

Discussion Papers
Department of Economics
University of Copenhagen

No. 12-18

The role of initial values in nonstationary fractional time series models

Søren Johansen and Morten Ørregaard Nielsen

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark

Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00

<http://www.econ.ku.dk>

ISSN: 1601-2461 (E)

The role of initial values in nonstationary fractional time series models*

Søren Johansen[†]
University of Copenhagen
and CREATES

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

November 8, 2012

Abstract

We consider the nonstationary fractional model $\Delta^d X_t = \varepsilon_t$ with ε_t i.i.d. $(0, \sigma^2)$ and $d > 1/2$. We derive an analytical expression for the main term of the asymptotic bias of the maximum likelihood estimator of d conditional on initial values, and we discuss the role of the initial values for the bias. The results are partially extended to other fractional models, and three different applications of the theoretical results are given.

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

JEL Classification: C22.

*We are grateful to seminar participants at various universities and conferences for comments, and to 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, funded by the Danish National Research Foundation) for financial support.

[†]Email: Soren.Johansen@econ.ku.dk

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

1 Introduction

Traditionally, inference in nonstationary autoregressive models is conditional on initial values, for example in the AR(k) model conditioning on k initial values implies that maximum likelihood estimation is equivalent to ordinary least squares. This 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, and in both cases assuming that the initial values are all zero. The conditional maximum likelihood estimator has been very widely applied in the literature, also for fractional time series models, where the initial values have typically been assumed to be zero.

Recently, inference conditional on (non-zero) initial values has been advocated for nonstationary fractional time series models by Johansen and Nielsen (2010, 2012)—henceforth JN (2010, 2012)—and Tschernig, Weber, and Weigand (2010) in theoretical work. In empirical work conditional inference has been applied by, for example, Carlini, Manzonei, and Mosconi (2010) and Bollerslev, Osterreider, Sizova, and Tauchen (2012) 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.

The purpose of this paper is to investigate the magnitude of the influence of initial values on the bias of the Gaussian (quasi-)maximum likelihood estimator of the fractional parameter, d , conditional on initial values. For analytic tractability we consider the simplest model for fractional processes, $\Delta^d X_t = \varepsilon_t$ with ε_t i.i.d.($0, \sigma^2$). In practice we have to decide how to split a given sample into initial values and observations. In order to discuss this we derive an analytical expression for the asymptotic second-order bias term via a higher-order stochastic expansion of the estimator.

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, \dots, X_T) . In particular, Lieberman and Phillips (2004) find simple 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 model and our main results. In Section 3 we give three applications of the theoretical results to (i) illustrate the bias numerically, (ii) discuss (non-)invariance of different fractional models to location and scale, (iii) an empirical data set. Section 4 concludes. Proofs of our main results and some mathematical details are given in the appendices.

2 Model and main results

We consider the model

$$\Delta^d X_t = \varepsilon_t, \quad \varepsilon_t \sim i.i.d.(0, \sigma^2), \quad t = 1, \dots, T, \quad (1)$$

where $d \geq 0$ and $\sigma^2 > 0$. To focus on estimation of d we consider σ^2 fixed at the true value $\sigma_0^2 > 0$. We denote the true value of d by d_0 .

The fractional coefficients $\pi_j(u)$ are defined as the coefficients in an expansion of $(1-z)^{-u}$, which are

$$\pi_j(u) = \frac{\Gamma(u+j)}{\Gamma(u)\Gamma(j+1)} = \frac{u(u+1)\dots(u+j-1)}{j!}, \quad (2)$$

where $\Gamma(d)$ denotes the Gamma function. The difference operators $\Delta^d X_t$, $\Delta_+^d X_t$, and $\Delta_-^d X_t$ are defined as

$$\Delta^d X_t = \sum_{n=0}^{\infty} \pi_n(-d) X_{t-n} = \sum_{n=0}^{t-1} \pi_n(-d) X_{t-n} + \sum_{n=0}^{\infty} \pi_{n+t}(-d) X_{-n} = \Delta_+^d X_t + \Delta_-^d X_t. \quad (3)$$

Thus infinitely many past values are needed to calculate the fractional differences. Several useful results for the fractional coefficients and their derivatives are collected in Appendix A.

The model (1) is a special case of several more general models. The univariate fractional autoregressive model of JN (2010) is

$$\Delta^d X_t = \pi \Delta^{d-b} L_b X_t + \sum_{j=1}^k \Gamma_j L_b^j \Delta^d X_t + \varepsilon_t,$$

where $L_b = 1 - \Delta^b$ denotes the fractional lag operator. For this model, the conditional likelihood depends on the residuals, see JN (2010, p. 52),

$$\begin{aligned} \varepsilon_t(d, \phi) &= \Delta^d X_t - \pi L_b \Delta^{d-b} X_t - \sum_{j=1}^k L_b^j \Delta^d X_t \\ &= \Delta^{-b} \Psi(L_b) \Delta^d X_t = a(\phi, L) \Delta^d X_t \end{aligned}$$

with $\phi = (b, \pi, \Gamma_1, \dots, \Gamma_k)$. Another well-known alternative model is the ARFIMA model,

$$A(L) \Delta^d X_t = B(L) \varepsilon_t,$$

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

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

For both the fractional autoregressive model and the ARFIMA model the analysis would depend on the derivatives of the conditional likelihood function, which would in turn be simple functions of the derivatives of the residuals. Again, to focus on estimation of d we consider the remaining parameters, ϕ and ψ , respectively, fixed at their true values. For a function $f(d)$ we denote the derivative of f with respect to d as $Df(d) = \frac{\partial}{\partial d} f(d)$ (Euler's notation), and the relevant derivatives are

$$\begin{aligned} D^m \varepsilon_t(d, \psi)|_{d_0, \psi_0} &= b(\psi_0, L) D^m \Delta^d X_t|_{d_0} \\ &= (\log \Delta)^m b(\psi_0, L) \Delta^{d_0} X_t = (\log \Delta)^m \varepsilon_t \end{aligned}$$

for the ARFIMA model, and the same argument applies for the fractional autoregressive model. Thus, for both these more general models, 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 (or ϕ_0). If ψ (or ϕ) has to be estimated, the analysis becomes much more complicated. We therefore focus our analysis on the simple model.

We consider maximum likelihood estimation of d_0 based on observations X_1, \dots, X_T generated by (1) for fixed bounded initial values, that is, conditional on $X_{-n}, n \geq 0$, as developed in JN (2010, 2012). For the asymptotic analysis we make the following assumptions.

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

Assumption 2 *The initial values $X_{-n}, n = 0, 1, \dots$, are bounded, i.e. $\sup_{n \geq 0} |X_{-n}| \leq c < \infty$.*

As remarked earlier, conditional maximum likelihood estimation has been very widely applied in the literature for fractional time series models, especially in the nonstationary case. However, to be able to calculate the fractional differences, the previous literature has typically assumed that the initial values are all zero, i.e. that $X_{-n} = 0, n \geq 0$.

For a general set of initial values the solution of model (1) is given in the following lemma.

Lemma 1 *Under Assumption 2, the solution of model (1) is*

$$X_t = \Delta_+^{-d_0} \varepsilon_t - \Delta_+^{-d_0} \Delta_-^{d_0} X_t, \quad t = 1, 2, \dots, T. \quad (4)$$

Proof. From (3) and Lemma A.2 we see that $\Delta_-^{d_0} X_t = \sum_{n=0}^{\infty} \pi_{n+t}(-d_0) X_{-n}$ is bounded in absolute value by $c \sum_{n=0}^{\infty} (n+t)^{-d_0-1} < \infty$ if $d_0 > 0$, so that $\Delta_-^{d_0} X_t$ and $\Delta_+^{-d_0} X_t$ are well defined. From (1) we then find $\varepsilon_t = \Delta_-^{d_0} X_t = \Delta_+^{d_0} X_t + \Delta_-^{d_0} X_t, t = 1, \dots, T$. The operator $\Delta_+^{d_0}$ only depends on X_t for $t \geq 1$, and is invertible on the sequences which are zero for $t \leq 0$. The inverse $\Delta_+^{-d_0}$ is given by $\Delta_+^{-d_0} \varepsilon_t = \sum_{n=0}^{t-1} \pi_n(d_0) \varepsilon_{t-n}$, which yields the solution for $t = 1, \dots, T$, when applied to $\varepsilon_t - \Delta_-^{d_0} X_t$. See also Johansen (2008, Corollary 6 and Theorem 8) and JN (2010, Lemma 1). ■

The Gaussian (quasi-)log-likelihood, conditional on initial values, is

$$L(d) = -\frac{T}{2} \log \sigma_0^2 - \frac{1}{2\sigma_0^2} \sum_{t=1}^T (\Delta^d X_t)^2 \quad (5)$$

apart from a constant, and is a function of fractional differences of X_t . Of course the likelihood function (5) depends on initial values through $\Delta^d X_t = \Delta_+^d X_t + \Delta_-^d X_t$. The first term is a function of the observations X_1, \dots, X_T , but the second is a function of infinitely many initial values which are not all observed. Thus, in order to calculate (an approximation to) the likelihood function we have to choose some initial values, say \tilde{X}_{-n} , and calculate $\Delta_-^d \tilde{X}_t$. A simple and commonly applied choice is $\tilde{X}_{-n} = 0, n \geq 0$, e.g. Hualde and Robinson (2011).

Another possibility is to set aside the first N observations as initial values as is usually done in the analysis of an AR(k) model, in which case $\tilde{X}_{-n} = X_{-n}$ for $0 \leq n \leq N-1$, and we analyze the effect of doing so on the bias of the estimator for d under different choices of the remaining $\tilde{X}_{-n}, n \geq N$. A simple choice is to set $\tilde{X}_{-n} = 0, n \geq N$, which corresponds to setting $\tilde{X}_{-n} = X_{-n}1_{\{n < N\}}$, where $1_{\{A\}}$ denotes the indicator function for the event A . A different choice is to use $\tilde{X}_{-n} = X_{-n}, n < N$, and $\tilde{X}_{-n} = X_{-N+1}, n \geq N$, i.e., setting the chosen initial values corresponding to unobserved initial values equal to the earliest observed initial value. This corresponds to setting $\tilde{X}_0 = X_0$ and $\Delta\tilde{X}_{-n} = \Delta X_{-n}1_{\{n < N-1\}}$. We summarize these assumptions below and apply them in our main results.

Assumption 3 *We set aside the N values $X_{-n}, n = 0, \dots, N-1$, as initial values, and choose the initial values for the calculation of the fractional differences according to one of the following possibilities:*

0. $\tilde{X}_{-n} = X_{-n}1_{\{n < N\}}, N \geq 0$,
1. $\tilde{X}_0 = X_0, \Delta\tilde{X}_{-n} = \Delta X_{-n}1_{\{n < N-1\}}, N \geq 1$.

With the chosen initial values we define $\tilde{\Delta}^d X_t = \Delta_+^d X_t + \Delta_-^d \tilde{X}_t$ and obtain the following approximation to the log-likelihood (5):

$$\tilde{L}(d) = -\frac{T}{2} \log \sigma_0^2 - \frac{1}{2\sigma_0^2} \sum_{t=1}^T (\tilde{\Delta}^d X_t)^2. \quad (6)$$

Thus, $\tilde{L}(d)$ can be considered a type of quasi-likelihood with respect to both the initial values and the distributional assumption. We also define the associated conditional (quasi-)maximum likelihood estimator,

$$\hat{d} = \arg \max_{d \geq 0} \tilde{L}(d). \quad (7)$$

Because maximizing $L(d)$ or $\tilde{L}(d)$ is the same as minimizing a sum of squared residuals, the estimator \hat{d} is sometimes referred to as the conditional sum-of-squares estimator.

