Finite sample behaviour of the modified conditional sum of squares estimator for fractionally integrated models

Mustafa Kilinc[1] and Michael Massmann[1]

[1]Department of Economics, WHU – Otto Beisheim School of Management, Vallendar, Germany

21st February, 2022

Abstract

In this paper, we remove the influence of estimating a constant term on the bias of the conditional maximum likelihood or conditional sum-of-squares estimator of the fractional parameter, d, in a stationary/invertible or nonstationary or non-invertible time series model. We consider a "type II" ARFIMA(0,d,0) process including a constant term and derive an expression for the second-order bias of $\hat{d}$. We show that we can remove the second-order bias in $\hat{d}$ that occurs due to the presence of a constant term by a simple modification on the conditional (profiled) likelihood. Consequently, in finite samples the estimated $\hat{d}$ behaves, on average, the same as if we have known the true value of the constant term. We call this estimator the modified conditional sum of squares estimator (MCSS).

1 Introduction

Fractional integrated time series models are applied widely in a wide range of fields, for example in finance, economics, political science and many more; see for illustrations Hassler (2019) and Hualde and Nielsen (2021a). In a parametric setting, one popular choice of estimating such models is the quasi-maximum likelihood estimator (QML), also called the conditional sum-of-squares estimator (CSS) or the truncated sum-of-squares estimator. This estimator has been widely applied in the empirical literature modeling fractional integrated time series, see for example Hualde and Robinson (2011) for aggregate income and consumption data and Johansen and Nielsen (2016)—henceforth JN (2016)—for opinion polls. The main appealing feature of this estimator is that the so-called memory parameter can be estimated consistently, as long as it lies in a compact large interval, on the real line allowing for nonstationary and noninvertible processes. Li and McLeod (1986) introduced this estimator for stationary ARFIMA models. Beran (1995) allowed for nonstationarity and gave a consistency proof, although Velasco and Robinson (2000) noticed that the argument for consistency was circular. Robinson (2006) fixed the consistency proof, albeit for the stationary region. Tanaka (1999) and Nielsen (2004) gave local consistency proofs. Finally, Hualde and Robinson (2011) and Nielsen (2015) provided global consistency proofs and derived the asymptotic behaviour of the estimators under quite general assumptions of the error component. The consistency proof is non-trivial because the objective function convergence non-uniformly when the range of the memory parameter is arbitrarily large. A drawback of the previously mentioned papers is that the asymptotic justification for the CSS estimator is only given for time series without deterministic components. Only recently, Hualde and Nielsen (2020, 2021b) introduced unknown deterministic components (such as a constant term and also additive generalized polynomial trend) and establish consistency and asymptotic normality of the estimator for the memory parameter.

In the presence of deterministic components, however, the CSS estimators might be severely deteriorated. This was already early noticed by Chung and Baillie (1993). Chung and Baillie (1993) showed, through a simulation study, that the estimation of the mean can make considerable difference to the small sample bias of memory parameter and other parameters in the error component. They evaluate three different estimators of the mean, and find that in some cases the sample median may perform better than the more usual sample mean or the QML of the mean. This bias issue was also documented by Nielsen and Frederiksen (2005), for example, for the ARFIMA(1,d,0) model with a positive autoregressive parameter.

We propose a simple way of handling the finite-sample bias issue that occurs due to estimating the deterministic component by directly modifying the (conditional) likelihood, based on the idea of Cox and Reid (1987). In an unpublished paper An and Bloomfield (1993)[1] applied this idea to remove the second-order bias from the maximum likelihood estimator due to the presence of unknown nuisance parameters in a stationary "type I" ARFIMA model. Ooms and Doornik (1999) and Doornik and Ooms (2004) showed the effectiveness of the bias correction for the estimation of the memory parameter. The literature does not yet contain, however, a theoretical justification of the modified likelihood approach in the context of fractionally integrated models. We fill this gap in the literature for a "type II" ARFIMA(0,d,0) model. We find almost the same modification term as in An and Bloomfield (1993). Despite the similarity of the modification term, in our setting we do not assume that the data is stationary nor invertible. Furthermore, our modification term is easy to calculate without the need for inverting the variance-covariance matrix. Finally, we do not impose that the data generation process is Gaussian such that any non-Gaussian processes with finite fourth moments is allowed.

This paper contributes to the literature in three ways. Firstly, we show analytically that the modified conditional sum-of-squares estimator (MCSS) removes the bias in the estimated memory parameter due to the presence of the unknown constant term. Secondly, we show that the asymptotically properties of the MCSS estimator are not affected by the correction term and therefore are asymptotically negligible and hence

[1]We thank Shu An and Peter Bloomfield for providing us with their paper.

behave the same as the CSS estimator. Thirdly, we find an analytical expression for the bias of the memory parameter that occurs due to the presence of the constant term. For non-stationary fractional time series Johansen and Nielsen (2016) derived an expression of the second-order bias using higher-order expansions. We extend their result by covering also the stationary domain. The remainder of the paper is organized as follows. In Section 2, we present the MCSS estimator and in Section 3 we present the main result. Section 4 present a simulation study. Section 5 concludes. Proofs of the main result are given in the appendix.

2 The Modified Conditional Sum of Squares Estimator

We consider the following simple fractional process with an unknown constant term given by

$$x_t = \mu_0 I(t \geq 1) + \Delta_0^{-d_0}\epsilon_t, \quad \epsilon_t \sim i.i.d.(0, \sigma_0^2), \quad t = 0, \pm 1, \pm 2, \ldots, \tag{1}$$

where Δ = 1 - L and L are the difference and lag operators, respectively, and where d_0 can take any value in $\mathbb{R}$. For any series u_t, real number ζ and $t \geq 1$, the operator Δ_0^{ζ} is defined by

$$\Delta_0^{\zeta} u_t = \Delta^{\zeta}\{u_t I(t \geq 1)\} = \sum_{i=0}^{t-1} \pi_i(-\zeta) u_{t-i}, \tag{2}$$

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

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

denoting the coefficients in the usual binomial expansion of $(1-z)^{-a} = \sum_{i=0}^{\infty} \pi_i(a) z^i$, where $\Gamma(\cdot)$ is the gamma function with the convention $\Gamma(i) = 0$ for $i =, -1, -2, \ldots$ and $\frac{\Gamma(0)}{\Gamma(0)} = 1$. Model (1) is known as a "type II"(or truncated) fractional model, since the operator in (2) implies that $x_t = 0$ for $t \leq 0$, see Marinucci and Robinson (1999).

We would like to consider the Gaussian log-likelihood, conditional[2] on $x_t = 0$ for $t \leq 0$, which is given by

$$-\frac{T}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(\Delta^d(x_t - \mu)\right)^2 = -\frac{T}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(\Delta_0^d(x_t - \mu)\right)^2. \tag{4}$$

It is clear from (4) that we can profile out σ^2 and find the maximum likelihood estimators of d and μ by minimizing

$$L(d, \mu) = \frac{1}{2}\sum_{t=1}^{T}\left(\Delta_0^d(x_t - \mu)\right)^2, \tag{5}$$

[2]Notice that actually the conditioning has been done implicitly in (1) by assuming a "type II" fractional process. Therefore, Hualde and Robinson (2011) and Hualde and Nielsen (2020) prefer to call the estimators that maximizes (4) the truncated sum-of-squares estimators instead of the conditional sum-of-squares estimators.

with respect to d and μ. Furthermore, we can profile out μ by rewriting the term

$$\begin{aligned}\Delta_0^d(x_t-\mu) &= \Delta_0^d x_t - \sum_{n=0}^{t-1}\pi_n(-d)\mu,\\ &= \Delta_0^d x_t - \pi_{t-1}(1-d)\mu = \Delta_0^d x_t - \kappa_{0t}(d)\mu,\end{aligned}$$

where $\kappa_{0t}(d) = \sum_{n=0}^{t-1}\pi_n(-d) = \pi_{t-1}(1-d)$ by JN (2016, Lemma A.4). We find that the estimator of μ for fixed d is given by

$$\hat{\mu}(d) = \frac{\sum_{t=1}^{T}(\Delta_0^d x_t)\kappa_{0t}(d)}{\sum_{t=1}^{T}\kappa_{0t}(d)^2}. \tag{6}$$

Then the maximum likelihood estimator of d can be found by minimizing the profiled likelihood function

$$L^*(d) = \frac{1}{2}\sum_{t=1}^{T}\left(\Delta_0^d x_t - \kappa_{0t}(d)\hat{\mu}(d)\right)^2 \tag{7}$$

with respect to d. Thus,

$$\hat{d} = \operatorname*{argmin}_{d\in\mathbb{D}} L^*(d), \tag{8}$$

for a parameter space $\mathbb{D}$ to be defined below. Formally (7) is a monotonic transformation of the likelihood function, but we prefer to call it the likelihood function and to make a distinction we place an asterisk.

Hualde and Nielsen (2020, Theorem 1 and Theorem 2)[3] showed that if x_t is generated by (1) and under Assumptions 3.1-3.2 (defined below), then, as $T\to\infty$, it holds that $\hat{d}$ is consistent and that

$$\sqrt{T}(\hat{d}-d_0) \xrightarrow{d} N(0,\zeta_2^{-1}), \tag{9}$$

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

A few remarks about the estimator $\hat{\mu}(d)$ in (6). For $d = d_0$ we have that

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

which has mean zero and variance $\sigma_0^2\left(\sum_{t=1}^{T}\kappa_{0t}(d)^2\right)^{-1}$. For the stationary region, e.g $d_0 < 1/2$, this variance goes to zero because then $\sum_{t=1}^{T}\kappa_{0t}(d)^2$ converges in T, see Lemma A.7. While for the non-stationary region, e.g. $d_0 > 1/2$, this variance does not go to zero because then $\sum_{t=1}^{T}\kappa_{0t}(d)^2$ is bounded in T, see Lemma A.2. This is exactly the reason that $\hat{\mu}(\hat{d})$ is consistent only when $d_0 < 1/2$, for the proof see Hualde and Nielsen (2020, Corollary 1).

[3] Hualde and Nielsen (2020) allow for short memory components in the DGP and an additive generalized polynomial trend with unknown exponent parameter. Therefore, our DGP is a special case of theirs.

The purpose of this paper is to remove the second-order bias that occurs due to estimating μ_0 and to find the analytic expression of this second-order bias. For this purpose, we analyze also the situation where μ_0 is known. In practice, this is not a feasible estimators since μ_0 is often not known. This CSS estimator can be formulated in (5) where $\mu = \mu_0$ and we get

$$\hat{d}_{\mu_0} = \underset{d \in \mathbb{D}}{\operatorname{argmin}} L^*_{\mu_0}(d) \tag{10}$$

$$L^*_{\mu_0}(d) = \frac{1}{2} \sum_{t=1}^{T} \left(\Delta_0^d (x_t - \mu_0) \right)^2 . \tag{11}$$

This estimator is considered by Hualde and Robinson (2011) and it is shown that $\hat{d}_{\mu_0}$ is consistent with the same limiting distribution as in (9).

