The modified conditional sum-of-squares estimator for fractionally integrated models*

Mustafa R. Kılınç[1] and Michael Massmann[†,1,2]

[1]WHU – Otto Beisheim School of Management,
Chair of Econometrics and Statistics, Vallendar, Germany

[2]Vrije Universiteit, Department of Econometrics and Data Science,
Amsterdam, The Netherlands

22nd April 2024

Abstract

In this paper, we analyse the influence of estimating a constant term on the bias of the conditional sum-of-squares (CSS) estimator in a stationary or non-stationary type-II ARFIMA (p_1, d, p_2) model. We derive expressions for the estimator's bias and show that the leading term can be easily removed by a simple modification of the CSS objective function. We call this new estimator the modified conditional sum-of-squares (MCSS) estimator. We show theoretically and by means of Monte Carlo simulations that its performance relative to that of the CSS estimator is markedly improved even for small sample sizes. Finally, we revisit three classical short datasets that have in the past been described by ARFIMA(p_1,d,p_2) models with constant term, namely the post-second World War real GNP data, the extended Nelson-Plosser data, and the Nile data.

Keywords: long memory, fractional integration, conditional sum-of-squares estimator, asymptotic expansion, small sample bias.

JEL Codes: C22.

*We are grateful to Jörg Breitung, Alexander Mayer, Philipp Sibbertsen and Dominik Wied for illuminating discussions and helpful suggestions. Valuable comments were also made by participants at the econometrics research seminar at Cologne University, at the European Meeting of the Econometric Society in Milan and at Statistische Woche in Dortmund. The financial support of Deutsche Forschungsgemeinschaft (project number 258395632) is gratefully acknowledged.

†corresponding author: `michael.massmann@whu.edu`

1 Introduction

Fractionally integrated autoregressive moving average (ARFIMA) models are applied in a wide range of fields for describing long-memory phenomena, witness inter alia the economic and political as well as the natural sciences; see Hassler (2019) and Hualde & Nielsen (2023) for general treatments. One particular variant of this model class that has recently gained popularity is the so-called type-II ARFIMA model, which sets the initial observations equal to zero and allows both stationary and non-stationary processes to be described, see for example Nielsen (2004), Robinson (2005) and Johansen (2008). A popular choice for estimating this model is the conditional sum-of-squares (CSS) estimator whose main appealing features are that it is computationally straightforward and that the memory parameter can be estimated consistently as long as it lies in an arbitrary compact interval on the real line. It was introduced by Li & McLeod (1986) in the context of stationary fractionally integrated models. Subsequent papers allowed for non-stationary models, see for instance Beran (1995) and Velasco & Robinson (2000). Local consistency proofs were provided by Tanaka (1999), Nielsen (2004) and Robinson (2006). Global consistency was proved by Hualde & Robinson (2011) and Nielsen (2015) in a model without deterministic components. Only recently, Hualde & Nielsen (2020, 2022) derived global consistency and the asymptotic normality of the CSS estimator in a model with deterministic components, such as a constant or a trending term. Empirical applications include Hualde & Robinson (2011) for aggregate income and consumption data and Johansen & Nielsen (2016) for opinion poll data.

While the literature dealing with asymptotic inferences in the context of parametric ARFIMA models is well-developed, some issues still require attention. One such concern pertains to the small sample performance of the CSS estimator. Despite the widespread use of the CSS estimator little is currently known about the impact deterministic terms have on the properties of the estimator of the memory parameter in small samples. Early on, Chung & Baillie (1993) and Cheung & Diebold (1994) conducted simulation studies and found that the inclusion of a constant term in the model can substantially increase the small-sample bias and mean squared error (MSE) of the estimated memory parameter. Lieberman & Phillips (2005) and Johansen & Nielsen (2016) are among the few theoretical contributions to shed light on the issue. Lieberman & Phillips (2005) derive the Edgeworth expansion of the memory parameter for the Gaussian maximum likelihood estimator in stationary fractional time series model. Johansen & Nielsen (2016) investigate the impact of observed and unobserved initial values on the bias of the memory parameter estimator in a non-stationary fractional time series model. Neither paper, however, includes short-run dynamics in its model. In addition, we are not aware of any related work that simultaneously tackles both stationary and non-stationary processes.

The purpose of the present paper is therefore to add to this literature and analyse the small-sample bias of the CSS estimator in a type-II fractional model with short-run dynamics and constant term from an analytical, empirical and simulation point of view. In particular, our analysis reveals that incorporating the level parameter into the model introduces an additional bias in the CSS estimator. This bias is due to a biased score which is particularly pronounced when the data is stationary. We will suggest what we call the modified conditional sum-of-squares (MCSS) estimator which is easy to compute, removes the leading bias term and allows much more accurate small-sample inference.

To do so, we will interpret the constant term as nuisance parameter and draw on a large literature on bias correction. Laskar & King (1998) provide an overview of this literature. We build on the approach to dealing with nuisance parameters initiated by Conniffe (1987) and McCullagh & Tibshirani (1990) and recently applied by Bartolucci, Bellio, Salvan & Sartori (2016) and by Martellosio & Hillier (2020), i.e. we adjust the score function so that its expectation equals zero. The idea is as follows: We find a stochastic higher-order expansion of the estimator as a function of the derivatives of the profile likelihood, cf. Johansen & Nielsen (2016) and Lawley (1956). The expansion is simplified by approximating the derivatives by their leading terms. This allows the expectation of the estimator to be taken. We notice that premultiplying the objective function by a modification term results in the expected score evaluated at the true parameter to be equal to zero, thereby mitigating the bias of the estimator.

The main contributions of this paper to the literature are threefold: First, we examine our MCSS estimator in type-II ARFIMA(p_1,d,p_2) models with constant term and compare it to the standard CSS estimator. In particular, we derive its exact bias and we show that it is consistent and asymptotically normally distributed. For expositional clarity, our treatment starts by covering the type-II ARFIMA(0,d,0) model before treating the general ARFIMA(p_1,d,p_2) case. The results generate new insights into bias correction of other models nested in our setup, such as stationary and invertible ARMA(p_1,p_2) models. Secondly, we re-visit three classical datasets that have in the past been described by ARFIMA(p_1,d,p_2) models with constant term, namely the post-second World War real GNP data, the extended Nelson-Plosser dataset, and the Nile data, by applying our MCSS estimator to estimate the long-memory parameter and the short-run dynamics. All three time series are short and therefore warrant the use of small-sample bias corrections. Our conclusion sheds new light on the interpretation of these datasets. Thirdly, this paper paves the way to extending the analysis of small-sample bias from univariate type-II ARFIMA modes to panel settings, see also the contributions of Robinson & Velasco (2015) and Schumann, Severini & Tripathi (2023).

The remainder of the paper is organised as follows. In Section 2 we present the MCSS estimator for ARFIMA$(0, d, 0)$ models. The extension to ARFIMA(p_1, d, p_2) models is covered in Section 3. Section 4 presents the empirical illustrations. Section 5 contains concluding remarks. All proofs are relegated to the appendix.

2 The modified conditional sum-of-squares estimator

In this section, we introduce the modified conditional sum-of-squares estimator, designed for estimating the ARFIMA(0,d,0) model with a constant term. In Section 3, we will expand our analysis to incorporate short-run dynamics, for instance covering the ARFIMA(p_1, d, p_2) model as a particular case. Covering the ARFIMA(0,d,0) model first serves two purposes: it allows a more straightforward explanation of our methodology, without the need for cumbersome notation, and affords a direct comparison with Lieberman & Phillips (2005) and Johansen & Nielsen (2016), both of which do not consider short-term dynamics. The theorems presented in this section are special cases of the theorems in Section 3, proofs of which are presented in the appendix.

We start with introducing the simple type-II ARFIMA(0,d,0) model with a constant term μ in Section 2.1. Building upon this, Section 2.2 introduces the conditional sum-of-squares (CSS) estimator and discusses its asymptotic properties, distinguishing between two scenarios: one where the constant parameter μ is either known or unknown. Subsequently, Section 2.3 shifts our focus towards examining score biases in the CSS estimators. It unveils a methodological approach to mitigating these biases: It discusses a well-established approach by McCullagh & Tibshirani (1990) and demonstrates that, in our setting, its application poses challenges. An adjustment of their approach resolves the issues. Section 2.4 introduces the modified conditional sum-of-squares (MCSS) estimator and delineates its asymptotic properties. In Section 2.5, we discuss how our MCSS estimator aligns with alternative bias-reduction methodologies. Section 2.6 assesses the asymptotic biases of the estimators, with specific attention to the performance of the MCSS estimator relative to the CSS estimator. Analytical expressions for these biases are derived, offering an understanding of their behaviour across distinct regions of d. Finally, in Section 2.7, we conduct a simulation study to examine the small sample properties of the estimators.

2.1 The model

Consider a so-called type II fractional process z_t, $t = 0, \pm 1, \pm 2, \ldots$, generated by the model

$$z_t = \Delta_+^{-d} \epsilon_t, \tag{1}$$

where $\Delta = 1 - L$ and L are the difference and lag operators, respectively, and where d can take any value in $\mathbb{R}$. For any series u_t, real number ζ and time index $t \geq 1$, the so-called truncation operator Δ_+^{ζ} is defined by

$$\Delta_+^{\zeta} u_t = \Delta^{\zeta}\{u_t I(t \geq 1)\} = \sum_{i=0}^{t-1} \pi_i(-\zeta) u_{t-i}, \tag{2}$$

with $I(\cdot)$ being the indicator function, and with $\pi_i(a) = 0$ for $i < 0$, $\pi_0(a) = 1$ as well as

$$\pi_i(a) = \frac{\Gamma(a+i)}{\Gamma(a)\Gamma(1+i)} = \frac{a(a+1)\ldots(a+i-1)}{i!}, \text{ for } i \geq 1, \tag{3}$$

denoting the coefficients in the usual binomial expansion of $\Delta^{-a} = \sum_{i=0}^{\infty} \pi_i(a) z^i$. $\Gamma(\cdot)$ is the gamma function with the convention that $\Gamma(i) = 0$ for $i = 0, -1, -2, \ldots$ and that $\Gamma(0)/\Gamma(0) = 1$. The parameter d in (1) is known as the memory parameter or the fractional parameter. The process z_t has been widely applied in the literature, see Marinucci & Robinson (2000, 2001), Robinson & Hualde (2003), Nielsen (2004), Shimotsu & Phillips (2005), Robinson (2005) and Johansen (2008), among others.

Two comments on the memory parameter are of interest: First, its range is commonly divided into a "stationary" and a "non-stationary" region: $d < 1/2$ and $d \geq 1/2$, respectively. Yet the definition in (2) implies that $z_t = 0$ for $t \leq 0$, which means that when $d < 1/2$ and $d \neq 0$, z_t is in fact not covariance stationary. However, it may be considered asymptotically stationary for any such d. To see this, consider the so-called type-I fractional process

$$\tilde{z}_t = \Delta^{-d} \epsilon_t \tag{4}$$

which is known to be covariance stationary for any $d < 1/2$. Marinucci & Robinson (1999) observe that for $|d| < 1/2$,

$$E(z_t - \tilde{z}_t)^2 = O(t^{2d-1}), \quad \text{as } t \to \infty, \tag{5}$$

and hence the difference to z_t vanishes. Although Marinucci & Robinson (1999) consider only $|d| < 1/2$, (5) actually holds for any $d < 1/2$. This follows from Stirling's approximation and Johansen & Nielsen (2016, Lemma A.1). This asymptotic equivalence prompts us to retain the terminological dichotomy between stationarity an non-stationarity. Secondly, it is worth noting that even for $d \geq 1/2$, i.e. in the non-stationary region, the truncation operator in (2) ensures that the process z_t is well-defined in the mean-square sense, see Johansen (2008, Section A.4) and Hualde & Robinson (2011).

While the model in (1) covers a wide range of dynamics, it seems unsuitable for many empirical applications because it implies that $E(z_t) = 0$. Nonetheless, a fair amount of theoretical work considers exclusively a purely random process, see for instance Hualde & Robinson (2011) and Nielsen (2015). In order to make our model more widely applicable, we complement the model in (1) by a constant term μ, to yield

$$x_t = \mu I(t \geq 1) + z_t \tag{6}$$

and hence $E(x_t) = \mu$ for $t \geq 1$. The level parameter μ has the added advantage of reducing the bias in the estimate of d arising from the pre-sample behaviour of x_t, as shown by Johansen & Nielsen (2016) for $d > 1/2$.

The model in (6) is the well-known ARFIMA(0,d,0) model plus a level parameter. It is considered as a special case in Hualde & Nielsen (2020) and Hualde & Nielsen (2022), both of which include short-run dynamics and a trending component in the model. We do not include a trend component in our analysis, but a discussion to that effect is presented in Section 5. We do extend our results, however, by adding short-run dynamics to (6) in Section 3.

2.2 The conditional sum-of-squares estimator

We now discuss the conditional sum-of-squares (CSS) estimator of the parameters in model (6). This is the estimator considered by e.g. Hualde & Robinson (2011) who, however, look at a model without the constant term. We distinguish the case in which μ is unknown from that in which it is known. As will be seen in Section 2.3 below, the CSS estimator may also be motivated as a maximum likelihood estimator under the assumption of Gaussian innovation terms ϵ_t, as in Johansen & Nielsen (2016) and Hualde & Nielsen (2020).

Following Johansen & Nielsen (2016), we make the following assumptions on the model's error term and the admissible parameter space. True parameter values are denoted by the subscript 0.

Assumption 2.1. *The errors ϵ_t are IID($0,\sigma_0^2$) with finite fourth moment.*

Assumption 2.2. *The parameter space for (d, μ) is $\mathbb{D} \times \mathbb{R}$, where $\mathbb{D} = [\nabla_1, \nabla_2]$, $-\infty < \nabla_1 < \nabla_2 < \infty$. The true value d_0 is in the interior of $\mathbb{D}$ and not equal to $1/2$.*

For any $(d, \mu) \in \mathbb{D} \times \mathbb{R}$, define the residuals $\epsilon_t(d, \mu) = \Delta_+^d (x_t - \mu)$. The CSS objective function is then given by

$$\begin{aligned} L(d, \mu) &= \frac{1}{2} \sum_{t=1}^{T} \epsilon_t^2(d, \mu), \\ &= \frac{1}{2} \sum_{t=1}^{T} \left(\Delta_+^d (x_t - \mu) \right)^2 . \end{aligned} \tag{7}$$

Since (7) is quadratic in μ we can concentrate it by writing

$$\begin{aligned} \Delta_+^d (x_t - \mu) &= \Delta_+^d x_t - \sum_{n=0}^{t-1} \pi_n(-d)\mu, \\ &= \Delta_+^d x_t - \pi_{t-1}(1-d)\mu = \Delta_+^d x_t - \kappa_{0t}(d)\mu, \end{aligned}$$

where

$$\begin{aligned} \kappa_{0t}(d) &= \Delta_+^d I(t \geq 1), \\ &= \sum_{n=0}^{t-1} \pi_n(-d) = \pi_{t-1}(1-d), \end{aligned} \tag{8}$$

the last line following from Johansen & Nielsen (2016, Lemma A.4). Unsurprisingly, the CSS estimator of μ for fixed d is given by

$$\hat{\mu}(d) = \frac{\sum_{t=1}^{T} (\Delta_+^d x_t) \kappa_{0t}(d)}{\sum_{t=1}^{T} \kappa_{0t}^2(d)}. \tag{9}$$

Substituting $\hat{\mu}(d)$ into (7) yields the profile (or concentrated) CSS function

$$L^*(d) = \frac{1}{2} \sum_{t=1}^{T} \left(\Delta_+^d (x_t - \hat{\mu}(d)) \right)^2 . \tag{10}$$

Note that we use asterisks to emphasise that we are dealing with a *profile* objective function. The resulting CSS estimator of d is given by

$$\hat{d} = \underset{d \in \mathbb{D}}{\operatorname{argmin}} \, L^*(d). \tag{11}$$

As discussed in Section 2.1, the model effectively conditions on $z_t = 0$, for $t \leq 0$. For this reason, Hualde & Robinson (2011) and Hualde & Nielsen (2020) prefer to call the estimator in (11) the truncated sum-of-squares estimator.

Hualde & Nielsen (2020, Theorem 1 and Theorem 2) show that if x_t is generated by (6) and if Assumption 2.1 and 2.2 hold, then, as $T \to \infty$,

$$\hat{d} \xrightarrow{p} d_0 \tag{12}$$

and

$$\sqrt{T}(\hat{d} - d_0) \xrightarrow{d} N(0, \zeta_2^{-1}), \tag{13}$$

where $\zeta_2^{-1} = 6/\pi^2$.

A few remarks about the estimator $\hat{\mu}(d)$ in (9) are instructive. For $d = d_0$ we have that

$$\hat{\mu}(d_0) - \mu_0 = \frac{\sum_{t=1}^{T} \epsilon_t \kappa_{0t}(d_0)}{\sum_{t=1}^{T} \kappa_{0t}^2(d_0)},$$

which has mean zero and variance $\sigma_0^2(\sum_{t=1}^{T} \kappa_{0t}^2(d_0))^{-1}$. In the stationary region, i.e. when $d_0 < 1/2$, this variance goes to zero because then $\sum_{t=1}^{T} \kappa_{0t}^2(d_0)$ diverges in T, see Lemma A.20. As opposed to that, in the non-stationary region, i.e. when $d_0 > 1/2$, this variance does not go to zero because then $\sum_{t=1}^{T} \kappa_{0t}^2(d_0)$ is bounded in T, see Lemma A.14. This is the reason why

$$\hat{\mu}(\hat{d}) \xrightarrow{p} \mu_0 \tag{14}$$

only if $d_0 < 1/2$, see Hualde & Nielsen (2020, Corollary 1) for the proof.

For comparison, we also analyse the situation where the true μ_0 is known. As mentioned earlier, this may often not be particularly realistic in practice. The CSS estimator for this model can be derived by substituting μ_0 into (7) to have

$$L_{\mu_0}^*(d) = \frac{1}{2} \sum_{t=1}^{T} \left(\Delta_+^d (x_t - \mu_0) \right)^2 \tag{15}$$

such that

$$\hat{d}_{\mu_0} = \operatorname*{argmin}_{d \in \mathbb{D}} L_{\mu_0}^*(d). \tag{16}$$

This estimator is considered by Hualde & Robinson (2011) and Nielsen (2015) who show that if x_t is generated by (6) and if Assumption 2.1 and 2.2 hold, then, as $T \to \infty$,

$$\hat{d}_{\mu_0} \xrightarrow{p} d_0 \tag{17}$$

and

$$\sqrt{T}(\hat{d}_{\mu_0} - d_0) \xrightarrow{d} N(0, \zeta_2^{-1}), \tag{18}$$

where $\zeta_2^{-1} = 6/\pi^2$. Remarkably, the asymptotic distribution of $\hat{d}_{\mu_0}$ is identical to that of $\hat{d}$ in (13). In other words, the distribution does not depend on whether μ is known or needs to be estimated. This contrasts to, for instance, unit root models in which the asymptotic distribution of the first-order serial correlation coefficient hinges on whether μ is known or not.

2.3 The modified profile likelihood

A central concern in this paper is to investigate the bias of $\hat{d}$ in (11) and of $\hat{d}_{\mu_0}$ in (16). This will be done in Section 2.6 below. It will turn out that the expectation of the CSS estimators is a function of the expectation of the score functions, or first derivatives, of $L^*(d)$ and $L_{\mu_0}^*(d)$ evaluated at $d = d_0$, respectively. The present section will therefore examine the bias of the two scores and builds on an approach by McCullagh & Tibshirani (1990) to correct for it.

To that end, it will be instructive to interpret the CSS objective in (7) as a log-likelihood function, as do Johansen & Nielsen (2016) and Hualde & Nielsen (2020). Assuming that $\epsilon_t \sim NID(0, \sigma^2)$, the Gaussian log-likelihood of x_t in (6), conditional on $x_t = 0$ for $t \leq 0$, is given by

$$\ell(d, \mu, \sigma^2) = -\frac{T}{2} \log\left(\sigma^2\right) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} \left(\Delta_+^d (x_t - \mu)\right)^2. \tag{19}$$

Throughout the paper, we omit additive constants in the likelihood functions for notational simplicity. Maximising (19) with respect to σ^2 yields

$$\hat{\sigma}^2(d, \mu) = \frac{1}{T} \sum_{t=1}^{T} \left(\Delta_+^d (x_t - \mu)\right)^2$$

and the profile log-likelihood

$$\begin{aligned} \ell(d, \mu) &= \ell(d, \mu, \hat{\sigma}^2(d, \mu)) \\ &= -\frac{T}{2} \log\left(\frac{1}{T} \sum_{t=1}^{T} \left(\Delta_+^d (x_t - \mu)\right)^2\right). \end{aligned} \tag{20}$$

Maximising (20) further with respect to μ results in $\hat{\mu}(d)$ in (9) and the profile log-likelihood function

$$\begin{aligned} \ell^*(d) &= \ell(d, \hat{\mu}(d), \hat{\sigma}^2(d, \hat{\mu}(d))) \\ &= -\frac{T}{2} \log\left(\frac{1}{T} \sum_{t=1}^{T} \left(\Delta_+^d (x_t - \hat{\mu}(d))\right)^2\right). \end{aligned} \tag{21}$$

Clearly, the estimator of d resulting from maximising (21) is identical to that obtained by minimising (10) since

$$L^*(d) = \frac{T}{2} \exp\left(-\frac{2}{T} \ell^*(d)\right). \tag{22}$$

So, the CSS objective $L^*(d)$ can be seen as a negative non-logged profile likelihood. As the maximum likelihood estimator of d is asymptotically efficient, see Hualde & Nielsen (2020), so is the CSS estimator $\hat{d}$ in (11). The same can of course be said of $\hat{d}_{\mu_0}$ in (16) since the profile CSS objective $L^*_{\mu_0}(d)$ in (15) can be obtained from (20) by replacing μ by its known value μ_0 such that

$$L^*_{\mu_0}(d) = \frac{T}{2} \exp\left(-\frac{2}{T} \ell(d, \mu_0)\right). \tag{23}$$

We will in the present section therefore interpret $L^*(d)$ in (10) as a profile likelihood. As such, it is not a genuine likelihood, for it is not directly based on observable quantities, see Barndorff-Nielsen (1983) and Severini (2000). Instead, it is a function of the maximum likelihood estimators of μ and σ^2 which are treated as if they were the true parameter values. In large samples, the concentration procedure has relatively minor effects, yet

Chung & Baillie (1993) showed in Monte Carlo simulations that in small samples it leads to a strong bias in $\hat{d}$. This is because profile likelihoods do not necessarily possess the same properties as genuine likelihoods. It is well-known that, under classical regularity conditions and with a fixed number of regressors, the score of the profile likelihood is biased. In particular, its expectation is $O(1)$, see Kalbfleisch & Sprott (1973), McCullagh & Tibshirani (1990) and Liang & Zeger (1995). The following theorem derives the bias of the score of $L^*(d)$. The proof will be given in Appendix A.4.1. Note that we adopt Euler's notation and denote the m^{th} derivative of a function $f(d)$ with respect to d by the operator D^m such that $D^m f(d) = \frac{\partial^m}{\partial d^m} f(d)$.

Theorem 2.1. *Let x_t, $t = 1,\ldots,T$, be given by* (6) *and let Assumption 2.1 be satisfied. Then, the expected score of $L^*(d)$, evaluated at the true parameter d_0, is given by*

$$E\left(DL^*(d_0)\right) = O(\log(T)I(d_0 < 1/2) + I(d_0 > 1/2)), \tag{24}$$

when $T \to \infty$.

Clearly, the score $DL^*(d)$ is biased. In addition, the bias is not uniform in d_0: The classical result that $E(DL^*(d_0)) = O(1)$ only holds for $d_0 > 1/2$. For $d_0 < 1/2$, however, the expectation of the score diverges at rate $\log(T)$. The competition between the stochastic and the deterministic component explains this difference in orders. In the non-stationary region, i.e. when $d_0 > 1/2$, we recall that $\hat{\mu}(\hat{d})$ is not consistently estimated, see the discussion in Section 2.2. The reason is that the stochastic component z_t in (6) dominates the deterministic component μ. Hence, the bias in the score is less influenced by $\hat{\mu}(\hat{d})$, resulting in the expected score being $O(1)$ for such d_0. On the other hand, if $d_0 < 1/2$, $\hat{\mu}(\hat{d})$ is consistently estimated and $\hat{\mu}(\hat{d})$ plays a more important role in the bias of the score. This is reflected in the expected score being $O(\log(T))$ for such d_0.

The order of magnitude in (24) also applies to the expectation of the score function of the profile log-likelihood function $\ell^*(d)$ in (21). To see this, note that (22) entails

$$D\ell^*(d) = -\frac{1}{2}\frac{DL^*(d)}{T^{-1}L^*(d)}. \tag{25}$$

From Theorem 2.1 we can then deduce the following corollary. The proof is omitted.

Corollary 2.1. *Let x_t, $t = 1,\ldots,T$, be given by* (6) *and let Assumption 2.1 be satisfied. Then, the expected score of $l^*(d)$, evaluated at the true parameter d_0, is given by*

$$E\left(D\ell^*(d_0)\right) = O(\log(T)I(d_0 < 1/2) + I(d_0 > 1/2)),$$

when $T \to \infty$.

The situation for $L^*_{\mu_0}(d)$ in (15) is somewhat different. Although, technically speaking, $L^*_{\mu_0}(d)$ is also a profile likelihood, it will be proved in Appendix A.4.1 that its score is unbiased despite the substitution of $\hat{\sigma}^2$ for σ^2. This is summarised in the following theorem.

Theorem 2.2. *Let x_t, $t = 1,\ldots,T$, be given by* (6) *and let Assumption 2.1 be satisfied. Then, the expected score of $L^*_{\mu_0}(d)$, evaluated at the true parameter d_0, is given by*

$$E\left(DL^*_{\mu_0}(d_0)\right) = 0 \tag{26}$$

when $T \to \infty$.

This discussion highlights the need for a modification of the profile likelihood function such that it behaves more like a genuine likelihood in terms of score unbiasedness. This modification will eliminate the bias of the CSS estimator $\hat{d}$ stemming from the presence of the unknown nuisance parameter, as will be seen in Section 2.6. The idea of modifying the profile likelihood to obtain score unbiasedness is in fact not new and was previously discussed by McCullagh & Tibshirani (1990). Martellosio & Hillier (2020), for instance, implement this idea for a spatial model.

To obtain an unbiased score, McCullagh & Tibshirani (1990) recenter the score of the profile log-likelihood function, yielding, say,

$$D\ell_a^*(d) = D\ell^*(d) - a(d), \tag{27}$$

where $\ell^*(d)$ denotes, as before, the profile log-likelihood function and where $a(d)$ is an adjustment function only depending on d. Then they require that

$$E\left(D\ell_a^*(d_0)\right) = 0, \tag{28}$$

which implies that

$$a(d_0) = E\left(D\ell^*(d_0)\right), \tag{29}$$

for all d_0. Finally, they call

$$\ell_a^*(d) = \int_d D\ell_a^*(t)dt, \tag{30}$$

the adjusted profile log-likelihood for d, which is subsequently maximised w.r.t. d.

Remark 2.3. McCullagh & Tibshirani (1990) further adjust $D\ell_a^*(d)$ to make it information unbiased, i.e. making its variance equals to the negative expectation of the derivative of the score. While these adjustments may improve the efficiency of the estimator, they are not addressed in this paper because they do not affect the location of the zeros of $D\ell_a^*(d)$.

The adjustment function $a(d)$ can in principle be computed from (29). Yet this calculation is challenging as can be seen by rewriting (29) as

$$a(d_0) = E\left(-\frac{1}{2}\frac{DL^*(d_0)}{T^{-1}L^*(d_0)}\right), \tag{31}$$

see (25). Evaluating the expectation of the fraction is not trivial. Martellosio & Hillier (2020) circumvent this problem by assuming ϵ_t to be Gaussian, thereby effectively using the profile log-likelihood $\ell^*(d)$ as basis for the adjustment. We, however, avoid this strong assumption and consider the profile CSS objective function $L^*(d)$ instead, as explained in Section 2.2. Indeed, McCullagh & Tibshirani (1990) in their Remark 3 allude to the possibility of using an objective function other than the profile likelihood for deriving an adjustment. To that end, we need to frame the approach of McCullagh & Tibshirani (1990) in terms of $L^*(d)$.

To do this, first note that (30) can be written as

$$\ell_a^*(d) = \int_d \left(D\ell^*(t) - a(t)\right)dt$$

$$= \ell^*(d) - A(d), \tag{32}$$

with $A(d) = \int_d a(t)dt$. Based on the relationship between $\ell^*(d)$ and $L^*(d)$ in (22), we can write (32) as

$$\ell_a^*(d) = -\frac{T}{2}\log\left(\frac{2}{T}L^*(d)\right) - A(d). \tag{33}$$

Using a similar argument as in (22), it is clear that maximizing the adjusted profile log-likelihood $\ell_a^*(d)$ in (33) is equivalent to minimising the adjusted profile CSS objective, defined as

$$L_a^*(d) = \frac{T}{2}\exp\left(-\frac{2}{T}\ell_a^*(d)\right). \tag{34}$$

Finally, replacing $\ell_a^*(d)$ in (34) by (33) yields

$$L_a^*(d) = \exp\left(\frac{2}{T}A(d)\right)L^*(d). \tag{35}$$

It is important to note that while the adjustment in (33) is additive, it is multiplicative in (35). Recall that $a(d)$ in (31), and thus $A(d)$ in (35), is difficult to compute. We therefore define, as an alternative, the modified profile CSS objective function

$$L_m^*(d) = m(d)L^*(d) \tag{36}$$

where the multiplicative modification term $m(d) > 0$ depends only on d. The corresponding score function is the first derivative of (36):

$$DL_m^*(d) = m(d)DL^*(d) + Dm(d)L^*(d). \tag{37}$$

As McCullagh & Tibshirani (1990), we now require that our objective function is score unbiased, i.e. that the score function in (37) satisfy

$$E\left(DL_m^*(d_0)\right) = 0, \tag{38}$$

cf. (28). Using (37) and the fact that $D\log(m(d)) = Dm(d)/m(d)$ and $m(d) > 0$ it follows that (38) is equivalent to the condition that

$$D\log\left(m(d_0)\right) = -\frac{E\left(DL^*(d_0)\right)}{E\left(L^*(d_0)\right)}. \tag{39}$$

It will be seen below that the evaluation of the right-hand side of (39) is straightforward, as opposed to the evaluation of (31). In particular, it avoids imposing an additional normality assumption.

2.4 The modified conditional sum-of-squares estimator

The condition in (39) is now used for finding the modification term $m(d)$ for the modified profile CSS objective in (36): First, it is shown in Lemma A.13 that the expectation of $DL^*(d_0)$ equals

$$E\left(DL^*(d_0)\right) = -\sigma_0^2 \frac{\sum_{t=1}^{T}\kappa_{0t}(d_0)\kappa_{1t}(d_0)}{\sum_{t=1}^{T}\kappa_{0t}^2(d_0)},$$

where $\kappa_{0t}(d) = \pi_{t-1}(1-d)$ and $\kappa_{1t}(d) = D\kappa_{0t}(d) = -D\pi_{t-1}(1-d)$ see also (8). It is also shown in Lemma A.13 that

$$E\left(L^*(d_0)\right) = \sigma_0^2 \frac{T-1}{2}.$$

Consequently, from (39), we have

$$D \log\left(m(d_0)\right) = \frac{2}{T-1} \frac{\sum_{t=1}^{T} \kappa_{0t}(d_0)\kappa_{1t}(d_0)}{\sum_{t=1}^{T} \kappa_{0t}^2(d_0)}. \tag{40}$$

Upon integrating (40) we obtain

$$\log\left(m(d)\right) = \log\left(\sum_{t=1}^{T} \kappa_{0t}^2(d)\right)^{\frac{1}{T-1}},$$

before, finally, exponentiation yields

$$m(d) = \left(\sum_{t=1}^{T} \kappa_{0t}^2(d)\right)^{\frac{1}{T-1}}. \tag{41}$$

The modified profile CSS objective function in (36) is thus given by the product of $m(d)$ in (41) and $L^*(d)$ in (10), i.e.

$$L_m^*(d) = \left(\sum_{t=1}^{T} \kappa_{0t}^2(d)\right)^{\frac{1}{T-1}} \frac{1}{2} \sum_{t=1}^{T} \left(\Delta_+^d(x_t - \hat{\mu}(d))\right)^2.$$

We call the estimator that minimises $L_m^*(d)$ the modified conditional sum-of-squares (MCSS) estimator and denote the estimator of d by $\hat{d}_m$, i.e. ,

$$\hat{d}_m = \underset{d \in \mathbb{D}}{\operatorname{argmin}}\, L_m^*(d). \tag{42}$$

Two important properties of the modification term $m(d)$ are stated in the following lemma. See Appendix A.4.2 for the proof.

Lemma 2.1. *For all $d \in \mathbb{R}$,*

$$m(d) \geq 1, \tag{43}$$

where equality holds if and only if $d = 1$. Also, it holds that, for $T \to \infty$,

$$m(d) = 1 + O(T^{-1}\log(T)I(d < 1/2) + T^{-1}I(d > 1/2)). \tag{44}$$

for all $d \in \mathbb{R}\backslash\{1/2\}$

The property in (43) implies that the modification term $m(d)$ acts as a penalisation in the minimisation of the modified profile likelihood $L_m^*(d)$ through inflating $L^*(d)$ by $m(d)$. The property in (44) ensures that $m(d) \to 1$ such that the asymptotic properties of the MCSS estimator $\hat{d}_m$ are the same as those of the CSS estimator $\hat{d}$ in (11). This is desirable because the CSS estimator is efficient under Gaussianity, as argued in Section 2.3. The asymptotic properties of $\hat{d}_m$ are summarised for completeness in the following theorem and are proved in Appendix A.4.3.

Theorem 2.4. *Let x_t, $t = 1,\ldots,T$, be given by* (6) *and let Assumption 2.1 and 2.2 be satisfied. Then, as $T \to \infty$,*

$$\hat{d}_m \xrightarrow{p} d_0, \tag{45}$$

and

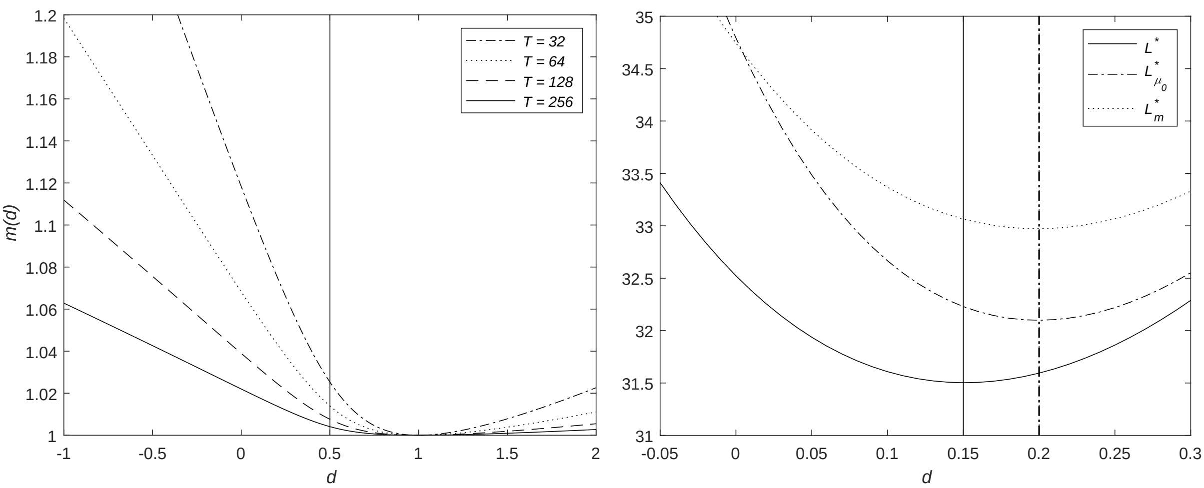
$$\sqrt{T}(\hat{d}_m - d_0) \xrightarrow{d} N(0, \zeta_2^{-1}), \tag{46}$$

where $\zeta_2^{-1} = 6/\pi^2$.

The intuitive explanation of Theorem 2.4 follows from noticing that

$$\begin{aligned} L_m^*(d_0) &= L^*(d_0) + O_P(1) && \text{for } d_0 > 1/2, \\ L_m^*(d_0) &= L^*(d_0) + O_P(\log(T)) && \text{for } d_0 < 1/2, \end{aligned} \tag{47}$$

where use was made of the definition $L_m^*(d_0)$ in (36) and the asymptotic behaviour of $m(d)$ in (44) of Lemma 2.1. Since $L^*(d_0)$ in (47) is $O_P(T)$, the second summands have no influence on the asymptotic distribution of $\hat{d}_m$. For the bias, however, the latter terms require further analysis, which is carried out below in Section 2.6.



(a) modification term (b) likelihoods

Figure 1: Panel (a) plots the modification term $m(d)$ in (41) for d between -1 and 2, and $T = 32$, 64, 128, 256. The value of $d = 1/2$ is added as a vertical line for clarity. Panel (b) shows the Monte Carlo average over 10,000 replications of $L^*(d)$, $L_{\mu_0}^*(d)$ and $L_m^*(d)$. The DGP is given in (6) with $\epsilon_t \sim NID(0, 1)$ with $d_0 = 0.2$, $\mu_0 = 0$ and $T = 64$.

The modification term is plotted in panel (a) of Figure 1 for some illustrative values of d and T. Four important observations can be made: First, recall from (43) that the modification term $m(d)$ penalises the CSS objective $L^*(d)$ through inflating it by the factor $m(d)$. It appears from the plot that in the stationary region, i.e. when $d < 1/2$, $m(d)$ inflates $L^*(d)$ more than in the non-stationary region, i.e. when $d \geq 1/2$. This is a reflection of the fact that the bias in the score is larger in the stationary region, as was argued in Theorem 2.1. Secondly, it is plain that when $d = 1$ the bias caused by estimating the constant term μ is the smallest, as predicted in Lemma 2.1. Thirdly, even for a moderately large sample of size $T = 256$, $m(d)$ still turns out to be substantial in the stationary region, implying that the corresponding bias in the score is large. Fourthly, the negative slope of the modification term $m(d)$ for $d < 1$ implies that the minimum of $L_m^*(d)$ is shifted to the right of that of $L^*(d)$. This is illustrated in panel (b) of Figure 1 which displays a Monte Carlo simulation of the CSS and MCSS objective functions. The DGP is stationary and corresponds to the model in (6) with $\epsilon_t \sim NID(0,1)$, $d_0 = 0.2$ and $\mu_0 = 0$. The sample size is $T = 64$ and the number of replications is 10,000. On display is the Monte Carlo average of the simulated $L^*(d)$, $L_{\mu_0}^*(d)$ and $L_m^*(d)$. The solid line represents the Monte Carlo average of $L^*(d)$: it can be seen that the CSS estimator underestimates the true $d_0 = 0.2$ on average. The dash-dotted line represents the Monte Carlo average of $L_{\mu_0}^*(d)$, which takes the constant term as known. This estimator is, on average, close to d_0. The dotted line represents the Monte Carlo average of $L_m^*(d)$, whose minimum is shifted to the right of that of $L^*(d)$. It therefore corrects for the distortion in $L^*(d)$ caused by estimating μ.

2.5 Relationship with other modifications

There is a large literature on correcting the bias of maximum likelihood caused by the presence of unknown nuisance parameters. Seminal contributions include Barndorff-Nielsen (1983) who proposed the modified likelihood function, and Cox & Reid (1987) who contributed the idea of the conditional profile likelihood by approximating the modified likelihood function. Both modifications result in modified profile likelihoods that are approximately score unbiased, see Liang (1987) and Cox & Barndorff-Nielsen (1994). It is therefore illuminating to investigate how our MCSS objective, with an expected score exactly equal to zero, relates to alternative approaches to bias-reduction, or how our modification term $m(d)$ compares to alternative adjustments. This section discusses two such ideas.

First, reconsider the adjusted profile log-likelihood $\ell_a^*(d)$ proposed by McCullagh & Tibshirani (1990) and derived in Section 2.3. Denote the corresponding estimator by

$$\hat{d}_a = \operatorname*{argmax}_{d \in \mathbb{D}} \ell_a^*(d).$$

The proof of the following corollary follows easily from (31) and (39) and is therefore omitted.

Corollary 2.2. *Let x_t, $t = 1,\ldots,T$, be given by* (6) *and let Assumption 2.1 and 2.2 be satisfied. Then if*

$$E\left(\frac{DL^*(d_0)}{L^*(d_0)}\right) = \frac{E\left(DL^*(d_0)\right)}{E\left(L^*(d_0)\right)} \tag{48}$$

it holds that $\hat{d}_m = \hat{d}_a$.

As mentioned earlier, it is not easy to evaluate the left-hand side of (48). One exception is if we assume ϵ_t to be NID. Yet it turns out that even in this special case (48) does not hold. The proof is omitted here but can be derived by applying Martellosio & Hillier (2020, Lemma S.7.1).

Secondly, a modification term closely related to $m(d)$ in (41) is the one discussed in An & Bloomfield (1993) who implement the idea of Cox & Reid (1987) to adjust the log-likelihood function. The setup in An & Bloomfield (1993) is different from ours, however: They consider a stationary Gaussian type-I fractional process $\tilde{x}_t$ generated by the model

$$\tilde{x}_t = \mu + \tilde{z}_t, \tag{49}$$

where $\tilde{z}_t$ is defined in (4) with $\epsilon_t \sim NID(0, \sigma^2)$ and $|d| < 1/2$. This contrasts to our type-II process whose d we also allow to lie in the non-stationary region and whose error term ϵ_t is not assumed to be Normally distributed.

It will prove helpful to phrase the approach by An & Bloomfield (1993) in matrix notation: Define the $T \times 1$ vector $\tilde{x} = (\tilde{x}_1, \ldots, \tilde{x}_T)'$ such that $\tilde{x} \sim N(\mu\iota, \sigma^2\Sigma(d))$ where ι is a $T \times 1$ vector of ones and $\Sigma(d)\sigma^2$ is the $T \times T$ variance-covariance matrix of $\tilde{x}$, see for instance Hosking (1981, Theorem 1) for the elements of $\Sigma(d)$. The log-likelihood function is then given by

$$\tilde{\ell}(d, \mu, \sigma^2) = -\frac{1}{2}\log(|\Sigma(d)|) - \frac{T}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\left(\tilde{x} - \iota\mu\right)' \Sigma(d)^{-1}\left(\tilde{x} - \iota\mu\right) \tag{50}$$

and the profile log-likelihood function for d by

$$\begin{aligned}\tilde{\ell}^*(d) &= \tilde{\ell}(d, \hat{\mu}(d), \hat{\sigma}^2(d, \hat{\mu}(d))) \\ &= -\frac{1}{2}\log(|\Sigma(d)|) - \frac{T}{2}\log\left(\frac{1}{T}\left(\tilde{x} - \iota\hat{\mu}(d)\right)' \Sigma(d)^{-1}\left(\tilde{x} - \iota\hat{\mu}(d)\right)\right),\end{aligned}$$

where $\hat{\mu}(d)$ and $\hat{\sigma}^2(d, \mu)$ are the maximum likelihood estimator of μ and σ, respectively. The modified profile log-likelihood function that An & Bloomfield (1993) find is

$$\tilde{\ell}_m^*(d) = \tilde{\ell}^*(d) + \frac{3}{2}\log\left(\hat{\sigma}^2(d, \hat{\mu}(d))\right) + \frac{1}{T}\log(|\Sigma(d)|) - \frac{1}{2}\log\left(\iota'\Sigma(d)^{-1}\iota\right). \tag{51}$$

It can be shown that the orders of magnitude of the last three summands on the right-hand side of (51) are $O_P(1)$, $O(1)$ and $O(\log(T))$, respectively. For the proof of the second summand, we refer to Dahlhaus (1989). The order of magnitude of the third summand follows from Hadamard's inequality, see Horn & Johnson (2013, Theorem 7.8.1), implying that $T^{-1}\log|\Sigma(d)| \leq \log(\mathrm{Var}(\tilde{x}_t)/\sigma^2)$ where the right-hand side is $O(1)$ because $|d| < 1/2$. The order of magnitude of the fourth summand follows from $\iota'\Sigma(d)^{-1}\iota = O(T^{1-2d+\varepsilon})$ for each $\varepsilon > 0$, cf. Adenstedt (1974, Theorem 5.2). The leading of the three summands is therefore the last one, its order of magnitude being $O(\log(T))$.

In order to compare $\tilde{\ell}_m^*(d)$ to our MCSS function $L_m^*(d)$ in (36) we transform the latter in a fashion similar to that in (22) or (34) and define

$$\ell_m^*(d) = -\frac{T}{2}\log\left(\frac{2}{T}L_m^*(d)\right),$$

such that

$$\ell_m^*(d) = -\frac{T}{2}\log\left(\frac{2}{T}m(d)L^*(d)\right),$$

$$= \ell^*(d) - \frac{T}{2}\log\left(m(d)\right),$$

with $\ell^*(d)$ given in (21). Using the definition of $m(d)$ in (41) yields

$$\begin{aligned}
\ell_m^*(d) &= \ell^*(d) - \frac{T}{(T-1)}\frac{1}{2}\log\left(\sum_{t=1}^{T}\kappa_{0t}^2(d)\right), \\
&= \ell^*(d) - \frac{1}{2}\log\left(\sum_{t=1}^{T}\kappa_{0t}^2(d)\right) + O_P(T^{-1}\log(T)), \qquad (52)
\end{aligned}$$

since $T/(T-1) = 1 + 1/(T-1)$. Clearly, the second summand is of order $O(\log T)$ as is shown in Lemma A.20.

Two observations are now instructive. First, the leading modification term in $\ell_m^*(d)$ is of the same order of magnitude as that in $\tilde{\ell}_m^*$ in (51), namely $O(\log(T))$. Second, we note that the Cholesky factor of $\Sigma(d)^{-1}$ is the GLS transformation matrix that filters out the correlation structure of the type-I error term $\tilde{z}_t$. Similarly, in our setting, Δ_+^d filters out the correlation structure of the type-II error z_t. Indeed, if we could replace $\Sigma(d)^{-1/2}\iota$ in (51) by $\Delta_+^d\iota$, we would obtain

$$\frac{1}{2}\log\left((\Delta_+^d\iota)'(\Delta_+^d\iota)\right) = \frac{1}{2}\log\left(\sum_{t=1}^{T}\kappa_{0t}^2(d)\right).$$

using the definition of $\kappa_{0t}(d)$ in (8). Let us emphasise again, however, that the approach by An & Bloomfield (1993), although asymptotically equivalent to ours, is based on a model that assumes stationary and Normally distributed data. In addition, it necessitates the computation of the $T \times T$ variance-covariance matrix, or its Cholesky factor, which is often onerous computationally.

2.6 Asymptotic biases

This section investigates the asymptotic biases of the estimators considered so far, with particular attention paid to the bias of the MCSS estimator $\hat{d}_m$ in (42). Two questions are of central interest. First, by how much does the MCSS estimator $\hat{d}_m$ reduce the bias of the CSS estimator $\hat{d}$ in (11)? Second, is the bias of the MCSS estimator $\hat{d}_m$ comparable to that of the CSS estimator with known μ_0? To address both questions, we find expressions for the asymptotic biases of $\hat{d}$, $\hat{d}_{\mu_0}$ and $\hat{d}_m$ based on a stochastic expansion of the estimators. For the non-stationary region, i.e. when $d_0 > 1/2$, the expansion of $\hat{d}$ and $\hat{d}_{\mu_0}$ as well as their asymptotic biases are derived in Johansen & Nielsen (2016). In the analysis below, we also consider $\hat{d}$ and $\hat{d}_{\mu_0}$ in the stationary region, i.e. when $d_0 < 1/2$. The bias of the MCSS estimator $\hat{d}_m$ is derived for $d_0 > 1/2$ and $d_0 < 1/2$.

Johansen & Nielsen (2016, Section 3.2) consider a second-order Taylor series expansion of $DL^*(\hat{d}) = 0$ around d_0, yielding

$$0 = DL^*(\hat{d}) = DL^*(d_0) + (\hat{d} - d_0)D^2L^*(d_0) + \frac{1}{2}(\hat{d} - d_0)^2D^3L^*(d^*), \qquad (53)$$

where d^* is an intermediate value satisfying $|d^* - d_0| \leq |\hat{d} - d_0| \xrightarrow{p} 0$. For $d_0 > 1/2$, they demonstrate that the derivatives satisfy $DL^*(d_0) = O_P(T^{1/2})$, $D^2L^*(d_0) = O_P(T)$, and $D^3L^*(d) = O_P(T)$ with d uniformly in a neighbourhood of d_0, see Johansen & Nielsen (2016, Lemma B.4), which allows them to obtain

$$T^{1/2}\left(\hat{d} - d_0\right) = -T^{1/2}\frac{DL^*(d_0)}{D^2L^*(d_0)} - \frac{1}{2}T^{-1/2}\frac{\left(DL^*(d_0)\right)^2 D^3L^*(d^*)}{\left(D^2L^*(d_0)\right)^3} + O_P(T^{-1}). \tag{54}$$

It is shown in Lemma A.23 below that the derivatives $DL^*(d_0)$, $D^2L^*(d_0)$ and $D^3L^*(d_0)$ are of the same orders of magnitude for $d_0 < 1/2$, implying that (54) holds for both the non-stationary and stationary region of d_0.

