

Discussion Papers
Department of Economics
University of Copenhagen

No. 10-15

Likelihood Inference for a Fractionally
Cointegrated Vector Autoregressive Model

Søren Johansen, Morten Ørregaard Nielsen

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
<http://www.econ.ku.dk>

ISSN: 1601-2461 (E)

Likelihood inference for a fractionally cointegrated vector autoregressive model*

Søren Johansen[†]
University of Copenhagen
and CREATES

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

May 18, 2010

Abstract

We consider model based inference in a fractionally cointegrated (or cofractional) vector autoregressive model based on the conditional Gaussian likelihood. The model allows the process X_t to be fractional of order d and cofractional of order $d - b$; that is, there exist vectors β for which $\beta'X_t$ is fractional of order $d - b$. The parameters d and b satisfy either $d \geq b \geq 1/2$, $d = b \geq 1/2$, or $d = d_0 \geq b \geq 1/2$. Our main technical contribution is the proof of consistency of the maximum likelihood estimators on the set $1/2 \leq b \leq d \leq d_1$ for any $d_1 \geq d_0$. To this end, we consider the conditional likelihood as a stochastic process in the parameters, and prove that it converges in distribution when errors are i.i.d. with suitable moment conditions and initial values are bounded. We then prove that the estimator of β is asymptotically mixed Gaussian and estimators of the remaining parameters are asymptotically Gaussian. We also find the asymptotic distribution of the likelihood ratio test for cointegration rank, which is a functional of fractional Brownian motion of type II.

Keywords: Cofractional processes, cointegration rank, fractional cointegration, likelihood inference, vector autoregressive model.

JEL Classification: C32.

*We are grateful to Uwe Hassler, James MacKinnon, Ilya Molchanov, and seminar participants at various universities and conferences for comments, and to the Danish Social Sciences Research Council (FSE grants 2114-04-0001 to Johansen and 275-05-0220 to Nielsen), the Social Sciences and Humanities Research Council of Canada (SSHRC grant 410-2009-0183), and the Center for Research in Econometric Analysis of Time Series (CREATES, funded by the Danish National Research Foundation) for financial support. A previous version of this paper was circulated under the title “Likelihood inference for a vector autoregressive model which allows for fractional and cofractional processes.”

[†]Address: Department of Economics, University of Copenhagen, Øster Farimagsgade 5, DK-1353 Copenhagen K, Denmark, Email: Soren.Johansen@econ.ku.dk

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

1 Introduction and motivation

We consider the p -dimensional time series X_t , $t = \dots, -1, 0, 1, \dots, T$, and model X_1, \dots, X_T conditional on the (infinitely many) initial values X_{-n} , $n = 0, 1, \dots$, by the fractional vector autoregressive model, $\text{VAR}_{d,b}(k)$, in error correction representation,

$$\mathcal{H}_r : \Delta^d X_t = \alpha\beta' \Delta^{d-b} L_b X_t + \sum_{i=1}^k \Gamma_i \Delta^d L_b^i X_t + \varepsilon_t, \quad t = 1, \dots, T, \quad (1)$$

where ε_t are i.i.d. $(0, \Omega)$, Ω is positive definite, $d \geq b \geq 1/2$, and α and β are $p \times r$, $0 \leq r \leq p$. The parameters $(d, b, \alpha, \beta, \Gamma_1, \dots, \Gamma_k, \Omega)$ are otherwise unrestricted. Here Δ^b is the fractional difference operator and $L_b = 1 - \Delta^b$ the fractional lag operator. We also consider the two submodels given by $d = d_0$ and $d = b$, respectively.

Model (1) can be expressed as $\Psi(L_b) \Delta^{d-b} X_t = \varepsilon_t$, where the polynomial $\Psi(y)$ is given by

$$\Psi(y) = (1 - y)I_p - \alpha\beta' y - \sum_{i=1}^k \Gamma_i (1 - y)y^i = \sum_{i=-1}^k \Psi_i (1 - y)^{i+1} \quad (2)$$

and the coefficients satisfy $\sum_{i=-1}^k \Psi_i = I_p$, $\Psi_{-1} = -\alpha\beta'$, and $\Psi_k = (-1)^{k+1} \Gamma_k$. That is, $\Delta^{d-b} X_t$ satisfies a vector autoregression (VAR) in the lag operator L_b rather than the standard lag operator $L = L_1$. The cointegrated VAR model analyzed by Johansen (1988) appears as the special case $d = b = 1$, and the interpretation of the model parameters is similar, i.e., the columns of β are the cointegrating (cofractional) relations and α are the adjustment or loading coefficients. Note that the expansion of $L_b = 1 - \Delta^b$ has no term in L^0 and thus only lagged disequilibrium errors appear in (1).

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

$$\Pi(z) = (1 - z)^{d-b} \Psi(1 - (1 - z)^b) = \sum_{i=-1}^k \Psi_i (1 - z)^{d+ib} = \sum_{n=0}^{\infty} \Pi_n z^n, \quad |z| < 1. \quad (3)$$

Conditions on the roots of the polynomial $\Psi(y)$ are given, see Johansen (2008), for X_t determined by (1) to be fractional of order d and $\beta' X_t$ fractional of order $d - b$.

The model considered here is derived from the usual cointegrated VAR model by replacing Δ by Δ^b and $L = 1 - \Delta$ by $L_b = 1 - \Delta^b$ and applying the model to $\Delta^{d-b} X_t$. The inspiration for the model comes from Granger (1986), who noted the special role of the fractional lag operator L_b and suggested the model

$$A^*(L) \Delta^d X_t = (1 - \Delta^b) \Delta^{d-b} \alpha\beta' X_{t-1} + d(L) \varepsilon_t,$$

see also Davidson (2002). In Johansen (2008) it was suggested to replace the polynomial $A^*(L)$ in the usual lag operator by a polynomial in the fractional lag operator in order to be able to analyze the stochastic properties of the solution. The univariate version of the resulting model (1) was analyzed by Johansen and Nielsen (2010), henceforth JN (2010), and we refer to that paper for some technical results.

We are interested in testing the rank of the coefficient to $\Delta^{d-b}L_bX_t$ and in conducting inference on the parameters of model (1). Because the processes are nonstationary, $d \geq 1/2$, we analyze the conditional likelihood function for (X_1, \dots, X_T) given initial values X_{-n} , $n = 0, 1, \dots$, under the assumption that ε_t is i.i.d. $N_p(0, \Omega)$. For the asymptotic analysis we assume only that ε_t is i.i.d. $(0, \Omega)$ with suitable moment conditions and that X_{-n} is bounded. Thus, the initial values are not modeled. In particular they are not assumed to be generated by equation (1). The asymptotic results show that the influence of initial values disappears in the limit provided they are bounded, an assumption that appears reasonable in practice, and which is needed to calculate the fractional differences $\Delta^d X_t$ for $d > 0$.

We consider throughout the case $b \leq d$, so that $\Delta^{d-b}X_t$ can be calculated, and furthermore $b \geq 1/2$ which is the “strong cointegration” case in the terminology of Hualde and Robinson (2010a). This also generates non-standard asymptotic theory for inference, which is perhaps the most interesting analysis because it involves the fractional Brownian motion.

For fixed orders of fractionality, d and b , model (1) can be estimated by reduced rank regression as in Johansen (1988), and the asymptotic analysis is not too complicated. If (d, b) are not known, the problem is more challenging. By reduced rank regression, calculation of maximum likelihood estimators is reduced to a two-parameter non-linear maximum likelihood problem, which is solved by numerical optimization. In JN (2010) we derived asymptotic theory for the univariate version of model (1), including asymptotic likelihood based inference and a fractional version of the Dickey-Fuller unit root test, although there we had to exclude certain parts of the parameter space in the consistency proof. In the present paper we analyze the multivariate model, but of course our results apply to the univariate model as well and therefore also complete the analysis in JN (2010). Specifically, the main technical contribution in this paper is the proof of existence and consistency of the MLE, which allows standard likelihood theory to be applied. This involves an analysis of the influence of initial values as well as proving tightness and uniform convergence of product moments of processes that can be close to critical processes of the form $\Delta^{-1/2}\varepsilon_t$. To tackle the latter we apply a truncation argument.

An attractive feature of the vector error correction model (1) is the straightforward interpretation of its parameters, and inference on these is thus of particular interest. We prove that for i.i.d. errors with sufficient moments finite, the estimated cointegrating vectors are asymptotically mixed Gaussian (LAMN), so that standard (chi-squared) asymptotic inference can be conducted on the cointegrating relations. Thus, for Gaussian errors we get asymptotically optimal inference, but the results hold more generally.

Although such results are well known from the standard (non-fractional) cointegration model, e.g. Johansen (1988, 1991), Phillips and Hansen (1990), Phillips (1991), and Saikkonen (1991) among others, they are novel for fractional models. Only recently, asymptotically optimal inference procedures have been developed for fractional processes, e.g. Jeganathan (1999), Robinson and Hualde (2003), Lasak (2008a,b), Avarucci and Velasco (2009), and Hualde and Robinson (2010a). Specifically, in a vector autoregressive context, but in a model with $d = 1$ and a different lag structure from ours, Lasak (2008a) analyzes a test for no cointegration and in Lasak (2008b) she analyzes maximum likelihood estimation and inference; in both cases assuming “strong cointegration”. In the same model as Lasak, but assuming “weak cointegration” ($b < 1/2$), Avarucci and Velasco (2009) extend the univariate test of Lobato and Velasco (2007) to analyze a Wald test for cointegration rank, see also

Marmol and Velasco (2004). However, the present paper seems to be the first to develop LAMN results for the MLE in a fractional cointegration model in a vector error correction framework and with two fractional parameter (d and b).

The analysis of the fractionally cointegrated (or cofractional; henceforth we shall use these terms synonymously) VAR model (1) generalizes the unit root test and related inference on fractional orders in the univariate fractional autoregressive model in the same way that the cointegrated VAR in Johansen (1988) generalizes the standard Dickey-Fuller test to the multivariate case. Hence, this paper at the same time generalizes the fractional unit root or fractional Dickey-Fuller tests and in particular that of JN (2010) to the multivariate case, and it generalizes the cointegrated VAR to models for fractional time series. This has far reaching implications for empirical research, where the cointegrated VAR is probably the most widely applied model for estimating and analyzing cointegrated time series.

The remainder of the paper is laid out as follows. In the next section we describe the solution of the cofractional vector autoregressive model and its properties. In Section 3 we derive the likelihood function and estimators and discuss asymptotic properties of both, and in Section 4 we find the asymptotic properties of the likelihood ratio test for cointegration rank. Section 5 concludes and technical material is presented in appendices.

A word on notation. For a symmetric matrix A we write $A > 0$ to mean that it is positive definite. The Euclidean norm of a matrix, vector, or scalar A is denoted $|A| = (tr\{A'A\})^{1/2}$ and the determinant of a square matrix is denoted $\det(A)$. Throughout, c denotes a generic positive constant which may take different values in different places.

2 Solution of the cofractional vector autoregressive model

We discuss the fractional difference operators Δ^d and Δ_+^d and calculation of $\Delta^d X_t$. We show how equation (1) can be solved for X_t as a function of initial values, parameters, and errors $\varepsilon_i, i = 1, \dots, t$, and give properties of the solution in Theorem 3. We then give assumptions for the asymptotic analysis and discuss briefly initial values and identification of parameters.

