QED

Queen’s Economics Department Working Paper No. 1300

The role of initial values in conditional sum-of-squares estimation of nonstationary fractional time series models

SÃÿren Johansen
University of Copenhagen and CREATES

Morten ÃŸrregaard Nielsen
Queen’s University and CREATES

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

11-2012

The role of initial values in conditional sum-of-squares estimation of nonstationary fractional time series models*

Søren Johansen† University of Copenhagen and CREATES

Morten Ørregaard Nielsen‡ Queen's University and CREATES

February 25, 2015

Abstract

In this paper we analyze the influence of observed and unobserved initial values on the bias of the conditional maximum likelihood or conditional sum-of-squares (CSS, or least squares) estimator of the fractional parameter, d, in a nonstationary fractional time series model. The CSS estimator is popular in empirical work due, at least in part, to its simplicity and its feasibility, even in very complicated nonstationary models.

We consider a process, X_t, for which data exist from some point in time, which we call $-N_0+1$, but we only start observing it at a later time, $t=1$. The parameter (d,μ,σ^2) is estimated by CSS based on the model $\Delta_0^d(X_t-\mu)=\varepsilon_t,\ t=N+1,\ldots,N+T$, conditional on $X_1,\ldots,X_N$. We derive an expression for the second-order bias of $\hat{d}$ as a function of the initial values, $X_t,\ t=-N_0+1,\ldots,N$, and we investigate the effect on the bias of setting aside the first N observations as initial values. We compare $\hat{d}$ with an estimator, $\hat{d}_c$, derived similarly but by choosing $\mu=C$. We find, both theoretically and using a data set on voting behavior, that in many cases, the estimation of the parameter μ picks up the effect of the initial values even for the choice $N=0$.

If $N_0=0$, we show that the second-order bias can be completely eliminated by a simple bias correction. If, on the other hand, $N_0>0$, it can only be partly eliminated because the second-order bias term due to the initial values can only be diminished by increasing N.

Keywords: Asymptotic expansion, bias, conditional inference, fractional integration, initial values, likelihood inference.

JEL Classification: C22.

---

*We are grateful to three anonymous referees, the co-editor, and seminar participants at various universities and conferences for comments and to James Davidson for sharing the opinion poll data set with us. We would like to thank the Canada Research Chairs program, the Social Sciences and Humanities Research Council of Canada (SSHRC), and the Center for Research in Econometric Analysis of Time Series (CREATES - DNRF78, funded by the Danish National Research Foundation) for research support.

†Email: `Soren.Johansen@econ.ku.dk`

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

1 Introduction

One of the most commonly applied inference methods in nonstationary autoregressive (AR) models, and indeed in all time series analysis, is based on the conditional sum-of-squares (CSS, or least squares) estimator, which is obtained by minimizing the sum of squared residuals. The estimator is derived from the Gaussian likelihood conditional on initial values and is often denoted the conditional maximum likelihood estimator. For example, in the AR(k) model we set aside k observations as initial values, and conditioning on these implies that Gaussian maximum likelihood estimation is equivalent to CSS estimation. This methodology was applied in classical work on ARIMA models by, e.g., Box and Jenkins (1970), and was introduced for fractional time series models by Li and McLeod (1986) and Robinson (1994), in the latter case for hypothesis testing purposes. The CSS estimator has been widely applied in the literature, also for fractional time series models. In these models, the initial values have typically been assumed to be zero, and as remarked by Hualde and Robinson (2011, p. 3154) a more appropriate name for the estimator may thus be the truncated sum-of-squares estimator. Despite the widespread use of the CSS estimator in empirical work, very little is known about its properties related to the initial values and specifically related to the assumption of zero initial values.

Recently, inference conditional on (non-zero) initial values has been advocated in theoretical work for univariate nonstationary fractional time series models by Johansen and Nielsen (2010) and for multivariate models by Johansen and Nielsen (2012a)—henceforth JN (2010, 2012a)—and Tschernig, Weber, and Weigand (2013). In empirical work, these methods have recently been applied by, for example, Carlini, Manzoni, and Mosconi (2010) and Bollerslev, Osterrieder, Sizova, and Tauchen (2013) to high-frequency stock market data, Hualde and Robinson (2011) to aggregate income and consumption data, Osterrieder and Schotman (2011) to real estate data, and Rossi and Santucci de Magistris (2013) to futures prices.

In this paper we assume the process X_t exists for $t \geq -N_0+1$, and we derive the properties of the process from the model given by the truncated fractional filter $\Delta_{-N_0}^{d_0}(X_t-\mu_0) = \varepsilon_t$ with $\varepsilon_t \sim i.i.d.(0, \sigma^2)$, for some $d_0 > 1/2$. However, we only observe X_t for $t = 1, \ldots, T_0 = N + T$, and so we estimate (d, μ, σ^2) from the conditional Gaussian likelihood for $X_{N+1}, \ldots, X_{N+T}$ given $X_1, \ldots, X_N$, which defines the CSS estimator $\hat{d}$. Our first result is to prove consistency and asymptotic normality of the estimator of d. This is of interest in its own right, not only because of the usual issue of non-uniform convergence of the objective function, but also because the estimator of μ is in fact not consistent when $d_0 > 1/2$. We then proceed to derive an analytical expression for the asymptotic second-order bias of $\hat{d}$ via a higher-order stochastic expansion of the estimator. We apply this to investigate the magnitude of the influence of observed and unobserved initial values, and to discuss the effect on the bias of setting aside a number of observations as initial values, i.e., of splitting a given sample of size $T_0 = N + T$ into N initial values and T observations for estimation. We compare $\hat{d}$ with an estimator, $\hat{d}_c$, derived from centering the data at C by restricting $\mu = C$. We find, both theoretically and using a data set on voting behavior as illustration, that in many cases, the parameter μ picks up the effect of the initial values even for the choice $N = 0$.

Finally, in a number of relevant cases, we show that the second-order bias can be eliminated, either partially or completely, by a bias correction. In the most general case, however, it can only be partly eliminated, and in particular the second-order bias term due to the initial values can only be diminished by increasing the number of initial values, N.

In the stationary case, $0 < d < 1/2$, there is a literature on Edgeworth expansions of the distribution of the (unconditional) Gaussian maximum likelihood estimator based on the joint density of the data, $(X_1, \ldots, X_T)$ in the model (1). In particular, Lieberman and Phillips (2004) find expressions for the second-order term, from which we can derive the main term of the bias in that case. We have not found any results on the nonstationary case, $d > 1/2$, for the estimator based on conditioning on initial values.

The remainder of the paper is organized as follows. In the next section we present the fractional models and in Section 3 our main results. In Section 4 we give an application of our theoretical results to a data set of Gallup opinion polls. Section 5 concludes. Proofs of our main results and some mathematical details are given in the appendices.

2 The fractional models and their interpretations

A simple model for fractional data is

$$\Delta^d (X_t - \mu) = \varepsilon_t, \ \varepsilon_t \sim i.i.d.(0, \sigma^2), \quad t = 1, \ldots, T, \tag{1}$$

where $d \geq 0$, $\mu \in \mathbb{R}$, and $\sigma^2 > 0$. The fractional filter $\Delta^d X_t$ is defined in terms of the fractional coefficients $\pi_n(u)$ from an expansion of $(1-z)^{-u} = \sum_{n=0}^{\infty} \pi_n(u) z^n$, i.e.,

$$\pi_n(u) = \frac{u(u+1)\ldots(u+n-1)}{n!} = \frac{\Gamma(u+n)}{\Gamma(u)\Gamma(n+1)} \sim \frac{n^{u-1}}{\Gamma(u)} \text{ as } n \to \infty, \tag{2}$$

where $\Gamma(u)$ denotes the Gamma function and "$\sim$" denotes that the ratio of the left- and right-hand sides converges to one. More results are collected in Appendix A.

For a given value of d such that $0 < d < 1/2$, we have $\sum_{n=0}^{\infty} \pi_n(d)^2 < \infty$. In this case, the infinite sum $X_t = \Delta^{-d} \varepsilon_t = \sum_{n=0}^{\infty} \pi_n(d) \varepsilon_{t-n}$ exists as a stationary process with a finite variance, and gives a solution to equation (1) because $\Delta^d \mu = \sum_{n=0}^{\infty} \pi_n(-d) \mu = 0$.

When $d > 1/2$, the solution to (1) is nonstationary. In that case, we discuss below two interpretations of equation (1) as a statistical model. First as an unconditional (joint) model of the stationary process $\Delta X_1, \ldots, \Delta X_T$ when $1/2 < d < 3/2$, and then as a conditional model for the nonstationary process $X_{N+1}, \ldots, X_{N+T}$ given initial values when $d > 1/2$. In the latter case we call X_t an initial value if $t \leq N$ and denote the initial values $X_n, n \leq N$, and we assume, see Section 2.2, that the variables we are measuring started at some point $-N_0 + 1$ in the past, and we truncate the fractional filter accordingly.

2.1 The unconditional fractional model and its estimation

One approach to the estimation of d from model (1) with nonstationary data is the difference-and-add-back approach based on Gaussian estimation for stationary processes. If we have the a priori information that $1/2 < d < 3/2$, say, then we could transform the data $X_0, \ldots, X_T$ to $\Delta \mathbb{X}_T = (\Delta X_1, \ldots, \Delta X_T)'$ and note that (1) can be written

$$\Delta^{d-1} \Delta (X_t - \mu) = \varepsilon_t,$$

so that ΔX_t is stationary and fractional of order $-1/2 < d - 1 < 1/2$. Note that $\Delta \mu = 0$, so the parameter μ does not enter. To calculate the unconditional Gaussian likelihood function, we then need to calculate the $T \times T$ variance matrix $\Sigma = \Sigma(d, \sigma^2) = Var(\Delta \mathbb{X}_T)$, its inverse, Σ^{-1}, and its determinant, $\det \Sigma$. This gives the Gaussian likelihood function,

$$-\frac{1}{2} \log \det \Sigma - \frac{1}{2} \Delta \mathbb{X}_T' \Sigma^{-1} \Delta \mathbb{X}_T. \tag{3}$$

A general optimization algorithm can then be applied to find the maximum likelihood estimator, $\hat{d}_{stat}$, if Σ can be calculated. This is possible by the algorithm in Sowell (1992). The estimator $\hat{d}_{stat}$ is not a CSS estimator, which is the class of estimators we study in this paper, but it was applied by Byers, Davidson, and Peel (1997) and Dolado, Gonzalo, and Mayoral (2002) in the analysis of the voting data, and by Davidson and Hashimzade (2009) to the Nile data.

The estimator $\hat{d}_{stat}$ was analyzed by Phillips and Lieberman (2004) for true value $d_0 < 1/2$. They derived an asymptotic expansion of the distribution function of $T^{1/2}(\hat{d}_{stat} - d_0)$, from which a second-order bias correction of the estimator can be derived, see Section 3.2.

In more complicated models than (1), the calculation of Σ may be computationally difficult. This is certainly the case in, say, the fractionally cointegrated vector autoregressive model of JN (2012a). However, even in much simpler models such as the usual autoregressive model, a conditional approach has been advocated for its computational simplicity, e.g., Box and Jenkins (1970), because conditional maximum likelihood estimation simplifies the calculation of estimators by reducing the numerical problem to least squares. For this reason, the conditional estimator has been very widely applied to many models, including (1). For a discussion and comparison of the numerical complexity of Gaussian maximum likelihood as in (3) and the CSS estimator, see e.g. Doornik and Ooms (2003).

2.2 The observations and initial values

It is difficult to imagine a situation where $\{X_s\}_{-\infty}^{T}$ is available, so that (1) could be applied. In general, we assume data could potentially be available from some (typically unknown) time in the past, $-N_0 + 1$, say. We therefore truncate the filter at time $-N_0$; that is, define $\Delta_{-N_0}^{d} X_t = \sum_{n=0}^{t+N_0-1} \pi_n(-d) X_{t-n}$, and consider

$$\Delta_{-N_0}^{d}(X_t - \mu) = \varepsilon_t, \ t = 1, \ldots, T_0. \tag{4}$$

as the model for the data we actually observe, namely X_t for $t = 1, \ldots, N + T = T_0$. In practice, when $N_0 > 0$, we do not observe all the data, and so we have to decide how to split a given sample of size $T_0 = N + T$ into (observed) initial values $\{X_n\}_{n=1}^{N}$ and observations $\{X_t\}_{t=N+1}^{T}$ to be modeled, and then calculate the likelihood based on the truncated filter Δ_0^d, as an approximation to the conditional likelihood based on (4). In the special case with $N_0 = 0$, the equations in (4) become

$$\begin{aligned} X_1 &= \mu + \varepsilon_1, \\ X_2 &= -\pi_1(-d) X_1 + \mu + \pi_1(-d)\mu + \varepsilon_2, \end{aligned} \tag{5}$$

etc., and μ can thus be interpreted as the initial mean or level of the observations. Clearly, if μ is not included in the model, the first observation is $X_1 = \varepsilon_1$ with mean zero and variance σ^2. The lag length builds up as more observations become available.

As an example we take (an updated version of) the Gallup poll data from Byers *et al.* (1997) to be analyzed in Section 4. The data is monthly from January 1951 to November

2000 for a total of 599 observations. In this case the data is not available for all t simply because the Labour party was founded in 1900, and the Gallup company was founded in 1935, and in fact the regular Gallup polls only started in January 1951, which is denoted $-N_0 + 1$.

As a second example, consider the paper by Andersen, Bollerslev, Diebold, and Ebens (2001) which analyzes log realized volatility for companies in the Dow Jones Industrial Average from January 2, 1993, to May 28, 1998. For each of these companies there is an earlier date, which we call $-N_0 + 1$, where the company became publicly traded and such measurements were made for the first time. The data analyzed in Andersen *et al.* (2001) was not from $-N_0+1$, but only from the later date when the data became available on CD-ROM, which was January 2, 1993, which we denote $t = 1$. We thus do not have observations from $-N_0 + 1$ to 0.

We summarize this in the following display, which we think is representative for most, if not all, data in economics:

$$\underbrace{\ldots, X_{-N_0}}_{\text{Data does not exist}}, \quad \underbrace{X_{-N_0+1}, \ldots, X_0}_{\substack{\text{Data exists}\\ \text{but is not observed}}}, \quad \underbrace{X_1, \ldots, X_N}_{\substack{\text{Data is observed}\\ \text{(initial values)}}}, \quad \underbrace{X_{N+1}, \ldots, X_{T_0}}_{\substack{\text{Data is observed}\\ \text{(estimation)}}} \tag{6}$$

Thus, we consider estimation of

$$\Delta_0^d (X_t - \mu) = \varepsilon_t, \; t = 1, \ldots, T_0, \tag{7}$$

as an approximation to model (4). Unlike for (4), the conditional likelihood for (7) can be calculated based on available data from 1 to T_0. For a fast algorithm to calculate the fractional difference, see Jensen and Nielsen (2014).

In summary, we use $\Delta_{-N_0}^d (X_t - \mu) = \varepsilon_t$ as the model we would like to analyze. However, because we only have data for $t = 1, \ldots, T_0$, we base the likelihood on the model $\Delta_0^d (X_t - \mu) = \varepsilon_t$, for which an approximation to the conditional likelihood from (4) can be calculated with the available data. We then try to mitigate the effect of the unobserved initial values by conditioning on $X_1, \ldots, X_N$.

2.3 The conditional fractional model

Let parameter subscript zero denote true values. In the conditional approach we interpret equation (4) as a model for X_t given the past $\mathcal{F}_{t-1} = \sigma(X_{-N_0+1}, \ldots, X_{t-1})$ and therefore solve the equation for X_t as a function of initial values, errors, and the initial level, μ_0. The solution to (1) is given in JN (2010, Lemma 1) under the assumption of bounded initial values, and we give here the solution of (4).

Lemma 1 *The solution of model (4) for $X_{N+1}, \ldots, X_{T_0}$, conditional on initial values $X_n, -N_0 < n \le N$, is, for $t = N + 1, \ldots, T_0$, given by*

$$X_t = \Delta_N^{-d_0} \varepsilon_t - \Delta_N^{-d_0} \sum_{n=t-N}^{t+N_0-1} \pi_n(-d_0) X_{t-n} + \Delta_N^{-d_0} \pi_{t+N_0-1}(-d_0 + 1)\mu_0. \tag{8}$$

We find the conditional mean and variance by writing model (4) as $X_t - \mu = (1 - \Delta_{-N_0}^d)(X_t - \mu) + \varepsilon_t$. Because $(1 - \Delta_{-N_0}^d)(X_t - \mu)$ is a function only of the past, we find

$$E(X_t - \mu | \mathcal{F}_{t-1}) = (1 - \Delta_{-N_0}^d)(X_t - \mu) \text{ and } Var(X_t | \mathcal{F}_{t-1}) = Var(\varepsilon_t) = \sigma^2.$$

As an example we get, for $d = 1$ and $\mu = 0$, the well-known result from the autoregressive model that $E(X_t|\mathcal{F}_{t-1}) = X_{t-1}$ and $Var(X_t|\mathcal{F}_{t-1}) = \sigma^2$. In model (4) this implies that the prediction error decomposition given $X_n, -N_0 < n \leq N$, is the conditional sum of squares,

$$\sum_{t=N+1}^{T_0} \frac{(X_t - E(X_t|\mathcal{F}_{t-1}))^2}{Var(X_t|\mathcal{F}_{t-1})} = \sigma^{-2} \sum_{t=N+1}^{T_0} (\Delta_{-N_0}^d (X_t - \mu))^2,$$

which is used in the conditional Gaussian likelihood function (9) below.

2.4 Estimation of the conditional fractional model

We would like to consider the conditional (Gaussian) likelihood of $\{X_t, N+1 \leq t \leq T_0\}$ given initial values $\{X_n, -N_0 + 1 \leq n \leq N\}$, which is given by

$$-\frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=N+1}^{T_0} (\Delta_{-N_0}^d (X_t - \mu))^2. \tag{9}$$

If in fact we have observed all available data, such that $N_0 = 0$ as in, e.g., the Gallup poll data we can use (9) for $N_0 = 0$. More commonly, however, data is not available all the way back to inception at time $-N_0 + 1$, so we consider the situation that the series exists for $t > -N_0$, but we only have observations for $t \geq 1$, as in the volatility data example. We therefore replace the truncated filter $\Delta_{-N_0}^d$ by Δ_0^d and suggest using the (quasi) likelihood conditional on $\{X_n, 1 \leq n \leq N\}$,

$$L(d, \mu, \sigma^2) = -\frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=N+1}^{T_0} (\Delta_0^d (X_t - \mu))^2. \tag{10}$$

That is, (10) is an approximation to the conditional likelihood (9), where (10) has the advantage that it can be calculated based on available data from $t = 1$ to $T_0 = N + T$. It is clear from (10) that we can equivalently find the (quasi) maximum likelihood estimators of d and μ by minimizing

$$L(d, \mu) = \frac{1}{2}\sum_{t=N+1}^{T_0} (\Delta_0^d (X_t - \mu))^2 \tag{11}$$

with respect to d and μ.

We find from (46) in Lemma A.4 that

$$\Delta_0^d (X_t - \mu) = \Delta_0^d X_t - \sum_{n=0}^{t-1} \pi_n(-d)\mu = \Delta_0^d X_t - \pi_{t-1}(-d+1)\mu = \Delta_0^d X_t - \kappa_{0t}(d)\mu,$$

where we have introduced $\kappa_{0t}(d) = \pi_{t-1}(-d+1)$. The estimator of μ for fixed d is

$$\hat{\mu}(d) = \frac{\sum_{t=N+1}^{T_0} (\Delta_0^d X_t)\kappa_{0t}(d)}{\sum_{t=N+1}^{T_0} \kappa_{0t}(d)^2},$$

provided $\sum_{t=N+1}^{T_0} \kappa_{0t}(d)^2 > 0$. The conditional quasi-maximum likelihood estimator of d can then be found by minimizing the concentrated objective function

$$L^*(d) = \frac{1}{2}\sum_{t=N+1}^{T_0} (\Delta_0^d X_t)^2 - \frac{1}{2}\frac{(\sum_{t=N+1}^{T_0} (\Delta_0^d X_t)\kappa_{0t}(d))^2}{\sum_{t=N+1}^{T_0} \kappa_{0t}(d)^2}, \tag{12}$$

which has no singularities at the points where $\sum_{t=N+1}^{T_0} \kappa_{0t}(d)^2 = 0$, see Theorem 1. Thus, the conditional quasi-maximum likelihood estimator $\hat{d}$ can be defined by

$$\hat{d} = \arg\min_{d \in \mathcal{D}} L^*(d) \tag{13}$$

for a parameter space $\mathcal{D}$ to be defined below.