The first-order asymptotic properties of \hat{d} under Assumptions 1 and 2 (but not necessarily Assumption 3) are given in the following lemma, based on results of JN (2012) and Nielsen (2012).

Lemma 2 *Let the process $X_t, t = 1, \dots, T$, be generated by model (1) and suppose Assumptions 1 and 2 are satisfied. Then the estimator \hat{d} in (7) exists and is consistent on a compact subset of $\mathbb{R}_+ = \{x \in \mathbb{R} : x > 0\}$, and furthermore $T^{1/2}(\hat{d} - d_0) \xrightarrow{D} N(0, (\pi^2/6)^{-1})$.*

Proof. From Lemma 1 we have that $\tilde{\Delta}^d X_t = \Delta_+^{d-d_0} \varepsilon_t - \Delta_+^{d-d_0} \Delta_-^{d_0} X_t + \Delta_-^d \tilde{X}_t$, so that we need to analyze product moments of the terms on the right-hand side, appropriately normalized, when d belongs to a compact subset of \mathbb{R}_+ . However, the deterministic terms $\Delta_+^{d-d_0} \Delta_-^{d_0} X_t$ and $\Delta_-^d \tilde{X}_t$ are shown to be asymptotically uniformly negligible under Assumption 2 in JN (2012, Lemma A.8(i)). This leaves the product moment $\sum_{t=1}^T (\Delta_+^{d-d_0} \varepsilon_t)^2$, which is analyzed in Nielsen (2012) under Assumption 1. Existence and consistency of \hat{d} on a compact subset of \mathbb{R}_+ follows.

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

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

where d^* is an intermediate value satisfying $|d^* - d_0| \leq |\hat{d} - d_0| \xrightarrow{P} 0$. The product moments in $\mathbf{D}^2\tilde{L}(d)$ are shown in JN (2010, Lemma C.4) and JN (2012, 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 $\mathbf{D}^2\tilde{L}(d^*) = \mathbf{D}^2\tilde{L}(d_0) + o_P(1)$, and we therefore analyze $\mathbf{D}\tilde{L}(d_0)$ and $\mathbf{D}^2\tilde{L}(d_0)$. From Lemmas B.1 and B.4 we find that $T^{-1/2}\mathbf{D}\tilde{L}(d_0) = -\sum_{j=1}^{t-1} j^{-1}\varepsilon_{t-j} + O_P(T^{-1/2})$ and $T^{-1}\mathbf{D}^2\tilde{L}(d_0) = -\pi^2/6 + O_P(T^{-1/2})$, and the result follows from Lemmas B.2 and B.3. ■

2.1 Asymptotic bias

Our main result holds only for $d_0 > 1/2$, that is, for nonstationary processes, which is therefore assumed in the remainder of the paper.

To analyze the asymptotic bias of the estimator for d , and in particular how initial values influence the bias, we need to examine higher-order terms in a stochastic expansion, see Lawley (1956), of \hat{d} . The conditional likelihood satisfies that $\mathbf{D}\tilde{L}(d_0) = O_P(T^{1/2})$, $\mathbf{D}^2\tilde{L}(d_0) = O_P(T)$, and $\mathbf{D}^3\tilde{L}(d) = O_P(T)$ uniformly in a neighborhood of d_0 and a Taylor series expansion of $\mathbf{D}\tilde{L}(\hat{d}) = 0$ around d_0 gives

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

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

$$T^{1/2}(\hat{d} - d_0) = -\frac{T^{-1/2}\mathbf{D}\tilde{L}(d_0)}{T^{-1}\mathbf{D}^2\tilde{L}(d_0)} - \frac{1}{2}T^{-1/2}\left(\frac{T^{-1/2}\mathbf{D}\tilde{L}(d_0)}{T^{-1}\mathbf{D}^2\tilde{L}(d_0)}\right)^2\frac{T^{-1}\mathbf{D}^3\tilde{L}(d^*)}{T^{-1}\mathbf{D}^2\tilde{L}(d_0)} + O_P(T^{-1}). \quad (8)$$

Based on this expansion we find another expansion $\hat{d} - d_0 = T^{-1/2}\tilde{A}_T + T^{-1}\tilde{B}_T + o_P(T^{-1})$ with the property that $(\tilde{A}_T, \tilde{B}_T) \xrightarrow{D} (A, B)$, where $E(\tilde{A}_T) = E(A) = 0$. Then the zero- and first-order terms of the bias are zero, and the second-order asymptotic bias term is $T^{-1}E(B)$.

We now state the main result on the asymptotic bias of \hat{d} , the proof of which is given in Appendix C. To describe the results we use Riemann's zeta function, $\zeta_s = \sum_{j=1}^{\infty} j^{-s}$, $s > 1$, and specifically

$$\zeta_2 = \sum_{j=1}^{\infty} j^{-2} = \frac{\pi^2}{6} \simeq 1.6449 \text{ and } \zeta_3 = \sum_{j=1}^{\infty} j^{-3} \simeq 1.2021. \quad (9)$$

Theorem 1 *Let the process $X_t, t = 1, \dots, T$, be generated by model (1) with $d_0 > 1/2$ and suppose Assumptions 1 and 2 are satisfied. Then the asymptotic bias of \hat{d} is*

$$-T^{-1}[3\zeta_3\zeta_2^{-2} + \xi(d_0)\zeta_2^{-1}] + o(T^{-1}), \quad (10)$$

where

$$\xi(d) = \sigma_0^{-2} \sum_{t=1}^{\infty} \left[\sum_{n=0}^{\infty} \pi_{t+n}(-d)(X_{-n} - \tilde{X}_{-n}) \right] \left[\sum_{n=0}^{\infty} \left(\sum_{k=1}^{t-1} k^{-1} \pi_{t-k+n}(-d)X_{-n} - \mathbf{D}\pi_{t+n}(-d)\tilde{X}_{-n} \right) \right]. \quad (11)$$

The main bias terms in (10) are of the same order of magnitude in T , namely $O(T^{-1})$. The first term, $3\zeta_3\zeta_2^{-2}$, is fixed and comes from the higher-order (second and third) derivatives of the likelihood and does not depend on initial values or on d_0 . The second term is a function of initial values and d_0 and can be made smaller by including more initial values. The bias due to initial values is quadratic in X_{-n}/σ_0 and \tilde{X}_{-n}/σ_0 .

If all initial values are observed, so that $\tilde{X}_{-n} = X_{-n}$ for all $n \geq 0$, then the second-order bias is $-T^{-1}3\zeta_3\zeta_2^{-2} \simeq -1.3328T^{-1}$, which does not depend on initial values or on d_0 . Thus, the estimator is asymptotically second-order pivotal in that case, suggesting that higher-order asymptotic refinements may be possible via the bootstrap also in our nonstationary setting, see Andrews, Lieberman, and Marmer (2006) for the stationary case. Furthermore, the fixed bias term could be used for a simple bias correction by considering the estimator $\tilde{d} = \hat{d} + T^{-1}3\zeta_3\zeta_2^{-2}$.

The fixed bias term, $3\zeta_3\zeta_2^{-2}$, is the same as the bias term derived by Lieberman and Phillips (2004) for the estimator, \hat{d}_{stat} , based on the joint (unconditional) likelihood of (X_1, \dots, X_T) in the stationary case, $0 < d < 1/2$. They show that the distribution function of $\zeta_2^{1/2}T^{1/2}(\hat{d}_{\text{stat}} - d_0)$ is approximated by

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

where $\Phi(x)$ and $\phi(x)$ denote the standard normal distribution and density functions, respectively. One can derive an approximation for the bias of \hat{d}_{stat} to be

$$E(\zeta_2^{1/2}T^{1/2}(\hat{d}_{\text{stat}} - d_0)) = \int_0^\infty (1 - F_T(x))dx - \int_{-\infty}^0 F_T(x)dx = -T^{-1/2}3\zeta_3\zeta_2^{-3/2} + O_P(T^{-1}),$$

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

Although the asymptotic bias of \hat{d} is of order $O(T^{-1})$ we note that the asymptotic standard deviation of \hat{d} is of order $O(T^{-1/2})$, see Lemma 2. That is, for testing purposes or for calculating confidence sets for d_0 the relevant quantity is in fact the bias relative to the asymptotic standard deviation, which is given by

$$-T^{-1}[3\zeta_3\zeta_2^{-2} + \xi(d_0)\zeta_2^{-1}]/(T\zeta_2)^{-1/2} = -T^{-1/2}[3\zeta_3\zeta_2^{-3/2} + \xi(d_0)\zeta_2^{-1/2}] \quad (12)$$

and is of order $O(T^{-1/2})$.

If we further use Assumption 3 we get the following expressions for $\xi(d)$, see Appendix D for the proof.

Theorem 2 *Let Assumptions 2 and 3. $\ell = 0, 1$, be satisfied and let $\xi_N(d)$ denote $\xi(d)$ in (11) for a fixed N chosen in Assumption 3. For $d > 1/2 + \ell$ it holds that*

$$\begin{aligned} \xi_N(d) &= \sigma_0^{-2} \sum_{t=1}^{\infty} \left[\sum_{n=N}^{\infty} \pi_{t+n}(-d + \ell) \Delta^\ell X_{-n} \right] \\ &\quad \times \left[\sum_{n=0}^{\infty} \sum_{k=1}^{t-1} k^{-1} \pi_{t-k+n}(-d + \ell) \Delta^\ell X_{-n} - \sum_{n=0}^{N-1} \mathbf{D} \pi_{t+n}(-d + \ell) \Delta^\ell X_{-n} \right], \end{aligned} \quad (13)$$

which is bounded by

$$|\xi_N(d)| \leq c(1+N)^{-\delta} \quad (14)$$

for any $\delta < \min(d - \ell, 2(d - \ell) - 1)$.

The formula (13) for $\ell = 0$ comes from inserting the choice $\tilde{X}_{-n} = X_{-n}1_{(n < N)}$ in the expression for $\xi(d)$, see (11). For d large it may be not be natural to choose the value zero for $\tilde{X}_{-n}, n \geq N$, but rather choose the first observed initial value, i.e. $\tilde{X}_{-n} = X_{-N+1}, n \geq N$, as for $\ell = 1$. This corresponds to setting $\Delta\tilde{X}_{-n} = 0, n \geq N - 1$, and therefore an expression for $\xi_N(d)$ is given involving ΔX_{-n} , see (13). Note, however, that the fractional coefficients are cumulated and $-d$ is replaced by $-d + 1$, so they decrease much slower and we only get the evaluation (14) if in fact $d > 1.5$.

We next discuss the bias terms $\xi(d)$ and $\xi_N(d)$ in more detail under additional assumptions.

2.2 Further results for special cases

The expressions for $\xi(d)$ and $\xi_N(d)$ in (11) and (13), respectively, show that both functions depend on d and on all initial values. In order to get an impression for what this dependence is, we derive simple expressions for $\xi(d)$ and $\xi_N(d)$ in special cases or under simplifying assumptions about the initial values.

