \boldsymbol{\varphi}) \boldsymbol{c}'_{j-1}(d_1, d_2, \boldsymbol{\varphi}) \right)^{-1} \boldsymbol{c}_{k-1}(d_1, d_2, \boldsymbol{\varphi}) \\
& -\sum_{l=1}^{T} \boldsymbol{c}'_{t-1}(d_1, d_2, \boldsymbol{\varphi}) \left(\sum_{j=1}^{T} \boldsymbol{c}_{j-1}(d_1, d_2, \boldsymbol{\varphi}) \boldsymbol{c}'_{j-1}(d_1, d_2, \boldsymbol{\varphi}) \right)^{-1} \boldsymbol{c}_{l-1}(d_1, d_2, \boldsymbol{\varphi}) \\
& \times \begin{pmatrix} 0 & \frac{\partial c_{l-1}(d_2, \boldsymbol{\varphi})}{\partial d_2} \end{pmatrix} \left(\sum_{j=1}^{T} \boldsymbol{c}_{j-1}(d_1, d_2, \boldsymbol{\varphi}) \boldsymbol{c}'_{j-1}(d_1, d_2, \boldsymbol{\varphi}) \right)^{-1} \boldsymbol{c}_{k-1}(d_1, d_2, \boldsymbol{\varphi}).
\end{aligned}
$$

By (176) it follows straightforwardly that

$$
\begin{aligned}
& \sup_{d_1 \in [\underline{d}_1, \overline{d}_1], d_2 \in [\underline{d}_2, \overline{d}_2], \boldsymbol{\varphi} \in \Psi} \left| \boldsymbol{c}'_{t-1}(d_1, d_2, \boldsymbol{\varphi}) \left(\sum_{j=1}^{T} \boldsymbol{c}_{j-1}(d_1, d_2, \boldsymbol{\varphi}) \boldsymbol{c}'_{j-1}(d_1, d_2, \boldsymbol{\varphi}) \right)^{-1} \boldsymbol{c}_{l-1}(d_1, d_2, \boldsymbol{\varphi}) \right| \\
& = \sup_{d_1 \in [\underline{d}_1, \overline{d}_1], d_2 \in [\underline{d}_2, \overline{d}_2], \boldsymbol{\varphi} \in \Psi} \left| \boldsymbol{h}'_{t-1, T}(d_1, d_2, \boldsymbol{\varphi}) \boldsymbol{h}_{l-1, T}(d_1, d_2, \boldsymbol{\varphi}) \right| \\
& \leq \sup_{d_1 \in [\underline{d}_1, \overline{d}_1], d_2 \in [\underline{d}_2, \overline{d}_2], \boldsymbol{\varphi} \in \Psi} \| \boldsymbol{h}_{t-1, T}(d_1, d_2, \boldsymbol{\varphi}) \| \sup_{d_1 \in [\underline{d}_1, \overline{d}_1], d_2 \in [\underline{d}_2, \overline{d}_2], \boldsymbol{\varphi} \in \Psi} \| \boldsymbol{h}_{l-1, T}(d_1, d_2, \boldsymbol{\varphi}) \| \\
& = O(t^{-1/2-\theta} l^{-1/2-\theta} T^{2\theta}), \qquad (275)
\end{aligned}
$$

so by (273)–(275),

$$
\sup_{d_1 \in [\underline{d}_1, \overline{d}_1], d_2 \in [\underline{d}_2, \overline{d}_2], \boldsymbol{\varphi} \in \Psi} |A_{3tk}| = O(t^{-1/2-\theta} k^{-1/2-\theta} T^{5\theta}),
$$

to conclude the proof of (179).