2.1 The fractional difference operator

The fractional coefficients, $\pi_n(-a)$, are defined by the expansion

$$(1 - z)^a = \sum_{n=0}^{\infty} \pi_n(-a) z^n = \sum_{n=0}^{\infty} \binom{a}{n} z^n = \sum_{n=0}^{\infty} \frac{a(a-1)\cdots(a-n+1)}{n!} z^n$$

and satisfy the evaluation $|\pi_n(-a)| \leq cn^{-a-1}$, $n \geq 1$, see Lemma A.5. The fractional difference operator applied to a process $Z_t, t = \dots, -1, 0, 1, \dots, T$, is defined by

$$\Delta^d Z_t = \sum_{n=0}^{\infty} \pi_n(-d) Z_{t-n},$$

provided the right hand side exists. We collect a few simple results in a lemma, where $D^m \Delta^d Z_t$ denotes the m 'th derivative with respect to d .

Lemma 1 (i) Let Z_t be a stochastic process with fixed and bounded initial values $Z_{-n}, n \geq 0$, then $D^m \Delta^d Z_t$ exists for $d \geq 0$.

Let $Z_t = \sum_{n=0}^{\infty} \xi_n \varepsilon_{t-n}$, where ξ_n is $m \times p$ and ε_t are p -dimensional i.i.d. $(0, \Omega)$ and $\sum_{n=0}^{\infty} |\xi_n| < \infty$. We next consider fractional differences of Z_t without fixing initial values.

- (ii) If $d \geq 0$ then $D^m \Delta^d Z_t$ is a stationary process with absolutely summable coefficients.
 (iii) If $d > -1/2$, then $D^m \Delta^d Z_t$ is a stationary process with square summable coefficients.

Proof. The proof is a simple consequence of the evaluation $|D^m \pi_n(-d)| \leq c(1 + \log n)^m n^{-d-1}$ for $n \geq 1$, see Lemma A.5, which implies that $D^m \pi_n(-d)$ is absolutely summable for $d \geq 0$ and square summable for $d > -1/2$. ■

For $d > -1/2$, an example of these results is the stationary linear process

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

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

$$\Delta_+^d \varepsilon_t = \sum_{n=0}^{t-1} (-1)^n \binom{d}{n} \varepsilon_{t-n} = \sum_{n=0}^{t-1} \pi_n(-d) \varepsilon_{t-n}, \quad t = 1, \dots, T.$$

Thus, for $d \leq -1/2$ we do not use Δ^d directly but apply instead Δ_+^d which is defined for all processes, see for instance Marinucci and Robinson (2000), who use the notation $\Delta^d \varepsilon_t 1_{\{t \geq 1\}}$ and call this a “type II” process.

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

$$T^{d+1/2} \Delta_+^d \varepsilon_{[Tu]} \implies W_{-d-1}(u) = \Gamma(-d)^{-1} \int_0^u (u-s)^{-d-1} dW(s) \text{ on } D[0, 1], \quad (4)$$

where W denotes p -dimensional Brownian motion (BM) generated by ε_t , W_{-d-1} is the corresponding fractional Brownian motion (fBM) of type II, and \implies is used for convergence in distribution as a process on $D^p[0, 1]$ or $C^p[0, 1]$, see Billingsley (1968) or Kallenberg (2001). We also have, see Jakubowski, Mémin, and Pages (1989),

$$T^d \sum_{t=1}^T \Delta_+^d \varepsilon_{t-1} \varepsilon_t' \xrightarrow{d} \int_0^1 W_{-d-1} dW', \quad (5)$$

where \xrightarrow{d} denotes convergence in distribution on $\mathbb{R}^{p \times p}$.

2.2 Solution of fractional autoregressive equations

We consider equation (1) written as $\Pi(L)X_t = \varepsilon_t$. Note that only fractional differences of positive order enter the expression for $\Pi(L)$ when $d \geq b$, and that means that if we consider initial values as fixed and bounded constants, then $\Pi(L)X_t$ is well defined. In order to derive a general expression for the solution in terms of initial values X_{-n} , $n = 0, 1, \dots$, and random shocks $\varepsilon_1, \dots, \varepsilon_t$, we define two operators, see Johansen (2008),

$$\Pi_+(L)X_t = 1_{\{t \geq 1\}} \sum_{i=0}^{t-1} \Pi_i X_{t-i} \text{ and } \Pi_-(L)X_t = \sum_{i=t}^{\infty} \Pi_i X_{t-i}.$$

Here the operator $\Pi_+(L)$ is defined for any sequence because it is a finite sum. Because $\Pi(0) = I_p$, $\Pi_+(L)$ is invertible on sequences that are zero for $t \leq 0$, and the coefficients of the inverse are found by expanding $\Pi(z)^{-1}$ around zero. The process $\Pi_-(L)X_t$ is defined if we assume initial values of X_t fixed and bounded.

The solution of the equation $\Pi(L)X_t = \varepsilon_t$ is found as follows. From

$$\varepsilon_t = \Pi(L)X_t = \Pi_+(L)X_t + \Pi_-(L)X_t,$$

we find by applying $\Pi_+(L)^{-1}$ on both sides that

$$X_t = \Pi_+(L)^{-1}\varepsilon_t - \Pi_+(L)^{-1}\Pi_-(L)X_t = \Pi_+(L)^{-1}\varepsilon_t + \mu_t, \quad t = 1, 2, \dots \quad (6)$$

The first term is the stochastic component generated by $\varepsilon_1, \dots, \varepsilon_t$, and the second a deterministic component generated by initial values. An example of this is the well known result that $X_t = \rho X_{t-1} + \varepsilon_t$ has the solution $X_t = \sum_{i=0}^{t-1} \rho^i \varepsilon_{t-i} + \rho^t X_0$ for any ρ .

The idea of conditioning on initial values is needed in the analysis of autoregressive models for nonstationary processes, and we modify the definition of a fractional process to take account of these. We let ε_t be i.i.d.(0, Ω) in p dimensions and consider $m \times p$ matrices ξ_n with the property that $\sum_{n=0}^{\infty} |\xi_n|^2 < \infty$, and define $C(z) = \sum_{n=0}^{\infty} \xi_n z^n$, $|z| < 1$.

Definition 2 *If $C(z)$ can be extended to a continuous function on the boundary $|z| = 1$ then the process $Z_t = C(L)\varepsilon_t = \sum_{n=0}^{\infty} \xi_n \varepsilon_{t-n}$ is fractional of order θ if $C(1) \neq 0$. A process Z_t is fractional of order $d > 0$ if $\Delta^d Z_t$ is fractional of order zero, and cofractional with cofractionality vector β if $\beta' Z_t$ is fractional of order $d - b \geq 0$ for some $b > 0$.*

The same definitions hold for the process Z_t^+ defined by

$$Z_t^+ = C_+(L)\varepsilon_t + \mu_t = 1_{\{t \geq 1\}} \sum_{n=0}^{t-1} \xi_n \varepsilon_{t-n} + \mu_t, \quad (7)$$

where μ_t is a deterministic term.

2.3 Properties of the solution of fractional autoregressive equations

The solution (6) of equation (1) is valid without any assumptions on the parameters. We next give results, see Johansen (2008, Theorem 8), which guarantee that the process is fractional of order d and cofractional from d to $d - b$. The conditions are given in terms of the roots of the polynomial $\det(\Psi(y))$ and the set \mathbb{C}_b , which is the image of the unit disk under the mapping $y = 1 - (1 - z)^b$.

The following result is Granger's Representation Theorem for the cofractional VAR model (1). It is related to previous representation theorems of Engle and Granger (1987) and Johansen (1988) for the cointegrated VAR, and Johansen (2008) for the fractional model.

Theorem 3 *Let $\Pi(z) = (1 - z)^{d-b} \Psi(1 - (1 - z)^b)$ be given by (3) and let $1/2 \leq b \leq d$. Assume that $\det(\Psi(y)) = 0$ implies that either $y = 1$ or $y \notin \mathbb{C}_b$ and that α and β have rank $r < p$. Let $\Gamma = I_p - \sum_{i=1}^k \Gamma_i$ and assume that $\det(\alpha'_\perp \Gamma \beta_\perp) \neq 0$, so that we can define*

$$C = \beta_\perp (\alpha'_\perp \Gamma \beta_\perp)^{-1} \alpha'_\perp. \quad (8)$$

Then

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

where $H(1) \neq 0$ and $H(y)$ is regular in a neighborhood of \mathbb{C}_b . It follows that the coefficient matrices τ_n defined by $F(z) = H(1 - (1-z)^b) = \sum_{n=0}^{\infty} \tau_n z^n$, $|z| < 1$, satisfy $\sum_{n=0}^{\infty} |\tau_n| < \infty$.

Equation (1) is solved by

$$X_t = C \Delta_+^{-d} \varepsilon_t + \Delta_+^{-(d-b)} Y_t^+ + \mu_t, \quad t = 1, \dots, T, \quad (9)$$

where $\mu_t = -\Pi_+(L)^{-1} \Pi_-(L) X_t$ and

$$Y_t = \sum_{n=0}^{\infty} \tau_n \varepsilon_{t-n} = \tau \varepsilon_t + \Delta^{1/2} \sum_{n=0}^{\infty} \tilde{\tau}_n \varepsilon_{t-n}, \quad \sum_{n=0}^{\infty} |\tilde{\tau}_n| < \infty, \quad (10)$$

where $\tau = \sum_{n=0}^{\infty} \tau_n \neq 0$ so that Y_t is fractional of order zero, $\sum_{h=-\infty}^{\infty} |E(Y_t Y'_{t-h})| < \infty$, and $\beta' \tau \alpha = -I_r$.

Thus X_t is fractional of order d , and because $\beta' C = 0$, X_t is cofractional since $\beta' X_t = \Delta_+^{-(d-b)} \beta' Y_t^+ + \beta' \mu_t$ is fractional of order $d-b$. If $r = 0$, then $\alpha = \beta = 0$, $\alpha_{\perp} = \beta_{\perp} = I_p$, and $C = \Gamma^{-1}$ has full rank, and thus X_t is fractional of order d and not cofractional.

