Queen’s Economics Department Working Paper No. 1172

Likelihood inference for a nonstationary fractional autoregressive model

Søren Johansen
University of Copenhagen and CREATES

Morten Ørregaard Nielsen
Queen’s University and CREATES

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

2-2010

Likelihood inference for a nonstationary fractional autoregressive model

Søren Johansen*
University of Copenhagen and CREATES

Morten Ørregaard Nielsen†
Queen's University and CREATES

February 2, 2010

Abstract

This paper discusses model-based inference in an autoregressive model for fractional processes which allows the process to be fractional of order d or $d-b$. Fractional differencing involves infinitely many past values and because we are interested in nonstationary processes we model the data $X_1, \ldots, X_T$ given the initial values $X_{-n},\ n = 0, 1, \ldots$, as is usually done. The initial values are not modeled but assumed to be bounded. This represents a considerable generalization relative to all previous work where it is assumed that initial values are zero. For the statistical analysis we assume the conditional Gaussian likelihood and for the probability analysis we also condition on initial values but assume that the errors in the autoregressive model are i.i.d. with suitable moment conditions.

We analyze the conditional likelihood and its derivatives as stochastic processes in the parameters, including d and b, and prove that they converge in distribution. We use the results to prove consistency of the maximum likelihood estimator for d, b in a large compact subset of $\{1/2 < b < d < \infty\}$, and to find the asymptotic distribution of the estimators and the likelihood ratio test of the associated fractional unit root hypothesis. The limit distributions contain the fractional Brownian motion of type II.

Key words and phrases: Dickey-Fuller test, fractional unit root, likelihood inference.

JEL Classification: C22.

1 Introduction and motivation

Nonstationary autoregressive models have been studied intensively over the past three decades. In the usual autoregressive model, we consider the nonstationarity implied by a unit root in the associated autoregressive polynomial. Recently, much attention has been given to alternative models of nonstationarity such as fractional models characterized by a fractional

*Address: Department of Economics, University of Copenhagen, Øster Farimagsgade 5, Bld. 26, DK-1353 Copenhagen K, Denmark. Email: soren.johansen@econ.ku.dk

†Corresponding author. Address: Morten Ørregaard Nielsen, Department of Economics, Dunning Hall Room 307, 94 University Avenue, Queen's University, Kingston, Ontario K7L 3N6, Canada. Email: mon@econ.queensu.ca

rather than integer integration operator, see e.g. Baillie (1996) or Henry and Zaffaroni (2003) for reviews and examples, and it appears important to allow fractional orders of integration (or fractionality) in the time series model. Formal statistical tests of the unit root hypothesis are of additional interest to economists because they can help to evaluate the nature of the nonstationarity that most macroeconomic and financial time series data exhibit.

The autoregressive model with $k+1$ lags for the univariate time series $X_t,\ t=1,\ldots,T$, conditional on initial values $X_{-n},\ n=0,\ldots,k$, written in levels and differences, is

$$\Delta X_t = \pi X_{t-1} + \sum_{i=1}^{k} \phi_i \Delta X_{t-i} + \varepsilon_t,\ t=1,\ldots,T,$$

where ε_t is i.i.d. $(0,\sigma^2)$.

A corresponding *fractional autoregressive model*, conditional on infinitely many initial values $X_{-n},\ n=0,\ldots,$ was obtained by Johansen (2008) by replacing the difference and lag operators Δ and $L=1-\Delta$ by their fractional counterparts Δ^b and $L_b = 1-\Delta^b$ and applying Δ^{d-b} to X_t :

$$\Delta^d X_t = \pi \Delta^{d-b} L_b X_t + \sum_{i=1}^{k} \phi_i \Delta^d L_b^i X_t + \varepsilon_t,\ t=1,\ldots,T. \tag{1}$$

In the statistical model defined by (1) the parameters $(d,b,\phi_1,\ldots,\phi_k,\pi,\sigma^2)$ are unrestricted except $d \geq b > 1/2$ and $\sigma^2 > 0$. Note that if $k=0$ in model (1) the parameter b is not identified under the unit root null $\pi = 0$, see section 2.3. This motivates study of the simpler model with $d=b$,

$$\Delta^d X_t = \pi L_d X_t + \sum_{i=1}^{k} \phi_i \Delta^d L_d^i X_t + \varepsilon_t,\ t=1,\ldots,T, \tag{2}$$

and in the simplest case with $k=0$, model (2) reduces to

$$\Delta^d X_t = \pi L_d X_t + \varepsilon_t,\ t=1,\ldots,T, \tag{3}$$

which we consider separately in some of our results.

The properties of the solution of model (1) depend on the characteristic function for the process

$$\begin{aligned}\Pi(z) &= (1-z)^d - \pi(1-z)^{d-b}(1-(1-z)^b) - \sum_{i=1}^{k} \phi_i (1-z)^d (1-(1-z)^b)^i \\ &= (1-z)^{d-b}\xi(1-(1-z)^b) = \sum_{j=-1}^{k} \rho_j (1-z)^{d+jb}, \qquad (4) \\ \xi(y) &= 1 - y - \pi y - \sum_{i=1}^{k} \phi_i (1-y) y^i = \sum_{j=0}^{k+1} \rho_{j-1} (1-z)^{jb},\end{aligned}$$

where $\sum_{j=-1}^{k} \rho_j = 1$ and $y = 1-(1-z)^b$. Note that $\Pi(z)$ is a polynomial in z if and only if d and b are non-negative integers, whereas $\xi(y)$ is a polynomial in y for any d, b. The parameter $\rho = (\rho_{-1},\ldots,\rho_k)'$ is a simple linear transformation of π and $\phi = (\phi_1,\ldots,\phi_k)'$ and satisfy $\rho_{-1} = -\pi, \rho_k = \phi_k$ and $\sum_{j=-1}^{k} \rho_j = 1$.

As statistical model we analyze the conditional distribution of $(X_1,\ldots,X_T)$ given the initial values $X_{-n},\ n = 0, 1, \ldots,$ under the assumption that ε_t is i.i.d. $N(0, \sigma^2)$. That is, we derive estimators and test statistics from the conditional likelihood function and their properties from the conditional distribution. For the asymptotic analysis we assume that ε_t is i.i.d. with suitable moment conditions and that X_{-n} is bounded. Thus, the initial values are not modeled but treated as (bounded) constants. The standard approach in the literature conducts inference conditional on initial values, which are set equal to zero. In this paper we develop methods for analysis of the nonstationary (unit root) case with $\pi = 0$ and $d \geq b > 1/2$. We call the test of $\pi = 0$ the fractional unit root test in our model.

Our main result is to find asymptotic properties of (Gaussian) maximum likelihood estimators of the parameters in model (1) under the assumption that $\pi = 0$, and the asymptotic distribution of the likelihood ratio (LR) test that $\pi = 0$. We show that if the initial values are bounded they have no influence on limit results, except that conditioning on initial values implies that some of the limit results are expressed in terms of the fractional Brownian motion (fBM) of type II, whereas fBM of type I plays no role in the analysis.

Thus, (1) and (2) are fractional versions of the augmented Dickey and Fuller (1979, 1981) regression model, and we provide a first analysis of a univariate model with two fractional parameters with a simple criterion for different orders of fractionality. The test that $\pi = 0$ is a test that the process is fractional of order d versus $d - b$, i.e. the fractional unit root test is also a test of the order of fractionality of X_t. Note that when $d > b$ the characteristic function (4) has a unit root also when $\pi \neq 0$. However, we still refer to the test of $\pi = 0$ as the unit root test in (1) since it is a test of a unit root in the polynomial $\xi(y)$. Other hypotheses of interest are linear hypotheses on the regression parameters $\phi = (\phi_1, \ldots, \phi_k)'$ and the fractionality parameters d and b. The analysis of model (1) is a first step towards the analysis of the multivariate model which allows cofractionality, that is, linear combinations of fractional processes that decrease the order of fractionality.

A prominent place in literature on models for fractional processes is held by the ARFIMA model

$$A(L)\Delta^d X_t = B(L)\varepsilon_t, \tag{5}$$

where $A(L)$ and $B(L)$ are the autoregressive and moving average polynomials. The ARFIMA model generalizes the well known ARIMA model by introducing the fractional (non-integer) order of differencing, d. The original Dickey and Fuller (1979, 1981) test is a LR test of $A(1) = 0$ within the autoregressive model with $d = 0$ and $B(L) = 1$. A Wald-type test of the same null was considered by, e.g. Chan and Terrin (1995) and Ling and Li (2001) in the ARFIMA model, where the null hypothesis $A(1) = 0$ implies that the process is fractional of order $d + 1$ versus order d under the alternative.[1]

The model we propose to analyze, (1), is different from the ARFIMA model (5) because it is characterized by two fractional parameters and because of the role of the lag operator L_b. Our model is not an ARFIMA model in L, unless $b = 0, 1, 2, \ldots,$ but an ARFIMA model in the new lag operator L_b. Thus, in (5) there is only one fractional parameter and the fractional order of the time series always differs by exactly one under the unit root null and the alternative, whereas the lag operator L_b implies that the difference in order

[1] Robinson (1994) proposed testing for a unit root using the LM-test in several different models, see also Tanaka (1999) and Nielsen (2004). However, these authors examined the properties of hypothesis tests of the form $d = d_0$ (against composite alternatives) in ARFIMA models, so these are not unit root tests in the sense defined above.

of fractionality of the process generated by model (1) under the unit root null and the alternative is b rather than one, see section 2.1 below. Note that when the roots of $A(z)$ are outside the unit circle, we find that X_t is fractional of order d. Thus a simple criterion exists for fractionality of the solution of the ARFIMA model. A similar simple condition exists in terms of $\xi(y)$ for the process generated by model (1), see Lemma 1.

Another strand of the literature analyzes regression-type statistics with the purpose of testing for a fractional unit root. An early contribution is Sowell (1990) who analyzed the regression $y_t = \phi y_{t-1} + u_t$, where $\Delta^d u_t = \varepsilon_t$ and $\phi_0 = 1$. He derived the asymptotic distribution of $\hat{\phi}_{FS}$, the regression estimator of y_t on y_{t-1}, instead of the maximum likelihood estimator for fixed d, $\hat{\phi}_{ML}$, that is, a regression of $\Delta^d y_t$ on $\Delta^d y_{t-1}$ as considered by Ling and Li (2001). Consequently, the asymptotic distribution of the estimator $\hat{\phi}_{FS}$ is discontinuous in d, in the sense that $T^{2d+1}(\hat{\phi}_{FS} - 1)$ converges in distribution to a fBM functional when $d \leq 0$ and $T(\hat{\phi}_{FS} - 1)$ converges in distribution to another such functional when $d \geq 0$. On the other hand, the distribution of $\hat{\phi}_{ML}$ is the same as that of the standard Dickey and Fuller (1979, 1981) statistic (see also the analysis in Phillips, 1987). The ideas in Sowell (1990) were further developed by Dolado, Gonzalo, and Mayoral (2002) who consider the statistical model $\alpha(L)\Delta y_t = \phi\Delta^{d_1} y_{t-1} + \varepsilon_t$ and test that $\phi = 0$, and Lobato and Velasco (2006) who consider the model $\alpha(L)\Delta y_t = \phi\alpha(L)(\Delta^{d_2-1} - 1)\Delta y_t + u_t$ and test that $\phi = 0$. Here $\alpha(L)$ is a lag polynomial. They indicate the properties of the process under the null and under the alternative[2]. In both cases they apply a t-ratio based on a regression equation and motivated by the model equations, rather than a test based upon an analysis of the likelihood function.

Model (1) proposed here has the advantage relative to that of Dolado, Gonzalo, and Mayoral (2002) and others, that one can give simple criteria for fractional integration of various orders in terms of the parameters of the model, see Johansen (2008, Theorem 8) and Lemma 1. In this way we have a platform for conducting model-based statistical inference on the parameters and on the fractional order of X_t, and the possibility of extending the results to the multivariate case.

The remainder of the paper is organized as follows. In section 2 we discuss the properties of the solution of model (1), including the role of initial values and assumptions for the asymptotic analysis. In section 3 we define the Gaussian likelihood function and profile likelihood function as a function of (d, b) and determine its probability limit. In section 4 we show that the asymptotic distribution of the maximum likelihood estimators $\hat{d}, \hat{b}$, and $\hat{\phi}$ are asymptotically Gaussian, whereas the asymptotic distribution of $\hat{\pi}$ and the LR test for $\pi = 0$ in section 5 involve fBM of type II. We conclude in section 6 and give some mathematical details in four appendices.

2 Assumptions and main properties of the fractional autoregressive model

We now discuss properties of the solution of the fractional autoregressive model (1) and the role of initial values, and formulate the assumptions needed for the asymptotic analysis.

[2]The condition given by Dolado, Gonzalo, and Mayoral (2002) for the roots of $\pi(z) = \alpha(z)(1-z)^{1-d_1} - \phi z = 0$ to be outside the unit circle is $\pi(0) = 1, \pi(1) > 0, \pi(-1) > 0$. This cannot be correct as the example $\pi(z) = 4(z - 1/2)^2 = (1 - 4z)(1 - z) + z$ shows. Indeed, the correct condition leads to the solution of an unpleasant transcendental equation, see the discussion in Johansen (2008), and thus it does not appear possible to give simple conditions for fractionality of various orders in terms of the parameters of their model.

2.1 Properties of the solution of the fractional autoregressive model

The binomial expansion of $(1-z)^{-d}$ defines the coefficients $\pi_n(d) = (-1)^n\binom{-d}{n}$ which are bounded in absolute value by $c(d)n^{d-1}$, $d \in \mathbb{R}$. For $d < 1/2$ and ε_t i.i.d. $(0, \sigma^2)$ we define the stationary process with finite variance $\Delta^{-d}\varepsilon_t = (1-L)^{-d}\varepsilon_t = \sum_{n=0}^{\infty}\pi_n(d)\varepsilon_{t-n}$. For $d \geq 1/2$ the infinite sum does not exist, but we define a nonstationary process by the operator Δ_+^{-d},

$$\Delta_+^{-d}\varepsilon_t = \sum_{n=0}^{t-1}\pi_n(d)\varepsilon_{t-n} = \varepsilon_t + d\varepsilon_{t-1} + \cdots + (-1)^{t-1}\binom{-d}{t-1}\varepsilon_1,\ t = 1, \ldots, T,$$

see for instance Dolado, Gonzalo, and Mayoral (2002) or Marinucci and Robinson (2000) who call this a "type II" process and use the notation $\Delta^{-d}\varepsilon_t 1_{\{t\geq 1\}}$, where $1_{\{A\}}$ denotes the indicator function of the event A.

We call a stationary linear process $Y_t = \sum_{n=0}^{\infty}\xi_n\varepsilon_{t-n}$ with continuous spectral density, $\phi(\lambda)$, fractional of order zero if $\phi(0) \neq 0$. We also call the process $Y_t^+ = \sum_{n=0}^{t-1}\xi_n\varepsilon_{t-n} + \mu_t$, where μ_t is a deterministic trend, fractional of order zero, and finally if $\Delta^d Z_t$ is fractional of order zero then Z_t is fractional of order d.

For the asymptotic analysis we apply the result, e.g. Davydov (1970) and Akonom and Gourieroux (1987), that when $d > 1/2$ and $E|\varepsilon_t|^q < \infty$ for some $q > 1/(d-1/2)$, then

$$T^{-d+1/2}\Delta_+^{-d}\varepsilon_{[Tu]} \Longrightarrow W_{d-1}(u) = \Gamma(d)^{-1}\int_0^u (u-s)^{d-1}dW(s) \text{ on } D[0,1], \tag{6}$$

where W denotes BM generated by ε_t and W_{d-1} is the corresponding fBM of type II. By the continuous mapping theorem,

$$T^{-2d}\sum_{t=1}^{T}(\Delta_+^{-d}\varepsilon_t)^2 \xrightarrow{d} \int_0^1 W_{d-1}^2 du. \tag{7}$$

We also have, see Jakubowski, Mémin, and Pages (1989),

$$T^{-d}\sum_{t=1}^{T}\Delta_+^{-d}\varepsilon_t\varepsilon_{t+1} \xrightarrow{d} \int_0^1 W_{d-1}dW. \tag{8}$$

The properties of the solution of model (1) depend on the characteristic function $\Pi(z)$ and the polynomial $\xi(y)$ in (4). If we represent $\Pi(z)$ as a power series $\sum_{n=0}^{\infty}\nu_n z^n$ with $\sum_{n=0}^{\infty}\nu_n^2 < \infty$, we define $\Pi_+(L)X_t = \sum_{n=0}^{t-1}\nu_n X_{t-n}$ and $\Pi_-(L)X_t = \sum_{n=t}^{\infty}\nu_n X_{t-n}$. An equation like $\Pi(L)X_t = \Pi_+(L)X_t + \Pi_-(L)X_t = \varepsilon_t$ is solved by expanding $\Pi(z)^{-1} = \sum_{n=0}^{\infty}\tau_n z^n$, $|z| < 1$, and using the coefficients to define $\Pi_+(L)^{-1}\varepsilon_t = \sum_{n=0}^{t-1}\tau_n\varepsilon_{t-n}$. Then we apply $\Pi_+(L)^{-1}$ and find

$$X_t = \Pi_+(L)^{-1}\varepsilon_t - \Pi_+(L)^{-1}\Pi_-(L)X_t,\ t = 1, \ldots, T. \tag{9}$$

This shows that without any conditions on the parameters, X_t is the sum of a stochastic term $\Pi_+(L)^{-1}\varepsilon_t$ and a deterministic term generated by initial values. To describe the stochastic properties of the solution we introduce the image of the unit circle under the function $z \mapsto 1-(1-z)^b$, $b > 0$, i.e. the set

$$\mathbb{C}_b = \{1-(1-z)^b : |z| \leq 1\}.$$

Lemma 1 *1. If the roots of ξ are outside $\mathbb{C}_b$, then $\xi(y)^{-1}$ is regular in a neighborhood of $\mathbb{C}_b$ and $\xi(1) \neq 0$, so that $F(z) = \xi(1-(1-z)^b)^{-1} = \sum_{n=0}^{\infty}\tau_n z^n$, $|z| < 1$, defines a stationary process $Y_t = \sum_{n=0}^{\infty}\tau_n\varepsilon_{t-n}$ with $\sum_{n=0}^{\infty}\tau_n^2 < \infty$, which is fractional of order zero. Equation (1)* *is solved by*

$$X_t = \Delta_+^{-(d-b)} Y_t^+ - \Pi_+(L)^{-1}\Pi_-(L) X_t = \Delta_+^{-(d-b)} Y_t^+ + \mu_t, \ t = 1, 2, \ldots, \tag{10}$$

where $Y_t^+ = \sum_{n=0}^{t-1} \tau_n \varepsilon_{t-n}$ *and* μ_t *is a function of initial values. Thus* X_t *is fractional of order* $d-b$.