First, when d is an integer, we find simple results for $\xi(d)$ and $\xi_N(d)$, and hence the asymptotic bias, as follows.

Theorem 3 *For $\xi(d)$ given in (11) it holds that:*

(i) *For $d = 1$ we have*

$$\xi(1) = \sigma_0^{-2}(X_0 - \tilde{X}_0) \sum_{n=0}^{\infty} \mathbf{D}\pi_{1+n}(-1)\tilde{X}_{-n} \quad (15)$$

such that $\xi(1) = 0$ if either (i) $\tilde{X}_{-n} = 0$ for all $n \geq 0$ or if (ii) $\tilde{X}_0 = X_0$. In those cases the asymptotic relative bias of \hat{d} is given by

$$-T^{-1/2}3\zeta_3\zeta_2^{-3/2} + O(T^{-1}) \simeq -1.7094T^{-1/2} + O(T^{-1}).$$

(ii) *Under Assumption 3.0 and $N \geq d = k$ for any integer $k \geq 1$ we have that $\xi_N(d) = 0$.*

Proof. From (43) we find

$$\begin{aligned} \sigma_0^2\xi(d) &= \sum_{t=1}^{\infty} \left[\sum_{n=0}^{\infty} \pi_{t+n}(-d)(X_{-n} - \tilde{X}_{-n}) \right] \left[\sum_{n=0}^{\infty} \sum_{k=1}^{t-1} k^{-1} \pi_{t-k+n}(-d)X_{-n} - \sum_{n=0}^{\infty} \mathbf{D}\pi_{t+n}(-d)\tilde{X}_{-n} \right] \\ &= \sum_{t=1}^{\infty} \eta_{0t}(d) [\eta_{1t}^{(1)}(d) + \eta_{1t}^{(2)}(d)], \end{aligned}$$

where

$$\eta_{1t}^{(1)}(d) = \sum_{n=0}^{\infty} \sum_{k=1}^{t-1} k^{-1} \pi_{t-k+n}(-d)X_{-n} = \sum_{k=1}^{t-1} (t-k)^{-1} \sum_{n=0}^{\infty} \pi_{k+n}(-d)X_{-n}$$

is non-zero only if $t \geq 2$.

Proof of (i): For $d = 1$, where $\pi_t(-1) = 0$ for $t \geq 2$, we only get a contribution to $\eta_{0t}(d)$ for $t + n = 1$. This shows that we must have $t = 1, n = 0$, and $\eta_{01}(d) = -(X_0 - \tilde{X}_0)$. Moreover, $\eta_{11}^{(1)}(d) = 0$ and $\eta_{11}^{(2)}(d) = -\sum_{n=0}^{\infty} \mathbf{D}\pi_{1+n}(-1)\tilde{X}_{-n}$.

Proof of (ii): When $d = k$ is a positive integer $\pi_t(-d) = 0$ for $t \geq k$ so that $\eta_{0t}(d) = 0$ under Assumption 3 with $N \geq k$. \blacksquare

It follows from Theorem 3 that for $d_0 = 1$ we need one initial value and for $d_0 = 2$ we need two initial values, etc. Alternatively, for $d_0 = 1$, one can, in fact, simply set $\tilde{X}_{-n} = 0$ for all $n \geq 0$, which gives no contribution from initial values to the second-order asymptotic bias. Since the bias term is continuous in d_0 , the same is true for a (small) neighborhood of $d_0 = 1$.

We next assume that the initial values are constant, and derive expressions for the initial values bias term $\xi_N(d)$ given by (13) in Theorem 1. Here, $\Psi(d) = \mathbf{D} \log \Gamma(d)$ denotes the Digamma function.

Theorem 4 *If $X_{-n} = C$ and $\tilde{X}_{-n} = 1_{\{n < N\}}C$ for $n \geq 0$, i.e. under Assumption 3.0, then $\xi_N(d)$ given in (13) is*

$$\xi_N(d) = \frac{C^2}{2\sigma_0^2} \mathbf{D} \sum_{n=N}^{\infty} \binom{d-1}{n}^2 = \xi_0(d) - \frac{C^2}{2\sigma_0^2} \mathbf{D} \sum_{n=1}^{N-1} \binom{d-1}{n}^2 \quad (16)$$

where

$$\xi_0(d) = \frac{C^2}{\sigma_0^2} \frac{\Gamma(2d-1)}{\Gamma(d)^2} (\Psi(2d-1) - \Psi(d)). \quad (17)$$

Proof. *Proof of (16):* The expression for $\xi_N(d)$ in (13) is found from (27),

$$\xi_N(d) = H_{01T}(d) = \sigma_0^{-2} \sum_{t=1}^{\infty} \eta_{0t}(d) \eta_{1t}(d),$$

where $\eta_{mt}(d)$ is given in (24). We therefore evaluate the deterministic term in $\tilde{\Delta}^d X_t$ and its derivatives at $d = d_0$. For $t \geq 1$ we find from (4) in Lemma 1 that the deterministic term of $\tilde{\Delta}^d X_t$ has the expression

$$-\Delta_+^{d-d_0} \Delta_-^{d_0} X_t + \Delta_-^d \tilde{X}_t = -\sum_{k=0}^{t-1} \pi_k(-d+d_0) \sum_{n=0}^{\infty} \pi_{t+n-k}(-d_0) X_{-n} + \sum_{n=0}^{\infty} \pi_{t+n}(-d) \tilde{X}_{-n}.$$

If $X_{-n} = C$ and $\tilde{X}_{-n} = 1_{\{n < N\}}C$ for $n \geq 0$, this is C times

$$\begin{aligned} & -\sum_{k=0}^{t-1} \pi_k(-d+d_0) \sum_{n=t-k}^{\infty} \pi_n(-d_0) + \sum_{n=t}^{t+N-1} \pi_n(-d) \\ & = \sum_{k=0}^{t-1} \pi_k(-d+d_0) \pi_{t-k-1}(-d_0+1) + \pi_{t+N-1}(-d+1) - \pi_{t-1}(-d+1) = \pi_{t+N-1}(-d+1) \end{aligned}$$

by application of Lemma A.4(d)-(f). Therefore it holds that $\eta_{mt}(d) = C(-1)^m \mathbf{D}^m \pi_{t+N-1}(-d+1)$, such that

$$\begin{aligned} \xi_N(d) &= -\frac{C^2}{\sigma_0^2} \sum_{t=1}^{\infty} \pi_{t+N-1}(-d+1) \mathbf{D} \pi_{t+N-1}(-d+1) \\ &= \frac{C^2}{2\sigma_0^2} \mathbf{D} \sum_{t=1}^{\infty} \pi_{t+N-1}(-d+1)^2 = \frac{C^2}{2\sigma_0^2} \mathbf{D} \sum_{n=N}^{\infty} \binom{d-1}{n}^2. \end{aligned}$$

Proof of (17): We find from Lemma A.5 that for $N = 0$,

$$\mathbf{D} \sum_{n=0}^{\infty} \binom{d-1}{n}^2 = 2 \frac{\Gamma(2d-1)}{\Gamma(d)^2} (\Psi(2d-1) - \Psi(d)).$$

A consequence of (16) in Theorem 4 is that for constant initial values we can find $\xi_N(d)$ simply by subtracting a finite summation from $\xi_0(d)$, which is given explicitly in (17). ■

3 Applications

3.1 Numerical illustration

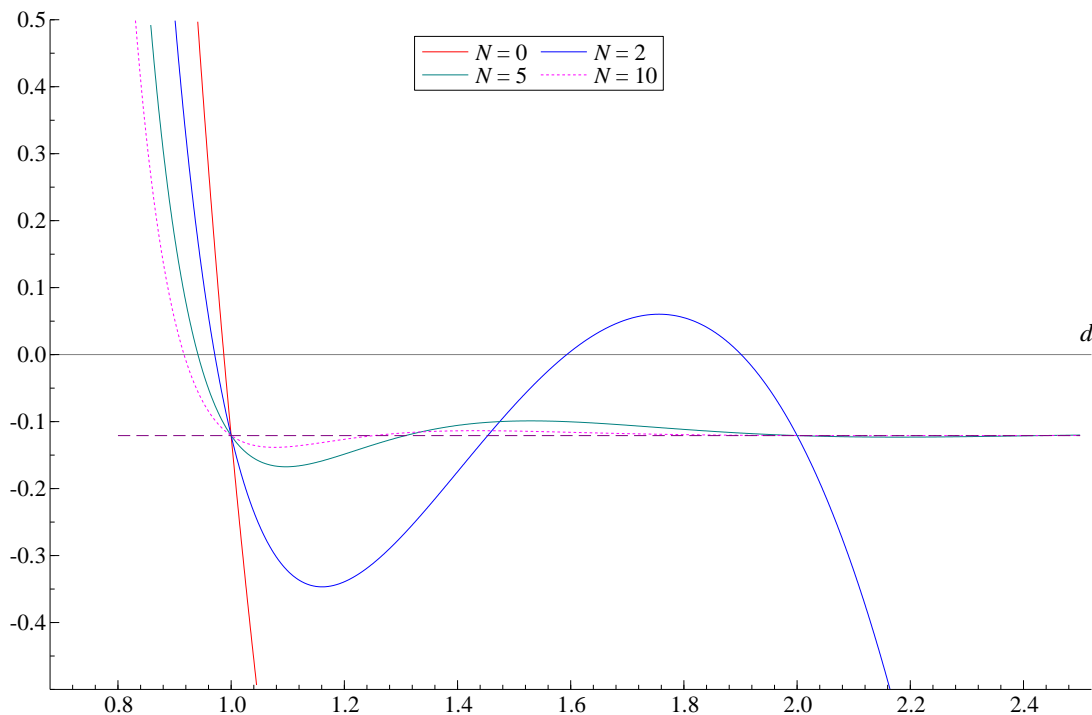
We illustrate the formulas above (under Assumption 3.0) by some numerical calculations in order to quantify the magnitude of the relative bias, and therefore the distortion of the quantiles (critical values). The relative bias is $b_T(d) = -T^{-1/2} \zeta_2^{-1/2} (3\zeta_3 \zeta_2^{-1} + \xi_N(d))$, where $\xi_N(d)$ is given in (13) and $\zeta_2 \simeq 1.6449$, $\zeta_3 \simeq 1.2021$ are given in (9). It follows from Theorem 3 that if we choose $\tilde{X}_{-n} = X_{-n} 1_{\{0 \leq n < N\}}$, then $\xi_N(d) = 0$ for $d = 1, \dots, N$. In these points the relative bias is $b_T(d) = -T^{-1/2} 3\zeta_3 \zeta_2^{-3/2} = -1.71T^{-1/2}$, which can only be made smaller by increasing T . For example, for $T = 200$ this value is $b_{200}(d) = -0.12$. Thus, when testing hypotheses in these cases the relative bias distorts the relevant quantile by -0.12 .