This is a type of conditional-sum-of-squares (CSS) estimator for d. The first term of (12) is standard, and the second takes into account the estimation of the unknown initial level μ at the inception of the series at time $-N_0 + 1$.

For $d = d_0$ and $\mu = \mu_0$ we find, provided $\sum_{t=N+1}^{T_0} \kappa_{0t}(d_0)^2 > 0$, that

$$\hat{\mu}(d_0) - \mu_0 = \frac{\sum_{t=N+1}^{T_0} \varepsilon_t \kappa_{0t}(d_0)}{\sum_{t=N+1}^{T_0} \kappa_{0t}(d_0)^2},$$

which has mean zero and variance $\sigma_0^2(\sum_{t=N+1}^{T_0} \kappa_{0t}(d_0)^2)^{-1}$ that does not go to zero when $d_0 > 1/2$ because then $\sigma_0^{-2} \sum_{t=N+1}^{T_0} \kappa_{0t}(d_0)^2$ is bounded in T_0, see (59) in Lemma B.1. Thus we have that, even if $d = d_0$, $\hat{\mu}(d_0)$ is not consistent.

In the following we also analyze another estimator, $\hat{d}_c$, constructed by choosing to center the observations by a known value rather than estimating μ as above. The known value, say C, used for centering, could be one of the observed initial values, e.g. the first one, or an average of these, or it could be any known constant. This can be formulated as choosing $\mu = C$ in the likelihood function (10) and defining

$$\hat{d}_c = \arg\min_{d \in \mathcal{D}} L_c^*(d), \tag{14}$$

$$L_c^*(d) = \frac{1}{2} \sum_{t=N+1}^{T_0} (\Delta_0^d(X_t - C))^2, \tag{15}$$

which is also a CSS estimator. A commonly applied estimator is the one obtained by not centering the observations, i.e. by setting $C = 0$. In that case, an initial non-zero level of the process is therefore not taken into account.

The introduction of centering and of the parameter μ, interpreted as the initial level of the process, thus allows analysis of the effects of centering the observations in different ways (and avoid the, possibly unrealistic, phenomenon described immediately after (5) when $\mu = 0$). We analyze the conditional maximum likelihood estimator, $\hat{d}$, where the initial level is estimated by maximum likelihood jointly with the fractional parameter, and we also analyze the more traditional CSS estimator, $\hat{d}_c$, where the initial level is "estimated" using a known value C, e.g. zero or the first available observation, X_1.

In practice we split a given sample of size $T_0 = N + T$ into (observed) initial values $\{X_n\}_{n=1}^N$ and observations $\{X_t\}_{t=N+1}^{N+T}$ to be modeled, and then calculate the likelihood (12) based on the truncated filter Δ_0^d as an approximation to the model (4) starting at $-N_0 + 1$. In order to discuss the error implied by using this choice in the likelihood function, we derive in Theorem 2 a computable expression for the asymptotic second-order bias term in the estimator of d via a higher-order stochastic expansion of the estimator. This bias term depends on all observed and unobserved initial values and the parameters. In Corollary 1 and Theorems 3 and 4 we further investigate the effect on the bias of setting aside the data from $t = 1$ to N as initial values.

2.5 A relation to the ARFIMA model

The simple model (1) is a special case of the well-known ARFIMA model,

$$A(L)\Delta^d X_t = B(L)\varepsilon_t,\ t = 1, \ldots, T,$$

where $A(L)$ and $B(L)$ depend on a parameter vector ψ and $B(z) \neq 0$ for $|z| \leq 1$. For this model, the conditional likelihood depends on the residuals

$$\varepsilon_t(d, \psi) = B(L)^{-1} A(L)\Delta^d X_t = b(\psi, L)\Delta^d X_t,$$

and when $b(\psi, L) = 1$ we obtain model (1) as a special case.

For the ARFIMA model the analysis would depend on the derivatives of the conditional likelihood function, which would in turn be functions of the derivatives of the residuals. Again, to focus on estimation of d we consider the remaining parameter ψ fixed at the true value ψ_0. For a function $f(d)$ we denote the derivative of f with respect to d as $\mathsf{D}f(d) = \frac{\partial}{\partial d} f(d)$ (Euler's notation), and the relevant derivatives are

$$\mathsf{D}^m \varepsilon_t(d, \psi)|_{d_0, \psi_0} = b(\psi_0, L)\mathsf{D}^m \Delta^d X_t|_{d_0} = (\log \Delta)^m b(\psi_0, L)\Delta^{d_0} X_t = (\log \Delta)^m \varepsilon_t.$$

Thus, for this more general model, the derivatives of the conditional likelihood with respect to d, when evaluated at the true values, are identical to those of the residuals from the simpler model (1). We can therefore apply the results from the simpler model more generally, but only if we know the parameter ψ_0. If ψ has to be estimated, the analysis becomes much more complicated. We therefore focus our analysis on the simple model.

3 Main results

Our main results hold only for the true value $d_0 > 1/2$, that is, for nonstationary processes, which is therefore assumed in the remainder of the paper. However, we maintain a large compact parameter set $\mathcal{D}$ for d in the statistical model, which does not assume *a priori* knowledge that $d_0 > 1/2$, see Assumption 2.

3.1 First-order asymptotic properties

The first-order asymptotic properties of the CSS estimators $\hat{d}$ and $\hat{d}_c$ derived from the likelihood functions $L^*(d)$ and $L_c^*(d)$ in (12) and (15), respectively, are given in the following theorem, based on results of JN (2012a) and Nielsen (2015). To describe the results, we use Riemann's zeta function, $\zeta_s = \sum_{j=1}^{\infty} j^{-s}$, $s > 1$, and specifically

$$\zeta_2 = \sum_{j=1}^{\infty} j^{-2} = \frac{\pi^2}{6} \simeq 1.6449 \text{ and } \zeta_3 = \sum_{j=1}^{\infty} j^{-3} \simeq 1.2021. \tag{16}$$

We formulate two assumptions that will be used throughout.

Assumption 1 *The errors ε_t are i.i.d.$(0, \sigma_0^2)$ with finite fourth moment.*

Assumption 2 *The parameter set for (d, μ, σ^2) is $\mathcal{D} \times \mathbb{R} \times \mathbb{R}_+$, where $\mathcal{D} = [\underline{d}, \bar{d}]$, $0 < \underline{d} < \bar{d} < \infty$. The true value is (d_0, μ_0, σ_0^2), where $d_0 > 1/2$ is in the interior of $\mathcal{D}$.*

Theorem 1 *Let the model for the data $X_t, t = 1, \ldots, N + T$, be given by (4) and let Assumptions 1 and 2 be satisfied. Then the functions $L^*(d)$ in (12) and $L_c^*(d)$ in (15) have no singularities for $d > 0$, and the estimators $\hat{d}$ and $\hat{d}_c$ derived from $L^*(d)$ and $L_c^*(d)$, respectively, are both $\sqrt{T}$-consistent and asymptotically distributed as $\mathcal{N}(0, \zeta_2^{-1})$.*

3.2 Higher-order expansions and asymptotic bias

To analyze the asymptotic bias of the CSS estimators for d, and in particular how initial values influence the bias, we need to examine higher-order terms in a stochastic expansion of the estimators, see Lawley (1956). The conditional (negative profile log) likelihoods $L^*(d)$ and $L_c^*(d)$ are given in (12) and (15). We find, see Lemma B.4, that the derivatives satisfy $\mathsf{D}L^*(d_0) = O_P(T^{1/2})$, $\mathsf{D}^2L^*(d_0) = O_P(T)$, and $\mathsf{D}^3L^*(d) = O_P(T)$ uniformly in a neighborhood of d_0, and a Taylor series expansion of $\mathsf{D}L^*(\hat{d}) = 0$ around d_0 gives

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

where d^* is an intermediate value satisfying $|d^* - d_0| \leq |\hat{d} - d_0| \overset{P}{\to} 0$. We then insert $\hat{d} - d_0 = T^{-1/2}\tilde{G}_{1T} + T^{-1}\tilde{G}_{2T} + O_P(T^{-3/2})$ and find $\tilde{G}_{1T} = -T^{1/2}\mathsf{D}L^*(d_0)/\mathsf{D}^2L^*(d_0)$ and $\tilde{G}_{2T} = -\frac{1}{2}T(\mathsf{D}L^*(d_0))^2\mathsf{D}^3L^*(d^*)/(\mathsf{D}^2L^*(d_0))^3$, which we write as

$$T^{1/2}(\hat{d} - d_0) = -\frac{T^{-1/2}\mathsf{D}L^*(d_0)}{T^{-1}\mathsf{D}^2L^*(d_0)} - \frac{1}{2}T^{-1/2}(\frac{T^{-1/2}\mathsf{D}L^*(d_0)}{T^{-1}\mathsf{D}^2L^*(d_0)})^2\frac{T^{-1}\mathsf{D}^3L^*(d^*)}{T^{-1}\mathsf{D}^2L^*(d_0)} + O_P(T^{-1}). \quad (17)$$

Based on this expansion, we find another expansion $T^{1/2}(\hat{d} - d_0) = G_{1T} + T^{-1/2}G_{2T} + o_P(T^{-1/2})$ with the property that $(G_{1T}, G_{2T}) \overset{D}{\to} (G_1, G_2)$ and $E(G_{1T}) = E(G_1) = 0$. Then the zero- and first-order terms of the bias are zero, and the second-order asymptotic bias term is defined as $T^{-1}E(G_2)$.

We next present the main result on the asymptotic bias of $\hat{d}$. In order to formulate the results, we define some coefficients that depend on N, N_0, T, and on initial values and (μ_0, σ_0^2, d) (we suppress some of these dependencies for notational convenience),

$$\eta_{0t}(d) = -\sum_{n=-N_0+1}^{0} \pi_{t-n}(-d)(X_n - \mu_0), \quad (18)$$

$$\eta_{1t}(d) = \sum_{k=1}^{t-1-N} k^{-1} \sum_{n=-N_0+1}^{N} \pi_{t-n-k}(-d)(X_n - \mu_0) - \sum_{n=1}^{N} \mathsf{D}\pi_{t-n}(-d)(X_n - \mu_0), \quad (19)$$

$$\kappa_{0t}(d) = \pi_{t-1}(-d+1), \text{ and } \kappa_{1t}(d) = -\mathsf{D}\pi_{t-1}(-d+1). \quad (20)$$

For two sequences $\{u_t, v_t\}_{t=1}^{\infty}$, we define the product moment $\langle u, v\rangle_T = \sigma_0^{-2}\sum_{t=N+1}^{T_0} u_t v_t$, see e.g. Lemma B.1. The main contributions to the bias are expressed for $d = d_0$ in terms of

$$\xi_{N,T}(d) = \langle\eta_0, \eta_1\rangle_T - \frac{\langle\eta_0, \kappa_0\rangle_T}{\langle\kappa_0, \kappa_0\rangle_T}(\langle\eta_0, \kappa_1\rangle_T + \langle\eta_1, \kappa_0\rangle_T) + \frac{\langle\eta_0, \kappa_0\rangle_T^2}{\langle\kappa_0, \kappa_0\rangle_T^2}\langle\kappa_1, \kappa_0\rangle_T, \quad (21)$$

$$\xi_{N,T}^C(d) = \langle\eta_0, \eta_1\rangle_T - (C - \mu_0)(\langle\eta_0, \kappa_1\rangle_T + \langle\eta_1, \kappa_0\rangle_T) + (C - \mu_0)^2\langle\kappa_1, \kappa_0\rangle_T, \quad (22)$$

$$\tau_{N,T}(d) = \sigma_0^{-2}\sum_{N\leq s<t\leq N+T-1} (t-s)^{-1}\pi_t(-d+1)\pi_s(-d+1)/\langle\kappa_0, \kappa_0\rangle_T. \quad (23)$$

Note that (21)–(23) are all invariant to scale because of the normalization by σ_0^2. Also note that, even if $\langle\kappa_0, \kappa_0\rangle_T = 0$, the ratio $\langle\eta_0, \kappa_0\rangle_T/\langle\kappa_0, \kappa_0\rangle_T$ as well as $\tau_{N,T}(d)$ are well defined, see Theorem 1 and Appendix C.1.

Theorem 2 *Let the model for the data $X_t, t = 1, \ldots, N+T$, be given by (4) and let Assumptions 1 and 2 be satisfied. Then the asymptotic biases of $\hat{d}$ and $\hat{d}_c$ are*

$$bias(\hat{d}) = -(T\zeta_2)^{-1}[3\zeta_3\zeta_2^{-1} + \xi_{N,T}(d_0) + \tau_{N,T}(d_0)] + o(T^{-1}), \tag{24}$$

$$bias(\hat{d}_c) = -(T\zeta_2)^{-1}[3\zeta_3\zeta_2^{-1} + \xi_{N,T}^C(d_0)] + o(T^{-1}), \tag{25}$$

where $\lim_{T\to\infty} |\xi_{N,T}(d_0)| < \infty$, $\lim_{T\to\infty} |\tau_{N,T}(d_0)| < \infty$, *and* $\lim_{T\to\infty} |\xi_{N,T}^C(d_0)| < \infty$.

The leading bias terms in (24) and (25) are of the same order of magnitude in T, namely $O(T^{-1})$. First, the fixed term, $3\zeta_3\zeta_2^{-2}$, derives from correlations of derivatives of the likelihood and does not depend on initial values or d_0. The second term in (24), $\xi_{N,T}(d_0)$, is a function of initial values and d_0, and can be made smaller by including more initial values (larger N) as shown in Corollary 1 below. The third term in (24), $\tau_{N,T}(d_0)$, only depends on (N, T, d_0). If we center the data by C, and do not correct for μ, we get the term $\xi_{N,T}^C(d_0)$ in (25). However, if we estimate μ we get $\xi_{N,T}(d_0) + \tau_{N,T}(d_0)$ in (24), where $\tau_{N,T}(d_0)$ is independent of initial values and only depends on (N, T, d_0). The coefficients $\eta_{0t}(d)$ and $\eta_{1t}(d)$ are linear in the initial values, and hence the bias terms $\xi_{N,T}(d)$ and $\xi_{N,T}^C(d)$ are quadratic in initial values scaled by σ_0.

The fixed bias term, $3\zeta_3\zeta_2^{-2}$, is the same as the bias derived by Lieberman and Phillips (2004) for the estimator $\hat{d}_{\text{stat}}$, based on the unconditional likelihood (3) in the stationary case, $0 < d_0 < 1/2$. They showed that the distribution function of $\zeta_2^{1/2}T^{1/2}(\hat{d}_{\text{stat}} - d_0)$ is

$$F_T(x) = P(\zeta_2^{1/2}T^{1/2}(\hat{d}_{\text{stat}} - d_0) \le x) = \Phi(x) + T^{-1/2}\zeta_3\zeta_2^{-3/2}\phi(x)(2 + x^2) + O(T^{-1}),$$

where $\Phi(x)$ and $\phi(x)$ denote the standard normal distribution and density functions, respectively. Using $\mathsf{D}\phi(x)(2 + x^2) = -\phi(x)x^3$, we find that an approximation to the expectation of $\zeta_2^{1/2}T^{1/2}(\hat{d}_{\text{stat}} - d_0)$, based on the first two terms, is given by

$$-T^{-1/2}\zeta_3\zeta_2^{-3/2}\int x\phi(x)x^3 dx = -T^{-1/2}3\zeta_3\zeta_2^{-3/2},$$

which shows that the second-order bias of $\hat{d}_{\text{stat}}$, derived for $0 < d_0 < 1/2$, is the same as the the second-order fixed bias term of $\hat{d}$ derived for $d_0 > 1/2$ in Theorem 2.

The dependence of the bias in Theorem 2 on the number of observed initial values, N, is explored in the following corollary.

Corollary 1 *Under the assumptions of Theorem 2, we obtain the following bounds for the components of the bias terms for $\hat{d}$ and $\hat{d}_c$ when $d > 1/2$,*

$$\max(|\xi_{N,T}^C(d)|, |\xi_{N,T}(d)|) \le c(1+N)^{-\min(d,2d-1)+\epsilon} \text{ for any } 0 < \epsilon < \min(d, 2d-1). \tag{26}$$

The result in Corollary 1 shows how the bias term arising from not observing all initial values decays as a function of the number of observed values set aside as initial values, N.

More generally, the results in this section shows that a partial bias correction is possible. That is, by adding the terms $(T\zeta_2)^{-1}3\zeta_3\zeta_2^{-1}$ and $(T\zeta_2)^{-1}\tau_{N,T}(\hat{d})$, the second-order bias in $\hat{d}$ and $\hat{d}_c$ can be partly eliminated, but the bias due to $(T\zeta_2)^{-1}\xi_{N,T}(d_0)$ can only be made smaller by increasing N.

A different type of bias correction was used by Davidson and Hashimzade (2009, eqn. (4.4)) in an analysis of the Nile data. They considered the CSS estimator when all initial values are set to zero in the stationary case. To capture the effect of the left-out initial values, they introduce a few extra regressors that are found as the first principal components of the variance matrix of the $n = 150$ variables $x^{**} = \{\sum_{k=s}^{\infty} \pi_k(-d) X_{s-k}\}_{s=1}^{n}$.

3.3 Further results for special cases

The expressions for $\xi_{N,T}(d)$, $\xi_{N,T}^C(d)$, and $\tau_{N,T}(d)$ in (21)–(23) show that they depend on (N, T, d) and, in the case of $\xi_{N,T}(d)$ and $\xi_{N,T}^C(d)$, also on all initial values. In order to get an impression of this dependence, we derive simple expressions for various special cases.

First, when d is an integer, we find simple results for $\xi_{N,T}(d)$, $\xi_{N,T}^C(d)$, and $\tau_{N,T}(d)$, and hence the asymptotic bias, as follows.

Theorem 3 *Under the assumptions of Theorem 2 it holds that $\xi_{N,T}^C(d) = \xi_{N,T}(d) = 0$ in the following two cases:*
(i) If $d = k$ for an integer k such that $1 \le k \le N$,
(ii) If $d = 1$ and $N \ge 0$.
In either case, the asymptotic biases of $\hat{d}$ and $\hat{d}_c$ are given by

$$bias(\hat{d}) = -(T\zeta_2)^{-1}(3\zeta_3\zeta_2^{-1} + \tau_{N,T}(d_0)) + o(T^{-1}),$$
$$bias(\hat{d}_c) = -(T\zeta_2)^{-1}3\zeta_3\zeta_2^{-1} + o(T^{-1}).$$

(iii) If $d_0 = N+1$ then $\tau_{N,T}(d_0) = 0$ and $bias(\hat{d}) = -(T\zeta_2)^{-1}(3\zeta_3\zeta_2^{-1} + \xi_{N,T}(N+1)) + o(T^{-1})$.

It follows from Theorem 3(i) that for $d = 1$ we need one initial value ($N \ge 1$) and for $d = 2$ we need two initial values ($N \ge 2$), etc., to obtain $\xi_{N,T}^C(d) = \xi_{N,T}(d) = 0$. Alternatively, for $d_0 = 1$, Theorem 3(ii) shows that there will be no contribution from initial values to the second-order asymptotic bias even if $N = 0$, and Theorem 3(iii) shows that when $N = 0$, it also holds that $\tau_{0,T}(1) = 0$ such that $bias(\hat{d}) = -(T\zeta_2)^{-1}3\zeta_3\zeta_2^{-1} + o(T^{-1})$. Since the bias term is continuous in d_0, the same is approximately true for a (small) neighborhood of $d_0 = 1$.

Note that the results in Theorem 3 show that in the cases considered, the estimators $\hat{d}$ and $\hat{d}_c$ can be bias corrected to have second-order bias equal to zero.