The difference between the likelihood in (11) compared to the likelihood in (7) is that μ_0 is replaced by the maximum likelihood estimator of μ keeping d fixed. Apart from the uncertainty of the replacement of σ^2 with the estimated, it is important to realize that $L^*(d)$ is not a genuine likelihood function, that is, it is not based on the density function of the random variables in (1), see for example Severini (2000). In large samples, this replacement has a relatively minor effect. In small samples, however, this replacement may have a large impact. This was already early noticed by Chung and Baillie (1993). As a consequence, the bias of $\hat{d}$ due to the presence of an unknown constant term is reflected in the expectation of the score function evaluated at d_0. It turns out that

$$E \left(DL^*(d_0) \right) = O(1),$$

where D stands for the first derivative, see JN (2016, Lemma B.4) and Lemma A.11. Hence, the poor performance of the profile likelihood is reflected in the expected score function evaluated at the true value of d. Whereas, if the μ_0 were known we would have that

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

by JN (2016, Lemma B.4) and Lemma A.12. The reason is that in this case the likelihood $L^*_{\mu_0}(d)$ is genuine, in the sense that it is based on the density function of the random variables in (1).

Higher-order asymptotic theory can provide corrections for the bias in the estimation of d caused by estimating the constant. We remove the second-order bias by modifying the the likelihood function as follows

$$L^*_{MCSS}(d) = m(d) L^*(d), \tag{12}$$

$$m(d) = \left(\sum_{t=1}^{T} \kappa_{0t}(d)^2 \right)^{\frac{1}{T-1}}, \tag{13}$$

such that

$$E \left(DL^*_{MCSS}(d_0) \right) = o(1)$$

from Lemma A.6 and A.12. We call this estimator the modified conditional sum-of-squares (MCSS) estimator and denote the estimated d by $\hat{d}_{MCSS}$, i.e,

$$\hat{d}_{MCSS} = \operatorname*{argmin}_{d \in \mathbb{D}} L^*_{MCSS}(d). \tag{14}$$

The asymptotic properties of the MCSS estimator are not affected by the correction term and therefore asymptotically negligible and hence the asymptotic distribution is given in (9). This follows from noticing that

$$\begin{aligned} L^*_{MCSS}(d_0) &= L^*(d_0) + O_P(1) \qquad\quad \text{for } d_0 > 1/2, \\ L^*_{MCSS}(d_0) &= L^*(d_0) + O_P(\log(T)) \text{ for } d_0 < 1/2, \end{aligned}$$

where the first term is $O_P(T)$, the next is $O_P(1)$ or $O_P(\log(T))$, so the second term has no influence on the asymptotic distribution of $\hat{d}_{MCSS}$, see Lemma A.3 and A.2. However, for the bias we need to analyze the second term further. Note that the MCSS is simple to implement and does not require a priori knowledge about d_0. The modification term $m(d)$ that we have is almost the same as the modification term derived by An and Bloomfield (1993). Who apply the idea in Cox and Reid (1987) to remove the bias of the maximum likelihood estimator due to nuisance parameters of the regression. The modification term that they find is

$$\iota' R^{-1} \iota \tag{15}$$

where ι is a $T \times 1$ vector of ones and R variance-covariance matrix of $x \sim N(\mu, R\sigma)$. We might want to write $R^{-1} = (R^{-1/2})' R^{-1/2}$ and then the expression in (15) becomes $(R^{-1/2}\iota)'\ (R^{-1/2}\iota)$, where $R^{-1/2}$ basically is filtering out the correlation structure in the error term. In our setting we have a truncated "type II" process such that $R^{-1/2}$ is replaced by Δ_0^d and we get $(\Delta_0^d \iota)'(\Delta_0^d \iota) = \sum_{t=1}^T \kappa_{0t}(d)^2$. Although the similarities between the two modifications, our setting does not require stationarity of x_t and neither we do not need to calculate the inverse of the variance-covariance matrix also we do not impose Gaussianity of x_t. The simplicity and the feasibility of the CSS estimator is also encountered in the modified CSS estimator. Doornik and Ooms (2004) showed the effectiveness of the bias correction for the estimation of d in ARFIMA models and we shall see that this effectiveness is carried out for the MCSS estimator as well.

Before we end this Section, we want to build some intuition of the modification term in (12). Therefore, Figure 1 plots $m(d)$ against different values of d, namely d between -1 and 2, and for $T = 64$, 128, 256.

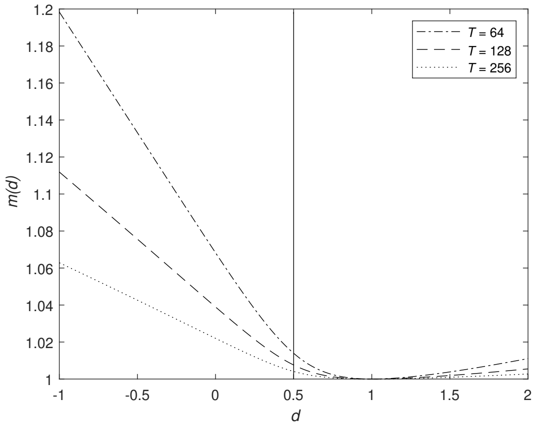

Figure 1: Plot of the modification term $m(d)$ in (12) for d between -1 and 2, and $T =$ 64, 128, 256.

A few remarks focusing only on the region $-1 \leq d \leq 2$. First, the modification term punishes the likelihood more severe for stationary time series, that is when $d < 1/2$, than for non-stationary time series, when $d > 1/2$. Therefore, one would except that the bias of $\hat{d}$ in the stationary region is more severe than the non stationary region and indeed that follows from Theorem 3.2. Second, around $d = 1$ the curve seems flat, implying that the bias caused by estimating the constant term is the lowest at that point. Third, even for a moderate sample of $T = 256$ the stationary part seems to have still a considerable punishment term.

3 Stochastic Expansions

Before we analyze the bias terms of the different estimators that we considered in Section 2 we first introduce two assumptions and then a theorem about the consistency and asymptotic normality of the MCCS estimator in (14).

Assumption 3.1. *The errors ϵ_t are i.i.d.$(0,\sigma_0^2)$ with finite fourth moment.*

Assumption 3.2. *The parameter set for (d, μ, σ^2) is $\mathbb{D} \times \mathbb{R} \times \mathbb{R}_+$, where $\mathbb{D} = [\underline{d}, \bar{d}]$, $0 < \underline{d} < \bar{d} < \infty$. The true value is d_0 in the interior of $\mathbb{D}$.*

Assumption 3.1 requires that the errors are i.i.d. with finite four moments. This assumption is stricter than in Hualde and Robinson (2011) and Hualde and Nielsen (2020) were they assume martingale difference errors. The literature for modified profiled likelihood, however, often assumes that the error term is known, and often assumed to be Gaussian, see An and Bloomfield (1993),Cox and Reid (1987) and Doornik and Ooms (2004). Therefore, compared to this branch of literature our Assumption 3.1 is mild. Assumption 3.2 allows d_0 to take any value inside a compact set $\mathbb{D}$. This assumption is the same as in Hualde and Robinson (2011) and Hualde and Nielsen (2020). The modified profiled likelihood of An and Bloomfield (1993), however, requires that $-1/2 < d_0 < 1/2$. In our case, we do not have any restrictions of d_0, therefore, we allow for non-invertible and non-stationary processes.

Before we move to our main result, we first show that our MCSS estimator is consistent and asymptotically normal in the follow theorem.

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

$$\hat{d}_{MCSS} \xrightarrow{p} d_0, \tag{16}$$

and

$$\sqrt{T}(\hat{d}_{MCSS} - d_0) \xrightarrow{d} N(0, \zeta_2^{-1}), \tag{17}$$

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

Theorem 3.1 shows that the MCSS estimator is consistent and asymptotically normal following the same distribution as the CSS estimator. We consider next the effects of the modification term in the likelihood function for the asymptotic bias of the MCSS estimator.

To analyze the asymptotic bias of the CSS estimators and MCSS estimator for d, we need to examine higher-order terms in a stochastic expansion of the estimators. JN (2016) use this technique to find the second-order bias of the CSS estimators in the nonstationary domain. We follow their approach closely. First, we apply a Taylor series expansion of $DL^*(\hat{d}) = 0$ around d_0 which gives

$$0 = DL^*(\hat{d}) = DL^*(d_0) + (\hat{d} - d_0)D^2L^*(d_0) + \frac{1}{2}(\hat{d} - d_0)^2 D^3 L^*(d^*) \tag{18}$$

where d^* is an intermediate value satisfying $|d^* - d_0| \le |\hat{d} - d_0| \xrightarrow{p} 0$. Applying Johansen and Nielsen (2010, Lemma A.3) allows us to replace $D^3L^*(d^*)$ by $D^3L^*(d_0)$. We then insert $\hat{d} - d_0 = T^{-1/2}\tilde{G}_{1T} + T^{-1}\tilde{G}_{2T} + O_P(T^{-3/2})$ into (18) and find

$$\begin{aligned}\tilde{G}_{1T} &= -T^{1/2}\frac{DL^*(d_0)}{D^2L^*(d_0)},\\ \tilde{G}_{2T} &= -\frac{1}{2}T\frac{\left(DL^*(d_0)\right)^2 D^3L^*(d_0)}{\left(D^2L^*(d_0)\right)^3}.\end{aligned}$$

Then the expansion of the bias becomes

$$T^{1/2}\left(\hat{d} - d_0\right) = -T^{1/2}\frac{DL^*(d_0)}{D^2L^*(d_0)} - \frac{1}{2}T^{-1/2}\frac{\left(DL^*(d_0)\right)^2 D^3L^*(d_0)}{\left(D^2L^*(d_0)\right)^3} + O_P(T^{-1}) \tag{19}$$

which depends on the asymptotic behavior of the derivatives $D^mL^*(d_0)$, see Lemma A.6 for the nonstationary region and Lemmas A.11, A.12 and A.13 for the stationary region.

The main result of this paper is summarized in the follow theorem. Results (20) and (21) are derived in JN (2016, Theorem 4) and mentioned here for completeness.

Theorem 3.2. *Let x_t, $t = 1,\ldots,T$, be given by (1) and let Assumption 3.1 and 3.2 be satisfied. For the nonstationary region, $d_0 > 1/2$, the biases of $\hat{d}$, $\hat{d}_{\mu_0}$ and $\hat{d}_{MCSS}$ are*

$$bias(\hat{d}) = -(T\zeta_2)^{-1}\left[3\zeta_3\zeta_2^{-1} + (\Psi(d_0) - \Psi(2d_0 - 1))\right] + o(T^{-1}), \tag{20}$$

$$bias(\hat{d}_{\mu_0}) = -(T\zeta_2)^{-1}\left[3\zeta_3\zeta_2^{-1}\right] + o(T^{-1}), \tag{21}$$

$$bias(\hat{d}_{MCSS}) = -(T\zeta_2)^{-1}\left[3\zeta_3\zeta_2^{-1}\right] + o(T^{-1}), \tag{22}$$

and for the stationary region, $d_0 < 1/2$, the biases of $\hat{d}$, $\hat{d}_{\mu_0}$ and $\hat{d}_{MCSS}$ are

$$\begin{aligned} bias(\hat{d}) = -(T\zeta_2)^{-1}&\left[3\zeta_3\zeta_2^{-1} - (\Psi(1 - d_0) + (1 - 2d_0)^{-1}) + \log(T)\right] \\ &+ o(T^{-1}\log(T)), \end{aligned} \tag{23}$$

$$bias(\hat{d}_{\mu_0}) = -(T\zeta_2)^{-1}\left[3\zeta_3\zeta_2^{-1}\right] + o(T^{-1}), \tag{24}$$

$$bias(\hat{d}_{MCSS}) = -(T\zeta_2)^{-1}\left[3\zeta_3\zeta_2^{-1}\right] + o(T^{-1}\log(T)). \tag{25}$$

where ζ_s is the Riemann's zeta function, $\zeta_s = \sum_{j=1}^{\infty} j^{-s}$, $s > 1$, and specially $\zeta_2 = \frac{\pi^2}{6} \approx 1.6449$ and $\zeta_3 \approx 1.2021$ and $\Psi(d)$ denotes the Digamma function.