Based on (54), the asymptotic bias of $\hat{d}$ can be found. In a first step, the approximations of the derivatives in Johansen & Nielsen (2016, Lemma B.4) for the region $d_0 > 1/2$ and in Lemma A.23 for the region $d_0 < 1/2$ are used to find

$$E\left(\hat{d} - d_0\right) = -(T\zeta_2)^{-1}\left[\sigma_0^{-2}E\left(DL^*(d_0)\right)\right] - (T\zeta_2)^{-1}\left[3\zeta_3\zeta_2^{-1}\right] + o(T^{-1}), \tag{55}$$

where ζ_s is Riemann's zeta function $\zeta_s = \sum_{j=1}^{\infty} j^{-s}$, $s > 1$. This shows the relationship between the bias of the estimator and that of the score function. The first term on the right-hand side of (55) dominates the remainder only if $d_0 < 1/2$, because then $E\left(DL^*(d_0)\right) = O(\log(T))$ from Theorem 2.1. As opposed to that, we have from the same theorem that $E\left(DL^*(d_0)\right) = O(1)$ if $d_0 > 1/2$ and hence the first term on the right-hand side of (55) is of the same order as the remainder.

It follows from Lemmata A.18, A.19, A.24 and A.25 that analogues of the derivation in the previous two paragraphs also hold for $\hat{d}_{\mu_0}$ and $\hat{d}_m$. Replacing in (55) $DL^*(d)$ by $DL^*_{\mu_0}(d)$ and using Theorem 2.2, it is clear that the expected score term in $E\left(\hat{d}_{\mu_0} - d_0\right)$ vanishes. Similarly, $E\left(DL^*_m(d_0)\right) = 0$ by construction, justifying the modification of the CSS objective function in Section 2.4 in order to obtain score unbiasedness.

In a second step, explicit expressions of the expected scores are found and substituted into (55), yielding one of the main results of this paper, summarised in the following theorem. The results in (56) and (57) are derived in Johansen & Nielsen (2016, Theorem 4) and mentioned here for completeness. The proof of the other cases is presented in Appendix A.4.4.

Theorem 2.5. *Let x_t, $t = 1,\ldots,T$, be given by* (6) *and let Assumption 2.1 and 2.2 be satisfied. For the non-stationary region, i.e. when $d_0 > 1/2$, the biases of $\hat{d}$, $\hat{d}_{\mu_0}$ and $\hat{d}_m$ are*

$$bias(\hat{d}) = -(T\zeta_2)^{-1}\left[3\zeta_3\zeta_2^{-1} + \left(\Psi(d_0) - \Psi(2d_0 - 1)\right)\right] + o(T^{-1}), \tag{56}$$

$$bias(\hat{d}_{\mu_0}) = -(T\zeta_2)^{-1}\left[3\zeta_3\zeta_2^{-1}\right] + o(T^{-1}), \tag{57}$$

$$bias(\hat{d}_m) = -(T\zeta_2)^{-1}\left[3\zeta_3\zeta_2^{-1}\right] + o(T^{-1}), \tag{58}$$

where ζ_s is Riemann's zeta function $\zeta_s = \sum_{j=1}^{\infty} j^{-s}$, $s > 1$, and $\Psi(d) = D\log\Gamma(d)$ denotes the Digamma function. For the stationary region, i.e. when $d_0 < 1/2$, the biases of $\hat{d}$, $\hat{d}_{\mu_0}$ and $\hat{d}_m$ are

$$bias(\hat{d}) = -(T\zeta_2)^{-1}\left[\log(T) + 3\zeta_3\zeta_2^{-1} - \left(\Psi(1 - d_0) + (1 - 2d_0)^{-1}\right)\right] + o(T^{-1}), \tag{59}$$

$$bias(\hat{d}_{\mu_0}) = -(T\zeta_2)^{-1}\left[3\zeta_3\zeta_2^{-1}\right] + o(T^{-1}), \tag{60}$$

$$bias(\hat{d}_m) = -(T\zeta_2)^{-1}\left[3\zeta_3\zeta_2^{-1}\right] + o(T^{-1}). \tag{61}$$

The order of the leading term in (56) and (59) varies depending on d_0: if $d_0 < 1/2$ it is $O(T^{-1}\log(T))$ while if $d_0 > 1/2$ it is $O(T^{-1})$. Therefore, in the stationary region, the bias of the CSS estimator is stronger than in the non-stationary region. The difference in the orders can be ascribed to the expectation of the score function, see (55) and the subsequent discussion. As opposed to that, the leading bias term of $\hat{d}_m$ does not depend on d_0 and is identical in both regions. Furthermore, the leading bias terms in (57) and (60) are the same as the leading bias terms in (58) and (61). That is, the estimated $\hat{d}_m$ behaves, on average, the same as if we had known the true value μ_0, discounting the higher order bias term. The remaining bias term $(T\zeta_2)^{-1}3\zeta_3\zeta_2^{-1}$ is due to the correlations of the derivatives of the likelihood and is not eliminated by our modification. The same bias term appears also in Lieberman & Phillips (2004) for the bias of the estimated memory parameter based on the maximum likelihood estimator in the type-I fractional model in (50) with $0 < d_0 < 1/2$ and σ^2 as well as μ known, see Johansen & Nielsen (2016, p. 1106) for a discussion. As Johansen & Nielsen (2016, p. 1108) note, the key factor to assess the distortion in testing or calculating confidence intervals for d_0 is the relative bias, i.e. the ratio of asymptotic bias to asymptotic standard deviation. The asymptotic standard deviation of the three estimators is equal to $(T\zeta_2)^{-1/2}$, see (13), (18) and (46). Then it follows from Theorem 2.5 that the relative bias for the three estimators is of order $O(T^{-1/2})$ in the non-stationary region. In the stationary region, it is of order $O(T^{-1/2}\log(T))$ for the CSS estimator with unknown μ_0 while it of order $O(T^{-1/2})$ for the CSS estimator with known μ_0 and the MCSS estimator.

We would like to point out that the remaining bias term of $\hat{d}_m$ is pivotal and can be easily eliminated. We refer to this estimator as the bias-corrected MCSS (*bcm*) estimator and denote it by

$$\hat{d}_{bcm} = \hat{d}_m + T^{-1}3\zeta_3\zeta_2^{-2}. \tag{62}$$

We obtain from (58) and (61) the following property of this estimator.

Corollary 2.3. *Let x_t, $t = 1,\ldots,T$, be given by* (6) *and let Assumption 2.1 and 2.2 be satisfied. Then,*

$$bias(\hat{d}_{bcm}) = o(T^{-1}).$$

The proof follows directly from Theorem 2.5 and is therefore omitted.

	bias($\hat{d}$)	bias($\hat{d}_{\mu_0}$)	bias($\hat{d}_m$)	bias($\hat{d}$)	bias($\hat{d}_{\mu_0}$)	bias($\hat{d}_m$)	bias($\hat{d}$)	bias($\hat{d}_{\mu_0}$)	bias($\hat{d}_m$)	bias($\hat{d}$)	bias($\hat{d}_{\mu_0}$)	bias($\hat{d}_m$)
$d_0 \setminus T$	32			64			128			256		
−0.2	−9.94	−4.16	−4.16	−5.63	−2.08	−2.08	−3.14	−1.04	−1.04	−1.74	−0.52	−0.52
−0.1	−9.97	−4.16	−4.16	−5.64	−2.08	−2.08	−3.15	−1.04	−1.04	−1.74	−0.52	−0.52
0.0	−9.95	−4.16	−4.16	−5.63	−2.08	−2.08	−3.14	−1.04	−1.04	−1.74	−0.52	−0.52
0.1	−9.81	−4.16	−4.16	−5.56	−2.08	−2.08	−3.11	−1.04	−1.04	−1.72	−0.52	−0.52
0.2	−9.42	−4.16	−4.16	−5.37	−2.08	−2.08	−3.01	−1.04	−1.04	−1.67	−0.52	−0.52
0.3	−8.32	−4.16	−4.16	−4.82	−2.08	−2.08	−2.74	−1.04	−1.04	−1.53	−0.52	−0.52
0.4	−4.18	−4.16	−4.16	−2.75	−2.08	−2.08	−1.70	−1.04	−1.04	−1.02	−0.52	−0.52
0.5	-	-	-	-	-	-	-	-	-	-	-	-
0.6	−11.29	−4.16	−4.16	−5.64	−2.08	−2.08	−2.82	−1.04	−1.04	−1.41	−0.52	−0.52
0.7	−6.71	−4.16	−4.16	−3.36	−2.08	−2.08	−1.68	−1.04	−1.04	−0.84	−0.52	−0.52
0.8	−5.26	−4.16	−4.16	−2.63	−2.08	−2.08	−1.31	−1.04	−1.04	−0.66	−0.52	−0.52
0.9	−4.56	−4.16	−4.16	−2.28	−2.08	−2.08	−1.14	−1.04	−1.04	−0.57	−0.52	−0.52
1.0	−4.16	−4.16	−4.16	−2.08	−2.08	−2.08	−1.04	−1.04	−1.04	−0.52	−0.52	−0.52
1.1	−3.91	−4.16	−4.16	−1.95	−2.08	−2.08	−0.98	−1.04	−1.04	−0.49	−0.52	−0.52
1.2	−3.73	−4.16	−4.16	−1.87	−2.08	−2.08	−0.93	−1.04	−1.04	−0.47	−0.52	−0.52

Table 1: (100 $\times$) Theoretical bias, up to $o(T^{-1})$ terms, of the CSS estimator of d with unknown and known μ_0 and of the MCSS estimator of d.

Analysing the theoretical bias terms through numerical comparisons may assist in building an intuition. Table 1 therefore presents the theoretical biases, up to $o(T^{-1})$ terms, of the CSS estimator with unknown and known μ_0 and of the MCSS estimator for selected values of d_0 and T. It is evident that the bias of the CSS estimator decreases in both the stationary and non-stationary region as d_0 decreases, and decreases everywhere as T increases. As Johansen & Nielsen (2016, p. 1107) note, the bias of $\hat{d}$ is equal to that of $\hat{d}_{\mu_0}$ when $d_0 = 1$. Yet it is curious to see that bias($\hat{d}$) is actually smaller than bias($\hat{d}_{\mu_0}$) for $d_0 = 1.1$ and $d_0 = 1.2$. In fact, this occurs for all $d_0 > 1$. The reason is that $\Psi(d_0) - \Psi(2d_0 - 1)$ becomes negative for $d_0 > 1$ and therefore reduces the term $3\zeta_3\zeta_2^{-1} \approx 2.1923$ in (56) of Theorem 2.5. Note also that the term $\Psi(d_0) - \Psi(2d_0 - 1)$ is monotonically decreasing in $d_0 > 1/2$. Furthermore, from Abramowitz & Stegun (1964, eqn. 6.3.18), for $d_0 \to \infty$, it holds that $\Psi(d_0) = \log(d_0) + O(d_0^{-1})$, implying that $\Psi(d_0) - \Psi(2d_0 - 1) \to -\log(2) \approx -0.6931$ for $d_0 \to \infty$. It then follows that, for all $d_0 > 1$,

$$-\log(2) < \Psi(d_0) - \Psi(2d_0 - 1) < 0,$$

implying that $0 < 3\zeta_3\zeta_2^{-1} + (\Psi(d_0) - \Psi(2d_0 - 1)) < 3\zeta_3\zeta_2^{-1} - \log(2) \approx 1.4992$. Therefore, the leading bias term of $\hat{d}$ is smaller than the leading bias term of $\hat{d}_{\mu_0}$ for all $d_0 > 1$.

2.7 Simulation

In this section, we report the results of a Monte Carlo simulation of the small sample properties of the various estimators considered so far. In particular, we look at the CSS estimator of the memory parameter with μ_0 known and unknown, see (11) and (16) respectively, the MCSS estimator in (42), together with the bias-corrected version thereof in (62). We take as our DGP the model (6) with $\epsilon_t \sim NID(0, 1)$. Without loss of generality, we assume that $\mu_0 = 0$, since all estimators are invariant to the value of μ_0. In all settings covered by our experiment, we generate x_t for $T = 32, 64, 128, 256$. We let the long memory parameter d_0 vary. In particular, we set $d_0 = -0.2, -0.1, \ldots, 1.1, 1.2$. We compute the estimates using the optimising interval $d \in [d_0 - 5, d_0 + 5]$. All results are based on 10,000 replications[1]. We use the fractional difference algorithm of Jensen & Nielsen

[1] All computations in this paper are done using MATLAB 2019a, see MathWorks Inc. (2019). The convergence criteria used for numerical optimisation are the default ones. The code for replicating the main results in this paper is available on request.

(2014) to generate the fractionally integrated series, as well as to filter the fractionally integrated series in order to evaluate the objective function of the estimators.

Table 2 shows the Monte Carlo bias (multiplied by 100) of $\hat{d}$, $\hat{d}_{\mu_0}$, $\hat{d}_m$ and $\hat{d}_{bcm}$. We also report the percentage increase of the bias from $|\text{bias}(\hat{d}_m)|$ to $|\text{bias}(\hat{d})|$ by $\Delta\%|\text{bias}|$ in the last column for each T. We now summarise the main findings. The Monte Carlo biases of the estimators are almost everywhere in accordance with the theoretical counterparts in Theorem 2.5, see Table 1. Nevertheless, some obvious differences are noticeable for $T = 32$. In particular, the Monte Carlo biases of $\hat{d}$ in the stationary region are smaller than the theoretical approximations. The theoretical biases of $\hat{d}$ in the neighbourhood of $d_0 = 0.5$ also seem to differ from the Monte Carlo biases. The reason for this is that the theoretical bias of $\hat{d}$ in the stationary and non-stationary region diverges for the case when $d_0 = 0.5$. Recall that the situation of $d_0 = 0.5$ needs a separate analysis which is not covered in our analysis. As can be seen from Table 2, the bias in the CSS estimator is stronger in the stationary region than for the non-stationary region as is expected from Theorem 2.5. However, for the MCSS estimator, the bias in the stationary and the non-stationary region is of an identical order of magnitude, since its leading bias terms are $-(T\zeta_2)^{-1}\zeta_3\zeta_2^{-1}$, regardless of d_0, see Theorem 2.5. For $T = 256$, the bias of the CSS estimator compared to the MCSS estimator increases between 277% and 384% in the stationary region while the increase is between -14% and 131% in the non-stationary region. This increase is negative for $d_0 = 1.1$ and $d_0 = 1.2$ and implies that the bias of $\hat{d}$ is smaller than that of $\hat{d}_m$, as well as $\hat{d}_{\mu_0}$. Indeed, as discussed in Section 2.6, this result is in line with the theoretical biases in Theorem 2.5. Importantly, our simulation results confirm that we can remove the bias that occurs due to the estimation of the constant term by using the MCSS estimator. The leading bias of the MCSS estimator is the same as the CSS estimator with known μ_0, neglecting higher order terms, resulting in similar values of the simulated biases. Notice also that, the bias-corrected MCSS estimator in (62) performs best since the order of magnitude for the bias of this estimator is $o(T^{-1})$, see Corollary 2.3.

In Table 3, the Monte Carlo MSEs (multiplied by 100) are reported. The last column for each T reports the percentage increase of the MSE from $\text{MSE}(\hat{d}_m)$ to $\text{MSE}(\hat{d})$ by $\Delta\%\text{MSE}$. A few relevant observations can be made from this table. First, it can be seen that the MCSS estimator is favourable as compared to the CSS estimator. For example, for $T = 32$ the MSE of the CSS estimator when compared to the MCSS estimator is between 44% and 62% higher in the stationary region and increases between 7% and 60% in the non-stationary region. Furthermore, even for a moderately large sample of $T = 256$ the MCSS estimator improves substantially upon the CSS estimator in terms of the MSE. Second, the MSE of $\hat{d}_m$ and $\hat{d}_{bcm}$ are close to each other. Although the bias of $\hat{d}_{bcm}$ is smaller than the bias of $\hat{d}_m$, the additional improvement in the bias does not lead to large improvements in MSE. Third, $\text{MSE}(\hat{d}_{\mu_0})$ is smaller than $\text{MSE}(\hat{d}_m)$. Yet, the difference diminishes for T increasing. Since the leading bias terms are the same, the difference is explained by the larger finite sample variance of $\hat{d}_m$ relative to $\hat{d}_{\mu_0}$.

	bias($\hat{d}$)	bias($\hat{d}_{\mu_0}$)	bias($\hat{d}_m$)	bias($\hat{d}_{bcm}$)	Δ%|bias|	bias($\hat{d}$)	bias($\hat{d}_{\mu_0}$)	bias($\hat{d}_m$)	bias($\hat{d}_{bcm}$)	Δ%|bias|	bias($\hat{d}$)	bias($\hat{d}_{\mu_0}$)	bias($\hat{d}_m$)	bias($\hat{d}_{bcm}$)	Δ%|bias|	bias($\hat{d}$)	bias($\hat{d}_{\mu_0}$)	bias($\hat{d}_m$)	bias($\hat{d}_{bcm}$)	Δ%|bias|
$d_0 \setminus T$			32					64					128					256		
−0.2	−14.45	−3.46	−3.52	0.65	310.49	−7.06	−1.81	−1.75	0.33	303.28	−3.57	−0.91	−0.86	0.18	313.79	−1.79	−0.41	−0.37	0.15	383.30
−0.1	−14.58	−3.46	−3.55	0.61	310.65	−7.10	−1.81	−1.75	0.33	304.45	−3.58	−0.91	−0.86	0.18	315.06	−1.80	−0.41	−0.37	0.15	384.17
0.0	−14.63	−3.46	−3.59	0.57	307.26	−7.10	−1.81	−1.76	0.32	302.66	−3.58	−0.91	−0.87	0.17	313.29	−1.80	−0.41	−0.37	0.15	382.02
0.1	−14.53	−3.46	−3.64	0.53	299.17	−7.03	−1.81	−1.77	0.31	296.34	−3.55	−0.91	−0.87	0.17	306.60	−1.78	−0.41	−0.38	0.15	374.83
0.2	−14.21	−3.46	−3.70	0.47	284.41	−6.86	−1.81	−1.79	0.29	282.77	−3.46	−0.91	−0.88	0.16	292.39	−1.74	−0.41	−0.38	0.14	359.08
0.3	−13.59	−3.46	−3.78	0.38	259.51	−6.52	−1.81	−1.82	0.27	258.74	−3.29	−0.91	−0.89	0.15	267.35	−1.64	−0.41	−0.38	0.14	328.95
0.4	−12.66	−3.46	−3.88	0.29	226.60	−5.97	−1.81	−1.85	0.23	222.04	−2.98	−0.91	−0.91	0.13	228.25	−1.47	−0.41	−0.39	0.13	277.33
0.5	−11.35	−3.46	−3.98	0.19	185.54	−5.21	−1.81	−1.89	0.19	175.14	−2.53	−0.91	−0.92	0.12	175.91	−1.22	−0.41	−0.40	0.12	205.10
0.6	−9.78	−3.46	−4.02	0.15	143.34	−4.31	−1.81	−1.91	0.17	125.58	−2.03	−0.91	−0.92	0.12	119.75	−0.95	−0.41	−0.41	0.11	130.70
0.7	−8.07	−3.46	−3.97	0.19	103.18	−3.45	−1.81	−1.91	0.17	80.79	−1.58	−0.91	−0.92	0.12	71.61	−0.72	−0.41	−0.42	0.10	73.36
0.8	−6.48	−3.46	−3.85	0.31	68.06	−2.76	−1.81	−1.89	0.19	45.95	−1.26	−0.91	−0.92	0.12	37.42	−0.57	−0.41	−0.42	0.10	36.58
0.9	−5.19	−3.46	−3.74	0.43	38.79	−2.27	−1.81	−1.87	0.22	21.62	−1.06	−0.91	−0.92	0.12	15.66	−0.47	−0.41	−0.41	0.11	14.51
1.0	−4.27	−3.46	−3.66	0.51	16.87	−1.95	−1.81	−1.85	0.23	5.54	−0.94	−0.91	−0.92	0.12	2.21	−0.42	−0.41	−0.41	0.11	1.05
1.1	−3.65	−3.46	−3.60	0.56	1.39	−1.75	−1.81	−1.84	0.24	−4.92	−0.86	−0.91	−0.92	0.12	−6.37	−0.38	−0.41	−0.41	0.11	−7.62
1.2	−3.23	−3.46	−3.56	0.60	−9.28	−1.61	−1.81	−1.83	0.25	−11.92	−0.81	−0.91	−0.92	0.12	−12.16	−0.35	−0.41	−0.41	0.11	−13.57

Table 2: (100 $\times$) Monte Carlo bias of the CSS estimator of d with unknown and known μ_0 and the MCSS estimator of d, together with the bias-corrected MCSS estimator of d. Δ%|bias| refers to a percentage increase from $|\text{bias}(\hat{d}_m)|$ to $|\text{bias}(\hat{d})|$.

	MSE($\hat{d}$)	MSE($\hat{d}_{\mu_0}$)	MSE($\hat{d}_m$)	MSE($\hat{d}_{bcm}$)	Δ%MSE	MSE($\hat{d}$)	MSE($\hat{d}_{\mu_0}$)	MSE($\hat{d}_m$)	MSE($\hat{d}_{bcm}$)	Δ%MSE	MSE($\hat{d}$)	MSE($\hat{d}_{\mu_0}$)	MSE($\hat{d}_m$)	MSE($\hat{d}_{bcm}$)	Δ%MSE	MSE($\hat{d}$)	MSE($\hat{d}_{\mu_0}$)	MSE($\hat{d}_m$)	MSE($\hat{d}_{bcm}$)	Δ%MSE
$d_0 \setminus T$			32					64					128					256		
−0.2	6.33	2.79	4.39	4.27	44.18	2.04	1.18	1.58	1.55	29.11	0.77	0.54	0.65	0.64	18.43	0.32	0.26	0.29	0.29	10.53
−0.1	6.44	2.79	4.40	4.28	46.43	2.06	1.18	1.58	1.55	30.04	0.77	0.54	0.65	0.64	18.85	0.32	0.26	0.29	0.29	10.72
0.0	6.54	2.79	4.39	4.26	49.09	2.07	1.18	1.58	1.55	31.12	0.78	0.54	0.65	0.64	19.32	0.32	0.26	0.29	0.29	10.92
0.1	6.62	2.79	4.35	4.22	52.18	2.08	1.18	1.57	1.54	32.39	0.78	0.54	0.65	0.64	19.86	0.32	0.26	0.29	0.29	11.14
0.2	6.64	2.79	4.27	4.13	55.53	2.08	1.18	1.55	1.52	33.90	0.78	0.54	0.64	0.64	20.47	0.32	0.26	0.29	0.29	11.39
0.3	6.60	2.79	4.14	4.00	59.28	2.06	1.18	1.52	1.49	35.46	0.77	0.54	0.64	0.63	21.01	0.32	0.26	0.29	0.29	11.62
0.4	6.47	2.79	3.99	3.84	62.24	2.01	1.18	1.48	1.44	36.37	0.76	0.54	0.63	0.62	21.05	0.32	0.26	0.29	0.28	11.60
0.5	6.22	2.79	3.82	3.66	62.68	1.92	1.18	1.42	1.39	35.25	0.73	0.54	0.61	0.60	19.63	0.31	0.26	0.28	0.28	10.65
0.6	5.77	2.79	3.62	3.46	59.24	1.78	1.18	1.36	1.33	30.79	0.68	0.54	0.59	0.58	16.01	0.30	0.26	0.27	0.27	8.28
0.7	5.18	2.79	3.42	3.26	51.47	1.62	1.18	1.31	1.27	23.70	0.63	0.54	0.57	0.56	11.15	0.28	0.26	0.27	0.27	5.35
0.8	4.54	2.79	3.22	3.08	40.77	1.47	1.18	1.27	1.23	16.20	0.59	0.54	0.56	0.55	6.84	0.27	0.26	0.27	0.26	3.06
0.9	3.99	2.79	3.09	2.95	29.21	1.35	1.18	1.23	1.20	9.99	0.57	0.54	0.55	0.54	3.91	0.27	0.26	0.26	0.26	1.68
1.0	3.58	2.79	3.00	2.87	19.55	1.29	1.18	1.21	1.18	5.80	0.56	0.54	0.55	0.54	2.17	0.26	0.26	0.26	0.26	0.93
1.1	3.31	2.79	2.95	2.82	12.37	1.24	1.18	1.20	1.17	3.23	0.55	0.54	0.54	0.54	1.17	0.26	0.26	0.26	0.26	0.51
1.2	3.13	2.79	2.91	2.79	7.51	1.22	1.18	1.20	1.17	1.71	0.55	0.54	0.54	0.53	0.58	0.26	0.26	0.26	0.26	0.26

Table 3: (100 $\times$) Monte Carlo MSE of the CSS estimator of d with unknown and known μ_0 and the MCSS estimator of d, together with the bias-corrected MCSS estimator of d. Δ%MSE refers to a percentage increase from MSE($\hat{d}_m$) to MSE($\hat{d}$).

3 Generalisation

We now consider an extension of the analysis to the case where the short-run dynamics are allowed to have a more general structure than the simple IID shocks assumed so far in model (6). Theoretical aspects are presented in Section 3.1, including the derivation of the MCSS estimator and of the analytical bias expression for the general model. Following that, Section 3.2 focuses on obtaining analytic expressions for the asymptotic biases of specific models. This encompasses bias expressions for the ARFIMA(1,d,0) model in Section 3.2.1, and bias expressions for short-memory models in Section 3.2.2. The analysis concludes with a simulation study presented in Section 3.3.

3.1 Asymptotic biases

The extended model is given by

$$x_t = \mu I(t \geq 1) + \Delta_+^{-d} u_t, \tag{63}$$

$$u_t = \omega(L;\varphi)\epsilon_t. \tag{64}$$

where $t = 0, \pm 1, \pm 2, \ldots$. The lag polynomial ω captures the short-run dependence structure parametrically and is given by

$$\omega(L;\varphi) = \sum_{j=0}^{\infty} \omega_j(\varphi) L^j, \tag{65}$$

where φ is an unknown $p \times 1$ vector and $\omega_0(\varphi) = 1$, $|\omega(s;\varphi)| \neq 0$ for $|s| \leq 1$, and $\sum_{j=0}^{\infty} |\omega_j(\varphi)| < \infty$. More precise conditions on ω will be specified below. The representation of u_t in (64) as a MA(∞) model in (65) is common in the literature and considered by, among others, Hualde & Robinson (2011) and Hualde & Nielsen (2020, 2022). One well-known special case of u_t in (64) is an ARMA(p_1, p_2) model which is given by

$$\omega(L;\varphi) = \frac{\alpha(L;\varphi)}{\beta(L;\varphi)}, \tag{66}$$

where $\beta(L;\varphi)$ is the AR polynomial of order p_1 and $\alpha(L;\varphi)$ is the MA polynomial of order p_2. It is assumed that the polynomials do not have common roots and that their roots lie outside the unit circle. Then (63), (64) and (66) is an ARFIMA(p_1, d, p_2) model. Another special case of (65) is Bloomfield's (1973) exponential spectrum model, see Robinson (1994) and Hassler (2019).

Following Hualde & Nielsen (2020), we make the following assumptions. We use the notation $\vartheta = (d, \varphi')'$, with true value denoted again by subscript 0, i.e. $\vartheta_0 = (d_0, \varphi_0')'$.

Assumption 3.1. *The errors ϵ_t are IID(0,σ_0^2) with finite fourth moments.*

Assumption 3.2. *The parameter space for $\vartheta = (d, \varphi')'$ is given by $\Theta = [\nabla_1, \nabla_2] \times \Phi$, with $-\infty < \nabla_1 < \nabla_2 < \infty$ and Φ being a compact and convex subset of $\mathbb{R}^p$. The true value $\vartheta_0 = (d_0, \varphi_0')' \in \Theta$ with d_0 not equal to 1/2. The parameter space for μ is $\mathbb{R}$.*

Assumption 3.3. *(i) For all $\varphi \in \Phi \backslash \{\varphi_0\}$, $|\omega(s;\varphi)| \neq |\omega(s;\varphi_0)|$ on a set $S \subset \{s : |s| = 1\}$ of positive Lebesgue measure.*

(ii) *For all* $\varphi \in \Phi$, $\omega(e^{i\lambda};\varphi)$ *is differentiable in* λ *with derivative in Lip*(ς) *for* $1/2 < \varsigma \leq 1$.

(iii) *For all* λ, $\omega(e^{i\lambda};\varphi)$ *is continuous in* φ.

(iv) *For all* $\varphi \in \Phi$, $|\omega(s;\varphi)| \neq 0$, $|s| \leq 1$.

(v) *The true value* ϑ_0 *is in the interior of* Θ.

(vi) *For all* λ, $\omega(e^{i\lambda};\varphi)$ *is thrice continuously differentiable in* φ *in a closed neighbourhood* $\mathcal{N}_\varepsilon(\varphi_0)$ *of radius* $\varepsilon \in (0,1/2)$ *about* φ_0. *For all* $\varphi \in \mathcal{N}_\varepsilon(\varphi_0)$ *these partial derivatives with respect to* φ *are themselves differentiable in* λ *with derivative in Lip*(ς) *for* $1/2 < \varsigma \leq 1$.

(vii) *The matrix*

$$A = \begin{pmatrix} \pi^2/6 & -\sum_{j=1}^{\infty} b_{\varphi' j}(\varphi_0)/j \\ -\sum_{j=1}^{\infty} b_{\varphi j}(\varphi_0)/j & \sum_{j=1}^{\infty} b_{\varphi j}(\varphi_0) b_{\varphi' j}(\varphi_0) \end{pmatrix} \tag{67}$$

is nonsingular, where

$$b_{\varphi j}(\varphi_0) = \sum_{k=0}^{j-1} \omega_k(\varphi_0) \partial \phi_{j-k}(\varphi_0)/\partial\varphi \tag{68}$$

and where ϕ_j *is defined by*

$$\phi(s;\varphi) = \omega^{-1}(s;\varphi) = \sum_{j=0}^{\infty} \phi_j(\varphi) s^j. \tag{69}$$

Assumption 3.1 is the same as Assumption 2.1. In fact, the IID assumption can be weakened to martingale difference series as in Hualde & Nielsen (2020, 2022) but for the sake of convenience we keep this condition simple. Assumption 3.2 extends Assumption 2.2 by including the parameter space of φ. Assumption 3.3(i)-(iv), which ensures the identification of short-term dynamics, is standard in the literature on parametric short-memory models since its introduction by Hannan (1973). Assumption 3.3(v)-(vii) serve as additional regulatory conditions necessary to establish the asymptotic distribution theory. We refer to the papers of Hualde & Robinson (2011), Nielsen (2015), Hualde & Nielsen (2020, 2022) for a discussion. Importantly, Assumption 3.3 is satisfied for the stationary and invertible ARMA model and also the exponential spectrum model of Bloomfield (1973).

For any (d,φ,μ) in the admissible parameter space, define the residuals $\epsilon_t(d,\varphi,\mu) = \phi(L;\varphi)\Delta_+^d(x_t - \mu)$. The CSS objective function is then given by

$$\begin{aligned} L(d,\varphi,\mu) &= \frac{1}{2}\sum_{t=1}^{T} \epsilon_t^2(d,\varphi,\mu), \\ &= \frac{1}{2}\sum_{t=1}^{T} \Big(\phi(L;\varphi)\Delta_+^d x_t - \mu c_t(d,\varphi)\Big)^2, \end{aligned} \tag{70}$$

where we define the convoluted coefficient

$$c_t(d,\varphi) = \phi(L;\varphi)\Delta_+^d I(t \geq 1) = \sum_{j=0}^{t-1} \phi_j(\varphi)\kappa_{0(t-j)}(d), \tag{71}$$

with $\kappa_{0t}(d)$ given in (8). Since (70) is quadratic in μ we can concentrate it. Differentiating with respect to μ and solving yields

$$\hat{\mu}(\vartheta) = \frac{\sum_{t=1}^{T}(\phi(L;\varphi)\Delta_+^d x_t)c_t(d,\varphi)}{\sum_{t=1}^{T} c_t^2(d,\varphi)}. \tag{72}$$

with the profile CSS function

$$L^*(\vartheta) = \frac{1}{2}\sum_{t=1}^{T}\left(\phi(L;\varphi)\Delta_+^d x_t - \hat{\mu}(d,\varphi)c_t(d,\varphi)\right)^2. \tag{73}$$

Note that, as in Section 2, we use asterisks to emphasise that we are dealing with a *profile* objective function. The resulting CSS estimator of $\vartheta = (d,\varphi')'$ is given by

$$\hat{\vartheta} = \operatorname*{argmin}_{\vartheta\in\Theta} L^*(\vartheta). \tag{74}$$

Hualde & Nielsen (2020) show that if x_t is generated by (63)-(65) and if Assumptions 3.1 to 3.3 hold, then, as $T \to \infty$,

$$\hat{\vartheta} \xrightarrow{p} \vartheta_0, \tag{75}$$

and

$$\sqrt{T}(\hat{\vartheta} - \vartheta_0) \xrightarrow{d} N(0_{p+1}, A^{-1}). \tag{76}$$

where A is given in (67).

For comparison, we also analyse the situation where the true μ_0 is known, as in Section 2.2. The CSS estimator for this model can be derived by substituting μ_0 into (70) to have

$$L^*_{\mu_0}(\vartheta) = \frac{1}{2}\sum_{t=1}^{T}\left(\phi(L;\varphi)\Delta_+^d x_t - \mu_0 c_t(d,\varphi)\right)^2, \tag{77}$$

such that

$$\hat{\vartheta}_{\mu_0} = \operatorname*{argmin}_{\vartheta\in\Theta} L^*_{\mu_0}(\vartheta). \tag{78}$$

Hualde & Robinson (2011) show that if x_t is generated by (63)-(65) and if Assumptions 3.1 to 3.3 hold, then this estimator is consistent, too, and attains the same limiting distribution as $\hat{\vartheta}$ in (74), see (76).

The CSS objectives in (73) and (77) may again be motivated in terms of a Gaussian likelihood, see Hualde & Nielsen (2020). Therefore, as in Section 2.3, it will be instructive to interpret the CSS objectives in (73) and (77) as profile likelihood functions. As such, they are not genuine likelihoods, for they are not based on observable quantities. Instead, they are functions of the maximum likelihood estimators of μ and σ^2 which are treated as if they were the true parameter values. This shows that profile likelihoods do not necessarily possess the same properties as genuine likelihoods. Accordingly, we show below that the score function of $L^*(d,\varphi)$ in (73) evaluated at the true parameters is biased. The score function of $L^*_{\mu_0}(d,\varphi)$ in (77) is unbiased.

For simplicity, we will assume in the sequel that the initial observations of the short-run dynamics are equal to 0 for $t \leq 0$. Alternatively, an argument of asymptotic negligibility of pre-sample observations as in Hualde & Robinson (2011, Lemma 2) could be made.

Assumption 3.4. *For all* $t \leq 0$*, we assume that* $\epsilon_t = 0$ *in* (64)*.*

The following theorem generalises Theorem 2.1 and Theorem 2.2. The proof will be given in Appendix A.4. Note that we use the notation $D_i f(d,\varphi)$ and $D_{ij} f(d,\varphi)$ to denote, respectively, the first derivative and second derivative of a function $f(d,\varphi)$ with respect to parameters $i,j \in \{d,\varphi\}$.

Theorem 3.1. *Let* x_t*,* $t = 1,\ldots,T$*, be given by* (63)-(65) *and let Assumptions 3.1 to 3.4 be satisfied. Then, the expected scores of* $L^*(d,\varphi)$*, evaluated at the true parameter* d_0 *and* φ_0*, are given by*

$$E(D_d L^*(d_0,\varphi_0)) = O(\log(T) I(d_0 < 1/2) + I(d_0 > 1/2)), \tag{79}$$

$$E(D_\varphi L^*(d_0,\varphi_0)) = O(1), \tag{80}$$

when $T \to \infty$*. The expected scores of* $L^*_{\mu_0}(d,\varphi)$*, evaluated at the true parameter* d_0 *and* φ_0*, are given by*

$$E(D_d L^*_{\mu_0}(d_0,\varphi_0)) = 0, \tag{81}$$

$$E(D_\varphi L^*_{\mu_0}(d_0,\varphi_0)) = 0_p. \tag{82}$$

The expectation of the score in (79), i.e. the score with respect to d, is not uniform in d_0. For $d_0 < 1/2$ it diverges at the rate of $\log(T)$, while it is $O(1)$ for $d_0 > 1/2$. This is not true for the expectation of the score in (80), i.e. the score with respect to φ, which is $O(1)$ uniformly in d_0. The rationale behind this is that the score bias measures the relative strength of the level parameter and the stochastic component. The score bias with respect to d gauges the strength of the level parameter relative to the fractional dynamics, whereas the score bias with respect to φ evaluates the strength of the level parameter in relation to short-run dynamics. Recall that in the pure fractional model the bias of the CSS estimator is a function of the score bias scaled by T, see (55) in Section 2.6. If an analogous relationship were to hold in the present setting one might be tempted to think, from (79) and (80), that in the stationary region the bias of $\hat{d}$ will be of a larger order of magnitude than the bias of $\hat{\varphi}$. Yet this turns out not to be true. As will be shown below, the biases of $\hat{d}$ and $\hat{\varphi}$ are functions of not only their own score biases but, instead, of a weighted sum of both score biases. This will lead to the bias of the short-run dynamics to behave the same as the order of the bias of the memory parameter. The expectation of the score of $L^*_{\mu_0}(d,\varphi)$ is equal to zero, see (81) and (82).

The discussion above again highlights the need for a modification of the CSS profile likelihood $L^*(d_0,\varphi_0)$ such that it behaves more like a genuine likelihood or, equivalently, more like $L^*_{\mu_0}(d_0, \varphi_0)$. Following the same arguments as in Section 2, we consider a multiplicative modification term for the profile CSS objective function in (73). We therefore define again the modified profile CSS objective function as

$$L^*_m(\vartheta) = m(\vartheta)L^*(\vartheta), \tag{83}$$

where the multiplicative modification term $m(\vartheta) > 0$ depends only on ϑ. The corresponding $(p+1) \times 1$ vector of scores is the first derivative of (83):

$$D_\vartheta L^*_m(\vartheta) = m(\vartheta)D_\vartheta L^*(\vartheta) + D_\vartheta m(\vartheta)L^*(\vartheta). \tag{84}$$

We again require that our objective function is score unbiased, i.e. that the score functions in (84) satisfy

$$E\left(D_\vartheta L^*_m(\vartheta_0)\right) = 0_{p+1}, \tag{85}$$

cf. (38). It is important to stress that (85) requires all the scores to be unbiased. As was alluded to above and will be shown below, this is due to the fact that the bias of the CSS estimator depends on the biases of all scores and not of their own scores only. Using (84) and the fact that $D_\vartheta \log(m(\vartheta)) = D_\vartheta m(\vartheta)/m(\vartheta)$ and $m(\vartheta) > 0$ it follows that (85) is equivalent to the condition that

$$D_\vartheta \log\left(m(\vartheta_0)\right) = -\frac{E\left(D_\vartheta L^*(\vartheta_0)\right)}{E\left(L^*(\vartheta_0)\right)}. \tag{86}$$

Evaluating the right-hand side of (86) yields

$$D_\vartheta \log\left(m(\vartheta_0)\right) = \frac{2}{T-1}\frac{\sum_{t=1}^T c_t(\vartheta_0)D_\vartheta c_t(\vartheta_0)}{\sum_{t=1}^T c_t^2(\vartheta_0)}, \tag{87}$$

see Lemma A.13. As can be easily seen, the primitive function of (87) is

$$\log\left(m(\vartheta)\right) = \log\left(\sum_{t=1}^T c_t^2(\vartheta)\right)^{\frac{1}{T-1}}.$$

Finally, exponentiation yields

$$m(\vartheta) = \left(\sum_{t=1}^T c_t^2(\vartheta)\right)^{\frac{1}{T-1}}, \tag{88}$$

cf. (41). The modified profile CSS objective function in (83) is thus given by the product of $m(\vartheta)$ in (88) and $L^*(\vartheta)$ in (73), i.e.

$$L^*_m(\vartheta) = \left(\sum_{t=1}^T c_t^2(d,\varphi)\right)^{\frac{1}{T-1}} \frac{1}{2}\sum_{t=1}^T \Big(\phi(L;\varphi)\Delta_+^d x_t - \hat{\mu}(d,\varphi)c_t(d,\varphi)\Big)^2.$$

We again call the estimator that minimises $L^*_m(\vartheta)$ the modified conditional sum-of-squares (MCSS) estimator, as in Section 2.4, and denote the estimator of ϑ by $\hat{\vartheta}_m$, i.e.

$$\hat{\vartheta}_m = \underset{\vartheta\in\Theta}{\operatorname{argmin}}\, L^*_m(\vartheta). \tag{89}$$

The following lemma shows that the two properties of the modification term $m(d)$ in Lemma 2.1 carry over, to $m(\vartheta)$ in (88). As such, the modification term $m(\vartheta)$ acts again as penalisation through inflating the CSS objective function $L^*(\vartheta)$ by the factor $m(\vartheta)$. See Appendix A.4.2 for the proof.

Lemma 3.1. *For all $d \in \mathbb{R}$ and $\varphi \in \Phi$,*

$$m(\vartheta) \geq 1. \tag{90}$$

Here, equality holds if $d = 1$ and $\varphi = 0$. Also, it holds that, for $T \to \infty$,

$$m(\vartheta) = 1 + O(T^{-1}\log(T)I(d < 1/2) + T^{-1}I(d > 1/2)) \tag{91}$$

for all $d \in \mathbb{R}\backslash\{1/2\}$.

The asymptotic properties of the MCSS estimator are the same as of the CSS estimator and are summarised for completeness by the following theorem. The proof is given in Appendix A.4. The result follows from the fact that the modification term $m(\vartheta) \to 1$, see (91) in Lemma 3.1.

Theorem 3.2. *Let x_t, $t = 1,\ldots,T$, be given by* (63)-(65) *and let Assumptions 3.1 to 3.3 be satisfied. Then, as $T \to \infty$,*

$$\hat{\vartheta}_m \xrightarrow{p} \vartheta_0, \tag{92}$$

and

$$\sqrt{T}(\hat{\vartheta}_m - \vartheta_0) \xrightarrow{d} N(0_{p+1}, A^{-1}), \tag{93}$$

where A is given in (67)*.*

However, the asymptotic bias of the estimator $\hat{\vartheta}_m$ behaves differently from that of $\hat{\vartheta}$. In order to find the asymptotic biases, we proceed in a similar fashion as in Section 2.6, involving two steps: finding the asymptotic expansion of $\hat{\vartheta}$ and approximating this expansion. The bottom line of the derivation is that

$$E\left(\hat{\vartheta} - \vartheta_0\right) = S_T(d_0, \varphi_0) + B_T(\varphi_0) + o(T^{-1}), \tag{94}$$

where we call $S_T(d_0, \varphi_0) = -A^{-1}T^{-1}\left[\sigma_0^{-2}E\left(D_\vartheta L^*(\vartheta_0)\right)\right]$ the score bias and $B_T(\varphi_0)$ the intrinsic bias. Detailed expressions are given in Appendix A.4.

Importantly, we refer to $S_T(d_0, \varphi_0)$ and $B_T(\varphi_0)$ as "exact" biases, as we evaluate the expectations terms without taking the limits. The asymptotic counterparts, for $T \to \infty$ with appropriate scaling, are denoted by $S(d_0, \varphi_0)$ and $B(\varphi_0)$. We refer to these asymptotic biases as "approximate". Lieberman & Phillips (2005) also make a comparable dichotomy concerning the Edgeworth expansions of the memory parameter. The rationale behind emphasising the distinction lies in the exact and approximate biases being potentially very different from each other. This is because the number of quantities in $B_T(\varphi_0)$ that are approximated by $B(\varphi_0)$ amounts to a maximum of $3 \times (p+1)^3 + (p+1)^2$. If T is relatively small, the difference between $B_T(\varphi_0)$ and $B(\varphi_0)$ can therefore be substantial.

It follows from Lemmata A.18, A.19, A.24 and A.25 that analogues of (94) also hold for $\hat{\vartheta}_{\mu_0}$ and $\hat{\vartheta}_m$. Replacing $D_\vartheta L^*(\vartheta_0)$ by $D_\vartheta L^*_{\mu_0}(\vartheta_0)$, it is clear that the expected score term gets eliminated. Similarly, $E(D_\vartheta L^*_m(\vartheta_0)) = 0$ by construction, justifying the MCSS objective function in order to obtain score unbiasedness in all parameters. Indeed, the leading bias of $\hat{\vartheta}_m$ is the same as that of $\hat{\vartheta}_{\mu_0}$. The following theorem is the main result of this paper and presents the approximate bias of $\hat{\vartheta}$, $\hat{\vartheta}_{\mu_0}$ and $\hat{\vartheta}_m$. The proof is given in Appendix A.4.

Theorem 3.3. *Let x_t, $t = 1,\ldots,T$, be given by* (63)-(65) *and let Assumptions 3.1 to 3.4 be satisfied. The approximate biases of $\hat{\vartheta}$, $\hat{\vartheta}_{\mu_0}$ and $\hat{\vartheta}_m$ are*

$$bias(\hat{\vartheta}) = S(d_0, \varphi_0) + B(\varphi_0) + o(T^{-1}), \tag{95}$$

$$bias(\hat{\vartheta}_{\mu_0}) = B(\varphi_0) + o(T^{-1}), \tag{96}$$

$$bias(\hat{\vartheta}_m) = B(\varphi_0) + o(T^{-1}), \tag{97}$$

where the intrinsic bias $B(\varphi_0)$ is given in (A.131) *of Appendix A.4. The score bias for $d_0 > 1/2$ is given by*

$$TS(d_0, \varphi_0) = A^{-1} \frac{\sum_{t=1}^{\infty} c_t(\vartheta) c_{\vartheta t}(\vartheta)}{\sum_{t=1}^{\infty} c_t^2(\vartheta)}$$

while the score bias for $d_0 < 1/2$ is given by

$$TS(d_0, \varphi_0) = A^{-1} \begin{bmatrix} -\log(T) + (\Psi(1-d_0) + (1-2d_0)^{-1}) \\ \frac{D_{\varphi_1}\phi(1;\varphi)}{\phi(1;\varphi)} \\ \vdots \\ \frac{D_{\varphi_p}\phi(1;\varphi)}{\phi(1;\varphi)} \end{bmatrix}$$

where A is given in (67)*, $c_t(\vartheta)$ in* (71) *and $\phi(1;\varphi)$ in* (69)*. Furthermore, the intrinsic bias $B(\varphi_0) = O(T^{-1})$, whereas the score bias $S(d_0, \varphi_0) = O(T^{-1}\log(T))$ in the stationary region and $S(d_0, \varphi_0) = O(T^{-1})$ in the non-stationary region.*

The approximate biases reveal two key points. Firstly, the bias in the estimators is again composed of two terms: the score bias and the intrinsic bias. As outlined earlier in the discussion below Theorem 3.1, the score bias measures the relative strength of the deterministic component when contrasted with the stochastic component. The intrinsic bias is what remains even if the true value of the deterministic component μ_0 were known. The score bias term will dominate as the leading bias term in the stationary region. Secondly, the value of the memory parameter solely affects the bias through the score bias, not through the intrinsic bias, the latter merely depending on the short-run dynamics φ_0.

As discussed in Section 2.6, to quantify the accuracy of test statistics or confidence intervals for $\hat{\vartheta}_0$, the relevant quantity is the relative bias, i.e. the ratio of asymptotic bias to the asymptotic standard deviation. Theorem 3.3 implies that the relative bias for the three estimators is of order $O(T^{-1/2})$ in the non-stationary region. In the stationary region, the relative bias is of order $O(T^{-1/2}\log(T))$ for the CSS estimator with unknown μ_0, while it is of order $O(T^{-1/2})$ for the CSS estimator with known μ_0 and the MCSS estimator. Thus, especially in the stationary region, the t-test for the memory parameter as well as the short-run dynamics are more accurate when using the MCSS estimator compared to the CSS estimator. Later in the empirical study, we will exploit this feature to our advantage.

3.2 Special cases

In this section, we find analytic expressions of the asymptotic biases for specific models. Section 3.2.1 presents bias expressions for the ARFIMA(1,d,0) model, while Section 3.2.2

covers bias expressions for short-memory models, concluding with the biases of the AR(1) model.

3.2.1 ARFIMA(1,d,0) model

Theorem 3.3 enables us to express the bias of particular models explicitly. One such model is the ARFIMA(1,d,0), commonly used for modelling fractional integrated time series. In a Monte Carlo simulation performed by Nielsen & Frederiksen (2005), this specific model underwent thorough analysis. This section will provide an explanation of their findings. In the following theorem, we will present the approximate bias expression for this particular model.