We finally consider the special case with $N_0 = 0$, where all available data is observed. We use the notation $\Psi(d) = \mathsf{D}\log\Gamma(d)$ to denote the Digamma function.

Theorem 4 *If $N_0 = 0$ and $N \ge 0$ then $\xi_{N,T}(d_0) = 0$ and the biases of $\hat{d}$ and $\hat{d}_c$ are given by*

$$bias(\hat{d}) = -(T\zeta_2)^{-1}[3\zeta_3\zeta_2^{-1} + \tau_{N,T}(d_0)] + o(T^{-1}), \tag{27}$$
$$bias(\hat{d}_c) = -(T\zeta_2)^{-1}[3\zeta_3\zeta_2^{-1} + \xi_{N,T}^C(d_0)] + o(T^{-1}), \tag{28}$$

where $\tau_{N,T}(d_0)$ is defined in (23) and $\xi_{N,T}^C(d_0)$ simplifies to

$$\xi_{N,T}^C(d_0) = -(C - \mu_0)\langle\kappa_0, \eta_1\rangle_T + (C - \mu_0)^2\langle\kappa_0, \kappa_1\rangle_T. \tag{29}$$

In particular, for $N_0 = N = 0$ we get the analytical expressions

$$bias(\hat{d}) = -(T\zeta_2)^{-1}[3\zeta_3\zeta_2^{-1} - (\Psi(2d_0 - 1) - \Psi(d_0))] + o(T^{-1}), \tag{30}$$
$$bias(\hat{d}_c) = -(T\zeta_2)^{-1}[3\zeta_3\zeta_2^{-1} - \frac{(C - \mu_0)^2}{\sigma_0^2}\binom{2d_0 - 2}{d_0 - 1}(\Psi(2d_0 - 1) - \Psi(d_0))] + o(T^{-1}). \tag{31}$$

It follows from Theorem 4 that if we have observed all possible data, that is $N_0 = 0$, then we get a bias of $\hat{d}$ in (27) and of $\hat{d}_c$ in (28) and (29). The bias of $\hat{d}$ comes from the estimation of μ and the bias of $\hat{d}_c$ depends on the distance $C - \mu_0$.

With $N_0 = 0$ as in Theorem 4, we note that the biases of $\hat{d}$ and $\hat{d}_c$ do not depend on unobserved initial values. It follows that (27) can be used to bias correct the estimator $\hat{d}$ and (28) to bias correct the estimator $\hat{d}_c$. For $\hat{d}$ this bias correction gives a second-order bias of zero, but for $\hat{d}_c$ the correction is only partial due to (29).

Although the asymptotic bias of $\hat{d}$ is of order $O(T^{-1})$, we note that the asymptotic standard deviation of $\hat{d}$ is $(T\zeta_2)^{-1/2}$, see Theorem 1. That is, for testing purposes or for calculating confidence intervals for d_0, the relevant quantity is in fact the bias relative to the asymptotic standard deviation, and this is of order $O(T^{-1/2})$. To quantify the distortion of the quantiles (critical values), we therefore focus on the magnitude of the relative bias, for which we obtain the following corollary by tabulation.

Corollary 2 *Letting $T_0 = N + T$ be fixed and tabulating the relative bias,*

$$(T\zeta_2)^{1/2} bias(\hat{d}) = -((T_0 - N)\zeta_2)^{-1/2}[3\zeta_3\zeta_2^{-1} + \tau_{N,T_0-N}(d_0)],$$

see (27), for $N = 0, \ldots, T_0 - 2$ and $d_0 > 1/2$, the minimum value is attained for $N = 0$. Thus, we achieve the smallest relative (and also absolute) bias of $\hat{d}$ by choosing $N = 0$.

4 Application to Gallup opinion poll data

As an application and illustration of the results, we consider the monthly Gallup opinion poll data on support for the Conservative and Labour parties in the United Kingdom. They cover the period from January 1951 to November 2000, for a total of 599 months. The two series have been logistically transformed, so that, if Y_t denotes an observation on the original series, it is mapped into $X_t = \log(Y_t/(100 - Y_t))$. A shorter version of this data set was analyzed by Byers *et al.* (1997) and Dolado *et al.* (2002), among others.

Using an aggregation argument and a model of voter behavior, Byers *et al.* (1997) show that aggregate opinion poll data may be best modeled using fractional time series methods. The basic finding of Byers *et al.* (1997) and Dolado *et al.* (2002) is that the ARFIMA(0,d,0) model, i.e. model (1), appears to fit both data series well and they obtain values of the integration parameter d in the range of 0.6–0.8.

4.1 Analysis of the voting data

In light of the discussion in Section 2.2, we work throughout under the assumption that X_t was not observed prior to January 1951 because the data series did not exist, and we truncate the filter correspondingly, i.e., we consider model (4). Because we observe all available data, we estimate d (and μ, σ) by the estimator $\hat{d}$ setting $N = N_0 = 0$ and take $T = 599$ following Theorem 4 and Corollary 2.

The results are presented in Table 1. Since we have assumed that $N = N_0 = 0$, we can bias correct the estimator using (30) in Theorem 4, and the resulting estimate is reported in Table 1 as $\hat{d}_{\text{bc}}$. Two conclusions emerge from the table. First, the estimates of d (and σ) are quite similar for the two data series, but the estimates of μ are quite different. Second, the bias corrections to the estimates are small. More generally, the estimates obtained in Table 1 are in line with those from the literature cited above.

Table 1: Estimation results for Gallup opinion poll data

	Conservative	Labour
$\hat{d}$	0.7718	0.6940
$\hat{d}_{\text{bc}}$	0.7721	0.6914
$\hat{\mu}$	0.0097	−0.4313
$\hat{\sigma}$	0.1098	0.1212

Note: The table presents parameter estimates for the Gallup opinion poll data with $T = 599$ and $N_0 = N = 0$. The subscript 'bc' denotes the bias corrected estimator, c.f. (30). The asymptotic standard deviation of $\hat{d}$ is given in Theorem 1 as $(T\pi^2/6)^{-1/2} \simeq 0.032$.

4.2 An experiment with unobserved initial values

We next use this data to conduct a small experiment with the purpose of investigating how the choice of N influences the bias of the estimators of d, if there were unobserved initial values. For this purpose, we assume that the econometrician only observes data starting in January 1961. That is, January 1951 through December 1960 are $N_0 = 120$ unobserved initial values. We then split the given sample of $T_0 = 479$ observations into initial values (N) and observations used for estimation (T), such that $N + T = 479$. We can now ask the questions (i) what is the consequence in terms of bias of ignoring initial values, i.e. of setting $N = 0$, and (ii) how sensitive is the bias to the choice of N for this particular data set.

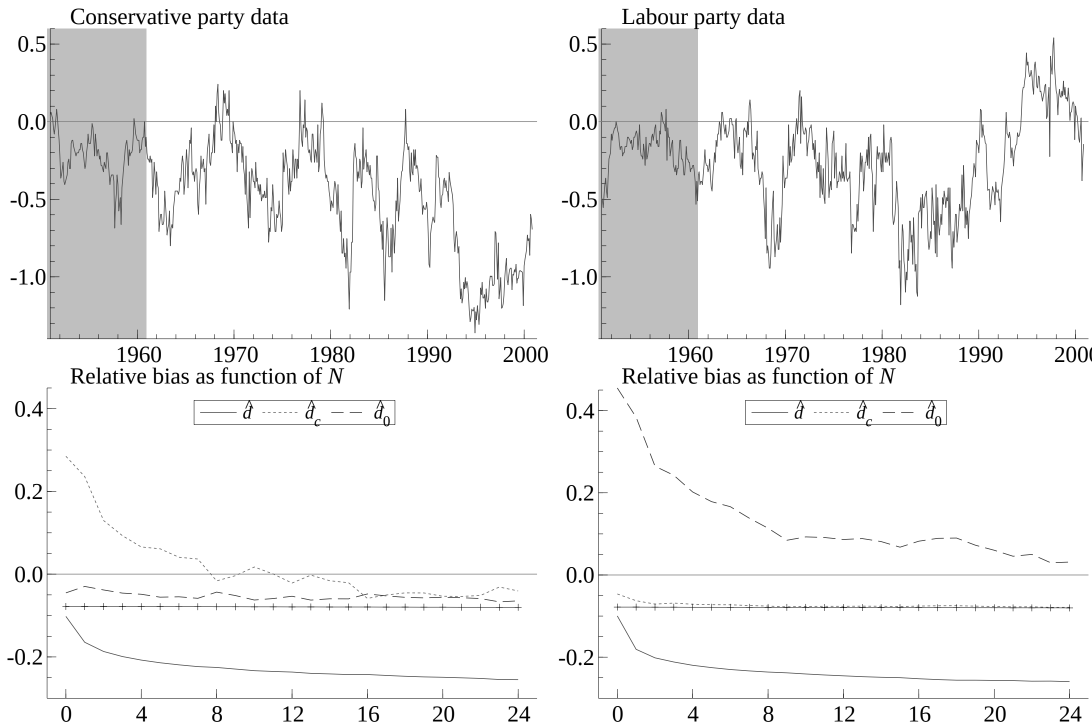
To answer these questions we apply (24) and (25) from Theorem 2. We note that $\xi_{N,T}(d)$ and $\xi^C_{N,T}(d)$ depend on the unobserved initial values, i.e. on $X_n, -N_0 < n \leq 0$, which in this example are the 120 observations from January 1951 to December 1960. To apply Theorem 2 we need (estimates of) d_0, μ_0, σ_0. For this purpose we use $(\hat{d}_{\text{bc}}, \hat{\mu}, \hat{\sigma})$ from Table 1.

The results are shown in Figure 1. The top panels show the logistically transformed opinion poll data for the Conservative (left) and Labour (right) parties. The shaded areas mark the unobserved initial values January 1951 to December 1960. The bottom panels show the relative bias in the estimators of d as a function of $N \in [0, 24]$, and the starred straight line denotes the value of the fixed (relative) bias term, $-(T_0 - N)^{-1/2} 3\zeta_3\zeta_2^{-3/2}$. The estimators are $\hat{d}$ in (13) and $\hat{d}_c$ in (14) either with C chosen as the average of the T_0 observations, denoted $\hat{d}_c$ in the graph, or with $C = 0$, denoted $\hat{d}_0$ in the graph. That is, for $\hat{d}_0$ the series have not been centered, and for $\hat{d}_c$ the series have been centered by the average of the T_0 observed values. The latter two estimators are the usual CSS estimators with and without centering of the series.

In Figure 1 we note that the relative bias of $\hat{d}_0$ is larger for the Labour party series because the last unobserved initial values are larger in absolute value than those of the Conservative party series. In particular, if one does not condition on initial values and uses $N = 0$, the relative bias of $\hat{d}_0$ is 0.45 for the Labour party series and -0.05 for the Conservative party series. It is clear from the figure that the relative bias of $\hat{d}_0$ for the Labour party series can be reduced substantially and be made much closer to the fixed bias value by conditioning on just a few initial values. The same conclusions can be drawn for $\hat{d}_c$ but reversing the roles of the two series. The reason is that, after centering the series by the average of the T_0 observations, it is now for the Conservative party series that the last unobserved initial values are different from zero, while those of the Labour party series are close to zero.

Finally, for $\hat{d}$, where the initial level or centering parameter, μ, is estimated jointly with

Figure 1: Application to Gallup opinion poll data



Note: The top panels show (logistically transformed) opinion poll time series and the bottom panels show the relative bias for three estimators of d as a function of the number of chosen initial values, N, when the first $N_0 = 120$ observations have been reserved as unobserved initial values (shaded area). The estimators are $\hat{d}$ in (13) and $\hat{d}_c$ in (14) either with C chosen as the average of the T_0 observations, denoted $\hat{d}_c$ in the graph, or with $C = 0$, denoted $\hat{d}_0$ in the graph. The starred line denotes the fixed (relative) bias, $-(T_0 - N)^{-1/2} 3\zeta_3\zeta_2^{-3/2}$.

d, we find that the relative bias is increasing in N. The reason for this is that $\tau_{N,T}(d)$ dominates $\xi_{N,T}(d)$, at least for this particular data series. With $N = 0$ the relative bias is very small and the estimator $\hat{d}$ is better than the other two estimators.

5 Conclusion

In this paper we have analyzed the effect of unobserved initial values on the asymptotic bias of the CSS estimators, $\hat{d}$ and $\hat{d}_c$, of the fractional parameter in a simple fractional model, for $d_0 > 1/2$. We assume that we have data X_t for $t = 1, \ldots, T_0 = N + T$, and model X_t by the truncated filter $\Delta_{-N_0}^{d_0}(X_t - \mu_0) = \varepsilon_t$ for $t = 1, \ldots, T_0$ and $N_0 \geq 0$. We derive estimators from the models $\Delta_0^d(X_t - \mu) = \varepsilon_t$ or $\Delta_0^d(X_t - C) = \varepsilon_t$ by maximizing the respective conditional Gaussian likelihoods of $X_{N+1}, \ldots, X_{T_0}$ given $X_1, \ldots, X_N$.

We give in Theorem 2 an explicit formula for the second-order bias of $\hat{d}$, consisting of three terms. The first is a constant, the second, $\xi_{N,T}(d_0)$, depends on initial values and decreases with N, and the third, $\tau_{N,T}(d_0)$, does not depend on initial values. The first and third terms can thus be used in general for a (partial) bias correction. In Theorem 4 we simplify the expressions for the case when $N_0 = 0$, so that all data are observed. In this case we can completely bias correct the estimator $\hat{d}$, at least to second order. We further find that for $\hat{d}$ the smallest bias appears for the choice $N = 0$. This choice is used for the analysis of the voting data in Section 4.1 where the bias correction is also illustrated.

In Section 4.2 we illustrate the general results with unobserved initial values, again using the voting data. Here we show that, when keeping $N_0 = 120$ observations for unobserved initial values, the estimator $\hat{d}$ with $N = 0$ has the smallest bias. Thus, the idea of letting the parameter μ capture the initial level of the process eliminates the effect of the unobserved initial values, at least in this example.

Appendix A The fractional coefficients

In this section we first give some results of Karamata. Because they are well known we sometimes apply them in the remainder without special reference.

Lemma A.1 *For $m \geq 0$ and $c < \infty$,*

$$\sum_{n=1}^{N}(1+\log n)^m n^\alpha \leq c(1+\log N)^m N^{\alpha+1} \text{ if } \alpha > -1, \tag{32}$$

$$\sum_{n=N}^{\infty}(1+\log n)^m n^\alpha \leq c(1+\log N)^m N^{\alpha+1} \text{ if } \alpha < -1. \tag{33}$$

Proof. See Theorems 1.5.8–1.5.10 of Bingham, Goldie, and Teugels (1987). ∎

We next present some useful results for the fractional coefficients (2) and their derivatives.

Lemma A.2 *Define the coefficient $a_j = 1_{\{j\geq 1\}}\sum_{k=1}^{j} k^{-1}$, where $1_{\{A\}}$ denotes the indicator function for the event A. The derivatives of $\pi_j(\cdot)$ are*

$$\mathsf{D}^m \log \pi_j(u) = (-1)^{m+1}\sum_{i=0}^{j-1}\frac{1}{(i+u)^m} \text{ for } u \neq 0,-1,\ldots,-j+1 \text{ and } m \geq 1, \tag{34}$$

$$\mathsf{D}\pi_j(u) = (-1)^{-u}\frac{(-u)!(j+u-1)!}{j!} \text{ for } u = 0,-1,\ldots,-j+1 \text{ and } j \geq 2, \tag{35}$$

$$\mathsf{D}^2\pi_j(u) = 2\mathsf{D}\pi_j(u)(a_{j+u-1} - a_{-u}) \text{ for } u = 0,-1,\ldots,-j+1 \text{ and } j \geq 2. \tag{36}$$

Proof of Lemma A.2. The result (34) follows by taking derivatives in (2) for $u \neq 0,-1,\ldots,-j+1$. For $u = -i$ and $i = 0,1,\ldots,j-1$ we first define

$$P(u) = u(u+1)\ldots(u+j-1), P_k(u) = \frac{P(u)}{u+k}, P_{kl}(u) = \frac{P(u)}{(u+k)(u+l)} \text{ for } k \neq l.$$

noting that $\pi_j(u) = P(u)/j!$, see (2). We then find

$$\mathsf{D}P(u) = \sum_{0\leq k\leq j-1} P_k(u) \text{ and } \mathsf{D}^2P(u) = \sum_{0\leq k\neq l\leq j-1} P_{kl}(u),$$

which we evaluate at $u = -i$ for $i = 0,1,\ldots,j-1$. However, for such i we find $P_k(-i) = 0$ unless $k = i$ and $P_{kl}(-i) = 0$ unless $k = i$ or $l = i$. Thus,

$$\mathsf{D}P(u)|_{u=-i} = P_i(-i) = (-i)(-i+1)\ldots(-1)\times(1)(2)\ldots(j-1-i) = (-1)^i i!(j-i-1)!$$

and (35) follows because $\mathsf{D}\pi_j(u) = \mathsf{D}P(u)/j!$, see (2). Similarly (36) follows from

$$\mathsf{D}^2 P(u)|_{u=-i} = \sum_{k\neq i} P_{ki}(-i) + \sum_{l\neq i} P_{il}(-i) = 2\sum_{k\neq i} P_{ki}(-i)$$
$$= 2\sum_{k\neq i}\frac{P_i(-i)}{k-i} = 2P_i(-i)\sum_{k\neq i}\frac{1}{k-i} = 2P_i(-i)(a_{j-i-1} - a_i).$$

■

For $u = 0, -1, -2, \ldots$, we note that $\pi_j(u) = 0$ for $j \geq -u+1$, but $\mathsf{D}^m\pi_j(u)$ remains non-zero even for such values of j where $\pi_j(u) = 0$.

Lemma A.3 *Let N be an integer and assume $j \geq N$, then*

$$\pi_j(u) = \prod_{i=1}^{j}\frac{i+u-1}{i} = \pi_N(u)\prod_{i=N+1}^{j}(1+(u-1)/i) = \pi_N(u)\alpha_{N,j}(u) \tag{37}$$

with $\alpha_{N,j}(u) = \prod_{i=N+1}^{j}(1+(u-1)/i)$ for $j > N$ and $\alpha_{N,j}(u) = 1$ for $j = N$.

For $m \geq 0$ and $j \geq 1$ it holds that

$$|\mathsf{D}^m\pi_j(u)| \leq c(1+\log j)^m j^{u-1}, \tag{38}$$
$$|\mathsf{D}^m\alpha_{N,j}(u)| \leq c(1+\log j)^m j^{u-1}. \tag{39}$$

For $m \geq 0$ and $j \geq 1$ we also have the more precise evaluations

$$\pi_j(u) = \frac{j^{u-1}}{\Gamma(u)}(1+\epsilon_{1j}(u)), \tag{40}$$

where $\sup_{u\in\mathcal{K}}|\epsilon_{1j}(u)| \to 0$ as $j\to\infty$ for any compact set $\mathcal{K}\subset \mathbb{R}\backslash\{0,-1,\ldots\}$, and

$$\alpha_{N,j}(u) = \frac{N!}{\Gamma(u+N)}j^{u-1}(1+\epsilon_{2j}(u)), \tag{41}$$

where $\sup_{v\in\mathcal{K}}|\epsilon_{2j}(u)| \to 0$ as $j\to\infty$ for any compact set $\mathcal{K}\subset \mathbb{R}\backslash\{-N,-(N+1),\ldots\}$.