To analyze what happens in between these values, we consider the situation that all initial values are constant, $X_{-n} = C$, and take $\tilde{X}_{-n} = C 1_{\{0 \leq n < N\}}$, where we can calculate the bias term explicitly for all d by using the expression, see (16),

$$\xi_N(d) = \frac{C^2}{\sigma_0^2} \frac{\Gamma(2d-1)}{\Gamma(d)^2} (\Psi(2d-1) - \Psi(d)) - \frac{C^2}{\sigma_0^2} \sum_{n=1}^{N-1} \pi_n(-d+1) \mathbf{D} \pi_n(-d+1).$$

The calculation with this value of $\xi_N(d)$ is illustrated in Figure 1, which depicts the relative bias as a function of d for $T = 200$, $C\sigma_0^{-1} = 10$, and $N = 0, 2, 5, 10$. It is seen that for $1 \leq d \leq 2$, the values $N = 0$ and $N = 2$ give a rather large relative bias, even though it is equal to the fixed value $-T^{-1/2} 3\zeta_3 \zeta_2^{-3/2} = -0.12$ (the dashed line) for $d = 1$ ($N \geq 0$) and $d = 2$ ($N \geq 2$). Note also that the absolute value of the relative bias for $N = 2$ increases rapidly for $d > 2$. For $N \geq 5$ and $1 \leq d \leq 2$, however, we get a relative bias close to the fixed value -0.12 .

Thus we see the effect, in this stylized example, of increasing the number of values set aside for initial values, N . The absolute relative bias is decreased by increasing N thereby forcing the bias to be -0.12 on the integer values $d = 1, 2, \dots, N$. A consequence of this

Figure 1: Relative bias as a function of d with constant initial values

Note: The relative bias $b_T(d) = -T^{-1/2}\zeta_2^{-1/2}(3\zeta_3\zeta_2^{-1} + \xi_N(d))$ is displayed as a function of d when the chosen initial values are $\tilde{X}_{-n} = X_{-n}1_{(n < N)}$ and the initial values of the process are constant $X_{-n} = C, n \geq 0$. We choose $T = 200$ and $C\sigma_0^{-1} = 10$. The dotted line indicates the fixed bias $-T^{-1/2}3\zeta_3\zeta_2^{-3/2} = -0.12$.

is that if we find, for a given value of $C\sigma_0^{-1}$, a value of N such that at (approximately) $d = 1.1$ we get a small absolute relative bias, then the relative bias remains small for all values $1 \leq d \leq N$. We note that this minimax procedure is conservative in the sense that for $d \neq 1.1$ we could do with a smaller value of N . The few extra initial values seem a small price to pay for the uncertainty in d . However, this calculation is only valid for the stylized illustrative example with constant initial values. In general, the bias term $\xi_N(d)$ depends on infinitely many initial values and it is difficult to quantify the influence of the initial values on the bias in the more general case.

3.2 Discussion of location and scale (non-)invariance of fractional models

Another application of our results is to the following situation. Consider the two models

$$\Delta^d X_t = \varepsilon_t, \quad t = 1, 2, \dots, \quad \text{with initial values } X_{-n}, n \geq 0, \quad (18)$$

$$\Delta^d X_t = \varepsilon_t, \quad t = 1, 2, \dots, \quad \text{with initial values } X_{-n} = 0, n \geq 0. \quad (19)$$

Model (18) is the model analyzed in this paper. Model (19) is the most commonly applied model in the literature for nonstationary fractional models and has been analyzed by, e.g., Hualde and Robinson (2011) among many others.

Both models (18) and (19) are clearly scale invariant. That is, the models are invariant under multiplication by a constant in the sense that changing units of, say, a price variable, X_t , does not change inference on d .

Both models are not, however, invariant to changes in location. Suppose, for example, that X_t is the log of a price variable. Then changing the unit from dollars to cents gives an additive constant of $\log 100$, i.e. $Y_t = X_t + \log 100$. This does not change the analysis of model (18) since $\Delta^d Y_t = \Delta^d (X_t + \log 100) = \Delta^d X_t$ because $\Delta^d 1 = 0$ for $d > 0$, so that Y_t also satisfies the equation $\Delta^d Y_t = \varepsilon_t$. The only difference is that now the initial values are $Y_{-n} = X_{-n} + \log 100$ for $n \geq 0$. Model (19), however, assumes initial values are zero, but if $X_{-n} = 0$ for $n \geq 0$ then for Y_t the initial values are $Y_{-n} = \log 100, n \geq 0$, and choosing $\tilde{Y}_{-n} = 0, n \geq 0$, we get a relative bias of

$$-T^{-1/2} \zeta_2^{-1/2} \left[\frac{3\zeta_3}{\zeta_2} + (\log 100)^2 \frac{\Gamma(2d_0 - 1)(\Psi(2d_0 - 1) - \Psi(d_0))}{\sigma_0^2 \Gamma(d_0)^2} \right],$$

see Theorem 4.

3.3 Data example

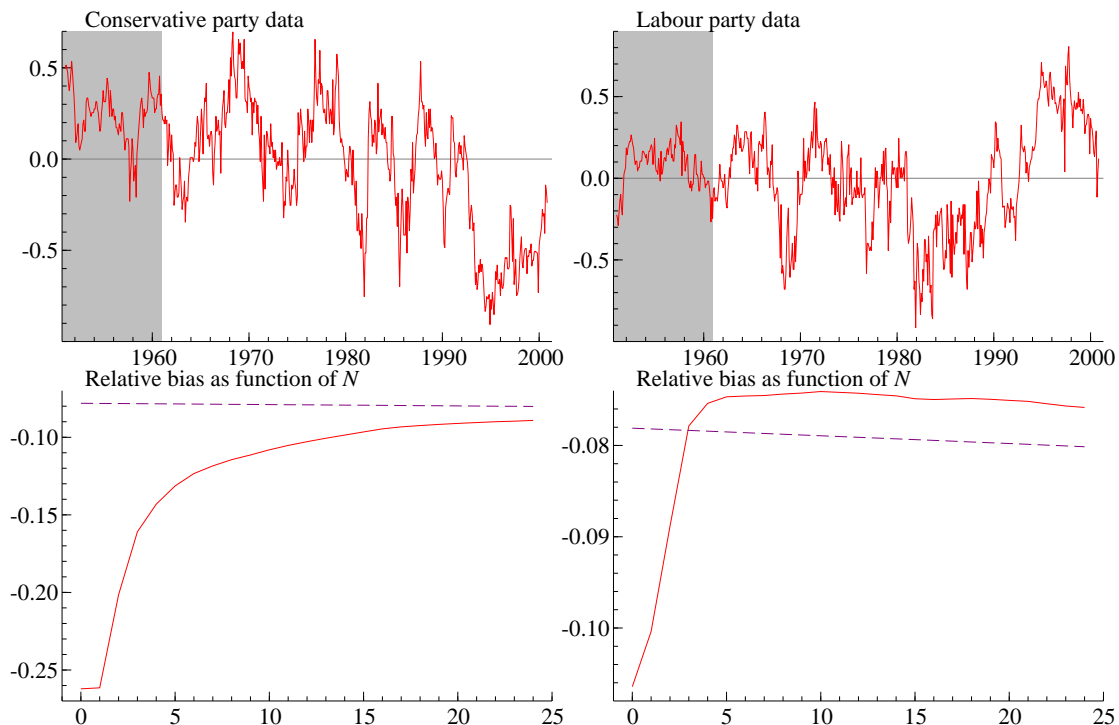
As the final application, we consider a specific data example. The data are monthly Gallup opinion poll data on support for the Conservative and Labour parties in the United Kingdom. They cover the period from January 1951 to November 2000, for a total of 599 months. The two series have been logistically transformed and centered, so that, if Y_t denotes an observation on the original series, it is mapped into $\log(Y_t/(100 - Y_t))$. A shorter version of this dataset was analyzed in Byers, Davidson, and Peel (1997) and Dolado, Gonzalo, and Mayoral (2002). In light of the discussion in Section 3.2, we consider the series centered by their sample averages as in Byers *et al.* (1997) and Dolado *et al.* (2002), but also the uncentered series.

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 integration methods. The basic findings of Byers *et al.* (1997) and Dolado *et al.* (2002) are 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.

Suppose, for illustration, that X_t (denoting either the centered or uncentered series) is in fact zero prior to January 1951, and that the econometrician only observes data starting in January 1961. That is, January 1951 through December 1960 are unobserved initial values. Following Assumption 3.0, the econometrician will then have to split the given sample of 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 relative bias of ignoring initial values, i.e. of setting $N = 0$, and (ii) how sensitive is the relative bias to the choice of N for this particular data set.

To answer these questions we apply Theorem 1, and in particular (10) and (13). We note that $\xi_N(d)$ in (13) depends only on the unobserved initial values, which in this example are the observations from January 1951 to December 1960. To apply Theorem 1 we need values of d_0 and σ_0^2 . For both the centered and uncentered series we use $d_0 = 0.76$ and $\sigma_0 = 0.12$ for the Conservative party series and $d_0 = 0.69$ and $\sigma_0 = 0.13$ for the Labour party series. These values were obtained as the estimators based on the data from January 1961 conditional on

Figure 2: Application to opinion poll data (centered series)



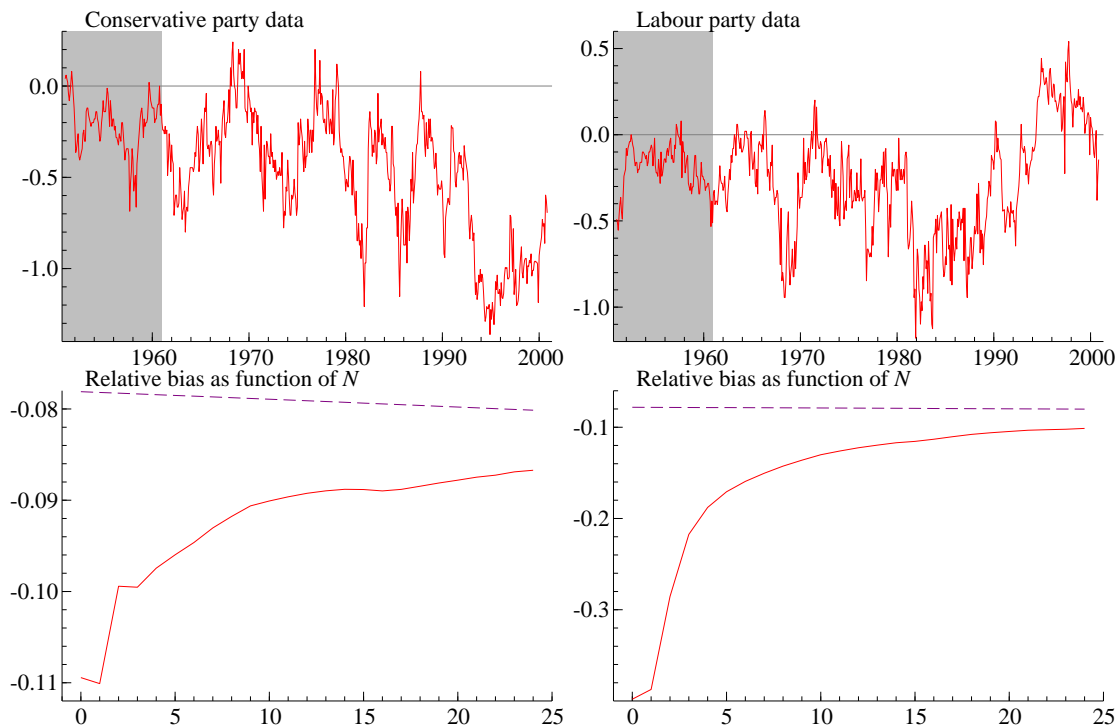
Note: The top panels show (transformed) opinion poll time series (centered by the sample average) and the bottom panels show the relative bias for the estimator of d as a function of the number of chosen initial values, N .

all initial values back to January 1951, and were the same (to two decimal places) whether based on centered or uncentered series.