It follows from Theorem 3.2 that the MCSS estimator indeed removes the bias that arises from estimating the constant parameter, irrespective whether the process is stationary or nonstationary. The remaining (fixed) 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 is derived by Lieberman and Phillips (2004) for the estimated memory parameter, based on the maximum likelihood estimator in the stationary case, $0 < d_0 < 1/2$, for a "type I" ARFIMA(0,d,0) process with μ known.

Some numerical comparison of the biases are of interest. Table 1 presents the theoretical biases of the CSS estimator with unknown and known μ_0 and the MCSS estimator for selected values of d_0 and T. We find that the bias($\hat{d}$) decreases in d and T for the stationary and non-stationary region separately. In the stationary region the bias of $\hat{d}$ is often more severe than in the non-stationary region. This finding coincides with the findings of Robinson and Velasco (2015). They consider a panel setting with individual components and fractionally integrated errors and estimate their model with the CSS estimator. However, the fixed effects causes a serious bias, and especially in the stationary region, as we can also observe from Theorem 3.2 and Table 1. A possible solution for their problem might be to use the MCSS estimator instead of the CSS estimator.

$d_0\backslash T$	bias($\hat{d}$)	bias($\hat{d}_{\mu_0}$)	bias($\hat{d}_{MCSS}$)	bias($\hat{d}$)	bias($\hat{d}_{\mu_0}$)	bias($\hat{d}_{MCSS}$)	bias($\hat{d}$)	bias($\hat{d}_{\mu_0}$)	bias($\hat{d}_{MCSS}$)
		64			128			256	
-0.2	-0.056	-0.021	-0.021	-0.031	-0.010	-0.010	-0.017	-0.005	-0.005
-0.1	-0.056	-0.021	-0.021	-0.031	-0.010	-0.010	-0.017	-0.005	-0.005
0.0	-0.056	-0.021	-0.021	-0.031	-0.010	-0.010	-0.017	-0.005	-0.005
0.1	-0.056	-0.021	-0.021	-0.031	-0.010	-0.010	-0.017	-0.005	-0.005
0.2	-0.054	-0.021	-0.021	-0.030	-0.010	-0.010	-0.017	-0.005	-0.005
0.3	-0.048	-0.021	-0.021	-0.027	-0.010	-0.010	-0.015	-0.005	-0.005
0.4	-0.028	-0.021	-0.021	-0.017	-0.010	-0.010	-0.010	-0.005	-0.005
0.5	-	-	-	-	-	-	-	-	-
0.6	-0.056	-0.021	-0.021	-0.028	-0.010	-0.010	-0.014	-0.005	-0.005
0.7	-0.034	-0.021	-0.021	-0.017	-0.010	-0.010	-0.008	-0.005	-0.005
0.8	-0.026	-0.021	-0.021	-0.013	-0.010	-0.010	-0.007	-0.005	-0.005
0.9	-0.023	-0.021	-0.021	-0.011	-0.010	-0.010	-0.006	-0.005	-0.005
1.0	-0.021	-0.021	-0.021	-0.010	-0.010	-0.010	-0.005	-0.005	-0.005
1.1	-0.020	-0.021	-0.021	-0.010	-0.010	-0.010	-0.005	-0.005	-0.005

Table 1: Theoretical bias of the estimated long memory parameter for ARFIMA(0,d_0,0) of CSS estimator with unknown and known μ_0 and the MCSS estimator.

From the results of Theorem 3.2 we might also introduce another estimator which we call the bias-corrected MCSS (BC-MCSS) estimator of d_0

$$\hat{d}_{BC-MCSS} = \hat{d}_{MCSS} + T^{-1}3\zeta_3(\zeta_2)^{-2}. \tag{26}$$

Note that the second-order bias of $\hat{d}_{BC-MCSS}$ is completely eliminated and that the correction term is pivotal.

4 Simulation Results

In this section we report the results of the simulation study of the small sample properties of the CSS estimator with known and unknown μ_0, see (8) and (10), respectively, and the modified version thereof (14) together with the bias-corrected version in (26). We take as our DGP the model in (1) where ϵ_t is $N(0,1)$ distributed. We take without loss of generality $\mu_0 = 0$, since the estimators are invariant against the value μ_0. In all settings covered by our experiment we generate the x_t for T = 64, 128, 256. We let the long memory parameter d_0 vary. In particular, we set d_0 = -0.2,-0.1,...,1.1. We computed the estimates using the optimizing interval $d \in [d_0 - 5, d_0 + 5]$ and for each we report the Monte Carlo bias. All results are based on 100000 replications.

Table 2 shows the Monte Carlo bias (scaled by 100) of $\hat{d}$, $\hat{d}_{\mu_0}$ and $\hat{d}_{MCCS}$ and the $\hat{d}_{BC-MCCS}$ estimator. The modified CSS estimator perform well and the Monte Carlo bias are in accordance with the theoretical counterparts in Theorem 3.2, see Table 1. As can be seen from the table, for the stationary region, $d_0 < 1/2$, the bias in the CSS estimator is more severe than for the nonstationary case, as expected from Theorem 3.2. However, for the MCSS estimator the bias in both the stationary and nonstationary are the same, namely $-(T\zeta_2)^{-1}(3\zeta_3\zeta_2^{-1})$. Our simulation results shows that we can remove the bias that occurs due to the estimation of the constant term by using the MCSS estimator and therefore the MCSS estimator outperforms the usual CSS estimator. Furthermore, notice that the BC-MCCS estimator in (26) performs the best since the second-order bias of $\hat{d}_{BC-MCSS}$ is completely eliminated

$d_0\backslash T$	bias($\hat{d}$) 64	bias($\hat{d}_{\mu_0}$) 64	bias($\hat{d}_{MCSS}$) 64	bias($\hat{d}_{BC-MCSS}$) 64	bias($\hat{d}$) 128	bias($\hat{d}_{\mu_0}$) 128	bias($\hat{d}_{MCSS}$) 128	bias($\hat{d}_{BC-MCSS}$) 128	bias($\hat{d}$) 256	bias($\hat{d}_{\mu_0}$) 256	bias($\hat{d}_{MCSS}$) 256	bias($\hat{d}_{BC-MCSS}$) 256
-0.2	−6.98	−1.70	−1.66	0.42	−3.60	−0.93	−0.89	0.15	−1.88	−0.48	−0.46	0.06
-0.1	−7.02	−1.77	−1.69	0.40	−3.64	−0.95	−0.92	0.12	−1.88	−0.48	−0.45	0.07
0.0	−7.09	−1.76	−1.75	0.33	−3.56	−0.88	−0.85	0.19	−1.86	−0.46	−0.44	0.08
0.1	−6.95	−1.71	−1.69	0.39	−3.59	−0.96	−0.91	0.13	−1.87	−0.49	−0.46	0.06
0.2	−6.81	−1.72	−1.74	0.34	−3.41	−0.86	−0.83	0.21	−1.79	−0.46	−0.43	0.09
0.3	−6.36	−1.65	−1.67	0.41	−3.26	−0.89	−0.88	0.17	−1.71	−0.46	−0.45	0.07
0.4	−5.92	−1.74	−1.81	0.27	−2.97	−0.89	−0.90	0.14	−1.55	−0.48	−0.47	0.05
0.5	−5.16	−1.73	−1.85	0.23	−2.54	−0.93	−0.93	0.11	−1.30	−0.48	−0.47	0.05
0.6	−4.30	−1.72	−1.89	0.19	−2.04	−0.90	−0.93	0.11	−1.02	−0.48	−0.48	0.04
0.7	−3.47	−1.77	−1.92	0.16	−1.56	−0.87	−0.90	0.14	−0.78	−0.46	−0.47	0.05
0.8	−2.69	−1.73	−1.83	0.25	−1.30	−0.94	−0.96	0.08	−0.66	−0.50	−0.50	0.02
0.9	−2.21	−1.73	−1.81	0.27	−1.08	−0.92	−0.94	0.11	−0.53	−0.47	−0.47	0.05
1.0	−1.85	−1.70	−1.75	0.33	−0.95	−0.92	−0.93	0.11	−0.51	−0.50	−0.51	0.01
1.1	−1.65	−1.71	−1.74	0.34	−0.86	−0.91	−0.92	0.12	−0.43	−0.46	−0.46	0.06

Table 2: 100 × Monte Carlo bias of the estimated long memory parameter for ARFIMA(0,d_0,0) of CSS estimator with unknown and known μ_0 and the MCSS estimator together with the BC-MCSS estimator.

Table 3 reports empirical coverage of 95 % confidence intervals for d_0 based on the asymptotic distribution of the estimators. The $\hat{d}_{\mu_0}$ estimator achieves the most accurate coverage, although the results are somewhat less accurate when d_0 is in the stationary region. The $\hat{d}_{MCSS}$ estimator also generally perform reasonably well, and improves for larger d_0, especially compared with intervals based on $\hat{d}$. Furthermore, we notice that the coverage rates of $\hat{d}_{MCSS}$ and $\hat{d}_{BF-MCSS}$ are close to each other.

$d_0\backslash T$	$\hat{d}$ 64	$\hat{d}_{\mu_0}$ 64	$\hat{d}_{MCSS}$ 64	$\hat{d}_{BF-MCSS}$ 64	$\hat{d}$ 128	$\hat{d}_{\mu_0}$ 128	$\hat{d}_{MCSS}$ 128	$\hat{d}_{BF-MCSS}$ 128	$\hat{d}$ 256	$\hat{d}_{\mu_0}$ 256	$\hat{d}_{MCSS}$ 256	$\hat{d}_{BC-MCSS}$ 256
-0.2	82.96	92.40	87.70	87.88	87.88	93.28	90.30	90.14	91.06	94.32	92.74	92.74
-0.1	82.60	92.46	87.68	87.86	87.40	93.08	90.28	90.00	90.76	94.18	92.28	92.46
0.0	82.44	91.74	86.76	86.96	87.48	93.20	90.44	90.52	89.60	93.30	92.00	92.26
0.1	82.66	92.38	87.80	87.64	87.44	93.40	90.48	90.80	90.90	93.88	92.34	92.30
0.2	81.94	91.86	87.90	88.22	86.84	92.78	90.04	90.08	91.02	94.40	92.68	92.92
0.3	82.10	91.74	87.08	87.46	88.02	93.70	91.22	91.60	90.66	93.86	92.28	92.54
0.4	82.98	91.62	88.38	88.54	87.28	92.64	90.20	90.60	91.66	94.16	93.22	93.30
0.5	84.24	91.98	89.58	90.16	89.06	93.24	92.08	92.20	91.86	94.58	93.52	93.40
0.6	86.26	92.52	90.66	91.10	89.80	93.30	92.18	92.24	92.08	93.86	93.42	93.64
0.7	87.30	92.04	90.48	90.86	90.78	92.90	92.04	92.44	93.12	93.94	93.92	93.94
0.8	89.18	92.18	91.12	91.80	92.50	93.74	93.28	93.64	93.28	93.80	93.58	93.62
0.9	90.52	91.94	91.72	91.62	91.68	92.70	92.20	92.74	93.58	93.84	93.84	93.60
1.0	91.06	91.82	91.80	92.18	92.32	93.02	92.56	92.92	94.12	94.30	94.26	94.50
1.1	91.58	92.00	91.84	92.42	92.74	92.92	92.82	93.08	94.26	94.12	94.30	94.58

Table 3: Empirical coverage of 95 % CI for the CSS estimator with unknown and known μ_0 and the MCSS estimator together with the BC-MCSS estimator.