Proof. To show (37), we first note that for $j = N$ the result is trivial. For $j > N$ we factor out the first N coefficients, $\prod_{i=1}^{N}(i+u-1)/i = \pi_N(u)$. The product of the remaining coefficients is denoted $\alpha_{N,j}(u)$. The results (38) and (40) for $\pi_j(u)$ can be found in JN (2012a, Lemma A.5), and the results (39) and (41) for $\alpha_{N,j}(u)$ can be found in the same way from a Taylor's expansion of $\sum_{i=j_0}^{j}\log(1+(u-1)/i)$ for $j > j_0 \geq 1-u$. ■

Lemma A.4 *Let $a_j = 1_{\{j\geq 1\}}\sum_{k=1}^{j}k^{-1}$. Then,*

$$\pi_0(u) = 1 \text{ and } \pi_1(u) = u \text{ for any } u, \tag{42}$$
$$\mathsf{D}^m\pi_0(u) = 0 \text{ and } \mathsf{D}^m\pi_1(u) = 1_{\{m=1\}} \text{ for } m\geq 1 \text{ and any } u, \tag{43}$$
$$\mathsf{D}\pi_j(0) = j^{-1}1_{\{j\geq 1\}} \text{ and } \mathsf{D}^2\pi_j(0) = 2j^{-1}a_{j-1}1_{\{j\geq 2\}}, \tag{44}$$
$$|\mathsf{D}^m\pi_j(0)| \leq cj^{-1}(1+\log j)^{m-1}1_{\{j\geq 1\}} \leq cj^{-1+\delta} \text{ for } m\geq 1 \text{ and any } \delta > 0, \tag{45}$$
$$\sum_{n=j}^{k}\mathsf{D}^m\pi_n(-u) = \mathsf{D}^m\pi_k(-u+1) - \mathsf{D}^m\pi_{j-1}(-u+1) \text{ for } m\geq 0 \text{ and any } u, \tag{46}$$

$$\sum_{n=j}^{\infty} \mathsf{D}^m \pi_n(-u) = -\mathsf{D}^m \pi_{j-1}(-u+1) \text{ for } m \geq 0 \text{ and } u > 0, \tag{47}$$

$$\sum_{n=0}^{k} \pi_n(u)\pi_{k-n}(v) = \pi_k(u+v) \text{ for any } u, v. \tag{48}$$

Proof of Lemma A.4. Result (42) is well known and follows trivially from (2), and (43) follows by taking derivatives in (42). Next, (44) and (45) follow from Lemmas A.2 and A.3. To prove (46) with $m = 0$ multiply the identity $\binom{u}{n} = \binom{u-1}{n} + \binom{u-1}{n-1}$ by $(-1)^n$ to get

$$(-1)^n \binom{u}{n} = (-1)^n \binom{u-1}{n} - (-1)^{n-1} \binom{u-1}{n-1}.$$

Summation from $n = j$ to $n = k$ yields a telescoping sum such that

$$\sum_{n=j}^{k} (-1)^n \binom{u}{n} = (-1)^k \binom{u-1}{k} - (-1)^{k-1} \binom{u-1}{j-1},$$

which in terms of the coefficients $\pi_n(\cdot)$ gives the result. Take derivatives to find (46) with $m \geq 1$. From (38) of Lemma A.3, $\mathsf{D}^m \pi_k(-u+1) \leq c(1+\log k)^m k^{-u} \to 0$ as $k \to \infty$ when $u > 0$ which shows (47). Finally, (48) follows from the Chu-Vandermonde identity, see Askey (1975, pp. 59–60). ■

Lemma A.5 *For any α, β it holds that*

$$\sum_{n=1}^{t-1} n^{\alpha-1}(t-n)^{\beta-1} \leq c(1+\log t) t^{\max(\alpha+\beta-1, \alpha-1, \beta-1)}. \tag{49}$$

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

$$\sum_{k=1}^{\infty} (k+h)^{\alpha-1} k^{\beta-1} (1+\log(k+h))^n \leq c h^{\alpha+\beta-1} (1+\log h)^n. \tag{50}$$

Proof of Lemma A.5. *(49)*: See JN (2010, Lemma B.4).
(50): We first consider the summation from $k = 1$ to h:

$$\begin{aligned} h^{1-\alpha-\beta} \sum_{k=1}^{h} (k+h)^{\alpha-1} k^{\beta-1} (1+\log(k+h))^n &\leq c(1+\log 2h)^n h^{-1} \sum_{k=1}^{h} (\frac{k}{h}+1)^{\alpha-1} (\frac{k}{h})^{\beta-1} \\ &\leq c(1+\log h)^n \int_0^1 (1+u)^{\alpha-1} u^{\beta-1} du. \end{aligned}$$

The integral is finite for $\beta > 0$ and all α because $1 \leq 1+u \leq 2$.

To evaluate the summation from $k = h+1$ to ∞ we note that $\log(k+h) \leq \log(2k) \leq c \log k$ for $h \leq k$. This gives the bound

$$\begin{aligned} \sum_{k=h+1}^{\infty} (k+h)^{\alpha-1} k^{\beta-1} (1+\log(k+h))^n &\leq c \sum_{k=h+1}^{\infty} (h+k)^{\alpha-1} k^{\beta-1} (1+\log k)^n \\ &\leq c \sum_{k=h}^{\infty} k^{\alpha+\beta-2} (1+\log k)^n \leq c h^{\alpha+\beta-1} (1+\log h)^n, \end{aligned}$$

see (33) of Lemma A.1. ■

Lemma A.6 *For $d > 1/2$ and $2d-1-u > 0$ it holds that*

$$\sum_{n=0}^{\infty}\binom{d-1}{n}\binom{d-1-u}{n} = \frac{\Gamma(2d-1-u)}{\Gamma(d)\Gamma(d-u)} = \binom{2d-2-u}{d-1},$$

$$\sum_{n=0}^{\infty}\binom{d-1}{n}\frac{\partial}{\partial u}\binom{d-1-u}{n}\bigg|_{u=0} = -\binom{2d-2}{d-1}(\Psi(2d-1)-\Psi(d)).$$

Proof of Lemma A.6. With the notation $a_{(n)} = a(a+1)\cdots(a+n-1)$, Gauss's Hypergeometric Theorem, see Abramowitz and Stegun (1964, p. 556, eqn. 15.1.20), shows that

$$\sum_{n=0}^{\infty}\frac{a_{(n)}b_{(n)}}{c_{(n)}n!} = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} \text{ for } c > a+b.$$

For $a = -d+1$, $b = -d+1+u$, and $c = 1$, we have $c-a-b = 2d-1-u > 0$ so that

$$\begin{aligned}\sum_{n=0}^{\infty}\binom{d-1}{n}\binom{d-1-u}{n} &= \sum_{n=0}^{\infty}\frac{(-d+1)_{(n)}}{n!}\frac{(-d+1+u)_{(n)}}{n!}\\ &= \frac{\Gamma(1)\Gamma(2d-1-u)}{\Gamma(d)\Gamma(d-u)} = \binom{2d-2-u}{d-1}\end{aligned}$$

with derivative with respect to u as given, using $\partial\log\Gamma(d+u)/\partial u|_{u=0} = \Psi(d)$. ■

Appendix B Asymptotic analysis of the derivatives

We first analyze $\Delta_0^d(X_t - C)$ and introduce some notation. From Lemma 1 we have an expression for $X_t, t = 1,\dots,N+T$, and we insert that into $\Delta_0^d X_t$ and find, using $\Delta_0^d X_t = \sum_{n=0}^{t-1}\pi_n(-d)X_{t-n}$ and (46), that for $t \geq N+1$ we have

$$\begin{aligned}\Delta_0^d(X_t - C) &= \Delta_N^d X_t + \sum_{n=t-N}^{t-1}\pi_n(-d)X_{t-n} - \sum_{n=0}^{t-1}\pi_n(-d)C\\ &= \Delta_N^{d-d_0}\varepsilon_t - \Delta_N^{d-d_0}\{\sum_{n=t-N}^{t+N_0-1}\pi_n(-d_0)X_{t-n} - \pi_{t+N_0-1}(-d_0+1)\mu_0\}\\ &\quad + \sum_{n=t-N}^{t-1}\pi_n(-d)X_{t-n} - \pi_{t-1}(-d+1)C\\ &= \Delta_N^{d-d_0}\varepsilon_t + \eta_t(d) - \kappa_{0t}(d)(C-\mu_0), \qquad (51)\end{aligned}$$

where

$$\begin{aligned}\eta_t(d) = &-\sum_{k=0}^{t-1-N}\pi_k(d_0-d)\sum_{n=-N_0+1}^{N}\pi_{t-n-k}(-d_0)X_n + \sum_{n=1}^{N}\pi_{t-n}(-d)X_n\\ &+\sum_{k=0}^{t-1-N}\pi_k(d_0-d)\pi_{t+N_0-k-1}(-d_0+1)\mu_0 - \pi_{t-1}(-d+1)\mu_0.\end{aligned}$$

The derivatives of $\Delta_0^d(X_t - C)$ with respect to d, evaluated at $d = d_0$, are of the form

$$\mathsf{D}^m \Delta_0^{d_0}(X_t - C) = S_{mt}^+ + \eta_{mt}(d_0) - \kappa_{mt}(d_0)(C - \mu_0), \tag{52}$$

where

$$\kappa_{mt}(d) = (-1)^m \mathsf{D}^m \pi_{t-1}(-d+1)$$

and the stochastic term S_{mt}^+ is defined, for $t \geq N+1$, as

$$S_{mt} = (-1)^m \sum_{k=0}^{\infty} \mathsf{D}^m \pi_k(0) \varepsilon_{t-k} = S_{mt}^+ + S_{mt}^-,$$

$$S_{mt}^+ = (-1)^m \sum_{k=0}^{t-1-N} \mathsf{D}^m \pi_k(0) \varepsilon_{t-k} \text{ and } S_{mt}^- = (-1)^m \sum_{k=t-N}^{\infty} \mathsf{D}^m \pi_k(0) \varepsilon_{t-k}.$$

The main deterministic term is

$$\eta_{mt}(d) = (-1)^{m+1} [\sum_{n=-N_0+1}^{N} \sum_{k=0}^{t-1-N} \mathsf{D}^m \pi_k(0) \pi_{t-k-n}(-d) X_n - \sum_{n=1}^{N} \mathsf{D}^m \pi_{t-n}(-d) X_n \tag{53}$$
$$- \sum_{k=0}^{t-1-N} \mathsf{D}^m \pi_k(0) \pi_{t+N_0-k-1}(-d+1)\mu_0 + \mathsf{D}^m \pi_{t-1}(-d+1)\mu_0].$$

We use the notation $\langle u, v \rangle_T = \sigma_0^{-2} \sum_{t=N+1}^{N+T} u_t v_t \to \sigma_0^{-2} \sum_{t=N+1}^{\infty} u_t v_t = \langle u, v \rangle$, if the limit exists.

We first give the order of magnitude of the deterministic terms and product moments containing these.

Lemma B.1 *The functions $\eta_{mt}(d)$ satisfy*

$$|\eta_{0t}(d)| \leq ct^{-d}, \tag{54}$$
$$|\eta_{mt}(d)| \leq c(t-N)^{-\min(1,d)+\delta} \text{ for } m \geq 1,\ t \geq N+1, \text{ and any } \delta > 0. \tag{55}$$

For $d > 1/2$ it follows that, for any $0 < \epsilon < \min(d, 2d-1)$,

$$\langle \eta_m, \eta_n \rangle_T \to \langle \eta_m, \eta_n \rangle < \infty,\ m, n \geq 0, \tag{56}$$
$$|\langle \eta_m, \kappa_n \rangle_T| \leq c(1+N)^{-\min(d,2d-1)+\epsilon},\ m, n \geq 0, \tag{57}$$
$$\max(|\langle \eta_0, \eta_1 \rangle_T|, |\langle \kappa_1, \kappa_0 \rangle_T|) \leq c(1+N)^{-\min(d,2d-1)+\epsilon}. \tag{58}$$

If $N = 0$ it holds that

$$\langle \kappa_0, \kappa_0 \rangle_T \to \sigma_0^{-2} \binom{2d-2}{d-1} \text{ and } \langle \kappa_0, \kappa_1 \rangle_T \to -\sigma_0^{-2} \binom{2d-2}{d-1} (\Psi(2d-1) - \Psi(d)). \tag{59}$$

If Assumption 1 holds then

$$\langle S_m^+, \eta_n \rangle_T \xrightarrow{P} \langle S_m^+, \eta_n \rangle,\ m, n \geq 0, \tag{60}$$
$$\langle S_m^+, \kappa_n \rangle_T \xrightarrow{P} \langle S_m^+, \kappa_n \rangle,\ m, n \geq 0, \tag{61}$$

where $E(\langle S_m^+, \eta_n \rangle_T) = E(\langle S_m^+, \kappa_n \rangle_T) = E(\langle S_m^+, \eta_n \rangle) = E(\langle S_m^+, \kappa_n \rangle) = 0$.

Proof of Lemma B.1. *(54)*: The expression for $\eta_{0t}(d)$ is

$$\eta_{0t}(d) = -\sum_{n=-N_0+1}^{0} \pi_{t-n}(-d)X_n + \pi_{t+N_0-1}(-d+1)\mu_0 - \pi_{t-1}(-d+1)\mu_0$$
$$= -\sum_{n=-N_0+1}^{0} \pi_{t-n}(-d)(X_n - \mu_0), \quad (62)$$

see (46) of Lemma A.4. Using the bound $|\pi_{t-n}(-d)| \leq c(t-n)^{-d-1}$ we find $\sum_{n=-N_0+1}^{0} |\pi_{t-n}(-d)| \leq ct^{-d}$ for $n \leq 0$, see (38) of Lemma A.3, and the result follows.
(55): The remaining deterministic terms with $m \geq 1$ are evaluated using $|(-1)^{m+1}\mathsf{D}^m \pi_k(0)| \leq ck^{-1+\delta}1_{\{k\geq 1\}}$ for $\delta > 0$, see (45) of Lemma A.4, and we find, for $t \geq N+1$,

$$|\eta_{mt}(d)| \leq c\sum_{n=-N}^{\infty}\sum_{k=1}^{t-1-N} k^{-1+\delta}(t-k+n)^{-d-1} + c\sum_{n=1}^{N}(t-n)^{-d-1+\delta}$$
$$+ c\sum_{k=1}^{t-1-N} k^{-1+\delta}(t+N_0-k-1)^{-d} + c(t-1)^{-d+\delta}$$
$$\leq c[\sum_{k=1}^{t-1-N} k^{-1+\delta}(t-k-N)^{-d} + (t-N)^{-d+\delta}]$$
$$\leq c[(1+\log(t-N))(t-N)^{-\min(1,d)+\delta} + (t-N)^{-d+\delta}] \leq c(t-N)^{-\min(1,d)+2\delta},$$

where we have used (49) of Lemma A.5.
(56): From (55) we find $|\eta_{mt}(d)\eta_{nt}(d)| \leq c(t-N)^{-2\min(1,d)+2\delta}$ so that $|\langle \eta_m, \eta_m\rangle| < \infty$ by choosing $2\delta < 2\min(1,d) - 1 = \min(1, 2d-1)$, which is possible for $d > 1/2$.
(57): Similarly we find $\langle \eta_n, \kappa_m\rangle_T \leq c\sum_{t=1}^{\infty} t^{-\min(1,d)+\delta}(t+N-1)^{-d}$. If $1/2 < d < 1$ we apply (50) of Lemma A.5 to obtain the result $\sum_{t=1}^{\infty}(t+N)^{-d}t^{-d+\epsilon} \leq c(1+N)^{1-2d+\epsilon}$, and if $d \geq 1$ we use $(t+N)^{-d} \leq (1+N)^{-d+2\epsilon}t^{-2\epsilon}$ for $2\epsilon < d$ and find

$$\sum_{t=1}^{\infty}(t+N)^{-d}t^{-1+\epsilon} \leq (1+N)^{-d+2\epsilon}\sum_{t=1}^{\infty} t^{-1-\epsilon} \leq c(1+N)^{-d+2\epsilon}.$$

(58): The proofs for $\langle \eta_0, \eta_1\rangle_T$ and $\langle \kappa_1, \kappa_0\rangle_T$ are the same as for (57).
(59): For $N = 0$ we find $\langle \kappa_0, \kappa_0\rangle_T = \sum_{n=0}^{T-1}\binom{d-1}{n}^2$ and $\langle \kappa_0, \kappa_1\rangle_T = \frac{1}{2}\mathsf{D}\sum_{t=1}^{T}\kappa_{0t}(d)^2$ such that the result follows from Lemma A.6.
(60): We have

$$\sum_{t=N+T+1}^{\infty} S_{mt}^{+}\eta_{nt}(d) = \sum_{t=N+T+1}^{\infty} \eta_{nt}(d)(-1)^{m+1}\sum_{k=1}^{t-1-N}\mathsf{D}^m\pi_{t-k}(0)\varepsilon_k \quad (63)$$
$$= \sum_{k=1}^{\infty}[\sum_{t=\max(T,k)+N+1}^{\infty} \eta_{nt}(d)(-1)^{m+1}\mathsf{D}^m\pi_{t-k}(0)]\varepsilon_k.$$

For some small $\delta > 0$ to be chosen subsequently, we use the evaluations $|\eta_{nt}(d)| \leq c(t-N)^{-\min(1,d)+\delta}$, $|\mathsf{D}^m\pi_{t-k}(0)| \leq c(t-k)^{-1+\delta}1_{\{t-k\geq 1\}}$, and $t^{-\min(1,d)+\delta} = (t-k+k)^{-\min(1,d)+\delta} \leq (t-k)^{-2\delta}k^{-\min(1,d)+3\delta}$. Then

$$Var(\sum_{t=N+T+1}^{\infty} S_{mt}^{+}\eta_{nt}(d)) \leq c\sum_{k=1}^{\infty}[\sum_{t=\max(T,k)+1}^{\infty} t^{-\min(1,d)+\delta}(t+N-k)^{-1+\delta}]^2$$
$$\leq c\sum_{k=1}^{\infty}k^{-2\min(1,d)+6\delta}[\sum_{t=\max(T,k)+1}^{\infty}(t-k)^{-1-\delta}]^2.$$

For $T\to\infty$ we have $\sum_{t=\max(T,k)+1}^{\infty}(t-k)^{-1-\delta}\to 0$, and because $\sum_{k=1}^{\infty}k^{-2\min(1,d)+6\delta} < \infty$ we get by dominated convergence that $Var(\sum_{t=N+T+1}^{\infty}S_{mt}^{+}\eta_{nt}(d)) \to 0$. This shows that $\langle S_m^+, \eta_n\rangle_T \overset{P}{\to} \langle S_m^+, \eta_n\rangle = \sum_{t=N+1}^{\infty}S_{mt}^{+}\eta_{nt}(d)$.
(61): We use (63) and find $\sum_{t=N+T+1}^{\infty}S_{mt}^{+}\kappa_{nt}(d) = \sum_{k=1}^{\infty}[\sum_{t=\max(T,k)+1}^{\infty}\kappa_{nt}(d)(-1)^{m+1}\mathsf{D}^m\pi_{t-k}(0)]\varepsilon_k$, and the proof is completed as for (60). ■

We next define the (centered) product moments of the stochastic terms,

$$M_{mnT}^{+} = \sigma_0^{-2}T^{-1/2}\sum_{t=N+1}^{N+T}(S_{mt}^{+}S_{nt}^{+} - E(S_{mt}^{+}S_{nt}^{+})), \tag{64}$$

as well as the product moments derived from the corresponding stationary processes,

$$M_{mnT} = \sigma_0^{-2}T^{-1/2}\sum_{t=N+1}^{N+T}(S_{mt}S_{nt} - E(S_{mt}S_{nt})).$$

The next two lemmas give their asymptotic behavior, where we note that

$$E(S_{0t}^{+}S_{mt}^{+}) = E(S_{0t}S_{mt}) = 0 \text{ for } m \geq 1. \tag{65}$$