Results are shown in Figures 2 and 3 for the centered and uncentered series, respectively. The top panels of each figure show the (logistically transformed) opinion poll data. The shaded areas mark the unobserved initial values January 1951 to December 1960. The bottom panels show the relative bias¹ in the estimator of d as a function of $N \in [0, 24]$, and the dashed straight line denotes the value of the fixed bias term, $-(479 - N)^{-1/2} 3\zeta_3 \zeta_2^{-3/2}$. In Figure 2 we note that the relative bias is larger for the Conservative party series because the last unobserved initial values are larger in absolute value than those of the Labour party series. In particular, if one does not condition on initial values and uses $N = 0$, the relative bias is -0.26 for the Conservative party series and -0.11 for the Labour party series. It is clear from the figure that the relative bias, particularly for the Conservative 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 from Figure 3 for the Labour party series.

¹In the calculation of $\xi_N(d)$, the infinite summation over t is truncated at 1000.

Figure 3: Application to opinion poll data (uncentered series)



Note: The top panels show (transformed) opinion poll time series (uncentered) and the bottom panels show the relative bias for the estimator of d as a function of the number of chosen initial values, N .

4 Conclusion

In this paper we have analyzed the effect of initial values on the asymptotic bias of the conditional maximum likelihood estimator, \hat{d} , of the fractional parameter, for $d_0 > 1/2$. This estimator is very often applied in practice, but although fractional models in principle depend on infinitely many initial values, the role of initial values has only been studied little in the literature.

We have shown that the asymptotic bias is of order $O(T^{-1})$, but since the asymptotic standard deviation is of order $O(T^{-1/2})$, the relevant quantity for testing and for constructing confidence sets is the relative bias, which is of order $O(T^{-1/2})$ and can be substantial.

When $d_0 = 1$ the choice $\tilde{X}_{-n} = 0$ for $n \geq 0$, which is commonly applied in practice, gives no contribution from initial values to the asymptotic second-order bias. Since the bias term is continuous in d_0 , the same is true for a (small) neighborhood of $d_0 = 1$.

In three applications of our theory we have demonstrated how to apply our theoretical results to (i) illustrate the bias numerically, (ii) discuss (non-)invariance of different fractional models to location and scale, and (iii) an empirical data set.

Appendix A The fractional coefficients

In this section we derive some useful results for the fractional coefficients (2) and their derivatives. The latter are given in the following lemma.

Lemma A.1 *Define the coefficient $a_j = 1_{\{j \geq 1\}} \sum_{k=1}^j k^{-1}$. The derivatives of $\pi_j(\cdot)$ are*

$$\mathbf{D}^m \log \pi_j(u) = (-1)^{m+1} \sum_{i=0}^{j-1} \frac{1}{(i+u)^m} \text{ for } u \neq 0, -1, \dots, -j+1 \text{ and } m \geq 1, \quad (20)$$

$$\mathbf{D}\pi_j(u)|_{u=-i} = (-1)^i \frac{i!(j-i-1)!}{j!} \text{ for } i = 0, 1, \dots, j-1 \text{ and } j \geq 2, \quad (21)$$

$$\mathbf{D}^2\pi_j(u)|_{u=-i} = 2\mathbf{D}\pi_j(u)|_{u=-i}(a_{j-i-1} - a_i) \text{ for } i = 0, 1, \dots, j-1 \text{ and } j \geq 2. \quad (22)$$

Proof. The result (20) follows by taking derivatives in (2) for $u \neq 0, -1, \dots, -j+1$. For $u = -i$ and $i = 0, 1, \dots, j-1$ we first define

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

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

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

which we evaluate at $u = -i$ for $i = 0, 1, \dots, j-1$. 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,

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

and (21) follows because $\mathbf{D}\pi_j(u) = \mathbf{D}P(u)/j!$, see (2). Similarly we find

$$\begin{aligned} \mathbf{D}^2P(u)|_{u=-i} &= \sum_{k \neq i} P_{ki}(-i) + \sum_{l \neq i} P_{il}(-i) = 2 \sum_{k \neq i} P_{ki}(-i) \\ &= 2 \sum_{k \neq i} \frac{P_i(-i)}{k-i} = 2P_i(-i) \sum_{k \neq i} \frac{1}{k-i} \\ &= 2P_i(-i)(a_{j-i-1} - a_i), \end{aligned}$$

which shows (22). ■

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

We next present some simple results for the fractional coefficients and their derivatives.

Lemma A.2 (a) *For $m \geq 0$, $j \geq 1$, and any u , it holds that*

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

(b) *For $u > 0$ and $j \rightarrow \infty$ it holds that*

$$\pi_j(u) = \frac{1}{\Gamma(u)} j^{u-1} (1 + o(1)).$$

Proof. *Proof of (a):* See JN (2010, Lemma B.3).

Proof of (b): To prove (b) we apply Stirling's formula,

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

where $|\epsilon_j(u)| \rightarrow 0$ as $j \rightarrow \infty$ for $u > 0$. ■

Lemma A.3 (a) *For any α, β it holds that*

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

(b) *For $\alpha + \beta < 1$ and $\beta > 0$ it holds that*

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

Proof. *Proof of (a):* See JN (2010, Lemma B.4).

Proof of (b): We first consider the summation from $k = 1$ to h :

$$\begin{aligned} h^{1-\alpha-\beta} \sum_{k=1}^h (k+h)^{\alpha-1} k^{\beta-1} (\log(k+h))^n &= h^{-1} \sum_{k=1}^h \left(\frac{k}{h} + 1\right)^{\alpha-1} \left(\frac{k}{h}\right)^{\beta-1} (\log(h) + \log(1 + \frac{k}{h}))^n \\ &\leq c(\log h)^n h^{-1} \sum_{k=1}^h \left(\frac{k}{h} + 1\right)^{\alpha-1} \left(\frac{k}{h}\right)^{\beta-1} \\ &= c(\log h)^n \int_0^1 (1+u)^{\alpha-1} u^{\beta-1} du (1 + o(1)) \text{ as } h \rightarrow \infty. \end{aligned}$$

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 choose $\varepsilon > 0$ such that $\varepsilon < 1 - (\alpha + \beta)$. Then $(k+h)^{\alpha-1} = (k+h)^{-\beta-\varepsilon} (k+h)^{\alpha+\beta-1+\varepsilon} \leq k^{-\beta-\varepsilon} h^{\alpha+\beta-1+\varepsilon}$ and $\log(k+h) \leq \log(2k) \leq c \log k$. It follows that

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

which is bounded by, see Lemma A.6, $ch^{\alpha+\beta-1+\varepsilon} h^{-\varepsilon} (\log h)^n = ch^{\alpha+\beta-1} (\log h)^n$. ■

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

(a) $\pi_0(u) = 1$ and $\pi_1(u) = u$,

(b) $\mathbf{D}^m \pi_0(u) = 0$ and $\mathbf{D}^m \pi_1(u) = 1_{\{m=1\}}$ for $m \geq 1$,

(c) $\mathbf{D} \pi_j(0) = j^{-1} 1_{\{j \geq 1\}}$, $\mathbf{D}^2 \pi_j(0) = 2j^{-1} a_{j-1} 1_{\{j \geq 2\}}$, and $|\mathbf{D}^m \pi_j(0)| \leq c j^{-1} (1 + \log j)^{m-1} 1_{\{j \geq 1\}}$ for $m \geq 1$,

(d) $\sum_{n=j}^k \mathbf{D}^m \pi_n(-u) = \mathbf{D}^m \pi_k(-u+1) - \mathbf{D}^m \pi_{j-1}(-u+1)$ for $m \geq 0$.

For $u > 0$,

(e) $\sum_{n=j}^{\infty} \mathbf{D}^m \pi_n(-u) = -\mathbf{D}^m \pi_{j-1}(-u+1)$ for $m \geq 0$.

For any u, v ,

(f) $\sum_{n=0}^k \pi_n(u) \pi_{k-n}(v) = \pi_k(u+v)$.

Proof. Result (a) is well known and follows trivially from (2), (b) follows by taking derivatives in (a), and (c) is a consequence of (21) and (22). To prove (d) 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 (d) with $m \geq 1$. From Lemma A.2, $\mathbf{D}^m \pi_k(-u+1) \leq c(1 + \log k)^m k^{-u} \rightarrow 0$ as $k \rightarrow \infty$ when $u > 0$ which shows (e). Finally, (f) follows from the Chu-Vandermonde identity, see Askey (1975, pp. 59-60). ■

Lemma A.5 For $d > 1/2$ it holds that

$$\begin{aligned} \sum_{n=0}^{\infty} \binom{d-1}{n}^2 &= \frac{\Gamma(2d-1)}{\Gamma(d)^2}, \\ \mathbf{D} \sum_{n=0}^{\infty} \binom{d-1}{n}^2 &= 2 \frac{\Gamma(2d-1)}{\Gamma(d)^2} (\Psi(2d-1) - \Psi(d)). \end{aligned}$$

Proof. With the notation $a_{(n)} = a(a+1)\dots(a+n-1)$, 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 $a = b = -(d-1)$ and $c = 1$, so that $c - a - b = 2d - 1 > 0$, it holds that

$$\sum_{n=0}^{\infty} \binom{d-1}{n}^2 = \sum_{n=0}^{\infty} \left(\frac{(d-1)(d-2)\dots(d-n)}{n!} \right)^2 = \sum_{n=0}^{\infty} \frac{a_{(n)} b_{(n)}}{c_{(n)} n!} = \frac{\Gamma(2d-1)}{\Gamma(d)^2},$$

with derivative $2(\Psi(2d-1) - \Psi(d))\Gamma(2d-1)/\Gamma(d)^2$. ■

The following summation results are special cases of Karamata's Theorem. Because they are well known we apply them in the remainder without special reference.

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

$$\begin{aligned} \sum_{n=1}^N (1 + \log n)^m n^\alpha &\leq c(1 + \log N)^m N^{\alpha+1} \text{ if } \alpha > -1, \\ \sum_{n=N}^{\infty} (1 + \log n)^m n^\alpha &\leq c(1 + \log N)^m N^{\alpha+1} \text{ if } \alpha < -1. \end{aligned}$$

Proof. See Theorems 1.5.8 and 1.5.10, respectively, in Bingham, Goldie, and Teugels (1987). ■

Appendix B Asymptotic analysis of the derivatives

The analysis of (8) and hence the proof of Theorem 1 requires asymptotic results for the first three derivatives of \tilde{L} evaluated at the true value, $d = d_0$. These in turn depend on the derivatives of $\tilde{\Delta}^d X_t = \Delta_+^d X_t + \Delta_-^d \tilde{X}_t$ for $d = d_0$.

