Discussion Papers
Department of Economics
University of Copenhagen

Likelihood Inference for a Nonstationary
Fractional Autoregressive Model

Søren Johansen
Morten Ørregaard Nielsen

Likelihood inference for a nonstationary fractional autoregressive model[*]

Søren Johansen[†]
University of Copenhagen and CREATES
and
Morten Ørregaard Nielsen[‡]
Cornell University and CREATES

August 31, 2007

Abstract

This paper discusses model based inference in an autoregressive model for fractional processes based on the Gaussian likelihood. The model allows for the process to be fractional of order d or $d-b$, where $d \geq b > 1/2$ are parameters to be estimated.

We model the data $X_1, \ldots, X_T$ given the initial values X_{-n}^0, $n = 0, 1, \ldots$, under the assumption that the errors are i.i.d. Gaussian. We consider the likelihood and its derivatives as stochastic processes in the parameters, and prove that they converge in distribution when the errors are i.i.d. with suitable moment conditions and the initial values are bounded. We use this to prove existence and consistency of the local likelihood estimator, and to find the asymptotic distribution of the estimators and the likelihood ratio test of the associated fractional unit root hypothesis, which contains the fractional Brownian motion of type II.

Keywords: Dickey-Fuller test, fractional unit root, likelihood inference.

JEL Classification: C22.

[†]Address: Department of Economics, University of Copenhagen, Studiestræde 6, DK-1455 Copenhagen K, Denmark. Email: sjo@math.ku.dk

[‡]Corresponding author. Address: Department of Economics, Cornell University, 486 Uris Hall, Ithaca, NY 14853. Email: mon2@cornell.edu

1 Introduction and motivation

We consider the univariate time series X_t, $t = \ldots, -1, 0, 1, \ldots, T$, and model $X_1, \ldots, X_T$ conditional on the initial values X^0_{-n}, $n = 0, 1, \ldots$, by the fractional autoregressive model

$$a(L_d)X_t = \varepsilon_t, \tag{1}$$

where ε_t is i.i.d. $(0, \sigma^2)$, $a(z)$ is a $(k+1)$'th order polynomial, and we have introduced the lag operator $L_d = 1 - \Delta^d$ (when $d = 1$ we have the usual lag operator $L_1 = L$). We rewrite the model as

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

where, in particular, $\pi = -a(1)$. The parameters $(\pi, \phi_1, \ldots, \phi_k, d, \sigma^2)$ are unrestricted except $\sigma^2 > 0$. In the simplest case with $k = 0$ the model is

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

which we shall consider separately in some of our results. We analyze the conditional likelihood function for $(X_1, \ldots, X_T)$ given the initial values X^0_{-n}, $n = 0, 1, \ldots$, under the assumption that ε_t is i.i.d. $N(0, \sigma^2)$. For the asymptotic analysis we assume that ε_t is i.i.d. with suitable moment conditions and that X^0_{-n} is bounded.

For given values of the parameters, the process X_t is determined by (2) as a function of parameters, initial values, and errors ε_i, $i = 1, \ldots, t$, but the properties of X_t depend on the properties of the characteristic function associated with (2),

$$\pi(z) = (1-z)^d - \pi(1 - (1-z)^d) - \sum_{i=1}^{k} \phi_i (1-z)^d (1 - (1-z)^d)^i = a(1 - (1-z)^d). \tag{4}$$

This is most easily analyzed by the substitution $y = 1 - (1-z)^d$. Note that $\pi(z)$ is a polynomial in z if and only if d is a non-negative integer, whereas $a(y)$ is a polynomial for any d. Clearly $a(y)$ is simpler to analyze, and conditions in terms of the roots of $a(y)$ are given under which the process determined by (2) is fractional of order 0, or fractional of order d when $\pi = -a(1) = 0$, in which case the characteristic function $\pi(z)$ has a unit root. In this paper we are primarily interested in the nonstationary (unit root) case with $\pi = 0$ and $d > 1/2$. Thus, the hypothesis of a unit root in the fractional autoregressive model (1), i.e. the hypothesis $a(1) = 0$, is most easily formulated in (2) where it is given simply by the restriction $\pi = 0$. We call the test of $\pi = 0$ the (fractional) unit root test in our model.

To allow even more generality we analyze the autoregressive model

$$\Delta^d X_t = \pi \Delta^{d-b} L_b X_t + \sum_{i=1}^{k} \phi_i \Delta^d L_b^i X_t + \varepsilon_t, \ t = 1, \ldots, T, \tag{5}$$

which allows the solution to be fractional of different orders, d or $d - b \geq 0$, depending on whether $\pi = 0$ or $\pi \neq 0$, assuming that the remaining roots are outside the set $\mathbb{C}_b$, see Johansen (2007) and section 2.1 below. The parameters $(\pi, \phi_1, \ldots, \phi_k, b, d, \sigma^2)$ are unrestricted except for $d \geq b$ and $\sigma^2 > 0$.

Thus, 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 of (5), $\pi(z) = (1-z)^{d-b}a(1-(1-z)^b)$ has a unit root also when $\pi \neq 0$. However, we shall still refer to the test of $\pi = 0$ as the unit root test in (5) since it is a test of a unit root in the polynomial $a(y)$. Other hypotheses of interest are linear hypotheses on the regression parameters $\phi = (\phi_1, \ldots, \phi_k)'$ and the fractionality parameters d and b.

1.1   Summary of main results

The main results of this paper are to find asymptotic properties of (local) maximum likelihood estimators of the parameters in model (5) under the assumption that $\pi = 0$, and the asymptotic distribution of the likelihood ratio 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 of type II, whereas fractional Brownian motion of type I plays no role in the analysis.

We show that the profile likelihood function converges in distribution as a continuous stochastic process in the parameters (d, b, ϕ) to a deterministic limit which is strictly convex in a neighborhood of the true value (d_0, b_0, ϕ_0). Using tightness, or stochastic equicontinuity, see Newey (1991) and Andrews (1992)[1], of the derivatives of the profile likelihood function, we show that it too is strictly convex in a small neighborhood with probability tending to one. Hence, the (local) likelihood estimator $(\hat{d}, \hat{b}, \hat{\phi})$ exists, is unique, and is consistent, a result which also holds for the estimators $\hat{\pi}$ and $\hat{\sigma}^2$. For model (3) we prove strict convexity of the likelihood function (with large probability) and hence existence and uniqueness of the (global) maximum likelihood estimator.

We find the asymptotic distribution of the estimators using the usual expansion of the profile score function with a remainder term, which is the second derivative evaluated at an intermediate point between $(\hat{d}, \hat{b}, \hat{\phi})$ and (d_0, b_0, ϕ_0). We use tightness of the second derivative to show that we can replace the intermediate point with the true value. We then find the asymptotic distribution of $\hat{\pi}$, and finally we apply an expansion of the log likelihood to find the limit distribution of the likelihood ratio test for the unit root hypothesis $\pi = 0$.

---

[1]Note that the stochastic equicontinuity conditions given by Newey (1991) and Andrews (1992) involve different conditions than the ones we apply. We prefer to use moment conditions usually presented in a tightness context, e.g. Billingsley (1968) and Kallenberg (2001), and thus we use the tightness terminology rather than stochastic equicontinuity. The moment conditions also have the useful feature that we need only analyze the second derivative of the likelihood function, and not higher-order derivatives.

1.2   Comparison with other models

Our paper fits into two related strands of the literature. The first group of papers considers statistical modeling of fractional processes. The statistical analysis is based upon the Gaussian likelihood function, and LM, Wald, and likelihood ratio tests are derived from this. The second group of papers focuses on the test for a unit root, but test statistics and estimators are motivated by regression equations.

It is important that in any case it is part of the methodology to test the underlying assumptions, e.g. on the error process or the constancy of the parameters, against the data to assess whether the model proposed is appropriate before conducting inference on the parameters of the model. Since we are concerned with model-based statistical inference, our paper clearly belongs to the former group, although parallels can be drawn between our unit root test and the fractional unit root tests in the latter group.

A prominent place in the first group is held by the ARFIMA model proposed by Granger and Joyeux (1980) and Hosking (1981), i.e.,

$$A(L)\Delta^d X_t = B(L)\varepsilon_t, \tag{6}$$

where $A(L)$ and $B(L)$ are the autoregressive and moving average polynomials and ε_t is a white noise process. The ARFIMA model generalizes the well known ARIMA model by introducing the fractional (non-integer) order of differentiation, d. The original Dickey and Fuller (1979, 1981) test can thus also be placed in this group, since it is a likelihood ratio (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 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.

Robinson (1991, 1994) proposed testing for the unit root using the LM-test in a number of 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, and thus these are not unit root tests in the sense defined above.[2]

The model we propose to analyze (5) is different from the ARFIMA model (6) because of the role of the lag operator L_b. The model is not an ARFIMA model in L, but an ARFIMA model in the new lag operator L_b, which implies that the difference in order of fractionality of the process under the unit root null and the alternative is b rather than one, see section 2.1 below.

---

[2]A slightly more general version of the ARFIMA model, see Hualde and Robinson (2005), is

$$\Delta^d X_t = u_t 1_{\{t \geq 1\}}, \quad u_t = \varphi(L, \theta)\varepsilon_t,$$

where the parameter vector θ does not contain the fractional parameter d. That is, the infinite lag polynomial φ is not allowed to depend on d. However, if we write our model in the similar way, the lag polynomial φ would depend on both d and b. This is an important difference, which complicates our analysis.

In the second group of papers we place those that analyze regression-type statistics with the purpose of testing for a fractional unit root, a topic that has received much attention recently.

An early contribution is Sowell (1990) who analyzed the behavior of the usual Dickey-Fuller-type regression when the errors are fractional. Specifically, Sowell (1990) considered the regression

$$y_t = \phi y_{t-1} + u_t, \quad \Delta^d u_t = \varepsilon_t,$$

where ε_t is i.i.d. $N(0, \sigma^2)$ 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 dichotomous, i.e. discontinuous in d, in the sense that $T^{2d+1}(\hat{\phi}_{FS} - 1)$ converges in distribution to a fractional Brownian motion 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). Clearly, the discontinuous distribution theory obtained for $\hat{\phi}_{FS}$ is a consequence of the simplicity of the estimator. Similar dichotomous or discontinuous distribution results for the same type of model and estimator are obtained by Tanaka (1999), see also Chan and Terrin (1995). Furthermore, these papers all assume that the parameter d is known, which is usually not desirable from a practical point of view, and which we do not assume in our analysis of (5) below.

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 Velasco and Lobato (2006) who consider the model $\alpha(L)\Delta_+^d X_t = \varepsilon_t$ and test that $d = 1$. Here $\alpha(L)$ is a lag polynomial. They indicate the properties of the process under the null and under the alternative[3]. In both cases they apply a t-ratio based on a regression equation, which is motivated by the model equations, rather than a test based upon an analysis of the likelihood function.

The model (5) 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 (2007). In this way we have a platform for conducting model-based statistical inference on the parameters and on the fractional order of X_t.

---

[3]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 are $\pi(0) = 1, \pi(1) > 0, \pi(-1) > 0$. This cannot be correct as the example $\pi(z) = 4(z - 1/2)^2 = (1 - 4z)(1 - z) + z$ shows. Indeed, the solution would lead to an unpleasant transcendental equation, see the discussion in Johansen (2007), and thus it does not appear possible to give general conditions for fractionality of various orders in terms of the parameters of the model.

1.3 An overview of the present paper

The remainder of the paper is organized as follows. We discuss in section 2 the properties of the solution of the model (5) including the role of the initial values and give the Gaussian likelihood function and the profile likelihood function as a function of (d, b, ϕ). In section 3 we give the results on the convergence of the product moments as functions of (d, b, ϕ). These results are applied in section 4 to prove consistency and to show that $\hat{d}, \hat{b}$, and $\hat{\phi}$ are asymptotically Gaussian, whereas the asymptotic distribution of $\hat{\pi}$ is a functional of Brownian motion B and fractional Brownian motion B_{b_0-1} of type II. In section 5 we show that the asymptotic distribution of the likelihood ratio test for a (fractional) unit root is a functional of B and B_{b_0-1}. We conclude in section 6, and give some mathematical details in three appendices.

1.4 Notation

For a symmetric matrix A we write $A > 0$ to mean that it is positive definite. For a function $f : \mathbb{R}^p \mapsto \mathbb{R}$ we sometimes denote the vector of derivatives $\mathsf{D}f$ and matrix of second derivatives $\mathsf{D}^2 f$. The Euclidean norm of a vector or scalar a is denoted $|a|$. For a real number a we denote the positive part $a^+ = \max(0, a)$ and the negative part $a^- = -\min(0, a)$. Throughout ε_t is a sequence of i.i.d. variables with mean zero and variance $\sigma^2 > 0$. For coefficients ξ_n with $\sum_{n=0}^{\infty} \xi_n^2 < \infty$ we define $F(z) = \sum_{n=0}^{\infty} \xi_n z^n$ and the linear process $Z_t = F(L)\varepsilon_t = \sum_{n=0}^{\infty} \xi_n \varepsilon_{t-n}$. We use the notation $Z_t^+ = F_+(L)\varepsilon_t = \sum_{n=0}^{t-1} \xi_n \varepsilon_{t-n}$ and $Z_t^- = F_-(L)\varepsilon_t = \sum_{n=t}^{\infty} \xi_n \varepsilon_{t-n}$ for the corresponding truncated processes.

For a random variable Z with $E|Z|^p < \infty$ we use the notation $||Z||_p = E(|Z|^p)^{1/p}$. The probability results (for the model (5)) are derived under the assumption that the true parameters are $d_0 > b_0 > 1/2$, $\phi_0, \pi_0 = 0$, and $\sigma_0^2 > 0$. Expectation with respect to the true values is denoted by E. We also use the notation $\psi = (d, b)'$. We let $\mathbb{C}_b$ denote the image of the unit disk under the function $f(z) = 1 - (1-z)^b$, $b > 0$. Throughout, c denotes a generic positive constant which may take different values in different places.

In the following we apply the theory of weak convergence of probability measures, see Billingsley (1968) and Kallenberg (2001). It is convenient to describe it in terms of stochastic variables and processes, and for a sequence of $k-$dimensional stochastic processes $X_T(u)$, $u \in [0, 1]$, we write $X_T \Longrightarrow X$ or $X_T(u) \Longrightarrow X(u)$ to indicate convergence in distribution of the sequence, either on $D^k[0, 1]$ or $C^k[0, 1]$, whereas $X_T(u) \xrightarrow{d} X(u)$ means convergence in distribution on $\mathbb{R}^k$ for a fixed u. We let W denote Brownian motion generated by ε_t, $B = \sigma^{-1}W$ denote standard Brownian motion, and B_{d-1} denote the corresponding fractional Brownian motion of type II, $B_{d-1}(t) = \Gamma(d)^{-1} \int_0^t (t-s)^{d-1} dB(s)$, $d > 1/2$.

2 The conditional likelihood and profile likelihood functions for the fractional process

This section contains some definitions of fractional processes and a discussion of the properties of the solution of the autoregressive model (5) and the role of initial values. We also define the conditional likelihood function and the profile likelihood function, where the regression parameter π and the variance have been eliminated.

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 cn^{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

$$\Delta^{-d}\varepsilon_t = (1-L)^{-d}\varepsilon_t = \sum_{n=0}^{\infty}(-1)^n \binom{-d}{n}\varepsilon_{t-n}.$$

For $d \geq 1/2$ the infinite sum does not exist, but we can define a nonstationary process by the operator Δ_+^{-d},

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

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

We apply the result, e.g. Davydov (1970) and Akonom and Gourieroux (1987), that on $D[0, 1]$

$$T^{-d+1/2}\Delta_+^{-d}\varepsilon_{[Tu]} \Longrightarrow W_{d-1}(u) = \Gamma(d)^{-1}\int_0^u (u-s)^{d-1}dW(s) \qquad (7)$$

for each $d > 1/2$, which by the continuous mapping theorem implies that

$$T^{-2d}\sum_{t=1}^{T}(\Delta_+^{-d}\varepsilon_t)^2 \xrightarrow{d} \int_0^1 W_{d-1}^2 du. \qquad (8)$$

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

$$T^{-d}\sum_{t=1}^{T}\Delta_+^{-d}\varepsilon_t\varepsilon_{t+1} \xrightarrow{d} \int_0^1 W_{d-1}dW. \qquad (9)$$

In this paper we apply these results to analyze model (5) which has as a solution a fractional process, see Johansen (2007, Theorem 8). We formulate the solution of (5) and some of its properties in the next result.

Lemma 1 *If $\pi = 0$ then $a(1) = 0$, and if the remaining roots are outside the set $\mathbb{C}_{b_0}$, then for $\pi(z)$ given by (5) it holds that*

$$(1 - z)^{d_0}\pi(z)^{-1} = \gamma_0 + (1 - z)^{b_0}H(1 - (1 - z)^{b_0}), \tag{10}$$

where $\gamma_0 = (1 - \sum_{i=1}^{k}\phi_{0i})^{-1}$ and $H(u)$ is regular in a neighborhood of $\mathbb{C}_{b_0}$, so that the coefficients defined by $F(z) = H(1 - (1 - z)^{b_0}) = \sum_{n=0}^{\infty}\tau_n z^n$, $|z| < 1$, define a stationary process $Y_t = \sum_{n=0}^{\infty}\tau_n\varepsilon_{t-n}$. Then $\sum_{n=0}^{\infty}|\tau_n| < \infty$, so that the covariance function $\gamma_Y(h) = E(Y_t Y_{t-h})$ satisfies $\sum_{h=-\infty}^{\infty}|\gamma_Y(h)| < \infty$. Equation (5) is solved by

$$X_t = \gamma_0\Delta_+^{-d_0}\varepsilon_t + \Delta_+^{-d_0+b_0}Y_t^+ + \mu_t^0, \ t = 1, 2, \ldots, \tag{11}$$

where $Y_t^+ = \sum_{n=0}^{t-1}\tau_n\varepsilon_{t-n}$ is asymptotically stationary and the trend generated by the initial values is $\mu_t^0 = -\pi_+(L)^{-1}\pi_-(L)X_t$.