Lemma B.2 *Suppose Assumption 1 holds and let* $\zeta_2 = \sum_{j=1}^{\infty}j^{-2} = \pi^2/6 \simeq 1.6449$ *and* $\zeta_3 = \sum_{j=1}^{\infty}j^{-3} \simeq 1.2021$*, see (16). Then*

$$E(M_{01T}^2) = \sigma_0^{-2}T^{-1}\sum_{t=N+1}^{N+T}E(S_{1t}^2) = \zeta_2, \tag{66}$$

$$E(M_{01T}M_{02T}) = \sigma_0^{-4}T^{-1}\sum_{s,t=N+1}^{N+T}E(S_{0t}S_{1t}S_{0s}S_{2s}) = \sigma_0^{-2}T^{-1}\sum_{t=N+1}^{N+T}E(S_{1t}S_{2t}) = -2\zeta_3, \tag{67}$$

$$E(M_{01T}M_{11T}) = \sigma_0^{-4}T^{-1}\sum_{s,t=N+1}^{N+T}E(S_{0t}S_{1t}S_{1s}^2) = -4\zeta_3, \tag{68}$$

$$E(\langle S_0^+, \kappa_0\rangle_T\langle S_1^+, \kappa_0\rangle_T = \sigma_0^{-4}\sum_{s,t=N+1}^{N+T}E(S_{0s}^{+}\kappa_{0s}(d)S_{1t}^{+}\kappa_{0t}(d)) \tag{69}$$
$$= -\sigma_0^{-2}\sum_{N\leq s<t\leq N+T-1}(t-s)^{-1}\pi_t(-d+1)\pi_s(-d+1).$$

It follows that, for $N = 0$ and $T \to \infty$,

$$\tau_{0,T}(d) = -\frac{E(\langle S_0^+, \kappa_0\rangle_T \langle S_1^+, \kappa_0\rangle_T)}{\langle \kappa_0, \kappa_0\rangle_T} \to -(\Psi(2d-1) - \Psi(d)). \tag{70}$$

Furthermore, for T fixed and $N \to \infty$, see also (37),

$$\tau_{N,T}(d) = \frac{\sum_{N \le s < t \le N+T-1}(t-s)^{-1}\alpha_{N,t}(-d+1)\alpha_{N,s}(-d+1)}{\sum_{N+1 \le t \le N+T}\alpha_{N,t-1}(-d+1)^2} \to \sum_{t=1}^{T-1} t^{-1} - (T-1)/T. \tag{71}$$

Proof of Lemma B.2. *(66)*: From $S_{0t} = \varepsilon_t$, $S_{1t} = -\sum_{k=1}^{\infty} k^{-1}\varepsilon_{t-k}$, and (65) we find

$$E(M_{01T}^2) = \sigma_0^{-4}E[T^{-1/2}\sum_{t=N+1}^{N+T}\varepsilon_t\sum_{k=1}^{\infty}k^{-1}\varepsilon_{t-k}]^2 = \sigma_0^{-2}T^{-1}\sum_{t=N+1}^{N+T}E[\sum_{k=1}^{\infty}k^{-1}\varepsilon_{t-k}]^2 = \sum_{k=1}^{\infty}k^{-2} = \zeta_2.$$

(67): We find using the expressions for S_{0t}, S_{1t}, and $S_{2t} = 2\sum_{j=2}^{\infty} j^{-1}a_{j-1}\varepsilon_{t-j}$, $a_j = 1_{\{j\ge 1\}}\sum_{k=1}^{j} k^{-1}$, together with (65) that

$$E(M_{01T}M_{02T}) = -2\sigma_0^{-4}T^{-1}E[\sum_{t=N+1}^{N+T}\varepsilon_t\sum_{k=1}^{\infty}k^{-1}\varepsilon_{t-k}][\sum_{s=N+1}^{N+T}\varepsilon_s\sum_{j=2}^{\infty}(j^{-1}a_{j-1})\varepsilon_{s-j}] = \sigma_0^{-2}T^{-1}\sum_{t=N+1}^{N+T}E(S_{1t}S_{2t})$$

and

$$\begin{aligned}\sigma_0^{-2}T^{-1}\sum_{t=N+1}^{N+T}E(S_{1t}S_{2t}) &= -2\sigma_0^{-2}T^{-1}\sum_{t=N+1}^{N+T}E[\sum_{k=1}^{\infty}k^{-1}\varepsilon_{t-k}\sum_{j=2}^{\infty}(j^{-1}a_{j-1})\varepsilon_{t-j}] \\ &= -2T^{-1}\sum_{t=N+1}^{N+T}\sum_{j=2}^{\infty}j^{-2}\sum_{k=1}^{j-1}k^{-1} = -2\sum_{j=2}^{\infty}j^{-2}\sum_{k=1}^{j-1}k^{-1} = -2\kappa_3 \end{aligned} \tag{72}$$

for $\kappa_3 = \sum_{j=2}^{\infty} j^{-2}\sum_{k=1}^{j-1}k^{-1}$. We thus need to show that $\kappa_3 = \zeta_3$.

Let $f(\lambda) = \log(1 - e^{\mathrm{i}\lambda}) = \frac{1}{2}c(\lambda) + \mathrm{i}\theta(\lambda)$, where $\mathrm{i} = \sqrt{-1}$ is the imaginary unit, $c(\lambda) = \log(2(1-\cos(\lambda))$, $\theta(\lambda) = \arg(1 - e^{\mathrm{i}\lambda}) = -(\pi - \lambda)/2$ for $0 < \lambda < \pi$, and $\theta(-\lambda) = -\theta(\lambda)$. The transfer function of S_{mt} is $\mathsf{D}^m(1-e^{\mathrm{i}\lambda})^{d-d_0}|_{d=d_0} = f(\lambda)^m$, so that the cross spectral density between S_{mt} and S_{nt} is $f(\lambda)^m f(-\lambda)^n$ and $E(S_{mt}S_{nt}) = \frac{\sigma_0^2}{2\pi}\int_{-\pi}^{\pi} f(\lambda)^m f(-\lambda)^n \mathsf{d}\lambda$. Because $c(\lambda)$ is symmetric around zero and $\theta(\lambda)$ is anti-symmetric around zero, i.e. $\theta(-\lambda) = -\theta(\lambda)$, it follows that

$$c(\lambda)^3 = (f(\lambda) + f(-\lambda))^3 = f(\lambda)^3 + 3f(\lambda)^2 f(-\lambda) + 3f(\lambda)f(-\lambda)^2 + f(-\lambda)^3.$$

Noting that the transfer function of $S_{0t} = \varepsilon_t$ is $f(\lambda)^0 = 1$ and integrating both sides we find

$$\frac{\sigma_0^2}{2\pi}\int_{-\pi}^{\pi} c(\lambda)^3 \mathsf{d}\lambda = E(S_{3t}S_{0t}) + 3E(S_{2t}S_{1t}) + 3E(S_{1t}S_{2t}) + E(S_{0t}S_{3t}).$$

The left-hand side is given as $-12\sigma_0^2\zeta_3$ in Lieberman and Phillips (2004, p. 478) and the right-hand side is $-12\sigma_0^2\kappa_3$ from (65) and (72), which proves the result.

(68): We find, using the expressions for S_{mt} and (65), that $E(M_{01T}M_{11T})$ is

$$\sigma_0^{-4}T^{-1}\sum_{s,t=N+1}^{N+T}E(S_{0t}S_{1t}S_{1s}^2) = -T^{-1}\sum_{s,t=N+1}^{N+T}E[\varepsilon_t\sum_{k=-\infty}^{t-1}(t-k)^{-1}\varepsilon_k\sum_{j=-\infty}^{s-1}(s-j)^{-1}\varepsilon_j\sum_{i=-\infty}^{s-1}(s-i)^{-1}\varepsilon_i].$$

The only contribution comes for $t = j > k = i$ or $t = i > k = j$ and therefore $t < s$. These two contributions are equal, so we find, using $s - k = s - t + t - k$,

$$2T^{-1}\sum_{t=N+1}^{N+T}\sum_{s=t+1}^{N+T}\sum_{k=-\infty}^{t-1}(t-k)^{-1}(s-t)^{-1}(s-k)^{-1} = 2T^{-1}\sum_{t=N+1}^{N+T}\sum_{s=t+1}^{N+T}\sum_{k=-\infty}^{t-1}[(t-k)^{-1}+(s-t)^{-1}](s-k)^{-2}.$$

Next we introduce $u = s - k$ [≥ 2] and $v = t - k$ [$1 \leq v < u$] and find

$$2\sum_{u=2}^{\infty}\sum_{v=1}^{u-1}[v^{-1} + (u-v)^{-1}]u^{-2} = 4\sum_{u=2}^{\infty}u^{-2}\sum_{v=1}^{u-1}v^{-1} = 4\kappa_3 = 4\zeta_3,$$

which proves (68).

(69): From $S_{0s}^+ = \varepsilon_s$, $S_{1t}^+ = -\sum_{k=1}^{t-1-N}k^{-1}\varepsilon_{t-k} = -\sum_{k=N+1}^{t-1}(t-k)^{-1}\varepsilon_k$, and $\kappa_{0t}(d) = \pi_{t-1}(-d+1)$ we get

$$\begin{aligned} E\langle S_0^+, \kappa_0\rangle_T\langle S_1^+, \kappa_0\rangle_T &= \sigma_0^{-4}\sum_{s,t=N+1}^{N+T}E(S_{0s}^+\kappa_{0s}(d)S_{1t}^+(d)\kappa_{0t}(d) \\ &= -\sigma_0^{-2}\sum_{N+1\leq s<t\leq N+T}(t-s)^{-1}\pi_{t-1}(-d+1)\pi_{s-1}(-d+1). \end{aligned}$$

(70): For $N = 0$ we use $\mathsf{D}\pi_{t-s}(u)|_{u=0} = (t-s)^{-1}1_{\{t-s\geq 1\}}$ and find the limit

$$\begin{aligned} &\sum_{0\leq s<t<\infty}\mathsf{D}\pi_{t-s}(u)|_{u=0}\pi_s(-d+1)\pi_t(-d+1) = \sum_{t=1}^{\infty}\mathsf{D}\pi_t(-d+1+u)|_{u=0}\pi_t(-d+1) \\ &= \sum_{t=1}^{\infty}\mathsf{D}\binom{d-1-u}{t}|_{u=0}\binom{d-1}{t} = -\binom{2d-2}{d-1}(\Psi(2d-1) - \Psi(d)) \end{aligned}$$

using (48) and Lemma A.6. From (59) we find the limit of $\langle\kappa_0, \kappa_0\rangle_T$.

(71): From (37) we find the representation in (71), where we have cancelled the factor $\pi_N(-d+1)^2$. Note that $\sum_{N+1\leq t\leq N+T}\alpha_{N,t-1}(-d+1)^2 \geq \alpha_{N,N}(-d+1)^2 = 1$ and $\alpha_{N,t}(-d+1) = \prod_{i=N+1}^{t}(1-d/i) \to 1$ for $N \to \infty$ and $t \geq N+1$, so that $\tau_{N,T}(d) \to T^{-1}\sum_{N\leq s<t\leq N+T-1}(t-s)^{-1} = \sum_{i=1}^{T-1}i^{-1} - (T-1)/T$. ∎

Lemma B.3 *Suppose Assumption 1 holds. Then, for $T \to \infty$, it holds that $\{M_{mnT}\}_{0\leq m,n\leq 3}$ are jointly asymptotically normal with mean zero, and some variances and covariances can be calculated from (66), (67), and (68) in Lemma B.2. It follows that the same holds for $\{M_{mnT}^+\}_{0\leq m,n\leq 3}$ with the same variances and covariances.*

Proof of Lemma B.3. $\{M_{mnT}\}$: We apply a result by Giraitis and Taqqu (1998) on limit distributions of quadratic forms of linear processes. We define the cross covariance function

$$r_{mn}(t) = E(S_{m0}S_{nt}) = \sigma_0^2(-1)^{m+n}\sum_{k=0}^{\infty} \mathsf{D}^m\pi_k(0)\mathsf{D}^n\pi_{t+k}(0)$$

and find $r_{00}(t) = \sigma_0^2 1_{\{t=0\}}$, $r_{m0}(t) = \sigma_0^2(-1)^m\mathsf{D}^m\pi_{-t}(0)1_{\{t\leq -1\}}$, and $r_{0n}(t) = \sigma_0^2(-1)^n\mathsf{D}^n\pi_t(0)$. For $m, n \geq 1$ we find that $|r_{mn}(t)|$ is bounded for a small $\delta > 0$ by

$$c\sum_{k=1}^{\infty}(1+\log(t+k))^{m-1}(1+\log k)^{n-1}(t+k)^{-1}k^{-1} \leq c\sum_{k=1}^{\infty}(t+k)^{-1+\delta}k^{-1+\delta} \leq ct^{-1+3\delta},$$

using the bound $(t+k)^{-1+\delta} \leq k^{-2\delta}t^{-1+3\delta}$. Thus $\sum_{t=-\infty}^{\infty} r_{mn}(t)^2 < \infty$, and joint asymptotic normality of $\{M_{mnT}\}_{0\leq m,n\leq 3}$ then follows from Theorem 5.1 of Giraitis and Taqqu (1998). The asymptotic variances and covariances can be calculated as in (66), (67), and (68) in Lemma B.2.

$\{M^+_{mnT}\}$: We show that $E(M_{mnT} - M^+_{mnT})^2 \to 0$. We find

$$M_{mnT} - M^+_{mnT} = \sigma_0^{-2}T^{-1/2}\sum_{t=N+1}^{N+T}(S^+_{mt}S^-_{nt} + S^-_{mt}S^+_{nt} + S^-_{mt}S^-_{nt} - E(S^+_{mt}S^-_{nt} + S^-_{mt}S^+_{nt} + S^-_{mt}S^-_{nt})), \tag{73}$$

and show that the expectation term converges to zero and that each of the stochastic terms has a variance tending to zero.

$T^{-1/2}\sum_{t=N+1}^{N+T} E(S^+_{mt}S^-_{nt} + S^-_{mt}S^+_{nt} + S^-_{mt}S^-_{nt}) \to 0$: The first two terms are zero because S^+_{mt} and S^-_{nt} are independent. For the last term we find using (45) of Lemma A.4 that

$$|E(S^-_{mt}S^-_{nt})| = \sigma_0^2\sum_{k=t-N}^{\infty}|\mathsf{D}^m\pi_k(0)\mathsf{D}^n\pi_k(0)| \leq c\sum_{k=t-N}^{\infty}k^{-2+\delta} \leq c(t-N)^{-1+\delta},$$

so that

$$T^{-1/2}\sum_{t=N+1}^{N+T}E(S^-_{mt}S^-_{nt}) \leq cT^{-1/2+\delta} \to 0. \tag{74}$$

$Var(T^{-1/2}\sum_{t=N+1}^{N+T} S^+_{mt}S^-_{nt}) \to 0$: The first two terms of (73) are handled the same way. We find because (S^+_{mt}, S^+_{ns}) is independent of (S^-_{mt}, S^-_{ns}) that

$$Var(T^{-1/2}\sum_{t=N+1}^{N+T}S^+_{mt}S^-_{nt}) = T^{-1}\sum_{s,t=N+1}^{N+T}E(S^+_{mt}S^-_{nt}S^+_{ms}S^-_{ns}) = T^{-1}\sum_{s,t=N+1}^{N+T}E(S^+_{mt}S^+_{ms})E(S^-_{nt}S^-_{ns}).$$

Then replacing the log factors by a small power, $\delta > 0$, we find for $|\mathsf{D}^m\pi_{t-i}(0)| \leq c(t-i)^{-1}(1+\log(t-i))^m \leq c(t-i)^{-1+\delta}$ that

$$\begin{aligned}|E(S^+_{mt}S^+_{ms})| &= |E(\sum_{i=1}^{t-1-N}\mathsf{D}^m\pi_{t-i}(0)\varepsilon_i\sum_{j=1}^{s-1-N}\mathsf{D}^m\pi_{s-j}(0)\varepsilon_j)| = \sigma_0^2\sum_{i=1}^{\min(s,t)-1-N}|\mathsf{D}^m\pi_{t-i}(0)\mathsf{D}^m\pi_{s-i}(0)| \\ &\leq c\sum_{i=1}^{\min(s,t)-1-N}(t-i)^{-1+\delta}(s-i)^{-1+\delta}.\end{aligned}$$

Now take $s > t$ and evaluate $(s-i)^{-1+\delta} = (s-t+t-i)^{-1+\delta} \leq (s-t)^{-1+3\delta}(t-i)^{-2\delta}$ and

$$|E(S_{mt}^{+}S_{ms}^{+})| \leq c(s-t)^{-1+3\delta}\sum_{i=1}^{t-1-N}(t-i)^{-1-\delta} \leq c(s-t)^{-1+3\delta}.$$

Similarly for

$$E(S_{nt}^{-}S_{ns}^{-}) = E(\sum_{i=-\infty}^{N}\mathsf{D}^{n}\pi_{t-i}(0)\varepsilon_i\sum_{j=-\infty}^{N}\mathsf{D}^{n}\pi_{s-j}(0)\varepsilon_j) = \sigma_0^2\sum_{i=-\infty}^{N}\mathsf{D}^{n}\pi_{t-i}(0)\mathsf{D}^{n}\pi_{s-i}(0)$$

we find

$$|E(S_{nt}^{-}S_{ns}^{-})| \leq c\sum_{i=-\infty}^{N}(t-i)^{-1+\delta}(s-i)^{-1+\delta} = c\sum_{i=-N}^{\infty}(t+i)^{-1+\delta}(s+i)^{-1+\delta}$$
$$\leq c(s-t)^{-1+3\delta}\sum_{i=-N}^{\infty}(t+i)^{-1-\delta} \leq c(s-t)^{-1+3\delta}(t-N)^{-\delta}.$$

Finally, we can evaluate the variance as

$$Var(T^{-1/2}\sum_{t=N+1}^{N+T}S_{mt}^{+}S_{nt}^{-}) \leq cT^{-1}\sum_{N+1\leq t<s\leq N+T}(s-t)^{-1+3\delta}(t-N)^{-\delta}(s-t)^{-1+3\delta}$$
$$= cT^{-1}\sum_{h=1}^{T-1}h^{-2+6\delta}\sum_{t=1}^{T-h}t^{-\delta} \leq cT^{-1}T^{1-\delta} \to 0.$$

$Var(T^{-1/2}\sum_{t=N+1}^{N+T}S_{mt}^{-}S_{nt}^{-}) \to 0$: The last term of (73) has variance

$$Var(T^{-1/2}\sum_{t=N+1}^{N+T}S_{mt}^{-}S_{nt}^{-}) = T^{-1}E([\sum_{t=N+1}^{N+T}S_{mt}^{-}S_{nt}^{-}]^2) - T^{-1}[\sum_{t=N+1}^{N+T}E(S_{mt}^{-}S_{nt}^{-})]^2$$

and the first term is $T^{-1}\sum_{s,t=N+1}^{N+T}E(S_{mt}^{-}S_{nt}^{-}S_{ms}^{-}S_{ns}^{-})$ which equals

$$T^{-1}\sum_{s,t=N+1}^{N+T}\sum_{i,j,k,p=-\infty}^{N}E(\mathsf{D}^{m}\pi_{t-i}(0)\varepsilon_i\mathsf{D}^{n}\pi_{t-j}(0)\varepsilon_j\mathsf{D}^{m}\pi_{s-k}(0)\varepsilon_k\mathsf{D}^{n}\pi_{s-p}(0)\varepsilon_p).$$

There are contributions from $E(\varepsilon_i\varepsilon_j\varepsilon_k\varepsilon_p)$ in four cases which we treat in turn.

Case 1, $i = j \neq k = p$: This gives the expectation squared, $T^{-1}[\sum_{t=N+1}^{N+T}E(S_{mt}^{-}S_{nt}^{-})]^2$, which is subtracted to form the variance.