2. If $\pi = 0$ *then* $\xi(1) = 0$, *and if the remaining roots of* ξ *are outside* $\mathbb{C}_b$, *then*

$$(1-z)^d \Pi(z)^{-1} = \gamma + (1-z)^b H(1-(1-z)^b), \tag{11}$$

where $\gamma = (1 - \sum_{i=1}^k \phi_i)^{-1}$ *and* $H(y)$ *is regular in a neighborhood of* $\mathbb{C}_b$ *with* $H(1) \neq 0$, *so that the coefficients defined by* $F(z) = H(1-(1-z)^b) = \sum_{n=0}^\infty \tau_n z^n$, $|z| < 1$, *define a stationary process* $Y_t = \sum_{n=0}^\infty \tau_n \varepsilon_{t-n}$ *with* $\sum_{n=0}^\infty \tau_n^2 < \infty$, *which is fractional of order zero. Equation (1) is solved by*

$$X_t = \Delta_+^{-d}(\gamma \varepsilon_t + \Delta_+^b Y_t^+) + \mu_t, \ t = 1, 2, \ldots, \tag{12}$$

where μ_t *is given in (10). Thus* X_t *is fractional of order* d. *Note that* $Z_t = \gamma\varepsilon_t + \Delta^b Y_t$ *satisfies* $(1 - \sum_{i=1}^k \phi_i L_b^i) Z_t = \varepsilon_t$.

3. If $b > 1/2$ *it holds in both cases 1 and 2 that* $\sum_{n=0}^\infty |\tau_n| < \infty$, *so that the covariance function* $E(Y_t Y_{t-h})$ *satisfies* $\sum_{h=-\infty}^\infty |E(Y_t Y_{t-h})| < \infty$.

Proof. The results (10), (11), and (12) follow from Johansen (2008, Corollary 6 and Theorem 8) and the equation is solved as in (9) using the expression (11) for $\Pi_+(L)^{-1}\varepsilon_t$.

We prove here that $b > 1/2$ implies $\sum_{n=0}^\infty |\tau_n| < \infty$. We give the proof for case 2 of a unit root. Because $H(y)$ is regular in a neighborhood of $\mathbb{C}_b$, the function $F(z) = \sum_{n=0}^\infty \tau_n z^n$, $|z| < 1$, can be extended by continuity to $|z| = 1$, and the continuous transfer function for Y_t has the form $\phi(\lambda) = H(1-(1-e^{\mathrm{i}\lambda})^b) = F(e^{\mathrm{i}\lambda})$, where $\mathrm{i} = \sqrt{-1}$ is the imaginary unit. The derivative of $\phi(\lambda)$ is, for $y = 1-(1-z)^b$,

$$\frac{\partial \phi}{\partial \lambda} = \frac{\partial H}{\partial y}\frac{\partial y}{\partial z}\frac{\partial z}{\partial \lambda} = \frac{\partial H}{\partial y}(1-(1-e^{\mathrm{i}\lambda})^b) b (1-e^{\mathrm{i}\lambda})^{b-1} \mathrm{i} e^{\mathrm{i}\lambda},$$

which is continuous for $b \geq 1$, and has a pole for $\lambda = 0$ when $b < 1$. It is, however, square integrable for $b > 1/2$. By Parseval's formula it then holds that the Fourier coefficients of $\partial\phi/\partial\lambda$, $\mathrm{i}n\tau_n$, are square summable so that $\sum_{n=0}^\infty n^2\tau_n^2 < \infty$, see Zygmund (2003, p. 37). Hence

$$(\sum_{n=1}^\infty |\tau_n|)^2 = (\sum_{n=1}^\infty |\tau_n n| n^{-1})^2 \leq \sum_{n=1}^\infty |\tau_n n|^2 \sum_{n=1}^\infty n^{-2} < \infty,$$

so that $\sum_{n=0}^\infty |\tau_n| < \infty$. Finally we evaluate $\sum_{h=-\infty}^\infty |E(Y_t Y_{t-h})|$ by

$$\sum_{h=-\infty}^\infty \sum_{i,j=0}^\infty |E(\tau_j \varepsilon_{t-j} \tau_i \varepsilon_{t+h-i})| \leq c \sum_{i=0}^\infty \sum_{h=-\infty}^\infty |\tau_{i-h}||\tau_i| \leq c(\sum_{i=0}^\infty |\tau_i|)^2 < \infty.$$

■

2.2 Assumptions for the asymptotic analysis

We give here the assumptions we use in the asymptotic analysis of our models and a discussion of initial values.

Assumption 1 *The process* X_t, $t = 1, \ldots, T$, *is generated by model (1) for some* $k = 1, 2, \ldots$ *and satisfies:*

ERRORS: *The errors* ε_t *are i.i.d.* $(0, \sigma^2)$ *with* $E|\varepsilon_t|^q < \infty$ *for some* $q > \max(6, 1/\eta_0)$ *and* $0 < \eta_0 < \min(1/2, b_0 - 1/2, d_0 - b_0)$.

TRUE VALUES: *The true values satisfy* $d_0 > b_0 > 1/2$, $\pi_0 = 0$, $\sigma_0^2 > 0$, *so that* $\xi(y)$ *has a unit root and the remaining roots of* $\xi(y)$ *are outside the set* $\mathbb{C}_{b_0}$, *so that* $\gamma_0 \neq 0$, *where* γ_0 *is the true value of* γ. *Finally* $\phi_{0k} \neq 0$ *so the asymptotic distribution of the estimators is nonsingular.*

INITIAL VALUES: *The initial values* $X_{-n}, n = 0, 1, \ldots$, *satisfy* $\max_{n \geq 0} |X_{-n}| \leq c$.

Assumption 2 *The process* X_t, $t = 1, \ldots, T$, *is generated by model (2) for some* $k = 0, 1, 2, \ldots$ *and satisfies the* ERRORS *condition for some* $q > \max(4, 1/(d_0 - 1/2))$, *together with the* TRUE VALUES *and* INITIAL VALUES *conditions.*

Importantly, the ERRORS assumption does not include Gaussianity for the asymptotic analysis, but only an i.i.d. condition with sufficient moments to apply a functional central limit theorem and our tightness arguments. Note that the moment condition in model (2) with only one fractional parameter is the same as that required to obtain the functional central limit theorem (6) at $d = d_0$, e.g. Akonom and Gourieroux (1987) or Marinucci and Robinson (2000), except we need a minimum of four moments to apply our tightness methods. The moment condition in model (1) with two fractional parameters is stronger since we need (6) to hold in a neighborhood of the true value and we also need more moments (six) to apply our tightness methods with two fractional parameters. The TRUE VALUES assumption is the unit root assumption, which ensures that X_t is nonstationary and fractional of order d_0.

The theory in this paper will be developed for observations $X_1, \ldots, X_T$ generated by (1) assuming that all initial values are observed, that is, conditional on X_{-n}, $n = 0, 1, \ldots$, and under the assumption that they are bounded, which seems a reasonable condition in practice. Thus, we follow the standard approach in the literature on inference for nonstationary autoregressive processes, where the initial values are observed but not modeled and inference is conditional on them. However, we do not set initial values equal to zero as is often done in the literature on fractional processes, but instead assume only that they are observed unmodelled bounded constants, which represents a significant generalization and makes the results more applicable, as we prove that initial values do not influence the asymptotic results.

In the INITIAL VALUES assumption the bound on $|X_{-n}|$ is needed so that $\Delta^v X_t = \sum_{n=0}^{\infty} \pi_n(-v) X_{t-n}$, and hence the likelihood function, can be calculated for any $v > 0$. In practice, however, one has to choose an approximation to $\Delta^v X_t$ by setting initial values to zero before some time $-T_0$. The approximation error in doing so, $D_{tT_0}(v) = \sum_{j=t+T_0}^{\infty} \pi_j(-v) X_{t-j}$, is bounded by, see (55),

$$|D_{tT_0}(v)| \leq c \sum_{j=t+T_0}^{\infty} j^{-v-1} \leq c(t + T_0)^{-v}.$$

This is small if either t or T_0 is large. For $T_0 = \lambda T, \lambda > 0$, the error in $T^{-1} \sum_{t=1}^{T} (\Delta^v X_t)^2$ can be evaluated by $(T^{-1} \sum_{t=1}^{T} D_{tT_0}(v)^2)^{1/2} = O(T^{-v})$. For $T_0 = 0$ the bound becomes $c(\log T)^{1/2} T^{-\min(v, 1/2)}$. Thus even if $T_0 = 0$, we get an increasing accuracy in the calculations the larger the sample size T, but when $v > 1/2$ we can improve the order of approximation by including initial values.

To study the consistency of the maximum likelihood estimator of (d,b) we next define a large compact parameter set $N(\eta_0)$, which is almost $\{1/2 < b < d < \infty\}$. We use the notation $\psi = (d,b)'$.

Definition 2 *For $0 < \eta_0 < \min(1/2, b_0 - 1/2, d_0 - b_0)$ and $d_0 < \eta_1 < \infty$, we define the compact parameter sets $\bar{N} = \{\psi : 1/2 + \eta_0 \leq b \leq d - \eta_0 \leq \eta_1/\eta_0\}$ and $N(\eta_0) = \cup_{n=0}^{k+2} N_n$, where*

$$\begin{aligned} N_0 &= \{\psi \in \bar{N} : -1/2 + \eta_0 \leq d - b - d_0\}, \qquad (13)\\ N_n &= \{\psi \in \bar{N} : d + (n-2)b - d_0 \leq -1/2 - \eta_0 < -1/2 + \eta_0 \leq d + (n-1)b - d_0\},\ n = 1, \ldots, k+1,\\ N_{k+2} &= \{\psi \in \bar{N} : d + kb - d_0 \leq -1/2 - \eta_0\}. \end{aligned}$$

Similarly, for $0 < \eta_0 < \min(1/2, d_0 - 1/2)$ and $d_0 < \eta_1 < \infty$, we define $\bar{J} = \{d : 1/2 + \eta_0 \leq d \leq \eta_1/\eta_0\}$ and $J(\eta_0) = \cup_{n=1}^{k+2} J_n$, where

$$\begin{aligned} J_1 &= \{d \in \bar{J} : -1/2 + \eta_0 \leq d - d_0\}, \qquad (14)\\ J_n &= \{d \in \bar{J} : d(n-1) - d_0 \leq -1/2 - \eta_0 < -1/2 + \eta_0 \leq dn - d_0\},\ n = 2, \ldots, k+1,\\ J_{k+2} &= \{d \in \bar{J} : d(k+1) - d_0 \leq -1/2 - \eta_0\}. \end{aligned}$$

The interpretation of N_n (or J_n) is that for $\psi \in N_n$ and $-1 \leq i < n-1$, the n processes $T^{d+ib-d_0+1/2}\Delta^{d+ib}X_{[Tu]}$ are nonstationary, and for $n-1 \leq i \leq k$ the $k-n+2$ processes $\Delta^{d+ib}X_t$ are (asymptotically) stationary. Note that some of the sets may be empty depending on d_0, and also note that $\psi_0 \in N_1$ where $\Delta^{d-b}X_t$ is the only nonstationary process.

For $\eta_0 \to 0$ and η_1 fixed, $N(\eta_0) \to \{1/2 < b < d < \infty\}$ except for the lines $d + ib = d_0 - 1/2, i = -1, \ldots, k$, which are critical values for the behavior of the processes because the main contribution to $\Delta^{d+ib}X_t$ on these lines is $\gamma_0 \Delta_+^{d+ib-d_0}\varepsilon_t = \gamma_0 \Delta_+^{-1/2}\varepsilon_t$. To prove uniform convergence of the likelihood function, we thus need to exclude a small neighborhood of these critical values.

2.3 Identification of the parameters in the statistical model

There is an indeterminacy between b, d, and k, as the following example shows. We use the formulation in terms of ρ, see (4). The two functions

$$\begin{aligned} \Pi^*(z) &= \rho_{-1}^*(1-z)^{d^*-b^*} + \rho_0^*(1-z)^{d^*} + \rho_1^*(1-z)^{d^*+b^*}\\ \Pi(z) &= \rho_{-1}^*(1-z)^{d-b} + 0(1-z)^{d} + \rho_0^*(1-z)^{d+b} + 0(1-z)^{d+2b} + \rho_1^*(1-z)^{d+3b} \end{aligned}$$

correspond to the same model if we take $d^* = d + b$ and $b^* = 2b$, but the first has $k = 1$ and the second $k = 3$. Note that $d - b = d^* - b^*$. We therefore discuss identification based on the following result.

Lemma 3 *Let $m \geq 1$ and $\sum_{i=0}^m \zeta_i (1-z)^{a+ib} = \sum_{i=0}^m \bar{\zeta}_i (1-z)^{\bar{a}+i\bar{b}},\ |z| < 1$, and assume that $\zeta_0 \neq 0$ and $\zeta_m \neq 0$. Then $(a, b, \zeta) = (\bar{a}, \bar{b}, \bar{\zeta})$.*

Proof. Because $\zeta_m \neq 0$, the highest order terms must be equal, which implies that $a + mb = \bar{a} + m\bar{b}$. Similarly $\zeta_0 \neq 0$ implies that $a = \bar{a}$. Therefore also $\zeta = \bar{\zeta}$. ■

Corollary 4 *Let two models be generated by (1) with the same lag length $k \geq 1$, but with parameters (ρ, d, b) and $(\bar{\rho}, \bar{d}, \bar{b})$, respectively. Then, under Assumption 1, the parameters are identified under the null hypothesis $\pi = 0$. The same holds under the alternative where $\pi \neq 0$.*

Proof. The assumption that the lag length is k means that $\rho_k \neq 0$, and the conclusion then follows from Lemma 3 under the null $\pi = -\rho_{-1} = 0$, where $\rho_0 \neq 0$, see Assumption 1. It holds under the alternative, because then $\pi = -\rho_{-1} \neq 0$ and $\rho_k \neq 0$. ∎

3 Conditional likelihood and profile likelihood

We consider model (1) with i.i.d. Gaussian errors in order to determine the score and information. The parameter set is expressed in terms of $\tau = (d, b, \phi', \pi, \sigma^2)'$ or $\tau = (d, b, \rho', \sigma^2)'$, where $\sum_{i=-1}^{k} \rho_i = 1$ or $\rho'\mathbf{1} = 1$ and $\mathbf{1}$ is a vector of ones. The likelihood function, conditional on initial values $\{X_{-n}, n \geq 0\}$, becomes

$$-2T^{-1} \log L_T(\tau) = \log \sigma^2 + \frac{1}{\sigma^2 T} \sum_{t=1}^{T} (\Delta^d X_t - \pi \Delta^{d-b} L_b X_t - \sum_{i=1}^{k} \phi_i \Delta^d L_b^i X_t)^2 \tag{15}$$

$$= \log \sigma^2 + \frac{1}{\sigma^2 T} \sum_{t=1}^{T} (\sum_{i=-1}^{k} \rho_i \Delta^{d+ib} X_t)^2 = \log \sigma^2 + \frac{1}{\sigma^2} \rho' \mathcal{M}_T(\psi) \rho, \tag{16}$$

where we define the product moments

$$\mathcal{M}_{ijT}(\psi) = T^{-1} \sum_{t=1}^{T} (\Delta^{d+ib} X_t)(\Delta^{d+jb} X_t) \text{ for } i, j = -1, \ldots, k. \tag{17}$$

Minimizing over σ^2 and ρ, the profile likelihood and the estimators $\hat{\sigma}^2(\psi)$ and $\hat{\rho}(\psi)$ are

$$-2T^{-1} \log L_{profile,T}(\psi) = 1 + \log \hat{\sigma}^2(\psi), \tag{18}$$

$$\hat{\sigma}^2(\psi) = \min_{\rho'\mathbf{1}=1} \rho' \mathcal{M}_T(\psi) \rho \tag{19}$$

$$\hat{\rho}(\psi) = \mathcal{M}_T(\psi)^{-1} \mathbf{1} / \mathbf{1}' \mathcal{M}_T(\psi)^{-1} \mathbf{1}. \tag{20}$$

The last relation is found by considering the first order condition for the constrained optimization problem $\rho' \mathcal{M}_T(\psi) \rho - \lambda(\rho'\mathbf{1} - 1)$. This can of course also be found, using (π, ϕ), by regressing $\Delta^d X_t$ on $\Delta^{d-b} L_b X_t$ and $\Delta^d L_b^i X_t, i = 1, \ldots, k$. Finally the maximum likelihood estimator $\hat{\psi}$ is found by optimizing the profile likelihood (18) and applying (19) and (20).