Theorem 3.4. *Let x_t, $t = 1,\ldots,T$, be given by (63) and let $u_t = \varphi u_{t-1} + \epsilon_t$. Let Assumptions 3.1 to 3.4 be satisfied. The approximate biases of $\hat{\vartheta}$, $\hat{\vartheta}_{\mu_0}$ and $\hat{\vartheta}_m$ are*

$$bias(\hat{\vartheta}) = S(d_0, \varphi_0) + B(\varphi_0) + o(T^{-1}), \tag{98}$$

$$bias(\hat{\vartheta}_{\mu_0}) = B(\varphi_0) + o(T^{-1}), \tag{99}$$

$$bias(\hat{\vartheta}_m) = B(\varphi_0) + o(T^{-1}), \tag{100}$$

where the intrinsic bias $B(\varphi_0)$ is given by

$$TB(\varphi) = A^{-1}\begin{bmatrix} \iota'\left(A^{-1}\odot(G_1+F_1)\right)\iota \\ \iota'\left(A^{-1}\odot(G_2+F_2)\right)\iota \end{bmatrix} - \frac{1}{2}A^{-1}\begin{bmatrix} \iota'\left((A^{-1}C_{01}A^{-1})\odot A\right)\iota \\ \iota'\left((A^{-1}C_{02}A^{-1})\odot A\right)\iota \end{bmatrix}$$

$$A = \begin{bmatrix} \pi^2/6 & -\varphi^{-1}\log(1-\varphi) \\ -\varphi^{-1}\log(1-\varphi) & (1-\varphi^2)^{-1} \end{bmatrix}$$

$$C_{01} = \begin{bmatrix} -6\zeta_3 & 2\varphi^{-1}Li_2(-\frac{\varphi}{1-\varphi}) - \varphi^{-1}\log^2(1-\varphi) \\ 2\varphi^{-1}Li_2(-\frac{\varphi}{1-\varphi}) - \varphi^{-1}\log^2(1-\varphi) & 2\frac{\log(1-\varphi)}{1-\varphi^2} \end{bmatrix}$$

$$C_{02} = \begin{bmatrix} 2\varphi^{-1}Li_2(-\frac{\varphi}{1-\varphi}) - \varphi^{-1}\log^2(1-\varphi) & 2\frac{\log(1-\varphi)}{1-\varphi^2} \\ 2\frac{\log(1-\varphi)}{1-\varphi^2} & 0 \end{bmatrix}$$

$$F_1 = \begin{bmatrix} -2\zeta_3 & -\varphi^{-1}\log^2(1-\varphi) \\ \varphi^{-1}Li_2(-\frac{\varphi}{1-\varphi}) & \frac{\log(1-\varphi)}{1-\varphi^2} \end{bmatrix}$$

$$F_2 = \begin{bmatrix} \varphi^{-1}Li_2(-\frac{\varphi}{1-\varphi}) & \frac{\log(1-\varphi)}{1-\varphi^2} \\ 0 & 0 \end{bmatrix}$$

$$G_1 = \begin{bmatrix} -4\zeta_3 & 2\varphi^{-1}Li_2(-\frac{\varphi}{1-\varphi}) \\ -\varphi^{-1}\log^2(1-\varphi) + \varphi^{-1}Li_2(-\frac{\varphi}{1-\varphi}) & \frac{\log(1-\varphi)}{1-\varphi^2} - \varphi^{-2}\left(\frac{\varphi}{1-\varphi} + \log(1-\varphi)\right) \end{bmatrix}$$

$$G_2 = \begin{bmatrix} -\varphi^{-1}\log^2(1-\varphi) + \varphi^{-1}Li_2(-\frac{\varphi}{1-\varphi}) & \frac{\log(1-\varphi)}{1-\varphi^2} - \varphi^{-2}\left(\frac{\varphi}{1-\varphi} + \log(1-\varphi)\right) \\ 2\log(1-\varphi)\frac{1}{1-\varphi^2} & -2\frac{\varphi}{(1-\varphi^2)^2} \end{bmatrix}.$$

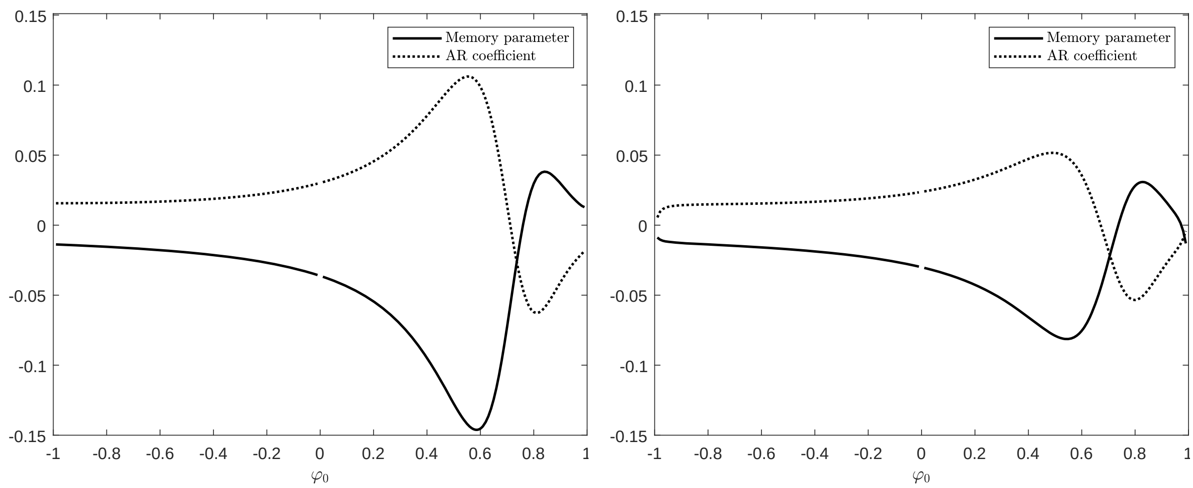
The score bias $S(d_0, \varphi_0)$ for $d_0 > 1/2$ is given by

$$TS(d,\varphi) = A^{-1}\left[(1-\varphi)^2\binom{2d-2}{d-1} + \varphi\binom{2d}{d}\right]^{-1} \times$$
$$\begin{bmatrix} (1-\varphi)^2\binom{2d-2}{d-1}\left(\Psi(2d-1)-\Psi(d)\right) + \varphi\binom{2d}{d}\left(\Psi(2d+1)-\Psi(d+1)\right) \\ (\varphi-1)\binom{2d-2}{d-1} + 0.5\binom{2d}{d} \end{bmatrix}$$

while the score bias $S(d_0, \varphi_0)$ for $d_0 < 1/2$ is given by

$$TS(d, \varphi) = A^{-1} \begin{bmatrix} -\log(T) + \Psi(1 - d_0) + (1 - 2d_0)^{-1} \\ -\frac{1}{1-\varphi} \end{bmatrix} \tag{101}$$

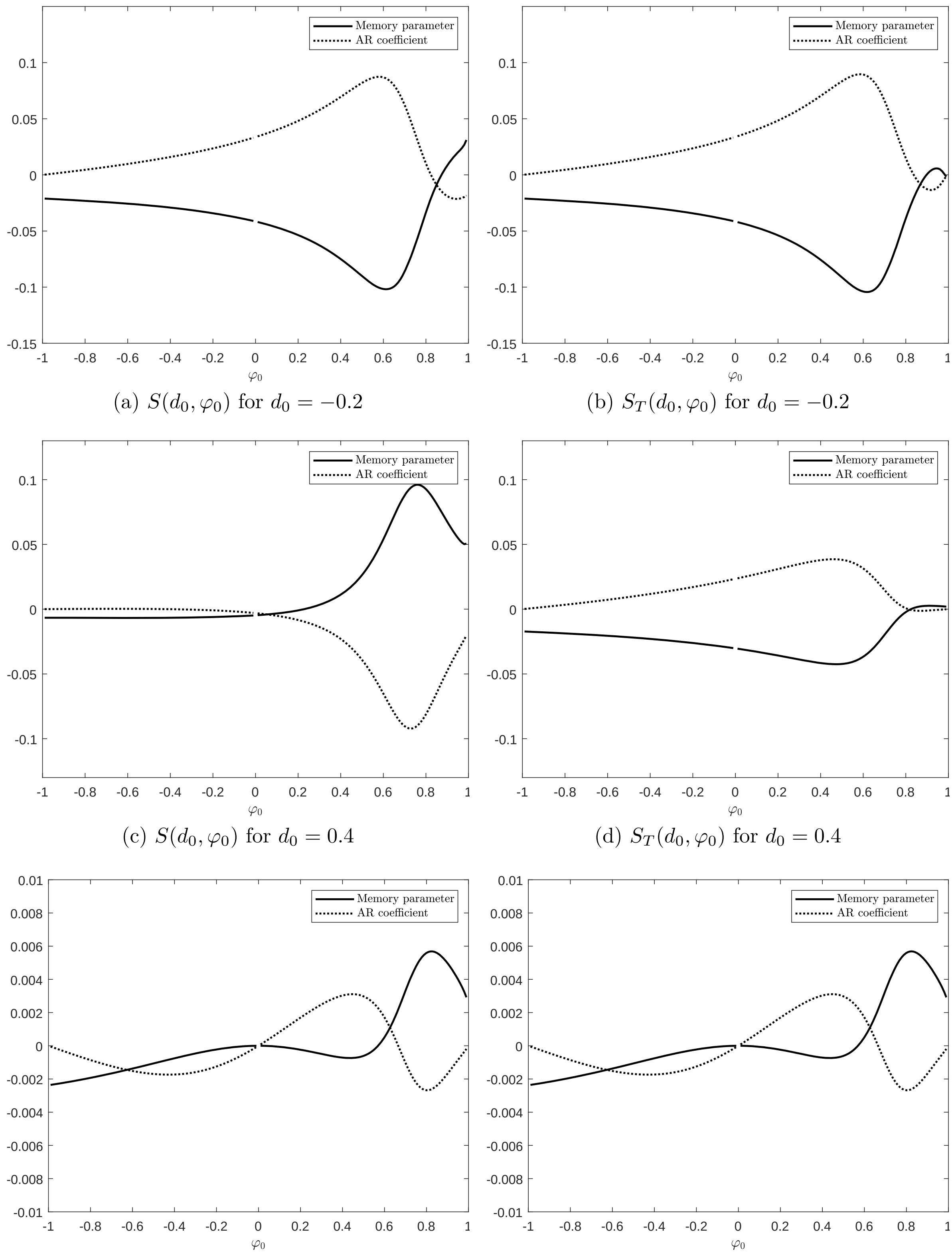
where ζ_s is the Riemann's zeta function $\zeta_s = \sum_{j=1}^{\infty} j^{-s}$, $s > 1$, and $\Psi(d) = D \log \Gamma(d)$ denotes the Digamma function and $Li_2(\varphi) = \sum_{i=1}^{\infty} i^{-2} \varphi^i$ is the dilogarithm function (Spence's integral). The binomial coefficients are represented using the notation $\binom{\cdot}{\cdot}$.



(a) $B(\varphi_0)$ (b) $B_T(\varphi_0)$

Figure 2: The approximate and exact intrinsic bias for the ARFIMA(1,d,0) model with $T = 128$ in panel (a) and (b), respectively.

Figure 2(a) shows the approximate intrinsic bias of both the memory parameter and the autoregressive coefficient. Interestingly, the biases for these two parameters exhibit near-symmetry. Specifically, the bias for the memory parameter tends to be negative, while the bias for the autoregressive parameter is typically positive. Furthermore, both biases display a degree of moderation and linearity when the true value of φ_0 is below 0. However, beyond this value, a pronounced acceleration in bias growth becomes evident until reaching a peak at around 0.5; beyond this point, the biases start to diminish rapidly. This same trend was noticed in a Monte Carlo simulation by Nielsen & Frederiksen (2005). They noted that the memory parameter's downward bias is particularly pronounced when the AR coefficient is either 0 or 0.4 and that the estimation methods seem robust against stronger positive AR coefficient, such as 0.8, which aligns with the bias expression in Theorem 3.4. Figure 2(b) displays the exact intrinsic bias of both the memory parameter and the autoregressive coefficient. The patterns of the biases closely resemble those of the approximated intrinsic bias in Figure 2(a). However, the specific values differ significantly, particularly in the range between 0 and 0.6. This discrepancy suggests that the asymptotic approximation can lead to notable distortion, especially within this range. Consequently, we suggest to use the exact intrinsic bias when correcting for it. As we will observe later, these biases also align with the findings of the simulation study. It is suboptimal to use the approximate intrinsic bias when the sample size is small.



(a) $S(d_0, \varphi_0)$ for $d_0 = -0.2$ (b) $S_T(d_0, \varphi_0)$ for $d_0 = -0.2$

(c) $S(d_0, \varphi_0)$ for $d_0 = 0.4$ (d) $S_T(d_0, \varphi_0)$ for $d_0 = 0.4$

(e) $S(d_0, \varphi_0)$ for $d_0 = 1$ (f) $S_T(d_0, \varphi_0)$ for $d_0 = 1$

Figure 3: The approximate and exact score bias for ARFIMA(1,d,0) model with $T = 128$ in the graphs on the left-hand side and right-hand side, respectively.

Figure 3 illustrates the approximate score bias as well as the exact score bias of the

ARFIMA(1,d_0,0) model with the memory parameter d_0 taking on values of -0.2, 0.4, and 1. Several observations can be drawn from these score biases. Firstly, the score bias tends to be more pronounced in the stationary region compared to the non-stationary region, which aligns with what we anticipated from Theorem 3.1. Secondly, there exists a noticeable symmetry between the memory parameter and the autoregressive coefficient. Thirdly, a close match is observed between the exact and approximate score biases for $d_0 = -0.2$ and $d = 1$, except when the autoregressive coefficient approaches 1 in the case of $d_0 = -0.2$. However, this correspondence breaks down when $d_0 = 0.4$; the approximate biases become distorted. This distortion arises due to the presence of the term $(1-2d_0)^{-1}$ in (101), which diverges as d_0 approaches 0.5. Robinson & Velasco (2015), who, in a panel setting, correct for the score bias of the CSS estimator, also observe that the precision of the approximate score bias expression diminishes unless the region is non-stationary. It is, therefore, recommendable to employ the exact score biases in practical empirical applications. Importantly, this bias is inherently eliminated by the MCSS estimator, obviating the need for additional correction.

3.2.2 Short-memory models

Theorem 3.3 also covers bias expressions in cases where long memory is absent in the model. This derivation is straightforward and hence the proof is not included; it involves truncating the matrix $B(\varphi_0)$ by removing the components related to long memory and by setting $d_0 = 0$ in $S(d_0, \varphi_0)$. The following theorem presents an approximate bias expression applicable to models characterised by short memory. It is important to emphasise that our model is not restricted to ARMA models alone; rather, it encompasses a broader category of short-memory models, with ARMA models being just one particular instance. Indeed, any representation that conforms to (65) is allowed, incorporating models like the Bloomfield exponential model.

Theorem 3.5. *Let x_t, $t = 1,\ldots,T$, be given by* (63) *with $d_0 = 0$ and let Assumptions 3.1 to 3.4 be satisfied. Furthermore, when d is set to zero in the respective objective functions, the approximate biases of $\hat{\varphi}$, $\hat{\varphi}_{\mu_0}$ and $\hat{\varphi}_m$ are given by*

$$bias(\hat{\varphi}) = \tilde{S}(\varphi_0) + \tilde{B}(\varphi_0) + o(T^{-1}), \tag{102}$$

$$bias(\hat{\varphi}_{\mu_0}) = \tilde{B}(\varphi_0) + o(T^{-1}), \tag{103}$$

$$bias(\hat{\varphi}_m) = \tilde{B}(\varphi_0) + o(T^{-1}), \tag{104}$$

where intrinsic bias is given by

$$T\tilde{B}(\varphi) = \tilde{A}^{-1} \begin{bmatrix} \iota' \left(\tilde{A}^{-1} \odot \left(\tilde{G}_1 + \tilde{F}_1 \right) \right) \iota \\ \vdots \\ \iota' \left(\tilde{A}^{-1} \odot \left(\tilde{G}_p + \tilde{F}_p \right) \right) \iota \end{bmatrix} - \frac{1}{2} \tilde{A}^{-1} \begin{bmatrix} \iota' \left(\left(\tilde{A}^{-1} \tilde{C}_{01} \tilde{A}^{-1} \right) \odot \tilde{A} \right) \iota \\ \vdots \\ \iota' \left(\left(\tilde{A}^{-1} \tilde{C}_{0p} \tilde{A}^{-1} \right) \odot \tilde{A} \right) \iota \end{bmatrix}$$

and the score bias is given by

$$T\tilde{S}(\varphi_0) = \tilde{A}^{-1} \frac{D_{\varphi_k} \phi(1; \varphi)}{\phi(1; \varphi)}$$

with

$$
\begin{aligned}
\tilde{A} &= \sum_{j=1}^{\infty} b_{\varphi j}(\varphi_0) b_{\varphi' j}(\varphi_0) \\
\tilde{F}_m &= \sum_{i=1}^{\infty} b_{\varphi\varphi_m i}(\varphi_0) b_{\varphi' i}(\varphi_0) \\
\tilde{G}_m &= \sum_{k=1}^{\infty}\sum_{s=1}^{\infty} \left(b_{\varphi_m s}(\varphi_0) b_{\varphi(s+k)}(\varphi_0) + b_{\varphi_m (s+k)}(\varphi_0) b_{\varphi s}(\varphi_0) \right) b_{\varphi' k}(\varphi_0) \\
\tilde{C}_{0m} &= \left[\sum_{i=1}^{\infty} b_{\varphi i}(\varphi_0) b_{\varphi'\varphi_m i}(\varphi_0) \right]' + \sum_{i=1}^{\infty} b_{\varphi i}(\varphi_0) b_{\varphi'\varphi_m i}(\varphi_0) + \sum_{i=1}^{\infty} b_{\varphi_m i}(\varphi_0) b_{\varphi\varphi' i}(\varphi_0)
\end{aligned}
$$

for $m = 1, \ldots, p$. *Here,* $b_{\cdot i}(\varphi_0)$ *is defined in* (68).

It is instructive to compare these expressions to analogous results for stationary and invertible ARFIMA models in Theorem 3.3. In that context, the bias of the short-run dynamics in the stationary region for the CSS estimator is of order $O(T^{-1}\log(T))$. This, however, is not true for the bias of the short-run dynamics of the CSS estimator for stationary and invertible ARMA models models, which is of order $O(T^{-1})$, as shown above in Theorem 3.5. An extension of a stationary ARMA model to an ARFIMA model increases the bias of the short-run dynamics to be of the same order of magnitude as that of the memory parameter. Nevertheless, the MCSS estimator effectively eliminates this bias through its construction, leading to a reduction in the bias order of $\hat{\varphi}$ to $O(T^{-1})$ for general ARFIMA or ARMA models.

Based on this theorem, we can deduce the analytic bias of an AR(1) model as an illustration. The following corollary presents the expressions describing the analytic biases of the three estimators. The proof of this corollary is omitted because it follows straightforwardly.

Corollary 3.1. *Let* x_t, $t = 1, \ldots, T$, *be given by* (63) *with* $d_0 = 0$ *and let* $u_t = \varphi u_{t-1} + \epsilon_t$. *Let Assumptions 3.1 to 3.4 be satisfied. Furthermore, when d is set to zero in the respective objective functions, the approximate biases of* $\hat{\varphi}$, $\hat{\varphi}_{\mu_0}$ *and* $\hat{\varphi}_m$ *are given by*

$$bias(\hat{\varphi}) = \tilde{S}(\varphi_0) + \tilde{B}(\varphi_0) + o(T^{-1}), \tag{105}$$

$$bias(\hat{\varphi}_{\mu_0}) = \tilde{B}(\varphi_0) + o(T^{-1}), \tag{106}$$

$$bias(\hat{\varphi}_m) = \tilde{B}(\varphi_0) + o(T^{-1}), \tag{107}$$

where intrinsic bias is given by $T\tilde{B}(\varphi) = -2\varphi$ *and the score bias is given by* $T\tilde{S}(\varphi_0) = -\varphi - 1$.

The bias expression is the same as the one discovered by Tanaka (1984). This observation is intriguing as Tanaka's focus was on maximum likelihood estimation, while we consider the CSS estimator. This suggests that the bias of the estimators remains unaffected in this basic model when the initial observation is set to zero.

3.3 Simulation

In this section we conduct a simulation study of the finite sample properties of the CSS estimators with known and unknown μ_0 and of the MCSS estimator. In Section 2 the focus was on the pure fractional case, here we incorporate an autoregressive component. In particular, we take as our DGP the model in (63) with u_t an AR(1) model, i.e.

$$u_t = \varphi_0 u_{t-1} + \epsilon_t,$$

with $\epsilon_t \sim NID(0,1)$. We choose the same values for d_0 as in Section 2.7 and for the autoregressive parameter we let $\varphi_0 \in \{-0.5, 0, 0.5\}$. We computed the estimates using the optimising interval $d \in [d_0 - 5, d_0 + 5]$, $\varphi \in [-0.9999, 0.9999]$. All results are based on 10,000 replications. In accordance with Assumption 3.4, we put the initial observation of the short-run dynamics equal to 0, i.e $u_0 = 0$. It should be noted, however, that this condition is not necessary to achieve consistency and asymptotic normality of the MCSS estimator.

Table 4 and 5 present the Monte Carlo biases (multiplied by 100) of the memory parameter and the autoregressive parameter, respectively. We also report the percentage increase of the bias of the CSS estimator relative to the bias of the MCSS estimator by $\Delta\%|\text{bias}|$ in the last column for each T. In addition, Table 6 and 7 present the Monte Carlo MSE (multiplied by 100) of the memory parameter and the autoregressive parameter, respectively. We now summarise the main findings. To explain the simulation results, it is useful to consider each case of $\varphi_0 \in \{-0.5, 0, 0.5\}$ separately.

We will first consider the case where $\varphi_0 = -0.5$. The addition of an autoregressive term to the pure fractional model considerably increases the bias of the CSS estimator of d, especially in the stationary region, cf. Table 2. The CSS estimator $\hat{d}$ clearly underestimates the true d_0, while the $\hat{\varphi}$ overestimates the true φ_0. The MCSS estimator, however, reduces a large part of the bias. The largest reduction occurs in the stationary region, which is also expected from Theorem 3.3. Importantly, the bias of the MCSS estimator and the bias of the CSS estimator with known μ_0 are close to each other, confirming our theoretical findings that the leading bias of $\hat{\vartheta}_m$ is the same as that of $\hat{\vartheta}_{\mu_0}$. It can also be seen that the MCSS estimator almost everywhere outperforms the CSS estimator in terms of MSE. The largest improvement occurs again in the stationary region, which is also where the largest bias reduction occurs according to Theorem 3.3. The CSS estimator with known μ_0 performs the best and outperforms the MCSS estimator, while the biases are somewhat similar, the variance of this estimator is significantly lower because μ_0 is known.

We next discuss the situation where $\varphi_0 = 0$. This situation is not covered in our theoretical analysis, since Assumption 3.3(i) does not allow for overspecification of the AR polynomial. Nevertheless, this is an interesting case and a realistic one. Usually the AR lags of the regression model are not known and often a lag selection procedure, such as the one by Box & Jenkins (1990) or an information criterion, is used to estimate the true number of lags, see for example Huang, Chan, Chen & Ing (2022). It is possible and not unlikely that the lag selection procedure or the information criterion overestimates the number of lags. Then, according to our simulation, the estimated parameters are strongly biased when the level parameter term is not known, so wrong conclusions may be drawn from these approaches. The MCSS estimator, as opposed to the CSS estimator, significantly reduces the bias and therefore seems a better alternative to base the lag selection procedure on, e.g. see Lee & Phillips (2015). We also note that the biases of the MCSS estimator and the bias of the CSS estimator with known μ_0 are close to each other. This result indicates that the leading bias terms are indeed the same for both estimators, which was also true for the purely fractional case, see Theorem 2.5. It can also be seen that the MCSS estimator outperforms everywhere the CSS estimator in terms of MSE. The largest improvement occurs again in the stationary region. The CSS estimator with known μ_0 performs again the best and outperforms the MCSS estimator.

Finally, we now discuss the situation where $\varphi_0 = 0.5$. The comments made above are also true for the non-stationary region. In particular, the bias of the MCSS estimator is close to that of the CSS estimator with known μ_0. In the stationary region, however, the two estimators behave differently. Nevertheless, the differences become small for $T = 256$. Furthermore, the CSS estimator performs the worst in terms of the bias, while the other two estimators significantly improve on this estimator. In terms of the MSE, we see that the MSE of the MCSS estimator and that of the CSS estimator with known μ_0 is significantly lower than that of the CSS estimator. As opposed to that, the MSE of the CSS estimator for φ is lower than that of the MCSS estimator and also the CSS estimator with known μ_0 when $T = 32$. In order to better understand the differences in the bias and MSE of the estimators, we have plotted in Figure 4 densities of the three estimators for the constellation $d_0 = -0.2$, $\varphi_0 = 0.5$ and $T = 32$ (upper panel) and $T = 256$ (lower panel). It can be seen that the CSS estimators strongly underestimate the true $d_0 = -0.2$ and strongly overestimate the true $\varphi_0 = 0.5$ for $T = 32$. This strong bias in the CSS estimator contrasts with less variation. The CSS estimator's poor performance extends somewhat to the case $T = 256$. The MCSS and CSS estimators are well-centred, but this centring comes at the cost of an increase in the variance. This explains the differences in the MSE of the CSS estimator relative to that of the MCSS estimator and the CSS estimator with known μ_0. Also, it appears that for small T the MCSS estimator recentres the memory parameter relatively more than the autoregressive component, explaining the differences with the bias of the CSS estimator with known μ_0. Nevertheless, these plots show that MCSS density estimates are more similar to CSS density estimates with known μ_0 than to CSS density estimates with unknown μ. The good finite sample performance of the MCSS estimator is again evident.

φ_0	$d_0 \setminus T$	bias($\hat{d}$)	bias($\hat{d}_{\mu_0}$)	bias($\hat{d}_m$)	$\Delta\%\lvert\text{bias}\rvert$	bias($\hat{d}$)	bias($\hat{d}_{\mu_0}$)	bias($\hat{d}_m$)	$\Delta\%\lvert\text{bias}\rvert$	bias($\hat{d}$)	bias($\hat{d}_{\mu_0}$)	bias($\hat{d}_m$)	$\Delta\%\lvert\text{bias}\rvert$	bias($\hat{d}$)	bias($\hat{d}_{\mu_0}$)	bias($\hat{d}_m$)	$\Delta\%\lvert\text{bias}\rvert$
		32				64				128				256			
	−0.2	−37.62	−9.02	−12.15	209.55	−13.70	−3.91	−4.50	204.65	−5.91	−1.89	−2.00	194.61	−2.88	−0.90	−0.90	219.88
	−0.1	−38.59	−9.02	−12.20	216.22	−13.86	−3.91	−4.54	205.57	−5.95	−1.89	−2.01	196.14	−2.89	−0.90	−0.90	220.69
	0.0	−39.26	−9.02	−12.35	217.98	−13.86	−3.91	−4.58	202.82	−5.96	−1.89	−2.01	195.97	−2.89	−0.90	−0.90	220.00
	0.1	−39.60	−9.02	−12.35	220.59	−13.92	−3.91	−4.65	199.07	−5.94	−1.89	−2.02	193.49	−2.87	−0.90	−0.91	216.75
	0.2	−39.80	−9.02	−12.49	218.60	−13.87	−3.91	−4.65	198.41	−5.83	−1.89	−2.03	186.95	−2.82	−0.90	−0.91	209.18
	0.3	−39.18	−9.02	−12.42	215.46	−13.54	−3.91	−4.67	189.92	−5.60	−1.89	−2.04	173.99	−2.70	−0.90	−0.92	194.59
	0.4	−38.01	−9.02	−12.35	207.92	−12.88	−3.91	−4.64	177.86	−5.24	−1.89	−2.05	155.29	−2.49	−0.90	−0.92	169.89
−0.5	0.5	−36.47	−9.02	−12.17	199.53	−11.83	−3.91	−4.58	158.39	−4.68	−1.89	−2.05	128.31	−2.18	−0.90	−0.93	134.69
	0.6	−33.69	−9.02	−12.10	178.53	−10.32	−3.91	−4.52	128.34	−3.98	−1.89	−2.03	96.41	−1.82	−0.90	−0.93	94.92
	0.7	−29.72	−9.02	−11.68	154.40	−8.73	−3.91	−4.43	97.08	−3.28	−1.89	−1.99	64.77	−1.49	−0.90	−0.93	59.79
	0.8	−25.26	−9.01	−10.96	130.49	−7.25	−3.91	−4.29	69.05	−2.72	−1.89	−1.95	39.25	−1.24	−0.90	−0.92	34.33
	0.9	−20.84	−9.02	−10.39	100.68	−5.99	−3.91	−4.14	44.64	−2.35	−1.89	−1.93	21.77	−1.08	−0.90	−0.91	17.89
	1.0	−16.97	−9.02	−9.82	72.78	−5.13	−3.91	−4.06	26.45	−2.10	−1.89	−1.91	9.60	−0.98	−0.90	−0.91	7.59
	1.1	−13.97	−9.02	−9.67	44.41	−4.46	−3.91	−4.00	11.50	−1.94	−1.89	−1.90	2.12	−0.91	−0.90	−0.90	1.00
	1.2	−11.89	−9.02	−9.51	25.08	−4.05	−3.91	−3.98	1.84	−1.85	−1.89	−1.90	−2.73	−0.87	−0.90	−0.90	−3.40
	−0.2	−57.19	−16.48	−16.99	236.65	−36.69	−10.32	−12.17	201.58	−17.91	−4.87	−6.18	189.68	−6.90	−2.09	−2.36	191.75
	−0.1	−58.58	−16.49	−17.18	241.02	−37.44	−10.31	−12.22	206.35	−18.19	−4.84	−6.19	194.03	−6.97	−2.09	−2.39	191.14
	0.0	−59.70	−16.48	−17.31	244.86	−37.95	−10.32	−12.28	208.95	−18.48	−4.86	−6.24	196.00	−7.01	−2.08	−2.40	192.48
	0.1	−60.38	−16.50	−17.42	246.56	−37.97	−10.32	−12.18	211.67	−18.62	−4.86	−6.28	196.74	−7.01	−2.09	−2.39	193.12
	0.2	−60.14	−16.48	−17.40	245.57	−37.64	−10.33	−12.17	209.18	−18.44	−4.85	−6.30	192.98	−6.84	−2.09	−2.39	186.62
	0.3	−58.88	−16.49	−17.67	233.21	−36.57	−10.31	−12.15	201.11	−17.77	−4.85	−6.21	186.04	−6.51	−2.09	−2.37	174.95
	0.4	−56.06	−16.49	−17.81	214.74	−34.64	−10.33	−12.06	187.11	−16.57	−4.87	−6.29	163.68	−6.00	−2.09	−2.36	154.47
0	0.5	−51.49	−16.49	−18.14	183.83	−31.31	−10.30	−11.86	163.92	−14.55	−4.83	−6.03	141.40	−5.22	−2.08	−2.30	126.77
	0.6	−45.50	−16.50	−18.20	149.98	−26.72	−10.32	−11.57	130.98	−12.12	−4.85	−5.78	109.84	−4.19	−2.08	−2.24	87.40
	0.7	−38.84	−16.50	−18.01	115.63	−21.77	−10.33	−11.23	93.86	−9.37	−4.85	−5.41	73.28	−3.26	−2.09	−2.19	48.88
	0.8	−31.76	−16.49	−17.34	83.18	−17.36	−10.32	−10.91	59.14	−7.27	−4.87	−5.20	39.92	−2.65	−2.09	−2.14	23.78
	0.9	−25.92	−16.49	−17.06	51.93	−14.10	−10.32	−10.76	31.10	−5.95	−4.84	−5.01	18.74	−2.34	−2.08	−2.12	10.22
	1.0	−21.81	−16.49	−16.80	29.85	−12.12	−10.33	−10.55	14.89	−5.32	−4.87	−4.95	7.49	−2.16	−2.09	−2.10	3.03
	1.1	−19.14	−16.51	−16.75	14.28	−11.04	−10.32	−10.42	5.95	−5.01	−4.87	−4.92	1.88	−2.07	−2.08	−2.06	0.44
	1.2	−17.51	−16.50	−16.60	5.44	−10.52	−10.32	−10.35	1.60	−4.91	−4.88	−4.93	−0.43	−2.06	−2.09	−2.08	−1.27
	−0.2	−33.35	−5.93	2.77	1105.96	−24.73	−5.95	−0.81	2957.56	−19.03	−5.82	−3.08	517.33	−14.34	−4.82	−4.17	243.92
	−0.1	−33.57	−5.93	2.56	1209.35	−24.85	−5.96	−0.63	3848.19	−19.15	−5.82	−3.02	534.82	−14.40	−4.82	−4.13	248.21
	0.0	−32.95	−5.92	2.15	1430.67	−24.58	−5.96	−0.70	3422.44	−19.02	−5.82	−3.01	531.20	−14.31	−4.82	−4.18	242.10
	0.1	−31.69	−5.93	1.45	2091.57	−23.63	−5.96	−1.00	2264.08	−18.51	−5.82	−3.11	495.28	−14.13	−4.82	−4.20	235.94
	0.2	−28.86	−5.93	0.37	7615.84	−21.52	−5.96	−1.75	1129.85	−17.44	−5.82	−3.38	415.34	−13.48	−4.82	−4.19	221.80
	0.3	−24.89	−5.94	−1.34	1755.66	−18.68	−5.95	−2.69	594.38	−15.41	−5.82	−3.98	287.05	−12.08	−4.82	−4.34	178.18
	0.4	−20.17	−5.93	−3.11	547.84	−15.00	−5.96	−3.83	291.96	−12.62	−5.83	−4.64	171.91	−10.05	−4.82	−4.66	115.78
0.5	0.5	−15.68	−5.94	−4.28	266.08	−11.55	−5.95	−4.88	136.53	−9.64	−5.82	−5.11	88.55	−7.84	−4.82	−4.83	62.35
	0.6	−11.66	−5.94	−4.93	136.48	−8.97	−5.95	−5.53	62.16	−7.65	−5.82	−5.51	39.03	−6.18	−4.82	−4.83	27.91
	0.7	−8.79	−5.95	−5.30	65.83	−7.36	−5.96	−5.82	26.52	−6.52	−5.82	−5.72	14.03	−5.37	−4.82	−4.89	9.83
	0.8	−6.97	−5.93	−5.48	27.06	−6.50	−5.95	−5.98	8.70	−6.04	−5.82	−5.82	3.83	−5.03	−4.82	−4.88	3.10
	0.9	−5.93	−5.94	−5.69	4.35	−6.02	−5.95	−6.03	−0.14	−5.86	−5.82	−5.85	0.07	−4.86	−4.82	−4.85	0.25
	1.0	−5.52	−5.94	−5.88	−6.15	−5.91	−5.96	−6.01	−1.67	−5.77	−5.83	−5.85	−1.35	−4.82	−4.82	−4.84	−0.34
	1.1	−5.32	−5.95	−5.96	−10.73	−5.84	−5.95	−6.05	−3.45	−5.74	−5.82	−5.84	−1.67	−4.81	−4.82	−4.81	−0.09
	1.2	−5.22	−5.94	−6.00	−13.00	−5.83	−5.96	−6.06	−3.76	−5.77	−5.82	−5.83	−0.99	−4.79	−4.82	−4.82	−0.69

Table 4: (100 ×) Monte Carlo bias of the estimated memory parameter for ARFIMA(1,d_0,0) of CSS estimator with unknown and known μ_0 and the MCSS estimator.

φ_0	$d_0 \setminus T$	bias($\hat{\varphi}$)	bias($\hat{\varphi}_{\mu_0}$)	bias($\hat{\varphi}_m$)	$\Delta\%\lvert\text{bias}\rvert$	bias($\hat{\varphi}$)	bias($\hat{\varphi}_{\mu_0}$)	bias($\hat{\varphi}_m$)	$\Delta\%\lvert\text{bias}\rvert$	bias($\hat{\varphi}$)	bias($\hat{\varphi}_{\mu_0}$)	bias($\hat{\varphi}_m$)	$\Delta\%\lvert\text{bias}\rvert$	bias($\hat{\varphi}$)	bias($\hat{\varphi}_{\mu_0}$)	bias($\hat{\varphi}_m$)	$\Delta\%\lvert\text{bias}\rvert$
		32				64				128				256			
	−0.2	26.93	8.25	9.89	172.41	9.43	3.67	3.96	137.96	3.95	1.86	1.90	107.99	1.97	0.95	0.95	107.64
	−0.1	27.70	8.24	9.88	180.42	9.55	3.67	3.99	139.20	3.97	1.86	1.90	109.35	1.98	0.95	0.95	108.25
	0.0	28.22	8.25	9.94	183.75	9.53	3.68	4.01	137.38	3.99	1.86	1.90	109.73	1.98	0.95	0.95	108.10
	0.1	28.48	8.25	9.87	188.51	9.59	3.67	4.07	135.59	3.97	1.86	1.90	108.76	1.96	0.95	0.95	106.58
	0.2	28.76	8.25	9.95	189.11	9.63	3.67	4.06	137.42	3.91	1.86	1.91	105.22	1.93	0.95	0.95	102.69
	0.3	28.40	8.24	9.82	189.15	9.46	3.67	4.06	132.98	3.77	1.86	1.91	97.22	1.86	0.95	0.95	94.89
	0.4	27.68	8.24	9.72	184.83	9.07	3.67	4.01	126.00	3.56	1.86	1.91	85.99	1.73	0.95	0.96	81.33
−0.5	0.5	26.80	8.24	9.57	180.01	8.43	3.67	3.97	112.61	3.23	1.86	1.91	68.64	1.55	0.95	0.96	61.61
	0.6	24.95	8.25	9.59	160.21	7.43	3.67	3.94	88.78	2.81	1.86	1.91	47.58	1.34	0.95	0.96	39.01
	0.7	22.16	8.24	9.39	135.92	6.40	3.67	3.90	63.94	2.40	1.86	1.89	26.62	1.15	0.95	0.96	19.10
	0.8	19.05	8.24	9.01	111.49	5.44	3.67	3.84	41.84	2.07	1.86	1.88	10.31	1.01	0.95	0.96	5.13
	0.9	15.91	8.24	8.72	82.36	4.62	3.67	3.76	22.86	1.87	1.86	1.87	0.26	0.93	0.95	0.96	−3.26
	1.0	13.15	8.24	8.43	56.02	4.09	3.67	3.73	9.73	1.74	1.86	1.86	−6.41	0.88	0.95	0.95	−7.88
	1.1	11.07	8.24	8.47	30.58	3.66	3.67	3.71	−1.14	1.68	1.86	1.86	−9.74	0.85	0.95	0.95	−10.29
	1.2	9.64	8.25	8.44	14.24	3.43	3.67	3.70	−7.41	1.65	1.86	1.86	−11.34	0.84	0.95	0.95	−11.42
	−0.2	41.35	12.08	10.83	281.83	30.07	8.43	9.35	221.49	15.59	4.18	5.18	200.96	6.19	1.89	2.10	195.07
	−0.1	42.40	12.09	10.93	288.02	30.70	8.41	9.39	227.04	15.85	4.16	5.19	205.38	6.26	1.89	2.13	194.01
	0.0	43.24	12.09	10.96	294.39	31.12	8.43	9.43	230.15	16.10	4.17	5.23	208.00	6.30	1.89	2.13	195.69
	0.1	43.72	12.10	10.99	297.66	31.10	8.44	9.31	234.19	16.22	4.17	5.24	209.35	6.30	1.89	2.12	196.65
	0.2	43.49	12.08	10.95	297.07	30.77	8.43	9.28	231.66	16.06	4.17	5.27	205.01	6.14	1.90	2.12	189.58
	0.3	42.45	12.08	11.21	278.73	29.82	8.42	9.25	222.42	15.44	4.17	5.19	197.74	5.82	1.89	2.09	177.94
	0.4	40.13	12.08	11.41	251.71	28.10	8.43	9.20	205.59	14.34	4.18	5.26	172.47	5.34	1.89	2.09	155.77
0	0.5	36.50	12.08	11.86	207.83	25.23	8.42	9.09	177.41	12.50	4.16	5.06	147.18	4.61	1.88	2.04	126.29
	0.6	31.89	12.09	12.16	162.37	21.32	8.43	8.95	138.07	10.31	4.17	4.85	112.49	3.67	1.88	1.99	84.58
	0.7	27.04	12.10	12.32	119.53	17.23	8.43	8.81	95.50	7.90	4.17	4.57	72.78	2.84	1.89	1.96	44.85
	0.8	21.97	12.09	12.06	82.11	13.69	8.43	8.67	57.96	6.09	4.18	4.41	38.20	2.32	1.89	1.92	20.42
	0.9	18.00	12.10	12.11	48.60	11.16	8.43	8.64	29.06	5.03	4.17	4.29	17.31	2.07	1.89	1.92	8.32
	1.0	15.37	12.09	12.07	27.32	9.71	8.43	8.52	13.97	4.55	4.19	4.24	7.36	1.95	1.89	1.89	3.00
	1.1	13.75	12.10	12.13	13.36	9.01	8.42	8.45	6.53	4.36	4.19	4.23	3.20	1.91	1.89	1.87	2.18
	1.2	12.88	12.09	12.08	6.65	8.74	8.44	8.42	3.78	4.33	4.20	4.24	2.23	1.92	1.90	1.89	1.97
	−0.2	15.50	−2.25	−8.38	84.89	14.41	0.66	−3.60	300.20	12.90	2.56	−0.09	13834.33	10.93	3.01	2.18	400.54
	−0.1	15.65	−2.24	−8.21	90.58	14.47	0.66	−3.78	283.08	12.98	2.56	−0.16	7966.49	10.97	3.01	2.15	410.69
	0.0	15.19	−2.25	−7.92	91.86	14.20	0.66	−3.72	281.65	12.83	2.55	−0.17	7432.74	10.88	3.01	2.19	396.56
	0.1	14.35	−2.24	−7.43	93.08	13.40	0.66	−3.47	286.12	12.36	2.55	−0.08	14628.76	10.69	3.01	2.21	383.96
	0.2	12.35	−2.24	−6.74	83.05	11.67	0.66	−2.82	313.86	11.43	2.56	0.18	6182.65	10.10	3.01	2.21	357.94
	0.3	9.54	−2.23	−5.51	73.04	9.49	0.66	−2.00	373.77	9.73	2.55	0.75	1199.28	8.87	3.01	2.38	273.09
	0.4	6.33	−2.24	−4.22	49.85	6.79	0.66	−1.05	547.99	7.53	2.56	1.40	438.16	7.15	3.01	2.71	163.48
0.5	0.5	3.47	−2.23	−3.44	0.83	4.41	0.66	−0.16	2683.50	5.29	2.55	1.87	182.76	5.36	3.01	2.93	83.22
	0.6	1.08	−2.23	−3.01	−64.13	2.75	0.65	0.36	669.70	3.89	2.55	2.25	73.17	4.07	3.01	2.97	36.98
	0.7	−0.41	−2.22	−2.76	−85.13	1.76	0.66	0.56	216.75	3.12	2.55	2.44	27.78	3.47	3.01	3.04	14.09
	0.8	−1.37	−2.24	−2.69	−49.04	1.27	0.66	0.66	92.05	2.83	2.55	2.53	11.85	3.24	3.01	3.04	6.45
	0.9	−1.81	−2.23	−2.54	−28.66	1.02	0.66	0.70	45.54	2.75	2.55	2.56	7.21	3.13	3.01	3.02	3.51
	1.0	−1.87	−2.24	−2.38	−21.41	1.05	0.66	0.68	54.67	2.73	2.56	2.57	6.19	3.12	3.01	3.02	3.38
	1.1	−1.84	−2.23	−2.31	−20.49	1.06	0.66	0.71	49.18	2.73	2.56	2.55	6.99	3.12	3.01	2.99	4.26
	1.2	−1.76	−2.23	−2.26	−22.26	1.11	0.67	0.73	53.33	2.78	2.55	2.54	9.37	3.11	3.01	3.00	3.72

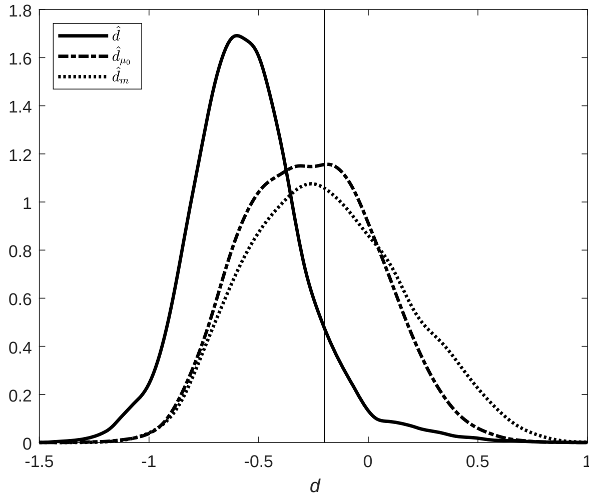
Table 5: (100 ×) Monte Carlo bias of the estimated AR coefficient for ARFIMA(1,d_0,0) of CSS estimator with unknown and known μ_0 and the MCSS estimator.