Proof. The expression for $(1-z)^{d_0}\pi(z)^{-1}$ and the definition of Y_t follows from Johansen (2007, Theorem 8). The transfer function for Y_t has the form

$$\phi(\lambda) = H(1 - (1 - e^{i\lambda})^b) = F(e^{i\lambda}),$$

where $F(z) = \sum_{n=0}^{\infty}\tau_n z^n$ and the series is convergent for $|z| \leq 1 + \delta$, for some $\delta > 0$, and $F(1) \neq 0$. It follows that $\phi(\lambda)$ is continuous, that $\phi(0) \neq 0$, and that $\phi(\lambda)$ has the expansion

$$\phi(\lambda) = \sum_{n=0}^{\infty}\tau_n e^{i\lambda n}$$

for some coefficients τ_n which are used to define the process Y_t. The derivative of $\phi(\lambda)$ is

$$\frac{d\phi}{d\lambda} = ibe^{i\lambda}\frac{dH}{dz}(1 - (1 - e^{i\lambda})^b)(1 - e^{i\lambda})^{b-1},$$

which 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 $d\phi/d\lambda$, $in\tau_n$, are square summable so that $\sum_{n=0}^{\infty}n^2\tau_n^2 < \infty$, see Zygmund (2003, p. 37). It follows from this that

$$(\sum_{n=0}^{\infty}|\tau_n|)^2 = (\sum_{n=0}^{\infty}|\tau_n n||n^{-1}|)^2 \leq \sum_{n=0}^{\infty}|\tau_n n|^2\sum_{n=0}^{\infty}n^{-2} < \infty,$$

so that $\sum_{n=0}^{\infty}|\tau_n| < \infty$. Finally

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

The expression (11) follows from (5) by applying $\pi_+^{-1}(L)$ to $\pi(L)X_t = \pi_+(L)X_t + \pi_-(L)X_t = \varepsilon_t$ to get

$$X_t = \pi_+^{-1}(L)\varepsilon_t - \pi_+^{-1}(L)\pi_-(L)X_t$$

and applying the expression (10) for $\pi_+^{-1}(L)\varepsilon_t$. $\blacksquare$

From (11) follow the properties of the process and we find in particular that

$$
\begin{aligned}
\Delta^{-u}X_t &= \gamma_0\Delta_+^{-d_0-u}\varepsilon_t + \Delta_+^{-d_0-u+b_0}Y_t^+ + \Delta_+^{-u}\mu_t^0 + \Delta_-^{-u}X_t, \ t = 1, 2, \dots, \quad (12) \\
\Delta^{-u}X_t &= \Delta_-^{-u}X_t, \ t = 0, -1, -2, \dots,
\end{aligned}
$$

where the first term of (12) is nonstationary, but asymptotically stationary for $d_0 + u < 1/2$. For $d_0 + u > 1/2$ it will, suitably normalized, converge to a fractional Brownian motion, see (7). The next term is asymptotically stationary when $d_0 + u - b_0 < 1/2$, and the last terms are deterministic functions of the initial values. The different processes will be studied in detail below.

In order to study the impact of the initial values on the process we apply the representations

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

for some coefficients ρ_j, and find

$$
\mu_t^0 = -\pi_+(L)^{-1}\pi_-(L)X_t = -(\gamma_0 + \Delta_+^{b_0}F_+(L))\sum_{j=0}^{k}\rho_j\Delta_+^{-d_0}\Delta_-^{d_0+jb_0}X_t. \quad (13)
$$

The theory in this paper will be developed for observations $X_1, \dots, X_T$ generated by (5) assuming that all initial values are observed, that is, conditional on X_{-n}^0, $n = 0, 1, \dots$ In practice, this is obviously not the case, and one will have to choose a value T_0 and base the calculations on setting $X_{-n}^0 = 0$, $n > T_0$. We call X_{-n}^0, $n = 0, \dots, T_0$, the observed initial values. One will then have to investigate the sensitivity to the initial values by choosing different values of T_0. For usual autoregressive models with k lags, the observed initial values will be $X_{-k+1}^0, \dots, X_0^0$.

Thus, the initial values are not modeled, and the asymptotic results show that the influence of the initial values disappears in the limit provided they are bounded, an assumption that appears reasonable in practice.

2.2 The conditional likelihood function

The model is

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

where ε_t is i.i.d. Gaussian $(0, \sigma^2)$ and we apply the lag operator $L_b = 1 - \Delta^b$. Note that the asymptotic properties are derived below without the Gaussianity assumption

but assuming ε_t is i.i.d. with mean zero and q finite moments. The parameter space is defined by assuming that $\sigma^2 > 0$, $d \geq b > 1/2$, and that the remaining parameters vary freely, but it is convenient to introduce the parameter $\theta = T^{d_0 - d + b - 1/2}\pi$. Note that the parameter space is unchanged. We use the notation $\tau = (d, b, \phi', \theta, \sigma^2)'$ and $\psi = (d, b)'$ so that the likelihood function, conditional on initial values $\{X^0_{-n}, n \geq 0\}$, becomes

$$-2T^{-1}\log L_T(\tau) \tag{14}$$

$$= \log\sigma^2 + \frac{1}{\sigma^2 T}\sum_{t=1}^{T}(\Delta^d X_t - T^{-d_0+d-b+1/2}\theta\Delta^{d-b}L_b X_t - \sum_{i=1}^{k}\phi_i\Delta^d L_b^i X_t)^2$$

$$= \log\sigma^2 + \frac{1}{\sigma^2}\begin{pmatrix} 1 \\ -\phi \\ -\theta \end{pmatrix}' \begin{pmatrix} \mathcal{B}_{00T}(\psi) & \mathcal{B}_{0*T}(\psi) & \mathcal{C}_{0T}(\psi) \\ \mathcal{B}_{*0T}(\psi) & \mathcal{B}_{**T}(\psi) & \mathcal{C}_{*T}(\psi)' \\ \mathcal{C}_{0T}(\psi) & \mathcal{C}_{*T}(\psi) & \mathcal{A}_T(\psi) \end{pmatrix} \begin{pmatrix} 1 \\ -\phi \\ -\theta \end{pmatrix}.$$

We here define the product moments

$$\mathcal{A}_T(\psi) = T^{-2(d_0-d+b)}\sum_{t=1}^{T}(\Delta^{d-b}L_b X_t)^2, \tag{15}$$

$$\mathcal{B}_{ijT}(\psi) = T^{-1}\sum_{t=1}^{T}(\Delta^d L_b^i X_t)(\Delta^d L_b^j X_t), \ i,j = 0, 1, \ldots, k, \tag{16}$$

$$\mathcal{C}_{iT}(\psi) = T^{-(d_0-d+b)-1/2}\sum_{t=1}^{T}(\Delta^{d-b}L_b X_t)(\Delta^d L_b^i X_t), \ i = 0, 1, \ldots, k. \tag{17}$$

The notation $\mathcal{B}_{*jT}(\psi)$ is used for the vector with components $\mathcal{B}_{ijT}(\psi)$, $i = 1, \ldots, k$, $\mathcal{B}_{i*T}(\psi) = \mathcal{B}_{*iT}(\psi)'$, and $\mathcal{B}_{**T}$ is the matrix with elements $\mathcal{B}_{ijT}(\psi)$, $i, j = 1, \ldots, k$. We also define the matrix

$$\mathcal{B}_T(\psi) = \begin{pmatrix} \mathcal{B}_{00T}(\psi) & \mathcal{B}_{0*T}(\psi) \\ \mathcal{B}_{*0T}(\psi) & \mathcal{B}_{**T}(\psi) \end{pmatrix}$$

and define $\mathcal{B}(\psi)$ as the probability limit of $\mathcal{B}_T(\psi)$, see Lemma 24.

We let $\mathcal{C}_{*T}(\psi)$ denote the vector with components $\mathcal{C}_{iT}(\psi)$, $i = 1, \ldots, k$ and $\mathcal{C}_T(\psi) = (\mathcal{C}_{0T}(\psi), \mathcal{C}_{*T}(\psi)')'$, and finally we need

$$\mathcal{C}_{0\varepsilon T}(\psi) = T^{-1/2}\sum_{t=1}^{T}(T^{-(d_0-d+b)}\Delta^{d-b}L_b X_t)\varepsilon_t. \tag{18}$$

2.3 The maximum likelihood estimators and profile likelihood function for fixed d, b, and ϕ

For fixed d, b, and ϕ we can find the maximum likelihood estimators and the profile likelihood function by regression,

$$\hat{\theta}(\psi, \phi) = (\mathcal{C}_{0T}(\psi) - \phi'\mathcal{C}_{*T}(\psi))/\mathcal{A}_T(\psi), \tag{19}$$

$$\hat{\sigma}^2(\psi, \phi) = \mathcal{B}_{00T}(\psi) - 2\phi'\mathcal{B}_{*0T}(\psi) + \phi'\mathcal{B}_{**T}(\psi)\phi - \frac{(\mathcal{C}_{0T}(\psi) - \phi'\mathcal{C}_{*T}(\psi))^2}{\mathcal{A}_T(\psi)}. \tag{20}$$

Finally we find the profile likelihood as

$$-2T^{-1} \log L_{\text{profile},T}(\psi, \phi) = -2T^{-1} \log \max_{\theta, \sigma^2} L_T(\psi, \phi, \theta, \sigma^2) = 1 + \log \hat{\sigma}^2(\psi, \phi). \quad (21)$$

In the following, we use this expression to investigate the profile likelihood function in a small neighborhood of the value (ψ_0, ϕ_0).

For model (2) with $b = d$, we find in the same way, writing $\mathcal{A}_T(d)$, $\mathcal{B}_{ijT}(d)$, and $\mathcal{C}_{iT}(d)$ for $\mathcal{A}_T(d, d)$, $\mathcal{B}_{ijT}(d, d)$, and $\mathcal{C}_{iT}(d, d)$, that

$$\hat{\sigma}^2(d, \phi) = \mathcal{B}_{00T}(d) - 2\phi' \mathcal{B}_{*0T}(d) + \phi' \mathcal{B}_{**T}(d)\phi - \frac{(\mathcal{C}_{0T}(d) - \phi' \mathcal{C}_{*T}(d))^2}{\mathcal{A}_T(d)},$$

$$-2T^{-1} \log L_{\text{profile},T}(d) = 1 + \log \hat{\sigma}^2(d, \phi).$$

We also define $\mathcal{B}(d)$ as the probability limit of $\mathcal{B}_T(d)$.

We conclude this section with the assumptions we shall use in the asymptotic analysis of our model.

Assumption 1 *The process X_t, $t = 1, \ldots, T$, is generated by model (5) 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 $q > \max(6, 2/(2b_0 - 1))$.*

TRUE VALUES: *The true values satisfy $d_0 > b_0 > 1/2$, $\pi_0 = 0$, $\sigma_0^2 > 0$, so that $a(u)$ has a unit root, and the remaining roots of $a(u)$ are outside the set $\mathbb{C}_{b_0}$.*

INITIAL VALUES: *The initial values $X_{-n}^0, n = 0, 1, \ldots$, are bounded, i.e. there exists a $c > 0$ such that $|X_{-n}^0| \le c$ for all $n \ge 0$.*

Assumption 2 *The process X_t, $t = 1, \ldots, T$, is generated by model (2) for some $k = 0, 1, 2, \ldots$ and satisfies:*

ERRORS: *The errors ε_t are i.i.d. $(0, \sigma^2)$ with $E|\varepsilon_t|^q < \infty$ for some $q > \max(4, 2/(2b_0 - 1))$.*

TRUE VALUES: *The true values satisfy $d_0 > 1/2$, $\pi_0 = 0$, $\sigma_0^2 > 0$, so that $a(u)$ has a unit root, and the remaining roots of $a(u)$ in (1) are outside the set $\mathbb{C}_{d_0}$.*

INITIAL VALUES: *The initial values $X_{-n}^0, n = 0, 1, \ldots$, are bounded, i.e. there exists a $c > 0$ such that $|X_{-n}^0| \le c$ for all $n \ge 0$.*

Importantly, the errors are not assumed Gaussian for the asymptotic analysis, but are only assumed to be i.i.d. with sufficient moments to apply a functional central limit theorem and our tightness arguments below. The TRUE VALUES assumption is the unit root assumption, which ensures that X_t is nonstationary and fractional of order d_0. The INITIAL VALUES assumption is needed so that $\Delta^d X_t$ can be calculated for any $d > 0$, and is sufficient for the asymptotic analysis of the conditional likelihood.

3 Weak convergence of the profile likelihood function

We first give a useful tightness criterion from Kallenberg (2001), generalizing a well known result from Billingsley (1968), and formulate and prove some simple consequences of tightness, convergence in distribution, and the continuous mapping theorem. We then give the result on the asymptotic behavior of the product moments $(\mathcal{A}_T(\psi), \mathcal{B}_T(\psi), \mathcal{C}_T(\psi))$ and their derivatives, and end this section with the weak limit of the profile likelihood function considered as a stochastic process in the parameters ψ and ϕ.

We apply the convergence result to processes defined on a compact set containing the true value (ψ_0, ϕ_0), but we formulate them, as is usually done, for the unit hypercube $[0, 1]^m$.

3.1 Some weak convergence results

Lemma 2 *If $X_n(s)$ is a sequence of p-dimensional continuous processes on $[0,1]^2$ for which $X_n(0)$ is tight and*

$$||X_n(s) - X_n(t)||_3 \leq c|s - t| \tag{22}$$

for some constant $c > 0$, which does not depend on n, s, or t, then $X_n(s)$ is tight on $C^p[0, 1]^2$.

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

Lemma 3 *If $X_n(s)$ satisfies (22) on $C^p[0, 1]^2$, $X_n(s_0)$ is tight on $\mathbb{R}^p$, and $f : [0, 1]^m \to \mathbb{R}^p$ is continuous, then $Z_n(s, u) = f(u)'X_n(s)$ is tight on $C[0, 1]^{2+m}$.*

Proof. Let

$$\omega_h(\delta) = \max_{|u-v| \leq \delta} |h(u) - h(v)|$$

denote the modulus of continuity of $h(u)$, which may be a deterministic function or a stochastic process. Then

$$Z_n(s, u) - Z_n(s^*, u^*) = f(u)'(X_n(s) - X_n(s^*)) + (f(u) - f(u^*))'X_n(s^*),$$

which shows that

$$\sup_n \omega_{Z_n}(\delta) \leq \max_{u \in [0,1]^m} |f(u)| \sup_n \omega_{X_n}(\delta) + \omega_f(\delta) \sup_n \max_{s \in [0,1]^2} |X_n(s)|.$$

By continuity of f, $\omega_f(\delta) \to 0$ as $\delta \to 0$, and tightness of X_n implies, by the Arzelà-Ascoli theorem see Kallenberg (2001, pp. 311 and 563), that $\sup_n \omega_{X_n}(\delta) \to 0$ as $\delta \to 0$. Finally $\max_{u \in [0,1]^m} |f(u)| < \infty$, and because the mapping $X_n(s) \longmapsto \max_{s \in [0,1]^2} |X_n(s)|$

is continuous and continuous mappings preserve compact sets (and thus tightness) it follows that also $\sup_n \max_{s \in [0,1]^2} |X_n(s)|$ is bounded on a set with large probability, so that $\sup_n \omega_{Z_n}(\delta) \xrightarrow{P} 0$ as $\delta \to 0$, which shows that $Z_n(s,u) = f(u)'X_n(s)$ is tight on $C[0,1]^{2+m}$. ■

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, ϕ) are all tight as processes in the parameters. Lemma 3 shows that this follows from the tightness of the product moments $\mathcal{A}_T$, $\mathcal{B}_T$, and $\mathcal{C}_T$.

In the next lemma we consider a sequence of univariate processes $X_n(s)$ where $s \in [0,1]^m$.

Lemma 4 *1. Assume $X_n(s_0) \xrightarrow{P} c > 0$ and $X_n(s)$ is tight on $C[0,1]^m$. Then for all $\eta > 0$ there is a $\delta > 0$ and an n_0 so that $P(\min_{|s-s_0| \le \delta} X_n(s) > 0) \ge 1 - \eta$ for $n \ge n_0$.*

2. Assume that $S_n \xrightarrow{P} s_0$ and $X_n(s)$ is tight on $C[0,1]^m$. Then $X_n(S_n) - X_n(s_0) \xrightarrow{P} 0$.