5 Final Comments

The CSS estimator is a popular choice among practitioners for estimating stationary and nonstationary ARFIMA models due to its simplicity and effectiveness. More recently, the asymptotic justification of the CSS estimator for regression model including deterministic components is given by Hualde and Nielsen (2020, 2021b). However, introducing deterministic components might cause the CSS estimators, $\hat{d}$, of the fractional parameter and also other parameters in the error component to be (severely) biased. We show, analytically and through simulations, for a simple fractional model that we can remove the bias that occurs due to the estimation of the constant term by using the MCSS estimator. This modified CSS estimator is easy to compute and implement and, therefore, can be used by practitioners for computing more precise estimators. It seems possible to generalize our framework by including short memory complements as well as additional deterministic components, such as polynomials of higher orders. A formal justification will be the object of future research.

References

An, S., & Bloomfield, P. (1993). *Cox and reid's modification in regression models with correlated errors* (Tech. Rep.).

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 (Methodological)*, *57*(4), 659–672.

Chung, C.-F., & Baillie, R. T. (1993). Small sample bias in conditional sum-of-squares estimators of fractionally integrated arma models. *Empirical economics*, *18*(4), 791–806.

Cox, D. R., & Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society: Series B (Methodological)*, *49*(1), 1–18.

Doornik, J. A., & Ooms, M. (2004). Inference and forecasting for arfima models with an application to us and uk inflation. *Studies in Nonlinear Dynamics & Econometrics*, *8*(2).

Hassler, U. (2019). *Time series analysis with long memory in view.* John Wiley & Sons.

Hualde, J., & Nielsen, M. Ø. (2020). Truncated sum of squares estimation of fractional time series models with deterministic trends. *Econometric Theory*, *36*(4), 751–772.

Hualde, J., & Nielsen, M. Ø. (2021a). Fractional integration and cointegration.

Hualde, J., & Nielsen, M. Ø. (2021b). Truncated sum-of-squares estimation of fractional time series models with generalized power law trend.

Hualde, J., & Robinson, P. M. (2011). Gaussian pseudo-maximum likelihood estimation of fractional time series models. *The Annals of Statistics*, *39*(6), 3152–3181.

Johansen, S., & Nielsen, M. Ø. (2010). Likelihood inference for a nonstationary fractional autoregressive model. *Journal of Econometrics*, *158*(1), 51–66.

Johansen, S., & Nielsen, M. Ø. (2016). The role of initial values in conditional sum-of-squares estimation of nonstationary fractional time series models. *Econometric Theory*, *32*(5), 1095–1139.

Li, W. K., & McLeod, A. I. (1986). Fractional time series modelling. *Biometrika*, *73*(1), 217–221.

Lieberman, O., & Phillips, P. C. (2004). Expansions for the distribution of the maximum likelihood estimator of the fractional difference parameter. *Econometric Theory*, *20*(3), 464–484.

Marinucci, D., & Robinson, P. M. (1999). Alternative forms of fractional brownian motion. *Journal of statistical planning and inference*, *80*(1-2), 111–122.

Nielsen, M. Ø. (2004). Efficient likelihood inference in nonstationary univariate models. *Econometric Theory*, *20*(1), 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*, *36*(2), 154–188.

Nielsen, M. Ø., & Frederiksen, P. H. (2005). Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration. *Econometric Reviews*, *24*(4), 405–443.
Ooms, M., & Doornik, J. A. (1999). *Inference and forecasting for fractional autoregressive integrated moving average models: With an application to us and uk inflation* (Tech. Rep.). Citeseer.
Robinson, P. M. (2006). Conditional-sum-of-squares estimation of models for stationary time series with long memory. In *Time series and related topics* (pp. 130–137). Institute of Mathematical Statistics.
Robinson, P. M., & Velasco, C. (2015). Efficient inference on fractionally integrated panel data models with fixed effects. *Journal of Econometrics*, *185*(2), 435–452.
Severini, T. A. (2000). *Likelihood methods in statistics*. Oxford University Press.
Tanaka, K. (1999). The nonstationary fractional unit root. *Econometric theory*, *15*(4), 549–582.
Velasco, C., & Robinson, P. M. (2000). Whittle pseudo-maximum likelihood estimation for nonstationary time series. *Journal of the American Statistical Association*, *95*(452), 1229–1243.

A Appendix

We use the same notation as in JN (2016). For the sake of completeness some of the lemmas in JN (2016) are also reported in this appendix. We first analyze $\Delta_0^d(x_t - \mu)$ and introduce some notations. Clearly, inserting (1) to $\Delta_0^d(x_t - \mu)$ gives us

$$\Delta_0^d(x_t - \mu) = \Delta_0^{d-d_0}\epsilon_t - \kappa_{0t}(d)\,(\mu - \mu_0)\,.$$

Taking derivatives of $\Delta_0^d(x_t - \mu)$ with respect to d and evaluated at $d = d_0$ are given by

$$D^m\Delta_0^d(x_t - \mu) = S_{mt}^+ - \kappa_{mt}(d_0)\,(\mu - \mu_0)\,,$$

where

$$\kappa_{mt}(d) = (-1)^m D^m \pi_{t-1}(1-d),$$
$$S_{mt}^+ = (-1)^m \sum_{k=0}^{t-1} D^m \pi_k(0)\epsilon_{t-k}.$$

In Section A.1 we investigate the order of magnitude of functions involving the deterministic term $\kappa_{mt}(d)$ and the stochastic term S_{mt}^+ and product moments containing these. This is divided into two part. In Section A.1.1 we investigate the nonstationary region ($d > 1/2$) and in Section A.1.2 the stationary region ($d < 1/2$). The need for a separate analysis is because the order of magnitude differs depending on the region.

A.1 Lemmas

The next lemma is taken from JN (2016) and gives the asymptotic behavior of the centered product moments of the stochastic term. It is stated here because the asymptotic behavior is not depend on d_0.

Define the centered product moments of the stochastic terms

$$M_{mnT}^{+} = \sigma_0^{-2} T^{-1/2} \sum_{t=1}^{T} \left(S_{mt}^{+} S_{nt}^{+} - E\left(S_{mt}^{+} S_{nt}^{+} \right) \right) \tag{27}$$

Notice the asymptotic properties are invariant against d_0 and $E\left(S_{0t}^{+} S_{mt}^{+}\right) = 0$ for $m \geq 1$.

Lemma A.1. *Suppose Assumption 3.1 holds and let* $\zeta_2 = \sum_{j=1}^{\infty} j^{-2} = \pi^2/6 \approx 1.6449$ *and* $\zeta_3 = \sum_{j=1}^{\infty} j^{-3} = \pi/6 \approx 1.2021$ *Then, for* $T \to \infty$*, it holds that* $\{M_{mnT}^{+}\}$ *are asymptotically normal, for* $0 \leq m, n \leq 3$*, with mean zero and some variances covariances given below*

$$\begin{aligned} E\left(\left(M_{01T}^{+} \right)^2 \right) &= \zeta_2, \\ E\left(M_{01T}^{+} M_{02T}^{+} \right) &= -2\zeta_3, \\ E\left(M_{01T}^{+} M_{11T}^{+} \right) &= -4\zeta_3, \end{aligned}$$

Proof of Lemma A.1. See JN (2016, Lemma B.2 and Lemma B.3). □

In Section A.1.1 we apply the Lemmas A.1-A.5 and find asymptotic results for the derivatives of $L_{MCSS}^{*}(d_0)$. Similarly, in Section A.1.2 we apply Lemmas A.1, A.2-A.10 and find asymptotic results for the derivatives of $L^{*}(d_0)$, $L_{\mu_0}^{*}$ and L_{MCSS}^{*}.

A.1.1 Nonstationary region

In Lemmas A.2, A.4 and A.5 we investigate the order of magnitude of functions involving the deterministic term $\kappa_{mt}(d)$ and the stochastic term S_{mt}^{+} and product moments containing these. In Lemma A.3 we investigate the order of magnitude involving the modification term $m(d)$ and derivatives of these. These lemmas are then used together with Lemmas A.1 to find asymptotic results for the derivatives L_{MCSS}^{*}.

Lemma A.2. *Let* $d > 1/2$ *and* $m, n \geq 0$*, then we have that:*

$$\sum_{t=1}^{T} \kappa_{0t}(d)^2 \to \binom{2d-2}{d-1} \tag{28}$$

$$\sum_{t=1}^{T} \kappa_{0t}(d)\kappa_{1t}(d) \to -\binom{2d-2}{d-1} \left(\Psi(2d-1) - \Psi(d) \right) \tag{29}$$

$$\sum_{t=1}^{T} \kappa_{mt}(d)\kappa_{nt}(d) = O(1) \tag{30}$$

Proof of Lemma A.2. Proof of (28) and (29): See JN (2016, Lemma B.1 line (B.9)). Proof of (30): Using JN (2016, Lemma A.3 line (A.7)) we have that

$$|\kappa_{mt}(d)| \leq c(1 + \log(t-1))^m (t-1)^{-d} 1_{(t-1 \leq 1)}$$

where c is a genetic constant. Take a small $\delta > 0$ then log term is bounded by $(t-1)^{\delta}$. Hence, it follows that

$$\left|\sum_{t=1}^{T} \kappa_{mt}(d)\kappa_{nt}(d)\right| \leq c\sum_{t=1}^{T} t^{2(\delta-d)}.$$

For $T \to \infty$ this term is bounded if $2(\delta - d) < -1$ or similarly by taking $\delta < d - 1/2$. □

Lemma A.3. *Let $d > 1/2$, then we have that:*

$$m(d) = \left(\sum_{t=1}^{T} \kappa_{0t}(d)^2\right)^{\frac{1}{T-1}} = 1 + O(T^{-1}) \tag{31}$$

$$m_d(d) = \frac{2}{T-1}\left(\sum_{t=1}^{T} \kappa_{0t}(d)^2\right)^{-\frac{T-2}{T-1}} \sum_{t=1}^{T} \kappa_{0t}(d)\kappa_{1t}(d) = O(T^{-1}) \tag{32}$$

$$\begin{aligned} m_{dd}(d) = &\frac{2}{T-1}\left(\sum_{t=1}^{T} \kappa_{0t}(d)^2\right)^{-\frac{T-2}{T-1}} \sum_{t=1}^{T}(\kappa_{0t}(d)\kappa_{2t}(d) + \kappa_{1t}(d)^2) \\ &- 4\frac{T-2}{(T-1)^2}\left(\sum_{t=1}^{T} \kappa_{0t}(d)^2\right)^{-\frac{2T-3}{T-1}} \left(\sum_{t=1}^{T} \kappa_{0t}(d)\kappa_{1t}(d)\right)^2 = O(T^{-1}) \end{aligned} \tag{33}$$

Proof of Lemma A.3. Proof of (31): Let $a \in \mathbb{R}$ then

$$a^{\frac{1}{T-1}} = 1 + \frac{\log(a)}{T} + O(T^{-2}) \tag{34}$$

Proof follows from the Taylor expension in (34) and (30).
Proof of (32) and (33): Follows from (30) .