Next, the proof of (180) is omitted as it is almost identical to that of (179) in view of (178).
Before proving (181), we first note that, for $m = 0, 1$,

$$
\sup_{d \geq 1/2+\theta, \boldsymbol{\varphi} \in \Psi} \left| \frac{\frac{\partial^m}{\partial d^m} \left(T^{-d} c_{t-1}(d, \boldsymbol{\varphi}) \right)}{\left(\sum_{j=1}^{T} T^{-2d} c_{j-1}^2(d, \boldsymbol{\varphi}) \right)^{1/2}} \right| \leq \frac{\sup_{d \geq 1/2+\theta, \boldsymbol{\varphi} \in \Psi} \left| \frac{\partial^m}{\partial d^m} \left(T^{-d} c_{t-1}(d, \boldsymbol{\varphi}) \right) \right|}{\left(\inf_{d \geq 1/2+\theta, \boldsymbol{\varphi} \in \Psi} \sum_{j=1}^{T} T^{-2d} c_{j-1}^2(d, \boldsymbol{\varphi}) \right)^{1/2}}
$$
$$
= O \left(t^{-1/2} (t/T)^{\theta} (1 + |\log(t/T)|)^m \right), \qquad (276)
$$

which is a better bound than (253) and (271), but on a smaller set. The bound (276) holds by (160) of Lemma 12 and (164) of Lemma 14.

To show (181) we define $\Upsilon_T = \mathrm{diag}(1, T^{-d_2})$ and note from (272) that

$$A_{1tk} = \frac{\partial}{\partial d_2}(\boldsymbol{c}'_{t-1}(d_1, d_2, \boldsymbol{\varphi})\Upsilon_T)\left(\Upsilon_T \sum_{j=1}^{T} \boldsymbol{c}_{j-1}(d_1, d_2, \boldsymbol{\varphi})\boldsymbol{c}'_{j-1}(d_1, d_2, \boldsymbol{\varphi})\Upsilon_T\right)^{-1}\Upsilon_T \boldsymbol{c}_{k-1}(d_1, d_2, \boldsymbol{\varphi})$$

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

$$(277)$$

$$A_{2tk} = \boldsymbol{c}'_{t-1}(d_1, d_2, \boldsymbol{\varphi})\Upsilon_T\left(\Upsilon_T \sum_{j=1}^{T} \boldsymbol{c}_{j-1}(d_1, d_2, \boldsymbol{\varphi})\boldsymbol{c}'_{j-1}(d_1, d_2, \boldsymbol{\varphi})\Upsilon_T\right)^{-1}\frac{\partial}{\partial d_2}(\Upsilon_T \boldsymbol{c}_{k-1}(d_1, d_2, \boldsymbol{\varphi})),$$

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

$$(278)$$

and

$$A_{3tk} = -\boldsymbol{c}'_{t-1}(d_1, d_2, \boldsymbol{\varphi})\Upsilon_T\left(\Upsilon_T \sum_{j=1}^{T} \boldsymbol{c}_{j-1}(d_1, d_2, \boldsymbol{\varphi})\boldsymbol{c}'_{j-1}(d_1, d_2, \boldsymbol{\varphi})\Upsilon_T\right)^{-1}$$

$$\times \sum_{l=1}^{T} \frac{\partial}{\partial d_2}(\Upsilon_T \boldsymbol{c}_{l-1}(d_1, d_2, \boldsymbol{\varphi}))\boldsymbol{c}'_{l-1}(d_1, d_2, \boldsymbol{\varphi})\Upsilon_T$$

$$\times \left(\Upsilon_T \sum_{j=1}^{T} \boldsymbol{c}_{j-1}(d_1, d_2, \boldsymbol{\varphi})\boldsymbol{c}'_{j-1}(d_1, d_2, \boldsymbol{\varphi})\Upsilon_T\right)^{-1}\Upsilon_T \boldsymbol{c}_{k-1}(d_1, d_2, \boldsymbol{\varphi})$$

$$- \boldsymbol{c}'_{t-1}(d_1, d_2, \boldsymbol{\varphi})\Upsilon_T\left(\Upsilon_T \sum_{j=1}^{T} \boldsymbol{c}_{j-1}(d_1, d_2, \boldsymbol{\varphi})\boldsymbol{c}'_{j-1}(d_1, d_2, \boldsymbol{\varphi})\Upsilon_T\right)^{-1}\sum_{l=1}^{T}\Upsilon_T \boldsymbol{c}_{l-1}(d_1, d_2, \boldsymbol{\varphi})$$