For $t \geq 1$ we find from (4) in Lemma 1 that $\tilde{\Delta}^d X_t$ has the expression

$$\begin{aligned} \tilde{\Delta}^d X_t &= \Delta_+^{d-d_0} \varepsilon_t - \Delta_+^{d-d_0} \Delta_-^{d_0} X_t + \Delta_-^d \tilde{X}_t \\ &= \sum_{k=0}^{t-1} \pi_k(-d + d_0) \varepsilon_{t-k} \\ &\quad - \sum_{k=0}^{t-1} \pi_k(-d + d_0) \sum_{n=0}^{\infty} \pi_{t+n-k}(-d_0) X_{-n} + \sum_{n=0}^{\infty} \pi_{t+n}(-d) \tilde{X}_{-n}. \end{aligned} \quad (23)$$

Hence, derivatives of $\tilde{\Delta}^d X_t$ for $d = d_0$ are of the form $\mathbf{D}^m \tilde{\Delta}^{d_0} X_t = S_{mt}^+ + \eta_{mt}(d_0)$, where the stochastic term is S_{mt}^+ defined as

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

with

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

The deterministic term is

$$\eta_{mt}(d_0) = (-1)^{m+1} \left[\sum_{n=0}^{\infty} \sum_{k=0}^{t-1} \mathbf{D}^m \pi_k(0) \pi_{t-k+n}(-d_0) X_{-n} - \sum_{n=0}^{\infty} \mathbf{D}^m \pi_{t+n}(-d_0) \tilde{X}_{-n} \right]. \quad (24)$$

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

Lemma B.1 Suppose Assumptions 1-2 hold then the functions $\eta_{mt}(d)$ satisfy

$$|\eta_{0t}(d)| \leq ct^{-d}, \quad (25)$$

$$|\eta_{mt}(d)| \leq c(1 + \log t)^m t^{-\min(1,d)}, \quad m \geq 1. \quad (26)$$

For $d > 1/2$ it follows that

$$H_{mnT}(d) = \sigma_0^{-2} \sum_{t=1}^T \eta_{mt}(d) \eta_{nt}(d) \rightarrow \sigma_0^{-2} \sum_{t=1}^{\infty} \eta_{mt}(d) \eta_{nt}(d) = H_{mn}(d) < \infty, \quad (27)$$

and

$$K_{mnT}(d) = \sigma_0^{-2} \sum_{t=1}^T S_{mt}^+ \eta_{nt}(d) \xrightarrow{P} \sigma_0^{-2} \sum_{t=1}^{\infty} S_{mt}^+ \eta_{nt}(d) = K_{mn}(d), \quad (28)$$

where $E(K_{mnT}(d)) = E(K_{mn}(d)) = 0$ and $K_{mn}(d) < \infty$ almost surely.

Proof. *Proof of (25):* The expression for $\eta_{0t}(d)$ is

$$\eta_{0t}(d) = - \sum_{n=0}^{\infty} \pi_{t+n}(-d) X_{-n} + \sum_{n=0}^{\infty} \pi_{t+n}(-d) \tilde{X}_{-n} = - \sum_{n=0}^{\infty} \pi_{t+n}(-d) (X_{-n} - \tilde{X}_{-n}). \quad (29)$$

Using boundedness of initial values, Assumption 2, and the bound $|\pi_{t+n}(-d)| \leq c(t+n)^{-d-1}$, see Lemma A.2(a), the result follows.

Proof of (26): The remaining deterministic terms with $m \geq 1$ are evaluated using $|(-1)^m \mathbf{D}^m \pi_k(0)| \leq ck^{-1}(1 + \log k)^{m-1} \mathbf{1}_{\{k \geq 1\}}$, see Lemma A.4(c), and we find

$$\begin{aligned} |\eta_{mt}(d)| &\leq c \sum_{n=0}^{\infty} \sum_{k=1}^{t-1} k^{-1} (1 + \log k)^{m-1} (t - k + n)^{-d-1} + c \sum_{n=0}^{\infty} (1 + \log(t+n))^m (n+t)^{-d-1} \\ &\leq c [(1 + \log t)^{m-1} \sum_{k=1}^{t-1} k^{-1} (t-k)^{-d} + (1 + \log t)^m t^{-d}] \\ &\leq c(1 + \log t)^m (t^{-\min(1,d)} + t^{-d}) \leq c(1 + \log t)^m t^{-\min(1,d)}. \end{aligned}$$

We have used the inequality, see Lemma A.3,

$$\sum_{k=1}^{t-1} k^{-1} (t-k)^{-d} \leq c(1 + \log t) t^{\max(-1, -d, -d)} \leq c(1 + \log t) t^{-\min(1,d)}.$$

Proof of (27): From (25) and (26) we find $|\eta_{mt}(d) \eta_{nt}(d)| \leq c(1 + \log t)^{m+n} t^{-2\min(1,d)}$ so that $\sum_{t=1}^{\infty} |\eta_{mt}(d) \eta_{nt}(d)| < \infty$ because $2\min(1, d) > 1$ for $d > 1/2$.

Proof of (28): We have

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

For some small $\delta > 0$ to be chosen subsequently, we use the evaluations $|\eta_{nt}(d)| \leq ct^{-\min(1,d)+\delta}$, $|\mathbf{D}^m \pi_{t-k}(0)| \leq c(t-k)^{-1+\delta}$, 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

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

For $k = 1, \dots, T-1$ we get $\sum_{t=T}^{\infty} (t-k)^{-1-\delta} \leq c(T-k)^{-\delta}$ and

$$c \sum_{k=1}^{T-1} k^{-2\min(1,d)+6\delta} (T-k)^{-\delta} \leq c(1+\log T) T^{\max(-\delta, -2\min(1,d)+6\delta, -2\min(1,d)+5\delta+1)} \rightarrow 0$$

if $5\delta < 2\min(1,d) - 1$, see Lemma A.3. For $k \geq T$ we find $\sum_{t=k+1}^{\infty} (t-k)^{-1-\delta} \leq c$ and then $\sum_{k=T}^{\infty} k^{-2\min(1,d)+6\delta} \rightarrow 0$ for $6\delta < 2\min(1,d) - 1$. This shows that $\sum_{t=1}^T S_{mt}^+ \eta_{nt}(d) \xrightarrow{P} \sum_{t=1}^{\infty} S_{mt}^+ \eta_{nt}(d)$. ■

We next define, for $m, n = 0, 1, 2, 3, m+n \leq 3$, the product moments of the stochastic terms, S_{mt}^+ , as

$$M_{mnT}^+ = \sigma_0^{-2} T^{-1/2} \sum_{t=1}^T (S_{mt}^+ S_{nt}^+ - E(S_{mt}^+ S_{nt}^+)), \quad (30)$$

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

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

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

$$E(S_{0t}^+ S_{mt}^+) = E(S_{0t} S_{mt}) = 0 \text{ for } m \geq 1. \quad (31)$$

Lemma B.2 *Suppose Assumption 1 holds. Then for $\zeta_2 = \sum_{j=1}^{\infty} j^{-2} = \frac{\pi^2}{6} \simeq 1.6449$ and $\zeta_3 = \sum_{j=1}^{\infty} j^{-3} \simeq 1.2021$, see (9),*

$$E(M_{01T}^2) = \sigma_0^{-2} T^{-1} \sum_{t=1}^T E(S_{1t}^2) = \zeta_2, \quad (32)$$

$$E(M_{01T} M_{02T}) = \sigma_0^{-4} T^{-1} \sum_{s,t=1}^T E(S_{0t} S_{1t} S_{0s} S_{2s}) = \sigma_0^{-2} T^{-1} \sum_{t=1}^T E(S_{1t} S_{2t}) = -2\zeta_3, \quad (33)$$

$$E(M_{01T} M_{11T}) = \sigma_0^{-4} T^{-1} \sum_{s,t=1}^T E(S_{0t} S_{1t} S_{1s}^2) = -4\zeta_3. \quad (34)$$

Proof. *Proof of (32):* From $S_{0t} = \varepsilon_t$, $S_{1t} = -\sum_{k=1}^{\infty} k^{-1} \varepsilon_{t-k}$, and (31) we find

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

Proof of (33): We find using the expressions for S_{0t} , S_{1t} , and $S_{2t} = 2 \sum_{j=2}^{\infty} j^{-1} a_{j-1} \varepsilon_{t-j}$, $a_j = 1_{\{j \geq 1\}} \sum_{k=1}^j k^{-1}$, together with (31) that

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

and

$$\begin{aligned} \sigma_0^{-2} T^{-1} \sum_{t=1}^T E(S_{1t} S_{2t}) &= -2\sigma_0^{-2} T^{-1} \sum_{t=1}^T E\left[\sum_{k=1}^{\infty} k^{-1} \varepsilon_{t-k} \sum_{j=2}^{\infty} (j^{-1} a_{j-1}) \varepsilon_{t-j}\right] \\ &= -2T^{-1} \sum_{t=1}^T \sum_{j=2}^{\infty} j^{-2} \sum_{k=1}^{j-1} k^{-1} = -2 \sum_{j=2}^{\infty} j^{-2} \sum_{k=1}^{j-1} k^{-1} = -2\kappa_3 \end{aligned} \quad (35)$$

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

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

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 (31) and (35), which proves the result.

Proof of (34): Next we find, using the expressions for S_{mt} and (31) that

$$\begin{aligned} E(M_{01T} M_{11T}) &= \sigma_0^{-4} T^{-1} \sum_{s,t=1}^T E(S_{0t} S_{1t} S_{1s}^2) \\ &= -T^{-1} \sum_{s,t=1}^T E\left[\varepsilon_t \sum_{k=-\infty}^{t-1} (t-k)^{-1} \varepsilon_k \sum_{j=-\infty}^{s-1} (s-j)^{-1} \varepsilon_j \sum_{i=-\infty}^{s-1} (s-i)^{-1} \varepsilon_i\right]. \end{aligned}$$

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 $s - k = s - t + t - k$,

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

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

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

This proves (34) and completes the proof of Lemma B.2. ■

Lemma B.3 *Suppose Assumption 1 holds. Then, for $T \rightarrow \infty$, 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 (32), (33), and (34) 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. *Proof for $\{M_{mnT}\}$:* We apply a result by Giraitis and Taqqu (1998) on limit distributions of quadratic forms of linear processes. We define the cross covariance function

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

and find $r_{00}(t) = \sigma_0^2 1_{\{t=0\}}$, $r_{m0}(t) = \sigma_0^2(-1)^m D^m \pi_{-t}(0) 1_{\{t \leq -1\}}$, and $r_{0n}(t) = \sigma_0^2(-1)^n D^n \pi_t(0)$. For $m, n \geq 1$ we find the following evaluation for a small $\delta > 0$,

$$\begin{aligned} |r_{mn}(t)| &\leq c \sum_{k=1}^{\infty} (1 + \log(t+k))^{m-1} (1 + \log k)^{n-1} (t+k)^{-1} k^{-1} \\ &\leq c \sum_{k=1}^{\infty} (t+k)^{-1+\delta} k^{-1+\delta} \leq ct^{-1+3\delta}, \end{aligned}$$

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 (32), (33), and (34) in Lemma B.2.