Proof. The expression for $(1-z)^d \Pi(z)^{-1}$ and the definition of Y_t follows from Johansen (2008, Theorem 8), see also formula (17) on page 665 for the result $\beta' H(1) \alpha = \beta' \tau \alpha = -I_r$. Because $H(y)$ is regular in a neighborhood of \mathbb{C}_b we can write $H(y) = H(1) + (1-y)H^*(y)$ where $H^*(y)$ is regular in a neighborhood of \mathbb{C}_b . It follows that $F(z)$ and $F^*(z) = H^*(1 - (1-z)^b)$ are regular in a neighborhood of the unit disk, and hence continuous and differentiable with a square integrable derivative. This implies in particular that the coefficients in the expansion of $H^*(1 - (1-z)^b)$, say τ^* , are summable, but then

$$(1-z)^b H^*(1 - (1-z)^b) = (1-z)^{1/2} \sum_{n=0}^{\infty} \left(\sum_{m=0}^n \pi_m (1/2 - b) \tau_{n-m}^* \right) z^n$$

has summable coefficients $\tilde{\tau}_n = \sum_{m=0}^n \pi_m (1/2 - b) \tau_{n-m}^*$ because $|\tau_n^*|$ and $|\pi_m (1/2 - b)|$ are both summable when $b \geq 1/2$. See also JN (2010, Lemma 1) for the univariate case. ■

2.4 Assumptions for asymptotic analysis

We next formulate some assumptions needed for statistical analysis of the model and for asymptotic analysis of estimators and the likelihood function. We define the parameter set \mathcal{N} for some $d_1 > 1/2$,

$$\mathcal{N} = \{d, b : 1/2 \leq b \leq d \leq d_1\}. \quad (11)$$

Assumption 1 *The process X_t , $t = 1, \dots, T$, is generated by model (1) for some $k \geq 1$, as a function of parameters $\alpha_0, \beta_0, d_0, b_0, \Gamma_{01}, \dots, \Gamma_{0k}, \Omega_0$, errors ε_t that are i.i.d. $(0, \Omega_0)$, and bounded initial values $X_{-n}, n \geq 0$. We assume that the true values satisfy $(d_0, b_0) \in \mathcal{N}$, $\Omega_0 > 0$, $\Gamma_{0k} \neq 0$, α_0 and β_0 are $p \times r$ of rank r , and that $\det(\alpha'_{0\perp} \Gamma_0 \beta_{0\perp}) \neq 0$, so that if $r < p$, $\det(\Psi(y)) = 0$ has $p-r$ unit roots, and the remaining roots of $\det(\Psi(y))$ are outside \mathbb{C}_{b_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 assumption about the true values includes the assumption of cofractionality when $r > 0$, which ensures that X_t is nonstationary and fractional of order d_0 and $\beta'_0 X_t$ is fractional of order $d_0 - b_0$. The assumption $\Gamma_{0k} \neq 0$ guarantees that the lag length is well defined, that the parameters are identified for a given lag length, and that the asymptotic distribution of the maximum likelihood estimator is nonsingular. The assumption about initial values is needed so that $\Delta^d X_t$ can be calculated for any $d \geq 0$, see Lemma 1.

2.5 Initial values

From (9) in Theorem 3 we find that $\Delta^u X_t = \Delta^u_+ X_t + \Delta^u_- X_t$ has the representation

$$\Delta^u X_t = \Delta^{u-d_0}_+(C_0 \varepsilon_t + \Delta^{b_0}_+ Y_t^+) + \Delta^u_+ \mu_t + \Delta^u_- X_t \quad (12)$$

for $t = 1, \dots, T$, where $u \geq 0$ and μ_t is a term generated by initial values X_{-n} , $n \geq 0$.

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

Alternatively, we could think of most phenomena described by fractional processes in economics as having a starting point in the past, say $-N_0$, before which the phenomenon was not defined. That is, we can reasonably set $X_{-n} = 0$, $n > N_0$. The initial values are then X_{-n} , $n = 0, \dots, N_0$, which are observed unmodeled bounded constants. In any case, in practice one would have to truncate the calculation of $\Delta^d X_t$ by setting $X_{-n} = 0$, $n > N_0$. We shall sometimes use this additional assumption in the asymptotic analysis.

We prove that, under either of these assumptions, initial values do not influence the limits of product moments and hence the asymptotic analysis of the likelihood function.

2.6 Identification of parameters

For a given set of parameters $\lambda = (d, b, \alpha, \beta, \Gamma_1, \dots, \Gamma_k, \Omega)$ the characteristic function is given by $\Pi_\lambda(z) = \Pi(z)$, see (3). For two different parameter sets λ and λ^* with $\Pi_\lambda(z) = \Pi_{\lambda^*}(z)$ for all z and the same lag length, $k = k^*$ and $(\Gamma_k, \Gamma_k^*) \neq 0$, it holds that the parameters $(d, b, \alpha, \beta, \Gamma_1, \dots, \Gamma_k, \Omega)$ are identified. See JN (2010, Section 2.3) for a fuller discussion in the univariate case and an example of the indeterminacy between d , b , and k .

3 Likelihood function and maximum likelihood estimators

In this section we first present the likelihood and profile likelihood functions and the maximum likelihood estimator (MLE). We then derive their asymptotic properties.

3.1 Profile likelihood function, its limit, and calculation of MLE

In (2) we eliminate $\Psi_k = I_p - \sum_{i=-1}^{k-1} \Psi_i$ and reparametrize the model equations as

$$\varepsilon_t(\lambda) = \Delta^{d+kb} X_t - \alpha\beta'(\Delta^{d-b} X_t - \Delta^{d+kb} X_t) + \sum_{i=0}^{k-1} \Psi_i(\Delta^{d+ib} X_t - \Delta^{d+kb} X_t), \quad (13)$$

where $\lambda = (d, b, \alpha, \beta, \Psi_0, \dots, \Psi_{k-1}, \Omega) = (d, b, \alpha, \beta, \Psi_*, \Omega)$ are freely varying parameters. The Gaussian likelihood function conditional on initial values $X_{-n}, n \geq 0$, is

$$-2T^{-1} \log L_T(\lambda) = \log \det(\Omega) + \text{tr}\{\Omega^{-1} T^{-1} \sum_{t=1}^T \varepsilon_t(\lambda) \varepsilon_t(\lambda)'\}. \quad (14)$$

For given values of $\psi = (d, b)$ we can calculate the processes $\Delta^{d+kb} X_t$ and $\{\Delta^{d+ib} X_t - \Delta^{d+kb} X_t\}_{i=-1}^{k-1}$ for $d \geq b \geq 1/2$, when initial values are bounded, see Lemma 1. MLEs $(\hat{\alpha}, \hat{\beta}, \hat{\Psi}_*, \hat{\Omega})$ for given ψ , and the partially maximized likelihood or likelihood profile,

$$\ell_{T,r}(\psi) = -2T^{-1} \log L_T(d, b, \hat{\alpha}, \hat{\beta}, \hat{\Psi}_*, \hat{\Omega}), \quad (15)$$

can then be calculated explicitly by reduced rank regression, Anderson (1951), of $\Delta^{d+kb} X_t$ on $\Delta^{d-b} X_t - \Delta^{d+kb} X_t$ corrected for $\{\Delta^{d+ib} X_t - \Delta^{d+kb} X_t\}_{i=0}^{k-1}$, see Johansen (1988). Finally the MLE and maximized likelihood can be calculated by optimizing $\ell_{T,r}(\psi)$ by a numerical procedure. Note that for $r = p$ the likelihood profile $\ell_{T,p}(\psi)$ is found by regression of $\Delta^{d+kb} X_t$ on $\{\Delta^{d+ib} X_t - \Delta^{d+kb} X_t\}_{i=0}^{k-1}$, i.e.

$$\ell_{T,p}(\psi) = \log \det(SSR_T(\psi)) = \log \det(T^{-1} \sum_{t=1}^T R_t R_t'), \quad (16)$$

where $R_t = (\Delta^{d+kb} X_t | \{\Delta^{d+ib} X_t - \Delta^{d+kb} X_t\}_{i=0}^{k-1})$ denotes the regression residual.

We next want to define the limit of the profile likelihood function, $\ell_{T,p}(\psi)$. We note that Theorem 3 gives the properties of $\Delta_+^{d_0} X_t$ at the true parameter point, but the likelihood function depends on product moments of $\Delta^{d+ib} X_t$, and their stationarity properties depend on d and b . We therefore introduce the processes

$$\begin{aligned} Z_{it} &= (\Delta^{d+ib} - \Delta^{d+kb}) \beta'_{0\perp} X_t, & Z_{kt} &= \Delta^{d+kb} \beta'_{0\perp} X_t, \\ W_{jt} &= (\Delta^{d+jb} - \Delta^{d+kb}) \beta'_0 X_t, & W_{kt} &= \Delta^{d+kb} \beta'_0 X_t, \end{aligned} \quad (17)$$

for $i, j = -1, \dots, k$. It is seen from (9) that $C_0 \varepsilon_t + \Delta^{b_0} Y_t$ determines the stochastic behavior of Z_{it} and W_{jt} , so that Z_{it} is asymptotically stationary if $d + ib - d_0 > -1/2$ and W_{jt} is asymptotically stationary if $d + jb - d_0 + b_0 > -1/2$. We denote the corresponding stationary processes S_{zit} and S_{wjt} ,

$$\begin{aligned} S_{zit} &= (\Delta^{d+ib-d_0} - \Delta^{d+kb-d_0}) \beta'_{0\perp} (C_0 \varepsilon_t + \Delta^{b_0} Y_t), & S_{zkt} &= \Delta^{d+kb-d_0} \beta'_{0\perp} (C_0 \varepsilon_t + \Delta^{b_0} Y_t), \\ S_{wjt} &= (\Delta^{d+jb-d_0} - \Delta^{d+kb-d_0}) \beta'_0 (C_0 \varepsilon_t + \Delta^{b_0} Y_t), & S_{wkt} &= \Delta^{d+kb-d_0} \beta'_0 (C_0 \varepsilon_t + \Delta^{b_0} Y_t), \end{aligned}$$

which we combine into

$$S_{it} = \bar{\beta}_0 S_{wit} + \bar{\beta}_{0\perp} S_{zit} = \Delta^{d+ib-d_0} (C_0 \varepsilon_t + \Delta^{b_0} Y_t) \quad (18)$$

if they are both stationary. Here $\bar{\beta}_0 = \beta_0(\beta'_0\beta_0)^{-1}$ and $\bar{\beta}_{0\perp} = \beta_{0\perp}(\beta'_{0\perp}\beta_{0\perp})^{-1}$.

We cover the parameter set \mathcal{N} by sets $\mathcal{N}_{mn} \subset \mathcal{N}$ defined for $-1 \leq m < n \leq k+1$ by

$$\mathcal{N}_{mn} = \{(Z_{mt}, W_{nt}) \text{ asymptotically stationary and } (Z_{m-1,t}, W_{n-1,t}) \text{ nonstationary}\}.$$

Note that if Z_{mt} is asymptotically stationary then $Z_{it}, i \geq m$, is also asymptotically stationary and if $Z_{m-1,t}$ is nonstationary then so is $Z_{it}, i \leq m-1$. Similarly for W_{nt} .

For $\psi \in \mathcal{N}_{mn}$ we define the sigma field generated by the stationary processes underlying Z_{it} and W_{jt} :

$$\mathcal{F}_{stat}(\psi) = \sigma(\{S_{zit}\}_{i=m}^{k-1} \text{ and } \{S_{wjt}\}_{j=n}^{k-1}) \text{ for } \psi \in \mathcal{N}_{mn},$$

and finally the uniform probability limit of $\ell_{T,p}(\psi)$, see Theorem 6,

$$\ell_p(\psi) = \begin{cases} \log \det(\text{Var}(S_{kt}|\mathcal{F}_{stat}(\psi))), & S_{kt} \text{ stationary,} \\ \infty, & \text{otherwise.} \end{cases} \quad (19)$$

Lemma 4 *The function $\ell_p(\psi), \psi \in \mathcal{N}$, has a strict minimum at $\psi = \psi_0$, that is*

$$\ell_p(\psi) \geq \ell_p(\psi_0) = \log \det(\Omega_0),$$

and equality holds if and only if $\psi = \psi_0$.