For the asymptotic analysis we normalize the processes as $T^{\min(d+ib-d_0+1/2,0)} \Delta^{d+ib} X_t$ and define the normalized product moment $\mathcal{M}_T^*(\psi)$. For model (2) we define the corresponding product moments $\mathcal{M}_T(d) = \mathcal{M}_T(d, d)$ and $\mathcal{M}_T^*(d) = \mathcal{M}_T^*(d, d)$.

The limit results are expressed in terms of fBM, see (6), and in terms of the stationary process $Z_t = \gamma_0 \varepsilon_t + \Delta^{b_0} Y_t$, see Lemma 1. We use these to define the matrices

$$\mathcal{M}_{ij}(\psi) = \gamma_0^2 \int_0^1 W_{d_0-d-ib-1} W_{d_0-d-jb-1} du, \ \psi \in N_n, \text{ for } -1 \leq i, j < n-1, \tag{21}$$

$$\mathcal{M}_{ij}(\psi) = E(\Delta^{d+ib-d_0} Z_t)(\Delta^{d+jb-d_0} Z_t), \ \psi \in N_n, \text{ for } n-1 \leq i, j \leq k, \tag{22}$$

$$\mathcal{M}_{ij}(\psi) = \mathcal{M}_{ji}(\psi) = 0, \ \psi \in N_n, \text{ for } -1 \leq j < n-1 \leq i \leq k. \tag{23}$$

We use the notation $\mathsf{D}^m \mathcal{M}_{ijT}(\psi)$ to denote derivatives with respect to ψ. We can now formulate the main limit result for the normalized product moments.

Theorem 5 *Let Assumption 1 be satisfied for model (1). Then the normalized product moments $\mathcal{M}_T^*(\psi)$ and their derivatives $\mathsf{D}^m \mathcal{M}_T^*(\psi)$ are tight as continuous processes on $N(\eta_0)$,* *and for $m = 0, 1, 2$, it holds jointly that for $\psi \in N_n$*

$$\mathcal{M}^*_{ijT}(\psi) \Longrightarrow \mathcal{M}_{ij}(\psi), -1 \leq i, j < n-1, \tag{24}$$

$$\mathsf{D}^m \mathcal{M}_{ijT}(\psi) \Longrightarrow \mathsf{D}^m \mathcal{M}_{ij}(\psi),\ n-1 \leq \max(i,j) \leq k. \tag{25}$$

For $\psi = \psi_0$

$$T^{-b_0} \sum_{t=1}^{T} \Delta^{d_0 - b_0} L_{b_0} X_t \varepsilon_t \xrightarrow{d} \gamma_0 \int_0^1 W_{b_0-1} dW. \tag{26}$$

Let Assumption 2 be satisfied for model (2). Then (24) and (25) hold for $\mathcal{M}^*_T(d)$ and their derivatives on $J(\eta_0)$ and (26) holds with $b_0 = d_0$.

The proof of Theorem 5 is given in Appendix D. We next apply these results to derive the uniform probability limit of the profile likelihood functions for models (1) and (2).

3.1 Convergence of the profile likelihood

Corollary 6 *Let Assumption 1 be satisfied for model (1) and let N_n be given in Definition 2 and define for $\psi \in N_n$, the $(k+2-n) \times (k+2-n)$ matrix $\mathcal{M}^{(n)}(\psi) = \{\mathcal{M}_{ij}(\psi)\}_{i,j=n-1}^k$. Then the profile likelihood function converges in distribution as a continuous process on $N(\eta_0)$ to a deterministic limit, so the convergence is uniform in probability:*

$$-2T^{-1} \log L_{profile,T}(\psi) \Longrightarrow \begin{cases} 1 + \log \sigma^2(\psi), & \psi \in \cup_{n=0}^{k+1} N_n, \\ \infty, & \psi \in N_{k+2}, \end{cases} \tag{27}$$

where

$$\sigma^2(\psi) = \min_{\zeta' \mathbf{1} = 1} \zeta' \mathcal{M}^{(n)}(\psi) \zeta \geq \sigma_0^2, \psi \in N_n, \tag{28}$$

and equality holds if and only if $\psi = \psi_0$. Note that $\psi_0 \in N_1$ so that $\min_{\psi \in N_n} \sigma^2(\psi) > \sigma_0^2$ for $n \neq 1$.

Proof. The profile likelihood function and the estimators $\hat{\sigma}^2(\psi)$ and $\hat{\rho}(\psi)$ are given in (18), (19), and (20).

For $\psi \in N_n, n \leq k+1$, the processes $\Delta^{d+ib} X_t$ are asymptotically stationary for $n-1 \leq i, j \leq k$ and $\mathcal{M}_{ijT}(\psi) \xrightarrow{P} \mathcal{M}_{ij}(\psi)$ by (25), but for $i, j < n-1$, $\mathcal{M}_{ijT}(\psi)$ diverges, see (24), and in order to minimize $\rho' \mathcal{M}_T(\psi) \rho$ over ρ, we have to choose $\rho_i = 0$, for $i < n-1$. Therefore $\hat{\sigma}^2(\psi) \xrightarrow{P} \sigma^2(\psi) = \min_{\zeta' \mathbf{1} = 1} \zeta' \mathcal{M}^{(n)}(\psi) \zeta,\ \psi \in N_n$.

For $\psi \in N_{k+2}$, all processes $\Delta^{d+ib} X_t$ are nonstationary and $\mathcal{M}_{ijT}(\psi)$ diverges for all i and j. It follows that $\hat{\sigma}^2(\psi)$ and hence the profile likelihood tends to infinity in probability. In all cases tightness implies that we have convergence in distribution and hence uniform convergence in probability on the compact set $N(\eta_0)$.

We next discuss the inequality in (28). For $\zeta = (\zeta_{n-1}, \ldots, \zeta_k)$ satisfying $\zeta' \mathbf{1} = 1$, we define for $\psi \in N_n$ the stationary linear process $S_t(\psi) = \sum_{i=n-1}^k \zeta_i \Delta^{d+ib-d_0}(\gamma_0 \varepsilon_t + \Delta^{b_0} Y_t)$, which has transfer function, see Lemma 1,

$$f(z) = \frac{(1-z)^{d-d_0} \sum_{i=n-1}^k \zeta_i (1-z)^{ib}}{\sum_{i=0}^k \rho_i^0 (1-z)^{ib_0}}. \tag{29}$$

We use the superscript in ρ_i^0 to denote true value to avoid confusion with the coefficient ρ_0, and we also define $\rho_* = (\rho_0, \ldots, \rho_k)'$. From $f(0) = \sum_{i=n-1}^k \zeta_i / \sum_{i=0}^k \rho_i^0 = 1$ the coefficient of ε_t in S_t is one, so that $S_t - \varepsilon_t$ is a function of lagged ε's. It follows that for any ζ such that

$\zeta'\mathbf{1} = 1$, we have that $\zeta'\mathcal{M}^{(n)}(\psi)\zeta = Var(S_t(\psi)) = Var(S_t(\psi)-\varepsilon_t)+Var(\varepsilon_t) \geq Var(\varepsilon_t) = \sigma_0^2$ with equality if and only if $S_t = \varepsilon_t$, which in turn is true if and only if $f(z) = 1$ for all z.

If $\psi = \psi_0$ (and thus $n = 1$), then $f(z) = (\sum_{i=0}^k \zeta_i(1-z)^{ib_0})/(\sum_{i=0}^k \rho_i^0(1-z)^{ib_0})$ which is clearly equal to one for all z if $\zeta = \rho_*^0$. That is, if $\psi = \psi_0$ then $\min_{\zeta'\mathbf{1}=1}\zeta'\mathcal{M}^{(1)}(\psi_0)\zeta = \rho_*^{0\prime}\mathcal{M}^{(1)}(\psi_0)\rho_*^0 = \sigma_0^2$. To prove the reverse implication, suppose the minimum is achieved for some value ζ_ψ for which $\zeta_\psi'\mathcal{M}^{(n)}(\psi)\zeta_\psi = \sigma_0^2$. Then it follows that $f(z) = ((1-z)^{d-d_0}\sum_{i=n-1}^k \zeta_{\psi,i}(1-z)^{ib})/(\sum_{i=0}^k \rho_i^0(1-z)^{ib_0}) = 1$ for all z. Because $\rho_k^0 \neq 0$ and $\rho_0^0 \neq 0$, this implies by Lemma 3 that $n = 1$ and $(\psi, \zeta_\psi) = (\psi_0, \rho_*^0)$. ■

Corollary 7 *Let Assumption 2 be satisfied for model (2) and J_n be given in Definition 2. Then:*

1. The profile likelihood function converges in distribution as a continuous process on $J(\eta_0)$:

$$-2T^{-1}\log L_{profile,T}(d) \Longrightarrow \begin{cases} 1+\log\sigma^2(d), & d \in \cup_{n=1}^{k+1} J_n, \\ \infty, & d \in J_{k+2}, \end{cases} \tag{30}$$

where $\sigma^2(d) = \min_{\zeta'\mathbf{1}=1}\zeta'\mathcal{M}^{(n)}(d)\zeta \geq \sigma_0^2$ for $d \in J_n$, and equality holds if and only if $d = d_0$. Note that $d_0 \in J_1$ so that $\min_{d\in J_n}\sigma^2(d) > \sigma_0^2$ for $n \neq 1$.

2. For $k = 0$, the convergence in (30) holds on $J(\eta_0)$, and the limit is strictly convex on J_1 with a minimum at $d = d_0$.

Proof. The proof of 1 follows as in Corollary 6. To prove 2, we note that when $k = 0$ and $\pi_0 = 0$, the model is $\Delta^d X_t = \varepsilon_t$, and for $d \in J_1$ we find $-2T^{-1}\log L_{profile,T}(d) \Longrightarrow 1+\log\sigma^2(d)$, where

$$\sigma^2(d) = E(\Delta^{d-d_0}\varepsilon_t)^2 = \frac{\sigma_0^2}{2\pi}\int_{-\pi}^{\pi}|1-e^{\mathrm{i}\lambda}|^{2(d-d_0)}d\lambda.$$

The convexity follows from the convexity of the function $d \mapsto |1-e^{\mathrm{i}\lambda}|^{2(d-d_0)}$. ■

3.2 Some moment relations

For $\psi = \psi_0$ we want to find an expression for σ_0^2 and ρ^0. We also prove that the asymptotic covariance for $(\hat\psi, \hat\phi)$ is nonsingular.

We define, for $\psi \in N_1$, the process

$$\varepsilon_t(\psi,\phi) = \Delta^{d-d_0}(1-\sum_{i=1}^k \phi_i L_b^i)\Delta^{d_0}X_t = \sum_{i=-1}^k \rho_i\Delta^{d+ib}X_t,$$

and the stationary process

$$\tilde\varepsilon_t(\psi,\phi) = \Delta^{d-d_0}(1-\sum_{i=1}^k \phi_i L_b^i)(\gamma_0\varepsilon_t + \Delta^{b_0}Y_t),$$

for which

$$\sigma^2(\psi,\phi) = \lim_{t\to\infty} Var(\varepsilon_t(\psi,\phi)) = Var(\tilde\varepsilon_t(\psi,\phi)).$$

For a symmetric matrix A we write $A > 0$ to mean that it is positive definite.

Lemma 8 *Under Assumption 1 it holds for model (1) that $\rho^0_{-1} = -\pi_0 = 0$ and for $\rho_* = (\rho_0, \dots, \rho_k)'$ and $\mathcal{M}^{(1)}(\psi_0) = \{\mathcal{M}_{ij}(\psi_0)\}_{i,j=0}^k$ we have*

$$\rho^0_* = (\mathcal{M}^{(1)}(\psi_0))^{-1}\mathbf{1}/(\mathbf{1}'(\mathcal{M}^{(1)}(\psi_0))^{-1}\mathbf{1}), \tag{31}$$

$$\sigma_0^2 = \rho^{0\prime}_*\mathcal{M}^{(1)}(\psi_0)\rho^0_* = 1/(\mathbf{1}'(\mathcal{M}^{(1)}(\psi_0))^{-1}\mathbf{1}), \tag{32}$$

$$\mathsf{D}^2\sigma^2(\psi_0,\phi_0) = 2\Sigma_0 = 2\begin{pmatrix} \Sigma_{0\psi\psi} & \Sigma_{0\psi\phi} \\ \Sigma_{0\phi\psi} & \Sigma_{0\phi\phi} \end{pmatrix} > 0 \text{ for } \phi_{0k} \neq 0. \tag{33}$$

Under Assumption 2 when $k \geq 0$, we find for model (2) that (31) and (32) hold with $b_0 = d_0$, and that

$$\mathsf{D}^2\sigma^2(d_0,\phi_0) = 2\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & I_k \end{pmatrix}\Sigma_0\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & I_k \end{pmatrix}' > 0 \text{ for } \phi_{0k} \neq 0. \tag{34}$$

Proof. Let $v = (v_0, \dots, v_k)'$ be such that $v'\mathbf{1} = 0$. The coefficient of ε_t in the process $U_t = \sum_{i=0}^k v_i\Delta^{d_0+ib_0}X_t$ is $\sum_{i=0}^k v_i = 0$ and it follows that ε_t is independent of U_t. Hence the probability limit of the product moment of U_t with $\sum_{i=0}^k \rho_i^0\Delta^{d_0+ib_0}X_t = \varepsilon_t$ gives $\rho^{0\prime}_*\mathcal{M}^{(1)}(\psi_0)v = 0$. Therefore ρ^0_* is proportional to $\mathcal{M}^{(1)}(\psi_0)^{-1}\mathbf{1}$, which shows (31) since also $\mathbf{1}'\rho^0_* = 1$. To prove (32) we take the average of $(\sum_{i=0}^k \rho_i^0\Delta^{d_0+ib_0}X_t)^2 = \varepsilon_t^2$ and find the limit $\rho^{0\prime}_*\mathcal{M}^{(1)}(\psi_0)\rho^0_* = \sigma_0^2$, which proves (32) using (31).

We let E_{t-1} denote the conditional expectation given the past, $\mathcal{F}_{t-1} = \sigma\{X_{-n},\ n \geq 0,\ \varepsilon_s,\ 1 \leq s \leq t-1\}$, and find that because $\varepsilon_t(\psi_0,\phi_0) = \varepsilon_t$ and $\mathsf{D}^m\varepsilon_t(\psi,\phi)$ only contains lagged ε_t for $m \geq 1$,

$$E_{t-1}(\varepsilon_t\mathsf{D}\varepsilon_t(\psi_0,\phi_0)) = 0,\ E_{t-1}(\varepsilon_t\mathsf{D}^2\varepsilon_t(\psi_0,\phi_0)) = 0, \tag{35}$$

showing that $\varepsilon_t\mathsf{D}\varepsilon_t(\psi_0,\phi_0)$ and $\varepsilon_t\mathsf{D}^2\varepsilon_t(\psi_0,\phi_0)$ are martingale difference sequences. To prove (33) we differentiate $\sigma^2(\psi,\phi)$ twice and find, for $(\psi,\phi) = (\psi_0,\phi_0)$ and using (35), that

$$\mathsf{D}^2\sigma^2(\psi_0,\phi_0) = 2\lim_{t\to\infty} E(\mathsf{D}\varepsilon_t(\psi_0,\phi_0)\mathsf{D}\varepsilon_t(\psi_0,\phi_0)') = 2E(\mathsf{D}\tilde{\varepsilon}_t(\psi_0,\phi_0)\mathsf{D}\tilde{\varepsilon}_t(\psi_0,\phi_0)') \tag{36}$$

which we call $2\Sigma_0$.

We next want to show that Σ_0 is positive definite when $\phi_{0k} \neq 0$. If Σ_0 were singular then there would be constants $\alpha, \beta, \lambda_1, \dots, \lambda_k$, such that the stationary linear process

$$(\alpha\frac{\partial}{\partial d} - \beta\frac{\partial}{\partial b} - \sum_{i=1}^k \lambda_i\frac{\partial}{\partial\phi_i})\tilde{\varepsilon}_t(\psi,\phi)|_{\psi=\psi_0,\phi=\phi_0} \tag{37}$$

is identically zero, or equivalently that its transfer function is identically zero. The transfer function for $\tilde{\varepsilon}_t(\psi,\phi)$ is, see also (29),

$$f(z) = \frac{(1-z)^{d-d_0}(1-\sum_{i=1}^k\phi_i(1-(1-z)^b)^i)}{1-\sum_{i=1}^k\phi_{0i}(1-(1-z)^{b_0})^i},$$

and that of the derivative process (37) can be expressed as

$$\frac{\alpha}{b_0}\log(1-y) - \frac{\beta}{b_0}(1-y)\log(1-y)\dot{\xi}_0(y)/\xi_0(y) - \sum_{i=1}^k\lambda_i y^i/\xi_0(y),$$

for $\xi_0(y) = 1 - \sum_{i=1}^k\phi_{0i}y^i$, $\dot{\xi}_0(y) = -\sum_{i=1}^k i\phi_{0i}y^{i-1}$ and $y = 1-(1-z)^{b_0}$. Thus singularity of Σ_0 is equivalent to

$$\log(1-y)(\alpha\xi_0(y)-\beta(1-y)\dot{\xi}_0(y))-b_0\sum_{i=1}^{k}\lambda_i y^i=0 \text{ for all } y.$$

The last term is a polynomial and the first is not, so this implies that $\lambda_i=0$ for all i, and that $\alpha\xi_0(y)-\beta(1-y)\dot{\xi}_0(y)=0$ for all y. Setting $y=1$ we find that $\alpha\xi_0(1)=\alpha(1-\sum_{i=1}^k\phi_{0i})=0$. Because there are no unit roots in $\xi_0(y)$, see Assumption 1, this implies that $\alpha=0$, and hence $\beta(1-y)\dot{\xi}_0(y)=0$ for all y. This implies that $\beta=0$ because $\xi_0(y)$ is of degree $k\geq 1$ and $\phi_{0k}\neq 0$. Thus we have proved that $\alpha,\beta,\lambda_1,\ldots,\lambda_k$ are all zero and the matrix Σ_0 is positive definite.