Proof. 1. We find for $|s - s_0| \le \delta$ that

$$X_n(s) = X_n(s_0) + (X_n(s) - X_n(s_0)) \ge X_n(s_0) - \omega_{X_n}(\delta).$$

By the Arzelà-Ascoli theorem, if X_n is tight on $C[0,1]^m$ and $X_n(s_0) \xrightarrow{P} c$, we can find for any $\eta > 0$, a $\delta > 0$ and an n_0 so that

$$
\begin{aligned}
P(A_{n1}) &= P(\omega_{X_n}(\delta) \le \frac{c}{3}) \ge 1 - \frac{\eta}{2}, \\
P(A_{n2}) &= P(|X_n(s_0) - c| \le \frac{c}{3}) \ge 1 - \frac{\eta}{2},
\end{aligned}
$$

for $n \ge n_0$. Let $A_n = A_{n1} \cap A_{n2}$. Then $P(A_n) \ge 1 - \eta$ for $n \ge n_0$, and on A_n we have the inequality

$$X_n(s) \ge c + X_n(s_0) - c - \omega_{X_n}(\delta) \ge c - \frac{c}{3} - \frac{c}{3} = \frac{c}{3} > 0,$$

for all $|s - s_0| \le \delta$.

2. To prove the second result we find

$$P(|X_n(S_n) - X_n(s_0)| > \varepsilon, \ |S_n - s_0| \ge \delta) \le P(|S_n - s_0| \ge \delta),$$

$$P(|X_n(S_n) - X_n(s_0)| > \varepsilon, \ |S_n - s_0| < \delta) \le P(\omega_{X_n}(\delta) \ge \varepsilon).$$

With the above δ and n_0 the last probability is less than η, and for n sufficiently large the first is less than η, which shows that

$$P(|X_n(S_n) - X_n(s_0)| > \varepsilon) \le P(|S_n - s_0| \ge \delta) + P(\omega_{X_n}(\delta) \ge \varepsilon) \le 2\eta,$$

which proves that $X_n(S_n) - X_n(s_0) \xrightarrow{P} 0$. ■

The first part of the lemma is used to show that with probability tending to one, the second derivative of the profile likelihood is positive definite, so that the profile likelihood itself is convex in a small δ neighborhood of the true value implying the existence of a local likelihood estimator.

The second part of Lemma 4 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 4 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 2.

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

Lemma 5 *For $u \in [0, 1]$ and $i = 1, 2$, let the processes V_{ut}^i, $t = 1, 2, \ldots$, be continuous in u and linear in the i.i.d. variables ε_t with finite sixth moment. If V_{ut}^i, $i = 1, 2$, satisfy*

$$||V_{ut}^i||_2 \leq c \text{ and } ||V_{ut}^i - V_{\tilde{u}t}^i||_2 \leq c|u - \tilde{u}|, \tag{23}$$

where the constants do not depend on $u \in [0, 1], \tilde{u} \in [0, 1]$, or $t \in [1, T]$, then, for $i, j = 1, 2$, the product moment

$$S_{uv} = T^{-1} \sum_{t=1}^{T} V_{ut}^i V_{vt}^j \tag{24}$$

is tight as a process in $(u, v) \in [0, 1]^2$.

If furthermore D_{ut}^i, $i = 1, 2$, are deterministic functions which are continuous in $u \in [0, 1]$ and satisfy

$$\max_{u \in [0,1]} |D_{ut}^i| \to 0 \text{ as } t \to \infty, \tag{25}$$

then

$$S_{uv}^* = T^{-1} \sum_{t=1}^{T} (V_{ut}^i + D_{ut}^i)(V_{vt}^j + D_{vt}^j)$$

is tight in $(u, v) \in [0, 1]^2$.

Proof. To prove (24) we apply the decomposition

$$S_{uv} - S_{\tilde{u}\tilde{v}} = T^{-1} \sum_{t=1}^{T} (V_{ut}^i - V_{\tilde{u}t}^i)V_{vt}^j + V_{\tilde{u}t}^i(V_{vt}^j - V_{\tilde{v}t}^j)$$

and the inequality (47) in Lemma 15 and find

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

This shows that the tightness criterion (22) from Lemma 2 holds.

In order to prove tightness of S^*_{uv} we note that

$$S^*_{uv} = S_{uv} + T^{-1} \sum_{t=1}^{T} V^i_{ut} D^j_{vt} + T^{-1} \sum_{t=1}^{T} D^i_{ut} V^j_{vt} + T^{-1} \sum_{t=1}^{T} D^i_{ut} D^j_{vt}.$$

We want to show tightness of the last three terms by showing that the supremum converges in probability to zero. We find using (25) that

$$\max_{(u,v) \in [0,1]^2} |T^{-1} \sum_{t=1}^{T} D^i_{ut} D^j_{vt}| \leq T^{-1} \sum_{t=1}^{T} |\max_{u \in [0,1]} D^i_{ut}||\max_{v \in [0,1]} D^j_{vt}| \to 0 \text{ as } T \to \infty$$

and

$$\max_{(u,v) \in [0,1]^2} (T^{-1} \sum_{t=1}^{T} V^i_{ut} D^j_{vt})^2 \leq \max_{v \in [0,1]} T^{-1} \sum_{t=1}^{T} (D^j_{vt})^2 \max_{u \in [0,1]} T^{-1} \sum_{t=1}^{T} (V^i_{ut})^2.$$

The first factor tends to zero by assumption (25), and, by (24), $T^{-1} \sum_{t=1}^{T} (V^i_{ut})^2$ is tight in u, so that $\max_{u \in [0,1]} T^{-1} \sum_{t=1}^{T} (V^i_{ut})^2$ is tight. Hence the product tends to zero in probability. ∎

Thus to establish tightness of product moments it is enough simply to check condition (23) for the stochastic parts of the involved processes and condition (25) for the deterministic parts of the processes.

3.2 Convergence of product moments and the likelihood profile

We are now ready to state the result on weak convergence of the product moments.

Theorem 6 *Let Assumption 1 be satisfied for model (5) and let $0 < \eta < \min(1/2, b_0 - 1/2, d_0 - b_0)$. We define $d_1 = \max(1/2, d_0 - 1/2) + \eta$, $d_2 > d_0$, and*

$$N_1 = \{(b,d) : d_1 \leq d \leq d_2, \ b \geq 1/2 + \eta, \ \eta \leq d - b \leq d_0 - 1/2 - \eta\}.$$

Then $\mathcal{A}_T(\psi)$, $\mathcal{B}_T(\psi)$, and $\mathcal{C}_T(\psi)$ and their derivatives are tight on $C(N_1)$, and for $m = 0, 1, 2$, it holds jointly that

$$\mathcal{A}_T(\psi) \Longrightarrow \mathcal{A}(\psi) = \gamma_0^2 \int_0^1 W^2_{d_0 - d + b - 1} du \quad \text{on } C(N_1), \tag{26}$$

$$\mathsf{D}^m \mathcal{B}_T(\psi) \Longrightarrow \mathsf{D}^m \mathcal{B}(\psi) \quad \text{on } C^{(k+1) \times (k+1)}(N_1), \tag{27}$$

$$\mathsf{D}^m \mathcal{C}_T(\psi) \Longrightarrow 0 \quad \text{on } C^{k+1}(N_1). \tag{28}$$

$$T^{1/2} \mathcal{C}_{0\varepsilon T}(\psi_0) \xrightarrow{d} \gamma_0 \int_0^1 W_{b_0 - 1} dW. \tag{29}$$

Let Assumption 2 be satisfied for model (2) and let $0 < \eta < \min(1/2, d_0 - 1/2)$ and $I_1 = [d_1, d_2]$. Then the same results hold for $\mathcal{A}_T(d)$, $\mathcal{B}_T(d)$, $\mathcal{C}_T(d)$, and $\mathcal{C}_{0\varepsilon T}(d_0) = \mathcal{C}_{0\varepsilon T}(d_0, d_0)$ and their derivatives, when N_1 is replaced by I_1, $d = b$, and $d_0 = b_0$.

The proof is given in appendix C.

We next apply these results to derive the weak limit of the profile likelihood functions for the models (5) and (2) as well as some useful properties.

Corollary 7 *Let Assumption 1 be satisfied for model (5), and define for $A > 0$ the set $N_2 = \{\phi : |\phi - \phi_0| \leq A\}$. Then, for any A, and $0 < \eta < \min(1/2, b_0 - 1/2)$ we have:*

1. The profile likelihood function converges weakly,

$$-2T^{-1} \log L_{profile,T}(\psi, \phi) \Longrightarrow 1 + \log \sigma^2(\psi, \phi) \ on \ C(N_1 \times N_2), \tag{30}$$

where $\sigma^2(\psi, \phi) = \mathcal{B}_{00}(\psi) - 2\phi' \mathcal{B}_{*0}(\psi) + \phi' \mathcal{B}_{**}(\psi)\phi$, see Lemma 24.

2. For $k = 0, 1, \ldots,$ $\mathsf{D}^2(-2T^{-1} \log L_{profile,T}(\psi, \phi))$ is tight on $C(N_1 \times N_2)$, and for $(\psi, \phi) = (\psi_0, \phi_0)$,

$$\mathsf{D}^2(-2T^{-1} \log L_{profile,T}(\psi_0, \phi_0)) \xrightarrow{d} \mathsf{D}^2 \log \sigma^2(\psi_0, \phi_0). \tag{31}$$

Proof. 1. The profile likelihood is given in (21) and involves the expression (20):

$$\hat{\sigma}^2(\psi, \phi) = \mathcal{B}_{00T}(\psi) - 2\phi' \mathcal{B}_{*0T}(\psi) + \phi' \mathcal{B}_{**T}(\psi)\phi - \frac{(\mathcal{C}_{0T}(\psi) - \phi' \mathcal{C}_{*T}(\psi))^2}{\mathcal{A}_T(\psi)}.$$

Because $(\mathcal{A}_T(\psi), \mathcal{B}_T(\psi), \mathcal{C}_T(\psi)) \Longrightarrow (\gamma_0^2 \int_0^1 W_{d_0-d+b-1}^2 du, \mathcal{B}(\psi), 0)$, see (26), (27), and (28), we find

$$(\mathcal{C}_{0T}(\psi) - \phi' \mathcal{C}_{*T}(\psi))^2 \mathcal{A}_T(\psi)^{-1} \Longrightarrow 0$$

and thus

$$\hat{\sigma}^2(\psi, \phi) \Longrightarrow \mathcal{B}_{00}(\psi) - 2\phi' \mathcal{B}_{*0}(\psi) + \phi' \mathcal{B}_{**}(\psi)\phi = \sigma^2(\psi, \phi),$$

which proves the first result.

2. The second derivative of the profile likelihood can be expressed in terms of $(\mathcal{A}_T(\psi), \mathcal{B}_T(\psi), \mathcal{C}_T(\psi))$ and their first two derivatives, and is therefore tight by Theorem 6 and Lemma 3. In order to determine the limit for $(\psi, \phi) = (\psi_0, \phi_0)$ we need the results (26) to (28) and the tightness of the second derivatives, and then we can apply Lemma 24. ∎

Corollary 8 *Let Assumption 2 be satisfied for model (2), and define the interval $I_1 = [d_1, d_2]$ for $d_1 = \max(1/2, d_0 - 1/2) + \eta$ and $d_2 > d_0$ with $0 < \eta < \min(1/2, d_0 - 1/2)$. Then we have:*

1. The profile likelihood function converges weakly,

$$-2T^{-1} \log L_{profile,T}(d, \phi) \Longrightarrow 1 + \log \sigma^2(d, \phi) \ on \ C(I_1 \times N_2), \tag{32}$$

where $\sigma^2(d, \phi) = \mathcal{B}_{00}(d) - 2\phi' \mathcal{B}_{*0}(d) + \phi' \mathcal{B}_{**}(d)\phi$, see Lemma 24.

2. The second derivative $\mathsf{D}^2(-2T^{-1} \log L_{profile,T}(d, \phi))$ is tight on $C(I_1 \times N_2)$, and for $(d, \phi) = (d_0, \phi_0)$,

$$\mathsf{D}^2(-2T^{-1} \log L_{profile,T}(d_0, \phi_0)) \xrightarrow{d} \mathsf{D}^2 \log \sigma^2(d_0, \phi_0). \tag{33}$$

3. For $k = 0$, the convergence in (32) holds on $C(I_1)$, and the limit $1 + \log \sigma^2(d) = 1 + \log \mathcal{B}_{00}(d)$ is strictly convex on $[d_1, d_2]$ with a minimum at $d = d_0$.

Proof. The proofs of 1. and 2. follow as in Corollary 7. To see 3., we note that when $k = 0$ and $\pi_0 = 0$, the process is $X_t = \Delta_+^{-d_0}\varepsilon_t + \mu_t^0$, see (11), and the limit of $\hat{\sigma}^2(d)$ is

$$
\begin{aligned}
\mathcal{B}_{00}(d) &= \lim_{T\to\infty} (T^{-1}\sum_{t=1}^{T} E(\Delta_+^{d-d_0}\varepsilon_t + \Delta_+^d\mu_t^0 + \Delta_-^d X_t)^2) \\
&= \lim_{T\to\infty} T^{-1}\sigma_0^2 \sum_{t=1}^{T}\sum_{j=0}^{t-1}\pi_j(d_0-d)^2 = \sigma_0^2 \frac{\Gamma(1-2(d_0-d))}{\Gamma(1-(d_0-d))^2}, \ d \in [d_1, d_2],
\end{aligned}
$$

where the second equality uses (59) with $u = d_0 - d \leq 1/2 - \eta$ and $v = d_0 + jb_0 \geq d_0 > d_1 \geq 1/2 + \eta$ for $j = 0, \ldots, k$ (noting that (d_0, b_0) is an interior point in N_1 and therefore $d_0 > d_1$), and (57) with $v = d \geq d_1 \geq 1/2 + \eta$ from Lemma 18 such that $T^{-1}\sum_{t=1}^{T}(\Delta_+^d\mu_t^0)^2 \to 0$ and $T^{-1}\sum_{t=1}^{T}(\Delta_-^d X_t)^2 \to 0$, and hence the initial values have no influence on the limit. We find

$$
\mathsf{D}\log\frac{\Gamma(1-2(d_0-d))}{\Gamma(1-(d_0-d))^2}\Big|_{d=d_0} = 2\psi(1-2(d_0-d))|_{d=d_0} - 2\psi(1-(d_0-d))|_{d=d_0} = 0,
$$

where $\psi(\cdot)$ is the digamma function,

$$
\psi(\zeta) = \mathsf{D}\log\Gamma(\zeta) = C - \frac{1}{\zeta} + \sum_{i=1}^{\infty}(\frac{1}{i} - \frac{1}{\zeta+i}), \tag{34}
$$

and C is Euler's constant. Using the multiplication formula $2^\zeta\Gamma(\frac{\zeta}{2})\Gamma(\frac{\zeta+1}{2}) = 2\Gamma(\zeta)\Gamma(\frac{1}{2})$, see Artin (1964, p. 24), we find that for $\zeta = 1 - 2(d_0 - d)$,

$$
\frac{\Gamma(1-2(d_0-d))}{\Gamma(1-(d_0-d))^2} = \frac{\Gamma(1/2-(d-d_0))}{\Gamma(1-(d_0-d))\Gamma(1/2)}2^{-2(d-d_0)}.
$$

Hence

$$
\mathsf{D}^2\log\frac{\Gamma(1-2(d_0-d))}{\Gamma(1-(d_0-d))^2} = \sum_{i=0}^{\infty}(\frac{1}{(1/2-(d_0-d)+i)^2} - \frac{1}{(1-(d_0-d)+i)^2}) > 0.
$$

∎

4 Asymptotic properties of the local likelihood estimator

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

4.1 Existence and consistency of the local likelihood estimator

We apply weak convergence of the profile likelihood function and its derivatives to show that there is a neighborhood of (ψ_0, ϕ_0) on which the likelihood profile is convex with probability tending to one, so that the (local) likelihood estimator for (ψ, ϕ) exists and is consistent.

Theorem 9 *Let Assumption 1 be satisfied for model (5) or Assumption 2 for model (2).*

1. Then for model (5) with $b_0 < d_0$, $k > 0$, and $\phi_0 \neq 0$, there exists a neighborhood $N(\psi_0, \phi_0)$ of (ψ_0, ϕ_0) and a sequence of sets K_T with probability tending to one, such that on K_T the local likelihood estimators of d, b, ϕ, θ, and σ^2 exist uniquely and are consistent.

2. The same result holds for the local likelihood estimators of d, ϕ, θ, and σ^2 in model (2) where $d_0 = b_0$ and $k \geq 0$.

3. If $k = 0$ in model (2), the limit of the profile likelihood is convex on any interval $[d_1, d_2]$ in $]\max(1/2, d_0 - 1/2), \infty[$, so that on K_T the maximum likelihood estimator exists uniquely and is consistent, for $d \in [d_1, d_2]$.

Proof. *Existence and uniqueness:* We give the proof for model (5). The limit $1 + \log \sigma^2(\psi, \phi)$ of the profile likelihood function on the set $N_1 \times N_2$ is given in Corollary 7. For $k > 0$ and $\phi_0 \neq 0$, the limit has a positive definite second derivative for $(\psi, \phi) = (\psi_0, \phi_0)$, see (72) and (76) in Lemma 24. Note that (ψ_0, ϕ_0) is an interior point in $N_1 \times N_2$ by definition of η and d_1.