Proof. We can assume that $S_{kt} = \Delta^{d+kb-d_0}(C_0\varepsilon_t + \Delta^{b_0}Y_t)$ is stationary since otherwise $\ell_p(\psi) = \infty$. The transfer function for $C_0\varepsilon_t + \Delta^{b_0}Y_t$ is $f_0(z)^{-1}$, where $f_0(z) = (1-z)^{-d_0}\Pi_0(z) = \sum_{i=-1}^k \Psi_{0i}(1-z)^{ib_0}$, see (3). For $\psi \in \mathcal{N}_{mn}$, where $\{S_{zit}\}_{i=m}^k, \{S_{wjt}\}_{j=n}^k$ are stationary, we define

$$S_t = S_{kt} - \sum_{i=m}^{k-1} \gamma_{zi} S_{zit} - \sum_{j=n}^{k-1} \gamma_{wj} S_{wjt} = f_{mn}(L)(C_0\varepsilon_t + \Delta^{b_0}Y_t),$$

where $\gamma_{zi} = \Psi_i\bar{\beta}_{0\perp}$ and $\gamma_{wj} = \Psi_j\bar{\beta}_0$, and

$$f_{mn}(L) = \Delta^{d-d_0}[\Delta^{kb}I_p - \sum_{i=m}^{k-1} \gamma_{zi}(\Delta^{ib} - \Delta^{kb})\beta'_{0\perp} - \sum_{j=n}^{k-1} \gamma_{wj}(\Delta^{jb} - \Delta^{kb})\beta'_0].$$

The transfer function of the stationary linear process S_t is $f_{mn}(z)f_0(z)^{-1}$, which has $f_{mn}(0)f_0(0)^{-1} = I_p$, so that S_t is of the form $S_t = \varepsilon_t + \tau_1\varepsilon_{t-1} + \dots$. It follows that $\text{Var}(S_t) \geq \Omega_0$ and equality holds only for $S_t = \varepsilon_t$ or $f_{mn}(z) = f_0(z)$, which implies that $(d, b) = (d_0, b_0)$. Note that for $\psi \in \mathcal{N}_{mn}$, $\text{Var}(S_t)$ is quadratic in the parameters $\{\gamma_{zi}\}_{i=m}^{k-1}, \{\gamma_{wj}\}_{j=n}^{k-1}$, and that minimizing over these, the residual variance satisfies the same inequality

$$\text{Var}(S_{kt}|\mathcal{F}_{stat}(\psi)) = \text{Var}(S_t|\mathcal{F}_{stat}(\psi)) \geq \Omega_0,$$

where equality holds only for $\psi = \psi_0$. This completes the proof of Lemma 4. ■

3.2 Convergence of the profile likelihood function and consistency of the MLE

We now show that the likelihood profile function $\ell_{T,p}(\psi)$ converges uniformly in probability to the deterministic limit $\ell_p(\psi)$ for $T \rightarrow \infty$. This implies that the maximum likelihood estimator in model \mathcal{H}_p exists with probability converging to one and is consistent, and that the same result holds for the submodel \mathcal{H}_r , see (1), and the models with $d = b$ or $d = d_0$.

We first present the result on uniform convergence of the profile likelihood function.

Theorem 5 (i) *Let Assumption 1 hold with $d_0 > b_0$. If in addition either*

$$d_0 - b_0 < 1/2 \text{ and } E|\varepsilon_t|^8 < \infty \quad (20)$$

or

$$E|\varepsilon_t|^q < \infty \text{ for all } q > 0, \quad (21)$$

then the likelihood function for \mathcal{H}_p satisfies

$$\ell_{T,p}(\psi) \implies \ell_p(\psi) \geq \ell_p(\psi_0) = \log \det(\Omega_0) \text{ on } \mathcal{N} \cap \{d - b \geq \delta_0\} \quad (22)$$

for any $\delta_0 \in (0, d_0 - b_0]$, and equality holds only for $\psi = \psi_0$.

(ii) *Let Assumption 1 hold with $X_{-n} = 0, n > N_0$. Then the uniform convergence (22) holds on \mathcal{N} under either of the assumptions (20) or (21).*

(iii) *The results (i) and (ii) also hold for the model with $d = d_0$.*

(iv) *Let Assumption 1 hold with $d_0 = b_0$. For the model with $d = b$ we get uniform convergence on $\mathcal{N} \cap \{d = b\}$ if $E|\varepsilon_t|^8 < \infty$.*

The proof is given in Appendix B. Note that $d_0 - b_0 < 1/2$ as in (20) appears to be perhaps the most empirically relevant range of values for $d_0 - b_0$, because in this case $\beta'_0 X_t$ is (asymptotically) stationary, see e.g. Henry and Zaffaroni (2003) and the references in the introduction. In this important case we have $d + nb - d_0 + b_0 \geq -d_0 + b_0 > -1/2$ so that $\Delta^{d+nb} \beta'_0 X_t$ is asymptotically stationary for all n . This simplifies the proof and allows the weaker moment condition (20). This condition is clearly also satisfied for the model with $d = b$ since then $d_0 = b_0$. Another case that is covered by assumption (20) is the univariate model with a unit root, see JN (2010), which has $\beta_0 = 0$ so that the behavior of $\Delta^{d+nb} \beta'_0 X_t$ is irrelevant.

The reason for the restriction to $\{d - b \geq \delta_0\}$ is the fact that, close to the boundary $\{d = b\}$, the contribution from initial values does not vanish uniformly. This uniformity can be obtained by setting $X_{-n} = 0, n > N_0$, and if $d = b$ the problem does not arise.

We now derive the important consequence of Theorem 5.

Theorem 6 (i) *Let Assumption 1 hold with $d_0 > b_0$, let $0 < \delta_0 \leq d_0 - b_0$, and suppose either (20) or (21) holds. Then with probability converging to one, the maximum likelihood estimator in model $\mathcal{H}_r, r = 0, \dots, p$, exists on $\mathcal{N} \cap \{d - b \geq \delta_0\}$ and is consistent.*

(ii) *Let Assumption 1 hold with $X_{-n} = 0, n > N_0$. Then the results hold on \mathcal{N} under either of the assumptions (20) or (21).*

(iii) *In the model with $d = d_0$ the results (i) and (ii) also hold.*

(iv) *Let Assumption 1 hold with $d_0 = b_0$. For the model with $d = b$, existence and consistency in model $\mathcal{H}_r, r = 0, \dots, p$, hold on $\mathcal{N} \cap \{d = b\}$.*

Proof. Assume that (22) holds. We start with model \mathcal{H}_p , see (1), where α and β are $p \times p$, and the convergence in distribution of the continuous process $\ell_{T,p}(\psi) = \log \det(SSR_T(\psi))$, see Theorem 5, shows that the probability limit $\ell_p(\psi)$ is continuous.

Let $O(\epsilon) = \{\psi : |\psi - \psi_0| < \epsilon\}$ be a small neighborhood around ψ_0 and denote $\mathcal{N}_0 = \mathcal{N} \cap \{d - b \geq \delta_0\}$. Because $\ell_p(\psi)$ is continuous and $> \ell_p(\psi_0)$ if $\psi \neq \psi_0$, see Lemma 4, and $\mathcal{N}_0 \setminus O(\epsilon)$ is compact and does not contain ψ_0 , then $\min_{\psi \in \mathcal{N}_0 \setminus O(\epsilon)} (\ell_p(\psi) - \ell_p(\psi_0)) \geq c_0 > 0$. By the uniform convergence of $\ell_{T,p}(\psi)$ to $\ell_p(\psi)$, we can take for any $\eta > 0$ a $T_0(\eta, \epsilon)$ such that for all $T \geq T_0(\eta, \epsilon)$ we have

$$P\left(\min_{\psi \in \mathcal{N}_0 \setminus O(\epsilon)} |\ell_{T,p}(\psi) - \ell_p(\psi)| \leq c_0/3\right) \geq 1 - \eta/2,$$

and therefore on this set we have

$$\begin{aligned} \min_{\psi \in \mathcal{N}_0 \setminus O(\epsilon)} (\ell_{T,p}(\psi) - \ell_p(\psi_0)) &= \min_{\psi \in \mathcal{N}_0 \setminus O(\epsilon)} [(\ell_{T,p}(\psi) - \ell_p(\psi)) + (\ell_p(\psi) - \ell_p(\psi_0))] \\ &\geq -c_0/3 + c_0 = 2c_0/3. \end{aligned}$$

For any $r \leq p$ we now get, because $\ell_p(\psi_0) = \log \det(\Omega_0)$, that on this set,

$$\min_{\psi \in \mathcal{N}_0 \setminus O(\epsilon)} (\ell_{T,r}(\psi) - \log \det(\Omega_0)) \geq \min_{\psi \in \mathcal{N}_0 \setminus O(\epsilon)} (\ell_{T,p}(\psi) - \log \det(\Omega_0)) \geq 2c_0/3.$$

On the other hand, at the point $\psi = \psi_0$ we have that for all $T \geq T_1(\epsilon, \eta)$,

$$P(|\ell_{T,r}(\psi_0) - \log \det(\Omega_0)| \leq c_0/3) \geq 1 - \eta/2,$$

which implies that the minimum of $\ell_{T,r}(\psi)$ is attained at a point in $O(\epsilon)$. Thus the maximum likelihood estimator of ψ in model \mathcal{H}_r exists with probability $1 - \eta$ and is contained in the set $O(\epsilon)$, which proves consistency, see also van der Vaart (1998, Theorem 5.7).

The model with $d = d_0$ is a submodel of \mathcal{H}_r so the same uniform convergence holds. If $X_{-n} = 0, n > N_0$, or if $d = b$ then the same proof can be used with \mathcal{N} and $\mathcal{N} \cap \{d = b\}$, respectively, instead of $\mathcal{N}_0 = \mathcal{N} \cap \{d - b \geq \delta_0\}$. ■

The result in Theorem 6 on existence and consistency of the MLE involves analyzing the likelihood function on the set of admissible values $1/2 \leq b \leq d \leq d_1$ for any $d_1 \geq d_0$. The likelihood depends on product moments of $\Delta^{d+ib} X_t$ for all such (d, b) , even if the true values are fixed at some b_0 and d_0 . Since the main term in X_t is $\Delta_+^{-d_0} \varepsilon_t$, see (9), analysis of the likelihood function leads to analysis of $\Delta_+^{d+ib-d_0} \varepsilon_t$, which may be asymptotically stationary, nonstationary, or it may be critical in the sense that it may be close to the process $\Delta_+^{-1/2} \varepsilon_t$. The possibility that $\Delta^{d+ib} X_t$ can be critical or close to critical, even if X_t is not, implies that we have to split up the parameter space around values where $\Delta^{d+ib} X_t$ is close to critical and give separate proofs of uniform convergence of the likelihood function in each subset of the parameter space.

This is true in general for any fractional model, where the main term in X_t is typically of the form $\Delta_+^{-d_0} \varepsilon_t$, and analysis of the likelihood function requires analysis of $\Delta^d X_t$ and therefore of a term like $\Delta_+^{d-d_0} \varepsilon_t$ which may be close to critical. To the best of our knowledge, all previous consistency results in the literature for parametric fractional models have either been of a local nature or have covered only the set where $\Delta^d X_t$ is asymptotically stationary,

due to the difficulties in proving uniform convergence of the likelihood function when $\Delta^d X_t$ is close to critical and hence on the whole parameter set, see the discussion in Hualde and Robinson (2010*b*, pp. 2-3).¹

Unlike previous consistency results, our Theorem 6 applies to an admissible parameter set so large that it includes values of (d, b) where $\Delta^{d+ib} X_t$ is asymptotically stationary, nonstationary, and critical. The inclusion of the near critical processes in the proof is made possible by a truncation argument, allowing us to show that when $v \in [-1/2 - \kappa, -1/2 + \kappa]$ for κ sufficiently small, then the inverse of appropriately normalized product moments of critical processes $\Delta_+^v \varepsilon_t$ is tight in v , and further that it is convergent uniformly to zero for $(T, \kappa) \rightarrow (\infty, 0)$, see (83) in Lemma A.11 below.