□

Lemma A.4. *Suppose Assumption 3.1 holds. Then*

$$E\left(\sum_{s=1}^{T} \kappa_{0s}(d)S_{0s}^{+}(d)\sum_{t=1}^{T} \kappa_{0t}(d)S_{1t}^{+}(d)\right) = \sigma_0^2\sum_{t=1}^{T} \kappa_{0t}(d)\kappa_{1t}(d)$$

Proof of Lemma A.4. Note that $S_{0s}^{+} = \epsilon_s$, $S_{1t}^{+} = -\sum_{k=1}^{t-1} k^{-1}\epsilon_{t-k}$, and $\kappa_{0t(d)} = \pi_{t-1}(1-d)$. It can be easily shown that

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

using $D\pi_{k-t}(u)|_{u=0} = \frac{1}{(k-t)} 1_{\{k-t\geq 1\}}$, see JN (2016, Lemma A.4 line (A.13)). Then we find that

$$
\begin{aligned}
E\left(\sum_{s=1}^{T} \kappa_{0s}(d) S_{0s}^{+}(d) \sum_{t=1}^{T} \kappa_{0t}(d) S_{1t}^{+}(d)\right) &= -\sigma_0^2 \sum_{t=1}^{T-1} \kappa_{0t}(d) \sum_{k=1}^{T} \kappa_{0k}(d) D\pi_{k-t}(u)|_{u=0} \\
&= -\sigma_0^2 \sum_{k=1}^{T} \kappa_{0k}(d) \sum_{t=1}^{T-1} \kappa_{0t}(d) D\pi_{k-t}(u)|_{u=0} \\
&= -\sigma_0^2 \sum_{k=1}^{T-1} \kappa_{0k}(d) \sum_{t=0}^{k-1} \pi_t(1-d) D\pi_{(k-1)-t}(u)|_{u=0} \\
&= -\sigma_0^2 \sum_{k=1}^{T} \kappa_{0k}(d) D\pi_{k-1}(1-d+u)|_{u=0}
\end{aligned}
$$

where the last equality holds by JN (2016, Lemma A.4 line (A.17)). The proof follows since

$$
D\pi_{t-1}(1-d+u)|_{u=0} = -\kappa_{1t}(1-d)
$$

by definition. $\square$

Lemma A.5. *Suppose Assumption 3.1 holds and let $d > 1/2$. Then*

$$
\zeta_{T,1}(d) = -\sigma_0^{-2} \frac{E\left(\sum_{t=1}^{T} S_{0t}^{+} \kappa_{0t} \sum_{s=1}^{T} S_{1s}^{+} \kappa_{0t}\right)}{\sum_{t=1}^{T} \kappa_{0t}\kappa_{0t}} \to -\left(\Psi(2d-1) - \Psi(d)\right) \tag{35}
$$

$$
\zeta_{T,2}(d) = \frac{\sum_{t=1}^{T} \kappa_{0t}\kappa_{1t}}{\left(\sum_{t=1}^{T} \kappa_{0t}\kappa_{0t}\right)^{\frac{T-2}{T-1}}} \to \left(\Psi(2d-1) - \Psi(d)\right) \tag{36}
$$

Proof of Lemma A.5. Proof of (35): Follows from Lemma A.4, (28) and (29) . Proof of (36): Proof follows from (28) and (29). $\square$

We now apply the previous Lemmas A.1-A.5 and find asymptotic results for the derivatives $L^{*}_{MCSS}(d_0)$.

Lemma A.6. *Let the model for the data x_t, $t = 1,\ldots,T$, be given by (1) and let Assumptions 3.1 and 3.2 be satisfied with $d_0 > 1/2$. Then the normalized derivatives of the modified likelihood function L^{*}_{MCSS}, see (12), satisfy*

$$
\sigma_0^{-2} T^{-1/2} D L^{*}_{MCSS}(d_0) = A_0 + T^{-1/2} A_1 + O_P(T^{-1}), \tag{37}
$$

$$
\sigma_0^{-2} T^{-1} D^2 L^{*}_{MCSS}(d_0) = B_0 + T^{-1/2} B_1 + O_P(T^{-1}\log(T)), \tag{38}
$$

$$
\sigma_0^{-2} T^{-1} D^3 L^{*}_{MCSS}(d_0) = C_0 + O_P(T^{-1/2}), \tag{39}
$$

where

$$
\begin{aligned}
A_0 &= M_{01T}^{+}, & E(A_1) &= \zeta_{T,1}(d_0) + \zeta_{T,2}(d_0), \\
B_0 &= \zeta_2, & B_1 &= M_{11T}^{+} + M_{02T}^{+}, \\
C_0 &= -6\zeta_3. & &
\end{aligned}
$$

Here, $\zeta_{T,1}(d_0)$, $\zeta_{T,2}(d_0)$ and M_{mnT}^{+}, are given in (35), (36), and (27), respectively, and $\zeta_2 = \pi^2/6$ and $\zeta_3 \approx 1.2021$.

Proof of Lemma A.6. Note that the derivatives of the modified likelihood function $L^*_{MCSS}(d) = m(d)L^*(d)$, see (12), with respect to d are equal to

$$
\begin{aligned}
DL^*_{MCSS}(d) &= m(d)DL^*(d) + m_d(d)L^*(d),\\
D^2L^*_{MCSS}(d) &= m(d)D^2L^*(d) + m_{dd}(d)L^*(d) + 2m_d(d)DL^*(d),\\
D^3L^*_{MCSS}(d) &= m(d)D^3L^*(d) + 3m_d(d)D^2L^*(d) + 3m_{dd}(d)DL^*(d) + m_{ddd}(d)L^*(d).
\end{aligned}
$$

In the following we find an expression for the derivatives with respect to d evaluated for $d = d_0$ and suppress the dependence on d_0.

Proof of (37): JN (2016, Lemma B.4 line (B.26)) states that

$$\sigma_0^{-2}T^{-1/2}DL^* = A_0 + T^{-1/2}\tilde{A}_1 + O_P(T^{-1}),$$

where $A_0 = M^+_{01T}$ and $E(\tilde{A}_1) = \zeta_{T,1}(d_0)$. We find from Lemma A.3 and JN (2016, Lemma B.4 line (B.26)) that

$$
\begin{aligned}
\sigma_0^{-2}T^{-1/2}DL^*_{MCSS} &= m\sigma_0^{-2}T^{-1/2}DL^* + \sigma_0^{-2}T^{-1/2}m_dL^*,\\
&= \sigma_0^{-2}T^{-1/2}DL^* + \sigma_0^{-2}T^{-1/2}m_dL^* + O_P(T^{-1}),\\
&= A_0 + T^{-1/2}\left(\tilde{A}_1 + \sigma_0^{-2}m_dL^*\right) + O_P(T^{-1}),\\
&= A_0 + T^{-1/2}A_1 + O_P(T^{-1}),
\end{aligned}
$$

where $m_dL^* = O_P(1)$, see Hualde and Nielsen (2020, Lemma S.1). The expectation of $A_1 = \tilde{A}_1 + \sigma_0^{-2}m_dL^*$ is equal to $\zeta_{T,1} + \zeta_{T,2}$ since

$$
\begin{aligned}
E(\sigma_0^{-2}m_dL^*) &= \sigma_0^{-2}m_d\frac{1}{2}\sum_{t=1}^{T}E\left(\Delta_0^{d_0}x_t - \kappa_{0t}\hat{\mu}\right)^2\\
&= \sigma_0^{-2}m_d\frac{1}{2}\sum_{t=1}^{T}E\left(\epsilon_t - \kappa_{0t}\frac{\sum_{t=1}^{T}\kappa_{0t}\epsilon_t}{\sum_{t=1}^{T}\kappa_{0t}^2}\right)^2\\
&= \sigma_0^{-2}m_d\frac{1}{2}\sigma_0^2(T-1)\\
&= \left(\sum_{t=1}^{T}\kappa_{0t}(d)^2\right)^{-\frac{T-2}{T-1}}\sum_{t=1}^{T}\kappa_{0t}(d)\kappa_{1t}(d)\\
&= \zeta_{T,2}
\end{aligned}
$$

where $\zeta_{T,1}$ and $\zeta_{T,2}$ are defined in Lemma A.5.

Proof of (38): JN (2016, Lemma B.4 line (B.27)) states that

$$\sigma_0^{-2}T^{-1/2}D^2L^* = B_0 + T^{-1/2}B_1 + O_P(T^{-1}\log(T)),$$

where $B_0 = \zeta_2$ and $B_1 = M^+_{11T} + M^+_{02T}$. We find from Lemma A.3 and JN (2016, Lemma B.4 line (B.27)) that

$$
\begin{aligned}
\sigma_0^{-2}T^{-1}D^2L^*_{MCSS} &= m\sigma_0^{-2}T^{-1}D^2L^* + m_{dd}\sigma_0^{-2}T^{-1}L^* + 2m_d\sigma_0^{-2}T^{-1}DL^*,\\
&= \sigma_0^{-2}T^{-1}D^2L^* + m_{dd}\sigma_0^{-2}T^{-1}L^* + 2m_dDL^* + O_P(T^{-1})\\
&= B_0 + T^{-1/2}B_1 + m_{dd}\sigma_0^{-2}T^{-1}L^* + \sigma_0^{-2}T^{-1}2m_dDL^*\\
&\quad + O_P(T^{-1}\log(T)).
\end{aligned}
$$

Furthermore the terms

$$m_{dd}\sigma_0^{-2}T^{-1}L^* = O_P(T^{-1})$$
$$\sigma_0^{-2}T^{-1}2m_dDL^* = O_P(T^{-3/2})$$

and thus can be ignored so that we have

$$\sigma_0^{-2}T^{-1}D^2L^*_{MCSS} = B_0 + T^{-1/2}B_1 + O_P(T^{-1}\log(T)).$$

Proof of (39): JN (2016, Lemma B.4 line (B.28)) states that

$$\sigma_0^{-2}T^{-1/2}D^3L^* = C_0 + O_P(T^{-1/2}),$$

where $C_0 = -6\zeta_3$. We find from Lemma A.3 and JN (2016, Lemma B.4 line (B.28)) that

$$\begin{aligned}\sigma_0^{-2}T^{-1}D^3L^*_{MCSS} &= \sigma_0^{-2}T^{-1}mD^3L^* \\ &\quad + \sigma_0^{-2}T^{-1}(3m_dD^2L^* + 3m_{dd}DL^* + m_{ddd}L^*) \\ &= C_0 + \sigma_0^{-2}T^{-1}(3m_dD^2L^* + 3m_{dd}DL^* + m_{ddd}L^*) + O_P(T^{-1/2}).\end{aligned}$$

Notice that $m_{ddd} = O(T^{-1})$, proof follows in a similar way as the proof of (33), and together with Lemma A.3 we get

$$\sigma_0^{-2}T^{-1}(3m_dD^2L^* + 3m_{dd}DL^* + m_{ddd}L^*) = O_P(T^{-1})$$

and hence can be ignored so we have that

$$\sigma_0^{-2}T^{-1}D^3L^*_{MCSS} = C_0 + O_P(T^{-1/2}).$$

□

A.1.2 Stationary region

In Lemmas A.7, A.4, A.9 and A.10 we investigate the order of magnitude of functions involving the deterministic term $\kappa_{mt}(d)$ and the stochastic term S^+_{mt} and product moments containing these. In Lemma A.8 we investigate the order of magnitude involving the modification term $m(d)$ and derivatives of these. These lemmas are then used together with Lemmas A.1 to find asymptotic results for the derivatives $L^*(d_0)$, $L^*_{\mu_0}$ and L^*_{MCSS}.