φ_0	$d_0 \setminus T$	MSE($\hat{d}$)	MSE($\hat{d}_{\mu_0}$)	MSE($\hat{d}_m$)	$\Delta\%\lvert$MSE$\rvert$	MSE($\hat{d}$)	MSE($\hat{d}_{\mu_0}$)	MSE($\hat{d}_m$)	$\Delta\%\lvert$MSE$\rvert$	MSE($\hat{d}$)	MSE($\hat{d}_{\mu_0}$)	MSE($\hat{d}_m$)	$\Delta\%\lvert$MSE$\rvert$	MSE($\hat{d}$)	MSE($\hat{d}_{\mu_0}$)	MSE($\hat{d}_m$)	$\Delta\%\lvert$MSE$\rvert$
		32				64				128				256			
	−0.2	38.58	7.89	14.57	164.83	7.04	2.21	3.37	108.65	1.50	0.86	1.10	36.42	0.54	0.39	0.45	18.83
	−0.1	40.16	7.89	14.52	176.55	7.17	2.21	3.42	109.90	1.52	0.86	1.10	38.22	0.54	0.39	0.45	19.14
	0.0	41.46	7.89	14.69	182.33	7.12	2.21	3.45	106.42	1.54	0.86	1.10	39.89	0.54	0.39	0.45	19.45
	0.1	42.41	7.89	14.54	191.68	7.36	2.21	3.57	106.19	1.56	0.86	1.10	42.15	0.54	0.39	0.45	19.73
	0.2	43.39	7.89	14.55	198.14	7.60	2.21	3.50	117.52	1.55	0.86	1.09	42.69	0.54	0.39	0.45	19.93
	0.3	43.27	7.89	14.08	207.44	7.68	2.21	3.45	122.40	1.51	0.86	1.07	40.28	0.53	0.39	0.45	19.95
	0.4	42.77	7.89	13.52	216.46	7.60	2.21	3.29	130.76	1.49	0.86	1.05	42.51	0.52	0.39	0.44	19.42
−0.5	0.5	42.53	7.89	12.89	230.00	7.24	2.21	3.06	136.20	1.43	0.86	1.01	40.59	0.50	0.39	0.43	17.67
	0.6	40.63	7.89	12.70	219.98	6.45	2.21	2.90	122.43	1.31	0.86	0.97	34.27	0.48	0.39	0.42	14.16
	0.7	36.63	7.89	11.93	207.16	5.64	2.21	2.76	104.61	1.16	0.86	0.94	23.98	0.45	0.39	0.41	9.69
	0.8	31.57	7.89	10.76	193.36	4.83	2.21	2.58	86.96	1.03	0.86	0.90	14.40	0.42	0.39	0.40	5.81
	0.9	26.42	7.89	9.91	166.44	3.98	2.21	2.41	65.47	0.98	0.86	0.88	10.69	0.41	0.39	0.39	3.22
	1.0	21.39	7.90	8.97	138.38	3.44	2.21	2.32	47.86	0.91	0.86	0.87	4.69	0.40	0.39	0.39	1.71
	1.1	17.14	7.89	8.81	94.52	2.88	2.21	2.27	26.62	0.89	0.86	0.87	2.38	0.39	0.39	0.39	0.86
	1.2	14.10	7.89	8.63	63.37	2.56	2.21	2.27	12.70	0.87	0.86	0.86	1.08	0.39	0.39	0.39	0.36
	−0.2	54.39	16.06	21.94	147.94	29.91	8.56	11.69	155.87	11.74	3.23	4.87	141.09	2.79	1.05	1.41	97.75
	−0.1	56.32	16.06	22.01	155.85	30.85	8.54	11.71	163.49	12.04	3.20	4.87	147.28	2.84	1.06	1.43	98.77
	0.0	58.19	16.06	21.87	166.08	31.56	8.55	11.75	168.62	12.35	3.21	4.89	152.54	2.88	1.05	1.43	101.76
	0.1	59.83	16.07	21.56	177.55	31.86	8.56	11.57	175.36	12.62	3.22	4.89	157.84	2.93	1.06	1.41	107.90
	0.2	60.51	16.07	20.92	189.20	31.96	8.56	11.44	179.41	12.68	3.21	4.89	159.47	2.88	1.06	1.40	105.33
	0.3	60.54	16.06	20.45	196.03	31.58	8.55	11.22	181.60	12.37	3.21	4.73	161.39	2.77	1.05	1.37	101.69
	0.4	59.17	16.06	19.82	198.56	30.57	8.56	10.87	181.31	11.80	3.22	4.75	148.55	2.68	1.05	1.34	99.97
0	0.5	55.90	16.06	19.39	188.28	28.37	8.55	10.40	172.73	10.61	3.20	4.41	140.38	2.48	1.05	1.24	99.57
	0.6	50.68	16.07	18.75	170.28	24.71	8.56	9.95	148.45	9.09	3.21	4.13	120.19	2.09	1.05	1.16	79.60
	0.7	44.09	16.07	18.04	144.38	20.53	8.56	9.50	116.02	7.08	3.21	3.75	88.63	1.67	1.05	1.12	49.09
	0.8	36.42	16.06	17.10	112.94	16.35	8.56	9.15	78.80	5.37	3.23	3.52	52.39	1.38	1.05	1.09	26.65
	0.9	29.58	16.07	16.80	76.14	13.01	8.56	9.00	44.56	4.25	3.21	3.35	26.85	1.24	1.05	1.08	14.15
	1.0	24.57	16.07	16.51	48.81	10.90	8.56	8.79	24.01	3.72	3.23	3.30	12.71	1.12	1.05	1.06	6.13
	1.1	20.98	16.08	16.49	27.23	9.69	8.56	8.66	11.84	3.45	3.23	3.28	5.28	1.07	1.05	1.03	3.95
	1.2	18.70	16.07	16.36	14.28	9.05	8.56	8.59	5.34	3.37	3.23	3.30	1.95	1.06	1.06	1.05	1.20
	−0.2	18.68	9.55	12.60	48.33	11.81	6.57	7.58	55.91	8.17	4.57	5.05	61.73	5.48	2.95	3.43	60.01
	−0.1	19.25	9.55	12.37	55.59	12.03	6.57	7.55	59.35	8.28	4.57	5.02	65.00	5.53	2.95	3.42	61.79
	0.0	19.73	9.55	11.91	65.71	12.20	6.57	7.45	63.68	8.34	4.57	5.00	66.82	5.53	2.95	3.43	61.53
	0.1	20.11	9.55	11.24	79.03	12.28	6.57	7.23	69.89	8.39	4.57	4.94	69.67	5.56	2.95	3.41	62.94
	0.2	19.91	9.55	10.54	88.96	12.09	6.57	6.91	75.04	8.29	4.57	4.84	71.17	5.50	2.95	3.35	64.22
	0.3	19.18	9.55	9.92	93.28	11.56	6.57	6.59	75.38	7.89	4.57	4.73	66.86	5.22	2.95	3.27	59.64
	0.4	17.74	9.55	9.55	85.77	10.49	6.57	6.41	63.58	7.12	4.57	4.61	54.63	4.71	2.95	3.20	47.53
0.5	0.5	15.91	9.55	9.39	69.42	9.18	6.57	6.39	43.54	6.10	4.57	4.53	34.67	4.03	2.95	3.10	30.02
	0.6	13.86	9.55	9.32	48.72	8.04	6.57	6.48	24.16	5.35	4.57	4.53	17.89	3.46	2.95	3.02	14.80
	0.7	12.04	9.55	9.34	28.99	7.29	6.57	6.52	11.74	4.88	4.57	4.57	6.87	3.16	2.95	3.00	5.31
	0.8	10.89	9.55	9.42	15.57	6.90	6.57	6.61	4.28	4.67	4.57	4.58	1.81	3.03	2.95	2.98	1.60
	0.9	10.10	9.55	9.49	6.44	6.67	6.57	6.61	0.87	4.59	4.57	4.60	−0.19	2.96	2.95	2.97	−0.28
	1.0	9.74	9.55	9.62	1.25	6.56	6.57	6.62	−0.83	4.54	4.57	4.59	−0.99	2.94	2.95	2.96	−0.72
	1.1	9.56	9.55	9.61	−0.51	6.50	6.57	6.62	−1.77	4.52	4.57	4.59	−1.54	2.93	2.95	2.95	−0.66
	1.2	9.41	9.55	9.64	−2.35	6.46	6.57	6.60	−1.99	4.53	4.57	4.59	−1.33	2.92	2.95	2.96	−1.10

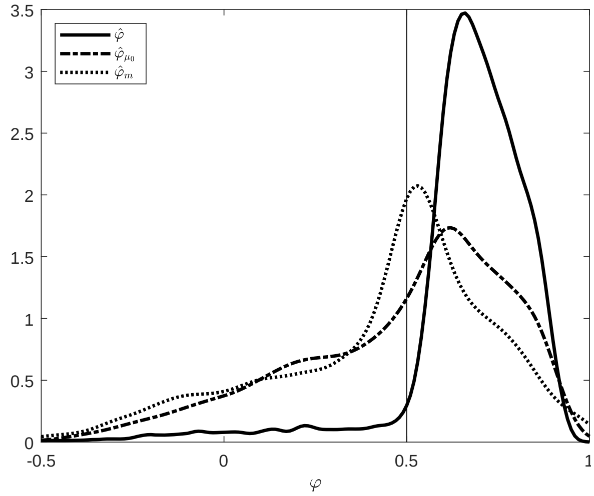
Table 6: (100 ×) Empirical MSE of the estimated memory parameter for ARFIMA(1,d_0,0) of CSS estimator with unknown and known μ_0 and the MCSS estimator.

φ_0	$d_0 \setminus T$	MSE($\hat{\varphi}$)	MSE($\hat{\varphi}_{\mu_0}$)	MSE($\hat{\varphi}_m$)	$\Delta\%\lvert$MSE$\rvert$	MSE($\hat{\varphi}$)	MSE($\hat{\varphi}_{\mu_0}$)	MSE($\hat{\varphi}_m$)	$\Delta\%\lvert$MSE$\rvert$	MSE($\hat{\varphi}$)	MSE($\hat{\varphi}_{\mu_0}$)	MSE($\hat{\varphi}_m$)	$\Delta\%\lvert$MSE$\rvert$	MSE($\hat{\varphi}$)	MSE($\hat{\varphi}_{\mu_0}$)	MSE($\hat{\varphi}_m$)	$\Delta\%\lvert$MSE$\rvert$
		32				64				128				256			
	−0.2	28.23	7.91	10.40	171.49	6.00	2.46	2.92	105.29	1.32	0.98	1.05	26.59	0.52	0.45	0.46	11.95
	−0.1	29.35	7.91	10.35	183.70	6.10	2.46	2.95	106.59	1.35	0.98	1.05	28.51	0.52	0.45	0.46	12.10
	0.0	30.10	7.91	10.43	188.64	6.00	2.46	2.98	101.49	1.35	0.98	1.05	29.41	0.52	0.45	0.46	12.21
	0.1	30.52	7.91	10.23	198.33	6.16	2.46	3.09	99.60	1.37	0.98	1.05	30.75	0.52	0.45	0.46	12.22
	0.2	31.28	7.91	10.32	202.98	6.41	2.46	3.04	110.82	1.36	0.98	1.04	30.67	0.52	0.45	0.46	12.07
	0.3	31.19	7.91	10.07	209.76	6.46	2.46	3.03	113.64	1.32	0.98	1.04	26.87	0.51	0.45	0.46	11.61
	0.4	30.82	7.91	9.81	214.04	6.42	2.46	2.93	119.38	1.32	0.98	1.03	28.24	0.51	0.45	0.46	10.62
−0.5	0.5	30.61	7.91	9.48	222.79	6.19	2.46	2.80	120.77	1.28	0.98	1.02	25.36	0.50	0.45	0.46	8.84
	0.6	29.10	7.91	9.50	206.32	5.53	2.46	2.73	102.23	1.21	0.98	1.01	19.70	0.48	0.45	0.45	6.34
	0.7	26.22	7.90	9.26	183.24	4.93	2.46	2.70	82.52	1.12	0.98	1.00	11.73	0.47	0.45	0.45	3.75
	0.8	22.88	7.91	8.73	162.01	4.35	2.46	2.62	66.02	1.05	0.98	0.99	5.74	0.46	0.45	0.45	1.78
	0.9	19.22	7.91	8.37	129.68	3.72	2.46	2.52	47.62	1.03	0.98	0.99	4.31	0.45	0.45	0.45	0.60
	1.0	15.81	7.90	7.98	98.09	3.34	2.46	2.49	34.10	0.99	0.98	0.98	0.84	0.45	0.45	0.45	−0.01
	1.1	13.17	7.91	8.18	61.05	2.89	2.46	2.46	17.31	0.98	0.98	0.98	−0.07	0.44	0.45	0.45	−0.28
	1.2	11.24	7.91	8.18	37.43	2.66	2.46	2.48	7.04	0.98	0.98	0.98	−0.44	0.44	0.45	0.45	−0.38
	−0.2	34.38	15.17	15.24	125.60	24.50	9.12	10.47	134.08	11.24	3.91	5.10	120.16	3.09	1.44	1.74	77.98
	−0.1	35.39	15.18	15.22	132.58	25.13	9.11	10.48	139.84	11.47	3.89	5.10	124.88	3.14	1.45	1.76	78.67
	0.0	36.33	15.18	15.10	140.55	25.59	9.11	10.48	144.28	11.74	3.90	5.12	129.15	3.18	1.44	1.75	81.17
	0.1	37.13	15.19	14.97	148.08	25.73	9.12	10.31	149.47	11.96	3.91	5.13	133.33	3.22	1.45	1.73	85.65
	0.2	37.41	15.18	14.71	154.32	25.73	9.12	10.23	151.63	12.00	3.90	5.11	134.85	3.16	1.45	1.72	83.48
	0.3	37.24	15.18	14.67	153.90	25.38	9.11	10.12	150.92	11.71	3.90	4.98	135.36	3.06	1.44	1.70	80.42
	0.4	36.08	15.18	14.55	147.91	24.52	9.12	9.91	147.40	11.19	3.91	5.02	122.88	2.96	1.44	1.67	77.75
0	0.5	33.88	15.18	14.66	131.05	22.72	9.12	9.66	135.32	10.12	3.89	4.77	112.32	2.77	1.44	1.59	74.58
	0.6	30.81	15.19	14.77	108.56	19.88	9.12	9.45	110.36	8.79	3.90	4.55	93.16	2.39	1.44	1.52	57.20
	0.7	27.44	15.18	14.92	83.82	16.80	9.12	9.30	80.52	7.03	3.90	4.27	64.81	2.00	1.45	1.49	33.84
	0.8	23.61	15.18	14.83	59.27	13.92	9.12	9.24	50.72	5.60	3.91	4.12	36.02	1.73	1.44	1.47	17.87
	0.9	20.62	15.19	15.03	37.16	11.79	9.12	9.27	27.08	4.70	3.90	4.01	17.39	1.60	1.44	1.47	9.12
	1.0	18.51	15.19	15.15	22.15	10.50	9.12	9.21	14.00	4.31	3.92	3.98	8.28	1.50	1.44	1.45	3.90
	1.1	17.14	15.19	15.28	12.13	9.84	9.12	9.18	7.19	4.11	3.92	3.97	3.64	1.47	1.44	1.43	2.87
	1.2	16.32	15.19	15.27	6.87	9.52	9.13	9.15	4.01	4.06	3.92	3.99	1.89	1.46	1.45	1.44	1.21
	−0.2	6.69	9.11	10.56	−36.62	6.41	6.22	6.86	−6.59	5.52	4.30	4.68	18.11	4.20	2.83	3.14	33.75
	−0.1	6.87	9.11	10.43	−34.14	6.49	6.22	6.86	−5.41	5.59	4.30	4.66	19.93	4.23	2.83	3.13	34.89
	0.0	7.07	9.11	10.27	−31.17	6.59	6.22	6.80	−3.15	5.62	4.30	4.64	21.25	4.24	2.83	3.14	34.99
	0.1	7.29	9.11	10.06	−27.49	6.70	6.22	6.67	0.46	5.67	4.30	4.59	23.50	4.27	2.83	3.13	36.23
	0.2	7.64	9.11	9.80	−22.05	6.78	6.22	6.48	4.62	5.67	4.30	4.50	25.98	4.24	2.83	3.08	37.67
	0.3	8.11	9.11	9.51	−14.76	6.79	6.22	6.32	7.43	5.51	4.30	4.40	25.17	4.08	2.83	3.02	34.97
	0.4	8.55	9.11	9.27	−7.81	6.72	6.22	6.23	7.92	5.22	4.30	4.32	20.74	3.78	2.83	2.96	27.46
0.5	0.5	8.89	9.10	9.19	−3.27	6.57	6.22	6.18	6.16	4.83	4.30	4.28	12.87	3.40	2.83	2.91	16.87
	0.6	9.09	9.10	9.20	−1.10	6.42	6.22	6.22	3.20	4.56	4.30	4.29	6.20	3.09	2.83	2.86	8.04
	0.7	9.15	9.10	9.21	−0.65	6.32	6.22	6.24	1.27	4.40	4.30	4.31	2.08	2.94	2.83	2.86	2.63
	0.8	9.20	9.11	9.23	−0.33	6.26	6.22	6.29	−0.34	4.32	4.30	4.32	0.14	2.87	2.83	2.85	0.70
	0.9	9.17	9.10	9.24	−0.80	6.22	6.22	6.27	−0.76	4.30	4.30	4.33	−0.67	2.83	2.83	2.84	−0.38
	1.0	9.10	9.11	9.25	−1.65	6.17	6.22	6.27	−1.57	4.27	4.30	4.32	−1.04	2.82	2.83	2.84	−0.62
	1.1	9.03	9.10	9.23	−2.12	6.14	6.22	6.27	−1.98	4.25	4.30	4.32	−1.45	2.82	2.83	2.83	−0.58
	1.2	8.96	9.10	9.22	−2.78	6.11	6.22	6.25	−2.30	4.26	4.30	4.32	−1.34	2.81	2.83	2.83	−0.95

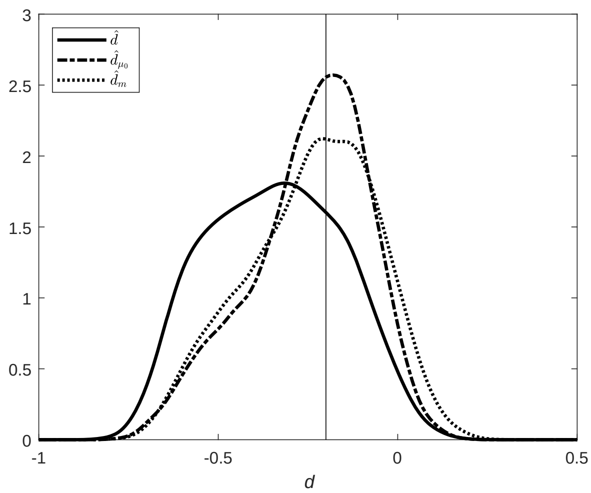
Table 7: (100 ×) Empirical MSE of the estimated AR coefficient for ARFIMA(1,d_0,0) of CSS estimator with unknown and known μ_0 and the MCSS estimator.



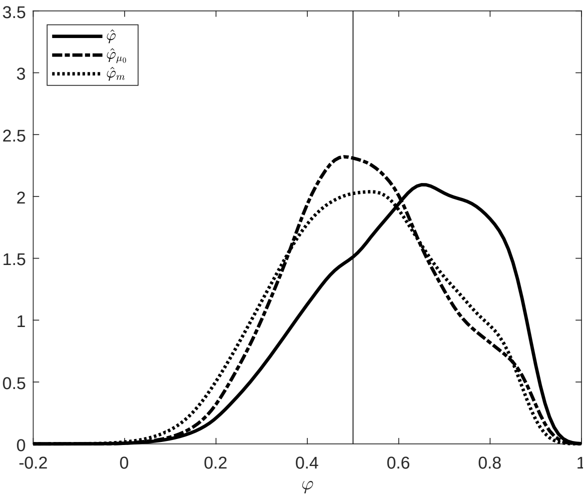
(a) Density plot of d estimates with $T = 32$



(b) Density plot of φ estimates with $T = 32$



(c) Density plot of d estimates with $T = 256$



(d) Density plot of φ estimates with $T = 256$

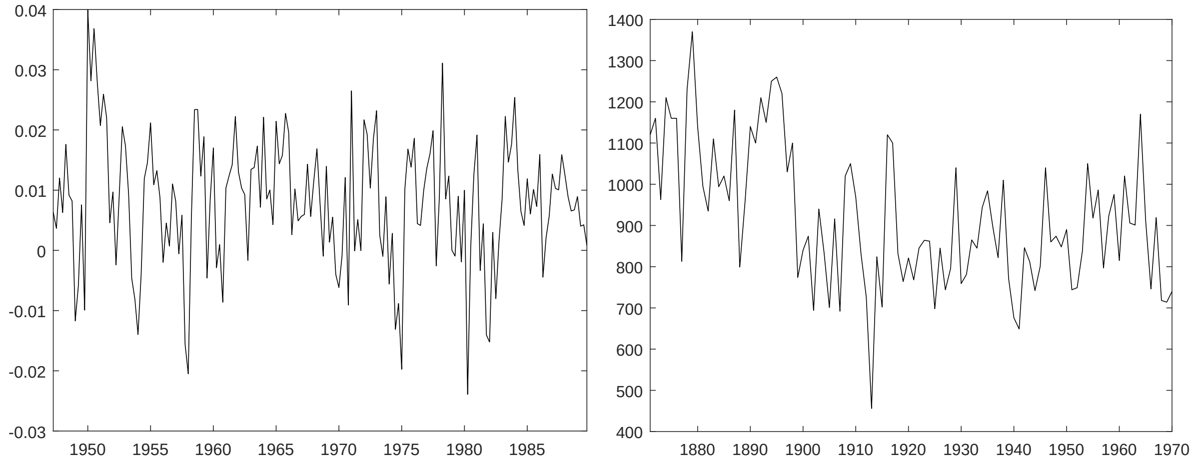
Figure 4: Density plots of the CSS estimator with unknown and known level parameter and of the MCSS estimator for the ARFIMA(1,d_0,0) model with $T = 32$ (upper panel) and 256 (lower panel) and $d_0 = -0.2$ and $\varphi_0 = 0.5$. The density estimates use a normal kernel.

4 Empirical examples

As an illustration of the results discussed in Sections 2 and 3, we now present three empirical applications, reconsidering the long-memory modelling of classical datasets: First, we examine the long-memory properties of U.S. post-Second World War real GNP. Secondly, we test for a unit root in the time series considered by Nelson & Plosser (1982). Last, we re-examine the issues of long memory and structural breaks in the well-known Nile data. What all three applications have in common is that the datasets consist of short time series of 79 to 171 observations each, warranting the use of our MCSS estimator to correct the small-sample bias of the received ML or CSS estimators[2].

4.1 Post-second World War real GNP

Sowell (1992) conducted a well-known empirical analysis of the long-memory behaviour of U.S. post-Second World War quarterly, seasonally adjusted, log real GNP. The data[3] comprise observations from 1947:2 to 1989:4 and are displayed in panel (a) of Figure 5.



(a) Post-Second World War real GNP data (b) Nile data

Figure 5: Panel (a) 171 quarterly observations on first differences of log quarterly U.S. real GNP for the time period 1947:2 to 1989:4, as in Sowell (1992). Panel (b) displays 100 annual observations of the volume of the Nile for the time period 1871 to 1970.

Sowell (1992) estimates an ARFIMA(3,d,2) type-I model of mean-adjusted first differences using full maximum likelihood (ML), basing the lag order on the Akaike information criterion. He obtains an estimated memory parameter of -0.59. However, Smith, Sowell & Zin (1997) assert that Sowell's results are substantially biased and especially the memory parameter is strongly underestimated. They propose a simulation-based bias correction of the profile maximum likelihood (BC-PML) estimator, resulting in $\hat{d} = -0.46$. The BC-PML estimator relies on the assumption that the bias is a linear function in the parameters. However, Lieberman & Phillips (2005)[4] show this not to be the case for a simple ARFIMA(0,d,0) type-I model. We circumvent this problem by using our MCSS estimator, which does not require the bias to be linear in the parameters. Table 8 presents the CSS and MCSS estimates of d for the ARFIMA(3,d,2) type-II model in (63), along with the ML estimate of Sowell (1992) and the profile maximum likelihood (PML) as well as BC-PML estimate of Smith, Sowell & Zin (1997). It can be noted, first, that the CSS

[2]The computations are again performed using MATLAB 2019a with the code to reproduce the empirical examples available on request.

[3]We use the data provided by Potter (1995) in the JAE Data Archive who mentions Citibase as his source, as does Sowell (1992). The dataset can be downloaded from `https://journaldata.zbw.eu/dataset/a-nonlinear-approach-to-us-gnp`.

[4]Lieberman & Phillips (2005) consider the profile plug-in maximum likelihood estimator instead of the profile maximum likelihood estimator for tractability reasons.

estimate is of similar order of magnitude as the maximum likelihood estimates, compare e.g. CSS and (P)ML. Secondly, the bias-correction increases both the CSS and PML estimates substantially, cf. MCSS and BC-PML. In fact, the CSS estimate is increased by a larger margin than the PML estimate. Thirdly, the type-II estimates are less significant than the type-I estimates, and the significance is reduced by the bias-correction. In conclusion, our results indicate that the long memory parameter is closer to zero than previously thought, even relative to its standard error.

	type-I			type-II	
	ML	PML	BC-PML	CSS	MCSS
$\hat{d}$	-0.59	-0.61	-0.46	-0.53	-0.26
SE	0.35	0.29	0.29	0.37	0.30

Table 8: The memory parameter estimates for the ARFIMA(3,d,2) model and their standard errors. The standard errors are calculated using the inverse of the empirical Hessian matrix.

4.2 Extended Nelson-Plosser dataset

There is a long-standing controversy on whether it is apt to describe the 14 time series in the well-known Nelson & Plosser (1982) dataset, as extended by Schotman & Van Dijk (1991)[5], by unit root processes. More recently, the literature on long memory processes has broadened the debate by considering a fractional integration parameter d that can take any value on the real line instead of merely zero or one. Yet the test statistics for the null hypothesis of $d = 1$ tend to be close to their critical values, impeding strong conclusions. Prominent papers are, amongst others, Crato & Rothman (1994), Gil-Alaña & Robinson (1997), Shimotsu (2010) and La Vecchia & Ronchetti (2019).

Our enquiry proceeds in two stages: First, we revisit Crato & Rothman (1994)[6] who use profile maximum likelihood (PML) to estimate an ARFIMA type-I model. We compare their PML to the CSS and MCSS estimates of d in our type-II setting, using either the model in (6) or (63). Secondly, we conduct unit root tests and relate them to the results obtained in the frequency-domain setting considered by Gil-Alaña & Robinson (1997) and Shimotsu (2010). This comparison is of interest because the MCSS estimator shares one interesting characteristic with frequency-domain estimators, namely that the leading bias of the estimator is not altered by an inclusion of a level parameter, a feature not present in PML or CSS.

[5]The dataset can be downloaded from `http://korora.econ.yale.edu/phillips/data/np&enp.dat` and is included in the R package 'tseries'.

[6]Unfortunately, we did not succeed in replicating the results of Crato & Rothman (1994). They use Sowell's Fortan program GQSTRFRAC, which is not available to us. Also, Hassler (2019, p. 110) mentions an error in the autocovariance formula of Sowell (1992, eq. (8)). We hence exercise caution in interpreting their results.

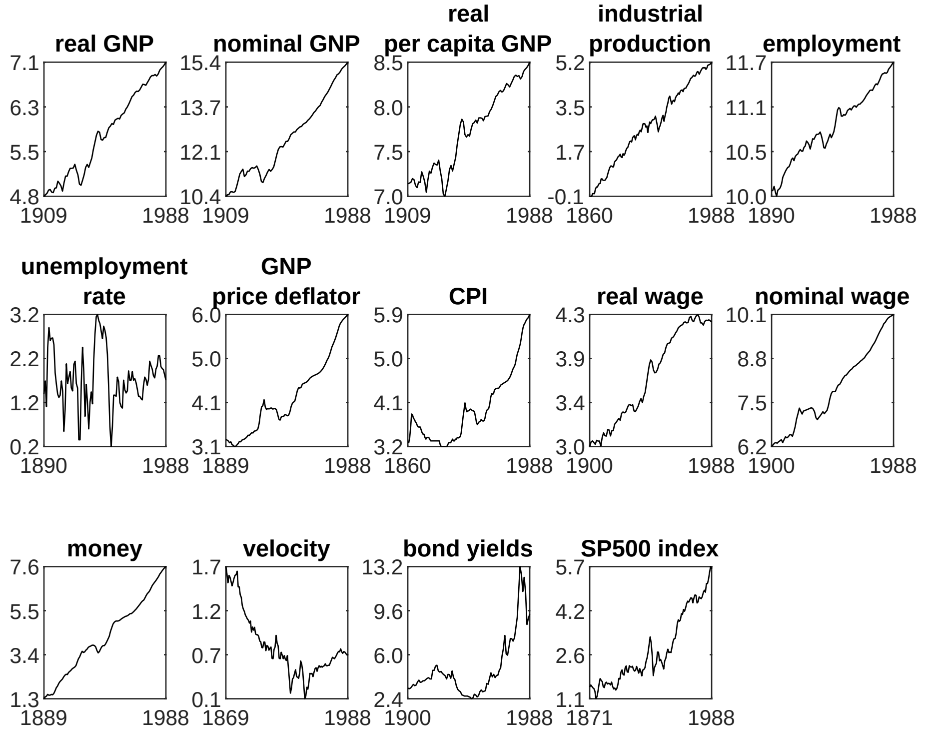


Figure 6: The extended Nelson-Plosser data in levels. All of the series are in logs, except for the bond yield.

			PML			CSS			MCSS		
series	T	BIC	$\hat{d}$	SE	t	$\hat{d}$	SE	t	$\hat{d}$	SE	t
real GNP	79	(1,d,0)	−0.41	0.21	−1.95	−0.43	0.21	−2.11	−0.32	0.23	−1.42
nominal GNP	79	(1,d,0)	−0.19	0.24	−0.80	−0.21	0.25	−0.85	−0.07	0.26	−0.27
real per capita GNP	79	(1,d,0)	−0.43	0.22	−1.96	−0.44	0.21	−2.10	−0.33	0.23	−1.41
industrial production	128	(1,d,0)	−0.64	0.33	−1.95	−0.59	0.24	−2.42	−0.46	0.21	−2.16
employment	98	(0,d,1)	−0.19	0.12	−1.60	−0.20	0.12	−1.65	−0.14	0.13	−1.09
unemployment rate	98	(0,d,1)	−0.58	0.11	−5.14	−0.57	0.11	−5.17	−0.52	0.11	−4.62
GNP price deflator	99	(1,d,0)	−0.39	0.21	−1.88	−0.40	0.20	−1.95	0.22	0.27	0.79
CPI	128	(0,d,1)	0.19	0.08	2.24	0.21	0.09	2.40	0.24	0.09	2.60
real wage	88	(0,d,0)	0.12	0.10	1.16	0.13	0.11	1.19	0.17	0.11	1.59
nominal wage	88	(1,d,0)	−0.21	0.25	−0.85	−0.23	0.25	−0.91	−0.07	0.28	−0.24
money	99	(1,d,1)	−0.50	0.22	−2.26	−0.52	0.21	−2.49	−0.44	0.26	−1.71
velocity	119	(0,d,0)	0.04	0.08	0.47	0.04	0.08	0.46	0.07	0.08	0.81
bond yields	88	(0,d,1)	−0.19	0.10	−1.81	−0.20	0.10	−1.92	−0.15	0.11	−1.42
SP500 index	117	(0,d,1)	−0.21	0.10	−2.21	−0.21	0.09	−2.21	−0.17	0.10	−1.76

Table 9: Estimated ARFIMA models of the extended Nelson-Plosser data. The time series are transformed into log-differences, merely bond yields are only in differences. The second column shows the length T of the individual series, the third column the model specifications based on the BIC for the profile maximum likelihood (PML) estimator. Subsequent columns then list the estimates of the memory parameter for PML, conditional sum-of-squares (CSS), and modified conditional sum-of-squares (MCSS). The empirical Hessian is used to calculate the standard errors, and the t-statistics are computed for the unit root null $H_0\colon d = 0$. The PML estimates are computed in R using the 'arfima' Package, see R Core Team (2023).

The extended Nelson-Plosser dataset consist of 14 annual macroeconomic series, starting between 1860 and 1909 and running to 1988, and are displayed in Figure 6. For the analysis, all of the series are log-differenced[7], except for the bond yield, which is merely in differences. Table 9 displays the PML, CSS and MCSS estimates of the memory parameter as well as their respective standard error and the t-statistics for testing the unit root null that $d = 0$. Following Crato & Rothman (1994), the model selection is based on the BIC[8] of the PML estimator. The table reveals that (a) four of the PML t-statistics are larger than the 5% critical values of a two-sided test, with a further five being borderline cases, (b) the MCSS estimates are consistently larger than the PML and CSS ones, and (c) of the MCSS t-statistics, only three lie in or close to the critical region. Another interesting point to note in Table 9 is that, for the GNP price deflator, PML and CSS provide a long memory estimate of -0.39 and -0.40, respectively, while the MCSS estimator yields a value of $+0.22$. This disparity may be attributed to the fact that CSS strongly underestimates the memory parameter when positive AR(1) dynamics are present whereas the MCSS estimator eliminates the bias, as demonstrated in Theorem 3.4 and the simulation study presented in Section 3.3. In summary, we find greater evidence than in the previous literature in favour of the unit root hypothesis in 11 out of the 14 Nelson-Plosser series.

Let us now turn to the second issue of interest, i.e. the comparison of our time-domain estimation to the frequency-domain approaches in Gil-Alaña & Robinson (1997) and Shimotsu (2010). For the unit root null hypothesis, Gil-Alaña & Robinson (1997) employ Robinson's (1994) LM-type test based on the Whittle (W) estimator, while Shimotsu (2010) uses a t-type statistic based on the extended local Whittle (ELW) objective function. Table 10 compares the results of the unit root test based on the frequency domain estimators W and ELW with those based on the time domain estimators in Table 9, a checkmark indicating that H_0 is (almost) rejected at the 5% level. Two important observations can be made from this table. First, the tests of Gil-Alaña & Robinson (1997) and Shimotsu (2010) give completely different outcomes, confirming the impression that there is presently no consensus in the literature on the unit root issue. A discussion of the relative merits of the W and ELW estimators is provided in, for instance, Hualde & Robinson (2011). Secondly, the test decisions of Gil-Alaña & Robinson (1997) are consistent with the majority of the MCSS tests. They only differ for real GNP, the unemployment rate and CPI, the reason for which could be that Gil-Alaña & Robinson (1997) capture the short-run dynamics solely through AR(k) components, which may be somewhat restrictive considering that the BIC also discovers MA lags.

[7] The "differencing and adding back" technique, a commonly used method to simplify estimation by removing drift through differencing, has been found to deliver inconsistent CSS estimates in type-II models when the data in levels exhibit a memory parameter of less than 0. As a solution to this problem, Hualde & Nielsen (2020) recommend modelling the data in levels instead of first-differences or, alternatively, employing a single dummy variable to capture the initial observation. Implementing this latter approach, our results remain qualitatively the same.

[8] Huang, Chan, Chen & Ing (2022) have recently shown the BIC criterion to provide consistent selection of the short-run dynamics when based on the CSS estimator in ARFIMA models without constant term.

	rejection of unit root hypothesis				
	time domain			frequency domain	
series	PML	CSS	MCSS	W	ELW
real GNP	(✓)	(✓)		✓	
nominal GNP					✓
real per capita GNP	(✓)	(✓)		(✓)	
industrial production	(✓)	✓	✓	✓	
employment					
unemployment rate	✓	✓	✓		
GNP price deflator	(✓)	(✓)			✓
CPI	✓	✓	✓		✓
real wage					
nominal wage					✓
money	✓	✓			✓
velocity					
bond yields	(✓)	(✓)			
SP500 index	✓	✓			

Table 10: Summary of the unit root tests, based on time-domain and frequency-domain estimators. W denotes the LM-type test of Gil-Alaña & Robinson (1997) based on the Whittle estimator, while ELW is the LM-type test of Shimotsu (2010) based on the extended local Whittle estimator. The presence of a checkmark shows that the null hypothesis of a unit root is rejected at a 5% significance level against a two-sided fractional alternative. A checkmark in parentheses means that the t-statistic is just outside the critical region.

4.3 Nile data

We now present an empirical application to the classical dataset[9] on the annual water flow volume of the Nile for the years 1871 to 1970. The 100 time-series observations are displayed in panel (b) of Figure 5. Several studies have analysed this dataset either in a long memory or short memory framework, with or without the presence of a break in the time series. Hosking (1984) and Boes, Davis & Gupta (1989) focus on long memory without considering a break. MacNeill, Tang & Jandhyala (1991), Wu & Zhao (2007), MacNeill, Jandhyala, Kaul & Fotopoulos (2020) examine breaks in a short memory time series context. Atkinson, Koopman & Shephard (1997) look at breaks in a unit root model. Shao (2011) and Betken (2017) address the testing and estimation of a break using a procedure that is robust to long memory although, after identifying a break, they do not proceed to estimating the fractional parameter. In summary, while there appears to be a consensus on including a break in the model, there is disagreement on whether the dynamics are better described by short or long memory. In particular, the literature currently does not consider the estimation of the memory parameter that is robust to a break. This is what we aim to achieve.

To that end, we proceed in two steps: First, we extend our model in (63) to incorporate

[9]The dataset used in this analysis can be obtained from the R package 'datasets'.

a break, i.e. μ in (63) is replaced by

$$\mu_t(\tau) = \mu + \beta I(t \leq \lfloor \tau T \rfloor), \tag{108}$$

where the break fraction $\tau \in (0,1)$ is assumed unknown. $\mu_t(\tau)$ can be consistently estimated in a type-II fractionally integrated model with $|d_0| < 1/2$, as shown by Chang & Perron (2016) and Iacone, Leybourne & Taylor (2019). It is, however, necessary to generalise our Assumption 3.1 such that $q > 1/(1+2d_0)$ moments exist, see Johansen & Nielsen (2012, Theorem 2). In a second step, we employ the filtered observations $\hat{x}_t = x_t - \hat{\mu}_t(\hat{\tau})$ to obtain the CSS estimates $\hat{\vartheta}$ in (74) and the MCSS estimate $\hat{\vartheta}_m$ in (89). The consistency of $\hat{\vartheta}$ in this model follows from similar arguments as in Robinson & Velasco (2015, Proposition 1), that of $\hat{\vartheta}_m$ in this model is easily obtained because of its asymptotic equivalence to the CSS estimator, see (91) in Lemma 3.1. The model selection procedure suggested by Hualde & Robinson (2011) is employed, consisting in a preliminary estimator $\tilde{d}$ of d obtained by local Whittle estimation as in Robinson (1995) before the procedure by Box & Jenkins (1990) is applied for selecting the short-run dynamics of $\Delta^{\tilde{d}}\{\hat{x}_t\}$. The Lobato & Robinson (1998) automatic selection rule of the bandwidth m is used.

In the first step, we find that $\hat{\tau} = 0.27$, translating into an estimated break in 1898. This is similar to what most of the aforementioned papers find, and it coincides with the beginning of the construction of the Lower Aswan Dam in 1899. The estimates of the level and break magnitudes of, respectively, $\hat{\mu}(\hat{\tau}) = 849.97$ and $\hat{\beta}(\hat{\tau}) = 247.78$ imply that the flow volume was reduced by 22%. Note that Hosking (1984) implements an alternative adjustment based on the recommendation of Todini & O'Connell (1979), namely that the pre-1903 flows are reduced by 8%.

In the second step, we find a bandwidth of $m = 22$, resulting in a preliminary estimate of $\tilde{d} = -0.05$. The Box-Jenkins procedure indicates that the short-run dynamics are best described by a MA(1) model. The resulting CSS and MCSS estimates are reported in Table 11, along with their standard errors and t-statistics. The results are unambiguous: the MCSS estimate does not provide evidence of long memory in the Nile data once the break is incorporated, with the point estimate of the memory parameter being -0.12. The CSS estimate supports this conclusion, with an estimate of -0.18. In terms of short-run dynamics, however, CSS and MCSS differ: CSS estimates the MA coefficient to be 0.30, an estimate that is statistically significant at the 5% level. On the other hand, the MCSS estimate of 0.26 is insignificant. Given the superior finite sample properties of MCSS, our conclusion is that after incorporating the break, the Nile data is characterised by IID shocks. This finding aligns with that of Atkinson, Koopman & Shephard (1997), supporting their argument that the series can be adequately described by a white noise process once the break is taken into consideration[10].

To corroborate our conclusion regarding the memory parameter, we employ the semiparametric t-type statistic in Iacone, Nielsen & Taylor (2022) to test the null hypothesis $H_0\colon d_0 = 0$ against the alternative hypothesis $H_1\colon d_0 \neq 0$. This test is designed to be robust against breaks and has the advantage that a parametric specification of the shocks is not needed. The test result, omitted to conserve space, is conclusive and supports our finding: after taking into account the break, there is no evidence that the Nile data

[10] Atkinson, Koopman & Shephard (1997) also identifies an outlier in the year 1913. However, even after removing this outlier, our results remain robust.

exhibits long memory.

Es	CSS estimate	SE	t	MCSS estimate	SE	t
d	-0.18	0.13	-1.44	-0.12	0.14	-0.81
φ	0.30	0.13	2.36	0.26	0.14	1.88

Table 11: CSS and MCSS estimates of the ARFIMA(0,d,1) model for the filtered observations $\hat{x}_t$ of the Nile data. The MA(1) coefficient is denoted by φ. The empirical Hessian matrix's inverse is used to calculate the standard errors.

5 Final comments

Practitioners like the CSS estimator due to its simplicity and effectiveness in estimating both stationary and non-stationary ARFIMA models. Recent work by Hualde & Nielsen (2020, 2022) provides the asymptotic justification for using the CSS estimator to estimate models that include deterministic components such as level and trend components. However, our analysis reveals that incorporating the level parameter into the model introduces an additional bias in the CSS estimator. This bias is due to a biased score which is particularly pronounced when the data is stationary. To address this issue, we propose modifying the CSS profile objective function to create an unbiased score, resulting in a new estimator which we call the modified CSS (MCSS) estimator. This new estimator is straightforward to compute and implement, enabling practitioners to obtain more accurate estimates and less distorted tests and confidence intervals. We illustrate the MCSS estimator by three classical empirical applications. Our analysis is for the general ARFIMA(p_1,d,p_2) model including a constant term. Various extensions are conceivable and of interest, yet beyond the scope of this paper:

First, further deterministic components could be included in the model, e.g. a linear time trend: Denoting by X a $T \times 2$ matrix of a constant and a linear trend then it can be shown that the modification term for the MCSS objective function turns out to be

$$m(d,\varphi) = \left|(\phi(L;\varphi)\Delta_+^d X)'(\phi(L;\varphi)\Delta_+^d X)\right|^{\frac{1}{T-2}}.$$

This modification term is again simple to calculate. Notably, it is equivalent to that in (88) if X is only a vector of ones and the degrees of freedom in the power term is replaced to $T-1$. We expect that the above modification term, corrected for the appropriate degrees of freedom, also holds for more general deterministic components in X. We conjecture that this MCSS estimator should improve on the CSS estimator and that its bias should be the same as that of the CSS estimator with known parameters of the deterministic components.

Secondly, while our paper focused solely on univariate fractional time series, the topic takes on added interest when extended to a panel setting. For instance, Robinson & Velasco (2015) extend the model presented in equations (63)-(64) to a panel framework. While the CSS estimator in a panel setting is consistent under large-T asymptotics, its finite sample properties are deficient due to the presence of fixed effects. To address this issue, the authors propose a bias correction that depends on the true parameters, necessitating the use of estimates to render this correction feasible. However, as the finite sample properties of the CSS estimator is unsatisfactory, substituting the true values with estimated ones leads to similarly non-optimal estimates. As an alternative to improving the small-sample properties of the CSS estimator a similar modification to the CSS objective can be made as in Section 3.1. The advantage of this approach is highlighted in the recent work of Schumann, Severini & Tripathi (2023).

References

Abramowitz, M. & I. A. Stegun (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Vol. 55. U.S. Government Printing Office.

Adenstedt, R. K. (1974). "On large-sample estimation for the mean of a stationary random sequence". *The Annals of Statistics* vol. 2, no. 6, pp. 1095–1107.

An, S. & P. Bloomfield (1993). *Cox and Reid's modification in regression models with correlated errors*. Mimeo, North Carolina State University Department of Statistics.

Atkinson, A. C., S.-J. Koopman & N. Shephard (1997). "Detecting shocks: Outliers and breaks in time series". *Journal of Econometrics* vol. 80, no. 2, pp. 387–422.

Barndorff-Nielsen, O. E. (1983). "On a formula for the distribution of the maximum likelihood estimator". *Biometrika* vol. 70, no. 2, pp. 343–365.

Bartolucci, F., R. Bellio, A. Salvan & N. Sartori (2016). "Modified profile likelihood for fixed-effects panel data models". *Econometric Reviews* vol. 35, no. 7, pp. 1271–1289.

Beran, J. (1995). "Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models". *Journal of the Royal Statistical Society: Series B* vol. 57, no. 4, pp. 659–672.

Betken, A. (2017). "Change point estimation based on Wilcoxon tests in the presence of long-range dependence". *Electronic Journal of Statistics* vol. 11, no. 2, pp. 3633–3672.

Bloomfield, P. (1973). "An exponential model for the spectrum of a scalar time series". *Biometrika* vol. 60, no. 2, pp. 217–226.

Boes, D. C., R. A. Davis & S. N. Gupta (1989). "Parameter estimation in low order fractionally differenced ARMA processes". *Stochastic Hydrology and Hydraulics* vol. 3, pp. 97–110.

Box, G. E. P. & G. Jenkins (1990). *Time Series Analysis, Forecasting and Control*. Holden-Day, Inc.

Chang, S. Y. & P. Perron (2016). "Inference on a structural break in trend with fractionally integrated errors". *Journal of Time Series Analysis* vol. 37, no. 4, pp. 555–574.

Cheung, Y.-W. & F. X. Diebold (1994). "On maximum likelihood estimation of the differencing parameter of fractionally-integrated noise with unknown mean". *Journal of Econometrics* vol. 62, no. 2, pp. 301–316.

Chung, C.-F. & R. T. Baillie (1993). "Small sample bias in conditional sum-of-squares estimators of fractionally integrated ARMA models". *Empirical Economics* vol. 18, no. 4, pp. 791–806.

Conniffe, D. (1987). "Expected maximum log likelihood estimation". *Journal of the Royal Statistical Society: Series D* vol. 36, no. 4, pp. 317–329.

Cox, D. R. & O. E. Barndorff-Nielsen (1994). *Inference and Asymptotics*. CRC Press.

Cox, D. R. & N. Reid (1987). "Parameter orthogonality and approximate conditional inference". *Journal of the Royal Statistical Society: Series B* vol. 49, no. 1, pp. 1–18.

Crato, N. & P. Rothman (1994). "Fractional integration analysis of long-run behavior for U.S. macroeconomic time series". *Economics Letters* vol. 45, no. 3, pp. 287–291.
Dahlhaus, R. (1989). "Efficient parameter estimation for self-similar processes". *The Annals of Statistics* vol. 17, no. 4, pp. 1749–1766.
Gil-Alaña, L. A. & P. M. Robinson (1997). "Testing of unit root and other nonstationary hypotheses in macroeconomic time series". *Journal of Econometrics* vol. 80, no. 2, pp. 241–268.
Gradshteyn, I. S. & I. M. Ryzhik (2014). *Table of Integrals, Series, and Products.* Academic Press.
Hannan, E. J. (1973). "The asymptotic theory of linear time-series models". *Journal of Applied Probability* vol. 10, no. 1, pp. 130–145.
Hassler, U. (2019). *Time Series Analysis with Long Memory in View.* John Wiley & Sons.
Horn, R. A. & C. R. Johnson (2013). *Matrix Analysis.* 2nd edition. Cambridge University Press.
Hosking, J. R. M. (1984). "Modeling persistence in hydrological time series using fractional differencing". *Water Resources Research* vol. 20, no. 12, pp. 1898–1908.
Hosking, J. R. M. (1981). "Fractional differencing". *Biometrika* vol. 68, no. 1, pp. 165–176.
Hualde, J. & M. Ø. Nielsen (2020). "Truncated sum of squares estimation of fractional time series models with deterministic trends". *Econometric Theory* vol. 36, no. 4, pp. 751–772.
Hualde, J. & M. Ø. Nielsen (2022). "Truncated sum-of-squares estimation of fractional time series models with generalized power law trend". *Electronic Journal of Statistics* vol. 16, no. 1, pp. 2884–2946.
Hualde, J. & M. Ø. Nielsen (2023). "Fractional Integration and Cointegration". *Oxford Research Encyclopedia of Economics and Finance.* Oxford University Press. DOI: 10.1093/acrefore/9780190625979.013.639.
Hualde, J. & P. M. Robinson (2011). "Gaussian pseudo-maximum likelihood estimation of fractional time series models". *The Annals of Statistics* vol. 39, no. 6, pp. 3152–3181.
Huang, H.-H., N. H. Chan, K. Chen & C.-K. Ing (2022). "Consistent order selection for ARFIMA processes". *The Annals of Statistics* vol. 50, no. 3, pp. 1297–1319.
Iacone, F., S. J. Leybourne & A. R. Taylor (2019). "Testing the order of fractional integration of a time series in the possible presence of a trend break at an unknown point". *Econometric Theory* vol. 35, no. 6, pp. 1201–1233.
Iacone, F., M. Ø. Nielsen & A. R. Taylor (2022). "Semiparametric tests for the order of integration in the possible presence of level breaks". *Journal of Business & Economic Statistics* vol. 40, no. 2, pp. 880–896.
Jensen, A. N. & M. Ø. Nielsen (2014). "A fast fractional difference algorithm". *Journal of Time Series Analysis* vol. 35, no. 5, pp. 428–436.
Johansen, S. (2008). "A representation theory for a class of vector autoregressive models for fractional processes". *Econometric Theory* vol. 24, no. 3, pp. 651–676.
Johansen, S. & M. Ø. Nielsen (2010). "Likelihood inference for a nonstationary fractional autoregressive model". *Journal of Econometrics* vol. 158, no. 1, pp. 51–66.
Johansen, S. & M. Ø. Nielsen (2012). "A necessary moment condition for the fractional functional central limit theorem". *Econometric Theory* vol. 28, no. 3, pp. 671–679.
Johansen, S. & M. Ø. Nielsen (2016). "The role of initial values in conditional sum-of-squares estimation of nonstationary fractional time series models". *Econometric Theory* vol. 32, no. 5, pp. 1095–1139.

Kalbfleisch, J. & D. Sprott (1973). “Marginal and conditional likelihoods”. *Sankhyā: The Indian Journal of Statistics, Series A* vol. 35, no. 3, pp. 311–328.

La Vecchia, D. & E. Ronchetti (2019). “Saddlepoint approximations for short and long memory time series: A frequency domain approach”. *Journal of Econometrics* vol. 213, no. 2, pp. 578–592.

Laskar, M. R. & M. L. King (1998). “Modified likelihood and related methods for handling nuisance parameters in the linear regression model”. Monash University Departement of Econometrics Working Paper, no. 5/98. DOI: 10.22004/ag.econ.267941.

Lawley, D. N. (1956). “A general method for approximating to the distribution of likelihood ratio criteria”. *Biometrika* vol. 43, no. 3/4, pp. 295–303.

Lee, Y. & P. C. Phillips (2015). “Model selection in the presence of incidental parameters”. *Journal of Econometrics* vol. 188, no. 2, pp. 474–489.

Li, W. K. & A. I. McLeod (1986). “Fractional time series modelling”. *Biometrika* vol. 73, no. 1, pp. 217–221.

Liang, K.-Y. (1987). “Estimating functions and approximate conditional likelihood”. *Biometrika* vol. 74, no. 4, pp. 695–702.

Liang, K.-Y. & S. L. Zeger (1995). “Inference based on estimating functions in the presence of nuisance parameters”. *Statistical Science* vol. 10, no. 2, pp. 158–173.

Lieberman, O. & P. C. Phillips (2004). “Expansions for the distribution of the maximum likelihood estimator of the fractional difference parameter”. *Econometric Theory* vol. 20, no. 3, pp. 464–484.