3.3 A reparametrization and the profile likelihood function for $d, b, \alpha, \Psi_*, \Omega$

We introduce the identified parameter $\theta = T^{d_0-d+b-1/2} \bar{\beta}'_{0\perp} (\beta - \beta_0)$, so that

$$\alpha\beta' = \alpha\beta'_0 + T^{d-b-d_0+1/2} \alpha\theta' \beta'_{0\perp},$$

where $\beta' \bar{\beta}_0$ is absorbed in α . Let $V_t = (W'_{-1t}, \{(\Delta^{d+ib} - \Delta^{d+kb})X'_t\}_{i=0}^{k-1}, \Delta^{d+kb} X'_t)'$ and define, for $\phi = (d, b, \alpha, \Psi_*)$,

$$\varepsilon_t(\phi, \theta) = -\alpha T^{d-b-d_0+1/2} \theta' Z_{-1t} + (-\alpha, \Psi_*, I_p) V_t, \quad (23)$$

see (13). The product moments needed to calculate the conditional likelihood function $-2T^{-1} \log L_T(\phi, \theta)$, see (14), are

$$\begin{pmatrix} \mathcal{A}_T & \mathcal{C}_T \\ \mathcal{C}'_T & \mathcal{B}_T \end{pmatrix} = T^{-1} \sum_{t=1}^T \begin{pmatrix} T^{d-b-d_0+1/2} Z_{-1t} \\ V_t \end{pmatrix} \begin{pmatrix} T^{d-b-d_0+1/2} Z_{-1t} \\ V_t \end{pmatrix}'. \quad (24)$$

Note that $\mathcal{A}_T, \mathcal{B}_T$, and \mathcal{C}_T depend on ψ . We indicate the values for $\psi = \psi_0$ by $\mathcal{A}_T(\psi_0), \mathcal{B}_T(\psi_0)$, and $\mathcal{C}_T(\psi_0)$. Finally we define

$$\mathcal{C}_{\varepsilon T}(\psi_0) = T^{-1/2} \sum_{t=1}^T T^{-b_0+1/2} (\Delta^{d_0-b_0} - \Delta^{d_0+kb_0}) \beta'_{0\perp} X_t \varepsilon'_t. \quad (25)$$

The conditional likelihood $-2T^{-1} \log L_T(\lambda)$ can now be expressed as

$$\log \det(\Omega) + tr\{\Omega^{-1}(\alpha\theta' \mathcal{A}_T \theta \alpha' + (-\alpha, \Psi_*, I_p) \mathcal{B}_T (-\alpha, \Psi_*, I_p)' + 2\alpha\theta' \mathcal{C}_T (-\alpha, \Psi_*, I_p)')\}. \quad (26)$$

For fixed $(d, b, \alpha, \Psi_*, \Omega)$ we estimate θ by regression and find

$$\hat{\theta}(\psi, \alpha, \Psi_*, \Omega) = -\mathcal{A}_T^{-1} \mathcal{C}_T (-\alpha, \Psi_*, I_p)' \Omega^{-1} \alpha (\alpha' \Omega^{-1} \alpha)^{-1}, \quad (27)$$

and the profile likelihood function $-2T^{-1} \log L_{\text{profile}, T}(\psi, \alpha, \Psi_*, \Omega)$ is then

$$\begin{aligned} & \log \det(\Omega) + tr\{\Omega^{-1}(-\alpha, \Psi_*, I_p) \mathcal{B}_T (-\alpha, \Psi_*, I_p)'\} \\ & - tr\{(-\alpha, \Psi_*, I_p) \mathcal{C}'_T \mathcal{A}_T^{-1} \mathcal{C}_T (-\alpha, \Psi_*, I_p)' \Omega^{-1} \alpha (\alpha' \Omega^{-1} \alpha)^{-1} \alpha' \Omega^{-1}\}. \end{aligned}$$

¹In independent and concurrent work, Hualde and Robinson (2010*b*) prove consistency for a large set of admissible values in a univariate fractional model. Their consistency proof applies only to the univariate case (see their discussion on p. 19 and p. 21), and even then it requires all moments finite, where our proof requires only 8 moments in the univariate case.

For (d, b) in a κ_0 -neighborhood of ψ_0 and $i = 0, 1, \dots, k$, the processes S_{it} and $S_{w,-1,t}$, see (18), and their derivatives with respect to (d, b) are stationary because $d + ib - d_0 \geq d - d_0 \geq -\kappa_0 > -1/2$. The process $\Delta_+^{d-b-d_0} C_0 \varepsilon_t$, however, is nonstationary, and when normalized by $T^{d-b-d_0+1/2}$ will converge to fBM provided $E|\varepsilon_t|^q < \infty$ for some $q > -1/(d - b - d_0 + 1/2)$, see (4).

The next theorem summarizes the asymptotic results for the product moments and their derivatives with respect to ψ , denoted \mathbf{D}^m , when ψ belongs to a small κ_0 -neighborhood of ψ_0 . We show that the contribution from initial values can be neglected asymptotically and that the stationary processes S_{zit}, S_{wjt} can replace Z_{it} and W_{jt} . This means that the limit of \mathcal{B}_T can be calculated as

$$\mathcal{B} = \text{Var}(S'_{w,-1,t}, S'_{0t}, \dots, S'_{kt})'.$$

Theorem 7 *Let Assumption 1 be satisfied with $(d_0, b_0) \in \text{int}(\mathcal{N})$ and assume that $E|\varepsilon_t|^q < \infty$ for some $q > (b_0 - 1/2)^{-1}$. Then, for κ_0 sufficiently small and $m \geq 0$, it holds that $\mathbf{D}^m \mathcal{A}_T$, $\mathbf{D}^m \mathcal{B}_T$, and $\mathbf{D}^m \mathcal{C}_T$ are tight and*

$$(\mathcal{A}_T, \mathbf{D}^m \mathcal{B}_T, \mathbf{D}^m \mathcal{C}_T) \Longrightarrow (\beta'_{0\perp} C_0 \int_0^1 W_{d_0-d+b-1} W'_{d_0-d+b-1} du C'_0 \beta_{0\perp}, \mathbf{D}^m \mathcal{B}, 0) \quad (28)$$

as continuous processes on $|\psi - \psi_0| \leq \kappa_0$. At the true value $\psi_0 = (d_0, b_0)$ we find

$$\mathcal{C}_{\varepsilon T}(\psi_0) \xrightarrow{d} \beta'_{0\perp} C_0 \int_0^1 W_{b_0-1} (dW)'. \quad (29)$$

The same results hold for the models with $d = b$ and $d = d_0$.

Proof. From Theorem 3, see also (12), we find that $\Delta^{d+ib} X_t$ has the representation

$$\Delta^{d+ib} X_t = \Delta_+^{d+ib} X_t + \Delta_-^{d+ib} X_t = \Delta_+^{d+ib-d_0} (C_0 \varepsilon_t + \Delta_+^{b_0} Y_t^+) + D_{it}(\psi), t = 1, \dots, T, \quad (30)$$

where $D_{it}(\psi) = \Delta_+^{d+ib} \mu_t + \Delta_-^{d+ib} X_t$ is the deterministic part generated by initial values. It follows from Lemma A.8 that $\mathbf{D}^m D_{it}(\psi)$, suitably normalized, is uniformly small so that it is enough to consider the stochastic parts of $\mathbf{D}^m \mathcal{A}_T, \mathbf{D}^m \mathcal{B}_T,$ and $\mathbf{D}^m \mathcal{C}_T$.

By Theorem 3, $C_0 \varepsilon_t + \Delta_+^{b_0} Y_t^+ \in \mathcal{Z}$, where the class \mathcal{Z} is given in Definition A.10. Lemma A.11 therefore applies directly to product moments of $\Delta_+^{d+ib-d_0} (C_0 \varepsilon_t + \Delta_+^{b_0} Y_t^+)$, because $q > (b_0 - 1/2)^{-1}$ and $(d_0, b_0) \in \text{int}(\mathcal{N})$ implies that we can choose κ_0 such that $q > (d_0 - d + b - 1/2)^{-1}$ for $|\psi - \psi_0| \leq \kappa_0$. Tightness of $\mathbf{D}^m \mathcal{A}_T, \mathbf{D}^m \mathcal{B}_T,$ and $\mathbf{D}^m \mathcal{C}_T$ and the convergence in (28) then follow from Lemma A.11.

To prove (29) we insert (9) into (25) and find

$$\begin{aligned} \mathcal{C}_{\varepsilon T}(\psi_0) &= T^{-b_0} \sum_{t=1}^T \beta'_{0\perp} C_0 (\Delta_+^{-b_0} \varepsilon_{t-1}) \varepsilon_t' + R_{1T} + R_{2T}, \\ R_{1T} &= T^{-b_0} \sum_{t=1}^T \beta'_{0\perp} ((1 - \Delta_+^{(k+1)b_0}) Y_t^+ - C_0 (\Delta_+^{kb_0} - 1) \varepsilon_t) \varepsilon_t', \\ R_{2T} &= T^{-b_0} \sum_{t=1}^T \beta'_{0\perp} C_0 (\Delta_+^{1-b_0} \varepsilon_t - \varepsilon_t) \varepsilon_t', \end{aligned}$$

in addition to an initial value term which is negligible by Lemma A.8.

From (5) the main term converges to the limit in (29). The summands in R_{1T} and R_{2T} are asymptotically stationary martingale difference sequences. In R_{1T} the sum of conditional variances of the summands is proportional to T , and because $b_0 > 1/2$ we find $R_{1T} \xrightarrow{P} 0$. In R_{2T} the sum of conditional variances of $T^{-b_0}(\Delta_+^{1-b_0}\varepsilon_t - \varepsilon_t)\varepsilon_t'$ is

$$\text{Var}(T^{-b_0} \sum_{t=1}^T \text{vec}((\Delta_+^{1-b_0}\varepsilon_t - \varepsilon_t)\varepsilon_t')) = \Omega_0 \otimes \Omega_0 T^{-2b_0} \sum_{t=1}^T \sum_{j=1}^{t-1} \pi_j^2 (b_0 - 1),$$

where

$$T^{-2b_0} \sum_{t=1}^T \sum_{j=1}^{t-1} \pi_j^2 (b_0 - 1) \leq cT^{-2b_0} \sum_{t=1}^T \sum_{j=1}^{t-1} j^{2(b_0-2)} \leq cT^{2\max(b_0-3/2,0)+1-2b_0} \rightarrow 0$$

because $b_0 > 1/2$. This completes the proof of (29). ■

We next define, for $\phi = (d, b, \alpha, \Psi_*)$ and $\theta_0 = 0$, the residuals $\varepsilon_t(\phi) = \varepsilon_t(\phi, 0) = (-\alpha, \Psi_*, I_p)V_t$, c.f. (23). For (d, b) close to (d_0, b_0) we define the corresponding stationary process

$$e_t(\phi) = S_{kt} - \alpha S_{w,-1,t} + \sum_{i=0}^{k-1} \Psi_i S_{it} = (-\alpha, \Psi_*, I_p)(S'_{w,-1,t}, S'_{*t}, S'_{kt})', \quad (31)$$

where S_{it} is given in (18). In the following we use D_ϕ and $D_{\phi\phi}^2$ to denote first- and second-order derivatives with respect to ϕ .