$$\times \frac{\partial}{\partial d_2}(\boldsymbol{c}'_{l-1}(d_1, d_2, \boldsymbol{\varphi})\Upsilon_T)\left(\Upsilon_T \sum_{j=1}^{T} \boldsymbol{c}_{j-1}(d_1, d_2, \boldsymbol{\varphi})\boldsymbol{c}'_{j-1}(d_1, d_2, \boldsymbol{\varphi})\Upsilon_T\right)^{-1}\Upsilon_T \boldsymbol{c}_{k-1}(d_1, d_2, \boldsymbol{\varphi}).$$

For A_{1tk}, we multiply by $T^{1-2d_2}\sum_{j=1}^{T} c_{j-1}^2(d_2, \boldsymbol{\varphi})$ in the numerator and denominator. Application of (172) then shows that the resulting denominator is bounded from below by $\epsilon > 0$, while (170) shows that the order of magnitude of the numerator is unaffected. For the numerator of A_{1tk} we apply the Cauchy-Schwarz inequality, $\left|\sum_{j=1}^{T} h_{j-1,T}(d_1, \boldsymbol{\varphi})h_{j-1,T}(d_2, \boldsymbol{\varphi})\right| \leq 1$, which together with (253), with α instead of θ, and (276) shows that

$$\sup_{\substack{d_2 \in [1/2+\theta, \bar{d}_2] \\ \boldsymbol{\varphi} \in \Psi, |d_1-d_2| \geq \kappa}} |A_{1tk}| = O\left(t^{-1/2+\theta}T^{-\theta}(1 + |\log(t/T)|)k^{-1/2-\alpha}T^{\alpha}\right). \tag{279}$$

Similarly, noting the symmetry of A_{1tk} and A_{2tk},

$$\sup_{\substack{d_2 \in [1/2+\theta, \bar{d}_2] \\ \boldsymbol{\varphi} \in \Psi, |d_1-d_2| \geq \kappa}} |A_{2tk}| = O\left(k^{-1/2+\theta}T^{-\theta}(1 + |\log(k/T)|)t^{-1/2-\alpha}T^{\alpha}\right). \tag{280}$$

Next, by (176) with α instead of θ, it holds that

$$\sup_{\substack{d_1\in[\underline{d}_1,\bar{d}_1],d_2\in[\underline{d}_2,\bar{d}_2]\\ \varphi\in\Psi,|d_1-d_2|\geq\kappa}}\left|\boldsymbol{h}'_{j-1,T}(d_1,d_2,\boldsymbol{\varphi})\boldsymbol{h}_{k-1,T}(d_1,d_2,\boldsymbol{\varphi})\right|=O(j^{-1/2-\alpha}k^{-1/2-\alpha}T^{2\alpha}). \tag{281}$$

It then follows straightforwardly by (279), (280), and (281) that

$$\sup_{\substack{d_2\in[1/2+\theta,\bar{d}_2]\\ \varphi\in\Psi,|d_1-d_2|\geq\kappa}}|A_{3tk}|=O\left(t^{-1/2-\alpha}k^{-1/2-\alpha}T^{3\alpha-\theta}\sum_{l=1}^{T}l^{-1+\theta-\alpha}(1+|\log(l/T)|)\right)$$

$$=O(t^{-1/2-\alpha}k^{-1/2-\alpha}T^{2\alpha})$$

because $\theta>\alpha$. Then (181) holds because, for any $s\leq T$, we have $|\log(s/T)|\leq K(T/s)^{\alpha}$ and

$$s^{-1/2+\theta}T^{-\theta}(1+|\log(s/T)|)=s^{-1/2-\alpha}T^{\alpha}(s/T)^{\alpha+\theta}(1+|\log(s/T)|)$$

$$\leq Ks^{-1/2-\alpha}T^{\alpha}(s/T)^{\alpha+\theta}(T/s)^{\alpha}\leq Ks^{-1/2-\alpha}T^{\alpha}.$$

Finally, the proof of (182) is omitted because it follows by identical arguments to that of (181) since it can be straightforwardly shown that

$$\sup_{d\geq1/2+\theta,\varphi\in\Psi}\left|\frac{\partial}{\partial d}(h_{t,T}(d,\boldsymbol{\varphi})-h_{t-1,T}(d,\boldsymbol{\varphi}))\right|=O(t^{-3/2+\theta}T^{-\theta}(1+|\log(t/T)|)),$$

see (253) and (270).