Cases 2 and 3, $i = k \neq j = p$ *and* $i = p \neq j = k$: These are treated the same way. We find for Case 2 the contribution

$$A_1 \leq cT^{-1}\sum_{s,t=N+1}^{N+T}\sum_{i=-N}^{\infty}(1+\log(t+i))^m(1+\log(s+i))^m(t+i)^{-1}(s+i)^{-1}$$
$$\times\sum_{j=-N}^{\infty}(1+\log(t+j))^n(1+\log(s+j))^n(s+j)^{-1}(t+j)^{-1}$$
$$\leq cT^{-1}\sum_{s,t=N+1}^{N+T}[\sum_{i=-N}^{\infty}(t+i)^{-1+\delta}(s+i)^{-1+\delta}]^2 \leq cT^{-1}\sum_{N+1\leq t<s\leq N+T}[\sum_{i=-N}^{\infty}(t+i)^{-1+\delta}(s+i)^{-1+\delta}]^2.$$

We evaluate $(s+i)^{-1+\delta} = (s-t+t+i)^{-1+\delta} \leq (s-t)^{-1+3\delta}(t+i)^{-2\delta}$ so that

$$\sum_{i=-N}^{\infty}(t+i)^{-1+\delta}(s+i)^{-1+\delta} \leq \sum_{i=-N}^{\infty}(s-t)^{-1+3\delta}(t+i)^{-1-\delta} \leq (s-t)^{-1+3\delta}(t-N)^{-\delta}$$

and hence

$$A_1 \leq cT^{-1}\sum_{N+1\leq t<s\leq N+T}(s-t)^{-2+6\delta}(t-N)^{-2\delta} = cT^{-1}\sum_{h=1}^{T-1}h^{-2+6\delta}\sum_{t=1}^{T-h}t^{-2\delta} \leq cT^{-1}T^{1-2\delta} \to 0.$$

Case 4, $i=j=p=k$: This gives in the same way the bound

$$T^{-1}\sum_{s,t=N+1}^{N+T}\sum_{i=-N}^{\infty}(t+i)^{-2+\delta}(s+i)^{-2+\delta} \leq cT^{-1}\sum_{i=-N}^{\infty}[\sum_{t=N+1}^{N+T}(t+i)^{-2-\delta}]^2 \leq cT^{-1}\sum_{i=1}^{\infty}i^{-2-2\delta} \to 0.$$

■

We now apply the previous Lemmas B.1, B.2, and B.3, and find asymptotic results for the derivatives $\mathsf{D}^m L^*(d_0)$.

Lemma B.4 *Let the model for the data $X_t, t=1,\ldots,N+T$, be given by (4) and let Assumptions 1 and 2 be satisfied. Then the (normalized) derivatives of the concentrated likelihood function $L^*(d)$, see (12), satisfy*

$$\sigma_0^{-2}T^{-1/2}\mathsf{D}L^*(d_0) = A_0 + T^{-1/2}A_1 + O_P(T^{-1}), \tag{75}$$
$$\sigma_0^{-2}T^{-1}\mathsf{D}^2L^*(d_0) = B_0 + T^{-1/2}B_1 + O_P(T^{-1}(\log T)), \tag{76}$$
$$\sigma_0^{-2}T^{-1}\mathsf{D}^3L^*(d_0) = C_0 + O_P(T^{-1/2}), \tag{77}$$

where

$$A_0 = M_{01T}^+,\ E(A_1) = \xi_{N,T}(d_0) + \tau_{N,T}(d_0), \tag{78}$$
$$B_0 = \zeta_2,\ B_1 = M_{11T}^+ + M_{02T}^+, \tag{79}$$
$$C_0 = -6\zeta_3. \tag{80}$$

Here $\xi_{N,T}(d_0)$, $\tau_{N,T}(d_0)$, and M_{mnT}^+, are given in (21), (23), and (64), respectively, and $\zeta_2 = \pi^2/6$ and $\zeta_3 \simeq 1.2021$, see (16).

The (normalized) derivatives of $L_c^*(d)$, see (15), satisfy (75)–(77) and (79)–(80), but (78) is replaced by

$$A_0 = M_{01T}^+,\ E(A_1) = \xi_{N,T}^C(d_0), \tag{81}$$

where $\xi_{N,T}^C(d_0)$ is given by (22).

Proof of Lemma B.4. The concentrated sum of squared residuals is given in (12). We note that the first term is $O_P(T)$, and from Lemmas B.1 and B.2 the next is $O_P(1)$, so the second term has no influence on the asymptotic distribution of $\hat{d}$. However, for the bias we need to analyze it further.

We need an expression for the derivatives of the concentrated likelihood, i.e., $\mathsf{D}^m L^*(d)$. Recall $L(d,\mu)$ from (11) and denote derivatives with respect to d and μ by subscripts. Then $L^*(d) = L(d,\mu(d))$ and therefore

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

but $\hat\mu$ is determined from $L_\mu(d,\mu(d)) = 0$, which implies $L_{d\mu}(d,\mu(d)) + L_{\mu\mu}(d,\mu(d))\mu_d(d) = 0$, and hence

$$\mathsf{D}L^*(d) = L_d(d,\mu(d)), \tag{82}$$

$$\mathsf{D}^2L^*(d) = L_{dd}(d,\mu(d)) - \frac{L_{d\mu}(d,\mu(d))^2}{L_{\mu\mu}(d,\mu(d))}. \tag{83}$$

We find the derivatives for $d = d_0$ and suppress the dependence on d_0 in the following. Thus $\kappa_{0t} = \kappa_{0t}(d_0)$ and $\kappa_{1t} = \kappa_{1t}(d_0)$, etc.
(75) and (78): We find from (52) that $\mathsf{D}^m\Delta_0^{d_0}(X_t - \mu) = S_{mt}^+ + \eta_{mt} - \kappa_{mt}(\mu - \mu_0)$, and therefore from (82),

$$\sigma_0^{-2}T^{-1/2}\mathsf{D}L^* = \sigma_0^{-2}T^{-1/2}\sum_{t=N+1}^{N+T}(S_{0t}^+ + \eta_{0t} - (\hat\mu - \mu_0)\kappa_{0t})(S_{1t}^+ + \eta_{1t} - (\hat\mu - \mu_0)\kappa_{1t}),$$

where $\hat\mu - \mu_0 = \hat\mu(d_0) - \mu_0 = (\langle S_0^+, \kappa_0\rangle_T + \langle \eta_0, \kappa_0\rangle_T)/\langle \kappa_0, \kappa_0\rangle_T$. Since $E(XY) = E(X)E(Y) + Cov(X,Y)$ and $E(\hat\mu - \mu_0) = \langle \eta_0, \kappa_0\rangle_T/\langle \kappa_0, \kappa_0\rangle_T$ we get

$$\begin{aligned}E(\sigma_0^{-2}T^{-1/2}\mathsf{D}L^*) &= \sigma_0^{-2}T^{-1/2}\sum_{t=N+1}^{N+T}(\eta_{0t} - \frac{\langle \eta_0, \kappa_0\rangle_T}{\langle \kappa_0, \kappa_0\rangle_T}\kappa_{0t})(\eta_{1t} - \frac{\langle \eta_0, \kappa_0\rangle_T}{\langle \kappa_0, \kappa_0\rangle_T}\kappa_{1t})\\ &\quad + \sigma_0^{-2}T^{-1/2}\sum_{t=N+1}^{N+T}Cov((S_{0t}^+ - \frac{\langle S_0^+, \kappa_0\rangle_T}{\langle \kappa_0, \kappa_0\rangle_T}\kappa_{0t}), (S_{1t}^+ - \frac{\langle S_0^+, \kappa_0\rangle_T}{\langle \kappa_0, \kappa_0\rangle_T}\kappa_{1t})).\end{aligned}$$

The first term is $T^{-1/2}\xi_{N,T}$, see (21). The second term is, using $Cov(S_{0t}^+, S_{1t}^+) = 0$, see (65), equal to $T^{-1/2}$ times

$$\begin{aligned}&-\frac{E\langle S_0^+, \kappa_0\rangle_T\langle S_0^+, \kappa_1\rangle_T}{\langle \kappa_0, \kappa_0\rangle_T} - \frac{E\langle S_1^+, \kappa_0\rangle_T\langle S_0^+, \kappa_0\rangle_T}{\langle \kappa_0, \kappa_0\rangle_T} + \frac{E\langle S_0^+, \kappa_0\rangle_T^2\langle \kappa_0, \kappa_1\rangle_T}{\langle \kappa_0, \kappa_0\rangle_T^2}\\ &= -\frac{\langle \kappa_0, \kappa_1\rangle_T}{\langle \kappa_0, \kappa_0\rangle_T} - \frac{E\langle S_1^+, \kappa_0\rangle_T\langle S_0^+, \kappa_0\rangle_T}{\langle \kappa_0, \kappa_0\rangle_T} + \frac{\langle \kappa_0, \kappa_1\rangle_T}{\langle \kappa_0, \kappa_0\rangle_T} = -\frac{E\langle S_1^+, \kappa_0\rangle_T\langle S_0^+, \kappa_0\rangle_T}{\langle \kappa_0, \kappa_0\rangle_T} = \tau_{N,T},\end{aligned}$$

see (69) and (23).
(76) and (79): The first term of $T^{-1}\mathsf{D}^2L^*$ in (83) is analyzed below and is of the order of 1 and $T^{-1/2}$. In the second term of (83) we find $L_{\mu\mu}(d_0,\mu(d_0)) = \sigma_0^2\langle \kappa_0, \kappa_0\rangle_T = O(1)$ and

$$L_{d\mu}(d_0,\mu(d_0)) = T^{-1}\sum_{t=N+1}^{N+T}(S_{0t}^+ + \eta_{0t} - (\hat\mu - \mu_0)\kappa_{0t})\kappa_{1t} + T^{-1}\sum_{t=N+1}^{N+T}\kappa_{0t}(S_{1t}^+ + \eta_{1t} - (\hat\mu - \mu_0)\kappa_{1t}) = O_P(1),$$

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

$$\sigma_0^{-2}T^{-1}\mathsf{D}^2L^* = \sigma_0^{-2}T^{-1}\sum_{t=N+1}^{N+T}(S_{1t}^+ + \eta_{1t} - (\hat{\mu}-\mu_0)\kappa_{1t})^2$$
$$+\sigma_0^{-2}T^{-1}\sum_{t=N+1}^{N+T}(S_{0t}^+ + \eta_{0t} - (\hat{\mu}-\mu_0)\kappa_{0t})(S_{2t}^+ + \eta_{2t} - (\hat{\mu}-\mu_0)\kappa_{2t}) + O_P(T^{-1}).$$

By Lemma B.1 it holds that $\langle \eta_m, \eta_n\rangle_T = O(1)$ and $\langle S_m^+, \eta_n\rangle_T = O_P(1)$ such that

$$\sigma_0^{-2}T^{-1}\mathsf{D}^2L^* = \sigma_0^{-2}T^{-1}\sum_{t=N+1}^{N+T}E(S_{1t}^+)^2 + T^{-1/2}(M_{11T}^+ + M_{02T}^+) + O_P(T^{-1})$$
$$= \zeta_2 + T^{-1/2}(M_{11T}^+ + M_{02T}^+) + O_P(T^{-1}(\log T))$$

using also (66) and (74).
(77) and (80): For the third derivative it can be shown that the extra terms involving derivatives $\mu_d(d_0)$ and $\mu_{dd}(d_0)$ can be ignored and we find

$$\sigma_0^{-2}T^{-1}\mathsf{D}^3L^* = \sigma_0^{-2}3T^{-1}\sum_{t=N+1}^{N+T}(S_{1t}^+ + \eta_{1t} - (\hat{\mu}-\mu_0)\kappa_{1t})(S_{2t}^+ + \eta_{2t} - (\hat{\mu}-\mu_0)\kappa_{2t})$$
$$+\sigma_0^{-2}T^{-1}\sum_{t=N+1}^{N+T}(S_{0t}^+ + \eta_{0t} - (\hat{\mu}-\mu_0)\kappa_{0t})(S_{3t}^+ + \eta_{3t} - (\hat{\mu}-\mu_0)\kappa_{3t}) + O_P(T^{-1})$$
$$= 3T^{-1/2}M_{12T}^+ + 3\sigma_0^{-2}T^{-1}\sum_{t=N+1}^{N+T}E(S_{1t}^+S_{2t}^+) + T^{-1/2}M_{03T}^+ + O_P(T^{-1}) = -6\zeta_3 + O_P(T^{-1/2}),$$

where the second-to-last equality uses Lemma B.1 and the last equality uses Lemmas B.2 and B.3, (67), and (74).
(81): For the function $L_c^*(d)$, see (15), we find

$$\sigma_0^{-2}T^{-1/2}\mathsf{D}L_c^* = \sigma_0^{-2}T^{-1/2}\sum_{t=N+1}^{N+T}(S_{0t}^+ + \eta_{0t} - (C-\mu_0)\kappa_{0t})(S_{1t}^+ + \eta_{1t} - (C-\mu_0)\kappa_{1t}),$$

with expectation given by

$$\sigma_0^{-2}T^{-1/2}\sum_{t=N+1}^{N+T}(\eta_{0t} - (C-\mu_0)\kappa_{0t})(\eta_{1t} - (C-\mu_0)\kappa_{1t}) + \sigma_0^{-2}T^{-1/2}\sum_{t=N+1}^{N+T}Cov(S_{0t}^+, S_{1t}^+)$$
$$= \sigma_0^{-2}T^{-1/2}\sum_{t=N+1}^{N+T}(\eta_{0t} - (C-\mu_0)\kappa_{0t})(\eta_{1t} - (C-\mu_0)\kappa_{1t}) = T^{-1/2}\xi_{N,T}^C(d_0).$$

The remaining derivatives give the same results as for L^*. Notice that the two factors in the sum in the score are independent so there is no term corresponding to $\tau_{N,T}$. ∎

Appendix C Proofs of main results

C.1 Proof of Theorem 1

We first show that the likelihood functions have no singularities. When $t \geq N+1$ we can use the decomposition $\pi_{t-1}(-d+1) = \pi_N(-d+1)\alpha_{N,t-1}(-d+1)$, see (37). We find in the second term of $L^*(d)$ in (12) that the factor $\pi_N(-d+1)^2$ cancels and

$$\frac{(\sum_{t=N+1}^{N+T}(\Delta_0^d X_t)\kappa_{0t}(d))^2}{\sum_{t=N+1}^{N+T}\kappa_{0t}(d)^2} = \frac{[\sum_{t=N+1}^{N+T}(\Delta_0^d X_t)\alpha_{N,t-1}(-d+1)]^2}{\sum_{t=N+1}^{N+T}\alpha_{N,t-1}(-d+1)^2}.$$

This is a differentiable function of d because $\sum_{t=N+1}^{N+T}\alpha_{N,t-1}(-d+1)^2 \geq \alpha_{N,N}(-d+1)^2 = 1$, see (37). Note, however, that $\hat{\mu}(d)$ has singularities at the points $d = 1, 2, \ldots, N$.

We next discuss the estimator $\hat{d}$. The proof for $\hat{d}_c$ is similar, but simpler because in that case $\hat{\mu}(d) = C$ does not depend on d. The arguments generally follow those of JN (2012a, Theorem 4) and Nielsen (2015, Theorem 1). To conserve space we only describe the differences in detail.

C.1.1 Existence and consistency of the estimator

The function $L^*(d)$ in (12) is the sum of squares of

$$\Delta_0^d(X_t - \hat{\mu}(d)) = \Delta_N^{d-d_0}\varepsilon_t + \eta_t(d) - (\hat{\mu}(d) - \mu_0)\kappa_{0t}(d),$$

see (51), so that we need to analyze product moments of the terms on the right-hand side, appropriately normalized. The deterministic term $\eta_t(d)$ was analyzed under the assumption of bounded initial values in JN (2012a, Lemma A.8(i)) as $D_{it}(\psi)$ with $b = d$, $i = k = 0$, and $\alpha_0 = \beta_0 = 0$, where it was shown that

$$\sup_{-1/2-\kappa \leq d-d_0 \leq \bar{d}-d_0} |\eta_t(d)| \to 0 \text{ and } \sup_{\underline{d}-d_0 \leq d-d_0 \leq -1/2-\kappa} \max_{1 \leq t \leq T} |t^{d-d_0+1/2}\eta_t(d)| \to 0 \text{ as } t \to \infty.$$

This shows that $\eta_t(d)$ is uniformly smaller than $\Delta_N^{d-d_0}\varepsilon_t$ (appropriately normalized on the intervals $-1/2-\kappa \leq d-d_0 \leq \bar{d}-d_0$ and $\underline{d}-d_0 \leq d-d_0 \leq -1/2-\kappa$), and is enough to show that in the calculation of product moments we can ignore $\eta_t(d)$, which will be done below.

The product moment of the stochastic term, $\sum_{t=N+1}^{N+T}(\Delta_N^{d-d_0}\varepsilon_t)^2$, is analyzed in Nielsen (2015) under Assumption 1 of finite fourth moment. Following that analysis, for some $0 < \kappa < 1/2$ to be determined, we divide the parameter space into intervals where $\Delta_N^{d-d_0}\varepsilon_t$ is nonstationary, "near critical", or (asymptotically) stationary according to $d - d_0 \leq -1/2 - \kappa$, $-1/2 - \kappa \leq d - d_0 \leq -1/2 + \kappa$, or $-1/2 + \kappa \leq d - d_0$.

Clearly, d_0 is contained in the interval $-1/2 + \kappa \leq d - d_0$, and we show that on this interval the contribution from the second term in the objective function

$$\begin{aligned} R_T(d) &= T^{-1}\sum_{t=N+1}^{N+T}(\Delta_N^{d-d_0}\varepsilon_t)^2 - T^{-1}\frac{[\sum_{t=N+1}^{N+T}\Delta_N^{d-d_0}\varepsilon_t\alpha_{N,t-1}(1-d)]^2}{\sum_{t=N+1}^{N+T}\alpha_{N,t-1}(1-d)^2} \\ &= T^{-1}\sum_{t=N+1}^{N+T}(\Delta_N^{d-d_0}\varepsilon_t)^2 - T^{-1}\frac{A_T(d)^2}{B_T(d)}, \end{aligned} \quad (84)$$

say, is negligible. It then follows that the objective function is only negligibly different from the objective function obtained without the parameter μ, see e.g. Nielsen (2015), and existence and consistency of $\hat{d}$ follows for the interval $d - d_0 \geq -1/2 + \kappa$.

The two intervals covering $d - d_0 \leq -1/2 + \kappa$ require a more careful analysis, which is given subsequently. Following the strategy of JN (2012a) and Nielsen (2015), we show that for any $K > 0$ there exists a (fixed) $\kappa > 0$ such that, for these intervals,

$$P(\inf R_T(d) > K) \to 1 \text{ as } T \to \infty. \tag{85}$$

This implies that $P(\hat{d} \in \{d : d - d_0 \geq -1/2 + \kappa\}) \to 1$ as $T \to \infty$, so that the relevant parameter space is reduced to $\{d : d - d_0 \geq -1/2 + \kappa\}$ on which existence and consistency has already been shown.

C.1.2 Tightness of product moments

We want to show that the remainder term, $T^{-1}A_T(d)^2/B_T(d)$, in (84) is dominated by the first term on various compact intervals. The function $B_T(d)$ is discussed below, and we want to find the supremum of the suitably normalized product moment $M_T(d) = T^{\alpha+\beta d}(\log T)^{\gamma} A_T(d)$ by considering it a continuous process on a compact interval $\mathcal{K}$; that is, we consider it a process in $\mathcal{C}(\mathcal{K})$, the space of continuous functions on $\mathcal{K}$ endowed with the uniform topology. The usual technique is then to prove that the process M_T is tight in $\mathcal{C}(\mathcal{K})$, which implies that also $\sup_{d\in\mathcal{K}} |M_T(d)|$ is tight, by the continuity of the mapping $f \mapsto \sup_{u\in\mathcal{K}} |f(u)|$, that is $\sup_{d\in\mathcal{K}} M_T(d) = O_P(1)$.

Tightness of M_T can be proved by applying Billingsley (1968, Theorem 12.3), which states that it is enough to verify the two conditions

$$EM_T(d_0)^2 \leq c, \tag{86}$$

$$E(M_T(d_1) - M_T(d_2))^2 \leq c(d_1 - d_2)^2 \text{ for } d_1, d_2 \in \mathcal{K}. \tag{87}$$

In one case we will also need the weak limit of the process M_T, and in that case we apply Billingsley (1968, Theorem 8.1), which states that if M_T is tight then convergence of the finite dimensional distributions implies weak convergence. Thus, instead of working with the processes themselves, we need only evaluate their second moments and finite dimensional distributions.