The function $\lambda_{\min}(\cdot)$ which to a symmetric matrix associates the smallest eigenvalue is a continuous function and we therefore have, see (31), that

$$\lambda_{\min}(-2T^{-1}\mathsf{D}^2 \log L_{\mathrm{profile},T}(\psi_0, \phi_0)) \xrightarrow{d} \lambda_{\min}(\mathsf{D}^2 \log \sigma^2(\psi_0, \phi_0)) > 0.$$

We then apply Lemma 4 which states that because $-2T^{-1}\mathsf{D}^2 \log L_{\mathrm{profile},T}(\psi, \phi)$ is tight on $N_1 \times N_2$, the set

$$\hat{K}_T = \{\min_{|(\psi,\phi)-(\psi_0,\phi_0)| \leq \delta_1} \lambda_{\min}(-2T^{-1}\mathsf{D}^2 \log L_{\mathrm{profile},T}(\psi, \phi)) > 0\}$$

has probability tending to one, where δ_1 has been chosen so small that

$$N_{\delta_1}(\psi_0, \phi_0) = \{(\psi, \phi) : |(\psi, \phi) - (\psi_0, \phi_0)| \leq \delta_1\} \subset N_1 \times N_2.$$

Define the minimum of $\log(\sigma^2(\psi, \phi)/\sigma^2(\psi_0, \phi_0))$ on the boundary of the neighborhood $N_{\delta_1}(\psi_0, \phi_0)$ as

$$\zeta(\delta_1) = \min_{|(\psi,\phi)-(\psi_0,\phi_0)|=\delta_1} \log \frac{\sigma^2(\psi, \phi)}{\sigma^2(\psi_0, \phi_0)}.$$

The continuous mapping theorem shows that, because the profile likelihood function converges in distribution,

$$Z_T = \min_{|(\psi,\phi)-(\psi_0,\phi_0)|=\delta_1} (-2T^{-1} \log L_{\mathrm{profile},T}(\psi, \phi) - 1 - \log \sigma^2(\psi_0, \phi_0))$$

converges in probability to

$$
\begin{aligned}
Z &= \min_{|(\psi,\phi)-(\psi_0,\phi_0)|=\delta_1} \left(1 + \log \sigma^2(\psi,\phi) - 1 - \log \sigma^2(\psi_0,\phi_0)\right) \\
&= \min_{|(\psi,\phi)-(\psi_0,\phi_0)|=\delta_1} \log \frac{\sigma^2(\psi,\phi)}{\sigma^2(\psi_0,\phi_0)} = \zeta(\delta_1).
\end{aligned}
$$

The distribution of Z is degenerate at the point $\zeta(\delta_1)$, so that all other points, in particular $\frac{1}{2}\zeta(\delta_1)$, are continuity points of the distribution function. Therefore $\tilde{K}_T = \{Z_T \geq \frac{1}{2}\zeta(\delta_1)\}$ satisfies $P(\tilde{K}_T) \to 1$.

By the same argument we have that in the neighborhood $N_{\delta_1}(\psi_0,\phi_0)$,

$$
Z_T^* = \min_{(\psi,\phi)\in N_{\delta_1}(\psi_0,\phi_0)} \left(-2T^{-1}\log L_{\mathrm{profile},T}(\psi,\phi) - 1 - \log \sigma^2(\psi_0,\phi_0)\right)
$$

converges in probability to

$$
Z^* = \min_{(\psi,\phi)\in N_{\delta_1}(\psi_0,\phi_0)} \log \frac{\sigma^2(\psi,\phi)}{\sigma^2(\psi_0,\phi_0)} \leq 0
$$

since the function $\log(\sigma^2(\psi,\phi)/\sigma^2(\psi_0,\phi_0))$ attains the value zero at $(\psi,\phi) = (\psi_0,\phi_0)$, so that $K_T^* = \{Z_T^* \leq \frac{1}{4}\zeta(\delta_1)\}$ satisfies $P(K_T^*) \to 1$.

For any observation in $\hat{K}_T$ the profile likelihood function is strictly convex. For any observation in $K_T = \hat{K}_T \cap \tilde{K}_T \cap K_T^*$ it attains its unique minimum in the interior of $N_{\delta_1}(\psi_0,\phi_0)$ because the function $-2T^{-1}\log L_{\mathrm{profile},T}(\psi,\phi) - 1 - \log \sigma^2(\psi_0,\phi_0)$ is no less than $\frac{1}{2}\zeta(\delta_1)$ on the boundary but attains a value no greater than $\frac{1}{4}\zeta(\delta_1)$ in the interior. Therefore there exists a unique minimizer, that is, a (local) solution $(\hat{\psi},\hat{\phi})$ of the likelihood equation exists uniquely for $(\psi,\phi) \in N_{\delta_1}(\psi_0,\phi_0)$ which satisfies $-2T^{-1}\mathsf{D}\log L_{\mathrm{profile},T}(\hat{\psi},\hat{\phi}) = 0$.

Consistency: The above arguments hold for all δ_1, and consistency of $(\hat{\psi},\hat{\phi})$ thus follows by taking δ_1 small.

Consistency of $(\hat{\psi},\hat{\phi})$ implies by the second part of Lemma 4 that, because of the tightness of $(\mathcal{A}_T, \mathcal{B}_T, \mathcal{C}_T)$ as processes indexed by ψ, we have that $(\mathcal{A}_T(\hat{\psi}), \mathcal{B}_T(\hat{\psi}), \mathcal{C}_T(\hat{\psi})) \overset{d}{\to} (\mathcal{A}(\psi_0), \mathcal{B}(\psi_0), 0)$, see also Theorem 6. Therefore

$$
\hat{\theta} = \frac{\mathcal{C}_{0T}(\hat{\psi}) - \hat{\phi}' C_{*T}(\hat{\psi})}{\mathcal{A}_T(\hat{\psi})} \overset{P}{\to} 0, \tag{35}
$$

so that $\hat{\theta}$ is consistent. By (20) and (35), $\hat{\sigma}^2$ has the same limit as

$$
\mathcal{B}_{00T}(\hat{\psi}) - 2\hat{\phi}'\mathcal{B}_{*0T}(\hat{\psi}) + \hat{\phi}'\mathcal{B}_{**T}(\hat{\psi})\hat{\phi}.
$$

Because $\mathcal{B}_T(\psi)$ is tight we can replace $\hat{\psi}$ by ψ_0, and find that $\hat{\sigma}^2$ converges in probability to $\sigma^2(\psi_0,\phi_0) = \sigma_0^2$, see Lemma 24.

The proof of the same result for model (2) is similar.

Finally, for $k = 0$ and $b = d$, we have weak convergence of the second derivative of the profile likelihood,

$$\mathsf{D}^2(-2T^{-1}\log L_{\mathrm{profile},T}(d)) \implies \mathsf{D}^2 \log \mathcal{B}_{00}(d) \text{ on } C(I_1),$$

see Theorem 6 and Corollary 8. In this case we can thus redefine the set K_T as $K_T = \{\min_{d\in I_1}(-2T^{-1}\mathsf{D}^2\log L_{\mathrm{profile},T}(d)) > 0\}$, and by the continuous mapping theorem

$$\min_{d\in I_1}(-2T^{-1}\mathsf{D}^2\log L_{\mathrm{profile},T}(d)) \implies \min_{d\in I_1}\mathsf{D}^2\log \mathcal{B}_{00}(d) > 0$$

such that $P\{K_T\} \to 1$. It follows that, for $d \in I_1$, the maximum likelihood estimator exists uniquely and is consistent. ∎

Note that for the simple model in (3) with $k = 0$ and $b = d$ we get global convergence of the profile likelihood function and can prove that it is convex on the interval $[d_1, d_2]$ using weak convergence the second derivative. For the general model we can prove convexity of the profile likelihood only in a small neighborhood of (d_0, b_0) using tightness. Thus, we obtain existence, uniqueness, and consistency of the estimators globally for the model (3) but only locally for the general model (5).

4.2 Asymptotic distribution of the local likelihood estimator

We first 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 4 we only need the information at $\breve{\tau}$ since the estimators are consistent (by Theorem 9) and the second derivatives are tight (by Theorem 6). Again 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 (5) at $\breve{\tau} = (d_0, b_0, \phi_0, 0, \hat{\sigma}^2)$ is given by*

$$\begin{pmatrix} T^{-1/2}\mathsf{D}\log L_T(\breve{\tau}) \\ T^{-1/2}\frac{\partial}{\partial\theta}\log L_T(\breve{\tau}) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} N_{2+k}\left(0, \sigma_0^{-2}\Sigma(\psi_0, \phi_0)\right) \\ \gamma_0 \int_0^1 B_{b_0-1}dB \end{pmatrix}, \tag{36}$$

where $\Sigma(\psi_0, \phi_0)$ is given in (72).

Proof. Let $\varepsilon_t(\psi, \phi) = \Delta^d X_t - \sum_{i=1}^k \phi_i \Delta^d L_b^i X_t$. Because $\varepsilon_t(\psi_0, \phi_0) = \varepsilon_t$, see (78), we find the score function for (ψ, ϕ) to be

$$T^{-1/2}\mathsf{D}\log L_T(\breve{\tau}) = -\hat{\sigma}^{-2}T^{-1/2}\sum_{t=1}^T \varepsilon_t\mathsf{D}\varepsilon_t(\psi_0, \phi_0). \tag{37}$$

Since $\varepsilon_t\mathsf{D}\varepsilon_t(\psi_0, \phi_0)$ is a stationary martingale difference, see (79), with finite third moment, we find the first result in (36) from the central limit theorem for martingale difference sequences, see Hall and Heyde (1980, chp. 3).

The score function for θ is

$$T^{-1/2}\frac{\partial}{\partial\theta}\log L_T(\check{\tau}) = \hat{\sigma}^{-2}T^{-b_0}\sum_{t=1}^{T}(\Delta^{d_0-b_0}L_{b_0}X_t)\varepsilon_t = \hat{\sigma}^{-2}T^{1/2}C_{0\varepsilon T}(\psi_0),$$

which converges as indicated, see (29). ∎

Lemma 11 *Under Assumption 1 the Gaussian information per observation for model (5) at $\check{\tau} = (d_0, b_0, \phi_0, 0, \hat{\sigma}^2)$ converges in distribution to*

$$\begin{pmatrix} \sigma_0^{-2}\Sigma(\psi_0, \phi_0) & 0 \\ 0 & \gamma_0^2\int_0^1 B_{b_0-1}^2 du \end{pmatrix}. \tag{38}$$

Proof. We find that

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

$$\xrightarrow{P} \sigma_0^{-2}\Sigma(\psi_0, \phi_0)$$

by (72), (79), and a law of large numbers. We also have that

$$-T^{-1}\frac{\partial^2}{\partial\theta\partial\phi}\log L_T(\check{\tau}) = \hat{\sigma}^{-2}C_{*T}(\psi_0) \xrightarrow{P} 0,$$

$$-T^{-1}\frac{\partial^2}{\partial\theta\partial\psi}\log L_T(\check{\tau}) = \hat{\sigma}^{-2}(\phi_0'\frac{\partial}{\partial\psi}C_{*T}(\psi_0) - \frac{\partial}{\partial\psi}\mathcal{C}_{0T}(\psi_0)) \xrightarrow{P} 0,$$

$$-T^{-1}\frac{\partial^2}{\partial\theta^2}\log L_T(\check{\tau}) = \hat{\sigma}^{-2}\mathcal{A}_T(\psi_0) \xrightarrow{d} \sigma_0^{-2}\gamma_0^2\int_0^1 W_{b_0-1}^2 du,$$

by Theorem 6. ∎

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

Theorem 12 *Under Assumption 1 the asymptotic distribution of the (local) Gaussian maximum likelihood estimators $(\hat{d}, \hat{b}, \hat{\phi}, \hat{\pi})$ for model (5) is given by*

$$\begin{pmatrix} T^{1/2}(\hat{d} - d_0) \\ T^{1/2}(\hat{b} - b_0) \\ T^{1/2}(\hat{\phi} - \phi_0) \\ T^{b_0}\hat{\pi} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} N_{2+k}\left(0, \sigma_0^2\Sigma(\psi_0, \phi_0)^{-1}\right) \\ \int_0^1 B_{b_0-1}dB(\gamma_0\int_0^1 B_{b_0-1}^2 du)^{-1} \end{pmatrix}. \tag{39}$$

For model (2) where $d = b$, we find

$$\begin{pmatrix} T^{1/2}(\hat{d} - d_0) \\ T^{1/2}(\hat{\phi} - \phi_0) \\ T^{b_0}\hat{\pi} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} N_{1+k}\left(0, \sigma_0^2(M'\Sigma(\psi_0, \phi_0)M)^{-1}\right) \\ \int_0^1 B_{b_0-1}dB(\gamma_0\int_0^1 B_{b_0-1}^2 du)^{-1} \end{pmatrix}, \tag{40}$$

see (77).

Proof. *Proof of (39):* To find the limit distributions of $\hat{d}$, $\hat{b}$, $\hat{\phi}$, and $\hat{\theta}$, we apply the usual expansion of the score function. We expand the first derivatives of

$$l_T(\hat{\tau}) = -T^{-1}\log L_T(\hat{\tau})$$

around the value $\breve{\tau} = (d_0, b_0, \phi_0, 0, \hat{\sigma}^2)$. Using Taylor's formula with remainder term we find (with subscripts denoting partial derivatives)

$$0 = \begin{pmatrix} T^{1/2}l_{T\psi}(\breve{\tau}) \\ T^{1/2}l_{T\phi}(\breve{\tau}) \\ T^{1/2}l_{T\theta}(\breve{\tau}) \end{pmatrix} + \begin{pmatrix} l_{T\psi\psi}(\breve{\tau}^*) & l_{T\psi\phi}(\breve{\tau}^*) & l_{T\psi\theta}(\breve{\tau}^*) \\ l_{T\phi\psi}(\breve{\tau}^{**}) & l_{T\phi\phi}(\breve{\tau}^{**}) & l_{T\phi\theta}(\breve{\tau}^{**}) \\ l_{T\theta\psi}(\breve{\tau}^{***}) & l_{T\theta\phi}(\breve{\tau}^{***}) & l_{T\theta\theta}(\breve{\tau}^{***}) \end{pmatrix} \begin{pmatrix} T^{1/2}(\hat{\psi} - \psi_0) \\ T^{1/2}(\hat{\phi} - \phi_0) \\ T^{1/2}\hat{\theta} \end{pmatrix}.$$

$$(41)$$

Here the asterisks indicate intermediate points between $\hat{\tau}$ and $\breve{\tau}$, which hence converge to τ_0 in probability.

The score functions normalized by $T^{1/2}$ and their weak limits are given by Lemma 10. Because the second derivatives are tight, see Theorem 6 and Lemma 3, and $(\breve{\tau}^*, \breve{\tau}^{**}, \breve{\tau}^{***}) \xrightarrow{P} (\tau_0, \tau_0, \tau_0)$ we apply Lemma 4 to replace the intermediate points by τ_0 and find the limit of the information per observation in Lemma 11, see (38). Premultiplying by its inverse we find (39).

Proof of (40): The same proof applies and we find the expression for the asymptotic variance from (77) in Lemma 24. ∎

We remark that the asymptotic distribution is normal for the estimators of the fractional and autoregressive parameters, whereas the asymptotic distribution of the estimator of the unit root is non-normal and of the Dickey-Fuller type, where some of the usual Brownian motions have been replaced by fractional Brownian motion. Similar results 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 Brownian motion since, in our notation, their $b = b_0 = 1$.

Note also that the order of the fractional Brownian motion depends on the distance between the fractional order of X_t when $\pi = 0$ (i.e. in the data generating process) and when $\pi \neq 0$. 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 in the sense that the rate of convergence is T^{b_0}, which is more than root-T-consistent because $b_0 > 1/2$.

5 The likelihood ratio test for a (fractional) unit root

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

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

where the restricted maximum likelihood estimators, $\tilde{\psi}$ and $\tilde{\phi}$ when $\pi = 0$, satisfy $\frac{\partial}{\partial \psi} \sigma^2(\tilde{\psi}, \tilde{\phi}) = 0$, $\frac{\partial}{\partial \phi} \sigma^2(\tilde{\psi}, \tilde{\phi}) = 0$, and $\tilde{\sigma}^2 = \sigma^2(\tilde{\psi}, \tilde{\phi})$. The consistency of the estimator $(\tilde{\psi}, \tilde{\phi})$ follows from the consistency of $(\hat{\psi}, \hat{\phi})$.