10.9 Proof of Lemma 17

The proof of Lemma 17 is given in Lemma S.17 of Hualde and Nielsen (2020).

10.10 Proof of Lemma 18

By summation by parts we find

$$\left|\sum_{t=1}^{T}\rho(L;\boldsymbol{\varphi})u_t(\delta-\delta_0)\boldsymbol{h}_{t-1,T}(1-\delta,\gamma-\delta,\boldsymbol{\varphi})\right|$$

$$\leq|\boldsymbol{h}_{T-1,T}(1-\delta,\gamma-\delta,\boldsymbol{\varphi})|\,|\rho(L;\boldsymbol{\varphi})u_T(\delta-\delta_0-1)|$$

$$+\sum_{t=1}^{T-1}|\boldsymbol{h}_{t,T}(1-\delta,\gamma-\delta,\boldsymbol{\varphi})-\boldsymbol{h}_{t-1,T}(1-\delta,\gamma-\delta,\boldsymbol{\varphi})|\,|\rho(L;\boldsymbol{\varphi})u_t(\delta-\delta_0-1)|.$$

Then (185) follows straightforwardly by (176), (178) of Lemma 16 and (183) of Lemma 17.
Next, again by summation by parts,

$$\left|\sum_{t=1}^{T}\rho(L;\boldsymbol{\varphi})u_t(\delta-\delta_0)\frac{\partial}{\partial\gamma}(\boldsymbol{h}'_{t-1,T}(1-\delta,\gamma-\delta,\boldsymbol{\varphi})\boldsymbol{h}_{k-1,T}(1-\delta,\gamma-\delta,\boldsymbol{\varphi}))\right|$$

$$\leq\left|\frac{\partial}{\partial\gamma}(\boldsymbol{h}'_{T-1,T}(1-\delta,\gamma-\delta,\boldsymbol{\varphi})\boldsymbol{h}_{k-1,T}(1-\delta,\gamma-\delta,\boldsymbol{\varphi}))\right|\,|\rho(L;\boldsymbol{\varphi})u_T(\delta-\delta_0-1)|$$

$$+\sum_{t=1}^{T-1}\left|\frac{\partial}{\partial\gamma}((\boldsymbol{h}'_{t,T}(1-\delta,\gamma-\delta,\boldsymbol{\varphi})-\boldsymbol{h}'_{t-1,T}(1-\delta,\gamma-\delta,\boldsymbol{\varphi}))\boldsymbol{h}_{k-1,T}(1-\delta,\gamma-\delta,\boldsymbol{\varphi}))\right|$$

$$\times|\rho(L;\boldsymbol{\varphi})u_t(\delta-\delta_0-1)|, \tag{282}$$

so (186) follows directly from (179), (180) of Lemma 16 and (183) of Lemma 17.

The proofs of (187) and (188) are nearly identical to those of (185) and (186), but using (184) instead of (183) of Lemma 17.

Next, by application of (181), (182), and (183) to (282), it is straightforward to show that, choosing any α such that $0 < \alpha < \theta$,

$$
\sup_{\delta_0-\delta\leq g,\gamma-\delta\geq 1/2+\theta}\left|\sum_{t=1}^{T}\rho(L;\boldsymbol{\varphi})u_t(\delta-\delta_0)\frac{\partial}{\partial\gamma}(\boldsymbol{h}'_{t-1,T}(1-\delta,\gamma-\delta,\boldsymbol{\varphi})\boldsymbol{h}_{k-1,T}(1-\delta,\gamma-\delta,\boldsymbol{\varphi}))\right|
$$
$$
=O_p(k^{-1/2-\alpha}T^{\alpha-1/2}(T^{g+1/2}+(\log T)\mathbb{I}(g\leq -1/2)))
$$
$$
+O_p(k^{-1/2-\alpha}T^{2\alpha}(T^{g-\alpha}\mathbb{I}(g>\alpha)+(\log T)\mathbb{I}(g\leq\alpha))),
$$

which proves (189).