Lemma A.7. *Let $d < 1/2$ and $m, n \geq 0$, then we have that:*

$$\frac{1}{T^{1-2d}}\sum_{t=1}^{T} k_{0t}(d)^2 \to \frac{1}{\Gamma(1-d)^2(1-2d)} \tag{40}$$

$$\begin{aligned}\frac{1}{T^{1-2d}\log(T)}\sum_{t=1}^{T} k_{0t}(d)k_{1t}(d) = &-\frac{1}{\Gamma(1-d)^2(1-2d)} \\ &+ \frac{1}{\log(T)}\frac{(1-2d)\Psi(1-d)+1}{(1-2d)^2\Gamma(1-d)^2} \\ &+ o(\log(T)^{-1})\end{aligned} \tag{41}$$

$$\frac{1}{T^{1-2d}\log^{m+n}(T)}\sum_{t=1}^{T} k_{mt}(d)k_{nt}(d) = O(1) \tag{42}$$

Proof of Lemma A.7. Proof of (40): Proof follows from Hualde and Nielsen (2020, Lemmas S.10 and S.11).
Proof of (41): First we take the derivative of (40) with respect to d on both sides:

$$\frac{2\log(T)}{T^{1-2d}}\sum_{t=1}^{T}\kappa_{0t}(d)^2+\frac{2}{T^{1-2d}}\sum_{t=1}^{T}\kappa_{0t}(d)\kappa_{1t}(d)=\frac{(2-4d)\Psi(1-d)+2}{(1-2d)^2\Gamma(1-d)^2}+o(1)$$

Dividing the left-hand side with $2\log(T)$ gives

$$\frac{1}{T^{1-2d}}\sum_{t=1}^{T}\kappa_{0t}(d)^2+\frac{1}{T^{1-2d}\log(T)}\sum_{t=1}^{T}\kappa_{0t}(d)\kappa_{1t}(d)=\frac{1}{\log(T)}\frac{(1-2d)\Psi(1-d)+1}{(1-2d)^2\Gamma(1-d)^2}$$
$$+o(\log(T)^{-1})$$

The proof follows directly by using (40).
Proof of (42): Using JN (2016, Lemma B.4 line (A.7)) we have that

$$\frac{1}{T^{1-2d}\log^{m+n}(T)}\left|\sum_{t=1}^{T}\kappa_{mt}\kappa_{nt}\right|\leq c\frac{1}{T^{1-2d}\log^{m+n}(T)}\sum_{t=1}^{T}\log^{m+n}(t)t^{-2d}$$

where c is a generic positive constant. Note that

$$\frac{1}{T^{1-2d}\log^{m+n}(T)}\sum_{t=1}^{T}\log^{m+n}(t)t^{-2d}=\frac{1}{T^{1-2d}}\sum_{t=1}^{T}t^{-2d}+o(1)\rightarrow\frac{1}{1-2d}\qquad(43)$$

by applying summation by parts and Hualde and Nielsen (2020, Lemma S.10) for $d<1/2$.

□

Lemma A.8. *Let $d<1/2$, then we have that:*

$$m(d)=\left(\sum_{t=1}^{T}\kappa_{0t}(d)^2\right)^{\frac{1}{T-1}}=1+O(T^{-1}\log(T))\qquad(44)$$

$$m_d(d)=\frac{2}{T-1}\left(\sum_{t=1}^{T}\kappa_{0t}(d)^2\right)^{-\frac{T-2}{T-1}}\sum_{t=1}^{T}\kappa_{0t}(d)\kappa_{1t}(d)=O(T^{-1}\log(T))\qquad(45)$$

$$m_{dd}(d)=\frac{2}{T-1}\left(\sum_{t=1}^{T}\kappa_{0t}(d)^2\right)^{-\frac{T-2}{T-1}}\sum_{t=1}^{T}(\kappa_{0t}(d)\kappa_{2t}(d)+\kappa_{1t}(d)^2)$$
$$-4\frac{T-2}{(T-1)^2}\left(\sum_{t=1}^{T}\kappa_{0t}(d)^2\right)^{-\frac{2T-3}{T-1}}\left(\sum_{t=1}^{T}\kappa_{0t}(d)\kappa_{1t}(d)\right)^2=O(T^{-1}\log^2(T))$$
$$(46)$$

Proof of Lemma A.8. Proof of (44): The proof follows from writing (44) into

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

and using Taylor's expansion in (34) twice together with (40) to conclude that:

$$(1 + O(T^{-1}))(1 + O(T^{-1}\log(T))) = 1 + O(T^{-1}\log(T))$$

Proof of (45): The proof follows directly from (40) and (41).
Proof of (46): By (42) we have that $\sum_{t=1}^{T} \kappa_{0t}(d)\kappa_{2t}(d) = O(T^{1-2d}\log^2(T))$. Then the proof follows from (40) and (41). □

Lemma A.9. *Suppose Assumption 3.1 holds and let* $d < \frac{1}{2}$. *Then*

$$\frac{1}{T^{1/2-d}\log^{m+n}(T)} \sum_{t=1}^{T} S_{mt}(d)k_{nt}(d) = O_P(1).$$

Proof of Lemma A.9. We note that for $m \geq 1$

$$\begin{aligned}\sum_{t=1}^{T} S_{mt}(d)\kappa_{nt}(d) &= \sum_{t=1}^{T-1} \epsilon_t \sum_{k=1}^{T} \kappa_{nk}(d)(-1)^m D^m \pi_{k-t}(0)\\ &= \sum_{t=1}^{T-1} \epsilon_t \sum_{k=1}^{T} (-1)^n D^n \pi_{k-1}(1-d)(-1)^m D^m \pi_{k-t}(0)\end{aligned}$$

Then

$$Var\left(\sum_{t=1}^{T} S_{mt}(d)\kappa_{nt}(d)\right) = \sigma_0^2 \sum_{t=1}^{T-1}\left(\sum_{k=1}^{T}(-1)^n D^n \pi_{k-1}(1-d)(-1)^m D^m \pi_{k-t}(0)\right)^2 \quad (47)$$

For $m \geq 1$ we have

$$\begin{aligned}\left|\sum_{k=1}^{T}(-1)^n D^n \pi_{k-1}(1-d)(-1)^m D^m \pi_{k-t}(0)\right| &\leq c\sum_{k=1}^{T} |D^n \pi_{k-1}(1-d)|\,|D^m \pi_{k-t}(0)|\\ &\leq c\sum_{k=1}^{T}(1+\log(k-1))^n (k-1)^{-d} 1_{\{k-1\geq 1\}}\\ &\quad (k-t)^{-1}(1+\log(k-t))^{m-1} 1_{\{k-t\geq 1\}}\\ &\leq c\sum_{k=t+1}^{T} \log^{n+m-1}(k-1)(k-1)^{-d}(k-t)^{-1}\end{aligned}$$

It can be readily shown that

$$c\sum_{k=t+1}^{T} \log^{n+m-1}(k-1)(k-1)^{-d}(k-t)^{-1} = O(\log^{n+m-1}(t)t^{-d}\log(T)) \quad (48)$$

using summation by parts. Then the proof follows by (47) and (48) together with (43). The proof for $m = 0$ follows from (42).

□

Lemma A.10. *Suppose Assumption 3.1 holds and let $d < 1/2$. Then*

$$\begin{aligned}
\zeta^*_{T,1}(d) &= -\log^{-1}(T)\sigma_0^{-2}\frac{E\left(\sum_{t=1}^T S^+_{0t}\kappa_{0t}\sum_{s=1}^T S^+_{1s}\kappa_{0t}\right)}{\sum_{t=1}^T \kappa_{0t}\kappa_{0t}} \\
&= -\log^{-1}(T)\frac{\sum_{t=1}^T \kappa_{0t}\kappa_{1t}}{\sum_{t=1}^T \kappa_{0t}\kappa_{0t}} \\
&= 1-\log^{-1}(T)\left(\psi(1-d)+(1-2d)^{-1}\right)+o(\log^{-1}(T)) \qquad (49)\\
\zeta^*_{T,2}(d) &= \log^{-1}(T)\frac{\sum_{t=1}^T \kappa_{0t}\kappa_{1t}}{\left(\sum_{t=1}^T \kappa_{0t}\kappa_{0t}\right)^{\frac{T-2}{T-1}}} \\
&= -1+\log^{-1}(T)\left(\psi(1-d)+(1-2d)^{-1}\right)+o(\log^{-1}(T)). \qquad (50)
\end{aligned}$$

Proof of Lemma A.10. Proof of (49): Follows from Lemma A.4 and A.2.
Proof of (50): Follows from Lemma A.2.

□

We now apply the previous Lemmas A.1, A.2-A.10 and find asymptotic results for the derivatives $L^*(d_0)$, $L^*_{\mu_0}$ and L^*_{MCSS}.

Lemma A.11. *Let the model for the data x_t, $t = 1,\ldots,T$, be given by (1) and let Assumptions 3.1 and 3.2 be satisfied with $d_0 < 1/2$. Then the normalized derivatives of the likelihood function L^*, see (7), satisfy*

$$\sigma_0^{-2}T^{-1/2}DL^*(d_0) = A_0 + T^{-1/2}\log(T)A_1, \qquad (51)$$

$$\sigma_0^{-2}T^{-1}D^2L^*(d_0) = B_0 + T^{-1/2}B_1 + O_P(T^{-1}\log^2(T)), \qquad (52)$$

$$\sigma_0^{-2}T^{-1}D^3L^*(d_0) = C_0 + O_P(T^{-1/2}), \qquad (53)$$

where

$$\begin{aligned}
A_0 &= M^+_{01T}, & E(A_1) &= \zeta^*_{T,1}(d_0),\\
B_0 &= \zeta_2, & B_1 &= M^+_{11T} + M^+_{02T},\\
C_0 &= -6\zeta_3.
\end{aligned}$$

Here, $\zeta^*_{T,1}(d_0)$ and M^+_{mnT} are given in (49) and (27), respectively, and $\zeta_2 = \pi^2/6$ and $\zeta_3 \approx 1.2021$.