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

$$-2 \log LR_T(\pi = 0) \xrightarrow{d} \frac{(\int_0^1 B_{b_0-1} dB)^2}{\int_0^1 B_{b_0-1}^2 du}. \tag{42}$$

Proof. Let $l_T(\tau) = -2T^{-1} \log L_T(\tau)$, and denote derivatives by subscripts. The 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 (41). The expansion of the likelihood ratio test of a simple hypothesis gives

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

say. With the notation $\eta = (d, b, \phi, \sigma^2)$ we then get

$$
\begin{aligned}
-2 \log \frac{L_T(\hat{\eta}, \hat{\theta})}{L_T(\eta_0, 0)} &= T \begin{pmatrix} l_{T\eta}(\tau_0) \\ l_{T\theta}(\tau_0) \end{pmatrix}' \begin{pmatrix} i_{T\eta\eta}^* & i_{T\eta\theta}^* \\ i_{T\theta\eta}^* & i_{T\theta\theta}^* \end{pmatrix}^{-1} \begin{pmatrix} l_{T\eta}(\tau_0) \\ l_{T\theta}(\tau_0) \end{pmatrix} \\
&= T l_{T\eta}(\tau_0)'(i_{T\eta\eta}^*)^{-1} l_{T\eta}(\tau_0) + T \frac{(l_{T\theta}(\tau_0) - i_{T\theta\eta}^*(i_{T\eta\eta}^*)^{-1} l_{T\eta}(\tau_0))^2}{(i_{T\theta\theta}^* - i_{T\theta\eta}^*(i_{T\eta\eta}^*)^{-1} i_{T\eta\theta}^*)}.
\end{aligned}
$$

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

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

so that the test for $\pi = 0$ becomes

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

Because $l_{T\eta\eta}$ is tight (see Lemma 3 and Theorem 6), $\hat{\tau}$ and $\tilde{\tau}$ are consistent, and $T^{1/2}l_{T\eta}(\tau_0)$ converges in distribution, we find that

$$
T^{1/2}l_{T\eta}(\tau_0)'[(i^*_{T\eta\eta})^{-1} - (i^{**}_{T\eta\eta})^{-1}]T^{1/2}l_{T\eta}(\tau_0) \xrightarrow{P} 0.
$$

Moreover we see from (38) that $i^*_{T\theta\eta}(i^*_{T\eta\eta})^{-1} \xrightarrow{P} 0$, 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 fractional Brownian motion functionals replacing the usual Brownian motion functionals as integrand. Note that a test of the $I(1)$ hypothesis in our framework would entail jointly testing $\pi = 0$ 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 (42) and a χ^2-distributed random variable with one degree of freedom.

Similar asymptotic distributions as those in our Theorem 13 are obtained by Dolado, Gonzalo, and Mayoral (2002) and Lobato and Velasco (2006), although these authors analyze other test statistics. On the other hand, Ling and Li (2001) obtain the usual Dickey-Fuller distribution since their model has $b = b_0 = 1$.

6 Conclusion

In this paper 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 in that it allows for solutions where the process is fractional of order d or $d - b$, where $d \geq b > 1/2$ are parameters to be estimated. 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 the initial values X^0_{-n}, $n = 0, 1, \ldots$, under the assumption that the errors are i.i.d. Gaussian. Our main technical tool is to consider the likelihood and its derivatives as stochastic processes in the parameters under the assumptions that the errors are i.i.d. with suitable moment conditions and that the initial values are bounded. Conditioning on initial values results in the use of the type II fractional Brownian motion for the asymptotic analysis. We apply these results to prove that the likelihood and its derivatives converge in distribution, and use this to discuss the existence, consistency, and asymptotic distribution of the local

likelihood estimator, as well as the distribution of the associated likelihood ratio test of the fractional unit root hypothesis.

A   Some inequalities

For a random variable X we define the norm $||X||_p = (E|X|^p)^{1/p}$ if $E|X|^p < \infty$ and note the properties

$$||X + Y||_p \leq ||X||_p + ||Y||_p, \ \ ||XY||_p \leq ||X||_{2p}||Y||_{2p}, \text{ for } p > 1. \tag{43}$$

The first inequality states that $|| \cdot ||_p$ is a norm (triangle inequality) and the second follows from the Cauchy-Schwarz inequality.

Lemma 14 *Let ε_t be i.i.d. with mean zero and finite n'th cumulant $\kappa_n(\varepsilon)$, 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 for $n = 1, 2, \ldots$*

$$|\kappa_n(Z)| \ \leq \ |\kappa_n(\varepsilon)|(\sum_{j=0}^{\infty} \xi_j^2)^{n/2}, \tag{44}$$

$$||Z||_n \ \leq \ c_n||Z||_2, \tag{45}$$

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

Proof. For $n = 1$, the results hold trivially because $E(Z) = \kappa_1(\varepsilon) = 0$. The characteristic function of Z is given by $\phi_Z(\lambda) = E(e^{i\lambda Z}) = \prod_{j=0}^{\infty} \phi_\varepsilon(\lambda\xi_j)$, so that the cumulants are

$$\kappa_n(Z) = (-i)^n \mathsf{D}^n \log \phi_Z(0) = \sum_{j=0}^{\infty} \xi_j^n (-i)^n \mathsf{D}^n \log \phi_\varepsilon(0) = \kappa_n(\varepsilon) \sum_{j=0}^{\infty} \xi_j^n.$$

We thus show the inequality

$$\sum_{j=0}^{\infty} |\xi_j|^n \leq (\sum_{j=0}^{\infty} \xi_j^2)^{n/2}, \ n = 2, 3, \ldots, \tag{46}$$

which will complete the proof of (44). Let first $n = 2m$, and note that

$$(\sum_{j=0}^{\infty} \xi_j^2)^m = \sum_{j_1,\ldots,j_m}^{\infty} \xi_{j_1}^2 \cdots \xi_{j_m}^2 \geq \sum_{j_1=\ldots=j_m}^{\infty} \xi_{j_1}^2 \cdots \xi_{j_m}^2 = \sum_{j=0}^{\infty} \xi_j^{2m}.$$

Next let $n = 2m + 1$ and apply the Cauchy-Schwarz inequality

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

Finally we want to prove (45). From Kendall and Stuart (1977, p. 70) we find a relation between moments and cumulants,

$$E(Z^n) = \sum_{m=1}^{n} \sum_{q} \frac{n!}{q_1! \cdots q_m!} \prod_{i=1}^{m} \left(\frac{\kappa_{p_i}(Z)}{p_i!}\right)^{q_i},$$

where the summation over q extends over all non-negative integers $(q_1, \ldots, q_m)$ and $(p_1, \ldots, p_m)$ such that $p_1 q_1 + \ldots + p_m q_m = n$. We then find

$$
\begin{aligned}
|E(Z^n)| &\leq c_n \sum_{m=1}^{n} \sum_{q} \prod_{i=1}^{m} \left|\frac{\kappa_{p_i}(Z)}{p_i!}\right|^{q_i} \leq c_n \sum_{m=1}^{n} \sum_{q} \prod_{i=1}^{m} |\kappa_{p_i}(\varepsilon)(\sum_{j=0}^{\infty} \xi_j^2)^{p_i/2}|^{q_i} \\
&\leq c_n \left(\sum_{j=0}^{\infty} \xi_j^2\right)^{\sum_{i=1}^{m} q_i p_i/2} = c_n \left(\sum_{j=0}^{\infty} \xi_j^2\right)^{n/2} \leq c_n E(Z^2)^{n/2},
\end{aligned}
$$

which proves (45). ∎

Lemma 15 *Let U_t, V_t, X_t, Y_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*

$$\|\sum_{t=1}^{T} X_t U_t - \sum_{t=1}^{T} Y_t V_t\|_3 \leq c \sum_{t=1}^{T} (\|X_t\|_2 \|U_t - V_t\|_2 + \|V_t\|_2 \|X_t - Y_t\|_2), \qquad (47)$$

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

Proof. The inequality follows by using the properties (43) with $p = 3$,

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

and then applying Lemma 14. ∎

Lemma 16 *We have*

$$\kappa(t, \alpha, \beta) = \sum_{j=0}^{t} |\pi_j(\alpha) \pi_{t-j}(\beta)| \leq c \begin{cases} t^{\alpha+\beta-1}, & \alpha \geq 0 \text{ and } \beta \geq 0, \\ t^{\max(\alpha,\beta)-1}, & \alpha < 0 \text{ or } \beta < 0, \end{cases} \qquad (48)$$

$$\xi(t, \alpha, \beta) = \max_{i,k} \sum_{j=\max(i,k)}^{t} |\pi_{j-i}(\alpha) \pi_{j-k}(\beta)| \leq c(\log t) \begin{cases} t^{(\alpha+\beta-1)^+}, & \alpha \leq 1 \text{ and } \beta \leq 1, \\ t^{\alpha^+ + \beta^+ - 1}, & \alpha > 1 \text{ or } \beta > 1. \end{cases}$$

$$(49)$$

Proof. *Proof of (48):* For $\alpha = \beta = 0$ we find $\kappa(t, 0, 0) = 1_{\{t=0\}}$, and for $\alpha = 0 \neq \beta$ we find $\kappa(t, 0, \beta) = |\pi_t(\beta)| \leq ct^{\beta-1}$, which shows (48) in case either α or β is zero. For $\alpha \neq 0$ and $\beta \neq 0$, we apply the inequality

$$\sum_{j=0}^{t} |\pi_j(\alpha)\pi_{t-j}(\beta)| \leq c \sum_{j=1}^{t-1} j^{\alpha-1}(t-j)^{\beta-1}.$$

For $\alpha > 0$ and $\beta > 0$ we normalize the product moment and find

$$t^{-1} \sum_{j=1}^{t-1} (\frac{j}{t})^{\alpha-1}(1 - \frac{j}{t})^{\beta-1} \to \int_0^1 u^{\alpha-1}(1-u)^{1-\beta}du = B(\alpha, \beta) \text{ as } t \to \infty,$$

where $B(\alpha, \beta)$ is the Beta function, which proves the first result in (48).

Next for $\alpha < 0$ or $\beta < 0$ we assume, by symmetry and without loss of generality, that $\beta \leq \alpha$ and $\beta < 0$, so that $\max(\alpha, \beta) = \alpha$, and split up according to $\alpha \geq 1$ or $\alpha < 1$. First, if $\alpha \geq 1$ then $j^{\alpha-1}$ is non-decreasing, $j^{\alpha-1} \leq t^{\alpha-1}$, and $\sum_{j=1}^{t-1}(t-j)^{\beta-1} \leq c$, so that

$$\sum_{j=1}^{t-1} j^{\alpha-1}(t-j)^{\beta-1} \leq ct^{\alpha-1},$$

Next consider $(0 \neq)\alpha < 1$, in which case $j^{\alpha-1}$ is decreasing and $(t-j)^{\beta-1}$ is increasing so that

$$\sum_{j=1}^{t-1} j^{\alpha-1}(t-j)^{\beta-1} \leq c \sum_{j \leq t/2} j^{\alpha-1}(\frac{t}{2})^{\beta-1} + c \sum_{j > t/2} (\frac{t}{2})^{\alpha-1}(t-j)^{\beta-1}$$
$$\leq ct^{\alpha^+}t^{\beta-1} + ct^{\alpha-1} \leq ct^{\alpha-1},$$

where the last inequality follows because $\beta \leq \alpha$ and $\beta < 0$. This completes the proof of (48).

Proof of (49): First take $\alpha \leq 1$ and $\beta \leq 1$, where we use $(j - i) \geq (j - \max(i, k))$ and $(j - k) \geq (j - \max(i, k))$ so that

$$\sum_{j=\max(i,k)+1}^{t-1} (j-i)^{\alpha-1}(j-k)^{\beta-1} \leq \sum_{j=\max(i,k)+1}^{t-1} (j - \max(i, k))^{\alpha+\beta-2} \leq c(\log t)t^{(\alpha+\beta-1)^+}.$$

Next let $\alpha > 1$ and $\beta > 1$, then $(j - i)^{\alpha-1} \leq t^{\alpha-1}$, $(j - k)^{\beta-1} \leq t^{\beta-1}$, and

$$\sum_{j=\max(i,k)+1}^{t-1} (j-i)^{\alpha-1}(j-k)^{\beta-1} \leq ct^{\alpha+\beta-1} \leq ct^{\alpha^++\beta^+-1}.$$

Finally for $\alpha > 1$ and $\beta \leq 1$ we use $(j - i)^{\alpha-1} \leq t^{\alpha-1}$ and $(j - k) \geq (j - \max(i, k))$, and find

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

The result for $\alpha \leq 1$ and $\beta > 1$ follows by symmetry. ∎

Let now D^m denote derivative(s) with respect to u and/or v.

Lemma 17 *For $u \geq u_0 > -1$ we have*

$$|\mathsf{D}^m \pi_j(u)| \leq c(u_0)(\log j)^m |\pi_j(u)|, \tag{50}$$

$$|\mathsf{D}^m T^{-u} \pi_j(u)| \leq c(u_0) T^{-u} |(\log \frac{j}{T})^m \pi_j(u)| \leq c(u_0)(\log T)^m T^{-u} |\pi_j(u)|. \tag{51}$$

For $-1 < v_0 \leq v < u$ we have

$$|\mathsf{D}^m \pi_j(u) - \mathsf{D}^m \pi_j(v)| \leq c(v_0)(u - v)(\log j)^{m+1} |\pi_j(u^*)|, \ u^* \in [v, u], \tag{52}$$

$$|\mathsf{D}^m T^{-u} \pi_j(u) - \mathsf{D}^m T^{-v} \pi_j(v)| \leq c(v_0)(u - v)(\log T)^{m+1} T^{-u^*} |\pi_j(u^*)|, \ u^* \in [v, u] \tag{53}$$

Proof. *Proof of (50):* For $\pi_j(u) = (-1)^j \binom{-u}{j} = \Gamma(u + j)/(\Gamma(u)\Gamma(j + 1))$ and $\psi(u) = \mathsf{D} \log \Gamma(u)$ we find for $m = 1$,

$$|\mathsf{D}\pi_j(u)| = |\pi_j(u)||\psi(u + j) - \psi(u)| = |\pi_j(u)||\sum_{i=0}^{j-1} \frac{1}{u + i}|,$$

where the second equality applies the recurrence relation $\psi(z + 1) - \psi(z) = z^{-1}$. Since $\pi_j(u) = u(u+1)\cdots(u+j-1)/j!$, the $i = 0$ term becomes $|\pi_j(u)|/|u| = |\pi_j(u+1)/(u+j)|$ so that

$$|\mathsf{D}\pi_j(u)| = |\frac{\pi_j(u + 1)}{(u + j)}| + |\pi_j(u)||\sum_{i=1}^{j-1} \frac{1}{u + i}| \leq c(u_0)|\pi_j(u)| \log j.$$

Further differentiation shows the derivatives are dominated by the term $c(u_0)(\log j)^m |\pi_j(u)|$.

Proof of (51): For $m = 1$ we find

$$
\begin{aligned}
|\mathsf{D}T^{-u}\pi_j(u)| &= T^{-u}|\pi_j(u)||\psi(u + j) - \psi(u) - \log T| \tag{54} \\
&= T^{-u}|\pi_j(u)||\sum_{i=0}^{j-1} \frac{1}{u + i} - \log T| \\
&\leq c(u_0)T^{-u}|\pi_j(u)||\log \frac{j}{T}| \leq c(u_0)(\log T)T^{-u}|\pi_j(u)|.
\end{aligned}
$$

Further differentiation shows that the bound is given by (51).

Proof of (52) and (53): These results follow from the mean value theorem using (50) and (51). ∎

B Variation bounds

In this appendix we prove a series of lemmas containing variation bounds of the type $||V_{ut}||_2 \leq c$ and $||V_{ut} - V_{vt}||_2 \leq c(u - v)$, which we shall use to verify condition (23) in Lemma 5 for relevant processes and product moments. The first lemma covers

the deterministic terms, the second the nonstationary processes, the third lemma deals with the (asymptotically) stationary processes, and the fourth lemma concerns product moments including both stationary and nonstationary processes.

The next lemma evaluates the influence of the initial values on (derivatives of) the differenced process $\Delta^{d+ib}X_t$, as given by the terms

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

and $\Delta_-^{d+ib}X_t$, see (12) and (13). A general form of such terms is $G_+(L)\mathsf{D}^m\Delta_+^{-u}\Delta_-^v X_t$ for $G_+(L) = -(\gamma_0 + \Delta_+^{b_0}F_+(L))$ or $G_+(L) = 1$ and various values of u and v. We let D^m, $m = 0, 1, 2$, denote derivatives with respect to the arguments u and/or v.

Lemma 18 *Let* $G_+(L)X_t = \sum_{n=0}^{t-1}g_n X_{t-n}$, *where* $\sum_{n=0}^{\infty}|g_n| < \infty$. *The initial values satisfy the relation*

$$\Delta_+^{-u}\Delta_-^v X_t = \sum_{n=0}^{\infty}[\sum_{j=0}^{t-1}\pi_j(u)\pi_{n+t-j}(-v)]X_{-n}^0. \tag{55}$$

If $\max_{n\geq 0}|X_{-n}^0| < \infty$ *and* $0 < \delta \leq v$, *then*

$$\left|\frac{\partial^{k+m}}{\partial u^k \partial v^m}\Delta_+^{-u}\Delta_-^v X_t\right| \leq c\begin{cases} t^{u-v+2\delta}, & u+\delta \geq 0 \text{ and } v-\delta \leq 1, \\ t^{\max(u+\delta, 1-v+\delta)-1}, & u+\delta < 0 \text{ or } v-\delta > 1. \end{cases} \tag{56}$$

It follows that for any positive η,

$$\sup_{v\geq\eta}|\mathsf{D}^m\Delta_-^v X_t| \to 0 \text{ as } t \to \infty, \tag{57}$$

$$\max_{\eta\leq v\leq d_0-1/2-\eta}\max_{1\leq t\leq T}T^{-d_0+1/2}|\mathsf{D}^m T^v\Delta_-^v X_t| \to 0 \text{ as } T \to \infty, \tag{58}$$

$$\sup_{u\leq 1/2-\eta}\sup_{v\geq 1/2+\eta}|G_+(L)\frac{\partial^{k+m}}{\partial u^k \partial v^m}\Delta_+^{-u}\Delta_-^v X_t| \to 0 \text{ as } t \to \infty, \tag{59}$$

$$\sup_{u\geq 1/2+\eta}\sup_{v\geq 1/2+\eta}\max_{1\leq t\leq T}|G_+(L)\frac{\partial^{k+m}}{\partial u^k \partial v^m}T^{-u+1/2}\Delta_+^{-u}\Delta_-^v X_t| \to 0 \text{ as } T \to \infty. \tag{60}$$

Proof. *Proof of (55):* We find

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

Proof of (56): We apply the inequalities $|X^0_{-n}| \le c$ and $|\mathsf{D}^m \pi_j(u)| \le c(\log j)^m j^{u-1}$, see (50), and find from (55) that

$$
|\frac{\partial^{k+m}}{\partial u^k \partial v^m} \Delta_+^{-u} \Delta_-^v X_t| \le c \sum_{n=0}^{\infty} \sum_{j=1}^{t-1} (\log j)^k (\log(n+t-j))^m j^{u-1}(n+t-j)^{-v-1}
$$

$$
\le c \sum_{n=0}^{\infty} \sum_{j=1}^{t-1} j^{u+\delta-1}(n+t-j)^{-v+\delta-1},
$$

where we have used that $\max_j(\log j)^k j^{-\delta} < \infty$ and $\max_j(\log j)^m j^{-\delta} < \infty$.