Finally, for $b=d=d_0$ and $\phi=\phi_0$ we find (34) from $\sigma^2(d,\phi)=\sigma^2(d,d,\phi)$ and the relation

$$\frac{\partial^2}{\partial d^2}\sigma^2(d,\phi)=\frac{\partial^2}{\partial d^2}\sigma^2(b,d,\phi)+\frac{\partial^2}{\partial b^2}\sigma^2(b,d,\phi)+2\frac{\partial^2}{\partial b\partial d}\sigma^2(b,d,\phi).$$

■

4 Asymptotic properties of the likelihood estimator

In this section we use the results of the previous sections to prove consistency and derive the asymptotic distribution of the maximum likelihood estimator.

4.1 Consistency of the likelihood estimator

Theorem 9 *Let Assumption 1 be satisfied for model (1) or Assumption 2 for model (2). Then:*

1. For model (1), $\hat{\psi}\in N(\eta_0)$ exists and is consistent and $(T^{b_0-1/2}\hat{\pi},\hat{\phi},\hat{\sigma}^2)\xrightarrow{P}(0,\phi_0,\sigma_0^2)$.

2. The same result holds for the likelihood estimators $\hat{d},\hat{\phi},\hat{\pi}$, and $\hat{\sigma}^2$ in model (2).

3. If $k=0$ in model (2), the limit of the profile likelihood is convex on the interval J_1, and $P\{\min_{d\in J_1}(-2T^{-1}\mathsf{D}^2\log L_{profile,T}(d))>0\}\to 1$, so the profile likelihood is convex and $\hat{\psi}$ exists uniquely in the set $J(\eta_0)$ with probability converging to one.

Proof. *Proof of 1 and 2:* We give the proof for model (1). Because $N(\eta_0)$ is compact and the profile likelihood function is continuous, $\hat{\psi}$ exists on $N(\eta_0)$, and hence also $(\hat{\rho},\hat{\sigma}^2)$, see (19) and (20). Convergence in distribution on $N(\eta_0)$ to a deterministic limit in (27), and continuity of the sup functional implies the uniform convergence $\sup_{\psi\in N_{k+2}}|-2T^{-1}\log L_{\text{profile},T}(\psi)|\xrightarrow{P}\infty$ and

$$\sup_{\psi\in\cup_{i=0}^{k+1}N_n}|-2T^{-1}\log L_{\text{profile},T}(\psi)-(1+\log\sigma^2(\psi))|\xrightarrow{P}0. \tag{38}$$

Let $m(\psi)$ be the probability limit given in (27) which has a a strict minimum at $\psi=\psi_0$. It follows that for any $\delta>0$,

$$\inf_{N(\eta_0)\cap\{\psi:|\psi-\psi_0|\geq\delta\}}m(\psi)>1+\log\sigma_0^2. \tag{39}$$

Consistency of $\hat{\psi}$ now follows from (38) and (39) by Theorem 5.7 of van der Vaart (1998).

Now take $\psi\in N_1$ which contains ψ_0 and let $\rho_*=(\rho_0,\ldots,\rho_k)'$. We express $\hat{\rho}(\psi)$, see (20), in terms of the normalized product moments $\mathcal{M}_T^*(\psi)$

$$\begin{pmatrix} T^{b_0-1/2}\hat{\rho}_{-1} \\ \hat{\rho}_* \end{pmatrix}=(\mathcal{M}_T^*(\hat{\psi}))^{-1}\begin{pmatrix} T^{-b_0+1/2} \\ \mathbf{1} \end{pmatrix}/(\begin{pmatrix} T^{-b_0+1/2} \\ \mathbf{1} \end{pmatrix}'(\mathcal{M}_T^*(\hat{\psi}))^{-1}\begin{pmatrix} T^{-b_0+1/2} \\ \mathbf{1} \end{pmatrix})$$

which by tightness of $\mathcal{M}_T^*(\psi)$ and Theorem 5 converges in probability towards, see (31),

$$\begin{pmatrix} 0 \\ (\mathcal{M}^{(1)}(\psi_0))^{-1}\mathbf{1}/(\mathbf{1}'(\mathcal{M}^{(1)}(\psi_0))^{-1}\mathbf{1}) \end{pmatrix} = \begin{pmatrix} 0 \\ \rho_*^0 \end{pmatrix},$$

where $\mathcal{M}^{(1)}(\psi_0) = \{\mathcal{M}_{ij}(\psi_0)\}_{i,j=0}^k$. The result for $(\hat{\pi}, \hat{\phi})$ follows from the result for $\hat{\rho}$.

Proof of 3: For $k = 0$ and $b = d$, the limit of the second derivative of the profile likelihood converges to $\mathsf{D}^2(1 + \log \sigma^2(d)) = \mathsf{D}^2 \log \sigma^2(d) > 0$ for $d \in J_1$. Because the convergence is uniform on the compact interval J_1, it holds that $K_T = \{\min_{d\in J_1}(-2T^{-1}\mathsf{D}^2 \log L_{\text{profile},T}(d)) > 0\}$ has probability converging to one and for such outcomes $-2T^{-1} \log L_{\text{profile},T}(d)$ is strictly convex, so the maximum likelihood estimator exists uniquely. ■

4.2 Asymptotic distribution of the likelihood estimator

Because $\hat{\pi}$ is superconsistent we rescale π as $\theta = T^{b_0-1/2}\pi$. We find the asymptotic distribution of the score functions and the limit of the information for $\tau = \breve{\tau} = (d_0, b_0, \phi_0, 0, \hat{\sigma}^2)$. By Lemma A.3 we only need the information at $\breve{\tau}$ since the estimators are consistent (by Theorem 9) and the second derivatives are tight (by Theorem 5). We let D denote the $2+k$ vector of derivatives with respect to ψ and ϕ.

Lemma 10 *Under Assumption 1 the limit distribution of the Gaussian score function for model (1) at $\breve{\tau} = (d_0, b_0, \phi_0, 0, \hat{\sigma}^2)$ is given by*

$$\begin{pmatrix} T^{-1/2}\mathsf{D}\log L_T(\breve{\tau}) \\ T^{-1/2}\frac{\partial}{\partial\theta}\log L_T(\breve{\tau}) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} N_{2+k}\left(0, \sigma_0^{-2}\Sigma_0\right) \\ \gamma_0 \int_0^1 B_{b_0-1} dB \end{pmatrix}, \tag{40}$$

where Σ_0 is given in (33), $B_{b_0-1} = \sigma_0^{-1}W_{b_0-1}$, $B = \sigma_0^{-1}W$, and the two components are independent.

Proof. Let $\varepsilon_t(\psi,\phi) = \Delta^d X_t - \sum_{i=1}^k \phi_i \Delta^d L_b^i X_t$. Because $\varepsilon_t(\psi_0, \phi_0) = \varepsilon_t$, we find the score functions for (ψ, ϕ) and θ to be

$$T^{-1/2}\mathsf{D}\log L_T(\breve{\tau}) = -\hat{\sigma}^{-2}T^{-1/2}\sum_{t=1}^T \varepsilon_t \mathsf{D}\varepsilon_t(\psi_0,\phi_0), \tag{41}$$

$$T^{-1/2}\frac{\partial}{\partial\theta}\log L_T(\breve{\tau}) = \hat{\sigma}^{-2}T^{-b_0}\sum_{t=1}^T (\Delta^{d_0-b_0}L_{b_0}X_t)\varepsilon_t. \tag{42}$$

Because $\varepsilon_t\mathsf{D}\varepsilon_t(\psi_0,\phi_0)$ is a stationary martingale difference, see (35), with finite third moment, we find the first result in (40) from the central limit theorem for martingale difference sequences, see Hall and Heyde (1980, chp. 3). The asymptotic variance is found as the limit $T^{-1}\sum_{t=1}^T \mathsf{D}\varepsilon_t(\psi_0,\phi_0)\mathsf{D}\varepsilon_t(\psi_0,\phi_0)' \xrightarrow{P} \Sigma_0$, see (33) and (36). The second result follows from (26). A similar argument shows, see (6), that we have joint convergence on $D[0,1]$ towards the Gaussian processes (U, W, W_{b_0-1}) :

$$(T^{-1/2}\sum_{t=1}^{[Tu]} \varepsilon_t \mathsf{D}\varepsilon_t(\psi_0,\phi_0), T^{-1/2}\Delta_+^{-1}\varepsilon_{[Tu]}, T^{-b_0+1/2}\Delta_+^{-b_0}\varepsilon_{[Tu]}) \Longrightarrow (U(u), W(u), W_{b_0-1}(u)),$$

where U is independent of (W, W_{b_0-1}). The independence of the two components of (40) follows because the stochastic integral $\int_0^1 W_{b_0-1}dW$, which is defined as the L_2 limit of the sums $N^{-1}\sum_{i=1}^N W_{b_0-1}(u_i)(W(u_{i+1}) - W(u_i))$, is measurable with respect to the sigma-field generated by the processes (W, W_{b_0-1}) and hence independent of U. ■

Lemma 11 *Under Assumption 1 the Gaussian information per observation for model (1) at* $\breve{\tau} = (d_0, b_0, \phi_0, 0, \hat{\sigma}^2)$ *converges in distribution to*

$$\begin{pmatrix} \sigma_0^{-2}\Sigma_0 & 0 \\ 0 & \gamma_0^2 \int_0^1 B_{b_0-1}^2 du \end{pmatrix}. \tag{43}$$

Proof. Let D denote derivatives with respect to (d, b, ϕ), then the probability limit of

$$-T^{-1}\mathsf{D}^2 \log L_T(\breve{\tau}) = \hat{\sigma}^{-2}T^{-1}\sum_{t=1}^T \varepsilon_t \mathsf{D}^2\varepsilon_t(\psi_0, \phi_0) + \hat{\sigma}^{-2}T^{-1}\sum_{t=1}^T \mathsf{D}\varepsilon_t(\psi_0, \phi_0)\mathsf{D}\varepsilon_t(\psi_0, \phi_0)'$$

is $\sigma_0^{-2}\Sigma_0$ by (33), (35), and a law of large numbers. For the remaining derivatives we find

$$\begin{aligned}
-T^{-1}\frac{\partial^2}{\partial\theta\partial\psi}\log L_T(\breve{\tau}) &= \hat{\sigma}^{-2}(T^{1/2-b_0}\frac{\partial}{\partial\psi}\mathcal{M}^*_{00T}(\psi_0) - \frac{\partial}{\partial\psi}\mathcal{M}^*_{-10T}(\psi_0)) \\
&\quad + \hat{\sigma}^{-2}\sum_{i=1}^k \phi_i \sum_{j=0}^i \binom{i}{j}(-1)^j(\frac{\partial}{\partial\psi}\mathcal{M}^*_{-1jT}(\psi_0) - T^{1/2-b_0}\frac{\partial}{\partial\psi}\mathcal{M}^*_{0jT}(\psi_0)) \xrightarrow{P} 0,
\end{aligned}$$

$$-T^{-1}\frac{\partial^2}{\partial\theta\partial\phi_i}\log L_T(\breve{\tau}) = \hat{\sigma}^{-2}\sum_{j=0}^i \binom{i}{j}(-1)^j(\mathcal{M}^*_{-1jT}(\psi_0) - T^{1/2-b_0}\mathcal{M}^*_{0jT}(\psi_0))) \xrightarrow{P} 0,$$

$$\begin{aligned}
-T^{-1}\frac{\partial^2}{\partial\theta\partial\psi}\log L_T(\breve{\tau}) &= \hat{\sigma}^{-2}(T^{1/2-b_0}\frac{\partial}{\partial\psi}\mathcal{M}^*_{00T}(\psi_0) - \frac{\partial}{\partial\psi}\mathcal{M}^*_{-10T}(\psi_0)) \\
&\quad + \hat{\sigma}^{-2}\sum_{i=1}^k \phi_i \sum_{j=0}^i \binom{i}{j}(-1)^j(\frac{\partial}{\partial\psi}\mathcal{M}^*_{-1jT}(\psi_0) - T^{1/2-b_0}\frac{\partial}{\partial\psi}\mathcal{M}^*_{0jT}(\psi_0)) \xrightarrow{P} 0,
\end{aligned}$$

$$\begin{aligned}
-T^{-1}\frac{\partial^2}{\partial\theta^2}\log L_T(\breve{\tau}) &= \hat{\sigma}^{-2}(\mathcal{M}^*_{-1-1T}(\psi_0) - 2T^{1/2-b_0}\mathcal{M}^*_{-10T}(\psi_0) + T^{1-2b_0}\mathcal{M}^*_{00T}(\psi_0)) \\
&\xrightarrow{d} \sigma_0^{-2}\gamma_0^2\int_0^1 W_{b_0-1}^2 du = \gamma_0^2\int_0^1 B_{b_0-1}^2 du.
\end{aligned}$$

Theorem 5 shows $\mathsf{D}^m\mathcal{M}^*_{-1jT}(\psi_0) \xrightarrow{P} 0,$ for $j = 0, \ldots, k$, and $\mathcal{M}^*_{-1-1T}(\psi_0) \xrightarrow{P} \gamma_0^2\int_0^1 W_{b_0-1}^2 du$. The factor $T^{1/2-b_0}$ ensures that the remaining terms converge to zero. ■

We now apply the previous two lemmas in the usual expansion of the score function to obtain the asymptotic distribution of the likelihood estimators.

Theorem 12 *Under Assumption 1 the asymptotic distribution of the Gaussian maximum likelihood estimators* $\hat{\kappa} = (\hat{d}, \hat{b}, \hat{\phi}')'$ *and* $\hat{\pi}$ *for model (1) is given by*

$$\begin{pmatrix} T^{1/2}(\hat{\kappa} - \kappa_0) \\ T^{b_0}\hat{\pi} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} N_{2+k}\left(0, \sigma_0^2\Sigma_0^{-1}\right) \\ \int_0^1 B_{b_0-1}dB/(\gamma_0\int_0^1 B_{b_0-1}^2 du) \end{pmatrix}, \tag{44}$$

where the two blocks are independent.

Under Assumption 2 for model (2) where $d = b$*, we define* $\hat{\kappa} = (\hat{d}, \hat{\phi}')'$ *and find*

$$\begin{pmatrix} T^{1/2}(\hat{\kappa} - \kappa_0) \\ T^{d_0}\hat{\pi} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} N_{1+k}\left(0, \sigma_0^2\left(Q\Sigma_0 Q'\right)^{-1}\right) \\ \int_0^1 B_{d_0-1}dB/(\gamma_0\int_0^1 B_{d_0-1}^2 du) \end{pmatrix}, \tag{45}$$

where $Q = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & I_k \end{pmatrix}$ *and the two blocks are independent.*

Proof. *Proof of (44):* To find the limit distributions of $\hat{d}$, $\hat{b}$, $\hat{\phi}$, and $\hat{\theta}$, we expand the first derivatives of $l_T(\hat{\tau}) = -T^{-1}\log L_T(\hat{\tau})$ around the value $\breve{\tau} = (\kappa_0, 0, \hat{\sigma}^2)$. Using Taylor's formula with remainder term we find (with subscripts denoting partial derivatives)

$$0 = \begin{pmatrix} T^{1/2} l_{T\kappa}(\breve{\tau}) \\ T^{1/2} l_{T\theta}(\breve{\tau}) \end{pmatrix} + \begin{pmatrix} l^*_{T\kappa\kappa} & l^*_{T\kappa\theta} \\ l^*_{T\theta\kappa} & l^*_{T\theta\theta} \end{pmatrix} \begin{pmatrix} T^{1/2}(\hat{\kappa} - \kappa_0) \\ T^{1/2}\hat{\theta} \end{pmatrix}. \tag{46}$$

Here the asterisks indicate that the information matrix is evaluated at intermediate points (for each row) between $\hat{\tau}$ and $\breve{\tau}$, which hence converge to τ_0 in probability by Theorem 9.

The score functions normalized by $T^{1/2}$ and their limits in distribution are given by Lemma 10. Because the second derivatives are tight, see Theorem 5 and Lemma A.2, and the intermediate points converge in probability to τ_0 we apply Lemma A.3 to replace the intermediate points by τ_0 and find the limit of the information per observation in Lemma 11, see (43). Premultiplying by its inverse we find (44). The independence of the two blocks follows from Lemma 10.