Proof of Lemma A.11. Recall that $L^*(d) = L(d,\mu(d))$, where from $L(d,\mu(d))$ is given in (5). We denote the partial derivatives with respect to d and μ of $L(d,\mu(d))$ by a subscript. Then the derivatives of the likelihood of $L^*(d)$, denoted by $D^mL^*(d)$ are equal to

$$\begin{aligned}
DL^*(d) &= L_d(d,\mu(d)) + L_\mu(d,\mu(d))\mu_d(d),\\
D^2L^*(d) &= L_{dd}(d,\mu(d)) + 2L_{d\mu}(d,\mu(d))\mu_d(d) + L_{\mu\mu}(d,\mu(d))\mu_d(d)^2 + L_\mu(d,\mu(d))\mu_{dd}(d).
\end{aligned}$$

We simply the expressions by noticing that $\hat{\mu}$ is determined from $L_\mu(d,\mu(d)) = 0$. Taking on both sides the derivative of with respect to d of $L_\mu(d,\mu(d)) = 0$ implies $L_{d\mu}(d,\mu(d)) + L_{\mu\mu}(d,\mu(d))\mu_d(d) = 0$ so that

$$\begin{aligned} DL^*(d) &= L_d(d,\mu(d)), \\ D^2L^*(d) &= L_{dd}(d,\mu(d)) - \frac{L_{d\mu}(d,\mu(d))^2}{L_{\mu\mu}(d,\mu(d))}. \end{aligned} \tag{54}$$

We evaluate the derivatives for $d = d_0$ and to make it easily readable we suppress the dependence.

Proof of (51): We find that

$$\begin{aligned} \sigma_0^{-2}T^{-1/2}DL^* &= \sigma_0^{-2}T^{-1/2}\sum_{t=1}^{T}(S_{0t}^+ - (\hat{\mu}-\mu_0)\kappa_{0t})(S_{1t}^+ - (\hat{\mu}-\mu_0)\kappa_{1t}) \\ &= \sigma_0^{-2}T^{-1/2}\sum_{t=1}^{T}S_{0t}^+S_{1t}^+ - \sigma_0^{-2}T^{-1/2}\log(T)T^{\frac{1}{2}-d}(\hat{\mu}-\mu_0)\frac{\sum_{t=1}^T S_{0t}^+\kappa_{1t}}{T^{1/2-d}\log(T)} \\ &\quad - \sigma_0^{-2}T^{-1/2}\log(T)T^{\frac{1}{2}-d}(\hat{\mu}-\mu_0)\frac{1}{T^{1/2-d}\log(T)}\sum_{t=1}^{T}S_{1t}^+\kappa_{0t} \\ &\quad + \sigma_0^{-2}T^{-1/2}\log(T)T^{1-2d}(\hat{\mu}-\mu_0)^2\frac{1}{T^{1-2d}\log(T)}\sum_{t=1}^{T}\kappa_{0t}\kappa_{1t} \end{aligned}$$

The first term is $O_P(1)$ and the rest terms are all of the order of $T^{-1/2}\log(T)$. Taking the expectation of the rest terms we get $T^{-1/2}\log(T)$ times

$$\begin{aligned} &-\sigma_0^{-2}\frac{\frac{1}{T^{1-2d_0}\log(T)}E\left(\sum_{t=1}^T S_{0t}^+k_{0t}\sum_{t=1}^T S_{0t}^+k_{1t}\right)}{\frac{1}{T^{1-2d_0}}\sum_{t=1}^T k_{0t}^2} \\ &-\sigma_0^{-2}\frac{\frac{1}{T^{1-2d_0}\log(T)}E\left(\sum_{t=1}^T S_{0t}^+k_{0t}\sum_{s=1}^T S_{1s}^+k_{0s}\right)}{\frac{1}{T^{1-2d_0}}\sum_{t=1}^T k_{0t}^2} \\ &+\sigma_0^{-2}\frac{\frac{1}{T^{2-4d_0}\log(T)}E\left(\left(\sum_{t=1}^T S_{0t}^+k_{0t}\right)^2\sum_{t=1}^T S_{0t}^+k_{1t}\right)}{\left(\frac{1}{T^{1-2d_0}}\sum_{t=1}^T k_{0t}^2\right)^2} \\ &= -\sigma_0^{-2}\frac{\frac{1}{T^{1-2d_0}\log(T)}\sum_{t=1}^T k_{0t}k_{1t}}{\frac{1}{T^{1-2d}}\sum_{t=1}^T k_{0t}^2} - \sigma_0^{-2}\frac{\frac{1}{T^{1-2d_0}\log(T)}E\left(\sum_{t=1}^T S_{0t}^+\kappa_{0t}\sum_{s=1}^T S_{1s}^+\kappa_{0t}\right)}{\frac{1}{T^{1-2d_0}}\sum_{t=1}^T \kappa_{0t}^2} \\ &+\sigma_0^{-2}\frac{\frac{1}{T^{1-2d_0}\log(T)}\sum_{t=1}^T k_{0t}k_{1t}}{\frac{1}{T^{1-2d}}\sum_{t=1}^T k_{0t}^2} \\ &= -\sigma_0^{-2}\frac{\frac{1}{T^{1-2d_0}\log(T)}E\left(\sum_{t=1}^T S_{0t}^+\kappa_{0t}\sum_{s=1}^T S_{1s}^+\kappa_{0t}\right)}{\frac{1}{T^{1-2d_0}}\sum_{t=1}^T \kappa_{0t}^2} \\ &= -\frac{\frac{1}{T^{1-2d_0}\log(T)}\sum_{t=1}^T \kappa_{0t}\kappa_{1t}}{\frac{1}{T^{1-2d_0}}\sum_{t=1}^T \kappa_{0t}^2} \\ &= \zeta_{T,1}^* \end{aligned}$$

where the second-to-last equality follows from Lemma A.4 and $\zeta^*_{T,1}$ is defined in Lemma A.10.

Proof of (52): We find for the second term of D^2L^* given in (54) that $L_{\mu\mu}(d_0,\mu(d_0))$ $= \sigma_0^2 \sum_{t=1}^T \kappa_{0t}^2 = O(T^{1-2d_0})$ and

$$\begin{aligned} L_{d\mu}(d_0,\mu(d_0)) &= -\sum_{t=1}^T \left(S_{0t}^+ - (\hat\mu-\mu_0)\kappa_{0t}\right)\kappa_{1t} \\ &\quad -\sum_{t=1}^T \left(S_{1t}^+ - (\hat\mu-\mu_0)\kappa_{1t}\right)\kappa_{0t} \\ &= O_P(T^{\frac{1}{2}-d_0}\log(T)) \end{aligned}$$

such that $T^{-1}\frac{L_{d\mu}(d_0,\mu(d_0))^2}{L_{\mu\mu}(d_0,\mu(d_0))} = O_P(T^{-1}\log^2(T))$ and hence can be ignored. Thus, we get

$$\begin{aligned} \sigma_0^{-2}T^{-1}D^2L^* = \sigma_0^{-2}T^{-1}\sum_{t=1}^T \left(S_{1t}^+ - (\hat\mu-\mu_0)\kappa_{1t}\right)^2 \\ + \sigma_0^{-2}T^{-1}\sum_{t=1}^T \left(S_{0t}^+ - (\hat\mu-\mu_0)\kappa_{0t}\right)\left(S_{2t}^+ - (\hat\mu-\mu_0)\kappa_{0t}\right) \\ + O_P(T^{-1}\log^2(T)) \end{aligned}$$

ignoring terms that are of order $T^{-1}\log^2(T)$ we get

$$\begin{aligned} \sigma_0^{-2}T^{-1}D^2L^* &= \sigma_0^{-2}T^{-1}\sum_{t=1}^T E\left(S_{1t}^+\right)^2 + T^{-1/2}\left(M_{11T}^+ + M_{02T}^+\right) + O_P(T^{-1}\log^2(T)) \\ &= \zeta_2 + T^{-1/2}\left(M_{11T}^+ + M_{02T}^+\right) + O_P(T^{-1}\log^2(T)) \end{aligned}$$

Proof of (53): For the third derivative it can be shown that the extra terms involving the derivatives $\mu_d(d_0)$ and $\mu_{dd}(d_0)$ can be ignored and we find

$$\begin{aligned} \sigma_0^{-2}T^{-1}D^3L^* &= \sigma_0^{-2}3T^{-1}\sum_{t=1}^T \left(S_{1t}^+ - (\hat\mu-\mu_0)\kappa_{1t}\right)\left(S_{2t}^+ - (\hat\mu-\mu_0)\kappa_{2t}\right) \\ &\quad + \sigma_0^{-2}T^{-1}\sum_{t=1}^T \left(S_{0t}^+ - (\hat\mu-\mu_0)\kappa_{1t}\right)\left(S_{3t}^+ - (\hat\mu-\mu_0)\kappa_{2t}\right) \\ &\quad + O_P(T^{-1}\log^3(T)) \\ &= 3\sigma_0^{-2}T^{-1}\sum_{t=1}^T E\left(S_{1t}^+S_{2t}^+\right) + 3T^{-1/2}M_{12T}^+ + T^{-1/2}M_{03T}^+ \\ &= -6\zeta_3 + O_P(T^{-1/2}) \end{aligned}$$

where last equality uses Lemma A.1. □

Lemma A.12. *Let the model for the data x_t, $t = 1,\ldots,T$, be given by* (1) *and let Assumptions 3.1 and 3.2 be satisfied with $d_0 < 1/2$. Then the normalized derivatives* *of the likelihood function* $L_{\mu_0}^*$*, see* (11)*, satisfy*

$$\sigma_0^{-2}T^{-1/2}DL_{\mu_0}^*(d_0) = A_0, \tag{55}$$
$$\sigma_0^{-2}T^{-1}D^2L_{\mu_0}^*(d_0) = B_0 + T^{-1/2}B_1 + O(T^{-1}\log(T)), \tag{56}$$
$$\sigma_0^{-2}T^{-1}D^3L_{\mu_0}^*(d_0) = C_0 + O_P(T^{-1/2}), \tag{57}$$

where

$$\begin{aligned} A_0 &= M_{01T}^+, \\ B_0 &= \zeta_2, \qquad\qquad B_1 = M_{11T}^+ + M_{02T}^+, \\ C_0 &= -6\zeta_3. \end{aligned}$$

Here M_{mnT}^+ *is given in* (27) *and* $\zeta_2 = \pi^2/6$ *and* $\zeta_3 \approx 1.2021$.

Proof of Lemma A.12. Note that derivatives of $\Delta_0^d(x_t - \mu_0)$ with respect to d, evaluated at $d = d_0$, are of the from

$$D^m\Delta_0^{d_0}(x_t - \mu_0) = S_{mt}^+.$$

So the derivatives only depend on the asymptotic properties of S_{mt}^+ which is given in Lemma A.1.
Proof of (55): Note that

$$\begin{aligned} \sigma_0^{-2}T^{-1/2}DL_{\mu_0}^* &= \sigma_0^{-2}T^{-1/2}\sum_{t=1}^{T} S_{0t}^+S_{1t}^+ \\ &= M_{01T}^+ \end{aligned}$$

by noticing that $E\left(S_{0t}^+S_{1t}^+\right) = 0$, see Lemma A.1.
Proof of (56): Note that

$$\begin{aligned} \sigma_0^{-2}T^{-1}D^2L_{\mu_0}^* &= \sigma_0^{-2}T^{-1}\sum_{t=1}^{T}\left(S_{1t}^+\right)^2 + \sigma_0^{-2}T^{-1}\sum_{t=1}^{T} S_{0t}^+S_{2t}^+ \\ &= \sigma_0^{-2}T^{-1}\sum_{t=1}^{T} E\left(S_{1t}^+\right)^2 + T^{-1/2}\left(M_{11T}^+ + M_{02T}^+\right) \\ &= \zeta_2 + T^{-1/2}\left(M_{11T}^+ + M_{02T}^+\right) + O(T^{-1}\log(T)) \end{aligned}$$

by using Lemma A.1 and noticing that

$$\sigma_0^{-2}T^{-1}\sum_{t=1}^{T} E\left(S_{1t}^+\right)^2 = \zeta_2 + O(T^{-1}\log(T)).$$

Proof of (57): We note that

$$\begin{aligned} \sigma_0^{-2}T^{-1}D^3L_{\mu_0}^* &= 3\sigma_0^{-2}T^{-1}\sum_{t=1}^{T} S_{1t}^+S_{2t}^+ + \sigma_0^{-2}T^{-1}\sum_{t=1}^{T} S_{0t}^+S_{3t}^+ \\ &= 3\sigma_0^{-2}T^{-1}\sum_{t=1}^{T} E\left(S_{1t}^+S_{2t}^+\right) + 3T^{-1/2}M_{12T}^+ + T^{-1/2}M_{03T}^+ \\ &= -6\zeta_3 + O_P(T^{-1/2}) \end{aligned}$$

where last equality uses Lemma A.1.