Lemma 8 *We find for $\phi = \phi_0$ that $e_t(\phi_0) = \varepsilon_t(\phi_0) = \varepsilon_t$, and furthermore we have*

$$T^{-1} \sum_{t=1}^T \varepsilon_t(\phi) \varepsilon_t(\phi)' \xrightarrow{P} E e_t(\phi) e_t(\phi)' = (-\alpha, \Psi_*, I_p) \mathcal{B}(-\alpha, \Psi_*, I_p)', \quad (32)$$

$$D_\phi E e_t(\phi_0)' \Omega_0^{-1} e_t(\phi_0) = E[D_\phi e_t(\phi_0)' \Omega_0^{-1} \varepsilon_t] + E[\varepsilon_t' \Omega_0^{-1} D_\phi e_t(\phi_0)] = 0, \quad (33)$$

$$D_{\phi\phi}^2 E e_t(\phi_0)' \Omega_0^{-1} e_t(\phi_0) = E[D_\phi e_t(\phi_0)' \Omega_0^{-1} D_\phi e_t(\phi_0)] = \Sigma_0, \quad (34)$$

where Σ_0 is positive definite if $\Psi_{0k} \neq 0$ or equivalently $\Gamma_{0k} \neq 0$.

Proof. The transfer function for the stationary process $C_0\varepsilon_t + \Delta^{b_0}Y_t$ is $f_0(z)^{-1} = (1 - z)^{d_0}\Pi_0(z)^{-1} = (1 - y)\Psi_0(y)^{-1}$ for $y = 1 - (1 - z)^{b_0}$, see (3), where subscripts indicate that we consider the characteristic and transfer functions for the process defined by the true parameter values. We then find the transfer function for $e_t(\phi)$ to be

$$f_\phi(z) = (1 - z)^{d-b-d_0+b_0} \Psi(1 - (1 - z)^b)|_{\beta=\beta_0} \Psi_0(y)^{-1}. \quad (35)$$

For $\phi = \phi_0$ we find $f_{\phi_0}(z) = 1$ so that $e_t(\phi_0) = \varepsilon_t$. The result (32) follows from (28) of Theorem 7. Differentiating the left hand side of (32), we find (33) and (34) using the results that $e_t(\phi_0) = \varepsilon_t$ and $D_\phi e_t(\phi_0)$ and $D_{\phi\phi}^2 e_t(\phi_0)$ are measurable with respect to $\varepsilon_1, \dots, \varepsilon_{t-1}$.

Finally we prove that if $\Psi_{0k} \neq 0$, then Σ_0 is positive definite. If Σ_0 were singular, there would exist a linear combination of the processes $D_\phi e_t(\phi_0)$ which had variance zero. We want

to show that this is not possible when $\Psi_{0k} \neq 0$. The statement that Σ_0 is singular translates into a statement that there is a linear combination of the derivatives of the transfer function $f_\phi(z)$ which, for $\phi = \phi_0$, is zero. That is, for some set of values $h = (d_1, b_1, A, G_*)$ of the same dimensions as $\phi = (d, b, \alpha, \Psi_*)$, the derivative $D_s f_{\phi_0+sh}(z)|_{s=0} = 0$. We find from (2) and (35) the derivatives, where we use $y = 1 - (1 - z)^{b_0}$,

$$\begin{aligned} D_d f_{\phi_0}(z) &= \log(1 - z)I_p = b_0^{-1} \log(1 - y)I_p, \\ D_b f_{\phi_0}(z) &= -b_0^{-1} \log(1 - y)(I_p + D_y \Psi_0(y)(1 - y)\Psi_0(y)^{-1}), \\ D_{\Psi_i} f_{\phi_0}(z) &= (1 - y)^i, i = 0, \dots, k - 1, \\ D_\alpha f_{\phi_0}(z) &= -\beta'_0 y. \end{aligned}$$

This gives the directional derivative $D_s f_{\phi_0+sh}(z)|_{s=0}$ in the direction $h = (d_1, b_1, A, G_*)$ which, multiplied by $\Psi_0(y)$, is

$$b_0^{-1} \log(1 - y) \{ (d_1 - b_1)\Psi_0(y) - b_1 D_y \Psi_0(y)(1 - y) \} - \{ A\beta'_0 y \Psi_0(y) + \sum_{i=0}^{k-1} G_i (1 - y)^i \Psi_0(y) \}.$$

This should be zero for all y for Σ_0 to be singular. Because $\log(1 - y)$ is not a polynomial we have $A\beta'_0 y \Psi_0(y) + \sum_{i=0}^{k-1} G_i (1 - y)^i \Psi_0(y) = 0$ for all y , and hence $A = 0$ and $G_i = 0, i = 0, \dots, k - 1$. We then find that the coefficient to $b_0^{-1} \log(1 - y)$ should be zero, so that

$$(d_1 - b_1)\Psi_0(y) - b_1 D_y \Psi_0(y)(1 - y) = 0 \text{ for all } y.$$

For $y = 1$ we find from (2) that $\Psi_0(1) = -\alpha_0 \beta'_0$ and therefore $(d_1 - b_1)\alpha_0 \beta'_0 = 0$, and hence $b_1 = d_1$, so that $(d_1 - b_1)\Psi_0(y) = 0$. The coefficient of the highest order term in the polynomial $b_1 D_y \Psi_0(y)$ is $(-1)^{k+1} b_1 (k + 1) \Psi_{0k}$ and for this to be zero when $\Psi_{0k} \neq 0$ we must have $b_1 = d_1 = 0$. Hence Σ_0 is positive definite. From (2) $\Psi_{0k} \neq 0$ is the same as $\Gamma_{0k} \neq 0$. ■

3.4 Asymptotic distribution of the MLE

We first find asymptotic distributions of the score functions and the limit of the information at the true value. We then expand the likelihood function in a neighborhood of the true value and find asymptotic distributions of MLEs. By Lemmas A.2 and A.3 we only need the information at the true value because the estimators are consistent (by Theorem 6) and first and second derivatives are tight (by Theorem 7).

Lemma 9 *Under Assumption 1 with $(d_0, b_0) \in \text{int}(\mathcal{N})$ and $E|\varepsilon_t|^q < \infty$ for some $q > (b_0 - 1/2)^{-1}$, the limit distribution of the Gaussian score function for model (1) at the true value is given by*

$$\begin{pmatrix} T^{-1/2} D_\phi \log L_T(\lambda_0) \\ T^{-1/2} D_\theta \log L_T(\lambda_0) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} N_{n_\phi}(0, \Sigma_0) \\ (\text{vec} \int_0^1 F_0(dG_0)')' \end{pmatrix}, \quad (36)$$

where Σ_0 is given in (34), $n_\phi = 1 + 1 + pr + kp^2$ is the number of parameters in $\phi = (d, b, \alpha, \Psi_*)$, $\theta = T^{-d+b+d_0+1/2} \beta'_{0\perp} (\beta - \beta_0)$, $F_0 = \beta'_{0\perp} C_0 W_{b_0-1}$, and $G_0 = \alpha'_0 \Omega_0^{-1} W$.

Proof. The score function for ϕ evaluated at the true value is

$$T^{-1/2} \mathbf{D}_\phi \log L_T(\lambda_0) = -T^{-1/2} \sum_{t=1}^T \varepsilon_t' \Omega_0^{-1} \mathbf{D}_\phi \varepsilon_t(\phi_0, 0),$$

where $T^{-1/2} \varepsilon_t' \Omega_0^{-1} \mathbf{D}_\phi \varepsilon_t(\phi_0, 0)$ is a martingale difference with sum of conditional variances

$$T^{-1} \sum_{t=1}^T \mathbf{D}_\phi \varepsilon_t(\phi_0, 0)' \Omega_0^{-1} \mathbf{D}_\phi \varepsilon_t(\phi_0, 0) \xrightarrow{P} \Sigma_0,$$

see Lemma 8. The result for the first block of (36) now follows from the central limit theorem for martingales, see Hall and Heyde (1980, chp. 3).

The score function for θ evaluated at the true value is

$$\begin{aligned} T^{-1/2} \mathbf{D}_\theta \log L_T(\lambda_0) &= -T^{1/2} \sum_{t=1}^T \varepsilon_t' \Omega_0^{-1} \mathbf{D}_\theta \varepsilon_t(\phi_0, 0) = T^{-b_0} \sum_{t=1}^T \varepsilon_t' \Omega_0^{-1} (\alpha_0 \otimes Z'_{-1t}) \\ &= T^{-b_0} \sum_{t=1}^T (\text{vec}(Z_{-1t} \varepsilon_t' \Omega_0^{-1} \alpha_0))' \xrightarrow{d} (\text{vec} \int_0^1 F_0(dG_0))', \end{aligned}$$

see (29) of Theorem 7, which proves the second block of (36). ■

Lemma 10 *Under Assumption 1 with $(d_0, b_0) \in \text{int}(\mathcal{N})$ and $E|\varepsilon_t|^q < \infty$ for some $q > (b_0 - 1/2)^{-1}$, the Gaussian information per observation in model (1) for $(\phi, \theta) = (\phi_0, 0)$ converges in distribution to*

$$\begin{pmatrix} \Sigma_0 & 0 \\ 0 & \alpha_0' \Omega_0^{-1} \alpha_0 \otimes \int_0^1 F_0 F_0' du \end{pmatrix}, \quad (37)$$

where Σ_0 is given in (34) and $F_0 = \beta'_{0\perp} C_0 W_{b_0-1}$.

Proof. The information matrices for the different parameters can be found from (26). From (28) of Theorem 7 it holds that $\mathbf{D}^m \mathcal{C}_T(\psi_0) \xrightarrow{P} 0$. Using this and (34) we find for $\theta_0 = 0$ that

$$\begin{aligned} -T^{-1} \mathbf{D}_{\phi\phi}^2 \log L_T(\lambda_0) &\xrightarrow{P} \Sigma_0, \\ -T^{-1} \mathbf{D}_{\theta\theta}^2 \log L_T(\lambda_0) &= \alpha_0' \Omega_0^{-1} \alpha_0 \otimes \mathcal{A}_T(\psi_0) \xrightarrow{d} \alpha_0' \Omega_0^{-1} \alpha_0 \otimes \int_0^1 F_0 F_0' du, \\ -T^{-1} \mathbf{D}_{\theta\phi}^2 \log L_T(\lambda_0) &= \mathbf{D}_{\theta\phi}^2 \text{tr}\{\Omega^{-1} 2\alpha\theta' \mathcal{C}_T(-\alpha, \Psi_*, I_p)'\} |_{\lambda=\lambda_0} \xrightarrow{P} 0. \end{aligned}$$

■

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

Theorem 11 *Under the assumptions of Theorems 5 and 6, $(d_0, b_0) \in \text{int}(\mathcal{N})$, and $E|\varepsilon_t|^q < \infty$ for some $q > (b_0 - 1/2)^{-1}$, the asymptotic distribution of the Gaussian maximum likelihood estimators $\hat{\phi} = (\hat{d}, \hat{b}, \hat{\alpha}, \hat{\Psi}_*)$ and $\hat{\beta}$ for model (1) is given by*

$$\begin{pmatrix} T^{1/2} \text{vec}(\hat{\phi} - \phi_0) \\ T^{b_0} \bar{\beta}'_{0\perp} (\hat{\beta} - \beta_0) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} N_{n_\phi}(0, \Sigma_0^{-1}) \\ (\int_0^1 F_0 F_0' du)^{-1} \int_0^1 F_0(dV_0)' \end{pmatrix}, \quad (38)$$

where $F_0 = \beta'_{0\perp} C_0 W_{b_0-1}$ and $V_0 = (\alpha'_0 \Omega_0^{-1} \alpha_0)^{-1} \alpha'_0 \Omega_0^{-1} W$ are independent. It follows that the asymptotic distribution of $T^{b_0} \text{vec}(\bar{\beta}'_{0\perp} (\hat{\beta} - \beta_0))$ is mixed Gaussian with conditional variance given by

$$(\alpha'_0 \Omega_0^{-1} \alpha_0)^{-1} \otimes \left(\int_0^1 F_0 F_0' du \right)^{-1}. \quad (39)$$

In the models with $d = d_0$ or $d = b$ the same results hold with the relevant restriction imposed.