Proof of (45): The same proof applies. The asymptotic variance follows from (34). ■

We remark that the asymptotic distribution is normal for the estimators of the fractional and autoregressive parameters. The asymptotic distribution of the estimator of the unit root is non-normal and of the Dickey-Fuller type, where the integrand BMs have been replaced by fBM. Similar distributions have been obtained previously in the literature. For instance, Tanaka (1999) and Nielsen (2004), among others, consider likelihood based inference in the ARFIMA model and obtain asymptotically normal distribution theory for the parameters. However, they do not allow for a unit root in the autoregressive polynomial and cannot consider the asymptotic distribution of an estimator of a unit root. On the other hand, Ling and Li (2001) do allow for a unit root in the autoregressive polynomial in the ARFIMA model, and obtain results similar to ours except their functionals are in fact functionals of BM since, in our notation, their $b = b_0 = 1$.

Note also that the order of the fBM depends on the distance between the fractional order of X_t when $\pi = 0$ (the data generating process) and $\pi \neq 0$. That is, it depends on the parameter b_0, but it does not depend on the fractional order of X_t itself, d_0. Finally, we remark that the estimator of π is super-consistent since the rate of convergence is T^{b_0}, which is more than root-T-consistent.

5 The likelihood ratio test for a fractional unit root

We next consider the likelihood ratio (LR) test of the unit root hypothesis $\pi = 0$, i.e. the Dickey and Fuller (1979, 1981) test in our model. The profile likelihood for (ψ, ϕ) when $\pi = 0$ is

$$-2T^{-1}\log L_{\text{profile},T}(\psi, \phi, \pi = 0) = -2T^{-1}\log \max_{\theta=0,\sigma^2} L_T(\psi, \phi, \theta, \sigma^2) = 1 + \log \tilde{\sigma}^2(\psi, \phi),$$

where the restricted maximum likelihood estimators, $\tilde{\psi}$ and $\tilde{\phi}$ when $\pi = 0$, satisfy $\frac{\partial}{\partial\psi}\tilde{\sigma}^2(\tilde{\psi}, \tilde{\phi}) = 0$, $\frac{\partial}{\partial\phi}\tilde{\sigma}^2(\tilde{\psi}, \tilde{\phi}) = 0$, and the estimator for σ^2 is $\tilde{\sigma}^2 = \tilde{\sigma}^2(\tilde{\psi}, \tilde{\phi})$. Consistency of the restricted estimator $(\tilde{\psi}, \tilde{\phi}, \tilde{\sigma}^2)$ can be proved the same way as consistency of $(\hat{\psi}, \hat{\phi}, \hat{\sigma}^2)$.

Theorem 13 *Under Assumption 1 for model (1) the asymptotic distribution of the Gaussian* *log likelihood ratio statistic for the hypothesis* $\pi = 0$ *is given by*

$$-2\log LR_T(\pi=0) \xrightarrow{d} \frac{(\int_0^1 B_{b_0-1}dB)^2}{\int_0^1 B_{b_0-1}^2 du}. \tag{47}$$

Under Assumption 2 for model (2) the same result holds with b_0 *replaced by* d_0.

Proof. We give the proof for model (1) only, since the same proof can be applied for model (2). Let $l_T(\tau) = -2T^{-1}\log L_T(\tau)$ and denote derivatives by subscripts. Expansion of $l_{T\tau}(\hat{\tau})$ around τ_0 gives

$$0 = l_{T\tau}(\hat{\tau}) = l_{T\tau}(\tau_0) + l^*_{T\tau\tau}(\hat{\tau}-\tau_0),$$

where $l^*_{T\tau\tau}$ is the matrix of second derivatives (the information per observation) with each row evaluated at an intermediate point, see (46). Expansion of the LR test for $\tau=\tau_0$ gives

$$\begin{aligned} -2\log LR_T(\tau=\tau_0)) &= 2\log(L_T(\hat{\tau})/L_T(\tau_0)) = T(\hat{\tau}-\tau_0)' l^{**}_{T\tau\tau}(\hat{\tau}-\tau_0) \\ &= Tl_{T\tau}(\tau_0)'(l^*_{T\tau\tau})^{-1} l^{**}_{T\tau\tau}(l^*_{T\tau\tau})^{-1} l_{T\tau}(\tau_0) = Tl_{T\tau}(\tau_0)'(i^*_{T\tau\tau})^{-1} l_{T\tau}(\tau_0), \end{aligned}$$

say. With the notation $\eta = (d, b, \phi, \sigma^2)$ we then get that $-2\log(L_T(\hat{\eta},\hat{\theta})/L_T(\eta_0, 0))$ is

$$\begin{aligned} &T\begin{pmatrix} l_{T\eta}(\tau_0) \\ l_{T\theta}(\tau_0) \end{pmatrix}' \begin{pmatrix} i^*_{T\eta\eta} & i^*_{T\eta\theta} \\ i^*_{T\theta\eta} & i^*_{T\theta\theta} \end{pmatrix}^{-1} \begin{pmatrix} l_{T\eta}(\tau_0) \\ l_{T\theta}(\tau_0) \end{pmatrix} \\ &= Tl_{T\eta}(\tau_0)'(i^*_{T\eta\eta})^{-1} l_{T\eta}(\tau_0) + T\frac{(l_{T\theta}(\tau_0) - i^*_{T\theta\eta}(i^*_{T\eta\eta})^{-1} l_{T\eta}(\tau_0))^2}{(i^*_{T\theta\theta} - i^*_{T\theta\eta}(i^*_{T\eta\eta})^{-1} i^*_{T\eta\theta})}. \end{aligned}$$

Similarly we find under the null hypothesis $\theta = 0$ that

$$-2\log\frac{L_T(\tilde{\eta},0)}{L_T(\eta_0,0)} = Tl_{T\eta}(\tau_0)'(l^{**}_{T\eta\eta})^{-1} l_{T\eta}(\tau_0) = Tl_{T\eta}(\tau_0)'(i^{**}_{T\eta\eta})^{-1} l_{T\eta}(\tau_0),$$

so that the test for $\pi = 0$, i.e. $-2\log(L_T(\tilde{\tau})/L_T(\hat{\tau}))$, becomes

$$-2\log\frac{L_T(\tilde{\tau})/L_T(\tau_0)}{L_T(\hat{\tau})/L_T(\tau_0)} = T\frac{(l_{T\theta}(\tau_0) - i^*_{T\theta\eta}(i^*_{T\eta\eta})^{-1} l_{T\eta}(\tau_0))^2}{(i^*_{T\theta\theta} - i^*_{T\theta\eta}(i^*_{T\eta\eta})^{-1} i^*_{T\eta\theta})} + Tl_{T\eta}(\tau_0)'((i^*_{T\eta\eta})^{-1} - (i^{**}_{T\eta\eta})^{-1}) l_{T\eta}(\tau_0).$$

Because $l_{T\eta\eta}$ is tight (see Lemma A.2 and Theorem 5), $\hat{\tau}$ and $\tilde{\tau}$ are consistent, and $T^{1/2} l_{T\eta}(\tau_0)$ converges in distribution, we find that $T^{1/2} l_{T\eta}(\tau_0)'((i^*_{T\eta\eta})^{-1} - (i^{**}_{T\eta\eta})^{-1}) T^{1/2} l_{T\eta}(\tau_0) = o_P(1)$. Finally, we see by (43) that $i^*_{T\theta\eta}(i^*_{T\eta\eta})^{-1} = o_P(1)$, so that $-2\log(L_T(\hat{\tau})/L_T(\tilde{\tau}))$ has the same limit as

$$\frac{(T^{1/2} l_{T\theta}(\tau_0))^2}{i_{T\theta\theta}(\tau_0)} \xrightarrow{d} \frac{(\int_0^1 B_{b_0-1}dB)^2}{\int_0^1 B_{b_0-1}^2 du}.$$

■

The asymptotic distribution of the LR test for a fractional unit root is of the Dickey-Fuller type, but with fBMs replacing the usual BMs as integrands. Critical values of the distribution (47) are easily obtained by simulation, see Table 1 for several values of b_0. In practice b_0 is unknown but can be replaced by $\hat{b}$, and critical values can be simulated on a case-by-case basis or obtained by interpolation in Table 1. Similar distributions are obtained by Sowell (1990), Chan and Terrin (1995), and Dolado, Gonzalo, and Mayoral (2002) for other test statistics. In contrast, Ling and Li (2001) obtain the usual Dickey-Fuller distribution since the ARFIMA model has $b = b_0 = 1$.

Insert Table 1 about here

Note that a test that X_t is fractional of order one (or $I(1)$) is a joint test of $\pi = 0$ and $d = 1$. The asymptotic distribution of the LR test of such a joint hypothesis is readily obtained from Theorems 12 and 13 as the sum of (47) and an independent χ^2_1-distributed random variable.

6 Conclusion

We have discussed likelihood based inference in an autoregressive model for a nonstationary fractional process based on the lag operator L_b. The model generalizes the usual autoregressive model to allow for solutions where the process is fractional of order d or $d-b$, where $d \geq b > 1/2$ are parameters to be estimated. The two fractional parameters and new lag operator, L_b, that characterize our model (1) allow the process to have fractional orders that differ by b under the unit root null hypothesis and the alternative. In the ARFIMA model there is a different lag structure and only one fractional parameter and consequently the fractional orders of the process always differ by exactly one under the unit root null and the alternative. Hence, our model allows substantially more generality than the ARFIMA model in this respect. Within this framework we have discussed model-based likelihood inference on the parameters and on the fractional order of the process.

We model the data $X_1, \ldots, X_T$ given initial values $X_{-n},\ n = 0, 1, \ldots,$ under the assumption that errors are i.i.d. Gaussian. The standard approach in analysis of fractional processes is to set initial values equal to zero. We assume that initial values are observed but not modeled, but we do not set initial values equal to zero. Thus our results are more useful for applications of fractional processes.

Our main technical tool is to consider the likelihood and its derivatives as stochastic processes in the parameters under the assumption that errors are i.i.d. with suitable moment conditions. We apply these tools to prove that the likelihood and its derivatives converge in distribution, and use this to discuss existence, consistency, and asymptotic distribution of the maximum likelihood estimator, as well as the distribution of the associated LR test of the unit root hypothesis. Conditioning on initial values results in the use of type II fBM for the asymptotic analysis.

Acknowledgements

We would like to thank the editor Peter Boswijk, two anonymous referees, Ilijan Georgiev, Marc Henry, Ilya Molchanov, Bent Nielsen, Paolo Paroulo, Anders Rahbek, Ib Skovgaard, and seminar participants at numerous universities and conferences. We are grateful to the Danish Social Sciences Research Council (FSE grants 2114-04-0001 to Johansen and 275-05-0220 to Nielsen), the Social Sciences and Humanities Research Council of Canada (SSHRC grant 410-2009-0183), and the Center for Research in Econometric Analysis of Time Series (CREATES, funded by the Danish National Research Foundation) for financial support. Part of this research was done while the second author was visiting the University of Aarhus, their hospitality is gratefully acknowledged.

Appendix A: Some results on convergence in distribution

We apply the results to processes indexed by the parameters defined on a large compact set, but formulate them, as is usually done, e.g. Kallenberg (2001), for processes $X_n(s)$ with

$s = (s_1, \ldots, s_m) \in [0,1]^m$. For a sequence of p-dimensional stochastic processes $X_n(s) \in \mathbb{R}^p$, $s = (s_1, \ldots, s_m) \in [0,1]^m$, we write $X_n \Longrightarrow X$ or $X_n(s) \Longrightarrow X(s)$ to indicate convergence in distribution of the sequence, either as continuous or cadlag processes on $[0,1]^m$, whereas $X_n(s) \xrightarrow{d} X(s)$ means convergence in distribution on $\mathbb{R}^p$ for a fixed s. When $E|X|^q < \infty$ we define the L_q-norm $||X||_q = (E|X|^q)^{1/q}$, where $|X| = (tr\{X'X\})^{1/2}$ denotes the Euclidean norm.

Lemma A.1 *If $X_n(s) \in \mathbb{R}^p$ with $s = (s_1, \ldots, s_m) \in [0,1]^m$ is a sequence of p-dimensional continuous processes on $[0,1]^m$ for which $X_n(0)$ is tight and*

$$||X_n(s) - X_n(t)||_{m+1} \leq c|s-t| \tag{48}$$

for some constant $c > 0$, which does not depend on n, s, or t, then $X_n(s)$ is tight.

Proof. This is a consequence of Kallenberg (2001, Corollary 16.9). ∎

Below we apply Lemma A.1 for the product moments as processes indexed by ψ and hence $m = 2$ for model (1), and for product moments indexed by d and hence $m = 1$ for model (2).

Lemma A.2 *If the continuous process $X_n(s) \in \mathbb{R}^p$ with $s \in [0,1]^m$ is tight on $[0,1]^m$ and $F : [0,1]^k \times \mathbb{R}^p \mapsto \mathbb{R}^q$ is continuously differentiable, then $Z_n(u,s) = F(u, X_n(s))$ is tight on $[0,1]^{k+m}$.*

Proof. From Taylor's formula we find $F(u,v) - F(\tilde{u},\tilde{v}) = (u-\tilde{u})'\frac{\partial F^*}{\partial u} + (v-\tilde{v})'\frac{\partial F^{**}}{\partial v}$, where $\partial F^*/\partial u$ and $\partial F^{**}/\partial v$ denote partial derivatives taken in suitable intermediate points. Because the partial derivatives are continuous and therefore bounded on compact sets, we find the inequality

$$\max_{u,\tilde{u}\in[0,1]^k, v,\tilde{v}\in[-A,A]^p} |F(u,v) - F(\tilde{u},\tilde{v})| \leq (|u-\tilde{u}| + |v-\tilde{v}|)M_A,$$

which we can use to evaluate the modulus of continuity of $Z_n(u,s)$ and thereby show tightness. ∎

Below we use that the likelihood function for $(\psi, \phi, \theta, \sigma^2)$, the profile likelihood function for (ψ, ϕ, θ), and the profile likelihood function for ψ are all tight as processes in the parameters. Lemma A.2 shows that this follows from the tightness of the product moment $\mathcal{M}_T$.

Lemma A.3 *Let the continuous process $X_n(s) \in \mathbb{R}^p$ with $s \in [0,1]^m$ be tight on $[0,1]^m$ and let $S_n \xrightarrow{P} s_0 \in [0,1]^m$. Then $X_n(S_n) - X_n(s_0) \xrightarrow{P} 0$.*

Proof. This result follows from $P(|X_n(S_n) - X_n(s_0)| > \varepsilon) \leq P(|S_n - s_0| \geq \delta) + P(\omega_{X_n}(\delta) \geq \varepsilon)$, where $\omega_{X_n}(\delta) = \max_{|s-\tilde{s}|\leq\delta} |X_n(s) - X_n(\tilde{s})|$ is the modulus of continuity of X_n. ∎

Lemma A.3 is especially useful when deriving the asymptotic distribution of the maximum likelihood estimators via an asymptotic expansion of the score function. The remainder term in the expansion is the second derivative of the likelihood function evaluated at an intermediate point, which we can replace by the true value by application of Lemma A.3 and an initial consistency proof. Thus, we avoid finding a uniform bound on the third derivative of the likelihood function and rely instead on showing tightness using the moment condition in Lemma A.1.

We conclude with a result which indicates how we are going to establish tightness in the application of the result of Kallenberg.

Lemma A.4 *We consider for $1 \leq t \leq T < \infty$ the processes $V_{utT}^{(i)}, i = 1, 2,$ which are continuous in $u \in [0, 1]$ and linear in the i.i.d. variables ε_t with finite sixth moment. We assume that*

$$||V_{utT}^{(i)}||_2 \leq c \text{ and } ||V_{utT}^{(i)} - V_{\tilde{u}tT}^{(i)}||_2 \leq c|u - \tilde{u}|, \tag{49}$$

where the constant c does not depend on $(u, \tilde{u}) \in [0, 1]^2$ or $1 \leq t \leq T < \infty$. Furthermore we consider deterministic functions $D_{utT}^{(i)}$, which are continuous in $u \in [0, 1]$ and satisfy the condition

$$\max_{u \in [0,1]} T^{-1} \sum_{t=1}^{T} (D_{utT}^{(i)})^2 \to 0 \text{ as } T \to \infty. \tag{50}$$

Then the sequence of product moments

$$S_{uvT}^{(i,j)} = T^{-1} \sum_{t=1}^{T} (V_{utT}^{(i)} + D_{utT}^{(i)})(V_{vtT}^{(j)} + D_{vtT}^{(j)}), \ T = 1, 2, \ldots,$$

is tight in $(u, v) \in [0, 1]^2$ for $i, j \in \{1, 2\}$.