Lieberman, O. & P. C. Phillips (2005). “Expansions for approximate maximum likelihood estimators of the fractional difference parameter”. *The Econometrics Journal* vol. 8, no. 3, pp. 367–379.

Lobato, I. N. & P. M. Robinson (1998). “A nonparametric test for I(0)”. *The Review of Economic Studies* vol. 65, no. 3, pp. 475–495.

MacNeill, I., V. Jandhyala, A. Kaul & S. Fotopoulos (2020). “Multiple change-point models for time series”. *Environmetrics* vol. 31, no. 1, e2593.

MacNeill, I., S. Tang & V. Jandhyala (1991). “A search for the source of the Nile’s changepoints”. *Environmetrics* vol. 2, no. 3, pp. 341–375.

Marinucci, D. & P. M. Robinson (1999). “Alternative forms of fractional Brownian motion”. *Journal of Statistical Planning and Inference* vol. 80, no. 1-2, pp. 111–122.

Marinucci, D. & P. M. Robinson (2000). “Weak convergence of multivariate fractional processes”. *Stochastic Processes and their Applications* vol. 86, no. 1, pp. 103–120.

Marinucci, D. & P. M. Robinson (2001). “Semiparametric fractional cointegration analysis”. *Journal of Econometrics* vol. 105, no. 1, pp. 225–247.

Martellosio, F. & G. Hillier (2020). “Adjusted QMLE for the spatial autoregressive parameter”. *Journal of Econometrics* vol. 219, no. 2, pp. 488–506.

MathWorks Inc. (2019). *MATLAB version: 9.6.0 (R2019a).* Natick, Massachusetts. URL: https://www.mathworks.com.

McCullagh, P. & R. Tibshirani (1990). “A simple method for the adjustment of profile likelihoods”. *Journal of the Royal Statistical Society: Series B* vol. 52, no. 2, pp. 325–344.

Nelson, C. R. & C. R. Plosser (1982). “Trends and random walks in macroeconmic time series: some evidence and implications”. *Journal of Monetary Economics* vol. 10, no. 2, pp. 139–162.

Nielsen, M. Ø. (2004). “Efficient likelihood inference in nonstationary univariate models”. *Econometric Theory* vol. 20, no. 1, pp. 116–146.

Nielsen, M. Ø. (2015). “Asymptotics for the conditional-sum-of-squares estimator in multivariate fractional time-series models”. *Journal of Time Series Analysis* vol. 36, no. 2, pp. 154–188.

Nielsen, M. Ø. & P. H. Frederiksen (2005). “Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration”. *Econometric Reviews* vol. 24, no. 4, pp. 405–443.

Potter, S. M. (1995). “A nonlinear approach to U.S. GNP”. *Journal of Applied Econometrics* vol. 10, no. 2, pp. 109–125.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna. URL: `https://www.R-project.org/`.

Robinson, P. M. (1994). “Efficient tests of nonstationary hypotheses”. *Journal of the American Statistical Association* vol. 89, no. 428, pp. 1420–1437.

Robinson, P. M. (1995). “Gaussian semiparametric estimation of long range dependence”. *The Annals of Statistics* vol. 23, no. 5, pp. 1630–1661.

Robinson, P. M. (2005). “The distance between rival nonstationary fractional processes”. *Journal of Econometrics* vol. 128, no. 2, pp. 283–300.

Robinson, P. M. (2006). “Conditional-sum-of-squares estimation of models for stationary time series with long memory”. *Time Series and Related Topics: In Memory of Ching-Zong Wei*. Ed. by H.-C. Ho, C.-K. Ing & T. L. Lai. Institute of Mathematical Statistics, pp. 130–137.

Robinson, P. M. & J. Hualde (2003). “Cointegration in fractional systems with unknown integration orders”. *Econometrica* vol. 71, no. 6, pp. 1727–1766.

Robinson, P. M. & C. Velasco (2015). “Efficient inference on fractionally integrated panel data models with fixed effects”. *Journal of Econometrics* vol. 185, no. 2, pp. 435–452.

Schotman, P. C. & H. K. Van Dijk (1991). “On Bayesian routes to unit roots”. *Journal of Applied Econometrics* vol. 6, no. 4, pp. 387–401.

Schumann, M., T. A. Severini & G. Tripathi (2023). “The role of score and information bias in panel data likelihoods”. *Journal of Econometrics* vol. 235, no. 2, pp. 1215–1238.

Severini, T. A. (2000). *Likelihood Methods in Statistics*. Oxford University Press.

Shao, X. (2011). “A simple test of changes in mean in the possible presence of long-range dependence”. *Journal of Time Series Analysis* vol. 32, no. 6, pp. 598–606.

Shimotsu, K. (2010). “Exact local Whittle estimation of fractional integration with unknown mean and time trend”. *Econometric Theory* vol. 26, no. 2, pp. 501–540.

Shimotsu, K. & P. C. Phillips (2005). “Exact local Whittle estimation of fractional integration”. *The Annals of Statistics* vol. 33, no. 4, pp. 1890–1933.

Smith, A. A., F. Sowell & S. E. Zin (1997). “Fractional integration with drift: estimation in small samples”. *Empirical Economics* vol. 22, no. 1, pp. 103–116.

Sowell, F. (1992). “Modeling long-run behavior with the fractional ARIMA model”. *Journal of Monetary Economics* vol. 29, no. 2, pp. 277–302.

Tanaka, K. (1984). “An asymptotic expansion associated with the maximum likelihood estimators in ARMA models”. *Journal of the Royal Statistical Society: Series B* vol. 46, no. 1, pp. 58–67.

Tanaka, K. (1999). “The nonstationary fractional unit root”. *Econometric Theory* vol. 15, no. 4, pp. 549–582.

Todini, E. & P. E. O’Connell, eds. (1979). *Hydrological Simulation of Lake Nasser*. Vol. 1. Italy and Wallingford, Oxfordshire, United Kingdom: IBM Italia Scientific Centers and Institute of Hydrology.

Velasco, C. & P. M. Robinson (2000). “Whittle pseudo-maximum likelihood estimation for nonstationary time series”. *Journal of the American Statistical Association* vol. 95, no. 452, pp. 1229–1243.

Wu, W. B. & Z. Zhao (2007). “Inference of trends in time series”. *Journal of the Royal Statistical Society: Series B* vol. 69, no. 3, pp. 391–410.

Zygmund, A. (1977). *Trigonometric Series*. Cambridge University Press.

A Proof of the results in Section 3

In this appendix, we give the proofs of the results in Section 3, i.e. the general model in (63)-(64). The results outlined in this Section 2 are special cases of that in Section 3. As such, they are also implicitly covered in this appendix.

The setup of this appendix is as follows: In Appendix A.1, we find expressions for the first three derivatives of the profile objective functions, namely $L^*(\vartheta)$, $L^*_{\mu_0}(\vartheta)$ and $L^*_m(\vartheta)$, evaluated at $\vartheta = \vartheta_0$. Appendix A.2 presents some preliminary results that play a central role in approximating these derivatives. In Appendix A.3, we analyse the terms involved in the derivatives and conclude with an asymptotic approximation of the derivatives. This approximation is divided into two parts: the non-stationary region, i.e. $d_0 > 1/2$, detailed in Appendix A.3.1, and the stationary region, i.e. $d_0 < 1/2$, detailed in Appendix A.3.2. The decision to partition the analysis is rooted in the dependency of the convergence order of these terms on their respective regions. We exclude the boundary case $d_0 = 0.5$ as it would necessitate a separate analysis which is beyond the scope of the present paper. Lastly, in Appendix A.4, we present the proofs of the main results in Section 3.

A.1 Derivatives of the objective functions

We first analyse the residuals $\epsilon_t(d, \varphi, \mu) = \phi(L;\varphi)\Delta^d_+ x_t - \mu c_t(d, \varphi)$ for $t \geq 1$ and introduce some notations. We use a subscript zero to represent the true parameters. Clearly, inserting the DGP in (63) into the expression $\epsilon_t(d, \varphi, \mu)$ yields

$$
\begin{aligned}
\epsilon_t(d, \varphi, \mu) &= \phi(L;\varphi)\Delta^d_+(\mu_0 + \Delta^{-d_0}_+ u_t) - \mu c_t(d, \varphi) \\
&= \phi(L;\varphi)\Delta^{d-d_0}_+ u_t - c_t(d, \varphi)\,(\mu - \mu_0) \\
&= S^+_t(\vartheta) - c_t(\vartheta)\,(\mu - \mu_0)\,,
\end{aligned}
\tag{A.1}
$$

where the stochastic term $S^+_t(\vartheta)$ is defined as

$$
S^+_t(\vartheta) = \phi(L;\varphi)\Delta^{d-d_0}_+ u_t \tag{A.2}
$$

and the deterministic term $c_t(\vartheta)$, see (71), is defined as

$$
c_t(\vartheta) = \phi(L;\varphi)\Delta^d_+ I(t \geq 1) = \phi(L;\varphi)\kappa_{0t}(d) = \sum_{j=0}^{t-1} \phi_j(\varphi)\kappa_{0(t-j)}(d), \tag{A.3}
$$

where $\kappa_{0t}(d)$ is defined in (8).

The derivative of $\epsilon_t(d, \varphi, \mu)$ with respect to $i \in \{\vartheta_k, \vartheta_k\vartheta_j, \vartheta_k\vartheta_j, \vartheta_k\vartheta_j\vartheta_l\}$, for $k, j, l = 1, \ldots, p+1$, evaluated at $\vartheta = \vartheta_0$, are of the form

$$
D_i\epsilon_t(d_0, \varphi_0, \mu) = S^+_{it}(\vartheta_0) - c_{it}(\vartheta_0)\,(\mu - \mu_0)\,, \tag{A.4}
$$

where

$$
S^+_{it}(\vartheta_0) = D_i S^+_t(\vartheta_0), \tag{A.5}
$$

and

$$c_{it}(\vartheta_0) = D_i c_t(\vartheta_0). \tag{A.6}$$

Throughout the appendix, we simplify notation by suppressing the dependence on ϑ_0. For instance, we write S_{it}^+ instead of $S_{it}^+(\vartheta_0)$. We follow the following convention: the derivative of a function $f(x, y(x))$ with respect to x is written as $D_x f(x, y(x))$, and the partial derivative with respect to x is written as $f_x(x, y(x))$.

The following lemma provides simple analytic expressions for the derivatives of the stochastic terms given in (A.5).

Lemma A.1. *Assume that Assumption 3.4 holds, then*

$$S_t^+ = \epsilon_t, \tag{A.7}$$

$$S_{mt}^+ = (-1)^{m^*} \sum_{k=0}^{t-1} D_m \pi_k(0) \epsilon_{t-k}, \tag{A.8}$$

$$S_{zt}^+ = \sum_{i=1}^{t-1} b_{zi}(\varphi_0) \epsilon_{t-i}, \tag{A.9}$$

$$S_{dzt}^+ = -\sum_{i=2}^{t-1} h_{dzi}(\varphi_0) \epsilon_{t-i}, \tag{A.10}$$

where $m \in \{d, dd, ddd\}$ and m^* denotes the number of times $S_t^+(\vartheta)$ is differenced with respect to d and where $z \in \{\varphi_k, \varphi_k\varphi_j, \varphi_k\varphi_j\varphi_l\}$ for $k, j, l = 1, \ldots, p$ and

$$h_{dzi}(\varphi_0) = \sum_{s=1}^{i-1} (i-s)^{-1} b_{zs}(\varphi_0), \tag{A.11}$$

$$b_{zi}(\varphi_0) = \sum_{s=0}^{i-1} \omega_s(\varphi_0) D_z \phi_{i-s}(\varphi_0). \tag{A.12}$$

Also,

$$D_d \pi_j(0) = j^{-1} I(j \geq 1), \tag{A.13}$$

$$D_{dd} \pi_j(0) = 2j^{-1} a_{j-1} I(j \geq 2), \tag{A.14}$$

where

$$a_j = I(j \geq 1) \sum_{k=1}^{j} k^{-1}. \tag{A.15}$$

Proof of Lemma A.2. Define $z_t(\varphi) = \phi(L;\varphi) u_t I(t \geq 1)$ and evaluating this expression at $\varphi = \varphi_0$ results in $z_t(\varphi_0) = \epsilon_t - \sum_{j=t}^{\infty} \phi_j(\varphi_0) u_{t-j} = \epsilon_t$, which follows from Assumption 3.4. Moreover, by observing that $S_t^+(d, \varphi) = \Delta_+^{d-d_0} z_t(\varphi)$, which can be used to conclude the proof of (A.7) and (A.8). Additionally, note that $z_t(\varphi) = \phi(L;\varphi)\omega(L;\varphi_0)\phi(L;\varphi_0) u_t I(t \geq 1) = \phi(L;\varphi)\omega(L;\varphi_0)\epsilon_t I(t \geq 1)$, which can be used to establish (A.9). By employing similar arguments and considering that (A.13) and (A.14) are provided in Johansen & Nielsen (2016, Lemma A.4), the remaining expression follows. □

Next, we find expressions for the first three derivatives of $L^*(\vartheta)$, $L^*_{\mu 0}(\vartheta)$ and $L^*_m(\vartheta)$ and evaluate them at $\vartheta = \vartheta_0$. We present them in the same order.

Recall that $L^*(\vartheta)$ in (73) equals $L(\vartheta, \mu(\vartheta))$, where $L(\vartheta, \mu)$ is given by

$$L(\vartheta, \mu) = \frac{1}{2} \sum_{t=1}^{T} \Big(\phi(L; \varphi) \Delta_+^d x_t - \mu c_t(d, \varphi) \Big)^2 ,$$

and $\mu(\vartheta) = \hat{\mu}(\vartheta)$ is given in (72). The first derivative of $L^*(\vartheta)$ with respect to ϑ_k equals

$$D_{\vartheta_k} L^*(\vartheta) = L_{\vartheta_k}(\vartheta, \mu(\vartheta)) + L_\mu(\vartheta, \mu(\vartheta)) \mu_{\vartheta_k}(\vartheta).$$

We simplify this expression by noticing that $\hat{\mu}(\vartheta)$ is determined from $L_\mu(\vartheta, \mu(\vartheta)) = 0$ such that

$$D_{\vartheta_k} L^*(\vartheta) = L_{\vartheta_k}(\vartheta, \mu(\vartheta)). \tag{A.16}$$

Next, we take the derivative of (A.16) with respect to ϑ_j to get an expression for $D_{\vartheta_k \vartheta_j} L^*(\vartheta)$. Using the chain rule we have that

$$D_{\vartheta_k \vartheta_j} L^*(\vartheta) = L_{\vartheta_k \vartheta_j}(\vartheta, \mu(\vartheta)) + L_{\vartheta_k \mu}(\vartheta, \mu(\vartheta)) \mu_{\vartheta_j}(\vartheta). \tag{A.17}$$

Taking on both sides the derivative with respect to ϑ_j of $L_\mu(\vartheta, \mu(\vartheta)) = 0$ implies $L_{\vartheta_j \mu}(\vartheta, \mu(\vartheta)) + L_{\mu\mu}(\vartheta, \mu(\vartheta)) \mu_{\vartheta_j}(\vartheta) = 0$ such that

$$\mu_{\vartheta_j}(\vartheta) = - \frac{L_{\vartheta_j \mu}(\vartheta, \mu(\vartheta))}{L_{\mu\mu}(\vartheta, \mu(\vartheta))}.$$

Lastly, we take the derivative of (A.17) with respect to ϑ_l to get an expression for $D_{\vartheta_k \vartheta_j \vartheta_l} L^*(\vartheta)$. We get that

$$\begin{aligned} D_{\vartheta_k \vartheta_j \vartheta_l} L^*(\vartheta) =& L_{\vartheta_k \vartheta_j \vartheta_l}(\vartheta, \mu(\vartheta)) + L_{\vartheta_k \vartheta_j \mu}(\vartheta, \mu(\vartheta)) \mu_{\vartheta_l}(\vartheta) + L_{\vartheta_k \vartheta_l \mu}(\vartheta, \mu(\vartheta)) \mu_{\vartheta_j}(\vartheta) \\ &+ L_{\vartheta_k \mu\mu}(\vartheta, \mu(\vartheta)) \mu_{\vartheta_l}(\vartheta) \mu_{\vartheta_j}(\vartheta) + L_{\vartheta_k \mu}(\vartheta, \mu(\vartheta)) \mu_{\vartheta_j \vartheta_l}(\vartheta). \end{aligned} \tag{A.18}$$

An expression for $\mu_{\vartheta_j \vartheta_l}(\vartheta)$ can then be easily found by taking on both sides the derivative with respect to ϑ_l of $L_{\vartheta_j \mu}(\vartheta, \mu(\vartheta)) + L_{\mu\mu}(\vartheta, \mu(\vartheta)) \mu_{\vartheta_j}(\vartheta) = 0$. We find that

$$\begin{aligned} 0 =& L_{\vartheta_j \vartheta_l \mu}(\vartheta, \mu(\vartheta)) + L_{\vartheta_j \mu\mu}(\vartheta, \mu(\vartheta)) \mu_{\vartheta_l}(\vartheta) \\ &+ \left(L_{\vartheta_l \mu\mu}(\vartheta, \mu(\vartheta)) + L_{\mu\mu\mu}(\vartheta, \mu(\vartheta)) \mu_{\vartheta_l}(\vartheta) \right) \mu_{\vartheta_j}(\vartheta) + L_{\mu\mu}(\vartheta, \mu(\vartheta)) \mu_{\vartheta_j \vartheta_l}(\vartheta), \end{aligned}$$

and by rewriting

$$\begin{aligned} \mu_{\vartheta_j \vartheta_l}(\vartheta) =& - \frac{L_{\vartheta_j \vartheta_l \mu}(\vartheta, \mu(\vartheta))}{L_{\mu\mu}(\vartheta, \mu(\vartheta))} - \mu_{\vartheta_l}(\vartheta) \frac{L_{\vartheta_j \mu\mu}(\vartheta, \mu(\vartheta))}{L_{\mu\mu}(\vartheta, \mu(\vartheta))} - \mu_{\vartheta_j}(\vartheta) \frac{L_{\vartheta_l \mu\mu}(\vartheta, \mu(\vartheta))}{L_{\mu\mu}(\vartheta, \mu(\vartheta))} \\ &- \mu_{\vartheta_j}(\vartheta) \mu_{\vartheta_l}(\vartheta) \frac{L_{\mu\mu\mu}(\vartheta, \mu(\vartheta))}{L_{\mu\mu}(\vartheta, \mu(\vartheta))}. \end{aligned}$$

Next, we find expressions for the first three derivatives of $L^*(\vartheta)$, given in (73), and present them in the following lemma. The derivatives are evaluated at $\vartheta = \vartheta_0$, and recall that we omit the explicit dependence.

Lemma A.2. *Let the model for the data x_t, $t = 1,\ldots,T$, be given by (63) and assume that Assumption 3.4 holds. Then the derivatives of $L^*(\vartheta)$, see (73), evaluated at $\vartheta = \vartheta_0$ are given by*

$$
\begin{aligned}
D_{\vartheta_k} L^* &= L_{\vartheta_k}, && \text{(A.19)}\\
D_{\vartheta_k\vartheta_j} L^* &= L_{\vartheta_k\vartheta_j} + L_{\vartheta_k\mu}\mu_{\vartheta_j}, && \text{(A.20)}\\
D_{\vartheta_k\vartheta_j\vartheta_l} L^* &= L_{\vartheta_k\vartheta_j\vartheta_l} + L_{\vartheta_k\vartheta_j\mu}\mu_{\vartheta_l} + L_{\vartheta_k\vartheta_l\mu}\mu_{\vartheta_j}\\
&\quad + L_{\vartheta_k\mu\mu}\mu_{\vartheta_l}\mu_{\vartheta_j} + L_{\vartheta_k\mu}\mu_{\vartheta_j\vartheta_l}, && \text{(A.21)}
\end{aligned}
$$

where

$$
\begin{aligned}
\mu_{\vartheta_j} &= -\frac{L_{\vartheta_j\mu}}{L_{\mu\mu}},\\
\mu_{\vartheta_j\vartheta_l} &= -\frac{L_{\vartheta_j\vartheta_l\mu}}{L_{\mu\mu}} - \mu_{\vartheta_l}\frac{L_{\vartheta_j\mu\mu}}{L_{\mu\mu}} - \mu_{\vartheta_j}\frac{L_{\vartheta_l\mu\mu}}{L_{\mu\mu}} - \mu_{\vartheta_j}\mu_{\vartheta_l}\frac{L_{\mu\mu\mu}}{L_{\mu\mu}}.
\end{aligned}
$$

for $k, j, l = 1, \ldots, p+1$. The partial derivatives of $L(\vartheta, \mu(\vartheta))$ evaluated at $\vartheta = \vartheta_0$ can be expressed as

$$
\begin{aligned}
L_{\vartheta_k} &= \sum_{t=1}^{T}\left(S_t^+ - c_t\left(\mu(\vartheta_0) - \mu_0\right)\right)\left(S_{\vartheta_k t}^+ - c_{\vartheta_k t}\left(\mu(\vartheta_0) - \mu_0\right)\right),\\
L_{\vartheta_k\vartheta_j} &= \sum_{t=1}^{T}\left(S_{\vartheta_j t}^+ - c_{\vartheta_j t}\left(\mu(\vartheta_0) - \mu_0\right)\right)\left(S_{\vartheta_k t}^+ - c_{\vartheta_k t}(\vartheta_0)\left(\mu(\vartheta_0) - \mu_0\right)\right)\\
&\quad + \sum_{t=1}^{T}\left(S_t^+ - c_t\left(\mu(\vartheta_0) - \mu_0\right)\right)\left(S_{\vartheta_k\vartheta_j t}^+ - c_{\vartheta_k\vartheta_j t}\left(\mu(\vartheta_0) - \mu_0\right)\right),\\
L_{\vartheta_k\vartheta_j\vartheta_l} &= \sum_{t=1}^{T}\left(S_{\vartheta_j\vartheta_l t}^+ - c_{\vartheta_j\vartheta_l t}\left(\mu(\vartheta_0) - \mu_0\right)\right)\left(S_{\vartheta_k t}^+ - c_{\vartheta_k t}\left(\mu(\vartheta_0) - \mu_0\right)\right)\\
&\quad + \sum_{t=1}^{T}\left(S_{\vartheta_j}^+ - c_{\vartheta_j t}\left(\mu(\vartheta_0) - \mu_0\right)\right)\left(S_{\vartheta_k\vartheta_l t}^+ - c_{\vartheta_k\vartheta_l t}\left(\mu(\vartheta_0) - \mu_0\right)\right)\\
&\quad + \sum_{t=1}^{T}\left(S_{\vartheta_l t}^+ - c_{\vartheta_l t}\left(\mu(\vartheta_0) - \mu_0\right)\right)\left(S_{\vartheta_k\vartheta_j t}^+ - c_{\vartheta_k\vartheta_j t}\left(\mu(\vartheta_0) - \mu_0\right)\right)\\
&\quad + \sum_{t=1}^{T}\left(S_t^+ - c_t\left(\mu(\vartheta_0) - \mu_0\right)\right)\left(S_{\vartheta_k\vartheta_j\vartheta_l t}^+ - c_{\vartheta_k\vartheta_j\vartheta_l t}\left(\mu(\vartheta_0) - \mu_0\right)\right),\\
L_{\vartheta_k\mu} &= -\sum_{t=1}^{T} c_t\left(S_{\vartheta_k t}^+ - c_{\vartheta_k t}\left(\mu(\vartheta_0) - \mu_0\right)\right)\\
&\quad - \sum_{t=1}^{T}\left(S_{0t}^+ - c_t\left(\mu(\vartheta_0) - \mu_0\right)\right)c_{\vartheta_k t},\\
L_{\vartheta_k\mu\mu} &= 2\sum_{t=1}^{T} c_t c_{\vartheta_k t},\\
L_{\vartheta_k\vartheta_j\mu} &= -\sum_{t=1}^{T} c_{\vartheta_j t}\left(S_{\vartheta_k t}^+ - c_{\vartheta_k t}\left(\mu(\vartheta_0) - \mu_0\right)\right)\\
&\quad - \sum_{t=1}^{T} c_t\left(S_{\vartheta_k\vartheta_j t}^+ - c_{\vartheta_k\vartheta_j t}\left(\mu(\vartheta_0) - \mu_0\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
&- \sum_{t=1}^{T} \left(S^{+}_{\vartheta_j t} - c_{\vartheta_j t} \left(\mu(\vartheta_0) - \mu_0 \right) \right) c_{\vartheta_k t} \\
&- \sum_{t=1}^{T} \left(S^{+}_{t} - c_{t} \left(\mu(\vartheta_0) - \mu_0 \right) \right) c_{\vartheta_k \vartheta_j t}, \\
L_{\mu} &= - \sum_{t=1}^{T} \left(S^{+}_{t} - c_{t} \left(\mu(\vartheta_0) - \mu_0 \right) \right) c_{t}, \\
L_{\mu\mu} &= \sum_{t=1}^{T} c_t^2, \\
L_{\mu\mu\mu} &= 0.
\end{aligned}
$$

Here, $\mu(\vartheta_0) = \hat{\mu}(\vartheta_0)$ and the stochastic term S_t^{+} is defined in (A.2) *and its derivatives are given in* (A.5)*. The deterministic term c_t is defined in* (A.3) *and its derivatives are given in* (A.6)*.*

Proof of Lemma A.2. The proof of (A.19), (A.20) and (A.21) is given in (A.16),(A.17) and (A.18), respectively. The partial derivatives of $L(\vartheta, \mu(\vartheta))$ follow from the relationship

$$
L(\vartheta, \mu) = \frac{1}{2} \sum_{t=1}^{T} \epsilon_t^2(d, \varphi, \mu),
$$

where $\epsilon_t(d, \varphi, \mu)$ is given in (A.1) and its derivatives are provided in (A.4). The proof follows easily by using these derivatives. □

Next, we find expressions for the first three derivatives of $L^{*}_{\mu_0}(\vartheta)$, given in (77), and evaluate them at $\vartheta = \vartheta_0$.

Lemma A.3. *Let the model for the data x_t, $t = 1,\ldots,T$, be given by* (63) *and assume that Assumption 3.4 holds. Then the derivatives of $L^{*}_{\mu_0}(\vartheta)$, see* (77)*, evaluated at $\vartheta = \vartheta_0$ are given by*

$$
D_{\vartheta_k} L^{*}_{\mu_0} = \sum_{t=1}^{T} S_t^{+} S^{+}_{\vartheta_k t}, \tag{A.22}
$$

$$
D_{\vartheta_k \vartheta_j} L^{*}_{\mu_0} = \sum_{t=1}^{T} S^{+}_{\vartheta_j t} S^{+}_{\vartheta_k t} + \sum_{t=1}^{T} S_t^{+} S^{+}_{\vartheta_k \vartheta_j t}, \tag{A.23}
$$

$$
D_{\vartheta_k \vartheta_j \vartheta_l} L^{*}_{\mu_0} = \sum_{t=1}^{T} S_t^{+} S^{+}_{\vartheta_k \vartheta_j \vartheta_l t} + \sum_{t=1}^{T} S^{+}_{\vartheta_j \vartheta_l t} S^{+}_{\vartheta_k t} + \sum_{t=1}^{T} S^{+}_{\vartheta_j} S^{+}_{\vartheta_k \vartheta_l t} + \sum_{t=1}^{T} S^{+}_{\vartheta_l t} S^{+}_{\vartheta_k \vartheta_j t}, \tag{A.24}
$$

for $k, j = 1, \ldots, p+1$. Here, the stochastic term S_t^{+} is defined in (A.2)*, and its derivatives are given in* (A.5)*.*

Proof of Lemma A.3. Recall that

$$
L^{*}_{\mu_0}(\vartheta) = \frac{1}{2} \sum_{t=1}^{T} \epsilon_t^2(d, \varphi, \mu_0),
$$

where $\epsilon_t(d, \varphi, \mu)$ is given in (A.1). The second term in (A.1) becomes zero when μ is equal to μ_0. This simplifies the proof, which can now be easily derived using (A.4). □

Finally, we find expressions for the first three derivatives of $L_m^*(\vartheta)$, see (83), and evaluate them $\vartheta = \vartheta_0$.

Lemma A.4. *Let the model for the data x_t, $t = 1,\ldots,T$, be given by* (63) *and assume that the assumption 3.4 holds. Then the derivatives of $L_m^*(\vartheta)$, given in* (83)*, evaluated at $\vartheta = \vartheta_0$ are given by*

$$D_{\vartheta_k} L^* = m D_{\vartheta_k} L^* + m_{\vartheta_k} L^*, \tag{A.25}$$

$$D_{\vartheta_k\vartheta_j} L^* = m D_{\vartheta_k\vartheta_j} L^* + m_{\vartheta_j} D_{\vartheta_k} L^* + m_{\vartheta_k\vartheta_j} L^* + m_{\vartheta_k} D_{\vartheta_j} L^*, \tag{A.26}$$

$$\begin{aligned} D_{\vartheta_k\vartheta_j\vartheta_l} L^* = {} & m D_{\vartheta_k\vartheta_j\vartheta_l} L^* + m_{\vartheta_l} D_{\vartheta_k\vartheta_j} L^* + m_{\vartheta_j\vartheta_l} D_{\vartheta_k} L^* + m_{\vartheta_j} D_{\vartheta_k\vartheta_l} L^* \\ & + m_{\vartheta_k\vartheta_j\vartheta_l} L^* + m_{\vartheta_k\vartheta_j} D_{\vartheta_l} L^* + m_{\vartheta_k\vartheta_l} D_{\vartheta_j} L^* + m_{\vartheta_k} D_{\vartheta_j\vartheta_l} L^*, \end{aligned} \tag{A.27}$$

where expression for the derivatives of L^* are given in Lemma A.2 and the modification term $m(\vartheta)$ is given in (88) *and the derivatives of $m(\vartheta)$, evaluated at $\vartheta = \vartheta_0$, are given by*

$$\begin{aligned}
m &= \left(\sum_{t=1}^T c_t^2\right)^{\frac{1}{T-1}}, \\
m_{\vartheta_k} &= \frac{2}{T-1}\left(\sum_{t=1}^T c_t^2\right)^{-\frac{T-2}{T-1}} \sum_{t=1}^T c_t c_{\vartheta_k t}, \\
m_{\vartheta_k\vartheta_j} &= \frac{2}{T-1}\left(\sum_{t=1}^T c_t^2\right)^{-\frac{T-2}{T-1}} \sum_{t=1}^T \left(c_{\vartheta_j t} c_{\vartheta_k t} + c_t c_{\vartheta_k\vartheta_j t}\right) \\
&\quad - 4\frac{T-2}{(T-1)^2}\left(\sum_{t=1}^T c_t^2\right)^{-\frac{2T-3}{T-1}} \sum_{t=1}^T c_t c_{\vartheta_k t} \sum_{t=1}^T c_t c_{\vartheta_j t}, \\
m_{\vartheta_k\vartheta_j\vartheta_l} &= \frac{2}{T-1}\left(\sum_{t=1}^T c_t^2\right)^{-\frac{T-2}{T-1}} \sum_{t=1}^T \left(c_{\vartheta_j\vartheta_l t} c_{\vartheta_k t} + c_{\vartheta_j t} c_{\vartheta_k\vartheta_l t} + c_{\vartheta_l t} c_{\vartheta_k\vartheta_j t} + c_t c_{\vartheta_k\vartheta_j\vartheta_l t}\right) \\
&\quad - 4\frac{T-2}{(T-1)^2}\left(\sum_{t=1}^T c_t^2\right)^{-\frac{2T-3}{T-1}} \sum_{t=1}^T c_t c_{\vartheta_l t} \sum_{t=1}^T \left(c_{\vartheta_j t} c_{\vartheta_k t} + c_t c_{\vartheta_k\vartheta_j t}\right) \\
&\quad - 4\frac{T-2}{(T-1)^2}\left(\sum_{t=1}^T c_t^2\right)^{-\frac{2T-3}{T-1}} \left(\left(\sum_{t=1}^T c_{\vartheta_l t} c_{\vartheta_k t} + \sum_{t=1}^T c_t c_{\vartheta_k\vartheta_l t}\right) \sum_{t=1}^T c_t c_{\vartheta_j t}\right. \\
&\quad \left. + \sum_{t=1}^T c_t c_{\vartheta_k t} \left(\sum_{t=1}^T c_{\vartheta_l t} c_{\vartheta_j t} + \sum_{t=1}^T c_t c_{\vartheta_j\vartheta_l t}\right)\right) \\
&\quad + 8\frac{(T-2)(2T-3)}{(T-1)^3}\left(\sum_{t=1}^T c_t^2\right)^{-\frac{3T-4}{T-1}} \sum_{t=1}^T c_t c_{\vartheta_l t} \sum_{t=1}^T c_t c_{\vartheta_k t} \sum_{t=1}^T c_t c_{\vartheta_j t}.
\end{aligned}$$

Proof of Lemma A.4. Proof is straightforward due to the multiplicative form of the MCSS objective function, see (83). □

A.2 Preliminary results

In this section, we present findings that play a central role in the approximation of the derivatives. Appendix A.2.1 contains usefull bounds, while Appendix A.2.2 presents results related to fractional coefficients in (3), and their derivatives, and the weights of the lag polynomial in (65) and the inverse of the lag polynomial in (69) along with their derivatives. In Appendix A.2.3, we investigate the limiting behavior of the centered product moments, which are particularly relevant in the later expression of the biases. Lastly, Appendix A.2.4 focuses on the expectation of the CSS score function, which is a part of the bias term in the CSS estimator.

A.2.1 Useful bounds

In this section, we provide some general results that are useful for finding the approximation of the derivatives. We sometimes apply them in the remainder without special reference.

Lemma A.5. *For any $d > -1$, as $T \to \infty$,*

$$\frac{1}{T^{d+1}} \sum_{t=1}^{T} t^d \to \frac{1}{d+1}. \tag{A.28}$$

Proof of Lemma A.5. See Hualde & Nielsen (2020, Lemma S.10). □

Next, we present some useful bounds that are frequently used in the remainder of the appendix.

Lemma A.6. *For $m \geq 0$ and $c < \infty$,*

$$\sum_{n=1}^{N} (1 + \log(n))^m n^\alpha \leq c(1 + \log(N))^m N^{\alpha+1} \text{ if } \alpha > -1, \tag{A.29}$$

$$\sum_{n=N}^{\infty} (1 + \log(n))^m n^\alpha \leq c(1 + \log(N))^m N^{\alpha+1} \text{ if } \alpha < -1. \tag{A.30}$$

For $\alpha < 0$, and any β it holds that

$$\sum_{n=1}^{t-1} n^{\alpha-1}(t-n)^{\beta-1} \leq ct^{\max(\alpha-1,\beta-1)}. \tag{A.31}$$

For $\alpha \geq 0$, and any β it holds that

$$\sum_{n=1}^{t-1} n^{\alpha-1}(t-n)^{\beta-1} \leq c(1 + \log(t))t^{\max(\alpha+\beta-1,\alpha-1,\beta-1)}. \tag{A.32}$$

For $\alpha + \beta < 1$ and $\beta > 0$ it holds that

$$\sum_{k=1}^{\infty} (k+h)^{\alpha-1} k^{\beta-1} (1 + \log(k+h))^n \leq ch^{\alpha+\beta-1}(1 + \log(h))^n. \tag{A.33}$$

Proof of Lemma A.6. Proof of (A.29) and (A.30): See Johansen & Nielsen (2016, Lemma A.1).

Proof of (A.31): The proof follows a similar approach to the proof in Hualde & Robinson (2011, Lemma 1). Cleary,

$$\begin{aligned}\sum_{n=1}^{t-1} n^{\alpha-1}(t-n)^{\beta-1} &\leq c\sum_{n=1}^{\lfloor t/2\rfloor} n^{\alpha-1}(t-n)^{\beta-1} + c\sum_{n=\lfloor t/2\rfloor}^{t-1} n^{\alpha-1}(t-n)^{\beta-1}\\ &\leq ct^{\beta-1}\sum_{n=1}^{\lfloor t/2\rfloor} n^{\alpha-1} + ct^{\alpha-1}\sum_{n=\lfloor t/2\rfloor}^{t-1}(t-n)^{\beta-1},\end{aligned}$$

because $\alpha < 0$ the first summand is $O(1)$ and the second summand is $O(1)$ if $\beta < 0$, $O(\log(t))$ if $\beta = 0$, and $O(t^{\beta})$ if $\beta > 0$.
Proof of (A.32) and (A.33): See Johansen & Nielsen (2016, Lemma A.5).

□

A.2.2 Bounds for the (derivates of) fractional coefficients and short-run dynamics

Next, we present findings concerning the fractional coefficients in (3) and their derivatives, as well as the weights of the lag polynomial in (65) and the weights of the inverse of the lag polynomial in (69) and their derivatives.

Lemma A.7. *For $m \geq 0$ and $j \geq 1$ it holds that*

$$|D^m\pi_j(u)| \leq c(1+\log(j))^m j^{u-1} \tag{A.34}$$

Under Assumption 3.2 and 3.3 it follows, as $j \to \infty$,

$$\sup_{\varphi\in\Phi}|\omega_j(\varphi)| = O(j^{-1-\varsigma}), \tag{A.35}$$

$$\sup_{\varphi\in\Phi}|\phi_j(\varphi)| = O(j^{-1-\varsigma}), \tag{A.36}$$

$$\sup_{\varphi\in\Phi}\left|\frac{\partial\phi_j(\varphi)}{\partial\varphi_i}\right| = O(j^{-1-\varsigma}), \tag{A.37}$$

$$\sup_{\varphi\in\Phi}\left|\frac{\partial^2\phi_j(\varphi)}{\partial\varphi_i\partial\varphi_l}\right| = O(j^{-1-\varsigma}), \tag{A.38}$$

$$\sup_{\varphi\in\Phi}\left|\frac{\partial^3\phi_j(\varphi)}{\partial\varphi_i\partial\varphi_l\partial\varphi_k}\right| = O(j^{-1-\varsigma}), \tag{A.39}$$

for $i, l, k = 1, \ldots, p$ and where $1/2 < \varsigma \leq 1$

Proof of Lemma A.7. Proof of (A.34): See Johansen & Nielsen (2016, Lemma A.3)
Proof of (A.35)-(A.39): See Zygmund (1977, page 46) and Hualde & Robinson (2011, page 3155 and page 3169). □

Next, we present bounds for the deterministic term in (A.3) and their derivatives and the terms in (A.11) and (A.12).

Lemma A.8. *For any integer $m \geq 0$ and under Assumption 3.2 and 3.3,*

$$\left|\frac{\partial^m c_t(\vartheta)}{\partial d^m}\right| = O(t^{\max(-d,-1-\varsigma)}\log^m(t)), \tag{A.40}$$

$$\left|\frac{\partial^{m+1} c_t(\vartheta)}{\partial d^m \partial\varphi_i(\varphi)}\right| = O(t^{\max(-d,-1-\varsigma)}\log^m(t)), \tag{A.41}$$

$$\left|\frac{\partial^{m+2} c_t(\vartheta)}{\partial d^m \partial\varphi_i\partial\varphi_l}\right| = O(t^{\max(-d,-1-\varsigma)}\log^m(t)), \tag{A.42}$$

$$\left|\frac{\partial^{m+3} c_t(\vartheta)}{\partial d^m \partial\varphi_i\partial\varphi_l\partial\varphi_k}\right| = O(t^{\max(-d,-1-\varsigma)}\log^m(t)), \tag{A.43}$$

for $i, l, k = 1, \ldots, p$ and where $1/2 < \varsigma \leq 1$. Also, for $j \to \infty$ it holds that

$$|h_{dzj}(\varphi_0)| = O(j^{-1}), \tag{A.44}$$

$$|b_{zj}(\varphi_0)| = O(j^{-1-\varsigma}), \tag{A.45}$$

where $h_{dzi}(\varphi_0)$ and $b_{zi}(\varphi_0)$ are, respectively, defined in (A.11) *and* (A.12) *and $z \in \{\varphi_k, \varphi_k\varphi_j, \varphi_k\varphi_j\varphi_l\}$ for $k, j, l = 1, \ldots, p$.*

Proof of Lemma A.8. Proof of (A.40)-(A.43): We give the proof of (A.43) only, as the bounds on derivatives of the weight of the inverse lag polynomials are the same order according to Lemma A.7, resulting in a similar proof. From $k_{0t}(d) = \pi_{t-1}(1-d)$ and Lemma A.7,

$$\begin{aligned}
\left|\frac{\partial^{m+3} c_t(\vartheta)}{\partial d^m \partial\varphi_i\partial\varphi_l\partial\varphi_k}\right| &= \sum_{j=0}^{t-1}\left|\frac{\partial^3\phi_j(\varphi)}{\partial\partial\varphi_i\partial\varphi_l\partial\varphi_k}\right|\left|\frac{\partial^m \kappa_{0(t-j)}(d)}{\partial d^m}\right|, \\
&\leq c\sum_{j=1}^{t-1}\log^m(j)j^{-1-\varsigma}(t-j-1)^{-d}, \\
&\leq c\log^m(t)\sum_{j=1}^{t-1}j^{-1-\varsigma}(t-j-1)^{-d},
\end{aligned}$$

because $\varsigma > 1/2$ this summand is $O(t^{\max(-d,-1-\varsigma)})$ from Lemma A.6.
Proof of (A.44) and (A.45): From Lemma A.7,

$$|b_{zi}(\varphi_0)| \leq \sum_{s=0}^{i-1}|\omega_s(\varphi_0)||D_z\phi_{i-s}(\varphi_0)| = O(\sum_{s=1}^{i-1}s^{-1-\varsigma}(i-s)^{-1-\varsigma}) = O(i^{-1-\varsigma}).$$

The last equality follows from Lemma A.6 because $\varsigma > 1/2$.

Then from Lemma A.6 and the bound above it follows that

$$|h_{dzi}(\varphi_0)| = O(\sum_{s=1}^{i-1}(i-s)^{-1}s^{-1-\zeta}) = O(i^{-1}).$$

□

A.2.3 Limit behaviour of the centered product moments

Let $k, j, l = 1, \dots, p+1$. Define the centered product moments of the derivative of the stochastic terms in (A.5) as

$$M^+_{0,\vartheta_k T} = \sigma_0^{-2} T^{-1/2} \sum_{t=1}^{T} \left(S^+_t S^+_{\vartheta_k t} - E\left(S^+_t S^+_{\vartheta_k t} \right) \right), \tag{A.46}$$

$$M^+_{0,\vartheta_k \vartheta_j T} = \sigma_0^{-2} T^{-1/2} \sum_{t=1}^{T} \left(S^+_t S^+_{\vartheta_k \vartheta_j t} - E\left(S^+_t S^+_{\vartheta_k \vartheta_j t} \right) \right), \tag{A.47}$$

$$M^+_{\vartheta_k, \vartheta_j T} = \sigma_0^{-2} T^{-1/2} \sum_{t=1}^{T} \left(S^+_{\vartheta_k t} S^+_{\vartheta_j t} - E\left(S^+_{\vartheta_k t} S^+_{\vartheta_j t} \right) \right), \tag{A.48}$$

$$M^+_{0,\vartheta_k \vartheta_j \vartheta_l T} = \sigma_0^{-2} T^{-1/2} \sum_{t=1}^{T} \left(S^+_t S^+_{\vartheta_k \vartheta_j \vartheta_l t} - E\left(S^+_t S^+_{\vartheta_k \vartheta_j \vartheta_l t} \right) \right), \tag{A.49}$$

$$M^+_{\vartheta_k, \vartheta_j \vartheta_l T} = \sigma_0^{-2} T^{-1/2} \sum_{t=1}^{T} \left(S^+_{\vartheta_k t} S^+_{\vartheta_j \vartheta_l t} - E\left(S^+_{\vartheta_k t} S^+_{\vartheta_j \vartheta_l t} \right) \right), \tag{A.50}$$

and define some of the corresponding vector forms of the centered product moments as

$$M^+_{0,\vartheta T} = (M^+_{0,\vartheta_1 T}, M^+_{0,\vartheta_2 T}, \dots, M^+_{0,\vartheta_{p+1} T})', \tag{A.51}$$

$$M^+_{0,\vartheta_k \vartheta T} = (M^+_{0,\vartheta_k \vartheta_1 T}, M^+_{0,\vartheta_k \vartheta_2 T}, \dots, M^+_{0,\vartheta_k \vartheta_{p+1} T})', \tag{A.52}$$

$$M^+_{\vartheta_k, \vartheta T} = (M^+_{\vartheta_k, \vartheta_1 T}, M^+_{\vartheta_k, \vartheta_2 T}, \dots, M^+_{\vartheta_k, \vartheta_{p+1} T})', \tag{A.53}$$

and matrix form as

$$M^+_{0,\vartheta\vartheta' T} = (M^+_{0,\vartheta_1 \vartheta T}, M^+_{0,\vartheta_2 \vartheta T}, \dots, M^+_{0,\vartheta_{p+1} \vartheta T}), \tag{A.54}$$

$$M^+_{\vartheta,\vartheta' T} = (M^+_{\vartheta_1, \vartheta T}, M^+_{\vartheta_2, \vartheta T}, \dots, M^+_{\vartheta_{p+1}, \vartheta T}). \tag{A.55}$$

By Lemma A.1, we have that $S^+_t = \epsilon_t$. As a direct consequence, we can observe that $E\left(S^+_t S^+_{\vartheta_k t}\right) = E\left(S^+_t S^+_{\vartheta_k \vartheta_j t}\right) = E\left(S^+_t S^+_{\vartheta_k \vartheta_j \vartheta_l t}\right) = 0$.

We next show the limiting behaviour of the centered product moments.