Specifically, by a Taylor series expansion of the coefficients we find

$$\begin{aligned}&\pi_m(d_0 - d_1)\alpha_{N,t+m-1}(1 - d_1) - \pi_m(d_0 - d_2)\alpha_{N,t+m-1}(1 - d_2)\\ &= -(d_1 - d_2)\{\mathsf{D}\pi_m(d_0 - d_{m,t}^*)\alpha_{N,t+m-1}(1 - d_{m,t}^*) + \pi_m(d_0 - d_{m,t}^*)\mathsf{D}\alpha_{N,t+m-1}(1 - d_{m,t}^*)\}\end{aligned}$$

for some $d_{m,t}^*$ between d_1 and d_2. It follows that if d_1 and d_2 are in the interval $\mathcal{K}$, then also $d_{m,t}^* \in \mathcal{K}$, so that any uniform bound we find for $E\mathsf{D}M_T(d)^2$ for $d \in \mathcal{K}$ will also be valid for $d_{m,t}^*$. This shows that to prove tightness of $M_T(d)$, it is enough to verify

$$\sup_{d\in\mathcal{K}} EM_T(d)^2 \leq c \text{ and } \sup_{d\in\mathcal{K}} E(\mathsf{D}M_T(d))^2 \leq c. \tag{88}$$

C.1.3 Evaluation of product moments

We evaluate product moments on intervals of the form $d \geq 1/2 - \xi$ or $d \leq 1/2 - \xi$, as well as $d - d_0 \geq -1/2 - \kappa$ or $d - d_0 \leq -1/2 - \kappa$. Some of these intervals may be empty, depending on $\underline{d}$ and $\bar{d}$, in which case the proof simplifies easily, so we proceed assuming all intervals are non-empty.

The product moment $B_T(d) = \sum_{t=N+1}^{N+T} \alpha_{N,t-1}(1-d)^2$. We first find that

$$\inf_{d\geq 0} B_T(d) \geq 1 \tag{89}$$

because $B_T(d) \geq \alpha_{N,N}(1-d) = 1$.

Next there are constants c_1, c_2 such that

$$0 < c_1 \leq \sup_{\underline{d}\leq d\leq 1/2-\xi} T^{2d-1}B_T(d) \leq c_2 < \infty. \tag{90}$$

This follows from (41) because

$$T^{2d-1}\sum_{t=N+1}^{N+T} \alpha_{N,j}(1-d)^2 = (\frac{N!}{\Gamma(1-d+N)})^2 T^{-1}\sum_{t=N+1}^{N+T}(\frac{t}{T})^{-2d}(1+\epsilon_{2t}(d)),$$

which converges uniformly in $d \in [\underline{d}, 1/2-\xi]$ to $(N!/\Gamma(1-d+N))^2/(1-2d)$ which is bounded between c_1 and c_2 because $2\xi \leq 1-2d \leq 1-2\underline{d}$.

Finally,

$$\inf_{1/2-\xi\leq d\leq 1/2+\xi} T^{2d-1}B_T(d) \geq c\frac{1-((N+1)/T)^{2\xi}}{2\xi}, \tag{91}$$

which again follows from (41) because $(t/T)^{-2d} \geq (t/T)^{2\xi-1}$ which implies that

$$T^{2d-1}B_T(d) \geq cT^{-1}\sum_{t=N+1}^{N+T}(\frac{t}{T})^{2\xi-1} \geq cT^{-2\xi}\int_{N+1}^{T} u^{2\xi-1}du = c\frac{1-((N+1)/T)^{2\xi}}{2\xi}.$$

The product moment $A_T(d) = \sum_{t=N+1}^{N+T} \Delta_N^{d-d_0}\varepsilon_t \alpha_{N,t-1}(1-d)$. We find that

$$A_T(d) = \sum_{t=N+1}^{N+T} \varepsilon_t \phi_{N,t}(d), \quad \phi_{N,t}(d) = \sum_{m=0}^{N+T-t} \pi_m(d_0-d)\alpha_{N,t+m-1}(1-d).$$

From (38) and (39) we find $|\phi_{N,t}(d)| \leq c\sum_{m=0}^{N+T-t} m^{d_0-d-1}(t+m)^{-d}$, and

$$EA_T(d)^2 = \sigma_0^2 \sum_{t=N+1}^{N+T} \phi_{N,t}(d)^2 \leq c\sum_{t=N+1}^{N+T}\{\sum_{m=0}^{N+T-t} m^{d_0-d-1}(t+m)^{-d}\}^2,$$

while $\mathsf{D}A_T(d)$ contains an extra factor $\log(m(t+m))$.

We give in Table 2 the bounds for $EA_T(d)^2$ for various intervals and normalizations. These follow from first using the inequalities $(t+m)^{-d} \leq (t+m)^{-1/2+\xi}$ when $d \geq 1/2-\xi$ and $T^d(m+t)^{-d} \leq ((m+t)/T)^{-1/2+\xi}$ when $d \leq 1/2-\xi$, and similarly for $d-d_0$. We then apply the result that

$$T^{-1}\sum_{t=N+1}^{N+T}\{T^{-1}\sum_{m=0}^{N+T-t}(\frac{m}{T})^{-1/2+\kappa}(\frac{t+m}{T})^{-1/2+\xi}\}^2 = O(1)$$

because the left-hand side converges to $\int_0^1\{\int_0^{1-v} u^{-1/2+\kappa}(u+v)^{-1/2+\xi}du\}^2 dv$.

Table 2: Bounds for $A_T(d)$

Second moment	d	$d-d_0$	Upper bound on second moment	Order
$EA_T(d)^2$	$\geq 1/2-\xi$	$\geq -1/2-\kappa$	$\sum(\sum m^{-1/2+\kappa}(t+m)^{-1/2+\xi})^2$	$T^{1+2\xi+2\kappa}$
$ET^{2d}A_T(d)^2$	$\leq 1/2-\xi$	$\geq -1/2-\kappa$	$\sum(\sum m^{-1/2+\kappa}(\frac{t+m}{T})^{-1/2+\xi})^2$	$T^{2+2\kappa}$
$ET^{2(d-d_0+1)}A_T(d)^2$	$\geq 1/2-\xi$	$\leq -1/2-\kappa$	$\sum(\sum (\frac{m}{T})^{-1/2+\kappa}(t+m)^{-1/2+\xi})^2$	$T^{2+2\xi}$
$ET^{4d-2d_0+2}A_T(d)^2$	$\leq 1/2-\xi$	$\leq -1/2-\kappa$	$\sum(\sum (\frac{m}{T})^{-1/2+\kappa}(\frac{t+m}{T})^{-1/2+\xi})^2$	T^3

Note: Uniform upper bounds on the normalized second moment of $A_T(d)$ for different restrictions on d and $d-d_0$. The bounds are also valid if we replace κ by $-\kappa$ or ξ by $-\xi$.

Table 3: Bounds for $C_{T,M}(d)$

Second moment	d	$d-d_0$	Upper bound on second moment	Order
$EC_{T,M}(d)^2$	$\geq 1/2-\xi$	$\geq -1/2-\kappa$	$\sum(\sum m^{-1/2+\kappa}(t+m)^{-1/2+\xi})^2$	$M^{1+2\kappa}T^{2\xi}$
$ET^{2d}C_{T,M}(d)^2$	$\leq 1/2-\xi$	$\geq -1/2-\kappa$	$\sum(\sum m^{-1/2+\kappa}(\frac{t+m}{T})^{-1/2+\xi})^2$	$M^{1+2\kappa}T$

Note: Uniform upper bounds on the normalized second moment of $C_{T,M}(d)$ for different restrictions on d and $d-d_0$.

The product moment $C_{M,T} = \sum_{t=N+M+1}^{N+T}\{\sum_{n=0}^{M-1}\pi_n(d_0-d)\varepsilon_{t-n}\}\alpha_{N,t-1}(1-d)$. Now we analyze another product moment, which we find by truncating the sum $\Delta_N^{d-d_0}\varepsilon_t = \sum_{n=0}^{t-N-1}\pi_n(d_0-d)\varepsilon_{t-n}$ at $M=T^\alpha$ for $\alpha<1$, and define

$$C_{T,M}(d) = \sum_{t=N+2}^{N+T}\varepsilon_t\psi_{N,M,t}(d),\quad \psi_{N,M,t}(d) = \sum_{m=\max(N+M+1-t,0)}^{\min(M-1,N+T-t)}\pi_m(d_0-d)\alpha_{N,t+m-1}(1-d). \quad (92)$$

The coefficients are the same as for $A_T(d)$, but the sum $\psi_{N,M,t}(d)$ only contains at most M terms. We give in Table 3 the bounds for the second moment of $C_{T,M}(d)$, which are derived using the same methods as for $A_T(d)$.

We now apply the above evaluations to study the objective function in the three intervals $d-d_0\geq -1/2+\kappa, -1/2-\kappa\leq d-d_0\leq -1/2+\kappa$, and $-1/2-\kappa\leq d-d_0$.

C.1.4 The stationarity interval: $\{d-d_0\geq -1/2+\kappa\}\cap\mathcal{D}$

We want to show that

$$\sup_{\{d-d_0\geq -1/2+\kappa\}\cap\mathcal{D}}|T^{-1}\frac{A_T(d)^2}{B_T(d)}| = o_P(1),$$

and consider two cases because of the different behavior of $B_T(d)$.

Case 1: If $d\geq 1/2-\xi$ we let $\mathcal{K}=\{d\geq 1/2-\xi, d-d_0\geq -1/2+\kappa\}\cap\mathcal{D}$ and use (89) to eliminate $B_T(d)$ and focus on $A_T(d)$. From Table 2 we find using $(\xi,-\kappa)$ that $\sup_{\mathcal{K}}E(T^{-1}A_T(d)^2) = O(T^{2\xi-2\kappa})$. For the derivative we get an extra factor $\log T$ in the coefficients and find $\sup_{\mathcal{K}}E(T^{-1}(\mathsf{D}A_T(d))^2) = O((\log T)^2T^{2\xi-2\kappa})$.

It then follows from (86) and (87) that $M_T(d) = T^{-1/2+\kappa-\xi}(\log T)^{-1}A_T(d)$ is tight. Because convergence in probability and tightness implies uniform convergence in probability it follows that

$$\sup_{d\in\mathcal{K}}T^{-1}A_T(d)^2 = O_P(T^{-2\kappa+2\xi}(\log T)^2) = o_P(1) \text{ for } \xi<\kappa.$$

Case 2: If $d \leq 1/2-\xi$ we define $\mathcal{K} = \{d \leq 1/2-\xi,\, d-d_0 \geq -1/2+\kappa\} \cap \mathcal{D}$. From (90) we find that $\sup_{\mathcal{K}} E(T^{-1}A_T(d)^2/B_T(d)) \leq c\sup_{\mathcal{K}} E(T^{2d-2}A_T(d)^2)$. From Table 2 we then find for $(\xi,-\kappa)$ that $\sup_{\mathcal{K}} E(T^{2d-2}A_T(d)^2) = O(T^{-2\kappa})$. For the derivative we get an extra factor $\log T$. Thus, $\sup_{\mathcal{K}} E(\mathsf{D}T^{d-1/2}A_T(d))^2 = O((\log T)^2T^{-2\kappa})$ and $(\log T)^{-1}T^{\kappa}T^{d-1}A_T(d)$ is tight, so that

$$\sup_{d\in\mathcal{K}}|T^{-1}\frac{A_T(d)^2}{B_T(d)}| \leq c\sup_{d\in\mathcal{K}} T^{2d-2}A_T(d)^2 = O_P((\log T)^2T^{-2\kappa}) = o_P(1).$$

C.1.5 The critical interval: $\{-1/2-\kappa \leq d-d_0 \leq -1/2+\kappa\} \cap \mathcal{D}$

For this interval we show that (85) holds by setting κ sufficiently small. As in JN (2012a) and Nielsen (2015) we apply a truncation argument. With $M = T^{\alpha}$, for some $\alpha > 0$ to be chosen below, let

$$\Delta_N^{d-d_0}\varepsilon_t = \sum_{n=0}^{M-1}\pi_n(d_0-d)\varepsilon_{t-n} + \sum_{n=M}^{t-n-1}\pi_n(d_0-d)\varepsilon_{t-n} = w_{1t}+w_{2t}, t \geq M+N+1,$$

such that the objective function (84) is

$$R_T(d) = T^{-1}\sum_{t=N+1}^{N+T}(\Delta_N^{d-d_0}\varepsilon_t - \alpha_{N,t-1}(1-d)\frac{A_T(d)}{B_T(d)})^2 \geq T^{-1}\sum_{t=N+M+1}^{N+T}(w_{1t}+v_t)^2\,, \qquad (93)$$

where $v_t = w_{2t} - \alpha_{N,t-1}(1-d)\frac{A_T(d)}{B_T(d)}$. We further find that

$$R_T(d) \geq T^{-1}\sum_{t=N+M+1}^{N+T}w_{1t}^2 + 2T^{-1}\sum_{t=N+M+1}^{N+T}w_{1t}w_{2t} - 2T^{-1}C_{T,M}(d)\frac{A_T(d)}{B_T(d)}, \qquad (94)$$

where $C_{T,M}(d)$ is given by (92). The first two terms in (94) are analyzed in Nielsen (2015), where it is shown that by setting κ sufficiently small, the first term can be made arbitrarily large while the second is $o_P(1)$, uniformly on $|d-d_0+1/2| \leq \kappa_1$ for some fixed $\kappa_1 > \kappa$. Thus it remains to be shown that the third term of (94) is asymptotically negligible, uniformly on the critical interval, that is,

$$\sup_{|d-d_0+1/2|\leq\kappa_1}|T^{-1}C_{T,M}(d)\frac{A_T(d)}{B_T(d)}| = o_P(1).$$

We consider two cases depending on d.

Case 1: Let $\mathcal{K} = \{1/2-\xi \leq d, -1/2-\kappa_1 \leq d-d_0 \leq -1/2+\kappa_1\} \cap \mathcal{D}$. From (89) we have $B_T(d)^{-1} \leq 1$ and from Table 2 we find for (ξ,κ_1) that $\sup_{\mathcal{K}} EA_T(d)^2 = O(T^{1+2\kappa_1+2\xi})$ and $\sup_{\mathcal{K}} E(\mathsf{D}A_T(d))^2 = O((\log T)^2T^{1+2\kappa_1+2\xi})$ such that $\sup_{\mathcal{K}}|A_T(d)| = O_P((\log T)T^{1/2+\kappa_1+\xi})$.

From Table 3 for (ξ,κ_1) we then find $\sup_{\mathcal{K}} EC_{T,M}(d)^2 = O(M^{1+2\kappa_1}T^{2\xi}) = O(T^{\alpha(1+2\kappa_1)+2\xi})$ and also $\sup_{\mathcal{K}} E(\mathsf{D}C_{T,M}(d))^2 = O((\log T)^2M^{1+2\kappa_1}T^{2\xi}) = O((\log T)^2T^{\alpha(1+2\kappa_1)+2\xi})$, such that $\sup_{\mathcal{K}}|C_{T,M}(d)| = O_P((\log T)T^{\alpha(1/2+\kappa_1)+\xi})$. This shows that

$$\sup_{d\in\mathcal{K}}|T^{-1}C_{T,M}(d)\frac{A_T(d)}{B_T(d)}| = O_P((\log T)^2T^{\alpha(1/2+\kappa_1)-(1/2-\kappa_1-2\xi)}) = o_P(1)$$

for $\alpha < (1/2-\kappa_1-2\xi)/(1/2+\kappa_1)$.

Case 2: Let $\mathcal{K} = \{d \leq 1/2-\xi, -1/2-\kappa_1 \leq d-d_0 \leq -1/2+\kappa_1\} \cap \mathcal{D}$. From (90) we find $\sup_{\mathcal{K}} |T^{1-2d}B_T(d)^{-1}| \leq c$, and we find from Table 2 that $\sup_{\mathcal{K}} E(T^{d-1}A_T(d))^2 = O(T^{2\kappa_1})$ and therefore $\sup_{\mathcal{K}} E(\mathsf{D}T^{d-1}A_T(d))^2 = O((\log T)^2 T^{2\kappa_1})$. From Table 3 we get $\sup_{\mathcal{K}} E(T^{d-1}C_{T,M}(d))^2 = O(M^{1+2\kappa_1}T^{-1}) = O(T^{\alpha(1+2\kappa_1)-1})$ and $\sup_{\mathcal{K}} E(\mathsf{D}T^{d-1}C_{T,M}(d))^2 = O((\log T)^2 M^{1+2\kappa_1}T^{-1}) = O((\log T)^2 T^{\alpha(1+2\kappa_1)-1})$. Hence

$$\sup_{d\in\mathcal{K}} |T^{-1}C_T(d)\frac{A_T(d)}{B_T(d)}| = O_P((\log T)^2 T^{\alpha(1/2+\kappa_1)-(1/2-\kappa_1)}) = o_P(1)$$

for $\alpha < (1/2-\kappa_1)/(1/2+\kappa_1)$.

C.1.6 The nonstationarity interval: $\{d-d_0 \leq -1/2-\kappa\} \cap \mathcal{D}$

We give different arguments for different intervals of d, and we distinguish three cases.

Case 1: Let $\mathcal{K} = \{1/2+\xi \leq d, d-d_0 \leq -1/2-\kappa\} \cap \mathcal{D}$. For this interval the main term of $R_T(d)$ in (84) has been shown by Nielsen (2015) to satisfy (85), and it is sufficient to show, with the normalization relevant to the nonstationarity interval, that

$$\sup_{d\in\mathcal{K}} T^{2(d-d_0)}\frac{A_T(d)^2}{B_T(d)} = o_P(1). \tag{95}$$

We use (89) to evaluate $B_T(d)^{-1} \leq 1$ and find from Table 2 for $(-\xi,\kappa)$ that $\sup_{\mathcal{K}} E(T^{2(d-d_0)}A_T(d)^2) = O(T^{-2\xi})$ so that $E(\mathsf{D}T^{d-d_0}A_T(d))^2 = O((\log T)^2 T^{-2\xi})$, which shows that

$$\sup_{d\in\mathcal{K}} T^{2(d-d_0)}\frac{A_T(d)^2}{B_T(d)} = O_P((\log T)^2 T^{-2\xi}) = o_P(1).$$

Case 2: Let $\mathcal{K} = \{1/2-\xi \leq d \leq 1/2+\xi, d-d_0 \leq -1/2-\kappa\} \cap \mathcal{D}$. Again the main term of $R_T(d)$ in (84) has been shown by Nielsen (2015) to satisfy (85), and we therefore want to show that

$$\sup_{d\in\mathcal{K}} |T^{2(d-d_0)}\frac{A_T(d)^2}{B_T(d)}| \leq \frac{\sup_{d\in\mathcal{K}} T^{2(d-d_0)}T^{2d-1}A_T(d)^2}{\inf_{d\in\mathcal{K}} |T^{2d-1}B_T(d)|} = O_P(1), \tag{96}$$

but can be made arbitrarily small by choosing ξ sufficiently small.

It follows from (91) that the denominator $T^{2d-1}B_T(d)$ of (96) can be made arbitrarily large by choosing ξ sufficiently small, because $(1-((N+1)/T)^{2\xi})/2\xi \to \log(T/(N+1))$ for $\xi \to 0$. We next prove that the numerator of (96) is uniformly $O_P(1)$, which proves the result for Case 2. From Table 2 for (κ,ξ) we find $\sup_{\mathcal{K}} E(T^{2(d-d_0)}T^{2d-1}A_T(d)^2) = O(1)$. The derivative of $T^{d-d_0}T^d\phi_{N,t}(d)$ is bounded by

$$T^{-1}\sum_{m=N+1}^{N+T-[Tv]}(\frac{m}{T})^{d_0-d-1}(\frac{[Tv]+m}{T})^{-d}\log(\frac{m}{T}(\frac{m+[Tv]}{T})),$$

which converges to $\int_0^{1-v} u^{d_0-d-1}(v+u)^{-d}\log(u(u+v))du < \infty$ for $d \in \mathcal{K}$. Thus no extra $\log T$ factor is needed in this case, and we find that $T^{d-d_0}T^{d-1/2}A_T(d)$ is tight, which proves that $\sup_{\mathcal{K}} |T^{d-d_0}T^{d-1/2}A_T(d)| = O_P(1)$.