We then use that $\sum_{n=0}^{\infty}(n+t-j)^{-v+\delta-1} = \sum_{n=t-j}^{\infty} n^{-v+\delta-1} \le c(t-j)^{-v+\delta}$ when $v - \delta \ge 0$, so that the bound becomes $c\sum_{j=1}^{t-1} j^{u+\delta-1}(t-j)^{-v+\delta}$. The result follows if we apply (48) of Lemma 16 with $\alpha = u + \delta$ and $\beta = 1 - v + \delta$.

Proof of (57): We find from (56) with $u = 0$, $k = 0$, and $\delta = \eta/3$, so that $v \ge \delta$, that

$$
|\mathsf{D}^m \Delta_-^v X_t| \le c \left\{ \begin{array}{ll} t^{-v+2\delta}, & v - \delta \le 1 \\ t^{\max(\delta, 1-v+\delta)-1}, & v - \delta > 1 \end{array} \right\} \le c \left\{ \begin{array}{ll} t^{-\delta}, & v - \delta \le 1, \\ t^{\delta-1}, & v - \delta > 1, \end{array} \right.
$$

which tends to zero uniformly in $v \ge \eta$ and thus proves (57).

Proof of (58): For $\eta \le v \le d_0 - 1/2 - \eta$ we find, using (57), that for $\delta = \eta/3$, we get

$$
T^{-d_0+1/2}|\mathsf{D}^m T^v \Delta_-^v X_t| = T^{-d_0+1/2+v}|\sum_{k=0}^{m} \binom{m}{k} (\log T)^k \mathsf{D}^{m-k} \Delta_-^v X_t|
$$

$$
\le T^{-\eta} \sum_{k=0}^{m} \binom{m}{k} (\log T)^k \sup_{v \ge \delta} |\mathsf{D}^{m-k} \Delta_-^v X_t|
$$

$$
\le cT^{-\eta} \sum_{k=0}^{m} \binom{m}{k} (\log T)^k \to 0,
$$

uniformly in $v \in [\eta, d_0 - 1/2 - \eta]$.

Proof of (59): We define $\delta = \eta/3$, and apply (56) to find that

$$
|G_+(L)\frac{\partial^{k+m}}{\partial u^k \partial v^m}\Delta_+^{-u}\Delta_-^v X_t|
$$

$$
\le \sum_{n=0}^{t-1} |g_n||\frac{\partial^{k+m}}{\partial u^k \partial v^m}\Delta_+^{-u}\Delta_-^v X_{t-n}|
$$

$$
\le \sum_{n=0}^{t-1} |g_n|c \left\{ \begin{array}{ll} t^{u-v+2\delta}, & u + \delta \ge 0 \text{ and } v - \delta \le 1, \\ t^{\max(u+\delta, 1-v+\delta)-1}, & u + \delta < 0 \text{ or } v - \delta > 1. \end{array} \right.
$$

Therefore, when $u \le 1/2-\eta$ and $v \ge 1/2+\eta$, we find $u-v+2\delta \le -2\eta+2\delta = -4\eta/3 < 0$ and

$$
\max(u + \delta, 1 - v + \delta) - 1 \le \max(1/2 - \eta + \delta, 1/2 - \eta + \delta) - 1 = -1/2 - 2\eta/3 < 0,
$$

so by application of the dominated convergence theorem and $\sum_{n=0}^{\infty} |g_n| < \infty$, it follows that $|G_+(L)\frac{\partial^{k+m}}{\partial u^k \partial v^m}\Delta_+^{-u}\Delta_-^v X_t| \to 0$ as $t \to \infty$ uniformly in $u \le 1/2 - \eta$ and $v \ge 1/2 + \eta$.

Proof of (60): We find

$$|G_+(L)\frac{\partial^{k+m}}{\partial u^k \partial v^m}T^{-u+1/2}\Delta_+^{-u}\Delta_-^v X_t| \le c\sum_{n=0}^{t-1}|g_n||\frac{\partial^{k+m}}{\partial u^k \partial v^m}T^{-u+1/2}\Delta_+^{-u}\Delta_-^v X_{t-n}|,$$

where

$$|T^{1/2}[\frac{\partial^k}{\partial u^k}T^{-u}\Delta_+^{-u}][\frac{\partial^m}{\partial v^m}\Delta_-^v X_t]|$$

$$\le T^{1/2-u}|\sum_{i=0}^{k}(-1)^i\binom{k}{i}(\log T)^i\frac{\partial^{k-i+m}}{\partial u^{k-i}\partial v^m}\Delta_+^{-u}\Delta_-^v X_t|$$

$$\le (\log T)^k T^{1/2-u}\sum_{i=0}^{k}\binom{k}{i}|\frac{\partial^{k-i+m}}{\partial u^{k-i}\partial v^m}\Delta_+^{-u}\Delta_-^v X_t|$$

$$\le c(\log T)^k T^{1/2-u}\begin{cases} t^{u-v+2\delta}, & v-\delta \le 1, \\ t^{\max(u+\delta,1-v+\delta)-1}, & v-\delta > 1. \end{cases}$$

by application of (56) with $\delta = \eta/3$. Thus, when $v - \delta \le 1$ we find the bound

$$c(\log T)^k T^{1/2-u}t^{u-v+2\delta} \le c(\log T)^k\left\{ \begin{array}{cc} T^{-\eta}, & u-v+2\delta \le 0 \\ T^{-\eta+2\delta} = T^{-\eta/3}, & u-v+2\delta > 0 \end{array} \right\} \to 0.$$

When $v - \delta > 1$ we find the bound

$$c(\log T)^k T^{1/2-u}t^{\max(u+\delta,1-v+\delta)-1} \le c(\log T)^k\left\{ \begin{array}{cc} T^{-\eta}, & \max(u+\delta,1-v+\delta) \le 1 \\ T^{\delta-1/2}, & \max(u+\delta,1-v+\delta) > 1 \end{array} \right\} \to 0$$

because

$$1/2 - u + \max(u+\delta,1-v+\delta) - 1 = -1/2 + \delta + \max(0,1-v-u) = -1/2 + \eta/3 < 0.$$

∎

Lemma 19 *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 $v_0 > 1/2$ and $m = 0,1,2$ it holds that*

$$||\mathsf{D}^m T^{-u+1/2}\Delta_+^{-u}Z_t^+||_2 \le c(v_0), \tag{61}$$

$$||\mathsf{D}^m T^{-u+1/2}\Delta_+^{-u}Z_t^+ - \mathsf{D}^m T^{-v+1/2}\Delta_+^{-v}Z_t^+||_2 \le c(v_0)(u-v), \tag{62}$$

uniformly in $u > v \ge v_0$.

Proof. We first note the evaluation

$$|Cov(Z_t^+, Z_{t+h}^+)| = \sigma^2 |\sum_{n=0}^{t-1-|h|} \xi_n \xi_{n+|h|}| \le \phi_Z(h),$$

so that

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

We apply the inequality (51) and find

$$
\begin{aligned}
&Var(\mathsf{D}^m T^{-u+1/2} \Delta_+^{-u} Z_t^+) \\
&\le c \sum_{i=1}^{t-1} \sum_{j=1}^{t-1} |T^{-u+1/2} \pi_i(u) \log(\frac{i}{T})^m||T^{-u+1/2} \pi_j(u) \log(\frac{j}{T})^m|\phi_Z(i-j) \\
&\le cT^{-2u+1} \sum_{h=0}^{t-1} \phi_Z(h) \sum_{j=1}^{t-1-h} (j+h)^{u-1} j^{u-1} |\log(\frac{j+h}{T})|^m |\log(\frac{j}{T})|^m.
\end{aligned}
$$

Now we evaluate $(j+h)^{u-1} \le j^{u-1}$ if $u < 1$ and $(j+h)^{u-1} \le (j+T)^{u-1}$ if $u \ge 1$, that is

$$(j+h)^{u-1} \le (j + T1_{\{u \ge 1\}})^{u-1}.$$

Then we find, because $\sum_{h=0}^{\infty} \phi_Z(h) < \infty$, that

$$
\begin{aligned}
Var(\mathsf{D}^m T^{-u+1/2} \Delta_+^{-u} Z_t) &\le cT^{-1} \sum_{j=1}^{T-1-h} (\frac{j}{T} + 1_{\{u \ge 1\}})^{u-1} (\frac{j}{T})^{u-1} |\log(\frac{j+T}{T})|^m |\log(\frac{j}{T})|^m \\
&\to c \int_0^1 (x + 1_{\{u \ge 1\}})^{u-1} x^{u-1} |\log(x+1)|^m |\log x|^m dx,
\end{aligned}
$$

which is integrable uniformly in $u \ge v_0 > 1/2$ because $x^{u-1}|\log x|^m$ is integrable when $u > 0$.

To prove (62), we apply the inequality (52) and then use the same proof. ■

Lemma 20 *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 $u_0 < 1/2$ and $m = 0, 1, 2$ it holds that*

$$
\begin{aligned}
||\mathsf{D}^m \Delta_+^{-u} Z_t^+||_2 &\le c(u_0), & (63) \\
||\mathsf{D}^m \Delta_+^{-u} Z_t^+ - \mathsf{D}^m \Delta_+^{-v} Z_t^+||_2 &\le c(u_0)(u-v), & (64)
\end{aligned}
$$

uniformly in $v < u \le u_0$.

Proof. We prove the results for $m = 0$. The results for the derivatives follow in the same way, using the evaluation (50). We find as in the proof of Lemma 19 the inequality

$$Var(\Delta_+^{-u}Z_t^+) \leq c \sum_{h=0}^{t-1} \phi_Z(h) \sum_{j=1}^{t-1-|h|} j^{u-1}(j+|h|)^{u-1} \leq c \sum_{j=1}^{t-1} j^{2u_0-2} \leq c(u_0),$$

when $u_0 < 1/2$, which gives (63) because $\sum_{h=0}^{\infty} \phi_Z(h) < \infty$. For $Var((\Delta_+^{-u} - \Delta_+^{-v})Z_t^+)$ we apply the inequality (52), and then use the same proof. ■

Lemma 21 *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* $Z_t^+ = \sum_{n=0}^{t-1} \xi_n \varepsilon_{t-n}$. *Then for* $v < 1/2 < u$ *we have*

$$T^{-u-1/2} \sum_{t=1}^{T} \Delta_+^{-u} Z_t^+ \Delta_+^{-v} Z_t^+ \xrightarrow{P} 0.$$

Proof. We show convergence in mean square. Let

$$I_T = \sum_{t=1}^{T} \Delta_+^{-u} Z_t^+ \Delta_+^{-v} Z_t^+ = \sum_{t=1}^{T} \sum_{i=0}^{t-1} \sum_{j=0}^{t-1} \pi_i(u) \pi_j(v) Z_{t-i}^+ Z_{t-j}^+.$$

The second moment of I_T is

$$
\begin{aligned}
V &= E(I_T^2) = \sum_{t=1}^{T} \sum_{i=1}^{t} \sum_{j=1}^{t} \pi_{t-i}(u) \pi_{t-j}(v) \sum_{s=1}^{T} \sum_{k=1}^{s} \sum_{l=1}^{s} \pi_{s-k}(u) \pi_{s-l}(v) E(Z_i^+ Z_k^+ Z_j^+ Z_l^+) \\
&\leq \sum_{i=1}^{T} \sum_{j=1}^{T} \sum_{t=\max(i,j)}^{T} \pi_{t-i}(u) \pi_{t-j}(v) \sum_{k=1}^{T} \sum_{l=1}^{T} \sum_{s=\max(k,l)}^{T} \pi_{s-k}(u) \pi_{s-l}(v) E(Z_i^+ Z_k^+ Z_j^+ Z_l^+) \\
&\leq \xi(T,u,v)^2 \sum_{i=1}^{T} \sum_{j=1}^{T} \sum_{k=1}^{T} \sum_{l=1}^{T} |E(Z_i^+ Z_k^+ Z_j^+ Z_l^+)|,
\end{aligned}
$$

see Lemma 16 for the definition of $\xi(T,u,v)$. We want to prove that $T^{-2u-1}V \to 0$ as $T \to \infty$.

Now,

$$
\begin{aligned}
E(Z_i^+ Z_k^+ Z_j^+ Z_l^+) &\leq \sigma_0^4 (\phi_Z(i-k)\phi_Z(j-l) + \phi_Z(i-j)\phi_Z(k-l) + \phi_Z(i-l)\phi_Z(j-k)) \\
&\quad + \kappa_4(\varepsilon) \sum_{n=0}^{\infty} \xi_{i-n} \xi_{k-n} \xi_{j-n} \xi_{l-n},
\end{aligned}
$$

see Anderson (1971, p. 467) for the case of stationary processes. Summing over $1 \leq (i,j,k,l) \leq T$ we find the bound

$$cT^2 \sigma_0^4 (\sum_{h=0}^{\infty} |\phi_Z(h)|)^2 + cT|\kappa_4(\varepsilon)|(\sum_{n=0}^{\infty} |\xi_n|)^4 \leq cT^2(\sum_{n=0}^{\infty} |\xi_n|)^4.$$

Thus we find from (49) in Lemma 16 that

$$T^{-2u-1}V \leq cT^{-2u+1}\xi(T,u,v)^2 \leq c(\log T) \begin{cases} T^{(u+v-1)^+-2u+1}, & u \leq 1 \text{ and } v \leq 1, \\ T^{v^+-u}, & u > 1 \text{ or } v > 1, \end{cases}$$

which tends to zero because $v < 1/2 < u$ implies that $v^+ - u < 0$ and

$$(u+v-1)^+ + 1 - 2u = \begin{cases} v - u < 0, & u + v - 1 \geq 0, \\ 1 - 2u < 0, & u + v - 1 < 0. \end{cases}$$

∎

C   Proof of Theorem 6

The derivatives of the likelihood function with respect to the various parameters are functions of $(\mathcal{A}_T(\psi), \mathcal{B}_T(\psi), \mathcal{C}_T(\psi))$, see (15), (16), and (17), and their derivatives with respect to d and b, which again are functions of the normalized product moments of the processes $\Delta^{d+ib}X_t$, $i = -1, 0, \ldots, k$, and their derivatives, and we discuss the properties of these processes below.

We prove tightness by showing that Lemma 5 is satisfied for the product moments entering $\mathcal{A}_T$, $\mathcal{B}_T$, and $\mathcal{C}_T$ and their derivatives, using conditions (23) and (25). Then we derive the limits of each of the product moments $\mathcal{A}_T$, $\mathcal{B}_T$, and $\mathcal{C}_T$ and the relevant derivatives.

C.1   Tightness of product moments

Lemma 22 *Under Assumption 1 for model (5) the product moments $\mathcal{A}_T(\psi)$, $\mathcal{B}_T(\psi)$, $\mathcal{C}_T(\psi)$, and their derivatives are tight on $C(N_1)$, where the compact set N_1 is defined as*

$$N_1 = \{(b,d) : d_1 \leq d \leq d_2, \ b \geq 1/2 + \delta, \ \eta \leq d - b \leq d_0 - 1/2 - \eta\},$$

with $d_1 = \max(1/2, d_0 - 1/2) + \eta$, $d_2 > d_0$, and $0 < \eta < \min(1/2, b_0 - 1/2, d_0 - b_0)$.

Under Assumption 2 for model (2) the product moments $\mathcal{A}_T(d)$, $\mathcal{B}_T(d)$, $\mathcal{C}_T(d)$, and their derivatives are tight on $C(I_1)$, where $I_1 = [d_1, d_2]$ with $d_1 = \max(1/2, d_0 - 1/2) + \eta$, $d_2 > d_0$, and $0 < \eta < \min(1/2, d_0 - 1/2)$.