Proof. We decompose $S_{uvT}^{(i,j)}$ as

$$S_{uvT}^{(i,j)} = T^{-1} \sum_{t=1}^{T} V_{utT}^{(i)} V_{vtT}^{(j)} + T^{-1} \sum_{t=1}^{T} V_{utT}^{(i)} D_{vtT}^{(j)} + T^{-1} \sum_{t=1}^{T} D_{utT}^{(i)} V_{vtT}^{(j)} + T^{-1} \sum_{t=1}^{T} D_{utT}^{(i)} D_{vtT}^{(j)}. \tag{51}$$

To prove that the first term is tight we apply the decomposition

$$T^{-1} \sum_{t=1}^{T} V_{utT}^{(i)} V_{vtT}^{(j)} - T^{-1} \sum_{t=1}^{T} V_{\tilde{u}tT}^{(i)} V_{\tilde{v}tT}^{(j)} = T^{-1} \sum_{t=1}^{T} (V_{utT}^{(i)} - V_{\tilde{u}tT}^{(i)}) V_{vtT}^{(j)} + V_{\tilde{u}tT}^{(i)} (V_{vtT}^{(j)} - V_{\tilde{v}tT}^{(j)})$$

and the inequality (54) in Lemma B.2 and find

$$\begin{aligned} ||T^{-1} \sum_{t=1}^{T} (V_{utT}^{(i)} V_{vtT}^{(j)} - V_{\tilde{u}tT}^{(i)} V_{\tilde{v}tT}^{(j)})||_3 &\leq c T^{-1} \sum_{t=1}^{T} ||V_{utT}^{(i)} - V_{\tilde{u}tT}^{(i)}||_2 ||V_{vtT}^{(j)}||_2 + ||V_{\tilde{u}tT}^{(i)}||_2 ||V_{vtT}^{(i)} - V_{\tilde{v}tT}^{(j)}||_2 \\ &\leq c(|u - \tilde{u}| + |v - \tilde{v}|) \leq c\sqrt{2}|(u - \tilde{u}, v - \tilde{v})|. \end{aligned}$$

This shows that the tightness criterion (48) from Lemma A.1 holds and that the first term of (51) is tight. The last term of (51) converges to zero uniformly in (u, v) by (50) and is therefore tight. The mixed moment is bounded by

$$\max_{(u,v) \in [0,1]^2} T^{-1} \sum_{t=1}^{T} |V_{utT}^{(i)} D_{vtT}^{(j)}| \leq \max_{v \in [0,1]} (T^{-1} \sum_{t=1}^{T} (D_{vtT}^{(j)})^2)^{1/2} \max_{u \in [0,1]} (T^{-1} \sum_{t=1}^{T} (V_{utT}^{(i)})^2)^{1/2}.$$

The first factor converges to zero by (50) and the second factor is $O_P(1)$ because $T^{-1} \sum_{t=1}^{T} (V_{utT}^{(i)})^2$ is tight. Therefore the product converges to zero in probability. ∎

Thus to establish tightness of product moments it is enough to check condition (49) for the stochastic components and (50) for the deterministic components of the processes.

Appendix B: Some inequalities

Note the following properties of the L_q-norm,

$$||X+Y||_q \leq ||X||_q + ||Y||_q,\ ||XY||_q \leq ||X||_{2q}||Y||_{2q},\ ||X||_q \leq ||X||_r, \text{ for } 1 \leq q \leq r. \quad (52)$$

The first inequality states that $||\cdot||_q$ is a norm (triangle inequality), the second follows from the Cauchy-Schwarz inequality, and the third from Jensen's inequality.

Lemma B.1 *Let ε_t be i.i.d. with mean zero and $E|\varepsilon_t|^{2q} < \infty$ and define $Z = \sum_{j=0}^{\infty} \xi_j \varepsilon_j$ for some coefficients ξ_j for which $\sum_{j=0}^{\infty} \xi_j^2 < \infty$. Then*

$$||Z||_{2q} \leq c_q ||Z||_2, \quad (53)$$

where the constant c_q does not depend on the coefficients ξ_j.

Proof. Because $||\lambda Z||_{2q} = |\lambda|||Z||_{2q}$, we can scale the ξ's so that $\sum_{j=0}^{\infty} \xi_j^2 = 1$. We first calculate the n'th cumulant $\kappa_n(Z), n \leq 2q$. This is additive for independent variables and homogenous of degree n, so that $\kappa_n(Z) = \sum_{j=0}^{\infty} \kappa_n(\xi_j \varepsilon_j) = \sum_{j=0}^{\infty} \xi_j^n \kappa_n(\varepsilon_j) = \kappa_n(\varepsilon) \sum_{j=0}^{\infty} \xi_j^n$. We next show the inequality $\sum_{j=0}^{\infty} |\xi_j|^n \leq 1,\ n = 1, \ldots, 2q$. For an even number $n = 2m$, this follows from $\sum_{j=0}^{\infty} \xi_j^{2m} \leq (\sum_{j=0}^{\infty} \xi_j^2)^m = 1$. For $n = 2m+1$ we apply the Cauchy-Schwarz inequality

$$(\sum_{j=0}^{\infty} |\xi_j|^{2m+1})^2 = (\sum_{j=0}^{\infty} |\xi_j| \xi_j^{2m})^2 \leq (\sum_{j=0}^{\infty} \xi_j^2)(\sum_{j=0}^{\infty} \xi_j^{4m}) \leq 1.$$

Thus the cumulants $\kappa_n(Z)$, $n = 1, \ldots, 2q$, are bounded independently of ξ_j when $\sum_{j=0}^{\infty} \xi_j^2 = 1$. Finally we find from Kendall and Stuart (1977, p. 70) that $E(Z^{2q})$ is a continuous function of the cumulants $\kappa_n(Z)$, $n = 1, \ldots, 2q$, and hence bounded when $\sum_{j=0}^{\infty} \xi_j^2 = 1$. ■

Lemma B.2 *Let U_t, V_t, X_t, Y_t, $t = 1, \ldots, T$, be processes of the form $\sum_{n=0}^{\infty} \xi_{tn} \varepsilon_n$, with finite sixth moments and $\sum_{n=0}^{\infty} \xi_{tn}^2 < \infty$, then*

$$||\sum_{t=1}^{T} X_t U_t - \sum_{t=1}^{T} Y_t V_t||_3 \leq c \sum_{t=1}^{T} (||X_t||_2 ||U_t - V_t||_2 + ||V_t||_2 ||X_t - Y_t||_2), \quad (54)$$

where the constant does not depend on the coefficients ξ_{tn}.

Proof. The inequality follows by the properties (52) with $q = 3$, and by using $X_t U_t - Y_t V_t = X_t(U_t - V_t) + V_t(X_t - Y_t)$ such that

$$\begin{aligned} ||\sum_{t=1}^{T} X_t U_t - \sum_{t=1}^{T} Y_t V_t||_3 &\leq \sum_{t=1}^{T} ||X_t(U_t - V_t)||_3 + ||V_t(X_t - Y_t)||_3 \\ &\leq \sum_{t=1}^{T} (||X_t||_6 ||U_t - V_t||_6 + ||V_t||_6 ||X_t - Y_t||_6), \end{aligned}$$

and then applying Lemma B.1. ■

Lemma B.3 *For $|u| \leq u_0$, $m \geq 0$, and $j \geq 1$ it holds that*

$$|\mathsf{D}^m \pi_j(u)| \leq c(u_0)(1 + \log j)^m j^{u-1}, \quad (55)$$

$$|\mathsf{D}^m T^{-u} \pi_j(u)| \leq c(u_0) T^{-u} (1 + |\log \frac{j}{T}|)^m j^{u-1}. \quad (56)$$

For $-u_0 \leq v \leq u \leq u_0$, $m \geq 0$, *and* $j \geq 1$ *we have*

$$|\mathsf{D}^m \pi_j(u) - \mathsf{D}^m \pi_j(v)| \leq c(u_0)(u-v)(1+\log j)^{m+1} j^{u-1}, \tag{57}$$

$$|\mathsf{D}^m T^{-u}\pi_j(u) - \mathsf{D}^m T^{-v}\pi_j(v)| \leq c(u_0)(u-v)(1+|\log \frac{j}{T}|)^{m+1} T^{-v} j^{v-1}. \tag{58}$$

Proof. *Proof of (55):* For $m=0$ we apply Stirling's formula

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

where $\max_{\delta_0 \leq u \leq u_0} |\epsilon(u,j)| \to 0$ as $j \to \infty$. This proves the result for $\delta_0 \leq u \leq u_0$, see also Lemma D.1 of Robinson and Hualde (2003), but we want to extend the result to negative u and $m > 0$.

For general u, possibly negative, we choose an integer $n_0 > u_0$ and take $j > n_0$, and define $g(u) = u(u+1)\cdots(u+n_0-1)$. Then

$$\pi_j(u) = \frac{g(u)}{(j-n_0+1)\cdots j} \pi_{j-n_0}(u+n_0). \tag{59}$$

Now we apply (55) to $\pi_{j-n_0}(u+n_0)$ because $u+n_0 \in [n_0-u_0, 2n_0]$, and therefore find, using $(j-n_0+1)\cdots j \geq (j-n_0)^{n_0}$, that

$$|\pi_j(u)| \leq (\max_{|u|\leq u_0} |g(u)|) \frac{1}{(j-n_0)^{n_0}} c(u_0)(j-n_0)^{u+n_0-1} \leq c(u_0)(j-n_0)^{u-1}$$

because $|g(u)|$ is uniformly bounded in $|u| \leq u_0$. Finally,

$$j^{-u+1}(j-n_0)^{u-1} = (1-n_0/j)^{u-1} \leq (1-n_0/j)^{-u_0-1} \to 1 \text{ as } j \to \infty,$$

such that $\max_{j>n_0} j^{-u+1}(j-n_0)^{u-1} \leq c(u_0)$, which proves (55) for $m=0$ and $j > n_0$. For $j \leq n_0$ we have finitely many terms and evaluate each of them as $\max_{|u|\leq u_0} |\pi_j(u) j^{-u+1}| \leq \max_{|u|\leq u_0} \frac{|u(u+1)\dots(u+j-1)|}{j!} j^{-u+1} \leq \pi_j(u_0) j^{u_0+1} \leq c(u_0)$. This completes the proof of (55) for $m=0$.

For $m > 0$ we apply the inequalities

$$\mathsf{D}\log \pi_j(u) = \sum_{i=0}^{j-1} \frac{1}{u+i} \leq \frac{1}{u} + \log j + \log(\frac{1}{u} + \frac{1}{j}) \leq c(u_0,\delta_0)(1+\log j),\ u > 0, \tag{60}$$

$$|\mathsf{D}^m \log \pi_j(u)| = |(-1)^{m+1} \sum_{i=0}^{j-1} \frac{1}{(u+i)^m}| \leq c(u_0,\delta_0),\ u > 0. \tag{61}$$

We find from (59) that

$$\mathsf{D}\pi_j(u) = \frac{\pi_{j-n_0}(u+n_0)}{(j-n_0+1)\cdots j}(ug(u)\mathsf{D}\log \pi_{j-n_0}(u+n_0) + \mathsf{D}ug(u)).$$

Here $ug(u)$ and $\mathsf{D}ug(u)$ are uniformly bounded for $|u| \leq u_0$, and $\mathsf{D}\log \pi_{j-n_0}(u+n_0)$ can be evaluated as in (60) and $\pi_{j-n_0}(u+n_0)$ by (55) because $u+n_0 \in [n_0-u_0, 2n_0]$. The result extends to all j as above. This proves the full result for $m=1$, and for $m>1$ we apply (61).

Proof of (56): We first prove the result for $0 < \delta_0 \leq u \leq u_0$. For $m=0$ the result follows from (55). For $m=1$ we find

$$\mathsf{D}T^{-u}\pi_j(u) = T^{-u}\pi_j(u)(\sum_{i=0}^{j-1} \frac{1}{u+i} - \log T).$$

We then apply the inequality, see (60),

$$\log \frac{j}{T} - c(u_0,\delta_0) \leq \sum_{i=0}^{j-1} \frac{1}{u+i} - \log T \leq \log \frac{j}{T} + c(u_0,\delta_0),$$

and find that

$$|\mathsf{D}T^{-u}\pi_j(u)| \leq c(u_0,\delta_0)T^{-u}j^{u-1}(1+|\log \frac{j}{T}|).$$

For $m>1$ we apply (61). This proves (56) for $0<\delta_0\leq u\leq u_0$. We extend the results to $|u|\leq u_0$ by applying (59) and noting that (56) holds for $\pi_{j-n_0}(u+n_0)$.

Proof of (57) and (58): This follows by the intermediate value theorem using (55) and (56). ■

Lemma B.4 *Define* $\kappa(t,\alpha,\beta)=\sum_{j=1}^{t-1} j^{\alpha-1}(t-j)^{\beta-1}$ *and* $\xi(T,\alpha,\beta)=\max_{i,k}\sum_{j=\max(i,k)}^{T}|\pi_{j-i}(\alpha)\pi_{j-k}(\beta)|$. *Then, uniformly for* $\max(|\alpha|,|\beta|)\leq a_0$, *it holds that*

$$\kappa(t,\alpha,\beta) \leq c(a_0)(1+\log t)t^{\max(\alpha+\beta-1,\alpha-1,\beta-1)}, \tag{62}$$

$$\xi(T,\alpha,\beta) \leq c(a_0)(1+\log T)T^{\max(\alpha+\beta-1,\alpha-1,\beta-1,0)}. \tag{63}$$

Proof. By symmetry we assume without loss of generality that $\beta\leq\alpha$. Note that Robinson and Hualde (2003, Lemma D.2) prove an analogue of (63) for $\alpha=\beta>1/2$, and that (62) can be proved for $\sum_{j=0}^{t}\pi_j(\alpha)\pi_{t-j}(\beta)$.

Proof of (62): We consider three cases. First if $\beta=1$, then $\max(\alpha+\beta-1,\alpha-1,\beta-1)=\max(\alpha,0)$ and $\kappa(t,\alpha,1)=\sum_{j=1}^{t-1}j^{\alpha-1}$. We compare $\sum_{j=1}^{t-1}j^{\alpha-1}$ with the integral $\int_1^t x^{\alpha-1}dx=\alpha^{-1}(t^\alpha-1)$, which is bounded by $\log t$ for $\alpha\leq 0$ and $t^\alpha\log t$ for $\alpha\geq 0$. We find that

$$\sum_{j=1}^{t-1} j^{\alpha-1} \leq \left\{ \begin{array}{cc} \alpha^{-1}(t^\alpha-1), & \alpha\geq 1 \\ \alpha^{-1}(t^\alpha-1)+1, & \alpha\leq 1 \end{array} \right. \leq t^{\max(\alpha,0)}(1+\log t), \tag{64}$$

which is the bound for $\beta=1$. Then we consider $\beta>1$ and find $\kappa(t,\alpha,\beta)\leq t^{\beta-1}\sum_{j=1}^{t-1}j^{\alpha-1}\leq t^{\max(\alpha+\beta-1,\beta-1)}(1+\log t)\leq t^{\max(\alpha+\beta-1,\beta-1,\alpha-1)}(1+\log t)$. Finally, if $\beta<1$ then also $\alpha<1$. In this case $j^{\alpha-1}$ is decreasing in j and $(t-j)^{\beta-1}$ is increasing in j. Therefore we have the inequality

$$\kappa(t,\alpha,\beta)=\sum_{j=1}^{t-1}j^{\alpha-1}(t-j)^{\beta-1} \leq (\frac{t}{2})^{\beta-1}\sum_{0\leq j\leq[t/2]}j^{\alpha-1}+(\frac{t}{2})^{\alpha-1}\sum_{[t/2]+1<j\leq t-1}(t-j)^{\beta-1}$$

so that, using the bound (64),

$$\kappa(t,\alpha,\beta)\leq c(t^{\beta-1}t^{\max(\alpha,0)}(1+\log t)+t^{\alpha-1}t^{\max(\beta,0)}(1+\log t)),$$

which is bounded by $t^{\max(\alpha+\beta-1,\beta-1,\alpha-1)}(1+\log t)$.

Proof of (63): We first consider the sum

$$\sum_{j=\max(i,k)+1}^{T}|\pi_{j-i}(\alpha)\pi_{j-k}(\beta)| \leq c(a_0)\sum_{j=\max(i,k)+1}^{T}(j-i)^{\alpha-1}(j-k)^{\beta-1}=c(a_0)R_T,$$

say. In this case we first take $\beta\leq\alpha\leq 1$, where we use $(j-i)\geq(j-\max(i,k))$ and $(j-k)\geq(j-\max(i,k))$. Then $R_T\leq\sum_{j=\max(i,k)+1}^{T}(j-\max(i,k))^{\alpha+\beta-2}\leq c(1+\log T)T^{\max(\alpha+\beta-1,0)}$. Next we let $\beta\leq\alpha$ and $\alpha>1$, where $(j-i)^{\alpha-1}\leq T^{\alpha-1}$. Then $R_T\leq T^{\alpha-1}\sum_{j=\max(i,k)+1}^{T}(j-$

$k)^{\beta-1} \leq c(1+\log T)T^{\alpha-1+\max(\beta,0)}$. The term with $j = \max(i,k)$ can be analyzed the same way. ■

Appendix C: Variation bounds

In this appendix we prove a series of lemmas containing variation bounds of the type $||V_{utT}||_2 \leq c$ and $||V_{utT} - V_{vtT}||_2 \leq c(u-v)$, which we use to verify condition (49) in Lemma A.4 for relevant processes and product moments. The first lemma covers the initial values, the second and third lemmas deal with the nonstationary processes and (asymptotically) stationary processes, respectively, and the fourth lemma concerns product moments including both stationary and nonstationary processes.