Lemma A.9. *Suppose that Assumptions 3.1-3.4 hold. Then, for $T \to \infty$, it holds that $M^+_{0,\vartheta T}$ is asymptotic normal with mean zero and the variance of $M^+_{0,\vartheta T}$ is*

$$E\left(M^+_{0,\vartheta T} (M^+_{0,\vartheta T})' \right) = A + O(T^{-1} \log(T)),$$

where A is the inverse of the variance-covariance matrix given in (67). *Furthermore, $M^+_{0,\vartheta_k T} = O_P(1)$, $M^+_{0,\vartheta_k \vartheta_j T} = O_P(1)$, $M^+_{\vartheta_k, \vartheta_j T} = O_P(1)$, $M^+_{0,\vartheta_k \vartheta_j \vartheta_l T} = O_P(1)$ and $M^+_{\vartheta_k, \vartheta_j \vartheta_l T} = O_P(1)$.*

Proof of Lemma A.9. The proof of asymptotic normality of $M^+_{0,\vartheta T}$ and the limiting variance is given in Hualde & Robinson (2011, (2.54) and (2.55)). The order the rest term comes from the (1,1)-th element of the matrix $E\left(M^+_{0,\vartheta T} (M^+_{0,\vartheta T})' \right)$,

$$\sigma_0^{-2} T^{-1} \sum_{t=1}^{T} E\left(S^+_{\vartheta_1 t} \right)^2 = T^{-1} \sum_{t=1}^{T} \sum_{k=1}^{t-1} \frac{1}{k^2}$$

$$= \sum_{k=1}^{\infty} \frac{1}{k^2} - T^{-1} \sum_{t=1}^{T} \sum_{k=t}^{\infty} \frac{1}{k^2}$$
$$= \zeta_2 + O(T^{-1} \log(T)),$$

using $\sum_{k=t}^{\infty} k^{-2} = O(t^{-1})$, see (A.30). It can be straightforwardly shown that the other elements in this matrix have a rest term of $O(T^{-1})$. We have that $S_t^+ = \epsilon_t$ and therefore the proof of $M_{0,\vartheta_k T}^+ = O_P(1)$, $M_{0,\vartheta_k\vartheta_j T}^+ = O_P(1)$ and $M_{0,\vartheta_k\vartheta_j\vartheta_l T}^+ = O_P(1)$ are straightforward and can be derived from Lemmata A.6, A.7 and A.8. The proofs of $M_{\vartheta_k,\vartheta_j T}^+ = O_P(1)$ and $M_{\vartheta_k,\vartheta_j\vartheta_l T}^+ = O_P(1)$ are analogous, so we will present the proof for $M_{\vartheta_k,\vartheta_j\vartheta_l T}^+ = O_P(1)$. It is sufficient to show that $E|M_{\vartheta_k,\vartheta_j\vartheta_l T}^+| = O(1)$. Bounding the right-hand side of (A.50) by taking the modulus yields

$$\sigma_0^{-2} T^{-1/2} \sum_{t=1}^{T} \left| S_{\vartheta_k t}^+ S_{\vartheta_j\vartheta_l t}^+ - E\left(S_{\vartheta_k t}^+ S_{\vartheta_j\vartheta_l t}^+\right)\right|.$$

First,

$$S_{\vartheta_k t}^+ S_{\vartheta_j\vartheta_l t}^+ - E\left(S_{\vartheta_k t}^+ S_{\vartheta_j\vartheta_l t}^+\right) = \sum_{n=1}^{t-1} \sum_{s=1}^{t-1} v_{\vartheta_k n} w_{\vartheta_j\vartheta_l s} \left(\epsilon_{t-n}\epsilon_{t-s} - E\left(\epsilon_{t-n}\epsilon_{t-s}\right)\right),$$

where the expression $v_{\vartheta_k t}$ and $w_{\vartheta_j\vartheta_l t}$ follow directly from Lemma A.1 and from Lemmata A.7 and A.8 it holds that $|v_{\vartheta_k t}| = O(t^{-\delta_1})$ and $|w_{\vartheta_j\vartheta_l t}| = O(t^{-\delta_2})$ for some $\delta_1, \delta_2 > 0$. Then,

$$E|M_{0,\vartheta_k\vartheta_j\vartheta_l T}^+| \leq T^{-1/2} \sum_{t=1}^{T} E \left| \sum_{n=1}^{t-1} \sum_{s=1}^{t-1} v_{\vartheta_k n} w_{\vartheta_j\vartheta_l s} \left(\epsilon_{t-n}\epsilon_{t-s} - E\left(\epsilon_{t-n}\epsilon_{t-s}\right)\right)\right|,$$

it follows readily that $\operatorname{Var}\left(\sum_{n=1}^{t-1} \sum_{s=1}^{t-1} v_{\vartheta_k n} w_{\vartheta_j\vartheta_l s} \epsilon_{t-n}\epsilon_{t-s}\right) = O(t)$, so

$$|M_{0,\vartheta_k\vartheta_j\vartheta_l T}^+| = O_P(T^{-1/2} \sum_{t=1}^{T} t^{1/2}) = O_P(1).$$

□

Lemma A.10. *Suppose that Assumptions 3.1-3.4 holds. The covariances of $M_{0,\vartheta_k T}^+$ and $M_{0,\vartheta_j\vartheta_l}^+$ are given by*

$$E\left(M_{0,\vartheta_k T}^+ M_{0,\vartheta_j\vartheta_l T}^+\right) = \sigma_0^{-2} T^{-1} \sum_{t=1}^{T} E\left(S_{\vartheta_k t}^+ S_{\vartheta_j\vartheta_l t}^+\right),$$

for $k, j, l \in \{1, \ldots, p+1\}$. For $T \to \infty$, it holds that

$$\sigma_0^{-2} T^{-1} \sum_{t=1}^{T} E\left(S_{\vartheta_1 t}^+ S_{\vartheta_1\vartheta_1 t}^+\right) = -2\zeta_3 + O(T^{-1}\log^4(T)), \tag{A.56}$$

$$\sigma_0^{-2} T^{-1} \sum_{t=1}^{T} E\left(S_{\vartheta_1 t}^+ S_{\vartheta_1\vartheta_l t}^+\right) = \sum_{i=2}^{\infty} i^{-1} h_{\vartheta_1\vartheta_l i}(\varphi_0) + O(T^{-1}\log(T)), \tag{A.57}$$

$$\sigma_0^{-2} T^{-1} \sum_{t=1}^{T} E\left(S_{\vartheta_1 t}^+ S_{\vartheta_j\vartheta_l t}^+\right) = -\sum_{i=1}^{\infty} i^{-1} b_{\vartheta_j\vartheta_l i}(\varphi_0) + O(T^{-1}), \tag{A.58}$$

$$\sigma_0^{-2}T^{-1}\sum_{t=1}^{T}E\left(S^+_{\vartheta_k t}S^+_{\vartheta_1\vartheta_1 t}\right)=\sum_{i=0}^{\infty}D_{dd}\pi_i(0)b_{\vartheta_k i}(\varphi_0)+O(T^{-1}), \tag{A.59}$$

$$\sigma_0^{-2}T^{-1}\sum_{t=1}^{T}E\left(S^+_{\vartheta_k t}S^+_{\vartheta_1\vartheta_l t}\right)=-\sum_{i=2}^{\infty}b_{\vartheta_k i}(\varphi_0)h_{\vartheta_1\vartheta_l i}(\varphi_0)+O(T^{-1}), \tag{A.60}$$

$$\sigma_0^{-2}T^{-1}\sum_{t=1}^{T}E\left(S^+_{\vartheta_k t}S^+_{\vartheta_j\vartheta_l t}\right)=\sum_{i=1}^{\infty}b_{\vartheta_k i}(\varphi_0)b_{\vartheta_j\vartheta_l i}(\varphi_0)+O(T^{-1}), \tag{A.61}$$

for $k,j,l\in\{2,\dots,p+1\}$.

Proof of Lemma A.10. From Lemma A.1 we have that $S^+_t=\epsilon_t$ and $S^+_{\vartheta_k t}$, $S^+_{\vartheta_j\vartheta_l t}$ are weighted sums of $\epsilon_1,\dots,\epsilon_{t-1}$, so that

$$\begin{aligned}E\left(M^+_{0,\vartheta_k T}M^+_{0,\vartheta_j\vartheta_l T}\right)&=\sigma_0^{-4}T^{-1}E\left(\sum_{t=1}^{T}S^+_tS^+_{\vartheta_k t}\sum_{s=1}^{T}S^+_sS^+_{\vartheta_j\vartheta_l s}\right)\\&=\sigma_0^{-4}T^{-1}\sum_{t=1}^{T}E\left(\left(S^+_t\right)^2S^+_{\vartheta_k t}S^+_{\vartheta_j\vartheta_l t}\right)\\&=\sigma_0^{-2}T^{-1}\sum_{t=1}^{T}E\left(S^+_{\vartheta_k t}S^+_{\vartheta_j\vartheta_l t}\right),\end{aligned} \tag{A.62}$$

where the last inequality uses the independence of $\left(S^+_t\right)^2$ and $S^+_{\vartheta_k t}S^+_{\vartheta_j\vartheta_l t}$.

Proof of (A.56): Consider the case $k=j=l=1$ for (A.62). We have

$$\begin{aligned}\sigma_0^{-2}T^{-1}\sum_{t=1}^{T}E\left(S^+_{\vartheta_1 t}S^+_{\vartheta_1\vartheta_1 t}\right)&=-\sigma_0^{-2}T^{-1}\sum_{t=1}^{T}E\left(\left(\sum_{k=0}^{t-1}D_d\pi_k(0)\epsilon_{t-k}\right)\left(\sum_{k=0}^{t-1}D_{dd}\pi_k(0)\epsilon_{t-k}\right)\right)\\&=-T^{-1}\sum_{t=1}^{T}\sum_{k=0}^{t-1}D_{dd}\pi_k(0)D_d\pi_k(0)\\&=-\sum_{k=0}^{\infty}D_{dd}\pi_k(0)D_d\pi_k(0)+T^{-1}\sum_{t=1}^{T}\sum_{k=t}^{\infty}D_{dd}\pi_k(0)D_d\pi_k(0),\end{aligned}$$

Then from (A.13) and (A.14)

$$\begin{aligned}-\sum_{k=0}^{\infty}D_{dd}\pi_k(0)D_d\pi_k(0)&=-2\sum_{k=2}^{\infty}k^{-2}\sum_{j=1}^{k-1}j^{-1}\\&=-2\zeta_3,\end{aligned}$$

where the last equality follows from Johansen & Nielsen (2016, Lemma B.2) and from Lemmata A.6 and A.7 we have

$$\begin{aligned}T^{-1}\sum_{t=1}^{T}\sum_{k=t}^{\infty}D_{dd}\pi_k(0)D_d\pi_k(0)&=O(T^{-1}\sum_{t=1}^{T}\sum_{k=t}^{\infty}(1+\log(k))^3k^{-2})\\&=O(T^{-1}\sum_{t=1}^{T}(1+\log(t))^3t^{-1})=O(T^{-1}\log^4(T)).\end{aligned}$$

Proof of (A.57): Consider the case $k = j = 1$ and $l \geq 2$ for (A.62). Then

$$\begin{aligned}
\sigma_0^{-2}T^{-1}\sum_{t=1}^{T} E\left(S^{+}_{\vartheta_1 t}S^{+}_{\vartheta_1\vartheta_l t}\right) &= \sigma_0^{-2}T^{-1}\sum_{t=1}^{T} E\left(\left(\sum_{k=0}^{t-1} D_d\pi_k(0)\epsilon_{t-k}\right)\left(\sum_{i=2}^{t-1} h_{\vartheta_1\vartheta_l i}(\varphi_0)\epsilon_{t-i}\right)\right)\\
&= T^{-1}\sum_{t=1}^{T}\sum_{i=2}^{t-1} i^{-1}h_{\vartheta_1\vartheta_l i}(\varphi_0)\\
&= \sum_{i=2}^{\infty} i^{-1}h_{\vartheta_1\vartheta_l i}(\varphi_0) - T^{-1}\sum_{t=1}^{T}\sum_{i=t}^{\infty} i^{-1}h_{\vartheta_1\vartheta_l i}(\varphi_0),
\end{aligned}$$

and

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T}\sum_{i=t}^{\infty} i^{-1}h_{\vartheta_1\vartheta_l i}(\varphi_0) &= O(T^{-1}\sum_{t=1}^{T}\sum_{i=t}^{\infty} i^{-2})\\
&= O(T^{-1}\sum_{t=1}^{T} t^{-1}) = O(T^{-1}\log(T)).
\end{aligned}$$

Proof of (A.58): Consider the case $k = 1$ and $j, l \geq 2$ for (A.62). We have

$$\begin{aligned}
\sigma_0^{-2}T^{-1}\sum_{t=1}^{T} E\left(S^{+}_{\vartheta_1 t}S^{+}_{\vartheta_j\vartheta_l t}\right) &= -\sigma_0^{-2}T^{-1}\sum_{t=1}^{T} E\left(\left(\sum_{k=0}^{t-1} D_d\pi_k(0)\epsilon_{t-k}\right)\left(\sum_{i=1}^{t-1} b_{\vartheta_j\vartheta_l i}(\varphi_0)\epsilon_{t-i}\right)\right)\\
&= -T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{t-1} i^{-1}b_{\vartheta_k\vartheta_l i}(\varphi_0)\\
&= -\sum_{i=1}^{\infty} i^{-1}b_{\vartheta_j\vartheta_l i}(\varphi_0) + T^{-1}\sum_{t=1}^{T}\sum_{i=t}^{\infty} i^{-1}b_{\vartheta_k\vartheta_l i}(\varphi_0),
\end{aligned}$$

and

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T}\sum_{i=t}^{\infty} i^{-1}b_{\vartheta_j\vartheta_l i}(\varphi_0) &= O(T^{-1}\sum_{t=1}^{T}\sum_{i=t}^{\infty} i^{-2-\varsigma})\\
&= O(T^{-1}\sum_{t=1}^{T} t^{-1-\varsigma}) = O(T^{-1}).
\end{aligned}$$

Proof of (A.59): Consider the case $k \geq 2$ and $j = l = 1$ for (A.62). We have

$$\begin{aligned}
\sigma_0^{-2}T^{-1}\sum_{t=1}^{T} E\left(S^{+}_{\vartheta_k t}S^{+}_{\vartheta_1\vartheta_1 t}\right) &= \sigma_0^{-2}T^{-1}\sum_{t=1}^{T} E\left(\left(\sum_{i=1}^{t-1} b_{\vartheta_k i}(\varphi_0)\epsilon_{t-i}\right)\left(\sum_{k=0}^{t-1} D_{dd}\pi_k(0)\epsilon_{t-k}\right)\right)\\
&= T^{-1}\sum_{t=1}^{T}\sum_{i=0}^{t-1} D_{dd}\pi_i(0)b_{\vartheta_k i}(\varphi_0)\\
&= \sum_{i=0}^{\infty} D_{dd}\pi_i(0)b_{\vartheta_k i}(\varphi_0) - T^{-1}\sum_{t=1}^{T}\sum_{i=t}^{\infty} D_{dd}\pi_i(0)b_{\vartheta_k i}(\varphi_0),
\end{aligned}$$

and

$$T^{-1}\sum_{t=1}^{T}\sum_{i=t}^{\infty} D_{dd}\pi_i(0)b_{\vartheta_k i}(\varphi_0) = O(T^{-1}\sum_{t=1}^{T}\sum_{i=t}^{\infty}(1+\log(k))^2 i^{-2-\varsigma})$$

$$= O(T^{-1}\sum_{t=1}^{T}(1+\log(t))^2 t^{-1-\varsigma}) = O(T^{-1}\sum_{t=1}^{T} t^{-1-\varsigma+2\lambda}) = O(T^{-1}),$$

where we use the bound $1+\log(t) < t^{\lambda}$ for small $\lambda > 0$.

Proof of (A.60): Consider the case $k, l \geq 2$ and $j = 1$ for (A.62). We have

$$\begin{aligned}\sigma_0^{-2}T^{-1}\sum_{t=1}^{T} E\left(S^{+}_{\vartheta_k t}S^{+}_{\vartheta_1\vartheta_l t}\right) &= -\sigma_0^{-2}T^{-1}\sum_{t=1}^{T} E\left(\left(\sum_{i=1}^{t-1} b_{\vartheta_k i}(\varphi_0)\epsilon_{t-i}\right)\left(\sum_{i=2}^{t-1} h_{\vartheta_1\vartheta_l i}(\varphi_0)\epsilon_{t-i}\right)\right)\\ &= -T^{-1}\sum_{t=1}^{T}\sum_{i=2}^{t-1} b_{\vartheta_k i}(\varphi_0)h_{\vartheta_1\vartheta_l i}(\varphi_0)\\ &= -\sum_{i=2}^{\infty} b_{\vartheta_k i}(\varphi_0)h_{\vartheta_1\vartheta_l i}(\varphi_0) + T^{-1}\sum_{t=1}^{T}\sum_{i=t}^{\infty} b_{\vartheta_k i}(\varphi_0)h_{\vartheta_1\vartheta_l i}(\varphi_0),\end{aligned}$$

and

$$\begin{aligned}T^{-1}\sum_{t=1}^{T}\sum_{i=t}^{\infty} b_{\vartheta_k i}(\varphi_0)h_{\vartheta_1\vartheta_l i}(\varphi_0) &= O(T^{-1}\sum_{t=1}^{T}\sum_{i=t}^{\infty} i^{-2-\varsigma})\\ &= O(T^{-1}\sum_{t=1}^{T} t^{-1-\varsigma}) = O(T^{-1}).\end{aligned}$$

Proof of (A.61): Consider the case $k, j, l \geq 2$ for (A.62). We have

$$\begin{aligned}\sigma_0^{-2}T^{-1}\sum_{t=1}^{T} E\left(S^{+}_{\vartheta_k t}S^{+}_{\vartheta_j\vartheta_l t}\right) &= \sigma_0^{-2}T^{-1}\sum_{t=1}^{T} E\left(\left(\sum_{i=1}^{t-1} b_{\vartheta_k i}(\varphi_0)\epsilon_{t-i}\right)\left(\sum_{i=1}^{t-1} b_{\vartheta_j\vartheta_l i}(\varphi_0)\epsilon_{t-i}\right)\right)\\ &= T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{t-1} b_{\vartheta_k i}(\varphi_0)b_{\vartheta_j\vartheta_l i}(\varphi_0)\\ &= \sum_{i=1}^{\infty} b_{\vartheta_k i}(\varphi_0)b_{\vartheta_j\vartheta_l i}(\varphi_0) - T^{-1}\sum_{t=1}^{T}\sum_{i=t}^{\infty} b_{\vartheta_k i}(\varphi_0)b_{\vartheta_j\vartheta_l i}(\varphi_0),\end{aligned}$$

and

$$\begin{aligned}T^{-1}\sum_{t=1}^{T}\sum_{i=t}^{\infty} b_{\vartheta_k i}(\varphi_0)b_{\vartheta_j\vartheta_l i}(\varphi_0) &= O(T^{-1}\sum_{t=1}^{T}\sum_{i=t}^{\infty} i^{-2-2\varsigma})\\ &= O(T^{-1}\sum_{t=1}^{T} t^{-1-2\varsigma}) = O(T^{-1}).\end{aligned}$$

□

Lemma A.11. *Suppose that Assumptions 3.1-3.4 holds. The covariances of $M^{+}_{0,\vartheta_1 T}$ and $M^{+}_{\vartheta_1,\vartheta_1}$ are given by*

$$E\left(M^{+}_{0,\vartheta_1 T}M^{+}_{\vartheta_1,\vartheta_1 T}\right) = -4\zeta_3 + O(T^{-1}\log^2(T)), \tag{A.63}$$

$$E\left(M^{+}_{0,\vartheta_1 T}M^{+}_{\vartheta_1,\vartheta_l T}\right) = \sum_{k=1}^{\infty} k^{-1}\sum_{s=1}^{\infty}\left(s^{-1}b_{\vartheta_l(s+k)}(\varphi_0) + (s+k)^{-1}b_{\vartheta_l s}(\varphi_0)\right)$$

$$+ O(T^{-1}\log(T)), \tag{A.64}$$

$$E\left(M^+_{0,\vartheta_1 T}M^+_{\vartheta_n,\vartheta_l T}\right) = -\sum_{k=1}^{\infty} k^{-1}\sum_{s=1}^{\infty}\left(b_{\vartheta_n s}(\varphi_0)b_{\vartheta_l (s+k)}(\varphi_0) + b_{\vartheta_n (s+k)}(\varphi_0)b_{\vartheta_l s}(\varphi_0)\right)$$
$$+ O(T^{-1}\log(T)), \tag{A.65}$$

$$E\left(M^+_{0,\vartheta_l T}M^+_{\vartheta_1,\vartheta_1 T}\right) = 2\sum_{k=1}^{\infty} b_{\vartheta_l k}(\varphi_0)\sum_{s=1}^{\infty} s^{-1}(s+k)^{-1} + O(T^{-1}\log(T)), \tag{A.66}$$

$$E\left(M^+_{0,\vartheta_m T}M^+_{\vartheta_1,\vartheta_l T}\right) = -\sum_{k=1}^{\infty} b_{\vartheta_m k}(\varphi_0)\sum_{s=1}^{\infty}\left(s^{-1}b_{\vartheta_l (s+k)}(\varphi_0) + (s+k)^{-1}b_{\vartheta_l s}(\varphi_0)\right)$$
$$+ O(T^{-1}\log(T)), \tag{A.67}$$

$$E\left(M^+_{0,\vartheta_m T}M^+_{\vartheta_n,\vartheta_l T}\right) = \sum_{k=1}^{\infty} b_{\vartheta_m k}(\varphi_0)\sum_{s=1}^{\infty}\left(b_{\vartheta_n s}(\varphi_0)b_{\vartheta_l (s+k)}(\varphi_0) + b_{\vartheta_n (s+k)}(\varphi_0)b_{\vartheta_l s}(\varphi_0)\right)$$
$$+ O(T^{-1}\log(T)), \tag{A.68}$$

for $k, j, l \in \{2, \dots, p+1\}$.

Proof of Lemma A.11. We have that

$$E\left(M^+_{0,\vartheta_k T}M^+_{\vartheta_j,\vartheta_l T}\right) = \sigma_0^{-4}T^{-1}E\left(\sum_{t=1}^{T} S_t^+ S^+_{\vartheta_k t}\sum_{s=1}^{T} S^+_{\vartheta_j s}S^+_{\vartheta_l s}\right).$$

The expectation of $S_t^+ S^+_{\vartheta_k t}S^+_{\vartheta_j s}S^+_{\vartheta_l}$ equals zero for $s \leq t$ so that what only matters is

$$\sigma_0^{-4}T^{-1}E\left(\sum_{t=1}^{T} S_t^+ S^+_{\vartheta_k t}\sum_{s=t+1}^{T} S^+_{\vartheta_j s}S^+_{\vartheta_l s}\right). \tag{A.69}$$

Now we consider the different cases.

Proof of (A.63): Consider the case $k, j, l = 1$ for (A.69). From Lemma A.1

$$-\sigma_0^{-4}T^{-1}E\left(\sum_{t=1}^{T}\epsilon_t\sum_{k=0}^{t-1} D_d\pi_k(0)\epsilon_{t-k}\sum_{s=t+1}^{T}\sum_{n=0}^{s-1} D_d\pi_n(0)\epsilon_{s-n}\sum_{a=0}^{s-1} D_d\pi_a(0)\epsilon_{s-a}\right).$$

Only the contributions of the form $\epsilon_t^2\epsilon_{t-k}^2$ are non-zero such that what only matter is if $s-n=t$ and $s-a=t-k$ or if $s-a=t$ and $s-n=t-k$ and since both contributions are equal we get

$$-2\sigma_0^{-4}T^{-1}\sum_{t=1}^{T}\sum_{k=0}^{t-1}\sum_{s=t+1}^{T} D_d\pi_k(0)D_d\pi_{s-t}(0)D_d\pi_{s-t+k}(0)E\left(\epsilon_t^2\epsilon_{t-k}^2\right).$$

Plugging in $D_d\pi_k(0)D_d = k^{-1}I(k \geq 1)$ and $D_{dd}\pi_j(0) = 2j^{-1}a_{j-1}I(j \geq 2)$, with $a_j = I(j \geq 1)\sum_{k=1}^{j} k^{-1}$, see (A.13) and (A.14), yields

$$-2T^{-1}\sum_{t=1}^{T}\sum_{k=1}^{t-1}\sum_{s=t+1}^{T} k^{-1}(s-t)^{-1}(s-t+k)^{-1},$$

or, equivalently,

$$-2T^{-1}\sum_{t=1}^{T}\sum_{s=t+1}^{T}\sum_{k=1}^{t-1}(t-k)^{-1}(s-t)^{-1}(s-k)^{-1},$$

which can be written as

$$
\begin{aligned}
&- 2T^{-1}\sum_{t=1}^{T}\sum_{s=t+1}^{T}\sum_{k=-\infty}^{t-1}(t-k)^{-1}(s-t)^{-1}(s-k)^{-1}\\
&+ 2T^{-1}\sum_{t=1}^{T}\sum_{s=t+1}^{T}(s-t)^{-1}\sum_{k=-\infty}^{0}(t-k)^{-1}(s-k)^{-1}.
\end{aligned}
$$

For the first term, we have

$$
-2T^{-1}\sum_{t=1}^{T}\sum_{s=t+1}^{T}\sum_{k=-\infty}^{t-1}(t-k)^{-1}(s-t)^{-1}(s-k)^{-1} = -4\zeta_3,
$$

see Johansen & Nielsen (2016, Lemma B.2). For the second term, we have

$$
\begin{aligned}
O(T^{-1}\sum_{t=1}^{T}\sum_{s=t+1}^{T}(s-t)^{-1}\sum_{k=-\infty}^{0}(t-k)^{-1}(s-k)^{-1}) &= O(T^{-1}\sum_{t=1}^{T}\sum_{s=t+1}^{T}(s-t)^{-1}\sum_{k=-\infty}^{0}(s-k)^{-2})\\
&= O(T^{-1}\sum_{t=1}^{T}\sum_{s=t+1}^{T}(s-t)^{-1}\sum_{k=0}^{\infty}(s+k)^{-2})\\
&= O(T^{-1}\sum_{t=1}^{T}\sum_{s=t+1}^{T}(s-t)^{-1}s^{-1})\\
&= O(T^{-1}\sum_{t=1}^{T}t^{-1}\log(T))\\
&= O(T^{-1}\log^2(T)).
\end{aligned}
$$

Proof of (A.64): Consider the case $k, j = 1$ and $l > 1$ for (A.69). From Lemma A.1

$$
\sigma_0^{-4}T^{-1}E\left(\sum_{t=1}^{T}\epsilon_t\sum_{k=0}^{t-1}D_d\pi_k(0)\epsilon_{t-k}\sum_{s=t+1}^{T}\sum_{n=0}^{s-1}D_d\pi_n(0)\epsilon_{s-n}\sum_{a=1}^{s-1}b_{\vartheta_l a}(\varphi_0)\epsilon_{s-a}\right).
$$

Only the contributions of the form $\epsilon_t^2\epsilon_{t-k}^2$ are non-zero such that

$$
\sigma_0^{-4}T^{-1}E\left(\sum_{t=1}^{T}\epsilon_t\sum_{k=0}^{t-1}D_d\pi_k(0)\epsilon_{t-k}\sum_{s=t+1}^{T}\left(\pi_{s-t}(0)\epsilon_t b_{\vartheta_l(s-t+k)}(\varphi_0)\epsilon_{t-k} + \pi_{s-t+k}(0)\epsilon_{t-k}b_{\vartheta_l(s-t)}(\varphi_0)\epsilon_t,\right)\right)
$$

and plugging in the definition $D_d\pi_k(0)$, see Lemma A.1, gives

$$
T^{-1}\sum_{t=1}^{T}\sum_{k=1}^{t-1}k^{-1}\sum_{s=t+1}^{T}\left((s-t)^{-1}b_{\vartheta_l(s-t+k)}(\varphi_0) + (s-t+k)^{-1}b_{\vartheta_l(s-t)}(\varphi_0)\right). \tag{A.70}
$$

The first term in (A.70) is

$$
\begin{aligned}
T^{-1}\sum_{t=1}^{T}\sum_{k=1}^{t-1}k^{-1}\sum_{s=t+1}^{T}(s-t)^{-1}b_{\vartheta_l(s-t+k)}(\varphi_0) &= T^{-1}\sum_{t=1}^{T}\sum_{k=1}^{t-1}k^{-1}\sum_{s=1}^{T-t}s^{-1}b_{\vartheta_l(s+k)}(\varphi_0)\\
&= \sum_{k=1}^{\infty}k^{-1}\sum_{s=1}^{\infty}s^{-1}b_{\vartheta_l(s+k)}(\varphi_0)
\end{aligned}
$$

$$- T^{-1}\sum_{t=1}^{T}\sum_{k=1}^{t-1} k^{-1} \sum_{s=T-t+1}^{\infty} s^{-1} b_{\vartheta_l(s+k)}(\varphi_0)$$

$$- T^{-1}\sum_{t=1}^{T}\sum_{k=t}^{\infty} k^{-1} \sum_{s=1}^{\infty} s^{-1} b_{\vartheta_l(s+k)}(\varphi_0),$$

where

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T}\sum_{k=1}^{t-1} k^{-1} \sum_{s=T-t+1}^{\infty} s^{-1} b_{\vartheta_l(s+k)}(\varphi_0) &= O(T^{-1}\sum_{t=1}^{T}\sum_{k=1}^{t-1} k^{-1} \sum_{s=T-t+1}^{\infty} s^{-1}(s+k)^{-1-\varsigma}) \\
&= O(T^{-1}\sum_{t=1}^{T}\sum_{k=1}^{t-1} k^{-1}(T-t+1)^{-1} \sum_{s=1}^{\infty}(s+k)^{-1-\varsigma}) \\
&= O(T^{-1}\sum_{t=1}^{T}(T-t+1)^{-1} \sum_{k=1}^{t-1} k^{-1-\varsigma}) \\
&= O(T^{-1}\log(T)),
\end{aligned}$$

and

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T}\sum_{k=t}^{\infty} k^{-1} \sum_{s=1}^{\infty} s^{-1} b_{\vartheta_l(s+k)}(\varphi_0) &= O(T^{-1}\sum_{t=1}^{T}\sum_{k=t}^{\infty} k^{-1} \sum_{s=1}^{\infty} s^{-1}(s+k)^{-1-\varsigma}) \\
&= O(T^{-1}\sum_{t=1}^{T}\sum_{k=t}^{\infty} k^{-1} \sum_{s=1}^{\infty} s^{-1+\lambda}(s+k)^{-1-\varsigma}) \\
&= O(T^{-1}\sum_{t=1}^{T}\sum_{k=t}^{\infty} k^{-2-\varsigma+\lambda}) \\
&= O(T^{-1}\sum_{t=1}^{T} t^{-1-\varsigma+\lambda}) \\
&= O(T^{-1}),
\end{aligned}$$

with $\lambda > 0$ is small constant. The second term in (A.70) is

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T}\sum_{k=1}^{t-1} k^{-1} \sum_{s=t+1}^{T} (s-t+k)^{-1} b_{\vartheta_l(s-t)}(\varphi_0) &= T^{-1}\sum_{t=1}^{T}\sum_{k=1}^{t-1} k^{-1} \sum_{s=1}^{T-t}(s+k)^{-1} b_{\vartheta_l s}(\varphi_0) \\
&= \sum_{k=1}^{\infty} k^{-1} \sum_{s=1}^{\infty}(s+k)^{-1} b_{\vartheta_l s}(\varphi_0) \\
&\quad - T^{-1}\sum_{t=1}^{T}\sum_{k=1}^{t-1} k^{-1} \sum_{s=T-t+1}^{\infty} (s+k)^{-1} b_{\vartheta_l s}(\varphi_0) \\
&\quad - T^{-1}\sum_{t=1}^{T}\sum_{k=t}^{\infty} k^{-1} \sum_{s=1}^{\infty}(s+k)^{-1} b_{\vartheta_l s}(\varphi_0),
\end{aligned}$$

where

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T}\sum_{k=1}^{t-1} k^{-1} \sum_{s=T-t+1}^{\infty} (s+k)^{-1} b_{\vartheta_l s}(\varphi_0) &= O(T^{-1}\sum_{t=1}^{T}\sum_{k=1}^{t-1} k^{-1} \sum_{s=T-t+1}^{\infty} (s+k)^{-1} s^{-1-\varsigma}) \\
&= O(T^{-1}\sum_{t=1}^{T}\sum_{k=1}^{t-1} k^{-1}(T-t+1)^{-1} \sum_{s=T-t+1}^{\infty} (s+k)^{-1} s^{-1+(1-\varsigma+\lambda)})
\end{aligned}$$

$$
\begin{aligned}
&= O(T^{-1}\sum_{t=1}^{T}\sum_{k=1}^{t-1} k^{-1}(T-t+1)^{-1}k^{-\varsigma+\lambda}) \\
&= O(T^{-1}\sum_{t=1}^{T} t^{-1}) \\
&= O(T^{-1}\log(T)),
\end{aligned}
$$

with $\lambda > 0$ is small constant, and

$$
\begin{aligned}
T^{-1}\sum_{t=1}^{T}\sum_{k=t}^{\infty} k^{-1}\sum_{s=1}^{\infty}(s+k)^{-1}b_{\vartheta_l s}(\varphi_0) &= O(T^{-1}\sum_{t=1}^{T}\sum_{k=t}^{\infty} k^{-1}\sum_{s=1}^{\infty}(s+k)^{-1}s^{-1-\varsigma}) \\
&= O(T^{-1}\sum_{t=1}^{T}\sum_{k=t}^{\infty} k^{-2}\sum_{s=1}^{\infty} s^{-1-\varsigma}) \\
&= O(T^{-1}\sum_{t=1}^{T} t^{-1}) \\
&= O(T^{-1}\log(T)).
\end{aligned}
$$

Proof of (A.65)-(A.68): The proof is omitted, as it follows a similar step as in the proof of (A.64).

□

A.2.4 Expectation of the score function

The following lemma will be used to calculate the expectation of the score function of $L^*(\vartheta)$.

Lemma A.12. *Suppose that Assumptions 3.1-3.4 holds. Then*

$$
E\left(\sum_{s=1}^{T} c_s S_s^+ \sum_{t=1}^{T} c_t S_{\vartheta_l t}^+\right) = \sigma_0^2 \sum_{t=1}^{T} c_t c_{\vartheta_l t},
$$

for $l \in \{1, \ldots, p+1\}$.

Proof of Lemma A.12. We first show the proof for $l = 1$, i.e. $\vartheta_1 = d$. From $S_s^+ = \epsilon_s$ and $S_{dt}^+ = -\sum_{k=0}^{t-1} k^{-1}\epsilon_{t-k}$, see Lemma A.1, we find

$$
\begin{aligned}
\sum_{t=1}^{T} c_t S_{dt}^+ &= -\sum_{t=1}^{T-1}\epsilon_t \sum_{k=t+1}^{T} c_k \frac{1}{k-t} \\
&= -\sum_{t=1}^{T-1}\epsilon_t \sum_{k=1}^{T} c_k D\pi_{k-t}(u)|_{u=0},
\end{aligned}
$$

where $D\pi_{k-t}(u)|_{u=0} = (k-t)^{-1}I(k-t \geq 1)$ and hence

$$
E(\sum_{s=1}^{T} c_s S_s^+ \sum_{t=1}^{T} c_t S_{dt}^+) = -\sigma_0^2 \sum_{t=1}^{T-1} c_t \sum_{k=1}^{T} c_k D\pi_{k-t}(u)|_{u=0}
$$

$$= -\sigma_0^2 \sum_{k=1}^{T} c_k \sum_{t=1}^{T-1} c_t D\pi_{k-t}(u)|_{u=0},$$

Next, we show that $\sum_{t=1}^{T-1} c_t D\pi_{k-t}(u)|_{u=0} = c_{dk}$. From $c_t = \sum_{j=0}^{t-1} \phi_j(\varphi)\kappa_{0(t-j)}(d)$ we find that

$$\begin{aligned}
\sum_{t=1}^{T-1} c_t D\pi_{k-t}(u)|_{u=0} &= \sum_{t=1}^{T-1} D\pi_{k-t}(u)|_{u=0} \sum_{j=0}^{t-1} \phi_j(\varphi)\kappa_{0(t-j)}(d) \\
&= \sum_{j=0}^{t-1} \phi_j(\varphi) \sum_{t=1}^{T-1} D\pi_{k-t}(u)|_{u=0}\kappa_{0(t-j)}(d) \\
&= \sum_{j=0}^{t-1} \phi_j(\varphi) \sum_{t=1}^{T-1} D\pi_{(k-j)-(t-j)}(u)|_{u=0}\kappa_{0(t-j)}(d) \\
&= \sum_{j=0}^{t-1} \phi_j(\varphi) \sum_{m=1-j}^{T-1} D\pi_{(k-j)-m}(u)|_{u=0}\kappa_{0m}(d) \\
&= \sum_{j=0}^{t-1} \phi_j(\varphi) \sum_{m=1}^{k-j} D\pi_{(k-j)-m}(u)|_{u=0}\kappa_{0m}(d) \\
&= \sum_{j=0}^{k-1} \phi_j(\varphi)\kappa_{1(k-j)}(d) \\
&= c_{dk},
\end{aligned}$$

where the second last equality follows from Johansen & Nielsen (2016, Lemma A.4). We next give a proof for $l \in \{2, \dots, p+1\}$, i.e. φ_n for $n \in \{1, \dots, p\}$. From Lemma A.1 it follows

$$\sum_{t=1}^{T} c_t S_{\varphi_n t}^{+} = \sum_{t=1}^{T} \epsilon_t \sum_{s=t}^{T} c_s b_{\varphi_n(s-t)},$$

so that

$$\begin{aligned}
E(\sum_{t=1}^{T} c_t S_{\varphi_n t}^{+} \sum_{s=1}^{T} c_s S_s^{+}) &= \sum_{t=1}^{T} c_t E(S_{\varphi_n t}^{+} \sum_{s=1}^{T} c_s S_s^{+}) \\
&= \sigma_0^2 \sum_{t=1}^{T} c_t \sum_{k=1}^{t-1} c_k b_{n(t-k)}.
\end{aligned}$$

We need to show that $\sum_{k=1}^{t-1} c_k b_{n(t-k)} = c_{\varphi_n t}$. We find that

$$\begin{aligned}
\sum_{k=1}^{t-1} c_k b_{n(t-k)} &= \sum_{k=1}^{t-1}\sum_{j=1}^{k} \phi_{k-j}(\varphi)\kappa_{0j}(d) \sum_{i=1}^{t-k} \omega_{t-k-i}(\varphi) D_{\varphi_n}\phi_i(\varphi) \\
&= \sum_{j=1}^{t-1} \kappa_{0j}(d) \sum_{i=1}^{t-1} D_{\varphi_n}\phi_i(\varphi) \sum_{k=1}^{t-1} \phi_{k-j}(\varphi)\omega_{t-k-i}(\varphi) \\
&= \sum_{j=1}^{t-1} \kappa_{0j}(d) \sum_{i=1}^{t-1} D_{\varphi_n}\phi_i(\varphi) \sum_{k=j}^{t-i} \phi_{k-j}(\varphi)\omega_{t-k-i}(\varphi)
\end{aligned}$$

$$= \sum_{j=1}^{t-1} \kappa_{0j}(d) \sum_{i=0}^{t-j} D_{\varphi_n} \phi_i(\varphi) \sum_{k=0}^{t-j-i} \phi_k(\varphi) \omega_{(t-j-i)-k}(\varphi)$$
$$= \sum_{j=1}^{t-1} \kappa_{0j}(d) D_{\varphi_n} \phi_{t-j}(\varphi),$$

since $\sum_{k=0}^{t-j-i} \phi_k(\varphi)\omega_{(t-j-i)-k}(\varphi) = 1$ and follows from the identity $\phi(L;\varphi)\omega(L;\varphi)I(t \leq k) = 1$.

□

The following lemma finds the expectation of $DL^*(\vartheta_0))$ and $L^*(\vartheta_0)$.

Lemma A.13. *Let the model for the data x_t, t = 1,...,T, be given by* (63) *and let Assumptions 3.1-3.4 be satisfied. Then*

$$E\left(D_{\vartheta_k} L^*(\vartheta_0)\right) = -\sigma_0^2 \frac{\sum_{t=1}^T c_t(\vartheta_0) c_{\vartheta_k t}(\vartheta_0)}{\sum_{t=1}^T c_t^2(\vartheta_0)}, \tag{A.71}$$
$$E\left(L^*(\vartheta_0)\right) = \sigma_0^2 \frac{T-1}{2}, \tag{A.72}$$

for $k \in \{1, \dots, p+1\}$.

Proof of Lemma A.13. The proofs are omitted since it follows straightforwardly from Lemmata A.2 and A.12.

□

A.3 Approximation of the derivatives

In this section, we provide approximations for the first three derivatives of $L^*(\vartheta)$, $L^*_{\mu_0}(\vartheta)$ and $L^*_m(\vartheta)$ evaluated at $\vartheta = \vartheta_0$. Before that, we present results that analyse the terms involved in these derivatives. Specifically, we examine the order of magnitude of functions that incorporate the derivatives of the deterministic term $c_t(\vartheta)$ and the derivatives of the stochastic term $S_t^+(\vartheta)$, as well as the product moments that contain these terms. This analysis is divided into two parts. In Section A.3.1, we focus on the non-stationary region, where $d_0 > 1/2$. Then, in Section A.3.2, we explore the stationary region, where $d_0 < 1/2$. The reason for conducting separate analyses is that the order of magnitude varies depending on the region. Each section concludes with an approximation of the derivatives.

A.3.1 Non-stationary region

In Lemmata A.14 and A.16, we investigate the order of magnitude of functions involving the deterministic term $c_t(\vartheta)$ and its derivatives and the stochastic term S_t^+ and its derivatives and the product moments containing these. In Lemma A.15, we investigate the order of magnitude involving the modification term $m(\vartheta)$ and derivatives of these. These lemmata are then used to find asymptotic results for the first three derivatives of L^*, $L^*_{\mu_0}$ and L^*_m in Lemmata A.17, A.18, and A.19, respectively.

Lemma A.14. *Suppose that Assumptions 3.2-3.4 holds. Let $d > 1/2$, then we have that:*

$$\sum_{t=1}^{T} c_t^2(\vartheta) = \sum_{t=1}^{\infty} c_t^2(\vartheta) + O(T^{\max(1-2d,-1-2\varsigma)}), \tag{A.73}$$

$$\begin{aligned}\sum_{t=1}^{T} c_t(\vartheta)c_{\vartheta_k t}(\vartheta) &= \sum_{t=1}^{\infty} c_t(\vartheta)c_{\vartheta_k t}(\vartheta) \\ &\quad + O(T^{\max(1-2d,-1-2\varsigma)}\log(T)I(k=1) + T^{\max(1-2d,-1-2\varsigma)}I(k>1)),\end{aligned} \tag{A.74}$$

$$\sum_{t=1}^{T} c_{st}(\vartheta)c_{it}(\vartheta) = O(1), \tag{A.75}$$

where $s \in \{0, \vartheta_{\tilde{k}}, \vartheta_{\tilde{k}}\vartheta_{\tilde{k}}, \vartheta_{\tilde{k}}\vartheta_j, \vartheta_{\tilde{k}}\vartheta_{\tilde{k}}\vartheta_{\tilde{l}}\}$, $i \in \{0, \vartheta_k, \vartheta_k\vartheta_j, \vartheta_k\vartheta_j, \vartheta_k\vartheta_j\vartheta_l\}$ and $\tilde{k}, \tilde{j}, \tilde{l}, k, j, l = 1, \dots, p+1$. Here, $c_{0t}(\vartheta)$ refers to $c_t(\vartheta)$.

Proof of Lemma A.14. Proof of (A.73): Given that

$$\sum_{t=1}^{T} c_t^2(\vartheta) = \sum_{t=1}^{\infty} c_t^2(\vartheta) - \sum_{t=T+1}^{\infty} c_t^2(\vartheta),$$

and using $c_t(\vartheta) = O(t^{\max(-d,-1-\varsigma)})$, see (A.40) in Lemma A.8, we can deduce that

$$\begin{aligned}\sum_{t=T+1}^{\infty} c_t^2(\vartheta) &= O\left(\sum_{t=T+1}^{\infty} t^{\max(-2d,-2-2\varsigma)}\right) \\ &= O\left(T^{\max(1-2d,-1-2\varsigma)}\right),\end{aligned}$$

where the last equality follows from (A.30) in Lemma A.6.
Proof of (A.74): Given that

$$\sum_{t=1}^{T} c_t(\vartheta)c_{\vartheta_k t}(\vartheta) = \sum_{t=1}^{\infty} c_t(\vartheta)c_{\vartheta_k t}(\vartheta) - \sum_{t=T+1}^{\infty} c_t(\vartheta)c_{\vartheta_k t}(\vartheta).$$

For ϑ_1, using $c_{\vartheta_1 t}(\vartheta) = O(\log(T)t^{\max(-d,-1-\varsigma)})$, see (A.40) in Lemma A.8, we can deduce that

$$\begin{aligned}\sum_{t=T+1}^{\infty} c_t(\vartheta)c_{\vartheta_1 t}(\vartheta) &= O\left(\sum_{t=T+1}^{\infty} \log(t)t^{\max(-2d,-2-2\varsigma)}\right) \\ &= O\left(\log(T)T^{\max(1-2d,-1-2\varsigma)}\right),\end{aligned}$$

where the last equality follows from (A.30) in Lemma A.6. Regarding ϑ_s, $s \geq 2$, it can be shown that $c_{\vartheta_s t}(\vartheta) = O(t^{\max(-d,-1-\varsigma)})$, see (A.41) in Lemma A.8. The proof follows similarly as in the proof of (A.73).
Proof of (A.75): We observe that we can establish an upper bound for $|c_s t(\vartheta)|$ as $c\log^3(t)t^{\max(-d,-1-\varsigma)}$,

see Lemma A.8, where c is a generic arbitrarily large positive constant. Consequently, we proceed to evaluate the summation

$$\begin{aligned}\sum_{t=1}^{T} c_{st}(\vartheta)c_{it}(\vartheta) &\leq c\sum_{t=1}^{T}\log^6(t)t^{\max(-2d,-2-2\varsigma)}\\ &\leq c\log^6(T)T^{\max(1-2d,-1-2\varsigma)},\end{aligned}$$

where the last inequality follows from (A.30) in Lemma A.6. Since d and ς are both greater than 1/2, this term is $O(1)$.

□

Lemma A.15. *Suppose that Assumptions 3.2-3.4 holds. Let $d > 1/2$, then we have that:*

$$m(\vartheta) = 1 + O(T^{-1}), \tag{A.76}$$
$$m_i(\vartheta) = O(T^{-1}), \tag{A.77}$$

where $i \in \{\vartheta_k, \vartheta_k\vartheta_j, \vartheta_k\vartheta_j, \vartheta_k\vartheta_j\vartheta_l\}$ and $k, j, l = 1, \ldots, p+1$.

Proof of Lemma A.15. Proof of (A.76): The expression for $m(\vartheta)$, as provided in (88), can be represented as

$$m(\vartheta) = e^{\frac{1}{T-1}\left(\sum_{t=1}^{T} c_t^2(\vartheta)\right)}.$$

By employing the expansion $e^b = \sum_{k=0}^{\infty}\frac{b^k}{k!}$ and considering (A.75) in Lemma A.14, we have that

$$m(\vartheta) = 1 + O\left(T^{-1}\right).$$

Proof of (A.77): The derivatives of $m(\vartheta)$ are given in Lemma A.4. Proof follows directly from (A.75) in Lemma A.14. □

Lemma A.16. *Suppose that Assumptions 3.1-3.4 hold. Let $d_0 > \frac{1}{2}$. Then*

$$\sum_{t=1}^{T} S_{st}^{+}c_{it} = O_P(1) \tag{A.78}$$

where $s \in \{0, \vartheta_{\tilde{k}}, \vartheta_{\tilde{k}}\vartheta_{\tilde{k}}, \vartheta_{\tilde{k}}\vartheta_j, \vartheta_{\tilde{k}}\vartheta_{\tilde{k}}\vartheta_{\tilde{l}}\}$, $i \in \{0, \vartheta_k, \vartheta_k\vartheta_j, \vartheta_k\vartheta_j, \vartheta_k\vartheta_j\vartheta_l\}$ and $\tilde{k}, \tilde{j}, \tilde{l}, k, j, l = 1, \ldots, p+1$. Here, $c_{0t}(\vartheta)$ refers to $c_t(\vartheta)$ and S_{0t}^{+} to S_t^{+}.

Proof of Lemma A.16. Proof of (A.78): Note that $S_t^{+}(\vartheta_0) = \epsilon_t$, and as a consequence, the results for $s = 0$ directly follow from (A.75) in Lemma A.14. Next, we provide a general proof. To begin with, we observe that from Lemma A.1, S_{st}^{+} can be expressed as

$$S_{st}^{+} = \sum_{k=1}^{t-1} v_{st}\epsilon_{t-k},$$

where the weights v_{st} depend on s. From (A.34) in Lemma A.7 and (A.45) in Lemma A.8, it follows that $|v_{st}| \leq c\log^3(t)t^{-1}$. Also, from the proof of (A.75), we have established a bound for $|c_{st}(\vartheta)|$ as $c\log^3(t)t^{\max(-d,-1-\varsigma)}$.

Firstly, we note that

$$\sum_{t=T+1}^{\infty} S_{st}^{+} c_{it} = \sum_{k=1}^{\infty} \epsilon_k \sum_{t=\max(T,k)+1}^{\infty} c_{it} v_{s(t-k)}$$

For small $\delta > 0$, we bound $\log^3(t) \leq ct^{\delta}$ and use the bounds $|c_{st}(\vartheta)| \leq c\log^3(t) t^{\max(-d,-1-\zeta)} \leq ct^{\max(-d,-1)+\delta}$ and $|v_{sk}| \leq c\log^3(k)k^{-1} \leq ck^{-1+\delta}$, $t^{\max(-d,-1)+\delta} \leq (t-k)^{-2\delta} k^{\max(-d,-1)+2\delta}$. Then, we obtain

$$\begin{aligned} Var\left(\sum_{t=T+1}^{\infty} S_{st}^{+} c_{it}\right) &\leq c \sum_{k=1}^{\infty} \left(\sum_{t=\max(T,k)+1}^{\infty} (t-k)^{-\delta-1} k^{\max(-d,-1)+2\delta} \right)^2 \\ &\leq c \sum_{k=1}^{\infty} k^{\max(-2d,-2)+4\delta} \left(\sum_{t=\max(T,k)+1}^{\infty} (t-k)^{-\delta-1} \right)^2. \end{aligned}$$

Since $\sum_{t=\max(T,k)+1}^{\infty} (t-k)^{-\delta-1} \to 0$ as $T \to \infty$ and because $\sum_{k=1}^{\infty} k^{\max(-2d,-2)+4\delta} < \infty$, we conclude, by the dominated convergence theorem, that this variance converges to zero. □

Lemma A.17. *Let the model for the data x_t, $t = 1,\dots,T$, be given by* (63) *and let Assumptions 3.1-3.4 be satisfied with $d_0 > 1/2$. Then the normalized derivatives of the likelihood function L^*, see* (73), *satisfy*

$$\sigma_0^{-2} T^{-1/2} D_{\vartheta} L^*(\vartheta_0) = A_0 + T^{-1/2} A_1, \tag{A.79}$$

$$\sigma_0^{-2} T^{-1} D_{\vartheta\vartheta'} L^*(\vartheta_0) = B_0 + T^{-1/2} B_1 + O_P(T^{-1}\log(T)), \tag{A.80}$$

$$\sigma_0^{-2} T^{-1} D_{\vartheta_i\vartheta\vartheta'} L^*(\vartheta_0) = C_{0i} + O_P(T^{-1/2}), \tag{A.81}$$

for $i = 1,\dots,p+1$ and where

$$\begin{aligned} &A_0 = M_{0\vartheta}^{+}, \quad E(A_1) = E(\sigma_0^{-2} D_{\vartheta} L^*(\vartheta_0)) = O(1), \\ &B_0 = A, \quad B_1 = M_{\vartheta,\vartheta' T}^{+} + M_{0,\vartheta\vartheta' T}^{+}, \end{aligned}$$

Here, $M_{0\vartheta}^{+}$, $M_{0,\vartheta\vartheta' T}^{+}$ and $M_{\vartheta,\vartheta' T}^{+}$ are given in (A.51), (A.54), (A.55) *respectively, and A is the inverse of the variance-covariance matrix given in* (67). *The expression for C_{0i}, $i = 1,\dots,p+1$, is given in* (A.82) *and* (A.83).