Proof. We give the proof for model (5) only. The same proof can be applied for model (2). For $\Delta^{d+ib}X_t$, $i = 0, 1, \ldots, k$, we have the representation, see (12),

$$\Delta^{d+ib}X_t = \Delta_+^{d+ib-d_0}(\gamma_0\varepsilon_t + \Delta_+^{b_0}Y_t^+) + \Delta_+^{d+ib}\mu_t^0 + \Delta_-^{d+ib}X_t, \ i = -1, 0, 1, \ldots, k. \quad (65)$$

For $i = 0, 1, \ldots, k$ and $(b,d) \in N_1$, condition (23) holds for the process $\mathsf{D}^m\Delta_+^{d+ib-d_0}(\gamma_0\varepsilon_t + \Delta_+^{b_0}Y_t^+)$ by (63) and (64) of Lemma 20 with $u = -d - ib + d_0 \leq d_0 - d \leq 1/2 - \eta$ and $Z_t^+ = \gamma_0\varepsilon_t + \Delta_+^{b_0}Y_t^+ = \sum_{n=0}^{t-1}\xi_n\varepsilon_{t-n}$ which satisfies $\sum_{h=0}^{\infty}|\xi_n| < \infty$, see Lemma 1.

We next show that the deterministic parts of the process, $\Delta_+^{d+ib}\mu_t^0$ and $\Delta_-^{d+ib}X_t$, satisfy condition (25). We find

$$\max_{(b,d)\in N_1} |\mathsf{D}^m\Delta_+^{d+ib}\mu_t^0| = \max_{d_1\leq d+ib\leq(1+k)d_2} |(\gamma_0 + \Delta_+^{b_0}F_+(L))\sum_{j=0}^{k}\rho_j\mathsf{D}^m\Delta_+^{d+ib-d_0}\Delta_-^{d_0+jb_0}X_t|,$$
(66)

which tends to zero by applying (59) with the choices $G_+(L) = (\gamma_0 + \Delta_+^{b_0}F_+(L))\rho_j$, $u = -d - ib + d_0 \leq -d + d_0 \leq 1/2 - \eta$, and $v = d_0 + jb_0 \geq d_0 > d_1 \geq 1/2 + \eta$ for $j = 0,\ldots,k$ (noting that (d_0, b_0) is an interior point in N_1 and therefore $d_0 > d_1$). Next we have

$$\max_{d_1\leq d+ib\leq(1+k)d_2} |\mathsf{D}^m\Delta_-^{d+ib}X_t|,$$

which tends to zero by (57) with the choice $v = d + ib \geq d \geq d_1 \geq 1/2 + \eta$.

For $i = -1$ we apply Lemma 5 for the process $\mathsf{D}^m T^{d-b-d_0+1/2}\Delta_-^{d-b}X_t$, see (65), and verify conditions (23) and (25). Condition (23) holds for $T^{d-b-d_0+1/2}\Delta_+^{d-b-d_0}(\gamma_0\varepsilon_t + \Delta_+^{b_0}Y_t^+)$ and its derivatives by Lemma 19 with $Z_t^+ = \gamma_0\varepsilon_t + \Delta_+^{b_0}Y_t^+$ and $u = d_0 - d + b \geq 1/2 + \eta$.

For the deterministic part we show condition (25). From (66) we see that the first term, $D^m T^{d-b-d_0+1/2}\Delta_+^{d-b}\mu_t^0$, is composed of terms of the form

$$(\gamma_0 + \Delta_+^{b_0}F_+(L))\mathsf{D}^m T^{d-b-d_0+1/2}\Delta_+^{d-b-d_0}\Delta_-^{d_0+jb_0}X_t,$$

which are investigated in (60). We take $G_+(L) = \gamma_0 + \Delta_+^{b_0}F_+(L)$, $u = -d + b + d_0 \geq 1/2 + \eta$, and $v = d_0 + jb_0 \geq d_0 > d_1 \geq 1/2 + \eta$, and find that condition (25) is satisfied. For the term $D^m T^{d-b-d_0+1/2}\Delta_-^{d-b}X_t$ we apply (58) with $v = d - b$ which satisfies $\eta \leq v \leq d_0 - 1/2 - \eta$. ∎

C.2 The product moment $\mathcal{B}_T$

These product moments, see (16), involve the processes $\Delta^d L_b^i X_t$, $i = 0, 1, \ldots, k$, which are linear combinations of the processes $\Delta^{d+ib}X_t$, $i = 0, 1, \ldots, k$, for which we have the representation, see (65),

$$\begin{aligned} \Delta^{d+ib}X_t &= S_{it}^+ + D_{1it}, \ i = 0, 1, \ldots, k, & (67)\\ S_{it}^+ &= \Delta_+^{d+ib-d_0}(\gamma_0\varepsilon_t + \Delta_+^{b_0}Y_t^+),\\ D_{1it} &= \Delta_+^{d+ib}\mu_t^0 + \Delta_-^{d+ib}X_t \end{aligned}$$

where S_{it}^+ is asymptotically stationary. In the next lemma, let D^m denote derivatives with respect to $\psi = (d, b)'$.

Lemma 23 *Under Assumption 1 for model (5) the representation (67) implies that*

$$\max_{d_1\leq d+ib\leq(k+1)d_2} |\mathsf{D}^m D_{1it}| \to 0 \ as \ t \to \infty,$$
(68)

$$\mathsf{D}^m\mathcal{B}_{ijT}(\psi) = T^{-1}\mathsf{D}^m\sum_{t=1}^{T}\Delta^{d+ib}X_t\Delta^{d+jb}X_t \xrightarrow{P} \mathsf{D}^m\mathcal{B}_{ij}(\psi) \ as \ T \to \infty,$$
(69)

when $(d, b) \in N_1$. If Assumption 2 holds for model (2) then the same conclusions hold with $d = b$ and $d \in I_1$.

Proof. *Proof of (68):* From Lemma 18 we see that we can apply (57) with $v = d + ib \geq d \geq d_1 \geq 1/2 + \eta$ to show that $\max_{d_1 \leq d+ib \leq (k+1)d_2} |\mathsf{D}^m \Delta_-^{d+ib} X_t| \leq \sup_{v \geq \eta} |\mathsf{D}^m \Delta_-^v X_t| \to 0$, and (59) with $u = d_0 - d - ib \leq 1/2 - \eta$ and $v = d_0 + jb_0 \geq d_0 > d_1 \geq 1/2 + \delta$ for $j = 0, \ldots, k$ to show that

$$\max_{d_1 \leq d+ib \leq (k+1)d_2} |\Delta_+^{d+ib} \mu_t^0| \leq \sup_{u \leq 1/2 - \eta} \sup_{v \geq 1/2 + \eta} |(\gamma_0 + \Delta_+^{b_0} F_+(L)) \sum_{j=0}^{k} \rho_j \Delta_+^{-u} \Delta_-^v X_t| \to 0.$$

Proof of (69): The process $S_{it}^+ = \sum_{n=0}^{t-1} \tau_{in} \varepsilon_{t-n}$ is asymptotically stationary in the sense that $Var(S_{it}^-) = Var(\sum_{n=t}^{\infty} \tau_{in} \varepsilon_{t-n}) = \sigma_0^2 \sum_{n=t}^{\infty} \tau_{in}^2 \to 0$ as $t \to \infty$. Let $S_{it} = \sum_{n=0}^{\infty} \tau_{in} \varepsilon_{t-n}$. From the law of large numbers we find

$$T^{-1} \sum_{t=1}^{T} S_{it} S_{jt} \xrightarrow{P} E(S_{it} S_{jt}).$$

It follows from $S_{it}^+ = S_{it} - S_{it}^-$ that also $T^{-1} \sum_{t=1}^{T} S_{it}^+ S_{jt}^+ \xrightarrow{P} E(S_{it} S_{jt})$, if $T^{-1} \sum_{t=1}^{T} (S_{it}^-)^2 \xrightarrow{P} 0$. But this is a consequence of

$$E(T^{-1} \sum_{t=1}^{T} (S_{it}^-)^2) = T^{-1} \sum_{t=1}^{T} (\sum_{n=t}^{\infty} \tau_{in}^2) \to 0.$$

The result (69) now follows from (68) using (67). The derivatives give rise to an extra factor $(\log T)^m$ which does not change the results, see (51). ∎

The results of Lemma 23 hold jointly for finitely many values of ψ in N_1 and we have shown tightness in Lemma 22, which proves (27) in Theorem 6.

C.3 Some moment relations for $\mathcal{B}(\psi)$

In this subsection we denote by D the $2 + k$ vector of derivatives with respect to the parameters ψ and ϕ. Similarly D^2 is the matrix of second derivatives. We define

$$\varepsilon_t(\psi, \phi) = \Delta^d X_t - \sum_{i=1}^{k} \phi_i \Delta^d L_b^i X_t, \tag{70}$$

$$\sigma^2(\psi, \phi) = \lim_{t \to \infty} E(\varepsilon_t(\psi, \phi)^2) = \mathcal{B}_{00}(\psi) - 2\phi' \mathcal{B}_{*0}(\psi) + \phi' \mathcal{B}_{**}(\psi)\phi, \tag{71}$$

see (69), and the $(2 + k) \times (2 + k)$ positive semidefinite matrix which enters the asymptotic distribution of the estimators $\hat{\psi}$ and $\hat{\phi}$:

$$\Sigma(\psi, \phi) = \begin{pmatrix} \Sigma_{00}(\psi, \phi) & \Sigma_{0*}(\psi, \phi) \\ \Sigma_{*0}(\psi, \phi) & \Sigma_{**}(\psi, \phi) \end{pmatrix} = \lim_{t \to \infty} E\left(\mathsf{D}\varepsilon_t(\psi, \phi)\mathsf{D}\varepsilon_t(\psi, \phi)'\right). \tag{72}$$

Lemma 24 *Under Assumption 1 we find for model (5) that the following identities hold*

$$\phi_0 = \mathcal{B}_{**}^{-1}(\psi_0)\mathcal{B}_{*0}(\psi_0), \tag{73}$$
$$\sigma_0^2 = \mathcal{B}_{00}(\psi_0) - 2\phi_0'\mathcal{B}_{*0}(\psi_0) + \phi_0'\mathcal{B}_{**}(\psi_0)\phi_0, \tag{74}$$

$$\mathsf{D}\sigma^2(\psi_0, \phi_0) = 0, \tag{75}$$
$$\mathsf{D}^2\sigma^2(\psi_0, \phi_0) = 2\Sigma(\psi_0, \phi_0) > 0, \ \phi_0 \neq 0. \tag{76}$$

It follows that for $\phi_0 \neq 0$, $\sigma^2(\psi, \phi)$ is strictly convex in a neighborhood of (ψ_0, ϕ_0) with a minimum at (ψ_0, ϕ_0).

Under Assumption 2 we find for model (2) that

$$\mathsf{D}^2\sigma^2(d, \phi)|_{d=d_0, \phi=\phi_0} = 2M'\Sigma(\psi_0, \phi_0)M > 0, \tag{77}$$

where D denotes derivatives with respect to $(d, \phi')'$ and $M' = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & I_k \end{pmatrix}$.

Proof. From equation (5) we find when $(d, b, \phi, \pi, \sigma^2) = (d_0, b_0, \phi_0, 0, \sigma_0^2)$ that

$$\Delta^{d_0} X_t = \sum_{i=1}^{k} \phi_{0i} \Delta^{d_0} L_{b_0}^i X_t + \varepsilon_t. \tag{78}$$

It follows in the same way as in (68) above, using Lemma 18, that the initial values have no influence on the calculation of the matrices $\mathcal{B}(\psi_0)$ and $\Sigma(\psi_0, \phi_0)$, and we therefore calculate them from the stationary processes $\Delta^{d_0} X_t = \gamma_0 \varepsilon_t + \Delta^{b_0} Y_t$ and its derivatives $\mathsf{D}^m \Delta^{d_0} X_t$. Multiplying (78) by the stationary process $\Delta^{d_0} L_{b_0}^j X_t$ and taking expectation we find

$$\mathcal{B}_{0j}(\psi_0) = \sum_{i=1}^{k} \phi_{0i} \mathcal{B}_{ij}(\psi_0),$$

which proves (73). Taking the variance in (78) we find (74).

From (70) and (78) it is seen that $\varepsilon_t(\psi_0, \phi_0) = \varepsilon_t$ and that the coefficient to ε_t in $\varepsilon_t(\psi, \phi)$ is one so that $\mathsf{D}\varepsilon_t(\psi, \phi)$ only contains lagged ε_t. We let E_{t-1} denote the conditional expectation given the past, $\mathcal{F}_{t-1} = \sigma\{X_{-n}^0, \ n \geq 0, \ \varepsilon_s, \ 1 \leq s \leq t-1\}$, and find

$$E_{t-1}(\varepsilon_t \mathsf{D}\varepsilon_t(\psi_0, \phi_0)) = 0, \ E_{t-1}(\varepsilon_t \mathsf{D}^2\varepsilon_t(\psi_0, \phi_0)) = 0, \tag{79}$$

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 (75) we differentiate (71) and find

$$\mathsf{D}\sigma^2(\psi_0, \phi_0) = 2 \lim_{t \to \infty} E(\varepsilon_t \mathsf{D}\varepsilon_t(\psi_0, \phi_0)) = 0,$$

using (79). To prove (76) we differentiate (71) twice and find, for $(\psi, \phi) = (\psi_0, \phi_0)$ and using (79), that

$$\mathsf{D}^2\sigma^2(\psi_0, \phi_0) = 2 \lim_{t \to \infty} E(\mathsf{D}\varepsilon_t(\psi_0, \phi_0)\mathsf{D}\varepsilon_t(\psi_0, \phi_0)') = 2\Sigma(\psi_0, \phi_0).$$

We next want to show that $\Sigma(\psi_0, \phi_0)$ is positive definite unless $\phi_0' = (0, 0, \ldots, 0)$.

The process $\Delta^{d_0} X_t = Z_t$, see (78), has transfer function

$$f_Z(z) = \sigma_0^2/(1 - \sum_{i=1}^{k} \phi_{0i}(1 - (1 - z)^{b_0})^i),$$

and

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

is stationary for d close to d_0 with transfer function

$$f_\varepsilon(z) = \sigma_0^2(1 - z)^{d-d_0}(1 - \sum_{i=1}^{k} \phi_i(1 - (1 - z)^b)^i)/(1 - \sum_{i=1}^{k} \phi_{0i}(1 - (1 - z)^{b_0})^i).$$

Let $\pi^*(u) = 1 - \sum_{i=1}^{k} \phi_{0i} u^i$, $\dot{\pi}^*(u) = -\sum_{i=1}^{k} i\phi_{0i} u^{i-1}$, and $u = 1 - (1 - z)^{b_0}$. The transfer function for the derivatives are

$$\frac{\partial}{\partial d} f_\varepsilon(z)|_{\psi=\psi_0,\phi=\phi_0} = \sigma_0^2 b_0^{-1} \log(1 - u),$$

$$\frac{\partial}{\partial b} f_\varepsilon(z)|_{\psi=\psi_0,\phi=\phi_0} = -\sigma_0^2 b_0^{-1}(1 - u) \log(1 - u)\dot{\pi}^*(u)/\pi^*(u),$$

$$\frac{\partial}{\partial \phi_i} f_\varepsilon(z)|_{\psi=\psi_0,\phi=\phi_0} = -\sigma_0^2 u^i/\pi^*(u).$$

If $\Sigma(\psi_0, \phi_0) = E(\mathsf{D}\varepsilon_t(\psi_0, \phi_0)\mathsf{D}\varepsilon_t(\psi_0, \phi_0)')$ is singular then there are constants $\alpha, \beta, \gamma_1, \ldots, \gamma_k$, so that

$$\alpha\frac{\partial}{\partial d}\varepsilon_t(\psi, \phi)|_{\psi=\psi_0,\phi=\phi_0} + \beta\frac{\partial}{\partial b}\varepsilon_t(\psi, \phi)|_{\psi=\psi_0,\phi=\phi_0} + \sum_{i=1}^{k} \gamma_i\frac{\partial}{\partial \phi_i}\varepsilon_t(\psi, \phi)|_{\psi=\psi_0,\phi=\phi_0} = 0.$$

The three derivatives are stationary linear processes, which are linearly dependent if and only if their transfer functions are linearly dependent. That is, the asymptotic variance matrix is singular if

$$\frac{\alpha}{b_0} \log(1 - u) + \frac{\beta}{b_0}(1 - u) \log(1 - u)\dot{\pi}^*(u)/\pi^*(u) + \sum_{i=1}^{k} \gamma_i u^i/\pi^*(u) = 0 \text{ for all } u,$$

or

$$\log(1 - u)(\alpha\pi^*(u) + \beta(1 - u)\dot{\pi}^*(u)) + b_0 \sum_{i=1}^{k} \gamma_i u^i = 0 \text{ for all } u.$$

The last term is a polynomial and the first is not, so this implies that $\gamma_i = 0$ for all i, and that

$$\alpha\pi^*(u) + \beta(1 - u)\dot{\pi}^*(u) = 0 \text{ for all } u.$$

Setting $u = 1$, we find that $\alpha \pi^*(1) = \alpha(1 - \sum_{i=1}^{k} \phi_{0i}) = 0$, so that $\alpha = 0$, and hence

$$\beta(1 - u)\dot{\pi}^*(u) = 0 \text{ for all } u.$$

This implies that either $\beta = 0$, in which case we have proved linear independence, or that

$$\dot{\pi}^*(u) = 0 \text{ for all } u,$$

which means that $\pi^*(u) = 1$, and hence $\phi'_0 = (0, 0, \ldots, 0)$.