Proof. To find limit distributions of $T^{1/2}(\hat{\phi} - \phi_0)$ and $T^{b_0} \bar{\beta}'_{0\perp} (\hat{\beta} - \beta_0) = T^{1/2} \hat{\theta}$, we apply the usual expansion of the score function around $\phi = \phi_0$, $\beta = \beta_0$ (or $\theta = 0$), and $\Omega = \hat{\Omega}$. Using Taylor's formula with remainder term we find for $l_T = \log L_T$ that

$$0 = \begin{pmatrix} T^{1/2} \mathbf{D}_\phi l_T(\phi_0, 0, \hat{\Omega}) \\ T^{1/2} \mathbf{D}_\theta l_T(\phi_0, 0, \hat{\Omega}) \end{pmatrix} + \begin{pmatrix} \mathbf{D}_{\phi\phi} l_T(\lambda^*) & \mathbf{D}_{\phi\theta} l_T(\lambda^*) \\ \mathbf{D}_{\theta\phi} l_T(\lambda^{**}) & \mathbf{D}_{\theta\theta} l_T(\lambda^{**}) \end{pmatrix} \begin{pmatrix} T^{1/2} \text{vec}(\hat{\phi} - \phi_0) \\ T^{1/2} \text{vec} \hat{\theta} \end{pmatrix}.$$

Here the asterisks indicate intermediate points between $(\hat{\phi}, \hat{\theta}, \hat{\Omega})$ and $(\phi_0, 0, \hat{\Omega})$, which therefore converge to $(\phi_0, 0, \Omega_0)$ in probability by Theorem 6.

Because the first and second derivatives are tight, see Theorem 7 and Lemma A.2, and $(\lambda^*, \lambda^{**}) \xrightarrow{P} (\lambda_0, \lambda_0)$, see Theorem 6, we apply Lemma A.3 to replace intermediate points by $(\phi_0, 0, \Omega_0)$. The score functions normalized by $T^{1/2}$ and their weak limits for $\lambda = \lambda_0$ are given in Lemma 9 and the limit of the information per observation in Lemma 10, see (37). Pre-multiplying by its inverse we find (38). The process

$$F_0 = \beta'_{0\perp} C_0 W_{b_0-1} = \int_0^u (u-s)^{b_0-1} \beta'_{0\perp} C_0 dW(s)$$

is a function of $\alpha'_{0\perp} W$, see (8), whereas $V_0 = (\alpha'_0 \Omega_0^{-1} \alpha_0)^{-1} \alpha'_0 \Omega_0^{-1} W$ depends only on $\alpha'_0 \Omega_0^{-1} W$, so that F_0 and V_0 are independent and the limit distribution of $T^{b_0} \bar{\beta}'_{0\perp} (\hat{\beta} - \beta_0)$ is mixed Gaussian.

If $d = d_0$ or $d = b$, the same expansions can be made and similar results derived. ■

The result in Theorem 11 shows under i.i.d. errors with suitable moments conditions, that $\hat{\phi}$ is asymptotically Gaussian, while the estimated cointegration vectors $\hat{\beta}$ are locally asymptotically mixed normal (LAMN). Results like these are well known from the standard (non-fractional) cointegration model, but are much less developed for fractional models, see the references in Section 1. These are important results, which allow asymptotically standard (chi-squared) inference on all parameters of the model – including the cointegrating relations and orders of fractionality – using Gaussian likelihood ratio tests.

Furthermore, this result has optimality implications for the estimation of β in the cofractional VAR. In our LAMN case with stochastic information matrix, $\hat{\beta}$ is asymptotically optimal under the additional assumption of Gaussian errors in the sense that it has asymptotic maximum concentration probability, see, e.g., Phillips (1991) and Saikkonen (1991) for the precise definitions in the context of the standard cointegration model.

4 Likelihood ratio test for cofractional rank

We consider the model

$$\Delta^d X_t = \Pi \Delta^{d-b} L_b X_t + \sum_{i=0}^k \Gamma_i \Delta^d L_b^i X_t + \varepsilon_t \quad (40)$$

and want to test the hypothesis $\mathcal{H}_r : \text{rank}(\Pi) = r$ against the alternative $\mathcal{H}_p : \text{rank}(\Pi) = p$. Let $\ell_{T,r}(\psi)$ be the profile likelihood function, where $\alpha, \beta, \Gamma_*, \Omega$ have been concentrated out by regression and reduced rank regression, see Section 3.1, and let $\hat{\psi}_r$ be the MLE of ψ in model $\mathcal{H}_r, r = 0, 1, \dots, p$. The likelihood ratio (LR) statistic is

$$-2 \log LR(\mathcal{H}_r | \mathcal{H}_p) = \ell_{T,p}(\hat{\psi}_p) - \ell_{T,r}(\hat{\psi}_r). \quad (41)$$

Theorem 12 *Under the assumptions of Theorem 11 the likelihood ratio statistic (41) in model (40) has asymptotic distribution*

$$-2 \log LR(\mathcal{H}_r | \mathcal{H}_p) \xrightarrow{d} \text{tr} \left\{ \int_0^1 (dB) B'_{b_0-1} \left(\int_0^1 B_{b_0-1} B'_{b_0-1} du \right)^{-1} \int_0^1 B_{b_0-1} (dB)' \right\}, \quad (42)$$

where B is $(p-r)$ -dimensional standard BM and B_{b_0-1} the corresponding fBM.

If we take an alternative $\Pi = \alpha\beta' + \alpha_1\beta_1' = (\alpha, \alpha_1)(\beta, \beta_1)'$, where α_1, β_1 are $p \times r_1$ of rank r_1 and (α, α_1) and (β, β_1) are of rank $r + r_1 > r$, and hence $\text{rank}(\Pi) > r$, and assume that Assumption 1 is satisfied under the alternative, then

$$-2 \log LR(\mathcal{H}_r | \mathcal{H}_p) \xrightarrow{P} \infty. \quad (43)$$

In the models with $d = d_0$ or $d = b$ the same results hold.

Proof. *Proof of (42):* We first derive the limit result (42) assuming that $\text{rank}(\Pi) = r$, so that $\Pi = \alpha\beta'$ where α and β are $p \times r$ of rank r . It is convenient to introduce the extra hypothesis that $\Pi = \alpha\beta'$ and $\beta = \beta_0$, or $\Pi = \alpha\beta'_0$, see Lawley (1956), and Johansen (2002) for an application to the cointegrated VAR model.

Then $LR(\mathcal{H}_r | \mathcal{H}_p)$ is

$$\frac{\max_{\Pi=\alpha\beta'} L}{\max L} = \frac{\max_{\Pi=\alpha\beta'_0} L}{\max L} / \frac{\max_{\Pi=\alpha\beta'_0} L}{\max_{\Pi=\alpha\beta'} L} = \frac{LR(\mathcal{H}_r, \beta = \beta_0 | \mathcal{H}_p)}{LR(\beta = \beta_0 | \mathcal{H}_r)}.$$

The statistic $LR(\mathcal{H}_r, \beta = \beta_0 | \mathcal{H}_p)$ is the test that $\Pi = \alpha\beta'_0$ (with rank r) against Π unrestricted, and $LR(\beta = \beta_0 | \mathcal{H}_r)$ is the test that $\beta = \beta_0$ in the model with $\Pi = \alpha\beta'$ and $\text{rank}(\Pi) = r$. We next find a first order approximation to each statistic and subtracting them we find the asymptotic distribution.

In both cases we apply the result that when, in a statistical problem with vector valued parameters ξ and η , the limiting observed information per observation is block diagonal and tight as a continuous process in a neighborhood of the true value, then a Taylor expansion of the log likelihood ratio statistic and the score function shows that

$$-2 \log LR(\xi = \xi_0) = \mathbf{D}_\xi \log L_T(\xi_0, \eta_0) [\mathbf{D}_{\xi\xi}^2 \log L_T(\xi_0, \eta_0)]^{-1} \mathbf{D}_\xi \log L_T(\xi_0, \eta_0)' + o_P(1), \quad (44)$$

see JN (2010, Theorem 14) for a detailed discussion of the univariate case.

A *first order approximation to* $-2\log LR(\beta = \beta_0|\mathcal{H}_r)$: It follows from Lemma 10 that, for $\xi = \theta$, $\eta = (d, b, \alpha, \Psi_*, \Omega)$, the asymptotic information per observation is block diagonal at the true value, and Theorem 7 and Lemma A.2 show that the information is tight as a process in the parameters. Thus we have that $-2\log LR(\beta = \beta_0|\mathcal{H}_r)$ is

$$\begin{aligned} & (\text{vec } \mathcal{C}_{\varepsilon T}(\psi_0)\Omega_0^{-1}\alpha_0)'(\alpha_0'\Omega_0^{-1}\alpha_0 \otimes \mathcal{A}_T(\psi_0))^{-1} \text{vec } \mathcal{C}_{\varepsilon T}(\psi_0)\Omega_0^{-1}\alpha_0 + o_P(1) \\ & = \text{tr}\{(\alpha_0'\Omega_0^{-1}\alpha_0)^{-1}\alpha_0'\Omega_0^{-1}\mathcal{C}_{\varepsilon T}(\psi_0)'\mathcal{A}_T(\psi_0)^{-1}\mathcal{C}_{\varepsilon T}(\psi_0)\Omega_0^{-1}\alpha_0\} + o_P(1), \end{aligned} \quad (45)$$

using the relation $\text{tr}\{ABCD\} = (\text{vec } B)'(A' \otimes C) \text{vec } D$.