Finally, the proof of (190) is almost identical to that of (189) but using (184) instead of (183).

10.11 Proof of Lemma 19

The results follow by direct application of (159), (160) of Lemma 12 and (176) of Lemma 16.

10.12 Proof of Lemma 20

By summation by parts we find

$$
\left|\sum_{t=1}^{T}\rho(L;\boldsymbol{\varphi})u_t(\delta-\delta_0)c_{t-1}(\gamma_0-\delta,\boldsymbol{\varphi})\right|
$$
$$
\leq \|c_{T-1}(\gamma_0-\delta,\boldsymbol{\varphi})\|\,|\rho(L;\boldsymbol{\varphi})u_T(\delta-\delta_0-1)|
$$
$$
+\left\|\sum_{t=1}^{T-1}c_{t-1}(\gamma_0-\delta-1,\boldsymbol{\varphi})\rho(L;\boldsymbol{\varphi})u_t(\delta-\delta_0-1)\right\|. \tag{283}
$$

The result (193) then follows by application of (159) of Lemma 12 and (183) of Lemma 17, while the result (194) follows by application of (160) of Lemma 12 and (184) of Lemma 17.

10.13 Proof of Lemma 21

These results are direct consequences of (159) and (160) of Lemma 12 and (184) of Lemma 17.

10.14 Proof of Lemma 22

The proof of Lemma 22 is given in Lemma S.12 of Hualde and Nielsen (2020); see also Lemma 2 of Johansen and Nielsen (2019).

References

1. Abadir, K.M., Distaso, W. and Giraitis, L. (2007). Nonstationarity-extended local Whittle estimation. *Journal of Econometrics* **141**, 1353–1384.

2. Abramowitz, M. and Stegun, I.A. (1970). *Handbook of Mathematical Functions*, National Bureau of Standards, Washington, D.C.

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. Cavaliere, G., Nielsen, M.Ø. and Taylor, A.M.R. (2015). Bootstrap score tests for fractional integration in heteroskedastic ARFIMA models, with an application to price dynamics in commodity spot and futures markets. *Journal of Econometrics* **187**, 557–579.

6. Cavaliere, G., Nielsen, M.Ø. and Taylor, A.M.R. (2017). Quasi-maximum likelihood estimation and bootstrap inference in fractional time series models with heteroskedasticity of unknown form. *Journal of Econometrics* **198**, 165–188.

7. Cavaliere, G., Nielsen, M.Ø. and Taylor, A.M.R. (2022). Adaptive inference in heteroscedastic fractional time series models. *Journal of Business and Economic Statistics* **40**, 50–65.

8. Chen, W.W. and Hurvich, C.M. (2003). Estimating fractional cointegration in the presence of polynomial trends. *Journal of Econometrics* **117**, 95–121.

9. Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *Annals of Statistics* **17**, 1749–1766.

10. 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.

11. 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.

12. Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and Its Application*, Academic Press, New York.

13. Hannan, E.J. (1973). The asymptotic theory of linear time series models. *Journal of Applied Probability* **10**, 130–145.

14. Hansen, B.E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413-430.

15. 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.

16. Hualde, J. and Nielsen, M.Ø. (2020). Truncated sum of squares estimation of fractional time series models with deterministic trends. *Econometric Theory* **36**, 393–412.

17. Hualde, J. and Robinson, P.M. (2011). Gaussian pseudo-maximum likelihood estimation of fractional time series models. *Annals of Statistics* **39**, 3152–3181.

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

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

20. Johansen, S. and Nielsen, M.Ø. (2012a). Likelihood inference for a fractionally cointegrated vector autoregressive model. *Econometrica* **80**, 2667–2732.

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

22. 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.