Proof for $\{M_{mnT}^+\}$: We show that $E(M_{mnT} - M_{mnT}^+)^2 \rightarrow 0$. We find

$$M_{mnT} - M_{mnT}^+ = \sigma_0^{-2} T^{-1/2} \sum_{t=1}^T (S_{mt}^+ S_{nt}^- + S_{mt}^- S_{nt}^+ + S_{mt}^- S_{nt}^- - E(S_{mt}^+ S_{nt}^- + S_{mt}^- S_{nt}^+ + S_{mt}^- S_{nt}^-)), \quad (36)$$

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

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

$$|E(S_{mt}^- S_{nt}^-)| = \sigma_0^2 \sum_{k=t}^{\infty} |D^m \pi_k(0) D^n \pi_k(0)| \leq c \sum_{k=t}^{\infty} (1 + \log k)^{m+n} k^{-2} \leq c(1 + \log t)^{m+n} t^{-1}$$

so that

$$T^{-1/2} \sum_{t=1}^T E(S_{mt}^- S_{nt}^-) \leq cT^{-1/2} (1 + \log T)^{m+n+1} \rightarrow 0. \quad (37)$$

Proof of $\text{Var}(T^{-1/2} \sum_{t=1}^T S_{mt}^+ S_{nt}^-) \rightarrow 0$: The first two terms of (36) are handled the same way. We find because (S_{mt}^+, S_{ns}^+) is independent of (S_{mt}^-, S_{ns}^-) that

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

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

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

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

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

Similarly for

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

we find

$$\begin{aligned} |E(S_{nt}^- S_{ns}^-)| &\leq c \sum_{i=-\infty}^0 (t-i)^{-1+\delta} (s-i)^{-1+\delta} = c \sum_{i=0}^{\infty} (t+i)^{-1+\delta} (s+i)^{-1+\delta} \\ &\leq c(s-t)^{-1+3\delta} \sum_{i=0}^{\infty} (t+i)^{-1-\delta} \leq c(s-t)^{-1+3\delta} t^{-\delta}. \end{aligned}$$

Finally, we can evaluate the variance as

$$\begin{aligned} \text{Var}(T^{-1/2} \sum_{t=1}^T S_{mt}^+ S_{nt}^-) &\leq cT^{-1} \sum_{1 \leq t < s \leq T} (s-t)^{-1+3\delta} t^{-\delta} (s-t)^{-1+3\delta} \\ &= cT^{-1} \sum_{h=1}^{T-1} h^{-2+6\delta} \sum_{t=1}^{T-h} t^{-\delta} \leq cT^{-1} T^{1-\delta} \rightarrow 0. \end{aligned}$$

Proof of $\text{Var}(T^{-1/2} \sum_{t=1}^T S_{mt}^- S_{nt}^-) \rightarrow 0$: The last term of (36) has variance

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

and the first term is

$$\begin{aligned} &T^{-1} \sum_{s,t=1}^T E(S_{mt}^- S_{nt}^- S_{ms}^- S_{ns}^-) \\ &= T^{-1} \sum_{s,t=1}^T \sum_{i,j,k,p=-\infty}^0 E(\mathbf{D}^m \pi_{t-i}(0) \varepsilon_i \mathbf{D}^n \pi_{t-j}(0) \varepsilon_j \mathbf{D}^m \pi_{s-k}(0) \varepsilon_k \mathbf{D}^n \pi_{s-p}(0) \varepsilon_p). \end{aligned}$$

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

$$\begin{aligned} A_1 &\leq cT^{-1} \sum_{s,t=1}^T \sum_{i=0}^{\infty} (1 + \log(t+i))^m (1 + \log(s+i))^m (t+i)^{-1} (s+i)^{-1} \\ &\quad \times \sum_{j=0}^{\infty} (1 + \log(t+j))^n (1 + \log(s+j))^n (s+j)^{-1} (t+j)^{-1} \\ &\leq cT^{-1} \sum_{s,t=1}^T [\sum_{i=0}^{\infty} (t+i)^{-1+\delta} (s+i)^{-1+\delta}]^2 \leq cT^{-1} \sum_{1 \leq t < s \leq T} [\sum_{i=0}^{\infty} (t+i)^{-1+\delta} (s+i)^{-1+\delta}]^2. \end{aligned}$$

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

$$\sum_{i=0}^{\infty} (t+i)^{-1+\delta} (s+i)^{-1+\delta} \leq \sum_{i=0}^{\infty} (s-t)^{-1+3\delta} (t+i)^{-1-\delta} \leq (s-t)^{-1+3\delta} t^{-\delta}$$

and hence

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

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

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

■

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

Lemma B.4 *Let the process $X_t, t = 1, \dots, T$, be generated by model (1) and suppose Assumptions 1-2 are satisfied. Then the derivatives satisfy*

$$T^{-1/2} \mathbf{D} \tilde{L}(d_0) = -M_{01T}^+ - T^{-1/2} (H_{01}(d_0) + K_{01}(d_0) + K_{10}(d_0)) + o_P(T^{-1/2}), \quad (38)$$

$$T^{-1} \mathbf{D}^2 \tilde{L}(d_0) = -\zeta_2 - T^{-1/2} (M_{11T}^+ + M_{02T}^+) + O_P(T^{-1}(\log T)), \quad (39)$$

$$T^{-1} \mathbf{D}^3 \tilde{L}(d_0) = 6\zeta_3 + O_P(T^{-1/2}). \quad (40)$$

Proof. *Proof of (38):* We find using Lemma B.1 that $H_{mnT}(d_0) = H_{mn}(d_0) + o(1)$ and $K_{mnT}(d_0) = K_{mn}(d_0) + o_P(1)$, so that

$$\begin{aligned} T^{-1/2} \mathbf{D} \tilde{L}(d_0) &= -T^{-1/2} \sigma_0^{-2} \sum_{t=1}^T (S_{0t}^+ + \eta_{0t}(d_0))(S_{1t}^+ + \eta_{1t}(d_0)) \\ &= -M_{01T}^+ - T^{-1/2} (H_{01T}(d_0) + K_{01T}(d_0) + K_{10T}(d_0)) \\ &= -M_{01T}^+ - T^{-1/2} (H_{01}(d_0) + K_{01}(d_0) + K_{10}(d_0)) + o_P(T^{-1/2}). \end{aligned}$$

Proof of (39): We find

$$\begin{aligned}
T^{-1}\mathbf{D}^2\tilde{L}(d_0) &= -\sigma_0^{-2}T^{-1}\sum_{t=1}^T[(\mathbf{D}\tilde{\Delta}^{d_0}X_t)^2 + (\tilde{\Delta}^{d_0}X_t)(\mathbf{D}^2\tilde{\Delta}^{d_0}X_t)] \\
&= -\sigma_0^{-2}T^{-1}\sum_{t=1}^T(S_{1t}^+ + \eta_{1t}(d_0))^2 - \sigma_0^{-2}T^{-1}\sum_{t=1}^T(S_{0t}^+ + \eta_{0t}(d_0))(S_{2t}^+ + \eta_{2t}(d_0)) \\
&= -T^{-1/2}M_{11T}^+ - \sigma_0^{-2}T^{-1}\sum_{t=1}^TE(S_{1t}^+)^2 - T^{-1/2}M_{02T}^+ \\
&\quad - T^{-1}(H_{11T}(d_0) + 2K_{11T}(d_0) + K_{02T}(d_0) + K_{20T}(d_0) + H_{02T}(d_0)).
\end{aligned}$$

Again using Lemma B.1 it holds that $H_{mnT}(d_0) = O(1)$ and $K_{mnT}(d_0) = O_P(1)$ such that

$$\begin{aligned}
T^{-1}\mathbf{D}^2\tilde{L}(d_0) &= -\sigma_0^{-2}T^{-1}\sum_{t=1}^TE(S_{1t}^+)^2 - T^{-1/2}(M_{11T}^+ + M_{02T}^+) + O_P(T^{-1}) \\
&= -\zeta_2 - T^{-1/2}(M_{11T}^+ + M_{02T}^+) + O_P(T^{-1}(\log T))
\end{aligned}$$

using also (32) and (37).

Proof of (40): We analyze the third derivative similarly,

$$\begin{aligned}
T^{-1}\mathbf{D}^3\tilde{L}(d_0) &= -\sigma_0^{-2}T^{-1}\sum_{t=1}^T[3(\mathbf{D}\tilde{\Delta}^{d_0}X_t)(\mathbf{D}^2\tilde{\Delta}^{d_0}X_t) + (\tilde{\Delta}^{d_0}X_t)(\mathbf{D}^3\tilde{\Delta}^{d_0}X_t)] \\
&= -3\sigma_0^{-2}T^{-1}\sum_{t=1}^T(S_{1t}^+ + \eta_{1t}(d_0))(S_{2t}^+ + \eta_{2t}(d_0)) - \sigma_0^{-2}T^{-1}\sum_{t=1}^T(S_{0t}^+ + \eta_{0t}(d_0))(S_{3t}^+ + \eta_{3t}(d_0)) \\
&= -3T^{-1/2}M_{12T}^+ - 3\sigma_0^{-2}T^{-1}\sum_{t=1}^TE(S_{1t}^+S_{2t}^+) - T^{-1/2}M_{03T}^+ + O_P(T^{-1}) \\
&= 6\zeta_3 + O_P(T^{-1/2}),
\end{aligned}$$

where the second-to-last equality uses Lemma B.1 and the last equality uses Lemmas B.2 and B.3, (33), and (37). \blacksquare

Appendix C Proof of Theorem 1

First we note that, because $|d^* - d_0| \leq |\hat{d} - d_0| \xrightarrow{P} 0$ and the product moments in (30) are tight (by Lemma C.4 of JN (2010)), we can apply JN (2010, Lemma A.3) to conclude that

$$\mathbf{D}^3\tilde{L}(d^*) = \mathbf{D}^3\tilde{L}(d_0) + o_P(1). \quad (41)$$

Using this result and Lemma B.4 we can approximate the second term on the right-hand side of (8) by replacing $T^{-1/2}\mathbf{D}\tilde{L}(d_0)$ by $-M_{01T}^+$, $T^{-1}\mathbf{D}^2\tilde{L}(d_0)$ by $-\zeta_2$, and $T^{-1}\mathbf{D}^3\tilde{L}(d_0)$ by $6\zeta_3$. Thus,

$$-\frac{1}{2}\left(\frac{T^{-1/2}\mathbf{D}\tilde{L}(d_0)}{T^{-1}\mathbf{D}^2\tilde{L}(d_0)}\right)^2\frac{T^{-1}\mathbf{D}^3\tilde{L}(d_0)}{T^{-1}\mathbf{D}^2\tilde{L}(d_0)} = (M_{01T}^+)^2\frac{3\zeta_3}{\zeta_2^3} + o_P(1).$$

In the first term on the right-hand side of (8) we use $1/(1+z) = 1 - z + O(z^2)$ to obtain the expansion

$$\frac{1}{T^{-1}\mathbf{D}^2\tilde{L}(d_0)} = \frac{1}{-\zeta_2} - T^{-1/2} \frac{T^{1/2}(T^{-1}\mathbf{D}^2\tilde{L}(d_0) + \zeta_2)}{\zeta_2^2} + O_P(T^{-1}).$$

We next evaluate $-T^{-1/2}\mathbf{D}\tilde{L}(d_0)/(T^{-1}\mathbf{D}^2\tilde{L}(d_0))$ using this expansion together with the expression (38) for $T^{-1/2}\mathbf{D}\tilde{L}(d_0)$, and find that it equals

$$\begin{aligned} & \frac{T^{-1/2}\mathbf{D}\tilde{L}(d_0)}{\zeta_2} + T^{-1/2} \frac{T^{-1/2}\mathbf{D}\tilde{L}(d_0)T^{1/2}(T^{-1}\mathbf{D}^2\tilde{L}(d_0) + \zeta_2)}{\zeta_2^2} + O_P(T^{-1}) \\ &= \frac{-M_{01T}^+}{\zeta_2} - T^{-1/2} \frac{H_{01}(d_0) + K_{01}(d_0) + K_{10}(d_0)}{\zeta_2} + T^{-1/2} \frac{M_{01T}^+(M_{02T}^+ + M_{11T}^+)}{\zeta_2^2} + o_P(T^{-1/2}). \end{aligned}$$