We find from (12) that for $v > 0$, $\Delta^v X_t = \Delta_+^v X_t + \Delta_-^v X_t$ has the representation

$$\begin{aligned} \Delta^v X_t &= \gamma_0 \Delta_+^{-d_0+v} \varepsilon_t + \Delta_+^{-d_0+v+b_0} Y_t^+ + \Delta_+^v \mu_t + \Delta_-^v X_t, \ t = 1, 2, \ldots, \\ \Delta^v X_t &= \Delta_-^v X_t, \ t = 0, -1, -2, \ldots. \end{aligned} \tag{65}$$

The first term of (65) is asymptotically stationary if $d_0 - v < 1/2$. For $d_0 - v > 1/2$ it will, suitably normalized and with sufficient moments, converge to fBM, see (6). The next term is asymptotically stationary if $d_0 - v - b_0 < 1/2$, and the last terms are deterministic trends that are functions of initial values. In order to study the impact of the initial values we use the representations,

$$\begin{aligned} \Pi_+(L)^{-1} &= \gamma_0 \Delta_+^{-d_0} + \Delta_+^{-d_0+b_0} F_+(L), \\ \Pi_-(L) &= (\Delta^{d_0} - \sum_{i=1}^{k} \phi_{0i} \Delta^{d_0}(1-\Delta^{b_0})^i)_- = \sum_{j=0}^{k} \rho_j^0 \Delta_-^{d_0+jb_0}, \end{aligned}$$

for some coefficients ρ_j^0, see Lemma 1 and (4), so that

$$\Delta_+^{d+ib} \mu_t = -\Delta_+^{d+ib} \Pi_+(L)^{-1} \Pi_-(L) X_t = -(\gamma_0 + \Delta_+^{b_0} F_+(L)) \sum_{j=0}^{k} \rho_j^0 \Delta_+^{d+ib-d_0} \Delta_-^{d_0+jb_0} X_t. \tag{66}$$

Lemma C.1 *(Initial values) If $|X_{-n}| \leq c$ and $|u| \leq u_0, 0 < \delta_0 \leq v \leq v_0$ then uniformly in u, v and for $c_0 = c(u_0, v_0, \delta_0)$ the initial values satisfy the relations*

$$\begin{aligned} |\frac{\partial^m}{\partial u^m} \Delta_+^u \Delta_-^v X_t| &= |\sum_{n=0}^{\infty} (\sum_{j=0}^{t-1} \frac{\partial^m}{\partial u^m} \pi_j(-u) \pi_{n+t-j}(-v)) X_{-n}| \\ &\leq c_0 (1+\log t)^{m+1} t^{-\min(u+v,u+1,v)}, \end{aligned} \tag{67}$$

$$|\frac{\partial^m}{\partial u^m} T^u \Delta_+^u \Delta_-^v X_t| \leq c_0 (1+\log T)^{m+1} T^{-\min(v,1,v-u,-u)}. \tag{68}$$

From this it follows that for $G(z) = \sum_{n=0}^{\infty} g_n z^n, \sum_{n=0}^{\infty} |g_n| < \infty$, we have

$$|G_+(L) \frac{\partial^m}{\partial u^m} \Delta_+^u \Delta_-^v X_t| \leq c_0 (1+\log t)^{m+1} \sum_{n=0}^{t-1} |g_{t-n}| n^{-\min(u+v,u+1,v)}, \tag{69}$$

$$|G_+(L) \frac{\partial^m}{\partial u^m} T^u \Delta_+^u \Delta_-^v X_t| \leq c_0 (1+\log T)^{m+1} T^{-\min(v,1,v-u,-u)}. \tag{70}$$

Similarly it holds that, uniformtly in $\delta_0 \leq v \leq v_0$,

$$|\frac{\partial^m}{\partial v^m}\Delta_-^v X_t| \to 0 \text{ as } t \to \infty, \tag{71}$$

$$\max_{1\leq t\leq T}|\frac{\partial^m}{\partial v^m}T^v\Delta_-^v X_t| \leq c_0(1+\log T)^m T^{v_0}. \tag{72}$$

Proof. *Proof of (67):* We find

$$\Delta_+^u\Delta_-^v X_t = \sum_{j=0}^{t-1}\pi_j(-u)\sum_{i=t-j}^{\infty}\pi_i(-v)X_{t-j-i} = \sum_{n=0}^{\infty}(\sum_{j=0}^{t-1}\pi_j(-u)\pi_{n+t-j}(-v))X_{-n}. \tag{73}$$

From the bound $|X_{-n}| \leq c$ we find by differentiating (73) and using $\mathsf{D}^m\pi_j(\delta) \leq c(1+\log j)^m j^{\delta-1}$, see (55), that the main terms of the derivatives of the right hand side of (67), corresponding to $j > 0$, are bounded by

$$c\sum_{n=0}^{\infty}\sum_{j=1}^{t-1}(1+\log j)^m j^{-u-1}(n+t-j)^{-v-1} \leq c(1+\log t)^m\sum_{n=0}^{\infty}\sum_{j=1}^{t-1}j^{-u-1}(n+t-j)^{-v-1}.$$

We then find $\sum_{n=0}^{\infty}(n+t-j)^{-v-1} = \sum_{n=t-j}^{\infty}n^{-v-1} \leq c(t-j)^{-v}$ because $-v \leq -\delta_0 < 0$, so that the bound becomes $c(1+\log t)^m\sum_{j=1}^{t-1}j^{-u-1}(t-j)^{-v}$. The result follows if we apply (62) of Lemma B.4 with $\alpha = -u$ and $\beta = 1-v$. The term with $j=0$ is treated similarly.

Proof of (68): We find, using (67), that

$$\begin{aligned}|\frac{\partial^m}{\partial u^m}T^u\Delta_+^u\Delta_-^v X_t| &= |T^u\sum_{k=0}^{m}\binom{m}{k}(\log T)^{m-k}\frac{\partial^k}{\partial u^k}\Delta_+^u\Delta_-^v X_t| \\ &\leq c_0T^u(1+\log T)^{m+1}t^{-\min(u+v,u+1,v)} \leq c_0(1+\log T)^{m+1}T^{-\min(v,1,v-u,-u)},\end{aligned}$$

where we used the evaluation $T^u t^{-\min(u+v,u+1,v)} \leq T^{-\min(v,1,v-u,-u)}$.

Proof of (69) and (70): Follows from (67) and (68).

Proof of (71) and (72): Uniformly for $\delta_0 \leq v \leq v_0$ we have from (55) that

$$|\frac{\partial^m}{\partial v^m}\Delta_-^v X_t| \leq |\sum_{n=t}^{\infty}\frac{\partial^m}{\partial v^m}\pi_n(-v)X_{t-n}| \leq c_0\sum_{n=t}^{\infty}(1+\log n)^m n^{-v-1} \leq c_0\sum_{n=t}^{\infty}(1+\log n)^m n^{-\delta_0-1} \to 0.$$

Similarly we find that

$$|\frac{\partial^m}{\partial v^m}T^v\Delta_-^v X_t| \leq |T^v\sum_{k=0}^{m}\binom{m}{k}(\log T)^k\frac{\partial^{m-k}}{\partial v^{m-k}}\Delta_-^v X_t| \leq c_0T^v(1+\log T)^m\sum_{n=t}^{\infty}(1+\log n)^m n^{-v-1},$$

which can be evaluated uniformly in $1 \leq t \leq T$ and $0 < \delta_0 \leq v \leq v_0$ by

$$c_0T^{v_0}(1+\log T)^m\sum_{n=0}^{\infty}(1+\log n)^m n^{-\delta_0-1} \leq c_0(1+\log T)^m T^{v_0}.$$

■

Lemma C.2 *The deterministic terms $D_{it} = \Delta_+^{d+ib}\mu_t + \Delta_-^{d+ib}X_t$, see (65), satisfy*

$$\max_{\psi\in N_n}|\mathsf{D}^m D_{it}| \to 0 \text{ as } t \to \infty \text{ for } n-1 \leq i \leq k, \tag{74}$$

$$\max_{\psi\in N_n}\max_{1\leq t\leq T}|\mathsf{D}^m T^{d+ib-d_0+1/2}D_{it}| \to 0 \text{ as } T \to \infty \text{ for } -1 \leq i < n-1. \tag{75}$$

Proof. *Proof of (74):* We first consider $\Delta_+^{d+ib}\mu_t$, see (66), and define $v = d_0 + jb_0 \in [d_0, v_0]$. For $\psi \in N_n$ and $n-1 \leq i \leq k$ we have $u = d + ib - d_0 \in [-1/2 + \eta_0, u_0]$. We therefore investigate $|G_+(L)\frac{\partial^m}{\partial u^m}\Delta_+^u\Delta_-^v X_t|$ for such (u, v). The result follows from (69) because the exponent is bounded as $\min(u + v, u + 1, v) \geq \min(d_0 - 1/2 + \eta_0, 1/2 + \eta_0, d_0) > 0$. Hence $|G_+(L)\mathsf{D}^m\Delta_+^{d+ib-d_0}\Delta_-^{d_0+jb_0}X_t| \to 0$ as $t \to \infty$ uniformly for $\psi \in N_n$ since also $\sum_{n=0}^{\infty}|g_n| < \infty$. For the term $\Delta_-^{d+ib}X_t$ we apply (71) with $v = d + ib \in [1/2, v_0]$.

Proof of (75): We first consider $T^{d+ib-d_0+1/2}\Delta_+^{d+ib}\mu_t$. Using (66) it is enough to investigate $|G_+(L)\mathsf{D}^m T^{d+ib-d_0+1/2}\Delta_+^{d+ib-d_0}\Delta_-^{d_0+jb_0}X_t|$. We define $v = d_0 + jb_0 \in [d_0, v_0]$ and $u = d + ib - d_0 \in [-d_0, -1/2 - \eta_0]$ for $\psi \in N_n$ and $-1 \leq i < n-1$. We apply (70) to evaluate $T^{1/2}|G_+(L)\mathsf{D}^m T^u\Delta_+^u\Delta_-^v X_t|$ uniformly in u, v, and t, and the exponent $\min(v - 1/2, 1/2, v - u - 1/2, -u - 1/2)$ is bounded by $\min(d_0 - 1/2, 1/2, d_0 + \eta_0, \eta_0) \geq \eta_0$. For the term $T^{d+ib-d_0+1/2}\Delta_-^{d+ib}X_t$ we define $v = d + ib$ which for $\psi \in N_n$ and $-1 \leq i < n-1$ satisfies $v = d + ib \in [\eta_0, d_0 - 1/2 - \eta_0]$. Then we consider $T^{1/2-d_0}|\mathsf{D}^m T^v\Delta_-^v X_t|$, and find from (72) the bound

$$c_0 T^{1/2-d_0}(1 + \log T)^m T^{d_0-1/2-\eta_0} \leq c_0(1 + \log T)^m T^{-\eta_0} \to 0.$$

■

Lemma C.3 *Let $Z_t = \sum_{n=0}^{\infty}\xi_n\varepsilon_{t-n}$ be a stationary linear process with finite variance and $\sum_{n=0}^{\infty}|\xi_n| < \infty$, and define $\phi_Z(h) = \sigma^2\sum_{n=0}^{\infty}|\xi_n||\xi_{n+h}|$ and $Z_t^+ = \sum_{n=0}^{t-1}\xi_n\varepsilon_{t-n}$. For $m = 0, 1, 2$ it holds that*

$$||\mathsf{D}^m T^{u+1/2}\Delta_+^u Z_t^+||_2 \leq c(v_0, u_0), \tag{76}$$

$$||\mathsf{D}^m T^{u+1/2}\Delta_+^u Z_t^+ - \mathsf{D}^m T^{v+1/2}\Delta_+^v Z_t^+||_2 \leq c(v_0, u_0)|u - v|, \tag{77}$$

uniformly in $v_0 \leq v \leq u \leq u_0 < -1/2$.

Proof. We first note the evaluation $|Cov(Z_t^+, Z_{t+h}^+)| = \sigma^2|\sum_{n=0}^{t-1-|h|}\xi_n\xi_{n+|h|}| \leq \phi_Z(h)$, so that

$$V_T = Var(\mathsf{D}^m T^{u+1/2}\Delta_+^u Z_t^+) = Var(\sum_{i=0}^{t-1}\mathsf{D}^m T^{u+1/2}\pi_i(-u)Z_{t-i}^+)$$
$$\leq \sum_{i=0}^{t-1}\sum_{j=0}^{t-1}|\mathsf{D}^m T^{u+1/2}\pi_i(-u)||\mathsf{D}^m T^{u+1/2}\pi_j(-u)|\phi_Z(i-j).$$

We apply the inequality (56) and find that V_T is bounded by

$$c(u_0, v_0)T^{2u+1}\sum_{i=1}^{t-1}\sum_{j=1}^{t-1}|i^{-u-1}(1 + |\log(\frac{i}{T})|)^m||j^{-u-1}(1 + |\log(\frac{j}{T})|)^m|\phi_Z(i-j) \tag{78}$$
$$\leq c(u_0, v_0)T^{2u+1}\sum_{h=0}^{t-1}\phi_Z(h)\sum_{j=1}^{t-1-h}(j+h)^{-u-1}j^{-u-1}(1 + |\log(\frac{j+h}{T})|)^m(1 + |\log(\frac{j}{T})|)^m.$$

Now we evaluate $|\log(\frac{j+h}{T})| \leq |\log(\frac{j}{T})|$ and for $-1 \leq u$ we find $(j+h)^{-u-1}j^{-u-1} \leq j^{-2u-2}$, so that when $-1 \leq u \leq u_0 < -1/2$, V_T is bounded by

$$c(u_0, v_0)\sum_{h=0}^{\infty}\phi_Z(h)T^{-1}\sum_{j=1}^{T}(\frac{j}{T})^{-2u_0-2}(1+|\log(\frac{j}{T})|)^{2m} \to c(u_0, v_0)\int_0^1 x^{-2u_0-2}(1+|\log x|)^{2m}dx < \infty.$$

For $u \leq -1$ we evaluate $(j+h)^{-u-1}j^{-u-1} \leq T^{-2u-2}$ and find

$$V_T \leq c(u_0,v_0)\sum_{h=0}^{\infty}\phi_Z(h)T^{-1}\sum_{j=1}^{T}(1+|\log(\frac{j}{T})|)^{2m} \to c(u_0,v_0)\int_0^1(1+|\log x|)^{2m}dx < \infty.$$

To prove (77), we apply the inequality (58) and then use the same proof. ∎

Lemma C.4 *Let $Z_t = \sum_{n=0}^{\infty}\xi_n\varepsilon_{t-n}$ be a stationary linear process with finite variance and $\sum_{n=0}^{\infty}|\xi_n| < \infty$, and define $\phi_Z(h) = \sigma^2\sum_{n=0}^{\infty}|\xi_n||\xi_{n+h}|$ and $Z_t^+ = \sum_{n=0}^{t-1}\xi_n\varepsilon_{t-n}$. For $m = 0,1,2$ it holds that*

$$||\mathsf{D}^m\Delta_+^u Z_t^+||_2 \leq c(u_0,v_0), \tag{79}$$

$$||\mathsf{D}^m\Delta_+^u Z_t^+ - \mathsf{D}^m\Delta_+^v Z_t^+||_2 \leq c(u_0,v_0)|u-v|, \tag{80}$$

uniformly in $-1/2 < v_0 \leq v \leq u \leq u_0$.

Proof. Using (55) we find, as in the proof of Lemma C.3, see (78), the inequality

$$V_T = Var(\mathsf{D}^m\Delta_+^u Z_t^+) \leq c(u_0,v_0)\sum_{h=0}^{t-1}\phi_Z(h)\sum_{j=1}^{t-1-h}(j+h)^{-u-1}j^{-u-1}(1+\log(j+h))^m(1+\log j)^m.$$

Let $\delta_0 < \min(1/2+v_0, 1/2)$ and note that $(j+h)^{-\delta_0}(1+\log(j+h))^m$ and $j^{-\delta_0}(1+\log j)^m$ are bounded in j and h and $(j+h)^{-u+\delta_0-1} \leq j^{-u+\delta_0-1}$ because $-u+\delta_0-1 \leq \delta_0-1/2 < 0$. Then

$$V_T \leq c(u_0,v_0)\sum_{h=0}^{t-1}\phi_Z(h)\sum_{j=1}^{t-1-h}(j+h)^{-u+\delta_0-1}j^{-u+\delta_0-1} \leq c(u_0,v_0)\sum_{h=0}^{\infty}\phi_Z(h)\sum_{j=1}^{\infty}j^{2(-v_0+\delta_0-1)}$$

which is bounded because $2(-v_0+\delta_0-1) < -1$, which gives (79) because $\sum_{h=0}^{\infty}\phi_Z(h) < \infty$. For $Var((\Delta_+^u - \Delta_+^v)Z_t^+)$ we apply the inequality (57), and then use the same proof. ∎

Lemma C.5 *For $i = 1,2$, let $Z_{it}^+ = \sum_{n=0}^{t-1}\xi_{in}\varepsilon_{t-n}$ satisfy $\sum_{n=0}^{\infty}|\xi_{in}| \leq c < \infty$, and define $\phi_{ij}(h) = \sigma^2\sum_{n=0}^{\infty}|\xi_{i,n}||\xi_{j,n+h}|$. Then for $u < -1/2 < v$ we have*

$$T^{u-1/2}\sum_{t=1}^{T}\Delta_+^u Z_{1t}^+\Delta_+^v Z_{2t}^+ = T^{u-1/2}\sum_{t=1}^{T}\sum_{i=0}^{t-1}\sum_{j=0}^{t-1}\pi_{t-i}(-u)\pi_{t-j}(-v)Z_{1i}^+Z_{2j}^+ \overset{P}{\to} 0.$$

Proof. For $\mu_{i,j,k,l} = E(Z_{1i}^+Z_{1k}^+Z_{2j}^+Z_{2l}^+)$, the second moment of the variable is

$$\begin{aligned}
V_T &= T^{2u-1}\sum_{t=1}^{T}\sum_{i=1}^{t}\sum_{j=1}^{t}\pi_{t-i}(-u)\pi_{t-j}(-v)\sum_{s=1}^{T}\sum_{k=1}^{s}\sum_{l=1}^{s}\pi_{s-k}(-u)\pi_{s-l}(-v)\mu_{i,j,k,l} \\
&\leq T^{2u-1}\sum_{i,j=1}^{T}\sum_{t=\max(i,j)}^{T}|\pi_{t-i}(-u)\pi_{t-j}(-v)|\sum_{k,l=1}^{T}\sum_{s=\max(k,l)}^{T}|\pi_{s-k}(-u)\pi_{s-l}(-v)||\mu_{i,j,k,l}| \\
&\leq T^{2u-1}\xi(T,-u,-v)^2\sum_{i,j,k,l=1}^{T}|\mu_{i,j,k,l}|,
\end{aligned}$$

see Lemma B.4 for the definition of $\xi(T,-u,-v)$. We want to prove that $V_T \to 0$ as $T \to \infty$.