Proof of Lemma A.17. Proof of (A.79): From Lemma A.2, we have that

$$\begin{aligned} \sigma_0^{-2} T^{-1/2} D_{\vartheta_k} L^* &= \sigma_0^{-2} T^{-1/2} \sum_{t=1}^{T} S_t^{+} S_{\vartheta_k t}^{+} - \sigma_0^{-2} T^{-1/2} \left(\mu(\vartheta_0) - \mu_0\right) \sum_{t=1}^{T} S_t^{+} c_{\vartheta_k t} \\ &\quad - \sigma_0^{-2} T^{-1/2} \left(\mu(\vartheta_0) - \mu_0\right) \sum_{t=1}^{T} S_{\vartheta_k t}^{+} c_t + \sigma_0^{-2} T^{-1/2} \left(\mu(\vartheta_0) - \mu_0\right)^2 \sum_{t=1}^{T} c_t c_{\vartheta_k t} \\ &= M_{0\vartheta_k}^{+} + T^{-1/2} A_{1k}, \end{aligned}$$

with elements of A_1 given by

$$A_1(k) = -\sigma_0^{-2} \left(\mu(\vartheta_0) - \mu_0\right) \sum_{t=1}^{T} S_t^{+} c_{\vartheta_k t}$$

$$- \sigma_0^{-2} \left(\mu(\vartheta_0) - \mu_0\right) \sum_{t=1}^{T} S_{\vartheta_k t}^{+} c_{0t} + \sigma_0^{-2} \left(\mu(\vartheta_0) - \mu_0\right)^2 \sum_{t=1}^{T} c_t c_{\vartheta_k t},$$

since $E(M_{0\vartheta_k}^{+}) = 0$ it follows that $E\left(A_1(k)\right) = E\left(\sigma_0^{-2} D_{\vartheta_k} L^*\right)$ and from Lemmata A.13 and A.14 we find that $E\left(\sigma_0^{-2} D_{\vartheta_k} L^*\right) = O(1)$.

Proof of (A.80): From Lemma A.2 we have that

$$\sigma_0^{-2} T^{-1} D_{\vartheta_k \vartheta_j} L^* = \sigma_0^{-2} T^{-1} L_{\vartheta_k \vartheta_j} - \sigma_0^{-2} T^{-1} \frac{L_{\mu\vartheta_j} L_{\mu\vartheta_k}}{L_{\mu\mu}},$$

where $\sigma_0^{-2} T^{-1} L_{\vartheta_k \mu} \mu_{\vartheta_j} / L_{\mu\mu} = O_P(T^{-1})$ from Lemmata A.14 and A.16. Thus we get

$$\begin{aligned}
\sigma_0^{-2} T^{-1} D_{\vartheta_k \vartheta_j} L^* = {} & \sigma_0^{-2} T^{-1} \sum_{t=1}^{T} \left(S_{\vartheta_j t}^{+} - c_{\vartheta_j t}(\vartheta_0) \left(\mu(\vartheta_0) - \mu_0\right)\right) \left(S_{\vartheta_k t}^{+} - c_{\vartheta_k t}(\vartheta_0) \left(\mu(\vartheta_0) - \mu_0\right)\right) \\
& + \sigma_0^{-2} T^{-1} \sum_{t=1}^{T} \left(S_t^{+} - c_t(\vartheta_0) \left(\mu(\vartheta_0) - \mu_0\right)\right) \left(S_{\vartheta_k \vartheta_j t}^{+} - c_{\vartheta_k \vartheta_j t}(\vartheta_0) \left(\mu(\vartheta_0) - \mu_0\right)\right) \\
& + O_P(T^{-1}),
\end{aligned}$$

ignoring terms that are of order T^{-1} we get

$$\begin{aligned}
\sigma_0^{-2} T^{-1} D_{\vartheta_k \vartheta_j} L^* &= \sigma_0^{-2} T^{-1} \sum_{t=1}^{T} S_{\vartheta_j t}^{+} S_{\vartheta_k t}^{+} + \sigma_0^{-2} T^{-1} \sum_{t=1}^{T} S_t^{+} S_{\vartheta_k \vartheta_j t}^{+} + O_P(T^{-1}) \\
&= \sigma_0^{-2} T^{-1} \sum_{t=1}^{T} E S_{\vartheta_j t}^{+} S_{\vartheta_k t}^{+} + T^{-1/2} \left(M_{\vartheta_j, \vartheta_k T}^{+} + M_{0, \vartheta_j \vartheta_k T}^{+}\right) + O_P(T^{-1}).
\end{aligned}$$

We notice that $\sigma_0^{-2} T^{-1} \sum_{t=1}^{T} E\left(S_{\vartheta_j t}^{+} S_{\vartheta_k t}^{+}\right) = E\left(M_{0,\vartheta_j} M_{0,\vartheta_k}\right)$ and is already convered in Lemma A.9.

Proof of (A.81): For the third derivative it can be shown from Lemmata A.14 and A.16 that the extra terms involving derivatives μ_{ϑ_k} and $\mu_{\vartheta_k \vartheta_k}$, see Lemma A.2, can be ignored and we find

$$\begin{aligned}
\sigma_0^{-2} T^{-1} D_{\vartheta_k \vartheta_j \vartheta_l} L^* = {} & \sigma_0^{-2} T^{-1} \sum_{t=1}^{T} \left(S_{\vartheta_j \vartheta_l t}^{+} - c_{\vartheta_j \vartheta_l t} \left(\mu(\vartheta_0) - \mu_0\right)\right) \left(S_{\vartheta_k t}^{+} - c_{\vartheta_k t} \left(\mu(\vartheta_0) - \mu_0\right)\right) \\
& + \sigma_0^{-2} T^{-1} \sum_{t=1}^{T} \left(S_{\vartheta_j}^{+} - c_{\vartheta_j t} \left(\mu(\vartheta_0) - \mu_0\right)\right) \left(S_{\vartheta_k \vartheta_l t}^{+} - c_{\vartheta_k \vartheta_l t} \left(\mu(\vartheta_0) - \mu_0\right)\right) \\
& + \sigma_0^{-2} T^{-1} \sum_{t=1}^{T} \left(S_{\vartheta_l t}^{+} - c_{\vartheta_l t} \left(\mu(\vartheta_0) - \mu_0\right)\right) \left(S_{\vartheta_k \vartheta_j t}^{+} - c_{\vartheta_k \vartheta_j t} \left(\mu(\vartheta_0) - \mu_0\right)\right) \\
& + O_P(T^{-1}) \\
= {} & \sigma_0^{-2} T^{-1} \sum_{t=1}^{T} E\left(S_{\vartheta_j \vartheta_l t}^{+} S_{\vartheta_k t}^{+}\right) + \sigma_0^{-2} T^{-1} \sum_{t=1}^{T} E\left(S_{\vartheta_j}^{+} S_{\vartheta_k \vartheta_l t}^{+}\right) \\
& + \sigma_0^{-2} T^{-1} \sum_{t=1}^{T} E\left(S_{\vartheta_l t}^{+} S_{\vartheta_k \vartheta_j t}^{+}\right) \\
& + T^{-1/2} \left(M_{0, \vartheta_k \vartheta_j \vartheta_l T}^{+} + M_{\vartheta_k, \vartheta_j \vartheta_l T}^{+} + M_{\vartheta_j, \vartheta_l \vartheta_k T}^{+} + M_{\vartheta_l, \vartheta_j \vartheta_k T}^{+}\right) + O_P(T^{-1})
\end{aligned}$$

$$
\begin{aligned}
&= \sigma_0^{-2} T^{-1} \sum_{t=1}^{T} E\left(S^{+}_{\vartheta_j \vartheta_l t} S^{+}_{\vartheta_k t}\right) + \sigma_0^{-2} T^{-1} \sum_{t=1}^{T} E\left(S^{+}_{\vartheta_j} S^{+}_{\vartheta_k \vartheta_l t}\right) \\
&\quad + \sigma_0^{-2} T^{-1} \sum_{t=1}^{T} E\left(S^{+}_{\vartheta_l t} S^{+}_{\vartheta_k \vartheta_j t}\right) + O_P(T^{-1/2}),
\end{aligned}
$$

where the second-to-last equality uses Lemmata A.14 and A.16 and the last equality uses Lemma A.9. The terms in this expression are given in Lemma A.10. In matrix notation, we can therefore define C_{0i} in (A.81) as follows

$$
C_{01} = \begin{pmatrix} C_{01}(1,1) & C_{01}(1,2) \\ C_{01}(2,1) & C_{01}(2,2) \end{pmatrix}, \tag{A.82}
$$

where the elements are given by

$$
\begin{aligned}
C_{01}(1,1) &= -6\zeta_3, \\
C_{01}(1,2) &= 2\sum_{i=2}^{\infty} i^{-1} h_{d\varphi' i}(\varphi_0) + \sum_{i=0}^{\infty} D_{dd}\pi_i(0) b_{\varphi' i}(\varphi_0), \\
C_{01}(2,1) &= 2\sum_{i=2}^{\infty} i^{-1} h_{d\varphi i}(\varphi_0) + \sum_{i=0}^{\infty} D_{dd}\pi_i(0) b_{\varphi i}(\varphi_0), \\
C_{01}(2,2) &= -\sum_{i=1}^{\infty} i^{-1} b_{\varphi\varphi' i}(\varphi_0) - \sum_{i=2}^{\infty} b_{\vartheta i}(\varphi_0) h_{d\vartheta' i}(\varphi_0) - \left(\sum_{i=2}^{\infty} b_{\vartheta i}(\varphi_0) h_{d\vartheta' i}(\varphi_0)\right)',
\end{aligned}
$$

and for $k = 1, \ldots, p$ we have that

$$
C_{0(k+1)} = \begin{pmatrix} C_{0(k+1)}(1,1) & C_{0(k+1)}(1,2) \\ C_{0(k+1)}(2,1) & C_{0(k+1)}(2,2), \end{pmatrix} \tag{A.83}
$$

where the elements are given by

$$
\begin{aligned}
C_{0(k+1)}(1,1) &= 2\sum_{i=2}^{\infty} i^{-1} h_{d\varphi_k i}(\varphi_0) + \sum_{i=0}^{\infty} D_{dd}\pi_i(0) b_{\varphi_k i}(\varphi_0), \\
C_{0(k+1)}(1,2) &= -\sum_{i=1}^{\infty} i^{-1} b_{\varphi'\varphi_k i}(\varphi_0) - \sum_{i=2}^{\infty} b_{\varphi_k i}(\varphi_0) h_{d\varphi' i}(\varphi_0) - \sum_{i=2}^{\infty} b_{\varphi' i}(\varphi_0) h_{d\varphi_k i}(\varphi_0), \\
C_{0(k+1)}(2,1) &= -\sum_{i=1}^{\infty} i^{-1} b_{\varphi\varphi_k i}(\varphi_0) - \sum_{i=2}^{\infty} b_{\varphi_k i}(\varphi_0) h_{d\varphi i}(\varphi_0) - \sum_{i=2}^{\infty} b_{\varphi i}(\varphi_0) h_{d\varphi_k i}(\varphi_0), \\
C_{0(k+1)}(2,2) &= \left(\sum_{i=1}^{\infty} b_{\varphi i}(\varphi_0) b_{\varphi'\varphi_k i}(\varphi_0)\right)' + \sum_{i=1}^{\infty} b_{\varphi i}(\varphi_0) b_{\varphi'\varphi_k i}(\varphi_0) + \sum_{i=1}^{\infty} b_{\varphi_k i}(\varphi_0) b_{\varphi\varphi' i}(\varphi_0).
\end{aligned}
$$

□

Lemma A.18. *Let the model for the data x_t, $t = 1,\ldots,T$, be given by* (63) *and let Assumptions 3.1-3.4 be satisfied with $d_0 > 1/2$. Then the normalized derivatives of the likelihood function $L^*_{\mu_0}$, see* (77), *satisfy*

$$
\sigma_0^{-2} T^{-1/2} D_{\vartheta} L^*_{\mu_0}(\vartheta_0) = A_0, \tag{A.84}
$$

$$
\sigma_0^{-2} T^{-1} D_{\vartheta\vartheta'} L^*_{\mu_0}(\vartheta_0) = B_0 + T^{-1/2} B_1 + O_P(T^{-1}\log(T)), \tag{A.85}
$$

$$
\sigma_0^{-2} T^{-1} D_{\vartheta_i\vartheta\vartheta'} L^*(\vartheta_0) = C_{0i} + O_P(T^{-1/2}), \tag{A.86}
$$

for $i = 1, \ldots, p+1$ and where

$$A_0 = M^{+}_{0\vartheta},$$
$$B_0 = A, \qquad B_1 = M^{+}_{\vartheta,\vartheta' T} + M^{+}_{0,\vartheta\vartheta' T},$$

Here, $M^{+}_{0\vartheta}$, $M^{+}_{0,\vartheta\vartheta' T}$ and $M^{+}_{\vartheta,\vartheta' T}$ are given in (A.51), (A.54) *and* (A.55)*, respectively, and A is the inverse of the variance-covariance marix given in* (67). *The expression for C_{0i}, $i = 1, \ldots, p+1$, is given in* (A.82) *and* (A.83)*.*

Proof of Lemma A.18. The proof is omitted and follows from the same approach as in the proof of Lemma A.17 but is much easier since the constant term is known. □

Lemma A.19. *Let the model for the data x_t, t = 1,…,T, be given by* (63) *and let Assumptions 3.1-3.4 be satisfied with $d_0 > 1/2$. Then the normalized derivatives of the likelihood function L^*_m, see* (83)*, satisfy*

$$\sigma_0^{-2} T^{-1/2} D_\vartheta L^*_m(\vartheta_0) = A_0 + T^{-1/2} A_1 + O(T^{-1}), \tag{A.87}$$
$$\sigma_0^{-2} T^{-1} D_{\vartheta\vartheta'} L^*_m(\vartheta_0) = B_0 + T^{-1/2} B_1 + O_P(T^{-1}\log(T)), \tag{A.88}$$
$$\sigma_0^{-2} T^{-1} D_{\vartheta_i\vartheta\vartheta'} L^*(\vartheta_0) = C_{0i} + O_P(T^{-1/2}), \tag{A.89}$$

for $i = 1, \ldots, p+1$ and where

$$A_0 = M^{+}_{0\vartheta}, \qquad E(A_1) = E(\sigma_0^{-2} D_\vartheta L^*(\vartheta_0)) = 0,$$
$$B_0 = A, \qquad B_1 = M^{+}_{\vartheta,\vartheta' T} + M^{+}_{0,\vartheta\vartheta' T},$$

Here, $M^{+}_{0\vartheta}$, $M^{+}_{0,\vartheta\vartheta' T}$ and $M^{+}_{\vartheta,\vartheta' T}$ are given in (A.51), (A.54) *and* (A.55)*, respectively, and A is the inverse of the variance-covariance matrix given in* (67). *The expression for C_{0i}, $i = 1, \ldots, p+1$, is given in* (A.82) *and* (A.83)*.*

Proof of Lemma A.19. The proof is omitted and follows from Lemma A.17 and the asymptotic behaviour of the modification term and its derivatives in Lemma A.15. □

A.3.2 Stationary region

In Lemmata A.20 and A.22, we investigate the order of magnitude of functions involving the deterministic term $c_t(\vartheta)$ and its derivatives and the stochastic term S^{+}_t and its derivatives and the product moments containing these. In Lemma A.21, we investigate the order of magnitude involving the modification term $m(\vartheta)$ and derivatives of these. These lemmata are then used to find asymptotic results for the first three derivatives of L^*, $L^*_{\mu_0}$ and L^*_m in Lemmata A.23, A.24, and A.25, respectively.

Lemma A.20. *Suppose that Assumptions 3.2-3.4 holds. Let $d < 1/2$, then we have that:*

$$\frac{1}{T^{1-2d}} \sum_{t=1}^{T} c_t^2(\vartheta) = \phi^2(1;\varphi) \frac{1}{T^{1-2d}} \sum_{t=1}^{T} \kappa_{0t}^2(d) + o(1), \tag{A.90}$$

$$\frac{1}{T^{1-2d}}\sum_{t=1}^{T}c_t(\vartheta)c_{dt}(\vartheta) = \phi^2(1;\varphi)\frac{1}{T^{1-2d}}\sum_{t=1}^{T}\kappa_{0t}(d)\kappa_{1t}(d) + o(1), \tag{A.91}$$

$$\frac{1}{T^{1-2d}}\sum_{t=1}^{T}c_t(\vartheta)c_{\varphi_k t}(\vartheta) = \phi(1;\varphi)D_{\varphi_k}\phi(1;\varphi)\frac{1}{T^{1-2d}}\sum_{t=1}^{T}\kappa_{0t}^2(d) + o(1), \tag{A.92}$$

$$\begin{aligned}\frac{1}{T^{1-2d}}\sum_{t=1}^{T}\kappa_{0t}(d)\kappa_{1t}(d) = &-(\log(T) - \Psi(1-d))\frac{1}{T^{1-2d}}\sum_{t=1}^{T}\kappa_{0t}^2(d)\\ &+ \frac{1}{\Gamma(1-d)^2(1-2d)^2} + o(1),\end{aligned} \tag{A.93}$$

$$\frac{1}{T^{1-2d}}\sum_{t=1}^{T}k_{0t}^2(d) \to \frac{1}{\Gamma(1-d)^2(1-2d)}, \tag{A.94}$$

$$\sum_{t=1}^{T}c_t(\vartheta)c_{\vartheta_k t}(\vartheta) = O(T^{1-2d}\log(T)), \tag{A.95}$$

$$\sum_{t=1}^{T}c_t(\vartheta)c_{\vartheta_k\vartheta_j t}(\vartheta) = O(T^{1-2d}\log^2(T)), \tag{A.96}$$

$$\sum_{t=1}^{T}c_t(\vartheta)c_{\vartheta_k\vartheta_j\vartheta_l t}(\vartheta) = O(T^{1-2d}\log^3(T)), \tag{A.97}$$

$$\sum_{t=1}^{T}c_{\vartheta_k t}(\vartheta)c_{\vartheta_j t}(\vartheta) = O(T^{1-2d}\log^2(T)), \tag{A.98}$$

$$\sum_{t=1}^{T}c_{\vartheta_k t}(\vartheta)c_{\vartheta_j\vartheta_l t}(\vartheta) = O(T^{1-2d}\log^3(T)), \tag{A.99}$$

for $k, j, l = 1, \dots, p+1$.

Proof of Lemma A.20. Proof of (A.90): See Hualde & Nielsen (2020, Lemma S.15).

Proof of (A.91): By summation by parts

$$\begin{aligned}c_t(\vartheta) = \sum_{j=0}^{t-1}\phi_j(\varphi)\kappa_{0(t-j)}(d) = \kappa_{0t}(d)&\sum_{j=0}^{t-1}\phi_j(\varphi)\\ -\sum_{j=0}^{t-2}\left(\kappa_{0(t-j)}(d) - \kappa_{0(t-j-1)}(d)\right)&\sum_{k=j+1}^{t-1}\phi_k(\varphi),\end{aligned}$$

From $\kappa_{0(t-j)}(d) - \kappa_{0(t-j-1)} = \pi_{t-j-1}(1-d) - \pi_{t-j-2}(1-d) = \pi_{t-j-1}(-d)$, see Johansen & Nielsen (2016, Lemma A.4), we have

$$\begin{aligned}c_t(\vartheta) = \kappa_{0t}(d)\sum_{j=0}^{\infty}\phi_j(\varphi) - \kappa_{0t}(d)&\sum_{j=t}^{\infty}\phi_j(\varphi)\\ -\sum_{j=0}^{t-2}\pi_{t-j-1}(-d)&\sum_{k=j+1}^{t-1}\phi_k(\varphi),\end{aligned}$$

Notice that

$$\sum_{j=0}^{t-2}\pi_{t-j-1}(-d)\sum_{k=j+1}^{t-1}\phi_k(\varphi) = \sum_{j=1}^{t-1}\pi_j(-d)\sum_{k=1}^{j}\phi_{t-k}(\varphi),$$

therefore

$$c_t(\vartheta) = \kappa_{0t}(d)\sum_{j=0}^{\infty}\phi_j(\varphi) - \kappa_{0t}(d)\sum_{j=t}^{\infty}\phi_j(\varphi) - \sum_{j=1}^{t-1}\pi_j(-d)\sum_{k=1}^{j}\phi_{t-k}(\varphi), \tag{A.100}$$

Taking the derivative of $c_t(\vartheta)$ with respect to d gives

$$D_d c_t(\vartheta) = \kappa_{1t}(d)\sum_{j=0}^{\infty}\phi_j(\varphi) - \kappa_{1t}(d)\sum_{j=t}^{\infty}\phi_j(\varphi) - \sum_{j=1}^{t-1}D_d\pi_j(-d)\sum_{k=1}^{j}\phi_{t-k}(\varphi).$$

The first term of $c_t(\vartheta)$ is bounded by $O(t^{-d})$ from (A.34). The second term is bounded by $O(t^{-d-\varsigma})$ from (A.34) and (A.35) and from the same arguments the third term is bounded by $O\left(\sum_{j=1}^{t-1} j^{-d-1}\sum_{k=1}^{j}(t-k)^{-1-\varsigma}\right) = O(t^{-d-\varsigma})$ which also involves employing Lemma A.5.

The first term of $D_d c_t(\vartheta)$ is bounded by $O(\log(t)t^{-d})$ from (A.34). The second term is bounded by $O(\log(t)t^{-d-\varsigma})$ from (A.34) and (A.35) and from the same arguments the third term is bounded by $O\left(\sum_{j=1}^{t-1}\log(j)j^{-d-1}\sum_{k=1}^{j}(t-k)^{-1-\varsigma}\right) = O(\log(t)t^{-d-\varsigma})$ which also involves employing Lemma A.5.

The leading term of $\sum_{t=1}^{T} c_t(\vartheta)c_{dt}(\vartheta)$ involves only the first term of $c_t(\vartheta)$ and $D_d c_t(\vartheta)$ and the remainder term is bounded by

$$O(\sum_{t=1}^{T}\log(t)t^{-2d-\varsigma}) = O(\log(T)\sum_{t=1}^{T}t^{-2d-\varsigma})$$

This term is $O(\log(T))$ when $-2d-\varsigma < -1$, $O(\log^2(T))$ when $-2d-\varsigma = -1$, and $O(\log(T)T^{1-2d-\varsigma})$ when $-2d-\varsigma > -1$. The proof is now completed.

Proof of (A.92): Taking the derivative of $c_t(\vartheta)$ in (A.100) with respect to φ_k gives

$$D_{\varphi_k} c_t(\vartheta) = \kappa_{0t}(d)\sum_{j=0}^{\infty}D_{\varphi_k}\phi_j(\varphi) - \kappa_{0t}(d)\sum_{j=t}^{\infty}D_{\varphi_k}\phi_j(\varphi) - \sum_{j=1}^{t-1}\pi_j(-d)\sum_{k=1}^{j}D_{\varphi_k}\phi_{t-k}(\varphi).$$

The first term of $D_{\varphi_k}c_t(\vartheta)$ is bounded by $O(t^{-d})$ from (A.34) and (A.37). The second term is bounded by $O(t^{-d-\varsigma})$ from (A.34) and (A.37) and from the same arguments the third term is bounded by $O\left(\sum_{j=1}^{t-1} j^{-d-1}\sum_{k=1}^{j}(t-k)^{-1-\varsigma}\right) = O(t^{-d-\varsigma})$ which also involves employing Lemma A.5. The leading term of $\sum_{t=1}^{T} c_t(\vartheta)c_{\varphi_k t}(\vartheta)$ involves only the first term of $c_t(\vartheta)$ and $D_{\varphi_k}c_t(\vartheta)$ and the remainder term is bounded by

$$O(\sum_{t=1}^{T}t^{-2d-\varsigma}) = O(\sum_{t=1}^{T}t^{-2d-\varsigma}).$$

This term is $O(1)$ when $-2d-\varsigma<-1$, $O(\log(T))$ when $-2d-\varsigma=-1$, and $O(T^{1-2d-\varsigma})$ when $-2d-\varsigma>-1$. The proof is now completed.

Proof of (A.93): We note that

$$\kappa_{1t}(d) = -\kappa_{0t}(d)\left(\Psi(t-d)-\Psi(1-d)\right),$$

then

$$\frac{1}{T^{1-2d}}\sum_{t=1}^{T}\kappa_{0t}(d)\kappa_{1t}(d) = -\frac{1}{T^{1-2d}}\sum_{t=1}^{T}\kappa_{0t}^2(d)\Psi(t-d)+\Psi(1-d)\frac{1}{T^{1-2d}}\sum_{t=1}^{T}\kappa_{0t}^2(d), \quad \text{(A.101)}$$

We evaluate the first term in (A.101). We have that

$$\begin{aligned}\frac{1}{T^{1-2d}}\sum_{t=1}^{T}\kappa_{0t}^2(d)\Psi(t-d) &= \Psi(T-d)\frac{1}{T^{1-2d}}\sum_{t=1}^{T}\kappa_{0t}^2(d)\\ &\quad+\frac{1}{T^{1-2d}}\sum_{t=1}^{T}k_{0t}^2(d)\left(\Psi(t-d)-\Psi(T-d)\right). \qquad \text{(A.102)}\end{aligned}$$

For a fixed d,

$$\Psi(t+d)=\log(t)+O(t^{-1}),$$

see Abramowitz & Stegun (1964, eqn. 6.3.18), hence the second term in (A.102) is

$$\frac{1}{T^{1-2d}}\sum_{t=1}^{T}\kappa_{0t}^2(d)\left(\Psi(t-d)-\Psi(T-d)\right)=\frac{1}{T^{1-2d}}\sum_{t=1}^{T}\kappa_{0t}^2(d)\log(t/T)+o(1).$$

The first term in (A.102) is

$$\Psi(T-d)\frac{1}{T^{1-2d}}\sum_{t=1}^{T}\kappa_{0t}^2(d)=\log(T)\frac{1}{T^{1-2d}}\sum_{t=1}^{T}\kappa_{0t}^2(d)+o(1).$$

Thus we find for (A.102) that

$$\begin{aligned}\frac{1}{T^{1-2d}}\sum_{t=1}^{T}\kappa_{0t}^2(d)\Psi(t-d) &= \log(T)\frac{1}{T^{1-2d}}\sum_{t=1}^{T}\kappa_{0t}^2(d)\\ &\quad+\frac{1}{T^{1-2d}}\sum_{t=1}^{T}\kappa_{0t}^2(d)\log(t/T)+o(1). \qquad \text{(A.103)}\end{aligned}$$

The second term in (A.103) is

$$\begin{aligned}\frac{1}{T^{1-2d}}\sum_{t=1}^{T}\kappa_{0t}^2(d)\log(t/T) &= \frac{1}{\Gamma\left(1-d\right)^2}\frac{1}{T^{1-2d}}\sum_{t=1}^{T}\log(t/T)t^{-2d}+o(1)\\ &\to -\frac{1}{\Gamma(1-d)^2(1-2d)^2},\end{aligned}$$

from Stirling's approximation, see Abramowitz & Stegun (1964, page 257 6.1.47),

$$\pi_t(d)\sim\frac{1}{\Gamma(d)}t^{d-1}+O(t^{d-2}), \qquad \text{(A.104)}$$

and the last line follows from Hualde & Nielsen (2020, Lemma S.10). Thus (A.103) equals

$$\frac{1}{T^{1-2d}}\sum_{t=1}^{T}\kappa_{0t}^{2}(d)\Psi(t-d)=\log(T)\frac{1}{T^{1-2d}}\sum_{t=1}^{T}\kappa_{0t}^{2}(d)$$
$$-\frac{1}{\Gamma(1-d)^{2}(1-2d)^{2}}+o(1). \tag{A.105}$$

Plugging in (A.105) to (A.101) gives

$$\frac{1}{T^{1-2d}}\sum_{t=1}^{T}\kappa_{0t}(d)\kappa_{1t}(d)=-\left(\log(T)-\Psi(1-d)\right)\frac{1}{T^{1-2d}}\sum_{t=1}^{T}\kappa_{0t}^{2}(d)$$
$$+\frac{1}{\Gamma(1-d)^{2}(1-2d)^{2}}+o(1),$$

completing the proof.

Proof of (A.94):

From Stirling's approximation (A.104)

$$\frac{1}{T^{1-2d}}\sum_{t=1}^{T}k_{0t}^{2}(d)=\frac{1}{\Gamma\left(1-d\right)^{2}}\frac{1}{T^{1-2d}}\sum_{t=1}^{T}t^{-2d}+o(1)$$
$$\to\frac{1}{\Gamma(1-d)^{2}(1-2d)},$$

and the last line follows from Lemma A.5.

Proof of (A.95)-(A.99): The proofs can be straightforwardly deduced from the provided bounds in Lemma A.8, together with the application of Lemma A.5. □

Lemma A.21. *Suppose that Assumptions 3.2-3.4 holds. Let $d<1/2$, then we have that:*

$$m(\vartheta)=1+O(T^{-1}\log(T)), \tag{A.106}$$
$$m_{\vartheta_k}(\vartheta)=O(T^{-1}\log(T)), \tag{A.107}$$
$$m_{\vartheta_k\vartheta_j}(\vartheta)=O(T^{-1}\log^{2}(T)), \tag{A.108}$$
$$m_{\vartheta_k\vartheta_j\vartheta_l}(\vartheta)=O(T^{-1}\log^{3}(T)), \tag{A.109}$$

for $k,j,l=1,\dots,p+1$.

Proof of Lemma A.21. Proof of (A.106): The expression for $m(\vartheta)$, as provided in (88), can be represented as

$$m(\vartheta)=\left(\frac{1}{T^{1-2d}}\sum_{t=1}^{T}c_{t}^{2}(\vartheta)\right)^{\frac{1}{T-1}}\left(T^{1-2d}\right)^{\frac{1}{T-1}}.$$

By employing the expansion $e^{b}=\sum_{k=0}^{\infty}\frac{b^{k}}{k!}$ and considering (A.90) in Lemma A.20, we have that

$$\left(\frac{1}{T^{1-2d}}\sum_{t=1}^{T}c_{t}^{2}(\vartheta)\right)^{\frac{1}{T-1}}=e^{(T-1)^{-1}\log\left(T^{-1+2d}\sum_{t=1}^{T}\kappa_{0t}^{2}(d)\right)}=1+O(T^{-1}).$$

From the same expansion we have that

$$\left(T^{1-2d}\right)^{\frac{1}{T-1}} = e^{(T-1)^{-1}(1-2d)\log(T)}$$
$$= 1 + O(T^{-1}\log(T)).$$

We conclude that:

$$m(\vartheta) = (1 + O(T^{-1}))(1 + O(T^{-1}\log(T))) = 1 + O(T^{-1}\log(T)).$$

Proof of (A.107)-(A.109): The derivatives of $m(\vartheta)$ are given in Lemma A.4. Proof follows directly from (A.95)-(A.99) in Lemma A.20.

□

Lemma A.22. *Suppose that Assumptions 3.1-3.4 hold. Let $d_0 < \frac{1}{2}$. Then*

$$\sum_{t=1}^{T} S_t^+ c_t = O_P(T^{1/2-d_0}), \quad \sum_{t=1}^{T} S_t^+ c_{\vartheta_k t} = O_P(T^{1/2-d_0}\log(T)),$$
$$\sum_{t=1}^{T} S_t^+(\vartheta) c_{\vartheta_k \vartheta_j t} = O_P(T^{1/2-d_0}\log^2(T)), \tag{A.110}$$
$$\sum_{t=1}^{T} S_{\vartheta_l t}^+ c_t = O_P(T^{1/2-d_0}\log(T)), \sum_{t=1}^{T} S_{\vartheta_l t}^+ c_{\vartheta_k t} = O_P(T^{1/2-d_0}\log^2(T)),$$
$$\sum_{t=1}^{T} S_{\vartheta_l t}^+ c_{\vartheta_k \vartheta_j t} = O_P(T^{1/2-d_0}\log^3(T)), \tag{A.111}$$
$$\sum_{t=1}^{T} S_{\vartheta_l \vartheta_n t}^+ c_t = O_P(T^{1/2-d_0}\log^2(T)), \sum_{t=1}^{T} S_{\vartheta_l \vartheta_n t}^+ c_{\vartheta_k t} = O_P(T^{1/2-d_0}\log^3(T)),$$
$$\sum_{t=1}^{T} S_{\vartheta_l \vartheta_n t}^+ c_{\vartheta_k \vartheta_j t} = O_P(T^{1/2-d_0}\log^4(T)). \tag{A.112}$$

for $l, n, k, j = 0, \ldots, p+1$.

Proof of Lemma A.22. Proof of (A.110): Note that $S_t^+ = \epsilon_t$ such that the results follow from Lemma A.20.
Proof of (A.111): Due to their similarity and relative simplicity, we exclusively show $\sum_{t=1}^{T} S_{\vartheta_l t}^+ c_{\vartheta_z \vartheta_j t} = O_P(T^{1/2-d_0}\log^3(T))$, omitting the proofs for $\sum_{t=1}^{T} S_{\vartheta_l t}^+ c_t = O_P(T^{1/2-d_0}\log(T))$ and $\sum_{t=1}^{T} S_{\vartheta_l t}^+ c_{\vartheta_z t} = O_P(T^{1/2-d_0}\log^2(T))$. First, consider $l = 1$. By Lemma A.1

$$S_{\vartheta_1 t}^+ = -\sum_{k=0}^{t-1} D\pi_k(0)\epsilon_{t-k},$$

resulting in

$$\sum_{t=1}^{T} S_{\vartheta_1 t}^+(\vartheta) c_{\vartheta_z \vartheta_j k} = -\sum_{t=1}^{T-1} \epsilon_t \sum_{k=t+1}^{T} c_{\vartheta_z \vartheta_j k} D\pi_{k-t}(0).$$

Now, we analyse three scenarios: the first case involves $z = 1$ and $j = 1$; the second case involves $z = 1$ and $j > 1$; and the third case encompasses $z > 1$ and $j > 1$.

Case I: $z = 1$ and $j = 1$. First let $0 < d_0 < 1/2$. We use the following bounds $|c_{\vartheta_1\vartheta_1 t}| \leq K t^{-d_0}\log^2(t)$ and $|D\pi_t(0)| \leq K t^{-1} I(t \geq 1)$. Then

$$\begin{aligned}
Var(\sum_{t=1}^{T} S^+_{\vartheta_1 t} c_{\vartheta_1\vartheta_1 t}) &\leq K \sum_{t=1}^{T-1}\left(\sum_{k=t+1}^{T}\log^2(k)k^{-d_0}(k-t)^{-1}\right)^2 \\
&\leq K \sum_{t=1}^{T-1}\left(\sum_{k=1}^{T-t}\log^2(t+k)(t+k)^{-d_0}k^{-1}\right)^2 \\
&\leq K \sum_{t=1}^{T} t^{-2d_0}\log^4(T)\left(\sum_{k=1}^{T}k^{-1}\right)^2 \\
&\leq K T^{1-2d_0}\log^6(T),
\end{aligned}$$

because $\sum_{k=1}^{T} t^{-2d} = O(T^{1-2d})$ for $d < 1/2$.

Second, let $d_0 \leq 0$. Then

$$\begin{aligned}
Var(\sum_{t=1}^{T} S^+_{\vartheta_1 t} c_{\vartheta_1\vartheta_1 t}) &\leq K \sum_{t=1}^{T-1}\left(\sum_{k=t+1}^{T}\log^2(k)k^{-d_0}(k-t)^{-1}\right)^2 \\
&\leq K \sum_{t=1}^{T-1}\left(\sum_{k=1}^{T-t}\log^2(t+k)(t+k)^{-d_0}k^{-1}\right)^2 \\
&\leq K T^{-2d_0}\log^4(T)\sum_{t=1}^{T}\left(\sum_{k=1}^{T-t}k^{-1}\right)^2 \\
&\leq K T^{-2d_0}\log^4(T)\sum_{t=1}^{T}\log^2(T-t+1) \\
&\leq K T^{1-2d_0}\log^6(T),
\end{aligned}$$

because

$$\begin{aligned}
\sum_{t=1}^{T}\log^2(T-t+1) &= \sum_{t=1}^{T}\log^2(t) \\
&= \sum_{t=1}^{T}\log^2(t/T) + \sum_{t=1}^{T}\log^2(T)\log^2(t) \\
&= O(T) + O(T\log^2(T)) \\
&= O(T\log^2(T)).
\end{aligned}$$

This shows that $\sum_{t=1}^{T} S^+_{\vartheta_1 t}(\vartheta) c_{\vartheta_1\vartheta_1 k} = O_P(T^{1/2-d_0}\log^3(T))$.

Case II: $z = 1$ and $j > 1$. The proof of this case follows in a similar way to that in *Case I* but now with the bound $|c_{\vartheta_1\vartheta_j t}| \leq K t^{-d_0}\log(t)$. To conclude, $\sum_{t=1}^{T} S^+_{\vartheta_1 t}(\vartheta) c_{\vartheta_1\vartheta_j k} = O_P(T^{1/2-d_0}\log^2(T))$.

Case III: $z > 1$ and $j > 1$. The proof of this case follows in a similar way to that in *Case I* but now with the bound $|c_{\vartheta_z\vartheta_j t}| \leq K t^{-d_0}$. To conclude, $\sum_{t=1}^{T} S^+_{\vartheta_1 t}(\vartheta) c_{\vartheta_z\vartheta_j k} = O_P(T^{1/2-d_0}\log(T))$.

Finally, consider $l > 1$. By Lemma A.1

$$S^+_{\vartheta_l t} = \sum_{i=1}^{t-1} b_{\vartheta_l i}\epsilon_{t-i},$$

resulting in

$$\sum_{t=1}^{T} S^+_{\vartheta_l t} c_{\vartheta_z\vartheta_j t} = \sum_{t=1}^{T-1}\epsilon_t \sum_{k=t+1}^{T} c_{\vartheta_z\vartheta_j k} b_{\vartheta_l (k-t)}.$$

Now, we analyse the same three scenarios again: the first case involves $z = 1$ and $j = 1$; the second case involves $z = 1$ and $j > 1$; and the third case encompasses $z > 1$ and $j > 1$.

Case I: $z = 1$ and $j = 1$. Given a small $\delta > 0$ to be chosen later, we use the following bounds $|c_{\vartheta_1\vartheta_1 t}| \leq Kt^{-d_0}\log^2(t)$ and $|b_{\vartheta_j t}| \leq Kt^{-1-\varsigma+\delta}$. Then the

$$\begin{aligned}
Var(\sum_{t=1}^{T} S^+_{\vartheta_l t}(\vartheta)c_{\vartheta_1\vartheta_1 t}(\vartheta)) &\leq K\sum_{t=1}^{T-1}\left(\sum_{k=t+1}^{T}\log^2(k)k^{-d_0}(k-t)^{-1-\varsigma+\delta}\right)^2 \\
&\leq K\sum_{t=1}^{T-1}\left(\sum_{k=1}^{T-t}\log^2(t+k)(t+k)^{-d_0}k^{-1-\varsigma+\delta}\right)^2 \\
&\leq K\sum_{t=1}^{T} t^{-2d_0}\log^4(T)\left(\sum_{k=1}^{T}k^{-1-\varsigma+\delta}\right)^2 \\
&\leq KT^{1-2d_0}\log^4(T),
\end{aligned}$$

because $\sum_{k=1}^{T} k^{-1-\varsigma+\delta} = O(1)$ since $-\varsigma+\delta < 0$ from choosing δ to satisfy $0 < \delta < \varsigma < 1/2$. This show that $\sum_{t=1}^{T} S^+_{\vartheta_l t}c_{\vartheta_1\vartheta_1 t} = O_P(T^{1/2-d_0}\log^2(T))$.

Case II: $z = 1$ and $j > 1$. The proof of this case follows in a similar way to that in *Case I* but now with the bound $|c_{\vartheta_1\vartheta_j t}| \leq Kt^{-d_0}\log(t)$. To conclude that $\sum_{t=1}^{T} S^+_{\vartheta_l t}(\vartheta)c_{\vartheta_1\vartheta_j k} = O_P(T^{1/2-d_0}\log(T))$.

Case III: $z > 1$ and $j > 1$. The proof of this case follows in a similar way to that in *Case I* but now with the bound $|c_{\vartheta_z\vartheta_j t}| \leq Kt^{-d_0}$. To conclude that $\sum_{t=1}^{T} S^+_{\vartheta_l t}c_{\vartheta_z\vartheta_j k} = O_P(T^{1/2-d_0})$.

Proof of (A.112): The proofs follow from similar arguments as in the proof of (A.111) and are therefore omitted.

□

Lemma A.23. *Let the model for the data x_t, $t = 1,\ldots,T$, be given by* (63) *and let Assumptions 3.1-3.4 be satisfied with $d_0 < 1/2$. Then the normalized derivatives of the likelihood function L^*, see* (73), *satisfy*

$$\sigma_0^{-2}T^{-1/2}D_\vartheta L^*(\vartheta_0) = A_0 + T^{-1/2}A_1, \tag{A.113}$$

$$\sigma_0^{-2}T^{-1}D_{\vartheta\vartheta'}L^*(\vartheta_0) = B_0 + T^{-1/2}B_1 + O_P(T^{-1}\log^2(T)), \tag{A.114}$$

$$\sigma_0^{-2} T^{-1} D_{\vartheta_i \vartheta \vartheta'} L^*(\vartheta_0) = C_{0i} + O_P(T^{-1/2}), \tag{A.115}$$

for $i = 1, \ldots, p+1$ and where

$$A_0 = M^+_{0\vartheta}, \qquad E(A_1) = E(\sigma_0^{-2} D_\vartheta L^*(\vartheta_0)) = O(\log(T)),$$
$$B_0 = A, \qquad B_1 = M^+_{\vartheta,\vartheta' T} + M^+_{0,\vartheta\vartheta' T},$$

Here, $M^+_{0\vartheta}$, $M^+_{0,\vartheta\vartheta' T}$ and $M^+_{\vartheta,\vartheta' T}$ are given in (A.51), (A.54) *and* (A.55), *respectively, and A is the inverse of the variance-covariance matrix given in* (67).*The expression for C_{0i}, $i = 1, \ldots, p+1$, is given in* (A.82) *and* (A.83).

Proof of Lemma A.23. Proof of (A.113): From Lemma A.2, we have that

$$\begin{aligned}
\sigma_0^{-2} T^{-1/2} D_{\vartheta_k} L^* &= \sigma_0^{-2} T^{-1/2} \sum_{t=1}^T S_t^+ S^+_{\vartheta_k t} - \sigma_0^{-2} T^{-1/2} \left(\mu(\vartheta_0) - \mu_0\right) \sum_{t=1}^T S_t^+ c_{\vartheta_k t} \\
&\quad - \sigma_0^{-2} T^{-1/2} \left(\mu(\vartheta_0) - \mu_0\right) \sum_{t=1}^T S^+_{\vartheta_k t} c_t + \sigma_0^{-2} T^{-1/2} \left(\mu(\vartheta_0) - \mu_0\right)^2 \sum_{t=1}^T c_t c_{\vartheta_k t} \\
&= M^+_{0\vartheta_k} + T^{-1/2} A_1,
\end{aligned}$$

with elements of A_1 given by

$$\begin{aligned}
A_1(k) &= -\sigma_0^{-2} \left(\mu(\vartheta_0) - \mu_0\right) \sum_{t=1}^T S_t^+ c_{\vartheta_k t} \\
&\quad - \sigma_0^{-2} \left(\mu(\vartheta_0) - \mu_0\right) \sum_{t=1}^T S^+_{\vartheta_k t} c_{0t} + \sigma_0^{-2} \left(\mu(\vartheta_0) - \mu_0\right)^2 \sum_{t=1}^T c_t c_{\vartheta_k t},
\end{aligned}$$

since $E(M^+_{0\vartheta_k}) = 0$ it follows that $E\left(A_1(k)\right) = E\left(\sigma_0^{-2} D_{\vartheta_k} L^*\right)$ and from Lemmata A.13 and A.20 we find that $E(A) = O(\log(T))$.

Proof of (A.114): From Lemma A.2 we have that

$$\sigma_0^{-2} T^{-1} D_{\vartheta_k \vartheta_j} L^* = \sigma_0^{-2} T^{-1} L_{\vartheta_k \vartheta_j} - \sigma_0^{-2} T^{-1} \frac{L_{\mu\vartheta_j} L_{\mu\vartheta_k}}{L_{\mu\mu}},$$

where $\sigma_0^{-2} T^{-1} L_{\vartheta_k \mu} \mu_{\vartheta_j} / L_{\mu\mu} = O_P(T^{-1} \log^2(T)$ from Lemmata A.20 and A.22. Thus we get

$$\begin{aligned}
\sigma_0^{-2} T^{-1} D_{\vartheta_k \vartheta_j} L^* &= \sigma_0^{-2} T^{-1} \sum_{t=1}^T \left(S^+_{\vartheta_j t} - c_{\vartheta_j t}(\vartheta_0) \left(\mu(\vartheta_0) - \mu_0\right)\right) \left(S^+_{\vartheta_k t} - c_{\vartheta_k t}(\vartheta_0) \left(\mu(\vartheta_0) - \mu_0\right)\right) \\
&\quad + \sigma_0^{-2} T^{-1} \sum_{t=1}^T \left(S_t^+ - c_t(\vartheta_0) \left(\mu(\vartheta_0) - \mu_0\right)\right) \left(S^+_{\vartheta_k \vartheta_j t} - c_{\vartheta_k \vartheta_j t}(\vartheta_0) \left(\mu(\vartheta_0) - \mu_0\right)\right) \\
&\quad + O_P(T^{-1} \log^2(T)),
\end{aligned}$$

ignoring terms that are of order $T^{-1} \log^2(T)$ we get

$$\sigma_0^{-2} T^{-1} D_{\vartheta_k \vartheta_j} L^* = \sigma_0^{-2} T^{-1} \sum_{t=1}^T S^+_{\vartheta_j t} S^+_{\vartheta_k t} + \sigma_0^{-2} T^{-1} \sum_{t=1}^T S_t^+ S^+_{\vartheta_k \vartheta_j t} + O_P(T^{-1})$$

$$= \sigma_0^{-2} T^{-1} \sum_{t=1}^{T} E S^+_{\vartheta_j t} S^+_{\vartheta_k t} + T^{-1/2} \left(M^+_{\vartheta_j, \vartheta_k T} + M^+_{0, \vartheta_j \vartheta_k T} \right) + O_P(T^{-1} \log^2(T)).$$

We notice that $\sigma_0^{-2} T^{-1} \sum_{t=1}^{T} E \left(S^+_{\vartheta_j t} S^+_{\vartheta_k t} \right) = E \left(M_{0,\vartheta_j} M_{0,\vartheta_k} \right)$ and is already convered in Lemma A.9.

Proof of (A.115): The proof strategy closely resembles that used in the derivation of (A.81) in Lemma A.17 and is thus omitted. It is worth noting that, for this approximation, the terms provided in Lemmata A.20 and A.22 can be considered negligible.