Case 3: Finally, we assume $\underline{d} \leq d \leq 1/2-\xi$ and $\underline{d}-d_0 \leq d-d_0 \leq -1/2-\kappa$. We note that on this set the term $T^{1-2d}B_T(d)^{-1}$ is uniformly bounded and uniformly bounded away from zero, see (90), so we factor it out of the objective function. We thus analyze the objective function

$$R_T^*(d) = T^{2d-2} \sum_{t,s=N+1}^{N+T} \left((\Delta_N^{d-d_0} \varepsilon_t)^2 \alpha_{N,s-1}(1-d)^2 - (\Delta_N^{d-d_0} \varepsilon_s) \alpha_{N,s-1}(1-d)(\Delta_N^{d-d_0} \varepsilon_t) \alpha_{N,t-1}(1-d) \right).$$

The most straightforward approach would be to obtain the weak limit of $T^{2(d-d_0+1/2)} R_T^*$ from the weak convergence of $T^{d-d_0+1/2} \Delta_N^{d-d_0} \varepsilon_t$ on $d - d_0 \leq 1/2 - \kappa$ and the uniform convergence of $T^d \alpha_{N,[Tu]-1}(1-d) \to \frac{N!}{\Gamma(1-d+N)} u^{-d}$. However, the former would require the existence of $E|\varepsilon_t|^q$ for $q > 1/(d - d_0 - 1/2) \geq 1/\kappa$ with κ arbitrarily small, see JN (2012b), which we have not assumed in Assumption 1. We therefore introduce $\Delta_N^{d-d_0-1} \varepsilon_t$, the cumulation of $\Delta_N^{d-d_0} \varepsilon_t$, to increase the fractional order sufficiently far away from the critical value $d - d_0 = -1/2$, so the number of moments needed is $q > 1/(1 + \kappa)$. To this end we first prove the following.

Lemma C.1 *Let $a_t, b_t, t = 1, \ldots, T$, be real numbers and $A_t = \sum_{s=1}^t a_s, B_t = \sum_{s=1}^t b_s$. Then*

$$\frac{2}{T(T-1)} \sum_{t,s=1}^{T} (a_t^2 b_s^2 - a_t b_s a_s b_t) \geq \left(\frac{2}{T(T-1)} \sum_{t=1}^{T-1} (b_t A_T - b_t A_t - b_{t+1} A_t) \right)^2.$$

Proof. We first find

$$\sum_{t=1}^{T} \sum_{s=1}^{T} (a_t^2 b_s^2 - a_t b_s a_s b_t) = \sum_{1 \leq s < t \leq T} (a_t^2 b_s^2 + a_s^2 b_t^2 - 2 a_t b_s a_s b_t) = \sum_{1 \leq s < t \leq T} (a_t b_s - a_s b_t)^2.$$

The proof is then completed by using the Cauchy-Schwarz inequality,

$$\left(\frac{2}{T(T-1)} \sum_{1 \leq s < t \leq T} (a_t b_s - a_s b_t) \right)^2 \leq \frac{2}{T(T-1)} \sum_{1 \leq s < t \leq T} (a_t b_s - a_s b_t)^2,$$

together with $\sum_{1 \leq s < t \leq T} (a_t b_s - a_s b_t) = \sum_{s=1}^{T-1} b_s (A_T - A_s) - \sum_{t=2}^{T} b_t A_{t-1}$. ■

Applying Lemma C.1 to $T^{2-2d} \frac{2}{T(T-1)} R_T^*(d)$ we find that for $a_t = \Delta_N^{d-d_0} \varepsilon_t$ and $b_t = \alpha_{N,t-1}(1-d)$ it holds that $R_T^*(d) \geq 2 T^{2(d_0-d)-1} Q_T(d)^2$ where

$$Q_T(d) = T^{2d-d_0-1/2} T^{-1} \sum_{t=N+1}^{N+T-1} (\alpha_{N,t-1}(1-d)(\Delta_N^{d-d_0-1} \varepsilon_{N+T-1}) - (\alpha_{N,t-1}(1-d) + \alpha_{N,t}(1-d))(\Delta_N^{d-d_0-1} \varepsilon_t)). \tag{97}$$

Following the arguments in JN (2012a) and Nielsen (2015), we show that $Q_T(d)$ converges weakly (in the space of continuous functions of d) to a random variable that is positive almost surely.

Let $\mathcal{K} = \{d - d_0 - 1 \leq -3/2 - \kappa, d \leq -1/2 - \xi\} \cap \mathcal{D}$. Assumption 1 ensures that we have enough moments, $q > \max(2, 1/(1+\kappa))$, to apply the fractional functional central limit theorem, e.g. Marinucci and Robinson (2000, Theorem 1), and find for each $d \in \mathcal{K}$ that

$$T^d \alpha_{N,[Tu]-1}(1-d) T^{d-d_0-1/2} \Delta_N^{d-d_0-1} \varepsilon_{[Tu]} \Rightarrow \frac{N!}{\Gamma(N+1-d)} u^{-d} W_{d_0-d}(u) \text{ as } T \to \infty \text{ on } D[0,1],$$

where "$\Rightarrow$" denotes weak convergence and $W_{d_0-d}(u) = (\Gamma(d_0-d+1))^{-1}\int_0^u (u-s)^{d_0-d}\mathsf{d}W(s)$ denotes fractional Brownian motion (of type II) and W denotes Brownian motion generated by ε_t.

Because the integral is a continuous mapping of $D[0,1]$ to $\mathbb{R}$ it holds that

$$Q_T(d) \Rightarrow Q(d) = \frac{N!}{\Gamma(N+1-d)}\int_0^1 u^{-d}(W_{d_0-d}(1) - 2W_{d_0-d}(u))\mathsf{d}u \text{ as } T\to\infty \tag{98}$$

for any fixed $d \in \mathcal{K}$. We can establish tightness of the continuous process $Q_T(d)$ by evaluating the second moment, using the methods above. For all terms we see that it has the same form as $A_T(d)$ except that $(d-d_0, d)$ is replaced by $(d-d_0-1, d)$ and hence the result follows as the results for $A_T(d)$. This establishes tightness of $Q_T(d)$ and hence strengthens the convergence in (98) to weak convergence in the space of continuous functions of d on $\mathcal{K}$ endowed with the uniform topology.

It thus holds that

$$\inf_{d\in\mathcal{K}} R_T^*(d) \geq 2\inf_{d\in\mathcal{K}} T^{2(d_0-d)-1}Q_T(d)^2 + o_P(1) \geq 2T^{2\kappa}\inf_{d\in\mathcal{K}} Q_T(d)^2 + o_P(1),$$

where $\inf_{d\in\mathcal{K}} Q_T(d)^2 > 0$ almost surely and $\kappa > 0$. It follows that, for any $K > 0$,

$$P(\inf_{d\in\mathcal{K}} R_T^*(d) > K) \to 1 \text{ as } T\to\infty,$$

which shows (85) and hence proves the result for Case 3.

C.1.7 Asymptotic normality of the estimator

To show asymptotic normality of $\hat{d}$ we apply the usual expansion of the score function,

$$0 = \mathsf{D}L^*(\hat{d}) = \mathsf{D}L^*(d_0) + (\hat{d}-d_0)\mathsf{D}^2L^*(d^*),$$

where d^* is an intermediate value satisfying $|d^*-d_0| \leq |\hat{d}-d_0| \xrightarrow{P} 0$. The product moments in $\mathsf{D}^2L^*(d)$ are shown in JN (2010, Lemma C.4) and JN (2012a, Lemma A.8(i)) to be tight, or equicontinuous, in a neighborhood of d_0, so that we can apply JN (2010, Lemma A.3) to conclude that $\mathsf{D}^2L^*(d^*) = \mathsf{D}^2L^*(d_0) + o_P(1)$, and we therefore analyze $\mathsf{D}L^*(d_0)$ and $\mathsf{D}^2L^*(d_0)$. From Lemma B.4 we find that $\sigma_0^{-2}T^{-1/2}\mathsf{D}L^*(d_0) = M_{01T}^+ + O_P(T^{-1/2})$ and $\sigma_0^{-2}T^{-1}\mathsf{D}^2L^*(d_0) = \zeta_2 + O_P(T^{-1/2}) = \pi^2/6 + O_P(T^{-1/2})$, and the result follows from Lemmas B.2 and B.3.

C.2 Proof of Theorem 2

First we note that, as in the proof of Theorem 1 in Appendix C.1.7, we can apply JN (2010, Lemma A.3) to conclude that $\mathsf{D}^3L^*(d^*) = \mathsf{D}^3L^*(d_0) + o_P(1)$. We thus insert the expressions (75), (76), and (77) into the expansion (17) and find

$$T^{1/2}(\hat{d}-d_0) = -\frac{A_0+T^{-1/2}A_1}{B_0+T^{-1/2}B_1} - \frac{1}{2}T^{-1/2}(\frac{A_0+T^{-1/2}A_1}{B_0+T^{-1/2}B_1})^2\frac{C_0}{B_0+T^{-1/2}B_1} + o_P(T^{-1/2}),$$

which, using the expansion $1/(1+z) = 1-z+z^2+\ldots$, reduces to

$$T^{1/2}(\hat{d}-d_0) = -\frac{A_0}{B_0} - T^{-1/2}(\frac{A_1}{B_0} - \frac{A_0B_1}{B_0^2} + \frac{1}{2}\frac{A_0^2C_0}{B_0^3}) + o_P(T^{-1/2}).$$

We find that $E(A_0) = E(M_{01T}^+) = 0$, so the bias of $T(\hat{d} - d_0)$ is, from (78)–(80),

$$-(\frac{E(A_1)}{B_0} - \frac{E(A_0 B_1)}{B_0^2} + \frac{1}{2}\frac{E(A_0^2)C_0}{B_0^3}) + o(1) \tag{99}$$
$$= -(\frac{\xi_{N,T}(d_0) + \tau_{N,T}(d_0)}{\zeta_2} - \frac{E(M_{01T}^+(M_{11T}^+ + M_{02T}^+)) + 3E(M_{01T}^{+2})\zeta_3\zeta_2^{-1}}{\zeta_2^2}) + o(1).$$

From Lemma B.2,

$$E(M_{01T}^+(M_{11T}^+ + M_{02T}^+)) + 3E(M_{01T}^{+2})\zeta_3\zeta_2^{-1} = -4\zeta_3 - 2\zeta_3 + 3\zeta_3 = -3\zeta_3,$$

see (66)–(68), so that we get the final result $-(\xi_{N,T}(d_0) + \tau_{N,T}(d_0) + 3\zeta_3\zeta_2^{-1})\zeta_2^{-1} + o(1)$.

For the estimator $\hat{d}_c$ we get the expansion (99), but use (81) instead of (80).

C.3 Proof of Corollary 1

We suppress the argument d and want to evaluate $\xi_{N,T}$ and $\xi_{N,T}^C$, see (21) and (22). From (57) and (58) we find that $\langle\eta_0, \eta_1\rangle_T$, $\langle\eta_0, \kappa_1\rangle_T$, $\langle\eta_1, \kappa_0\rangle_T$, and $\langle\kappa_1, \kappa_0\rangle_T$ are all bounded by $(1+N)^{-\min(d,2d-1)+\epsilon}$, which shows the result for $\xi_{N,T}^C$. To find the result for $\xi_{N,T}$, it only remains to be shown that $\langle\eta_0, \kappa_0\rangle_T / \langle\kappa_0, \kappa_0\rangle_T$ is bounded. We find from (62) that $|\eta_{0t}(d)| \leq c\sum_{j=0}^{N_0-1} |\pi_{t+j}(-d)|$. We apply (37) and note that, for a given d and $t > N > d$, the coefficients $\pi_{t+j}(-d) = \pi_N(-d)\alpha_{N,t+j}(-d) = \pi_N(-d)\prod_{i=N+1}^{t+j}(1-(d+1)/i)$ are all of the same sign for $j \geq 0$. If this is positive, we have, see (47),

$$\sum_{j=0}^{N_0-1} |\pi_{t+j}(-d)| \leq \sum_{j=0}^{\infty} \pi_{t+j}(-d) = -\pi_{t-1}(-d+1) > 0$$

because $t - 1 \geq N$, and a similar relation holds if the coefficients are negative. Thus, $|\eta_{0t}(d)| \leq c|\kappa_{0t}(d)|$ and therefore

$$|\langle\eta_0, \kappa_0\rangle_T| = \sigma_0^{-2}|\sum_{t=N+1}^{N+T} \eta_{0t}(d)\kappa_{0t}(d)| \leq c\sigma_0^{-2}\sum_{t=N+1}^{N+T} \kappa_{0t}(d)^2 = c\langle\kappa_0, \kappa_0\rangle_T.$$

C.4 Proof of Theorem 3

(i): We note that, because $t \geq N+1$, we have $\kappa_{0t}(d) = \pi_{t-1}(-d+1) = 0$ for $d = 1, \ldots, N$. Similarly, because $t + n \geq N + 1$ for $n \geq 0$, we have

$$\eta_{0t}(d) = \sum_{n=0}^{N_0-1} \pi_{t+n}(-d)(\mu_0 - X_{-n}) = 0 \text{ for } d = 0, 1, \ldots, N+1,$$

and hence $\langle\eta_0, \eta_1\rangle_T = \langle\eta_0, \kappa_1\rangle_T = \langle\eta_1, \kappa_0\rangle_T = \langle\kappa_1, \kappa_0\rangle_T = 0$ for $d = 1, \ldots, N$. This implies that $\xi_{N,T}$ and $\xi_{N,T}^C$ are zero.

(ii): Next assume $d = 1$. The case with $N \geq 1$ is covered by part (i), so we only need to show the result for $N = 0$. For $N = 0$ we have $\kappa_{0t}(1) = \pi_{t-1}(0) = 1_{(t=1)}$ and $\kappa_{1t}(1) = -\mathsf{D}\pi_{t-1}(0) = -(t-1)^{-1}1_{\{t-1\geq 1\}}$, see (44). From (18) we find $\eta_{0t}(1) = \sum_{n=-N_0+1}^{0} 1_{(t-n=1)}(X_n - \mu_0) = 1_{(t=1)}(X_0 - \mu_0)$, whereas $\eta_{1t}(1)$ is non-zero only for $t \geq 2$ because otherwise the summation over k in (19) is empty. Thus, $\eta_{0t}(1)$ and $\kappa_{0t}(1)$ are non-zero only if $t = 1$, but $\eta_{1t}(1)$ and $\kappa_{1t}(1)$ are non-zero only if $t \geq 2$, and therefore $\xi_{0,T}^C(1) = \xi_{0,T}(1) = 0$.

(iii): From (37) it follows that $\alpha_{N,t}(-d+1)|_{d=N+1} = \prod_{i=N+1}^{t}(i-N-1)/i = 0$ for $t \geq N+1$ and therefore (71) shows that $\tau_{N,T}(d) = 0$ for $d = N+1$.

C.5 Proof of Theorem 4

(27): For $N_0 = 0$ we find from (18) that $\eta_{0t}(d_0) = \sum_{n=0}^{-1} \pi_{t+n}(-d_0)(\mu_0 - X_{-n}) = 0$, and that is enough to show that $\xi_{N,T}(d_0) = 0$, see (21).
(28) and (29): We also find for $\eta_{0t}(d_0) = 0$ that $\xi^C_{N,T}(d_0)$ simplifies to

$$\xi^C_{N,T}(d_0) = \sigma_0^{-2} \sum_{t=N+1}^{N+T} (-(C-\mu_0)\kappa_{0t})(\eta_{1t}-(C-\mu_0)\kappa_{1t}) = -(C-\mu_0)(\langle\kappa_0,\eta_1\rangle_T-(C-\mu_0)\langle\kappa_0,\kappa_1\rangle_T).$$

(30): The result follows from (27) and (70).
(31): If further $N = 0$, then both summations over n in (19) are empty, and hence zero, such that $\eta_{1t}(d_0) = 0$. It then follows from (28) that $\xi^C_{N,T}(d_0) = (C - \mu_0)^2\langle\kappa_0, \kappa_1\rangle_T$, which can be replaced by its limit, see (59).

References

1. Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions*, National Bureau of Standards, Washington D.C.
2. Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001). The distribution of daily realized stock return volatility. *Journal of Financial Economics* **61**, pp. 43–76.
3. Askey, R. (1975). *Orthogonal Polynomials and Special Functions*, SIAM, Philadelphia.
4. Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1987). *Regular Variation*, Cambridge University Press, Cambridge.
5. Bollerslev, T., Osterrieder, D., Sizova, N., and Tauchen, G. (2013) Risk and return: long-run relationships, fractional cointegration, and return predictability. *Journal of Financial Economics* **108**, pp. 409–424.
6. Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis, Forecasting, and Control,* Holden-Day, San Francisco.
7. Byers, J. D., Davidson, J., and Peel, D. (1997). Modelling political popularity: an analysis of long-range dependence in opinion poll series. *Journal of the Royal Statistical Society, Series A* **160**, pp. 471–490.
8. Carlini, F., Manzoni, M., and Mosconi, R. (2010). The impact of supply and demand imbalance on stock prices: an analysis based on fractional cointegration using Borsa Italiana ultra high frequency data. Working paper, Politecnico di Milano.
9. Davidson, J. and Hashimzade, N. (2009). Type I and type II fractional Brownian motions: a reconsideration. *Computational Statistics & Data Analysis* **53**, pp. 2089–2106.
10. Dolado, J. J., Gonzalo, J., and Mayoral, L. (2002). A fractional Dickey-Fuller test for unit roots. *Econometrica* **70**, pp. 1963–2006.
11. Doornik, J. A. and Ooms, M. (2003). Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models. *Computational Statistics & Data Analysis* **42**, pp. 333–348.
12. Giraitis, L. and Taqqu, M. (1998). Central limit theorems for quadratic forms with time domain conditions. *Annals of Probability* **26**, pp. 377–398.
13. Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and its Application*, Academic Press, New York.

14. Hualde, J. and Robinson, P. M. (2011). Gaussian pseudo-maximum likelihood estimation of fractional time series models. *Annals of Statistics* **39**, pp. 3152–3181.
15. Jensen, A. N. and Nielsen, M. Ø. (2014). A fast fractional difference algorithm. *Journal of Time Series Analysis* **35**, pp. 428–436.
16. Johansen, S. (2008). A representation theory for a class of vector autoregressive models for fractional processes. *Econometric Theory* **24**, pp. 651–676.
17. Johansen, S. and Nielsen, M. Ø. (2010). Likelihood inference for a nonstationary fractional autoregressive model. *Journal of Econometrics* **158**, pp. 51–66.
18. Johansen, S. and Nielsen, M. Ø. (2012a). Likelihood inference for a fractionally cointegrated vector autoregressive model. *Econometrica* **80**, pp. 2667–2732.
19. Johansen, S. and Nielsen, M. Ø. (2012b). A necessary moment condition for the fractional functional central limit theorem. *Econometric Theory* **28**, pp. 671–679.
20. Lawley, D. N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* **43**, pp. 295–303.
21. Li, W. K. and McLeod, A. I. (1986). Fractional time series modelling. *Biometrika* **73**, pp. 217–221.
22. Lieberman, O. and Phillips, P. C. B. (2004). Expansions for the distribution of the maximum likelihood estimator of the fractional difference parameter. *Econometric Theory* **20**, pp. 464–484.
23. Marinucci, D. and Robinson, P. M. (2000). Weak convergence of multivariate fractional processes. *Stochastic Processes and their Applications* **86**, pp. 103–120.
24. Nielsen, M. Ø. (2014). Asymptotics for the conditional-sum-of-squares estimator in multivariate fractional time series models. *Journal of Time Series Analysis* **36**, pp. 154–188.
25. Osterrieder, D. and Schotman, P. C. (2011). Predicting returns with a co-fractional VAR model. CREATES research paper, Aarhus University.
26. Robinson, P. M. (1994). Efficient tests of nonstationary hypotheses, *Journal of the American Statistical Association* **89**, pp. 1420–1437.
27. Rossi, E. and Santucci de Magistris, P. (2013). A no-arbitrage fractional cointegration model for futures and spot daily ranges. *Journal of Futures Markets* **33**, pp. 77–102.
28. Sowell, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models, *Journal of Econometrics* **53**, pp. 165–188.
29. Tschernig, R., Weber, E., and Weigand, R. (2013). Long-run identification in a fractionally integrated system. *Journal of Business and Economic Statistics* **31**, pp. 438-450.