23. Johansen, S. and Nielsen, M.Ø. (2019). Nonstationary cointegration in the fractionally cointegrated VAR model. *Journal of Time Series Analysis* **40**, 519–543.

24. Li, W.K. and McLeod, A.I. (1986). Fractional time series modelling. *Biometrika* **73**, 217–221.

25. 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.

26. Phillips, P.C.B. (2007). Regression with slowly varying regressors and nonlinear trends. *Econometric Theory* **23**, 557–614.

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

28. Robinson, P.M. (2005). Efficiency improvements in inference on stationary and nonstationary fractional time series. *Annals of Statistics* **33**, 1800–1842.

29. Robinson, P.M. (2012). Inference on power law spatial trends. *Bernoulli* **18**, 644–677.

30. Robinson, P.M. and Hualde, J. (2003). Cointegration in fractional systems with unknown integration orders. *Econometrica* **71**, 1727–1766.

31. Robinson, P.M. and Iacone, F. (2005). Cointegration in fractional systems with deterministic trends. *Journal of Econometrics* **129**, 263–298.

32. Robinson, P.M. and Marinucci, D. (2000). The averaged periodogram for nonstationary vector time series. *Statistical Inference for Stochastic Processes* **3**, 149–160.

33. Velasco, C. (1999a). Non-stationary log-periodogram regression. *Journal of Econometrics* **91**, 325–371.

34. Velasco, C. (1999b). Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis* **20**, 87–127.

35. White, H. and Granger, C.W.J. (2011). Consideration of trends in time series. *Journal of Time Series Econometrics* **3**, issue 1, article 2.

36. Wu, C.-F. (1981). Asymptotic theory of nonlinear least squares estimation. *Annals of Statistics* **9**, 501–513.

Research Papers

2021-07:   Matei Demetrescu and Robinson Kruse-Becher: Is U.S. real output growth really non-normal? Testing distributional assumptions in time-varying location-scale models

2021-08:   Luisa Corrado, Stefano Grassi and Aldo Paolillo: Modelling and Estimating Large Macroeconomic Shocks During the Pandemic

2021-09:   Leopoldo Catania, Alessandra Luati and Pierluigi Vallarino: Economic vulnerability is state dependent

2021-10:   Søren Johansen and Anders Ryghn Swensen: Adjustment coefficients and exact rational expectations in cointegrated vector autoregressive models

2021-11:   Bent Jesper Christensen, Mads Markvart Kjær and Bezirgen Veliyev: The incremental information in the yield curve about future interest rate risk

2021-12:   Mikkel Bennedsen, Asger Lunde, Neil Shephard and Almut E. D. Veraart: Inference and forecasting for continuous-time integer-valued trawl processes and their use in financial economics

2021-13:   Anthony D. Hall, Annastiina Silvennoinen and Timo Teräsvirta: Four Australian Banks and the Multivariate Time-Varying Smooth Transition Correlation GARCH model

2021-14:   Ulrich Hounyo and Kajal Lahiri: Estimating the Variance of a Combined Forecast: Bootstrap-Based Approach

2021-15:   Salman Huseynov: Long and short memory in dynamic term structure models

2022-01:   Jian Kang, Johan Stax Jakobsen, Annastiina Silvennoinen, Timo Teräsvirta and Glen Wade: A parsimonious test of constancy of a positive definite correlation matrix in a multivariate time-varying GARCH model

2022-02:   Javier Haulde and Morten Ørregaard Nielsen: Fractional integration and cointegration

2022-03:   Yue Xu: Spillovers of Senior Mutual Fund Managers' Capital Raising Ability

2022-04:   Morten Ørregaard Nielsen, Wonk-ki Seo and Dakyung Seong: Inference on the dimension of the nonstationary subspace in functional time series

2022-05:   Kristoffer Pons Bertelsen: The Prior Adaptive Group Lasso and the Factor Zoo

2022-06:   Ole Linnemann Nielsen and Anders Merrild Posselt: Betting on mean reversion in the VIX? Evidence from ETP flows

2022-07:   Javier Hualde and Morten Ørregaard Nielsen: Truncated sum-of-squares estimation of fractional time series models with generalized power law trend