From Anderson (1971, p. 467), for the case of stationary processes, we have that

$$|\mu_{i,j,k,l}| \leq \sigma_0^4(\phi_{11}(i-k)\phi_{22}(j-l) + \phi_{12}(i-j)\phi_{12}(k-l) + \phi_{12}(i-l)\phi_{12}(j-k)) + \kappa_4(\varepsilon) \sum_{n=0}^{\min(i,j,k,l)} \xi_{1,i-n}\xi_{1,k-n}\xi_{2,j-n}\xi_{2,l-n}.$$

Summing over $1 \leq (i,j,k,l) \leq T$ and using the bounds

$$\sum_{i,k=1}^{T} \phi_{ij}(i-k) \leq cT \sum_{h=0}^{\infty} \phi_{ij}(h) \leq cT, \quad \sum_{i,j,k,l=1}^{T} \sum_{n=0}^{\min(i,j,k,l)} \xi_{1,i-n}\xi_{1,k-n}\xi_{2,j-n}\xi_{2,l-n} \leq cT,$$

we find $V_T \leq T^{2u-1}\xi(T,-u,-v)^2(c_1T^2 + c_2T)$. Thus we find from (63) in Lemma B.4 that

$$V_T \leq cT^{2u+1}\xi(T,-u,-v)^2 \leq c(\log T)^2 T^{\max(-1-2v,-1,2u-1-2v,2u+1)}$$

The exponent is negative because $u < -1/2 < v$ which shows that $V_T \to 0$. ■

Appendix D: Proof of Theorem 5

The derivatives of the likelihood function are functions of $\mathsf{D}^m\mathcal{M}_T(\psi)$, see (16) and (17). These product moments are functions of $\Delta^{d+ib}X_t,\ i=-1,0,\ldots,k$, and their derivatives, and we discuss the properties of these processes below. We next prove tightness by applying Lemma A.4 for the normalized product moments, $\mathcal{M}_T^*(\psi)$, and using conditions (49) and (50). Then we derive the limits of each of the product moments and the relevant derivatives, and this completes the proof of Theorem 5.

We give the proof only for the general model (1). The same proof can be applied for model (2).

D.1 Representation and tightness of the processes and product moments

From (65) we find for $i=-1,0,\ldots,k$, and $t=1,2,\ldots$

$$\Delta^{d+ib}X_t = \Delta_+^{d+ib-d_0}(\gamma_0\varepsilon_t + \Delta_+^{b_0}Y_t^+) + \Delta_+^{d+ib}\mu_t + \Delta_-^{d+ib}X_t = S_{it}^+ + D_{it}, \tag{81}$$

$$S_{it}^+ = \Delta_+^{d+ib-d_0}(\gamma_0\varepsilon_t + \Delta_+^{b_0}Y_t^+), \tag{82}$$

$$D_{it} = \Delta_+^{d+ib}\mu_t + \Delta_-^{d+ib}X_t. \tag{83}$$

For $\psi \in N_n$, $S_{it} = \Delta^{d+ib-d_0}(\gamma_0\varepsilon_t + \Delta^{b_0}Y_t)$ is stationary if $n-1 \leq i \leq k$, and nonstationary for $-1 \leq i < n-1$.

Lemma D.1 *Let Assumption 1 be satisfied for model (1) and let $m=0,1,2$. Then $\mathsf{D}^m\Delta^{d+ib}X_t$ satisfies condition (49) if $u_0 \geq d+ib-d_0 \geq -1/2+\eta_0$, and $\mathsf{D}^mT^{d+ib-d_0+1/2}\Delta^{d+ib}X_t$ satisfies condition (49) if $-d_0 \leq d+ib-d_0 \leq -1/2-\eta_0$. In the same cases D^mD_{it} and $\mathsf{D}^mT^{d+ib-d_0+1/2}D_{it}$ satisfy (50). It follows that $\mathsf{D}^m\mathcal{M}_T^*(\psi)$ is tight.*

Proof. We write $\Delta^{d+ib}X_t = S_{it}^+ + D_{it}$, see (81). The results for the deterministic terms follow from (74) and (75) in Lemma C.2. Next let $Z_t^+ = \gamma_0\varepsilon_t + \Delta_+^{b_0}Y_t^+ = \sum_{n=0}^{t-1}\tau_n\varepsilon_{t-n}$ which satisfies $\sum_{n=0}^{\infty}|\tau_n| < \infty$, see Lemma 1. If $d+ib-d_0 \geq -1/2+\eta_0$ then $u = d+ib-d_0 \in [-1/2+\eta_0, u_0]$ and $\mathsf{D}^mS_{it}^+ = \mathsf{D}^m\Delta^{d+ib-d_0}Z_t^+$ satisfies (49) by Lemma C.4. If $d+ib-d_0 \leq -1/2-\eta_0$ then $u = d+ib-d_0 \in [-d_0, -1/2-\eta_0]$ and $\mathsf{D}^mT^{d+ib-d_0+1/2}S_{it}^+$ satisfies (49) by Lemma C.3. ■

D.2 Convergence of product moments

We want to find the limits of $\mathcal{M}_T^*(\psi)$ and its derivatives and thereby prove the results (24), (25), and (26) of Theorem 5.

Lemma D.2 *Under Assumption 1 where $q > 1/\eta_0$ we find for $T \to \infty$ that*

$$T^{d+ib-d_0+1/2}\Delta^{d+ib}X_{[Tu]} \Longrightarrow \gamma_0 W_{d_0-d-ib-1}(u),\ \psi \in N_n,\ -1 \le i < n. \tag{84}$$

Proof. We write $\Delta^{d+ib}X_t = \Delta_+^{d+ib-d_0}\gamma_0\varepsilon_t + \Delta_+^{d+ib-d_0}\Delta_+^{b_0}Y_t^+ + D_{it}$. It follows from (6) and $q > 1/\eta_0 \ge 1/(d_0 - d - ib - 1/2)$ when $\psi \in N_n$ and $i < n-1$, that for $t = [Tu]$ the first term normalized by $T^{d+ib-d_0+1/2}$ converges in distribution on $D[0,1]$ to $\gamma_0 W_{d_0-d-ib-1}(u)$. Next let $Z_t^+ = \Delta_+^{d+ib-d_0+b_0}Y_t^+$ for which $\max_{1\le t\le T} E(|Z_t^+|^q) \le a$, and define $u = d + ib - d_0 + 1/2 \le -\eta_0$ when $\psi \in N_n$ and $-1 \le i < n-1$. Then

$$P(\max_{1\le t\le T} T^u|Z_t^+| \ge c) \le \sum_{t=1}^T P(|Z_t^+| \ge cT^{-u}) \le \sum_{t=1}^T \frac{E(|Z_t^+|^q)}{c^q T^{-qu}} \le ac^{-q}T^{1+qu}.$$

This converges to zero because $1+qu \le 1-q\eta_0 < 0$, which shows that $T^{d+ib-d_0+1/2}\Delta_+^{d+ib-d_0-b_0}Y_t^+$ converges to zero uniformly in $t \le T$. Finally $\max_{\psi\in N_n}\max_{1\le t\le T} T^{d+ib-d_0+1/2}|D_{it}| \to 0$ for $-1 \le i < n-1$ by (75). ∎

Proof of (24): The continuous mapping theorem and (84) imply that for a fixed ψ

$$\mathcal{M}_{ijT}^*(\psi) \xrightarrow{d} \gamma_0^2 \int_0^1 W_{d_0-d-ib-1}W_{d_0-d-jb-1}du = \mathcal{M}_{ij}(\psi),\ \psi \in N_n,\ -1 \le i,j < n-1.$$

Since this holds jointly for finitely many ψ and $\mathcal{M}_{ijT}^*(\psi)$ is tight, we have proved convergence in distribution.

Proof of (25): We first take $m = 0$. Because $S_{it} = \Delta^{d+ib-d_0}(\gamma_0\varepsilon_t + \Delta^{b_0}Y_t) = \sum_{j=0}^\infty \tau_{ij}\varepsilon_{t-j} = S_{it}^+ + S_{it}^-$ is stationary for $\psi \in N_n$ and $n-1 \le i \le k$, we find $E(T^{-1}\sum_{t=1}^T (S_{it}^-)^2) = \sigma_0^2 T^{-1}\sum_{t=1}^T(\sum_{n=t}^\infty \tau_{in}^2) \to 0$, so that the probability limit of $T^{-1}\sum_{t=1}^T S_{it}^+S_{jt}^+$ equals that of $T^{-1}\sum_{t=1}^T S_{it}S_{jt}$ which exists by the law of large numbers. The deterministic terms have no influence in the limit because of (74). Thus, see (22),

$$\mathcal{M}_{ijT}(\psi) \xrightarrow{P} \mathcal{M}_{ij}(\psi),\ \psi \in N_n,\ n-1 \le i,j \le k. \tag{85}$$

The convergence in probability (and distribution) holds jointly for finitely many values of $\psi \in N_n$, so we have proved (25) for $n-1 \le i,j \le k$ because $\mathcal{M}_{ijT}(\psi)$ is tight. Next we take $-1 \le i < n-1 \le j \le k$ and write

$$\mathcal{M}_{ijT}^*(\psi) = T^{-1}\sum_{t=1}^T T^{d+ib-d_0+1/2}(S_{it}^+ + D_{it})(S_{jt}^+ + D_{jt}).$$

From (74), (75), and $\max_{1\le t\le T}||S_{jt}^+||_2 \le c$, $\max_{1\le t\le T}||T^{d+ib-d_0+1/2}S_{it}^+||_2 \le c$, see (76) and (79), it follows that we only have to show that $T^{-1}\sum_{t=1}^T T^{d+ib-d_0+1/2}S_{it}^+S_{jt}^+ \xrightarrow{P} 0$. This follows by Lemma C.5 for $Z_{1t}^+ = Z_{2t}^+ = \gamma_0\varepsilon_t + \Delta_+^{b_0}Y_t^+$ and $u = d + ib - d_0 \le -1/2 - \eta_0$, and $v = d + jb - d_0 \ge -1/2 + \eta_0$. The convergence in probability holds jointly for finitely many values of $\psi \in N_n$ and tightness holds by Lemma D.1.

The derivatives D^m give rise to an extra factor $(\log T)^m$, which does not change the proof.

Proof of (26): We write

$$\Delta^{d_0-b_0}L_{b_0}X_t = \Delta^{d_0-b_0}X_t - \Delta^{d_0}X_t = \gamma_0(\Delta_+^{-b_0}\varepsilon_t - \varepsilon_t) + Y_t^+ - \Delta_+^{b_0}Y_t^+ + D_{-1t}^0 - D_{0t}^0,$$

see (81), where D^0_{-1t} and D^0_{0t} are calculated for $\psi = (d_0, b_0)$. We decompose the product moment of $\Delta^{d_0-b_0}L_{b_0}X_t$ and ε_t as

$$T^{-b_0}\sum_{t=1}^T \gamma_0(\Delta_+^{-b_0}\varepsilon_t - \varepsilon_t)\varepsilon_t + T^{-b_0}\sum_{t=1}^T (Y_t^+ - \Delta_+^{b_0}Y_t^+)\varepsilon_t + T^{-b_0}\sum_{t=1}^T (D^0_{-1t} - D^0_{0t})\varepsilon_t. \qquad (86)$$

For the last term we find

$$Var(T^{-b_0}\sum_{t=1}^T (D^0_{-1t} - D^0_{0t})\varepsilon_t) \leq T^{-2b_0+1}T^{-1}\sigma_0^2\sum_{t=1}^T (D^0_{-1t} - D^0_{0t})^2 \to 0,$$

by (74) and (75). The second term of (86) is $T^{-b_0}\sum_{t=1}^T (Y_t^+ - \Delta_+^{b_0}Y_t^+)\varepsilon_t$, which is a sum of a martingale difference sequence, so that

$$Var(T^{-b_0}\sum_{t=1}^T (Y_t^+ - \Delta_+^{b_0}Y_t^+)\varepsilon_t) = \sigma_0^2 T^{-2b_0}\sum_{t=1}^T Var(Y_t^+ - \Delta_+^{b_0}Y_t^+) \leq cT^{1-2b_0} \to 0.$$

Finally, the first term of (86) is compared with the product moment of $\Delta_+^{-b_0}\varepsilon_t$ and ε_{t+1} for which we have the convergence in (26), see (8), by showing that the difference converges to zero. We find $\sum_{t=1}^T \varepsilon_t(\Delta_+^{-b_0}\varepsilon_{t-1} - (\Delta_+^{-b_0}\varepsilon_t - \varepsilon_t)) = \sum_{t=1}^T \varepsilon_t(\varepsilon_t - \Delta_+^{-b_0+1}\varepsilon_t)$, with mean zero and variance

$$Var(\sum_{t=1}^T \varepsilon_t(\varepsilon_t - \Delta_+^{-b_0+1}\varepsilon_t)) = \sigma_0^4\sum_{t=1}^T\sum_{j=1}^{t-1}\pi_j^2(b_0-1) \leq c\sum_{t=1}^T\sum_{j=1}^{t-1} j^{2(b_0-2)} \leq cT^{2\max(b_0-3/2,0)+1}.$$

Normalized by T^{-2b_0} this converges to zero because $b_0 > 1/2$, and proves (26) of Theorem 5.

References

Akonom, J., Gourieroux, C., 1987. A functional central limit theorem for fractional processes. Technical report 8801, CEPREMAP, Paris.

Anderson, T. W., 1971. The Statistical Analysis of Time Series, Wiley, New York.

Baillie, R. T., 1996. Long memory processes and fractional integration in econometrics. Journal of Econometrics 73, 5–59.

Billingsley, P., 1968. Convergence of Probability Measures, Wiley, New York.

Chan, N. H., Terrin, N., 1995. Inference for unstable long-memory processes with applications to fractional unit root autoregressions. Annals of Statistics 23, 1662–1683.

Davydov, Yu. A., 1970. The invariance principle for stationary processes. Theory of Probability and its Applications 15, 487–498.

Dickey, D. A., Fuller, W. A., 1979. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association 74, 427–431.

Dickey, D. A., Fuller, W. A., 1981. Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica 49, 1057–1072.

Dolado, J. J., Gonzalo, J., Mayoral, L., 2002. A fractional Dickey-Fuller test for unit roots. Econometrica 70, 1963–2006.

Hall, P., Heyde, C. C. 1980. Martingale Limit Theory and its Application, Academic Press, New York.

Henry, M., Zaffaroni, P., 2003. The long range dependence paradigm for macroeconomics and finance. In: Doukhan, P., Oppenheim, G., Taqqu, M. S. (Eds.), Theory and Applications of Long-Range Dependence, Birkhäuser, Boston, 417–438.

Jakubowski, A., Mémin, J., Pages, G., 1989. Convergence en loi des suites d'intégrales stochastiques sur l'espace D^1 de Skorokhod. Probability Theory and Related Fields 81, 111–137.
Johansen, S., 2008. A representation theory for a class of vector autoregressive models for fractional processes. Econometric Theory 24, 651–676.
Kallenberg, O., 2001. Foundations of Modern Probability Theory, 2nd edition, Springer, New York.
Kendall, M., G. Stuart, A., 1977. The Advanced Theory of Statistics, Vol. I, 4th ed., Charles Griffin, London.
Ling, S., Li, W. K., 2001. Asymptotic inference for nonstationary fractionally integrated autoregressive moving-average models. Econometric Theory 17, 738–764.
Lobato, I. N., Velasco, C., 2006. Efficient Wald tests for fractional unit roots. Econometrica 75, 575–589.
Marinucci, D,. Robinson, P. M., 2000. Weak convergence of multivariate fractional processes. Stochastic Processes and their Applications 86, 103–120.
Nielsen, M. Ø., 2004. Efficient likelihood inference in nonstationary univariate models. Econometric Theory 20, 116–146.
Phillips, P. C. B., 1987. Time series regression with a unit root. Econometrica 55, 277–301.
Robinson, P. M., 1994. Efficient tests of nonstationary hypotheses. Journal of the American Statistical Association 89, 1420–1437.
Robinson, P. M., Hualde, J., 2003. Cointegration in fractional systems with unknown integration orders. Econometrica 71, 1727-1766.
Sowell, F. B., 1990. The fractional unit root distribution. Econometrica 58, 495–505.
Tanaka, K., 1999. The nonstationary fractional unit root. Econometric Theory 15, 549–582.
van der Vaart, A. W., 1998. Asymptotic Statistics, Cambridge University Press, Cambridge.
Zygmund, A., 2003. Trigonometric Series, Vol. I and II, 3rd rev. ed., Cambridge University Press, Cambridge.

Table 1: Simulated quantiles of the distribution (47)

	$b_0 = 0.6$	$b_0 = 0.8$	$b_0 = 1.0$	$b_0 = 1.2$	$b_0 = 1.4$
90%	2.52	2.70	2.95	3.18	3.33
95%	3.60	3.83	4.10	4.37	4.53
99%	6.34	6.60	6.92	7.23	7.42

Note: Based on $100,000$ replications and $T = 1000$.