This gives the expansion

$$\begin{aligned} \hat{d} - d &= T^{-1/2}\tilde{A}_T + T^{-1}\tilde{B}_T + o_P(T^{-1}), \\ \tilde{A}_T &= -\frac{M_{01T}^+}{\zeta_2}, \\ \tilde{B}_T &= -\frac{H_{01}(d_0)}{\zeta_2} - \frac{K_{01}(d_0) + K_{10}(d_0)}{\zeta_2} + \frac{M_{01T}^+(M_{02T}^+ + M_{11T}^+)}{\zeta_2^2} + (M_{01T}^+)^2 \frac{3\zeta_3}{\zeta_2^3}. \end{aligned} \tag{42}$$

The expectation of \tilde{A}_T and its limit in distribution is zero, see Lemma B.3, and it therefore does not contribute to the asymptotic bias. The first term in \tilde{B}_T is deterministic and its contribution to the bias is simply

$$-\frac{\xi(d_0)}{\zeta_2} = -\frac{H_{01}(d_0)}{\zeta_2} = -\frac{\sigma_0^2}{\zeta_2} \sum_{t=1}^T \eta_{0t}(d_0)\eta_{1t}(d_0), \tag{43}$$

see Lemma B.1. Both $K_{01}(d_0)$ and $K_{10}(d_0)$ are mean zero random variables and do not contribute to the bias. The terms $M_{01T}^+M_{02T}^+$ and $M_{01T}^+M_{11T}^+$ both converge in distribution by Lemma B.3, and the expectations of the limit distributions are $-2\zeta_3$ and $-4\zeta_3$, respectively, see (33) and (34) of Lemma B.2. Hence, the contribution from the third term is $-6\zeta_3/\zeta_2^2$. For the last term in \tilde{B}_T we find the limit in distribution,

$$(M_{01T}^+)^2 \frac{3\zeta_3}{\zeta_2^3} \xrightarrow{D} Z_{01}^2 \frac{3\zeta_3}{\zeta_2^2},$$

where $Z_{01} \sim N(0, 1)$, see Lemma B.3 and (32), and the expectation of the limit is $3\zeta_3/\zeta_2^2$.

Appendix D Proof of Theorem 2

Proof of (13) with $\ell = 0$: Under Assumption 3.0, where $\tilde{X}_{-n} = X_{-n}$, $n < N$, and $\tilde{X}_{-n} = 0$, $n \geq N$, and $d > 1/2$, we find that (11) reduces to the expression for $\xi_N(d)$ for $\ell = 0$.

Proof of (13) with $\ell = 1$: Next consider Assumption 3.1, where $\tilde{X}_0 = X_0$, $\Delta\tilde{X}_{-n} = \Delta X_{-n}1_{\{n < N\}}$, and $d > 3/2$. The expression (43) shows that $\xi_N(d) = \sigma_0^{-2} \sum_{t=1}^{\infty} \eta_{0t}(d)\eta_{1t}(d)$, and we discuss $\eta_{0t}(d)$ and $\eta_{1t}(d)$ separately.

We apply the well known Beveridge-Nelson expansion,

$$A(z) = \sum_{n=0}^{\infty} A_n z^n = A + (1-z) \sum_{n=0}^{\infty} A_n^* z^n,$$

where $\sum_{n=0}^{\infty} A_n = A$ and $A_n^* = -\sum_{m=n+1}^{\infty} A_m$, $n = 0, 1, 2, \dots$. For $\eta_{0t}(d)$ given in (29) we find

$$\eta_{0t}(d) = \sum_{n=0}^{\infty} \pi_{t+n}(-d)(X_{-n} - \tilde{X}_{-n}) = A(X_0 - \tilde{X}_0) + \sum_{n=0}^{\infty} A_n^* \Delta(X_{-n} - \tilde{X}_{-n}), \quad (44)$$

where $A = \sum_{n=0}^{\infty} \pi_{t+n}(-d) = -\pi_{t-1}(-d+1)$ and $A_n^* = -\sum_{m=n+1}^{\infty} \pi_{t+m}(-d) = \pi_{t+n}(-d+1)$, $n = 0, 1, 2, \dots$, see Lemma A.4(e). If $X_0 = \tilde{X}_0$ and $\Delta\tilde{X}_{-n} = \Delta X_{-n} 1_{\{n < N\}}$, we get

$$\eta_{0t}(d) = \sum_{n=N}^{\infty} \pi_{t+n}(-d+1) \Delta X_{-n}.$$

The expression for $\eta_{1t}(d)$ is found from (24),

$$\eta_{1t}(d) = \sum_{k=0}^{t-1} \mathbf{D}\pi_k(d)|_{d=0} \sum_{n=0}^{\infty} \pi_{t-k+n}(-d) X_{-n} - \sum_{n=0}^{\infty} \mathbf{D}\pi_{t+n}(-d) \tilde{X}_{-n}.$$

The same analysis, see (44), implies that this equals

$$\begin{aligned} \eta_{1t}(d) &= \sum_{k=0}^{t-1} \mathbf{D}\pi_k(d)|_{d=0} [-\pi_{t-k-1}(-d+1)X_0 + \sum_{n=0}^{\infty} \pi_{t-k+n}(-d+1)\Delta X_{-n}] \\ &\quad + \mathbf{D}\pi_{t-1}(-d+1)\tilde{X}_0 - \sum_{n=0}^{\infty} \mathbf{D}\pi_{t-1+n}(-d+1)\Delta\tilde{X}_{-n} \\ &= -\mathbf{D}\pi_{t-1}(-d+1)X_0 + \sum_{k=0}^{t-1} \mathbf{D}\pi_k(d)|_{d=0} \sum_{n=0}^{\infty} \pi_{t-k+n}(-d+1)\Delta X_{-n} \\ &\quad + \mathbf{D}\pi_{t-1}(-d+1)\tilde{X}_0 - \sum_{n=0}^{\infty} \mathbf{D}\pi_{t-1+n}(-d+1)\Delta\tilde{X}_{-n}, \end{aligned}$$

where the second equality uses $\sum_{k=0}^{t-1} \mathbf{D}\pi_k(d)|_{d=0} \pi_{t-k-1}(-d+1) = \mathbf{D}\pi_{t-1}(-d+1)$, see Lemma A.4(f). If $X_0 = \tilde{X}_0$ and $\Delta\tilde{X}_{-n} = \Delta X_{-n} 1_{\{n < N\}}$ we thus find

$$\eta_{1t}(d) = \sum_{k=0}^{t-1} \mathbf{D}\pi_k(0) \sum_{n=0}^{\infty} \pi_{t-k+n}(-d+1)\Delta X_{-n} - \sum_{n=0}^{N-1} \mathbf{D}\pi_{t-1+n}(-d+1)\Delta X_{-n},$$

which gives the expression (13) for $\ell = 1$, see Lemma A.4(c).

Proof of (14): We define $d_\ell = d - \ell$ and evaluate $\xi_N(d) = \sigma_0^{-2} \sum_{t=1}^{\infty} \eta_{0t}(d_\ell) \eta_{1t}(d_\ell)$ by first evaluating

$$|\eta_{0t}(d_\ell)| = \left| \sum_{n=N}^{\infty} \pi_{t+n}(-d_\ell) \Delta^\ell (X_{-n} - \tilde{X}_{-n}) \right| \leq c \sum_{n=N}^{\infty} (t+n)^{-d_\ell-1} \leq c(t+N)^{-d_\ell} \quad (45)$$

using Assumptions 2 and 3 and Lemma A.2(a). Applying this bound and (26) in Lemma B.1 we find that $\xi_N(d)$ is bounded as

$$|\xi_N(d)| \leq \sigma_0^{-2} \sum_{t=1}^{\infty} |\eta_{0t}(d_\ell) \eta_{1t}(d_\ell)| \leq c \sum_{t=1}^{\infty} (t+N)^{-d_\ell} (1+\log t) t^{-\min(1, d_\ell)}.$$

From the relation $(t+N)^{-d_\ell} \leq t^{\min(1, d_\ell)-1-\epsilon} (1+N)^{-d_\ell+1+\epsilon-\min(1, d_\ell)}$ for $0 < \epsilon < \min(d_\ell, 2d_\ell-1)$, we find the bound

$$c(1+N)^{-\min(d_\ell, 2d_\ell-1)+\epsilon} \sum_{t=1}^{\infty} (1+\log t) t^{-1-\epsilon} \leq c(1+N)^{-\delta}$$

with $\delta = \min(d_\ell, 2d_\ell-1) - \epsilon$, which proves (14) for $\ell = 0, 1$.

References

1. Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions*, National Bureau of Standards, Washington D.C.
2. Andrews, D. W. K., Lieberman, O., and Marmer, V. (2006). Higher-order improvements of the parametric bootstrap for long-memory Gaussian processes. *Journal of Econometrics* **133**, pp. 673–702.
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., Osterreider, D., Sizova, N., and Tauchen, G. (2012) Risk and return: long-run relationships, fractional cointegration, and return predictability. CREATES research paper, Aarhus University. Forthcoming in *Journal of Financial Economics*.
6. Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis, Forecasting, and Control*, Holden-Day, San Francisco.
7. Carlini, F., Manzonei, 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.
8. 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.
9. Dolado, J. J., Gonzalo, J., and Mayoral, L. (2002). A fractional Dickey-Fuller test for unit roots. *Econometrica* **70**, pp. 1963–2006.
10. Giraitis, L. and Taqqu, M. (1998). Central limit theorems for quadratic forms with time domain conditions. *Annals of Probability* **26**, pp. 377–398.
11. Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and its Application*, Academic Press, New York.
12. Hualde, J. and Robinson, P. M. (2011). Gaussian pseudo-maximum likelihood estimation of fractional time series models. *Annals of Statistics* **39**, pp. 3152–3181.
13. Johansen, S. (2008). A representation theory for a class of vector autoregressive models for fractional processes. *Econometric Theory* **24**, pp. 651–676.
14. Johansen, S. and Nielsen, M. Ø. (2010). Likelihood inference for a nonstationary fractional autoregressive model. *Journal of Econometrics* **158**, pp. 51–66.

15. Johansen, S. and Nielsen, M. Ø. (2012). Likelihood inference for a fractionally cointegrated vector autoregressive model. *Econometrica* **80**, pp. 2667–2732.
16. Lawley, D. N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* **43**, pp. 295–303.
17. Li, W. K. and McLeod, A. I. (1986). Fractional time series modelling. *Biometrika* **73**, pp. 217–221.
18. 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.
19. Nielsen, M. Ø. (2012). Asymptotics for the conditional-sum-of-squares estimator in fractional time series models. QED working paper 1259, Queen’s University.
20. Osterrieder, D. and Schotman, P. C. (2011). Predicting returns with a co-fractional VAR model. CREATES research paper, Aarhus University.
21. Robinson, P. M. (1994). Efficient tests of nonstationary hypotheses, *Journal of the American Statistical Association* **89**, pp. 1420–1437.
22. 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.
23. Tschernig, R., Weber, E., and Weigand, R. (2010). Long-run identification in a fractionally integrated system. Working paper, University of Regensburg.