$\square$

Lemma A.13. *Let the model for the data x_t, $t = 1,\dots,T$, be given by (1) and let Assumptions 3.1 and 3.2 be satisfied with $d_0 < 1/2$. Then the normalized derivatives of the modified likelihood function L^*_{MCSS}, see (12), satisfy*

$$\sigma_0^{-2}T^{-1/2}DL^*_{MCSS}(d_0) = A_0 + T^{-1/2}\log(T)A_1 + O_P(T^{-1}\log(T)), \tag{58}$$
$$\sigma_0^{-2}T^{-1}D^2L^*_{MCSS}(d_0) = B_0 + T^{-1/2}B_1 + O_P(T^{-1}\log^2(T)), \tag{59}$$
$$\sigma_0^{-2}T^{-1}D^3L^*_{MCSS}(d_0) = C_0 + O_P(T^{-1/2}), \tag{60}$$

where

$$\begin{aligned} A_0 &= M^+_{01T}, & E(A_1) &= \zeta^*_{T,1}(d_0) + \zeta^*_{T,2}(d_0), \\ B_0 &= \zeta_2, & B_1 &= M^+_{11T} + M^+_{02T}, \\ C_0 &= -6\zeta_3. \end{aligned}$$

Here, $\zeta^*_{T,1}(d_0)$, $\zeta^*_{T,2}(d_0)$ and M^+_{mnT}, are given in (49), (50), and (27), respectively, and $\zeta_2 = \pi^2/6$ and $\zeta_3 \approx 1.2021$.

Proof of A.13. The proof is omitted as it follows from the same approach as in the proof of Lemma A.6 and is therefore straightforward to proof. $\square$

A.2 Proof of Theorem 3.1

We note that the MCSS estimator is equal to

$$\begin{aligned} \hat{d}_{MCSS} &= \operatorname*{argmin}_{d\in\mathbb{D}} L^*_{MCSS}(d), \\ &= \operatorname*{argmin}_{d\in\mathbb{D}} \log\left(m(d)\frac{2}{T}L^*(d)\right) \end{aligned}$$

so that the objective function equals $\tilde{L}(d) = \log\left(m(d)\frac{2}{T}L^*(d)\right) = \log(m(d)) + \log\left(\frac{2}{T}L^*(d)\right)$. We also note that $R(d) = \frac{2}{T}L^*(d)$ is the same objective function as in Hualde and Nielsen (2020). Fix $\epsilon > 0$ and let $M_\epsilon = \{d \in \mathbb{D} : |d - d_0| < \epsilon\}$ and $\bar{M}_\epsilon = \{d \in \mathbb{D} : |d - d_0| \geq \epsilon\}$. Then

$$\begin{aligned} \Pr\left(\hat{d}_{MCSS} \in \bar{M}\right) &= \Pr\left(\inf_{d\in\bar{M}_\epsilon} \tilde{L}(d) \leq \inf_{d\in M_\epsilon} \tilde{L}(d)\right), \\ &\leq \Pr\left(\inf_{d\in\bar{M}_\epsilon} \tilde{L}(d) \leq \tilde{L}(d_0)\right), \\ &\leq \Pr\left(\inf_{d\in\bar{M}_\epsilon} \log(R(d)) - \log(R(d_0)) \leq \log(m(d_0)) - \inf_{d\in\mathbb{D}} \log(m(d))\right), \end{aligned}$$

From Hualde and Nielsen (2020) we know that, as $T \to \infty$, we have that

$$\Pr\left(\inf_{d \in \bar{M}_\epsilon} \log(R(d)) - \log(R(d_0)) \leq 0\right) \to 0$$

So to proof consistency it remains to show that

$$\log(m(d_0)) - \inf_{d \in \mathbb{D}} \log(m(d)) \to 0.$$

and the proof follow directly from Lemma A.3 and Lemma A.8.

To show asymptotic normality of $\hat{d}_{MCSS}$ we proceed with an usual Taylor expansion of the score function,

$$0 = DL^*_{MCSS}(\hat{d}_{MCSS}) = DL^*_{MCSS}(d_0) + \left(\hat{d}_{MCSS} - d_0\right) D^2 L^*_{MCSS}(d^*),$$

where d^* is an intermediate value satisfying $|d^* - d_0| \leq |\hat{d}_{MCSS} - d_0| \xrightarrow{p} 0$. We note that $D^2 L^*_{MCSS}(d^*) = m(d^*) D^2 L^*(d^*) + m_{dd}(d^*) L^*(d^*) + 2m_d(d^*) DL^*(d^*)$. Replacement of $D^2 L^*_{MCSS}(d^*)$ by $D^2 L^*_{MCSS}(d_0)$ can then be argued in a similar way as in Hualde and Nielsen (2020). From Lemma A.13 and A.6 we find that $\sigma_0^{-2} T^{-1/2} DL^*_{MCSS}(d_0) = M^+_{01} + O_P(T^{-1/2}\log(T))$ and $\sigma_0^{-2} T^{-1} D^2 L^*_{MCSS}(d_0) = \pi^2/6 + O_P(T^{-1/2})$ and the result follow from Lemma A.1.

A.3 Proof of Theorem 3.2

The proof of (20) and (21) is given in JN (2016, Theorem 4). We proceed with the proof of (22). First we need to show that $D^3 L^*_{MCSS}(d^*) = D^3 L^*_{MCSS}(d_0) + o_P(1)$. Following the same steps as Hualde and Nielsen (2020, Theorem 1) we can show that $\hat{d}_{MCSS}$ is consistent. Then applying Johansen and Nielsen (2010, Lemma A.3) the proof follows. We proceed by first insert the expressions of Lemma A.6 into (19) and find

$$\begin{aligned} T^{1/2}(\hat{d}_{MCSS} - d_0) = &- \frac{A_0 + T^{-1/2} A_1}{B_0 + T^{-1/2} B_1} - \frac{1}{2} T^{-1/2} \left(\frac{A_0 + T^{-1/2} A_1}{B_0 + T^{-1/2} B_1}\right)^2 \frac{C_0}{B_0 + T^{-1/2} B_1} \\ &+ o_P(T^{-1/2}) \end{aligned}$$

using the expansion

$$\frac{1}{B_0 + T^{-1/2} B_1} = \frac{1}{B_0} - T^{-1/2} \frac{B_1}{B_0^2} + T^{-1} \frac{B_1^3}{B_0^3} + \cdots \tag{61}$$

simplifies to

$$T^{1/2}(\hat{d}_{MCSS} - d_0) = -\frac{A_0}{B_0} - T^{-1/2} \left(\frac{A_1}{B_0} - \frac{A_0 B_1}{B_0^2} + \frac{1}{2} \frac{A_0^2 C_0}{B_0^3}\right) + o_P(T^{-1/2}).$$

Notice that $E(A_0) = 0$. Then the bias of $T(\hat{d}_{MCSS} - d_0)$ is

$$-\left(\frac{E(A_1)}{B_0} - \frac{E(A_0 B_1)}{B_0^2} + \frac{1}{2} \frac{E(A_0^2) C_0}{B_0^3}\right) + o(1) = -3\zeta_3 \zeta_2^{-2} + o(1)$$

since

$$-\frac{E(A_0B_1)}{B_0^2}+\frac{1}{2}\frac{E(A_0^2)C_0}{B_0^3}=3\zeta_3\zeta_2^{-2}$$

using Lemma A.1 and by noticing that

$$E(A_1)=\zeta_{T,1}(d_0)+\zeta_{T,2}(d_0)=o(1)$$

by Lemma A.5.

Next we proof (24): The proofs follow similarly as above but by noticing that $A_1=0$, see Lemma A.12. Then the bias of $T(\hat{d}_{\mu_0}-d_0)$ is equal to

$$-\left(-\frac{E(A_0B_1)}{B_0^2}+\frac{1}{2}\frac{E(A_0^2)C_0}{B_0^3}\right)+o(1)=-3\zeta_3\zeta_2^{-2}+o(1)$$

Finally we proof (23) and (25): We first work out the proof of (23). By similar argument as in the proof of (22) we have that

$$\begin{aligned}
T^{1/2}(\hat{d}-d_0)&=-\frac{A_0+T^{-1/2}\log(T)A_1}{B_0+T^{-1/2}B_1}\\
&\quad-\frac{1}{2}T^{-1/2}\left(\frac{A_0+T^{-1/2}\log(T)A_1}{B_0+T^{-1/2}B_1}\right)^2\frac{C_0}{B_0+T^{-1/2}B_1}+o_P(T^{-1/2}\log(T))\\
&=-\frac{A_0}{B_0}-T^{-1/2}\left(\frac{\log(T)A_1}{B_0}-\frac{A_0B_1}{B_0^2}+\frac{1}{2}\frac{A_0^2C_0}{B_0^3}\right)+o_P(T^{-1/2}\log(T)).
\end{aligned}$$

where the last equality follows from the expansion in (61). Notice that $E(A_0)=0$. Then the bias of $T(\hat{d}-d_0)$ is

$$\begin{aligned}
&-\left(\frac{\log(T)E(A_1)}{B_0}-\frac{E(A_0B_1)}{B_0^2}+\frac{1}{2}\frac{E(A_0^2)C_0}{B_0^3}\right)+o(1)\\
&=-3\zeta_3\zeta_2^{-2}+\frac{-\log(T)+(\Psi(1-d_0)+(1-2d_0)^{-1})}{\zeta_2}+o(1)
\end{aligned}$$

using Lemma A.1 and by noticing that

$$E(A_1)=\zeta_{T,1}^{*}(d_0)=1-\log^{-1}(T)\left(\Psi(1-d_0)+(1-2d_0)^{-1}\right)+o(\log^{-1}(T))$$

by Lemma A.10 and Lemma A.11. The proof of (25) is now trivial since $E(A_1)=o(\log^{-1}(T))$, see Lemma A.13, we have that the bias of $T(\hat{d}_{MCSS}-d_0)$ is

$$-\left(\frac{E(A_1)}{B_0}-\frac{E(A_0B_1)}{B_0^2}+\frac{1}{2}\frac{E(A_0^2)C_0}{B_0^3}\right)+o(1)=-3\zeta_3\zeta_2^{-2}+o(1)$$