Finally we find (77) from the relation

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

for $b = d = d_0$, and $\phi = \phi_0$. ∎

C.4 The product moment $\mathcal{A}_T$

The product moment $\mathcal{A}_T$, see (15), is

$$\mathcal{A}_T(\psi) = T^{-1} \sum_{t=1}^{T} (T^{d-b-d_0+1/2}\Delta^{d-b}L_b X_t)^2$$

and involves the process $\Delta^{d-b}L_b X_t = \Delta^{d-b}X_t - \Delta^d X_t$. Using the result for $\mathcal{B}_{00T}$ in Lemma 23 we find that $T^{-2(d_0-d+b)}\sum_{t=1}^{T}(\Delta^d X_t)^2 \implies 0$, when $(d, b) \in N_1$ because $-2(d_0 - d + b) \leq -1 - 2\eta < -1$. We therefore only consider the sum of squares of the process

$$\begin{aligned}
T^{d-b-d_0+1/2}\Delta^{d-b}X_t &= N_t^+ + M_t^+ + D_{2t}, & (80) \\
N_t^+ &= T^{d-b-d_0+1/2}\gamma_0\Delta_+^{d-b-d_0}\varepsilon_t, \\
M_t^+ &= T^{d-b-d_0+1/2}\Delta_+^{d-b-d_0}\Delta_+^{b_0}Y_t^+, \\
D_{2t} &= T^{d-b-d_0+1/2}\Delta_+^{d-b}\mu_t^0 + T^{d-b-d_0+1/2}\Delta_-^{d-b}X_t.
\end{aligned}$$

We first show that $T^{d-b-d_0+1/2}\Delta^{d-b}X_{[Tu]}$ converges in distribution on $D[0, 1]$ and then show that the product moment $\mathcal{A}_T(\psi)$ converges in distribution on $C(N_1)$.

Lemma 25 *Under Assumption 1 for model (5) we find that, as $T \to \infty$,*

$$\max_{1 \leq t \leq T} |D_{2t}| \to 0, \tag{81}$$

$$\max_{1 \leq t \leq T} |M_t^+| \xrightarrow{P} 0, \tag{82}$$

$$N_{[Tu]}^+ \implies \gamma_0 W_{d_0-d+b-1}(u), \tag{83}$$

from which it follows that

$$T^{d-b-d_0+1/2}\Delta^{d-b}X_{[Tu]} \implies \gamma_0 W_{d_0-d+b-1}(u) \text{ on } D[0, 1]$$

for fixed $(d, b) \in N_1$. If Assumption 2 holds for model (2) then the same conclusions hold with $d = b$ and $d \in I_1$.

Proof. *Proof of (81):* We apply Lemma 18. We find that the convergence of the term $\max_{1 \leq t \leq T} T^{d-b-d_0+1/2}|\Delta^{d-b}\mu_t^0| \to 0$ follows from (60) with $G_+(L) = \gamma_0 + \Delta_+^{b_0}F_+(L)$, $u = -d + b + d_0 \geq 1/2 + \eta$, and $v = d_0 + jb_0 \geq d_0 > d_1 \geq 1/2 + \eta$ for $j = 0, \ldots, k$. From (58) with $v = d - b \geq \eta$ it follows that $\max_{1 \leq t \leq T} T^{d-b-d_0+1/2}|\Delta_-^{d-b}X_t| \to 0$.

Proof of (82): We find

$$\max_{1 \leq t \leq T} |\Delta_+^{-(d_0-d+b)}\Delta_+^{b_0}Y_t^+| \leq \max_{1 \leq t \leq T} \sum_{j=0}^{t-1} |\pi_j(d_0 - d + b)||\Delta_+^{b_0}Y_{t-j}^+|$$

$$\leq c \max_{1 \leq t \leq T} |\Delta_+^{b_0}Y_t^+| \sum_{j=1}^{T} j^{d_0-d+b-1} \leq cT^{d_0-d+b} \max_{1 \leq t \leq T} |\Delta_+^{b_0}Y_t^+|.$$

Then

$$P(\max_{1 \leq t \leq T} |M_t| \geq c) \leq P(\max_{1 \leq t \leq T} |\Delta_+^{b_0}Y_t^+| \geq T^{-1/2}c)$$

$$\leq \sum_{t=1}^{T} P(|\Delta_+^{b_0}Y_t^+| \geq T^{-1/2}c) \leq T\frac{E|\Delta_+^{b_0}Y_t^+|^4}{T^2c^4}.$$

Now

$$E|\Delta_+^{b_0}Y_t^+|^4 \leq c(\sum_{n=0}^{t-1} \tau_n^2)^2 \leq c,$$

which proves (82).

Proof of (83): For the nonstationary process $N_{[Tu]} = \gamma_0\Delta_+^{-(d_0-d+b)}\varepsilon_{[Tu]}$ we can apply (7) and find the main result

$$T^{-(d_0-d+b)+1/2}\gamma_0\Delta_+^{-(d_0-d+b)}\varepsilon_{[Tu]} \implies \gamma_0 W_{d_0-d+b-1}(u).$$

∎

Lemma 26 *Under Assumption 1 for model (5) we find that for each fixed $\psi \in N_1$ it holds that*

$$\mathcal{A}_T(\psi) = T^{-2(d_0-d+b)} \sum_{t=1}^{T} (\Delta^{d-b}L_bX_t)^2 \xrightarrow{d} \gamma_0^2 \int_0^1 W_{d_0-d+b-1}^2 du.$$

If Assumption 2 holds for model (2) then the same result holds with $d = b$ and $d \in I_1$.

Proof. This follows from Lemma 25 and the continuous mapping theorem, see (8). ∎

Finally we want to prove the result (26) in Theorem 6. For a finite number of values $\psi_1, \ldots, \psi_m$ in N_1, we get joint convergence from Lemma 26, and we have shown tightness in Lemma 22. Thus we have proved (26) in Theorem 6.

C.5 The product moment $\mathcal{C}_T$

We find, for $i = 0, 1, \ldots, k$, the product moment $\mathcal{C}_{iT}(\psi)$, see (17) and (16),

$$\mathcal{C}_{iT}(\psi) = T^{-1} \sum_{t=1}^{T} (T^{-(d_0-d+b)+1/2} \Delta^{d-b} X_t)(\Delta^d L_b^i X_t) - T^{-1/2-(d_0-d+b)} \mathcal{B}_{0iT}(\psi).$$

Lemma 27 *Under Assumption 1 for model (5) we find that for each fixed $\psi \in N_1$ we have*

$$\mathsf{D}^m \mathcal{C}_{iT}(\psi) \xrightarrow{P} 0, \ i = 0, 1, \ldots, k.$$

If Assumption 2 holds for model (2) then the same result holds with $d = b$ and $d \in I_1$.

Proof. From (27) we find that $\mathsf{D}^m T^{-1/2-(d_0-d+b)} \mathcal{B}_{0iT}(\psi) \implies 0$, so we only need to prove that

$$\mathsf{D}^m T^{-1} \sum_{t=1}^{T} (T^{-(d_0-d+b)+1/2} \Delta^{d-b} X_t)(\Delta^{d+ib} X_t) \xrightarrow{P} 0.$$

For $i = 0, 1, \ldots, k$ we decompose the processes as

$$\begin{aligned}
\Delta^{d+ib} X_t &= S_{it}^+ + D_{1it}, \\
T^{d-b-d_0+1/2} \Delta^{d-b} X_t &= N_t^+ + M_t^+ + D_{2t},
\end{aligned} \tag{84}$$

see (67) and (80). For now let $m = 0$. We consider the product moment

$$T^{-1} \sum_{t=1}^{T} (N_t^+ + M_t^+ + D_{2t})(S_{it}^+ + D_{1it}).$$

Let $Z_t^+ = \gamma_0 \varepsilon_t + \Delta_+^{b_0} Y_t^+$, then

$$(N_t^+ + M_t^+) S_{it}^+ = T^{-(d_0-d+b)+1/2} (\Delta_+^{-(d_0-d+b)} Z_t^+)(\Delta_+^{d+ib-d_0} Z_t^+).$$

We apply Lemma 21 with $v = -(d+ib-d_0) \leq -(d-d_0) \leq 1/2 - \eta$, and $u = d_0 - d + b \geq 1/2 + \eta$, and find that

$$T^{-1} \sum_{t=1}^{T} (N_t^+ + M_t^+) S_{it}^+ \xrightarrow{P} 0.$$

The remaining product moments of the form $T^{-1} \sum_{t=1}^{T} A_t B_t$ are evaluated using the Cauchy-Schwarz inequality,

$$|T^{-1} \sum_{t=1}^{T} A_t B_t|^2 \leq (T^{-1} \sum_{t=1}^{T} A_t^2)(T^{-1} \sum_{t=1}^{T} B_t^2),$$

so that it is enough to show that $||A_t||_2 \to 0$, which gives $T^{-1} \sum_{t=1}^{T} A_t^2 \xrightarrow{P} 0$, and that $T^{-1} \sum_{t=1}^{T} B_t^2$ is bounded in probability. It is therefore enough to show that $||N_t^+ + M_t^+||_2$ and $||S_{it}^+||_2$ are bounded and that $T^{-1} \sum_{t=1}^{T} D_{1it}^2 \to 0$ and $T^{-1} \sum_{t=1}^{T} D_{2t}^2 \to 0$.

The result for $N_t^+ + M_t^+$ follows from (61) with $u = b - d + d_0 \geq 1/2 + \eta$ and $Z_t^+ = \gamma_0 \varepsilon_t + \Delta_+^{b_0} Y_t^+$, and for S_{it}^+ it comes from (63) with $u = -d - ib + d_0 \leq d_0 - d \leq 1/2 - \eta$. The result for D_{1it} follows from (68) and the result for D_{2t} from (81).

Finally, consider the derivatives D^m with respect to (d, b). We have from (51) that it only amounts to an extra factor $(\log T)^m$, which does not change the proof. ∎

We can now prove the result in Theorem 6 on $\mathcal{C}_T(\psi)$. Finite-dimensional convergence is obtained from Lemma 27. Tightness was proved in Lemma 22 and that proves (28) in Theorem 6.

C.6 The product moment $\mathcal{C}_{0\varepsilon T}$

Finally we investigate (for model (5) only, since the same proof can be applied for model (2)) $T^{1/2}\mathcal{C}_{0\varepsilon T}(\psi_0) = T^{-b_0} \sum_{t=1}^{T} (\Delta^{d_0 - b_0} L_{b_0} X_t)\varepsilon_t$, where

$$
\begin{aligned}
\Delta^{d_0 - b_0} L_{b_0} X_t &= \Delta^{d_0 - b_0} X_t - \Delta^{d_0} X_t = \gamma_0(\Delta_+^{-b_0}\varepsilon_t - \varepsilon_t) + Y_t^+ - \Delta_+^{b_0} Y_t^+ \\
&\quad + \Delta_+^{d_0 - b_0} \mu_t^0 - \Delta_+^{d_0} \mu_t^0 + \Delta_-^{d_0 - b_0} X_t - \Delta_-^{d_0} X_t,
\end{aligned}
$$

see (65). We decompose $\mathcal{C}_{0\varepsilon T}$ as

$$
T^{-b_0} \sum_{t=1}^{T} \gamma_0(\Delta_+^{-b_0}\varepsilon_t - \varepsilon_t)\varepsilon_t + T^{-b_0} \sum_{t=1}^{T} (Y_t^+ - \Delta_+^{b_0} Y_t^+)\varepsilon_t \tag{85}
$$

$$
+ T^{-b_0} \sum_{t=1}^{T} (\Delta_+^{d_0 - b_0} \mu_t^0 - \Delta_+^{d_0} \mu_t^0 + \Delta_-^{d_0 - b_0} X_t - \Delta_-^{d_0} X_t)\varepsilon_t. \tag{86}
$$

For the last term we find

$$
T^{-b_0} \sum_{t=1}^{T} ((\Delta_+^{d_0 - b_0} \mu_t^0 - \Delta_+^{d_0} \mu_t^0 + \Delta_-^{d_0 - b_0} X_t - \Delta_-^{d_0} X_t)\varepsilon_t \xrightarrow{P} 0,
$$

because the expectation is zero and

$$
T^{-2b_0} \sum_{t=1}^{T} (\Delta_+^{d_0 - b_0} \mu_t^0 + \Delta_-^{d_0 - b_0} X_t)^2 = T^{-1} \sum_{t=1}^{T} D_{2t}^2|_{d=d_0, b=b_0} \to 0
$$

$$
T^{-2b_0} \sum_{t=1}^{T} (\Delta_+^{d_0} \mu_t^0 + \Delta_-^{d_0} X_t)^2 = T^{-1} \sum_{t=1}^{T} D_{1it}^2|_{d=d_0, b=b_0, i=0} \to 0
$$

by (81) and (68).

The second term of (85) is

$$
T^{-b_0} \sum_{t=1}^{T} (Y_t^+ - \Delta_+^{b_0} Y_t^+)\varepsilon_t = -T^{-b_0} \sum_{t=1}^{T} \sum_{j=1}^{t-1} \pi_j(-b_0) Y_{t-j}\varepsilon_t.
$$

Because $(Y_t^+ - \Delta_+^{b_0} Y_t^+)\varepsilon_t = \sum_{j=1}^{t-1} \pi_j(-b_0) Y_{t-j}\varepsilon_t$ is a martingale difference sequence we find

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

Finally, the first term of (85) is compared with a product moment for which we have the convergence in (29), namely

$$T^{-b_0} \sum_{t=1}^{T} \gamma_0 \varepsilon_t \Delta_+^{-b_0} \varepsilon_{t-1} \xrightarrow{d} \sigma_0^2 \gamma_0 \int_0^1 B_{b_0-1} dB,$$

see Jakubowski, Mémin, and Pages (1989). We therefore show that the difference tends 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

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

Hence $Var(T^{-b_0} \sum_{t=1}^{T} \varepsilon_t(\Delta_+^{-b_0}\varepsilon_{t-1} - (\Delta_+^{-b_0}\varepsilon_t - \varepsilon_t))) \leq cT^{2(b_0-3/2)^++1-2b_0} \to 0$, which proves (29) of Theorem 6.

References

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

2. Anderson, T. W. (1971). *The Statistical Analysis of Time Series*, Wiley, New York.

3. Andrews, D. W. K. (1992). Generic uniform convergence. *Econometric Theory* **8**, pp. 241–257.

4. Artin, E. (1964). *The Gamma Function*, Springer, New York.

5. Billingsley, P. (1968). *Convergence of Probability Measures,* Wiley, New York.

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

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

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

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

10. Dolado, J. J., Gonzalo, J., and Mayoral, L. (2002). A fractional Dickey-Fuller test for unit roots. *Econometrica* **70**, pp. 1963–2006.

11. Granger, C. W. J. and Joyeux, J. (1980). An introduction to long memory time series mode ls and fractional differencing. *Journal of Time Series Analysis* **1**, pp. 15–29.

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

13. Hosking, J. R. M. (1981). Fractional differencing. *Biometrika* **68**, pp. 165–176.

14. Hualde, J. and Robinson, P. M. (2005). Gaussian pseudo-maximum likelihood estimation of fractional time series models. Working paper, University of Navarra.

15. Jakubowski, A., Mémin, J., and Pages, G. (1989). Convergence en loi des suites d'intégrales stochastiques sur l'espace $\mathbb{D}^1$ de Skorokhod. *Probability Theory and Related Fields* **81**, pp. 111–137.

16. Johansen, S. (2007). A representation theory for a class of vector autoregressive models for fractional processes. To appear in *Econometric Theory*.

17. Kallenberg, O. (2001). *Foundations of Modern Probability Theory*, 2nd edition, Springer, New York.

18. Kendall, M. G. and Stuart, A. (1977). *The Advanced Theory of Statistics,* Vol. I, 4th ed., Charles Griffin, London.

19. Ling, S. and Li, W. K. (2001). Asymptotic inference for nonstationary fractionally integrated autoregressive moving-average models. *Econometric Theory* **17**, pp. 738–764.

20. Lobato, I. N. and Velasco, C. (2006). Efficient Wald tests for fractional unit roots. *Econometrica* **75**, pp. 575–589.

21. Marinucci, D. and Robinson, P. M. (2000). Weak convergence of multivariate fractional processes. *Stochastic Processes and their Applications* **86**, pp. 103–120.

22. Newey, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica* **59**, pp. 1161–1167.

23. Nielsen, M. Ø. (2004). Efficient likelihood inference in nonstationary univariate models. *Econometric Theory* **20**, pp. 116–146.

24. Phillips, P. C. B. (1987). Time series regression with a unit root. *Econometrica* **55**, pp. 277–301.

25. Robinson, P. M. (1991). Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regressions. *Journal of Econometrics* **47**, pp. 67–84.

26. Robinson, P. M. (1994). Efficient tests of nonstationary hypotheses. *Journal of the American Statistical Association* **89**, pp. 1420–1437.

27. Sowell, F. B. (1990). The fractional unit root distribution. *Econometrica* **58**, pp. 495–505.

28. Tanaka, K. (1999). The nonstationary fractional unit root. *Econometric Theory* **15**, pp. 549–582.

29. Zygmund, A. (2003). *Trigonometric Series,* Vol. I and II, 3rd rev. ed., Cambridge University Press, Cambridge.