□

Lemma A.24. *Let the model for the data x_t, $t = 1,\ldots,T$, be given by* (63) *and let Assumptions 3.1-3.4 be satisfied with $d_0 < 1/2$. Then the normalized derivatives of the likelihood function $L^*_{\mu_0}$, see* (77)*, satisfy*

$$\sigma_0^{-2} T^{-1/2} D_\vartheta L^*_{\mu_0}(\vartheta_0) = A_0, \tag{A.116}$$

$$\sigma_0^{-2} T^{-1} D_{\vartheta\vartheta'} L^*_{\mu_0}(\vartheta_0) = B_0 + T^{-1/2} B_1 + O_P(T^{-1} \log(T)), \tag{A.117}$$

$$\sigma_0^{-2} T^{-1} D_{\vartheta_i \vartheta\vartheta'} L^*(\vartheta_0) = C_{0i} + O_P(T^{-1/2}), \tag{A.118}$$

for $i = 1, \ldots, p+1$ and where

$$A_0 = M^+_{0\vartheta},$$
$$B_0 = A, \qquad B_1 = M^+_{\vartheta,\vartheta' T} + M^+_{0,\vartheta\vartheta' T},$$

Here, $M^+_{0\vartheta}$, $M^+_{0,\vartheta\vartheta' T}$ and $M^+_{\vartheta,\vartheta' T}$ are given in (A.51), (A.54) *and* (A.55)*, respectively, and A is the inverse of the variance-covariance marix given in* (67)*. The expression for C_{0i}, $i = 1, \ldots, p+1$, is given in* (A.82) *and* (A.83)*.*

Proof of Lemma A.24. The proof is omitted and follows from the same approach as in the proof of Lemma A.23 but is much easier since the constant term is known. □

Lemma A.25. *Let the model for the data x_t, $t = 1,\ldots,T$, be given by* (63) *and let Assumptions 3.1-3.4 be satisfied with $d_0 < 1/2$. Then the normalized derivatives of the likelihood function L^*_m, see* (83)*, satisfy*

$$\sigma_0^{-2} T^{-1/2} D_\vartheta L^*_m(\vartheta_0) = A_0 + T^{-1/2} A_1 + O_P(T^{-1} \log(T)), \tag{A.119}$$

$$\sigma_0^{-2} T^{-1} D_{\vartheta\vartheta'} L^*_m(\vartheta_0) = B_0 + T^{-1/2} B_1 + O_P(T^{-1} \log^2(T)), \tag{A.120}$$

$$\sigma_0^{-2} T^{-1} D_{\vartheta_i \vartheta\vartheta'} L^*(\vartheta_0) = C_{0i} + O_P(T^{-1/2}), \tag{A.121}$$

for $i = 1, \ldots, p+1$ and where

$$A_0 = M^+_{0\vartheta}, \qquad E(A_1) = E(\sigma_0^{-2} D_\vartheta L^*(\vartheta_0)) = 0,$$
$$B_0 = A, \qquad B_1 = M^+_{\vartheta,\vartheta' T} + M^+_{0,\vartheta\vartheta' T},$$

Here, $M^+_{0\vartheta}$, $M^+_{0,\vartheta\vartheta' T}$ and $M^+_{\vartheta,\vartheta' T}$ are given in (A.51), (A.54) *and* (A.55)*, respectively, and A is the inverse of the variance-covariance matrix given in* (67)*. The expression for C_{0i}, $i = 1, \ldots, p+1$, is given in* (A.82) *and* (A.83)*.*

Proof of Lemma A.25. The proof is omitted and follows from Lemma A.23 and the asymptotic behaviour of the modification term and its derivatives in Lemma A.21. □

A.4 Proof of the main results

In this section, we provide the proofs for the main results presented in Section 3.

A.4.1 Proof of Theorem 3.1

The proof for the CSS score follows directly from Lemmata A.13, A.14 and A.20. The proof for the CSS score with known μ_0 follows directly from Lemma A.2 and from $E(S_t^+ S_{\vartheta_k t}^+) = 0$.

A.4.2 Proof of Lemma 3.1

The first property is readily established since

$$\sum_{t=1}^{T} c_t^2(\vartheta) \geq c_1^2 = \phi_0(\varphi)\kappa_{01}(d) = 1,$$

from $\kappa_{01}(d) = 1$ and $\phi_0(\varphi) = 1$.

In a special case where the short-run dynamics $\varphi = 0$, we have $c_t(\vartheta) = \kappa_{0t}(d)$. Then

$$\sum_{t=1}^{T} \kappa_{0t}^2 \leq \kappa_{01}^2 + \kappa_{02}^2 = 1 + (1-d)^2,$$

because $\kappa_{01} = 1$ for all d and $\kappa_{02} = \pi_1(1-d) = 1-d$ for all d, see Johansen & Nielsen (2016, Lemma A.4). Thus $\kappa_{02} = 0$ only if $d = 1$ and from the recursive relationship $\pi_j(a) = \frac{j-1+a}{j}\pi_{j-1}(a)$ for $j \geq 1$ and for all a, see for instance p.96 in Hassler (2019), it follows that $\kappa_{0n} = 0$ for all $n \geq 2$ when $d = 1$. Thus $m(d, \varphi) = 1$ if $d = 1$ and $\varphi = 0$.

The proof of the second property is given in Lemmata A.15 and A.21.

A.4.3 Proof of Theorem 3.2

We note that the MCSS estimator is equal to

$$\begin{aligned}\hat{\vartheta}_m &= \operatorname*{argmin}_{\vartheta \in \Theta} L_m^*(\vartheta) \\ &= \operatorname*{argmin}_{\vartheta \in \Theta} \log\left(m(\vartheta)\frac{2}{T}L^*(\vartheta)\right),\end{aligned}$$

so that the objective function equals $\tilde{L}(\vartheta) = \log\left(m(\vartheta)\frac{2}{T}L^*(\vartheta)\right) = \log(m(\vartheta)) + \log\left(\frac{2}{T}L^*(\vartheta)\right)$. We also note that $R(\vartheta) = \frac{2}{T}L^*(\vartheta)$ is the same objective function as in Hualde & Nielsen

(2020). Fix $\epsilon > 0$ and let $M_\epsilon = \{\vartheta \in \Theta : |\vartheta - \vartheta_0| < \epsilon\}$ and $\bar{M}_\epsilon = \{\vartheta \in \Theta : |\vartheta - \vartheta_0| \geq \epsilon\}$. Then

$$\begin{aligned}
\Pr\left(\hat{\vartheta}_m \in \bar{M}\right) &= \Pr\left(\inf_{\vartheta \in \bar{M}_\epsilon} \tilde{L}(\vartheta) \leq \inf_{\vartheta \in M_\epsilon} \tilde{L}(\vartheta)\right), \\
&\leq \Pr\left(\inf_{\vartheta \in \bar{M}_\epsilon} \tilde{L}(\vartheta) \leq \tilde{L}(\vartheta_0)\right), \\
&\leq \Pr\left(\inf_{\vartheta \in \bar{M}_\epsilon} \log(R(\vartheta)) - \log(R(\vartheta_0)) \leq \log(m(\vartheta_0)) - \inf_{\vartheta \in \Theta} \log(m(\vartheta))\right),
\end{aligned}$$

From Hualde & Nielsen (2020), as $T \to \infty$, we have that

$$\Pr\left(\inf_{\vartheta \in \bar{M}_\epsilon} \log(R(\vartheta)) - \log(R(\vartheta_0)) \leq 0\right) \to 0.$$

So to prove consistency, it remains to show that

$$\log(m(\vartheta_0)) - \inf_{\vartheta \in \Theta} \log(m(\vartheta)) \to 0, \tag{A.122}$$

which is already established in Lemmata A.14 and A.20.

To show the asymptotic normality of $\hat{\vartheta}_m$, we proceed with a usual Taylor expansion of the score function,

$$0 = D_\vartheta L_m^*(\hat{\vartheta}_m) = D_\vartheta L_m^*(\vartheta_0) + \left(\hat{\vartheta}_m - \vartheta_0\right) D_{\vartheta\vartheta'} L_m^*(\vartheta^*),$$

where ϑ^* is an intermediate value satisfying $|\vartheta^* - \vartheta_0| \leq |\hat{\vartheta}_m - \vartheta_0| \xrightarrow{p} 0$. The product moments within $D_{\vartheta\vartheta'} L^*(\vartheta)$ have been demonstrated in Johansen & Nielsen (2010, Lemma C.4) and Johansen & Nielsen (2012, Lemma A.8(i)) to exhibit tightness or equicontinuity in a neighborhood of ϑ_0. This allows us to apply Johansen & Nielsen (2010, Lemma A.3) and conclude that $D_{\vartheta\vartheta'} L_m^*(\vartheta^*) = D_{\vartheta\vartheta'} L_m^*(\vartheta_0) + o_P(1)$. Consequently, we proceed to analyse $D_\vartheta L_m^*(\vartheta_0)$ and $D_{\vartheta\vartheta'} L_m^*(\vartheta_0)$. According to Lemmata A.19 and A.25, we find that $\sigma_0^2 T^{-1/2} D_\vartheta L_m^*(\vartheta_0) = M_{0\vartheta}^+ + O_P(T^{-1/2} \log(T))$ and $\sigma_0^2 T^{-1} D_{\vartheta\vartheta'} L_m^*(\vartheta_0) = A + O_P(T^{-1/2})$ so that the final result follows from Lemma A.9.

A.4.4 Proof of Theorem 3.3

First, we consider the bias of $\hat{\vartheta}$. A Taylor series expansion of $D_\vartheta L^*(\hat{\vartheta}) = 0$ around ϑ_0 gives

$$0 = D_\vartheta L^*(\hat{\vartheta}) = D_\vartheta L^*(\vartheta_0) + D_{\vartheta\vartheta'} L^*(\vartheta_0)(\hat{\vartheta} - \vartheta_0) + \frac{1}{2}\begin{bmatrix} (\hat{\vartheta} - \vartheta_0)' D_{\vartheta_1\vartheta\vartheta'} L(\vartheta^*)(\hat{\vartheta} - \vartheta_0) \\ \vdots \\ (\hat{\vartheta} - \vartheta_0)' D_{\vartheta_{p+1}\vartheta\vartheta'} L(\vartheta^*)(\hat{\vartheta} - \vartheta_0) \end{bmatrix},$$

where ϑ^* is an intermediate value which is allowed to vary across the different rows of $D_{\vartheta_i\vartheta\vartheta'} L(\vartheta^*)$ for $i = 1, \ldots, p+1$ and satisfies $|\vartheta^* - \vartheta_0| \leq |\hat{\vartheta} - \vartheta_0| \xrightarrow{p} 0$. We then insert $\hat{\vartheta} - \vartheta_0 = T^{-1/2} G_{1T} + T^{-1} G_{2T} + O_p(T^{-3/2})$ and find

$G_{1T} = -T^{1/2}(D_{\vartheta\vartheta'} L^*(\vartheta_0))^{-1} D_\vartheta L^*(\vartheta_0),$

$$G_{2T} = -\frac{1}{2}T(D_{\vartheta\vartheta'}L^*(\vartheta_0))^{-1}\begin{bmatrix} D_{\vartheta}L^*(\vartheta_0)'(D_{\vartheta\vartheta'}L^*(\vartheta_0))^{-1}D_{\vartheta_1\vartheta\vartheta'}L(\vartheta^*)(D_{\vartheta\vartheta'}L^*(\vartheta_0))^{-1}D_{\vartheta}L^*(\vartheta_0) \\ \vdots \\ D_{\vartheta}L^*(\vartheta_0)'(D_{\vartheta\vartheta'}L^*(\vartheta_0))^{-1}D_{\vartheta_{p+1}\vartheta\vartheta'}L(\vartheta^*)(D_{\vartheta\vartheta'}L^*(\vartheta_0))^{-1}D_{\vartheta}L^*(\vartheta_0) \end{bmatrix},$$

which we write as

$$\begin{aligned} T^{1/2}(\hat{\vartheta}-\vartheta_0) = &-(T^{-1}D_{\vartheta\vartheta'}L^*(\vartheta_0))^{-1}T^{-1/2}D_{\vartheta}L^*(\vartheta_0) - \frac{1}{2}T^{-1/2}(T^{-1}D_{\vartheta\vartheta'}L^*(\vartheta_0))^{-1} \\ &\begin{bmatrix} T^{-1/2}D_{\vartheta}L^*(\vartheta_0)'(T^{-1}D_{\vartheta\vartheta'}L^*(\vartheta_0))^{-1}T^{-1}D_{\vartheta_1\vartheta\vartheta'}L(\vartheta^*)(T^{-1}D_{\vartheta\vartheta'}L^*(\vartheta_0))^{-1}T^{-1/2}D_{\vartheta}L^*(\vartheta_0) \\ \vdots \\ T^{-1/2}D_{\vartheta}L^*(\vartheta_0)'(T^{-1}D_{\vartheta\vartheta'}L^*(\vartheta_0))^{-1}T^{-1}D_{\vartheta_{p+1}\vartheta\vartheta'}L(\vartheta^*)(T^{-1}D_{\vartheta\vartheta'}L^*(\vartheta_0))^{-1}T^{-1/2}D_{\vartheta}L^*(\vartheta_0) \end{bmatrix} \\ &+ o_P(T^{-1/2}). \end{aligned} \tag{A.123}$$

First we note that, as in Appendix A.4.3, we can apply Johansen & Nielsen (2010, Lemma A.3) to conclude that $D_{\vartheta_i\vartheta\vartheta'}L(\vartheta^*) = D_{\vartheta_i\vartheta\vartheta'}L(\vartheta_0) + o_P(1)$ for $i = 1, \ldots, p+1$. Consequently, we plug in the derivatives in Lemma A.17 and A.23 into the expansion (A.123) and find

$$\begin{aligned} T^{1/2}(\hat{\vartheta}-\vartheta_0) = &-(B_0 + T^{-1/2}B_1)^{-1}\left(A_0 + T^{-1/2}A_1\right) - \frac{1}{2}T^{-1/2}(B_0 + T^{-1/2}B_1)^{-1} \\ &\begin{bmatrix} \left(A_0 + T^{-1/2}A_1\right)'(B_0 + T^{-1/2}B_1)^{-1}C_{01}(B_0 + T^{-1/2}B_1)^{-1}\left(A_0 + T^{-1/2}A_1\right) \\ \vdots \\ \left(A_0 + T^{-1/2}A_1\right)'(B_0 + T^{-1/2}B_1)^{-1}C_{0(p+1))}(B_0 + T^{-1/2}B_1)^{-1}\left(A_0 + T^{-1/2}A_1\right) \end{bmatrix} \\ &+ o_P(T^{-1/2}). \end{aligned} \tag{A.124}$$

Using the Woodbury matrix identity

$$\begin{aligned} (B_0 + T^{-1/2}B_1)^{-1} &= B_0^{-1} - T^{-1/2}B_0^{-1}(I + T^{-1/2}B_1B_0^{-1})^{-1}B_1B_0^{-1} \\ &= B_0^{-1} - T^{-1/2}B_0^{-1}B_1B_0^{-1} + O_P(T^{-1}), \end{aligned}$$

and hence (A.124) reduces to

$$\begin{aligned} T^{1/2}(\hat{\vartheta}-\vartheta_0) = -B_0^{-1}A_0 - T^{-1/2}&\left(B_0^{-1}A_1 - B_0^{-1}B_1B_0^{-1}A_0 + \frac{1}{2}B_0^{-1}\begin{bmatrix} A_0'B_0^{-1}C_{0,1}B_0^{-1}A_0 \\ \vdots \\ A_0'B_0^{-1}C_{0,p+1}B_0^{-1}A_0 \end{bmatrix}\right) \\ &+ o_P(T^{-1/2}). \end{aligned}$$

We find that $E(A_0) = E(M_{0\vartheta}^+) = 0$ so that

$$\begin{aligned} TE(\hat{\vartheta}-\vartheta_0) = -&\left(B_0^{-1}E(A_1) - B_0^{-1}E(B_1B_0^{-1}A_0) + \frac{1}{2}B_0^{-1}\begin{bmatrix} E(A_0'B_0^{-1}C_{0,1}B_0^{-1}A_0) \\ \vdots \\ E(A_0'B_0^{-1}C_{0,p+1}B_0^{-1}A_0) \end{bmatrix}\right) \\ &+ o(1). \end{aligned} \tag{A.125}$$

We rewrite

$$E(A_0'B_0^{-1}C_{0,i}B_0^{-1}A_0) = \iota'\left(\left(B_0^{-1}C_{0,i}B_0^{-1}\right) \odot E(A_0A_0')\right)\iota,$$

and from Lemma A.9 we have that

$$E(A_0A_0') = E\left(M_{0,\vartheta T}^{+}(M_{0,\vartheta T}^{+})'\right) = A + o(1)$$

We also rewrite

$$\begin{aligned}
E\left(B_1B_0^{-1}A_0\right) &= E\left(\left(M_{\vartheta,\vartheta' T}^{+} + M_{0,\vartheta\vartheta' T}^{+}\right)A^{-1}M_{0,\vartheta T}^{+}\right) \\
&= E\left(M_{\vartheta,\vartheta' T}^{+}A^{-1}M_{0,\vartheta T}^{+}\right) + E\left(M_{0,\vartheta\vartheta' T}^{+}A^{-1}M_{0,\vartheta T}^{+}\right) \\
&= \begin{bmatrix} \iota'\left(A^{-1}\odot E\left(M_{\vartheta_1,\vartheta T}^{+}\left(M_{0,\vartheta T}^{+}\right)'\right)\right)\iota \\ \vdots \\ \iota'\left(A^{-1}\odot E\left(M_{\vartheta_{p+1},\vartheta T}^{+}\left(M_{0,\vartheta T}^{+}\right)'\right)\right)\iota \end{bmatrix} \\
&\quad + \begin{bmatrix} \iota'\left(A^{-1}\odot E\left(M_{0,\vartheta_1\vartheta T}^{+}\left(M_{0,\vartheta T}^{+}\right)'\right)\right)\iota \\ \vdots \\ \iota'\left(A^{-1}\odot E\left(M_{0,\vartheta_{p+1}\vartheta T}^{+}\left(M_{0,\vartheta T}^{+}\right)'\right)\right)\iota \end{bmatrix}.
\end{aligned}$$

From Lemma A.10 we have that

$$E\left(M_{0,\vartheta_k\vartheta T}^{+}\left(M_{0,\vartheta T}^{+}\right)'\right) = F_k + o(1),$$

for $k = 1,\dots,p+1$, with

$$F_1 = \begin{pmatrix} -2\zeta_3 & \sum_{i=0}^{\infty} D_{dd}\pi_i(0)b_{\varphi' i}(\varphi_0) \\ \sum_{i=2}^{\infty} i^{-1}h_{d\varphi i}(\varphi_0) & -\sum_{i=2}^{\infty} h_{d\varphi i}(\varphi_0)b_{\varphi' i}(\varphi_0) \end{pmatrix}, \tag{A.126}$$

and for $m = 1,\dots,p$ it follows that

$$F_{m+1} = \begin{pmatrix} \sum_{i=2}^{\infty} i^{-1}h_{d\varphi_m i}(\varphi_0) & -\sum_{i=2}^{\infty} h_{d\varphi_m i}(\varphi_0)b_{\varphi' i}(\varphi_0) \\ -\sum_{i=1}^{\infty} i^{-1}b_{\varphi\varphi_m i}(\varphi_0) & \sum_{i=1}^{\infty} b_{\varphi\varphi_m i}(\varphi_0)b_{\varphi' i}(\varphi_0) \end{pmatrix}, \tag{A.127}$$

From Lemma A.11 we have that

$$E\left(M_{\vartheta_k,\vartheta T}^{+}\left(M_{0,\vartheta T}^{+}\right)'\right) = G_k + o(1)$$

for $k = 1,\dots,p+1$, with

$$G_1 = \begin{pmatrix} G_1(1,1) & G_1(1,2) \\ G_1(2,1) & G_1(2,2) \end{pmatrix}, \tag{A.128}$$

where the elements are given by

$$\begin{aligned}
G_1(1,1) &= -4\zeta_3, \\
G_1(1,2) &= 2\sum_{k=1}^{\infty} b_{\varphi' k}(\varphi_0)\sum_{s=1}^{\infty} s^{-1}(s+k)^{-1}, \\
G_1(2,1) &= \sum_{k=1}^{\infty} k^{-1}\sum_{s=1}^{\infty}\left(s^{-1}b_{\varphi(s+k)}(\varphi_0) + (s+k)^{-1}b_{\varphi s}(\varphi_0)\right),
\end{aligned}$$

$$G_1(2,2) = -\sum_{k=1}^{\infty}\sum_{s=1}^{\infty}\left(s^{-1}b_{\varphi(s+k)}(\varphi_0) + (s+k)^{-1}b_{\varphi s}(\varphi_0)\right)b_{\varphi' k}(\varphi_0),$$

and for $m = 1, \ldots, p$ it follows that

$$G_{m+1} = \begin{pmatrix} G_{m+1}(1,1) & G_{m+1}(1,2) \\ G_{m+1}(2,1) & G_{m+1}(2,2) \end{pmatrix} \tag{A.129}$$

where the elements are given by

$$\begin{aligned}
G_{m+1}(1,1) &= \sum_{k=1}^{\infty} k^{-1}\sum_{s=1}^{\infty}\left(s^{-1}b_{\varphi_m(s+k)}(\varphi_0) + (s+k)^{-1}b_{\varphi_m s}(\varphi_0)\right), \\
G_{m+1}(1,2) &= -\sum_{k=1}^{\infty} b_{\varphi' k}(\varphi_0)\sum_{s=1}^{\infty}\left(s^{-1}b_{\varphi_m(s+k)}(\varphi_0) + (s+k)^{-1}b_{\varphi_m s}(\varphi_0)\right), \\
G_{m+1}(2,1) &= -\sum_{k=1}^{\infty} k^{-1}\sum_{s=1}^{\infty}\left(b_{\varphi_m s}(\varphi_0)b_{\varphi(s+k)}(\varphi_0) + b_{\varphi_m(s+k)}(\varphi_0)b_{\varphi s}(\varphi_0)\right), \\
G_{m+1}(2,2) &= \sum_{k=1}^{\infty}\sum_{s=1}^{\infty}\left(b_{\varphi_m s}(\varphi_0)b_{\varphi(s+k)}(\varphi_0) + b_{\varphi_m(s+k)}(\varphi_0)b_{\varphi s}(\varphi_0)\right)b_{\varphi' k}(\varphi_0),
\end{aligned}$$

Then

$$\begin{aligned}
TE(\hat{\vartheta} - \vartheta_0) = {} & -A^{-1}E(\sigma_0^{-2}D_\vartheta L^*(\vartheta_0)) + A^{-1}\begin{bmatrix} \iota'\left(A^{-1}\odot G_1\right)\iota \\ \vdots \\ \iota'\left(A^{-1}\odot G_{p+1}\right)\iota \end{bmatrix} + A^{-1}\begin{bmatrix} \iota'\left(A^{-1}\odot F_1\right)\iota \\ \vdots \\ \iota'\left(A^{-1}\odot F_{p+1}\right)\iota \end{bmatrix} \\
& -\frac{1}{2}A^{-1}\begin{bmatrix} \iota'\left((A^{-1}C_{0,1}A^{-1})\odot A\right)\iota \\ \vdots \\ \iota'\left((A^{-1}C_{0,p+1}A^{-1})\odot A\right)\iota, \end{bmatrix} + o(1)
\end{aligned}$$

Hence we can write

$$E\left(\hat{\vartheta} - \vartheta_0\right) = S(d_0, \varphi_0) + B(\varphi_0) + o(T^{-1}), \tag{A.130}$$

where

$$\begin{aligned}
TS(d_0, \varphi_0) &= -A^{-1}\left[\sigma_0^{-2}E\left(D_\vartheta L^*(\vartheta_0)\right)\right] \\
&= A^{-1}\frac{\sum_{t=1}^{T} c_t(\vartheta_0)c_{\vartheta t}(\vartheta_0)}{\sum_{t=1}^{T} c_t^2(\vartheta_0)},
\end{aligned}$$

from Lemma A.13 and

$$TB(\varphi_0) = A^{-1}\begin{bmatrix} \iota'\left(A^{-1}\odot(G_1+F_1)\right)\iota \\ \vdots \\ \iota'\left(A^{-1}\odot(G_{p+1}+F_{p+1})\right)\iota \end{bmatrix} - \frac{1}{2}A^{-1}\begin{bmatrix} \iota'\left((A^{-1}C_{0,1}A^{-1})\odot A_T\right)\iota \\ \vdots \\ \iota'\left((A^{-1}C_{0,p+1}A^{-1})\odot A_T\right)\iota \end{bmatrix}. \tag{A.131}$$

For $d_0 > 1/2$ it follows from Lemma A.14 that

$$TS(d_0, \varphi_0) = A^{-1}\frac{\sum_{t=1}^{\infty} c_t(\vartheta)c_{\vartheta t}(\vartheta)}{\sum_{t=1}^{\infty} c_t^2(\vartheta)},$$

and for $d_0 < 1/2$ it follows from Lemma A.20 that

$$TS(d_0, \varphi_0) = A^{-1} \begin{bmatrix} -\log(T) + (\Psi(1-d_0) + (1-2d_0)^{-1}) \\ \frac{D_{\varphi_1}\phi(1;\varphi)}{\phi(1;\varphi)} \\ \vdots \\ \frac{D_{\varphi_p}\phi(1;\varphi)}{\phi(1;\varphi)} \end{bmatrix},$$

completing the proof for the bias of $\hat{\vartheta}$. It follows from Lemmata A.18, A.19, A.24 and A.25 that analogues of (A.130) also hold for $\hat{\vartheta}_{\mu_0}$ and $\hat{\vartheta}_m$. Replacing $D_\vartheta L^*(\vartheta_0)$ by $D_\vartheta L^*_{\mu_0}(\vartheta_0)$, it is clear that the expected score term $S(d_0, \varphi_0)$ gets eliminated. Similarly, $E(D_\vartheta L^*_m(\vartheta_0)) = 0$ by construction, and the proof is completed.

Observe that we additionally write $S_T(d_0, \varphi_0)$ and $B_T(\varphi_0)$, where we use the exact expectations of the expressions in (A.125) —referred to as $F_{T,i}$, $G_{T,i}$, A_T and $C_{T,0i}$—instead of the asymptotic expectations of F_i, G_i, A and C_{0i}. Then, the "exact" bias is given by

$$E\left(\hat{\vartheta} - \vartheta_0\right) = S_T(d_0, \varphi_0) + B_T(\varphi_0) + o(T^{-1}),$$

where

$$\begin{aligned} TS_T(d_0, \varphi_0) &= -A^{-1}\left[\sigma_0^{-2} E\left(D_\vartheta L^*(\vartheta_0)\right)\right] \\ &= A^{-1}\frac{\sum_{t=1}^T c_t(\vartheta_0) c_{\vartheta t}(\vartheta_0)}{\sum_{t=1}^T c_t^2(\vartheta_0)}, \end{aligned}$$

and

$$TB_T(\varphi_0) = A^{-1}\begin{bmatrix} \iota'\left(A^{-1} \odot (G_{T,1} + F_{T,1})\right)\iota \\ \vdots \\ \iota'\left(A^{-1} \odot (G_{T,p+1} + F_{T,p+1})\right)\iota \end{bmatrix} - \frac{1}{2}A^{-1}\begin{bmatrix} \iota'\left((A^{-1}C_{T,0,1}A^{-1}) \odot A_T\right)\iota \\ \vdots \\ \iota'\left((A^{-1}C_{T,0,p+1}A^{-1}) \odot A_T\right)\iota \end{bmatrix}.$$

With $F_{T,k} = E\left(M^+_{0,\vartheta_k\vartheta T}\left(M^+_{0,\vartheta T}\right)'\right)$ such that

$$F_{T,1} = \begin{pmatrix} -T^{-1}\sum_{t=1}^T\sum_{i=0}^{t-1} D_{dd}\pi_i(0)D_d\pi_i(0) & T^{-1}\sum_{t=1}^T\sum_{i=0}^{t-1} D_{dd}\pi_i(0)b_{\varphi' i}(\varphi_0) \\ T^{-1}\sum_{t=1}^T\sum_{i=2}^{t-1} i^{-1}h_{d\varphi i}(\varphi_0) & -T^{-1}\sum_{t=1}^T\sum_{i=2}^{t-1} h_{d\varphi i}(\varphi_0)b_{\varphi' i}(\varphi_0) \end{pmatrix},$$

and for $m = 1, \ldots, p$ we have

$$F_{T,m+1} = \begin{pmatrix} T^{-1}\sum_{t=1}^T\sum_{i=2}^{t-1} i^{-1}h_{d\varphi_m i}(\varphi_0) & -T^{-1}\sum_{t=1}^T\sum_{i=2}^{t-1} h_{d\varphi_m i}(\varphi_0)b_{\varphi' i}(\varphi_0) \\ -T^{-1}\sum_{t=1}^T\sum_{i=1}^{t-1} i^{-1}b_{\varphi\varphi_m i}(\varphi_0) & T^{-1}\sum_{t=1}^T\sum_{i=1}^{t-1} b_{\varphi\varphi_m i}(\varphi_0)b_{\varphi' i}(\varphi_0) \end{pmatrix},$$

With $G_{T,k} = E\left(M^+_{\vartheta_k,\vartheta T}\left(M^+_{0,\vartheta T}\right)'\right)$ such that

$$G_{T,1} = \begin{pmatrix} G_{T,1}(1,1) & G_{T,1}(1,2) \\ G_{T,1}(2,1) & G_{T,1}(2,2) \end{pmatrix},$$

where the elements are given by

$$G_{T,1}(1,1) = -2T^{-1}\sum_{t=1}^T\sum_{k=1}^{t-1}\sum_{s=t+1}^T k^{-1}(s-t)^{-1}(s-t+k)^{-1},$$

$$G_{T,1}(1,2) = 2T^{-1}\sum_{t=1}^{T}\sum_{k=1}^{t-1}\sum_{s=t+1}^{T} b_{\varphi' k}(\varphi_0)(s-t)^{-1}(s-t+k)^{-1},$$
$$G_{T,1}(2,1) = T^{-1}\sum_{t=1}^{T}\sum_{k=1}^{t-1}\sum_{s=t+1}^{T} k^{-1}\left((s-t)^{-1}b_{\varphi(s-t+k)}(\varphi_0) + (s-t+k)^{-1}b_{\varphi(s-t)}(\varphi_0)\right),$$
$$G_{T,1}(2,2) = -T^{-1}\sum_{t=1}^{T}\sum_{k=1}^{t-1}\sum_{s=t+1}^{T}\left((s-t)^{-1}b_{\varphi(s-t+k)}(\varphi_0) + (s-t+k)^{-1}b_{\varphi(s-t)}(\varphi_0)\right) b_{\varphi' k}(\varphi_0),$$

and for $m = 1,\dots,p$ we have

$$G_{m+1,T} = \begin{pmatrix} G_{T,m+1}(1,1) & G_{T,m+1}(1,2) \\ G_{T,m+1,T}(2,1) & G_{T,m+1}(2,2) \end{pmatrix},$$

where the elements are given by

$$G_{T,m+1}(1,1) = T^{-1}\sum_{t=1}^{T}\sum_{k=1}^{t-1} k^{-1}\sum_{s=t+1}^{T}\left((s-t)^{-1}b_{\varphi_m(s-t+k)}(\varphi_0) + (s-t+k)^{-1}b_{\varphi_m(s-t)}(\varphi_0)\right),$$
$$G_{T,m+1}(1,2) = -T^{-1}\sum_{t=1}^{T}\sum_{k=1}^{t-1} b_{\varphi' k}(\varphi_0)\sum_{s=t+1}^{T}\left((s-t)^{-1}b_{\varphi_m(s-t+k)}(\varphi_0) + (s-t+k)^{-1}b_{\varphi_m(s-t)}(\varphi_0)\right),$$
$$G_{T,m+1}(2,1) = -T^{-1}\sum_{t=1}^{T}\sum_{k=1}^{t-1} k^{-1}\sum_{s=t+1}^{T}\left(b_{\varphi_m(s-t)}(\varphi_0)b_{\varphi(s-t+k)}(\varphi_0) + b_{\varphi_m(s-t+k)}(\varphi_0)b_{\varphi(s-t)}(\varphi_0)\right),$$
$$G_{T,m+1}(2,2) = T^{-1}\sum_{t=1}^{T}\sum_{k=1}^{t-1}\sum_{s=t+1}^{T}\left(b_{\varphi_m(s-t)}(\varphi_0)b_{\varphi(s-t+k)}(\varphi_0) + b_{\varphi_m(s-t+k)}(\varphi_0)b_{\varphi(s-t)}(\varphi_0)\right) b_{\varphi' k}(\varphi_0),$$

With $A_T = E\left(M^+_{0,\vartheta T}(M^+_{0,\vartheta T})'\right)$ such that

$$A_T = \begin{pmatrix} T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{t-1}\frac{1}{j^2} & -T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{t-1} b_{\varphi' j}(\varphi_0)/j \\ -T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{t-1} b_{\varphi j}(\varphi_0)/j & T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{t-1} b_{\varphi j}(\varphi_0)b_{\varphi' j}(\varphi_0). \end{pmatrix}$$

Lastly, $C_{T,0i}$ is defined as follows

$$C_{T,01} = \begin{pmatrix} C_{T,01}(1,1) & C_{T,01}(1,2) \\ C_{T,01}(2,1) & C_{T,01}(2,2) \end{pmatrix},$$

where the elements are given by

$$C_{T,01}(1,1) = -3T^{-1}\sum_{t=1}^{T}\sum_{i=0}^{t-1} D_{dd}\pi_i(0)D_d\pi_i(0),$$
$$C_{T,01}(1,2) = 2T^{-1}\sum_{t=1}^{T}\sum_{i=2}^{t-1} i^{-1}h_{d\varphi' i}(\varphi_0) + T^{-1}\sum_{t=1}^{T}\sum_{i=0}^{t-1} D_{dd}\pi_i(0)b_{\varphi' i}(\varphi_0),$$
$$C_{T,01}(2,1) = 2T^{-1}\sum_{t=1}^{T}\sum_{i=2}^{t-1} i^{-1}h_{d\varphi i}(\varphi_0) + T^{-1}\sum_{t=1}^{T}\sum_{i=0}^{t-1} D_{dd}\pi_i(0)b_{\varphi i}(\varphi_0),$$
$$C_{T,01}(2,2) = -T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{t-1} i^{-1}b_{\varphi\varphi' i}(\varphi_0) - T^{-1}\sum_{t=1}^{T}\sum_{i=2}^{t-1} b_{\vartheta i}(\varphi_0)h_{d\vartheta' i}(\varphi_0)$$

$$-\left(T^{-1}\sum_{t=1}^{T}\sum_{i=2}^{t-1}b_{\vartheta i}(\varphi_0)h_{d\vartheta' i}(\varphi_0)\right)',$$

and for $k=1,\ldots,p$ we have that

$$C_{T,0(k+1)}=\begin{pmatrix}C_{T,0(k+1)}(1,1) & C_{T,0(k+1)}(1,2)\\ C_{T,0(k+1)}(2,1) & C_{T,0(k+1)}(2,2),\end{pmatrix}$$

where the elements are given by

$$\begin{aligned}
C_{T,0(k+1)}(1,1) &= 2T^{-1}\sum_{t=1}^{T}\sum_{i=2}^{t-1}i^{-1}h_{d\varphi_k i}(\varphi_0)+T^{-1}\sum_{t=1}^{T}\sum_{i=0}^{t-1}D_{dd}\pi_i(0)b_{\varphi_k i}(\varphi_0),\\
C_{T,0(k+1)}(1,2) &= -T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{t-1}i^{-1}b_{\varphi'\varphi_k i}(\varphi_0)-T^{-1}\sum_{t=1}^{T}\sum_{i=2}^{t-1}b_{\varphi_k i}(\varphi_0)h_{d\varphi' i}(\varphi_0)\\
&\quad -T^{-1}\sum_{t=1}^{T}\sum_{i=2}^{t-1}b_{\varphi' i}(\varphi_0)h_{d\varphi_k i}(\varphi_0),\\
C_{T,0(k+1)}(2,1) &= -T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{t-1}i^{-1}b_{\varphi\varphi_k i}(\varphi_0)-T^{-1}\sum_{t=1}^{T}\sum_{i=2}^{t-1}b_{\varphi_k i}(\varphi_0)h_{d\varphi i}(\varphi_0)\\
&\quad -T^{-1}\sum_{t=1}^{T}\sum_{i=2}^{t-1}b_{\varphi i}(\varphi_0)h_{d\varphi_k i}(\varphi_0),\\
C_{T,0(k+1)}(2,2) &= \left(T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{t-1}b_{\varphi i}(\varphi_0)b_{\varphi'\varphi_k i}(\varphi_0)\right)'+T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{t-1}b_{\varphi i}(\varphi_0)b_{\varphi'\varphi_k i}(\varphi_0)\\
&\quad +T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{t-1}b_{\varphi_k i}(\varphi_0)b_{\varphi\varphi' i}(\varphi_0).
\end{aligned}$$

A.4.5 Proof of Corollary 3.4

The lag polynomial for AR(1) specification is given by

$$\omega(L;\varphi)=(1-\varphi L)^{-1}=\sum_{j=0}^{\infty}\omega_j(\varphi)L^j,$$

where $\omega_j(\varphi)=\varphi^j$, $j\geq 0$. The inverse lag polynomial is given by

$$\phi(L;\varphi)=\omega^{-1}(L;\varphi)=(1-\varphi L)=\sum_{j=0}^{\infty}\phi_j(\varphi)L^j,$$

where $\phi_0(\varphi)=1$ and $\phi_1(\varphi)=-\varphi$, $\phi_s(\varphi)=0$, $s\geq 2$. Taking the derivatives of $\phi_s(\varphi)$ with respect to φ yields: $\partial\phi_0(\varphi)/\partial\varphi=0$, $\partial\phi_1(\varphi)/\partial\varphi=-1$, and $\partial\phi_s(\varphi)/\partial\varphi=0$, $s\geq 2$. Then

$$\begin{aligned}
b_{\varphi j}(\varphi_0) &= \sum_{k=0}^{j-1}\omega_k(\varphi_0)\partial\phi_{j-k}(\varphi_0)/\partial\varphi\\
&= -\varphi^{j-1},
\end{aligned}$$

for $j \geq 1$. Since $\partial^2\phi_j(\varphi)/\partial\varphi^2 = 0$ for $j \geq 1$ we get

$$\begin{aligned} b_{\varphi\varphi j}(\varphi_0) &= \sum_{k=0}^{j-1} \omega_k(\varphi_0)\partial^2\phi_{j-k}(\varphi_0)/\partial\varphi^2, \\ &= 0 \end{aligned}$$

To find an expression for the bias we need the following expansion

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}, \tag{A.132}$$

$$\sum_{k=1}^{\infty} k^{-1}x^k = -\log(1-x), \tag{A.133}$$

$$\sum_{k=1}^{\infty}(k+1)^{-1}x^{k+1}\sum_{n=1}^{k} n^{-1} = \frac{1}{2}\log^2(1-x), \tag{A.134}$$

$$\sum_{k=1}^{\infty} x^k \sum_{s=1}^{\infty} s^{-1}(s+k)^{-1} = -Li_2(-\frac{x}{1-x}), \tag{A.135}$$

where $|x| < 1$. The first three expansions are well known and can be found in Gradshteyn & Ryzhik (2014) on pages 7 (0.231-1), 44 (1.513-4), and 45 (1.516-1) respectively. The last expansion makes use of a couple of results. First, we have the expression:

$$\begin{aligned} \sum_{s=1}^{\infty} s^{-1}(s+k)^{-1} &= k^{-1}\left(\Psi(k+1)+\gamma\right), \\ \Psi(k+1)+\gamma &= \int_0^1 (1-t)^{-1}(1-t^k)dt, \end{aligned}$$

These results can be found in Abramowitz & Stegun (1964) on page 259, 6.3.16 and 6.3.22 respectively. Using these results and (A.133), we get

$$\begin{aligned} \sum_{k=1}^{\infty} x^k \sum_{s=1}^{\infty} s^{-1}(s+k)^{-1} &= \sum_{k=1}^{\infty} x^k k^{-1}\int_0^1 (1-t)^{-1}(1-t^k)dt \\ &= \int_0^1 (1-t)^{-1}\sum_{k=1}^{\infty} x^k k^{-1}(1-t^k)dt \\ &= \int_0^1 (1-t)^{-1}\log\left(\frac{1-xt}{1-x}\right)dt, \end{aligned}$$

Then a change of variable of integration to get

$$\begin{aligned} \int_0^1 (1-t)^{-1}\log\left(\frac{1-xt}{1-x}\right)dt &= -\int_1^{(1-x)^{-1}} (1-u)^{-1}\log(u)du \\ &= -Li_2(-\frac{x}{1-x}), \end{aligned}$$

where $Li_2(\varphi) = \sum_{i=1}^{\infty} i^{-2}\varphi^i$, or alternatively $Li_2(1-v) = \int_1^v (1-t)^{-1}\log(t)dt$, is the dilogarithm function (Spence's integral), see Abramowitz & Stegun (1964, page 1004, 27.7.1) for the integral representation where a slight different definition of the dilogarithm function is used, namely $f(x) = Li_2(1-x)$.

Next, we find the expression for A in (67), and its inverse A^{-1} by using (A.132) and (A.133):

$$A = \begin{pmatrix} \pi^2/6 & -\varphi^{-1}\log(1-\varphi) \\ -\varphi^{-1}\log(1-\varphi) & (1-\varphi^2)^{-1} \end{pmatrix},$$

and

$$A^{-1} = \frac{\varphi}{\pi^2\varphi^2 - 6(1-\varphi^2)\log^2(1-\varphi)} \begin{pmatrix} 6\varphi & 6\log(1-\varphi)(1-\varphi^2) \\ 6\log(1-\varphi)(1-\varphi^2) & \pi^2\varphi(1-\varphi^2) \end{pmatrix}.$$

Next, we find the expression for C_{01} and C_{02} in respectively, (A.82) and (A.83). Using (A.133) and (A.134) we find

$$C_{01} = \begin{pmatrix} -6\zeta_3 & 2\varphi^{-1}Li_2(-\frac{\varphi}{1-\varphi}) - \varphi^{-1}\log^2(1-\varphi) \\ 2\varphi^{-1}Li_2(-\frac{\varphi}{1-\varphi}) - \varphi^{-1}\log^2(1-\varphi) & 2\frac{\log(1-\varphi)}{1-\varphi^2} \end{pmatrix},$$

and

$$C_{02} = \begin{pmatrix} 2\varphi^{-1}Li_2(-\frac{\varphi}{1-\varphi}) - \varphi^{-1}\log^2(1-\varphi) & 2\frac{\log(1-\varphi)}{1-\varphi^2} \\ 2\frac{\log(1-\varphi)}{1-\varphi^2} & 0 \end{pmatrix}.$$

Next, we find expression for F_1 and F_2 in respectively, (A.126) and (A.127). Using (A.133) and (A.134) we find

$$F_1 = \begin{pmatrix} -2\zeta_3 & -\varphi^{-1}\log^2(1-\varphi) \\ \varphi^{-1}Li_2(-\frac{\varphi}{1-\varphi}) & \frac{\log(1-\varphi)}{1-\varphi^2} \end{pmatrix},$$

and

$$F_2 = \begin{pmatrix} \varphi^{-1}Li_2(-\frac{\varphi}{1-\varphi}) & \frac{\log(1-\varphi)}{1-\varphi^2} \\ 0 & 0 \end{pmatrix}.$$

Next, we find expression for G_1 and G_2 in respectively, (A.128) and (A.129). Using (A.132),(A.133) and (A.135) we find

$$G_1 = \begin{pmatrix} -4\zeta_3 & 2\varphi^{-1}Li_2(-\frac{\varphi}{1-\varphi}) \\ -\varphi^{-1}\log^2(1-\varphi) + \varphi^{-1}Li_2(-\frac{\varphi}{1-\varphi}) & \frac{\log(1-\varphi)}{1-\varphi^2} - \varphi^{-2}\left(\frac{\varphi}{1-\varphi} + \log(1-\varphi)\right) \end{pmatrix},$$

and

$$G_2 = \begin{pmatrix} -\varphi^{-1}\log^2(1-\varphi) + \varphi^{-1}Li_2(-\frac{\varphi}{1-\varphi}) & \frac{\log(1-\varphi)}{1-\varphi^2} - \varphi^{-2}\left(\frac{\varphi}{1-\varphi} + \log(1-\varphi)\right) \\ 2\log(1-\varphi)\frac{1}{1-\varphi^2} & -2\frac{\varphi}{(1-\varphi^2)^2} \end{pmatrix}.$$

Next, we find an expression for the score bias term. First we consider the non-stationary region, i.e. $d_0 > 1/2$. We have that

$$c_t(\vartheta) = \sum_{j=0}^{t-1} \phi_j(\varphi)\kappa_{0(t-j)}(d)$$

$$= \kappa_{0t}(d) - \varphi\kappa_{0(t-1)}(d)I(t \geq 2),$$

and, therefore,

$$\begin{aligned}
\sum_{t=1}^{\infty} c_t^2(d,\varphi) &= (1+\varphi^2)\sum_{t=1}^{\infty}\kappa_{0t}^2(d) - 2\varphi\sum_{t=1}^{\infty}\kappa_{0t}(d)\kappa_{0(t+1)}(d) \\
&= (1+\varphi^2)\sum_{t=1}^{\infty}\kappa_{0t}^2(d) + 2\varphi\sum_{t=1}^{\infty}\Big(\kappa_{0(t+1)}(d) - \kappa_{0t}(d)\Big)\kappa_{0(t+1)}(d) - 2\varphi\sum_{t=1}^{\infty}\kappa_{0(t+1)}^2(d) \\
&= (1+\varphi^2-2\varphi)\sum_{t=1}^{\infty}\kappa_{0t}^2(d) + 2\varphi\sum_{t=1}^{\infty}\kappa_{0t}(d)\kappa_{0t}(1+d), \qquad \text{(A.136)}
\end{aligned}$$

where we used the properties $\pi_0(u) = 1$ and $\pi_t(u) - \pi_{t-1}(u) = \pi_t(u-1)$ for any u, see Johansen & Nielsen (2016, Lemma A.4), and approximation by Stirling's Formula impling that $\kappa_{0t}(d) = O(t^{-d})$. The first summand in this expression is given in Johansen & Nielsen (2016, Lemma B.1), i.e. ,

$$\sum_{t=1}^{\infty}\kappa_{0t}^2(d) = \binom{2d-2}{d-1},$$

where $d > 1/2$. The second term can be derived using a similar proof strategy. We have that

$$\begin{aligned}
\sum_{t=1}^{\infty}\kappa_{0t}(d)\kappa_{0t}(1+d) &= \frac{1}{\Gamma(1-d)\Gamma(-d)}\sum_{t=0}^{\infty}\frac{\Gamma(1-d+t)\Gamma(-d+t)}{\Gamma(t)t!}, \\
&= 0.5\binom{2d}{d},
\end{aligned}$$

where the last equality follows from Abramowitz & Stegun (1964, p. 556, eqn. 15.1.20)). We conclude that

$$\sum_{t=1}^{\infty} c_t^2 = (1-\varphi)^2\binom{2d-2}{d-1} + \varphi\binom{2d}{d}.$$

Next, we find an expression for $\sum_{t=1}^{\infty} c_t(d,\varphi)D_d c_t(d,\varphi)$. Taking the derivative of the left and right-hand side of (A.136) with respect to d gives

$$2\sum_{t=1}^{\infty} c_t(d,\varphi)D_d c_t(d,\varphi) = 2(1-\varphi)^2\sum_{t=1}^{\infty}\kappa_{1t}(d)\kappa_{0t}(d) + 2\varphi\sum_{t=1}^{\infty}\left(\kappa_{1t}(d)\kappa_{0t}(1+d) + \kappa_{0t}(d)\kappa_{1t}(1+d)\right).$$

The first summand in this expression is given in Johansen & Nielsen (2016, Lemma B.1) (there is a small typo; the minus sign should be a plus sign), i.e. ,

$$\sum_{t=1}^{\infty}\kappa_{0t}(d)\kappa_{1t}(d) = \binom{2d-2}{d-1}\left(\Psi(2d-1) - \Psi(d)\right),$$

where $d > 1/2$. The second term can be obtained using a similar approach as in Johansen & Nielsen (2016, Lemma B.1), and is given by

$$D_d\sum_{t=1}^{\infty}\kappa_{0t}(d)\kappa_{0t}(1+d) = \binom{2d}{d}\left(\Psi(2d+1) - \Psi(d+1)\right).$$

We conclude that

$$\sum_{t=1}^{\infty} c_t(d,\varphi) D_d c_t(d,\varphi) = (1-\varphi)^2 \binom{2d-2}{d-1} \left(\Psi(2d-1) - \Psi(d)\right) + \varphi \binom{2d}{d} \left(\Psi(2d+1) - \Psi(d+1)\right).$$

Next, we find an expression for $\sum_{t=1}^{\infty} c_t(d,\varphi) D_\varphi c_t(d,\varphi)$. Taking the derivative of the left-hand side and right-hand side of (A.136) with respect to φ gives

$$2\sum_{t=1}^{\infty} c_t(d,\varphi) D_\varphi c_t(d,\varphi) = (2\varphi - 2)\sum_{t=1}^{\infty} \kappa_{0t}^2(d) + 2\sum_{t=1}^{\infty} \kappa_{0t}(d)\kappa_{0t}(1+d).$$

Note that the expressions for the two summands are given above. We conclude that

$$\sum_{t=1}^{\infty} c_t(d,\varphi) D_\varphi c_t(d,\varphi) = (\varphi - 1)\binom{2d-2}{d-1} + 0.5\binom{2d}{d}.$$

Therefore the score bias $S(d_0,\varphi_0)$ for $d_0 > 1/2$ is given by

$$TS(d,\varphi) = A^{-1}\left[(1-\varphi)^2\binom{2d-2}{d-1} + \varphi\binom{2d}{d}\right]^{-1} \times \left[\begin{matrix} (1-\varphi)^2\binom{2d-2}{d-1}\left(\Psi(2d-1)-\Psi(d)\right) + \varphi\binom{2d}{d}\left(\Psi(2d+1)-\Psi(d+1)\right) \\ (\varphi-1)\binom{2d-2}{d-1} + 0.5\binom{2d}{d} \end{matrix}\right],$$

The score bias $S(d_0,\varphi_0)$ for $d_0 < 1/2$ is given by

$$TS(d,\varphi) = A^{-1}\left[\begin{matrix} -\log(T) + \Psi(1-d_0) + (1-2d_0)^{-1} \\ -\frac{1}{1-\varphi} \end{matrix}\right],$$

because $\phi(1;\varphi) = 1 - \varphi$ and the proof is complete.