A *first order approximation to* $-2\log LR(\mathcal{H}_r, \beta = \beta_0|\mathcal{H}_p)$: In model (40) we introduce a convenient reparametrization by $\alpha = \Pi\bar{\beta}_0$ and $\xi' = T^{-\delta-1-1/2}\Pi\bar{\beta}_{0\perp}$, so that $\Pi = \alpha\beta_0' + T^{\delta-1+1/2}\xi'\beta_{0\perp}'$ and the equations are

$$\Delta^{d+kb}X_t = \alpha W_{-1t} + T^{\delta-1+1/2}\xi'Z_{-1t} + \sum_{i=1}^k \Psi_i(\Delta^{d+ib} - \Delta^{d+kb})X_t + \varepsilon_t.$$

The likelihood function $-2T^{-1}\log L_T(\xi, \eta)$ conditional on initial values becomes

$$\log \det(\Omega) + \text{tr}\{\Omega^{-1}(\xi'\mathcal{A}_T\xi + (-\alpha, \Psi_*, I_p)\mathcal{B}_T(-\alpha, \Psi_*, I_p)' + 2\xi'\mathcal{C}_T(-\alpha, \Psi_*, I_p)')\},$$

where $\eta = (d, b, \alpha, \Psi_*, \Omega)$. This expression is the same as the conditional likelihood (26) except that $\alpha\theta'$ is replaced by ξ' . The properties of the likelihood function and its derivatives can be derived from those of \mathcal{A}_T , \mathcal{B}_T , and \mathcal{C}_T , and it is seen that the second derivative as a function of the parameters is tight and that the limit is block diagonal. It follows as above that

$$-2\log LR(\mathcal{H}_r, \beta = \beta_0|\mathcal{H}_p) = \text{tr}\{\Omega_0^{-1}\mathcal{C}_{\varepsilon T}(\psi_0)'\mathcal{A}_T(\psi_0)^{-1}\mathcal{C}_{\varepsilon T}(\psi_0)\} + o_P(1). \quad (46)$$

A *first order approximation to* $-2\log LR(\mathcal{H}_r|\mathcal{H}_p)$: Subtracting (45) from (46) and applying the identity

$$\Omega_0^{-1} - \Omega_0^{-1}\alpha_0(\alpha_0'\Omega_0^{-1}\alpha_0)^{-1}\alpha_0'\Omega_0^{-1} = \alpha_{0\perp}(\alpha_{0\perp}'\Omega_0\alpha_{0\perp})^{-1}\alpha_{0\perp}'$$

we find that $-2\log LR(\mathcal{H}_r|\mathcal{H}_p)$ has the same limit as

$$\begin{aligned} & \text{tr}\{\alpha_{0\perp}(\alpha_{0\perp}'\Omega_0\alpha_{0\perp})^{-1}\alpha_{0\perp}'\mathcal{C}_{\varepsilon T}(\psi_0)'\mathcal{A}_T(\psi_0)^{-1}\mathcal{C}_{\varepsilon T}(\psi_0)\} \\ & \xrightarrow{d} \text{tr}\{\alpha_{0\perp}(\alpha_{0\perp}'\Omega_0\alpha_{0\perp})^{-1}\alpha_{0\perp}' \int_0^1 (dW)F_0' \left(\int_0^1 F_0F_0' du \right)^{-1} \int_0^1 F_0(dW)'\}, \end{aligned} \quad (47)$$

which is the desired result for $B = (\alpha_{0\perp}'\Omega_0\alpha_{0\perp})^{-1/2}\alpha_{0\perp}'W$.

Proof of (43): We want to analyze the alternative that $\Pi = \alpha\beta' + \alpha_1\beta_1' = (\alpha, \alpha_1)(\beta, \beta_1)'$, where $\text{rank}(\Pi) > r$, and apply the same methods as in the proof of (42). Under the alternative there are more parameters and therefore the information matrix is larger, but still asymptotically block diagonal. The information for the parameters $(d, b, \alpha, \beta, \Psi_*, \Omega)$ is therefore also asymptotically block diagonal so that (44) also holds under the alternative.

Without loss of generality we can set $\beta_1 = \beta_{0\perp}\zeta_0$ for a conforming matrix ζ_0 , so that $\zeta_0'\beta_{0\perp}'X_t$ is $\mathcal{F}(d_0 - b_0)$ under the alternative. Moreover, Assumption 1 holds under the alternative, and in particular $\det((\alpha_0, \alpha_1)_{\perp}\Gamma_0(\beta_0, \beta_1)_{\perp}) \neq 0$, so that $\zeta_{0\perp}'\beta_{0\perp}'X_t$ is still $\mathcal{F}(d_0)$.

Under the alternative we do not have $\varepsilon_t(\phi_0) = \varepsilon_t$ but instead

$$\varepsilon_t^{alt}(\phi_0) = \varepsilon_t + \alpha_1 \zeta_0' Z_{-1t}^0, \quad (48)$$

where $\zeta_0' Z_{-1t}^0 = (\Delta_+^{d_0-b_0} - \Delta_+^{d_0+kb_0}) \zeta_0' \beta_{0\perp}' X_t$ is an asymptotically stationary $\mathcal{F}(0)$ process. To analyze the approximation (47) we define $\mathcal{C}_{\varepsilon T}^{alt} = T^{-b_0} \sum_{t=1}^T Z_{-1t}^0 \varepsilon_t^{alt}(\phi_0)'$ and consider the matrix

$$\mathcal{C}_{\varepsilon T}^{alt} \mathcal{A}_T(\psi_0)^{-1} \mathcal{C}_{\varepsilon T}^{alt} \geq (\zeta_0' \mathcal{C}_{\varepsilon T}^{alt})' (\zeta_0' \mathcal{A}_T(\psi_0) \zeta_0)^{-1} (\zeta_0' \mathcal{C}_{\varepsilon T}^{alt}) \quad (49)$$

and want to show that the right hand side tends to infinity in probability. We find from (78) of Lemma A.11 and (48) that

$$T^{b_0-1} \zeta_0' \mathcal{C}_{\varepsilon T}^{alt} = T^{-1} \sum_{t=1}^T \zeta_0' Z_{-1t}^0 \varepsilon_t' + T^{-1} \sum_{t=1}^T \zeta_0' Z_{-1t}^0 Z_{-1t}^{0'} \zeta_0 \alpha_1',$$

which converges in probability to $E(\zeta_0' S_{z,-1,t}^0 S_{z,-1,t}^{0'} \zeta_0 \alpha_1') = Var(\zeta_0' S_{z,-1,t}^0) \alpha_1'$, where $\zeta_0' S_{z,-1,t}^0 = (1 - \Delta^{(k+1)b_0})(\beta_{0\perp} \zeta_0)' Y_t$ is a stationary $\mathcal{F}(0)$ process. We also find that

$$T^{2b_0-1} \zeta_0' \mathcal{A}_T(\psi_0) \zeta_0 = T^{-1} \sum_{t=1}^T \zeta_0' Z_{-1t}^0 Z_{-1t}^{0'} \zeta_0 \xrightarrow{P} E(\zeta_0' S_{z,-1,t}^0 S_{z,-1,t}^{0'} \zeta_0),$$

because under the alternative $\zeta_0' Z_{-1t}^0$ is an asymptotically stationary $\mathcal{F}(0)$ process. Inserting both these expressions into (49) we see that the right hand side multiplied by T^{-1} converges in probability to the deterministic limit $\alpha_1 Var(\zeta_0' S_{z,-1,t}^0) \alpha_1' > 0$, which proves (43). ■

The distribution (42) of the LR test for cointegration rank is a fractional version of the distribution of the trace test in the cointegrated I(1) VAR model, see Johansen (1988). Note that it is the parameter b_0 , describing the “strength” of the cofractional relations, which determines the order of the fBMs in the distribution. The parameter d_0 does not appear in the distribution. For a given b_0 , either hypothesized or estimated (\hat{b}_r), the distribution (42) can easily be simulated to obtain critical values. Table 1 provides quantiles of (42) for a range of values of b_0 .

To find the cofractional rank a sequence of tests, for a given size δ , can be conducted in the usual way: test \mathcal{H}_r for $r = 0, 1, \dots$ until rejection, and the estimated rank is then the last value of r which is not rejected by the sequence of tests. If the true rank is r_0 , then the consistency of the LR rank test in Theorem 12 shows that any test of $r < r_0$ will reject with probability one as $T \rightarrow \infty$. Thus, $P_{r_0}(\hat{r} < r_0) \rightarrow 0$. Since the asymptotic size of the test for rank is δ we also have that $P_{r_0}(\hat{r} = r_0) \rightarrow 1 - \delta$ and it follows that $P_{r_0}(\hat{r} > r_0) \rightarrow \delta$. This shows that \hat{r} is almost consistent, in the sense that it attains the true value with probability $1 - \delta$ as $T \rightarrow \infty$.

5 Conclusion

We have generalized the well known likelihood based inference results for the cointegrated VAR model,

$$\Delta X_t = \alpha \beta' X_{t-1} + \sum_{i=1}^k \Gamma_i \Delta X_{t-i} + \varepsilon_t,$$

Table 1: Simulated quantiles of the distribution (42)

b_0	level	$p-r=1$	$p-r=2$	$p-r=3$	$p-r=4$	$p-r=5$	$p-r=6$	$p-r=7$	$p-r=8$
0.6	0.90	2.57	8.01	15.63	25.58	38.21	53.11	70.39	90.34
	0.95	3.71	9.58	17.81	28.35	41.59	57.36	75.23	95.35
	0.99	6.34	13.50	22.79	34.30	48.36	65.35	84.38	105.68
0.8	0.90	2.77	9.17	18.35	30.89	46.74	65.37	87.02	112.04
	0.95	4.01	10.91	20.79	33.92	49.90	69.75	92.20	117.76
	0.99	6.63	14.74	26.17	39.69	57.27	78.64	102.03	128.76
1.0	0.90	2.98	10.54	21.56	36.83	56.03	79.17	105.80	136.87
	0.95	4.20	12.40	24.09	39.84	59.54	83.59	111.11	142.57
	0.99	6.98	16.17	29.70	45.74	66.95	92.47	121.14	154.65
1.2	0.90	3.21	11.42	24.08	41.46	63.26	90.08	120.81	156.11
	0.95	4.41	13.42	26.75	44.52	67.08	94.52	125.63	162.28
	0.99	7.27	17.47	32.06	50.40	74.80	103.25	135.81	173.12
1.4	0.90	3.39	12.04	25.64	43.93	67.30	95.97	128.75	166.90
	0.95	4.57	13.98	28.18	47.03	71.09	100.71	133.96	172.74
	0.99	7.58	18.25	33.65	54.01	79.21	109.34	143.92	184.27

Note: Simulations are based on 10,000 replications and sample size 1000.

to the cointegrated fractional VAR model,

$$\Delta^d X_t = \alpha\beta' \Delta^{d-b} L_b X_t + \sum_{i=1}^k \Gamma_i \Delta^d L_b^i X_t + \varepsilon_t, \quad 1/2 \leq b \leq d.$$

We have analyzed the conditional Gaussian likelihood given initial values, which we assumed bounded. We have shown existence and consistency and derived the asymptotic distribution of the maximum likelihood estimator. In the asymptotic analysis we assumed i.i.d. errors with suitable moment conditions. We have derived the asymptotic distribution of the test for the rank of $\alpha\beta'$ and shown that it is expressed in terms of fractional Brownian motion B_{b-1} , that inference on β is asymptotically mixed Gaussian, and finally that the estimators of the remaining parameters are asymptotically Gaussian. The same type of results are valid for the two submodels with $1/2 \leq d = b$ and $1/2 \leq b \leq d = d_0$, respectively.

The main technical contribution in this paper is the proof of existence and consistency of the maximum likelihood estimator, which allows standard likelihood theory to be applied. This involves an analysis of the influence of initial values as well as proving tightness and uniform convergence of product moments of processes that can be critical and nearly critical, and this was made possible by a truncation argument.

Appendix A Product moments

In this appendix we evaluate product moments of stochastic and deterministic terms and find their limits based on results for convergence in distribution of probability measures on $C^p[0, 1]^m$ and $D^p[0, 1]^m$.

A.1 Results on convergence in distribution

For a multivariate random variable Z with $E|Z|^q < \infty$ the L_q norm is $\|Z\|_q = E(|Z|^q)^{1/q}$.

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

$$\|X_n(s)\|_4 \leq c, \quad \text{and} \quad \|X_n(s) - X_n(t)\|_4 \leq c|s - t| \quad (50)$$

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

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

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

Proof. JN (2010, Lemma A.2). ■

Lemma A.3 *Assume that $S_n \xrightarrow{P} s_0 \in [0, 1]^m$ and that the $p \times p$ matrix-valued continuous process $X_n(s)$ is tight on $[0, 1]^m$. Then $X_n(S_n) - X_n(s_0) \xrightarrow{P} 0$.*

Proof. See JN (2010, Lemma A.3) for the univariate ($p = 1$) result. ■

A.2 Bounds on product moments

Our proof of tightness applies the result of Kallenberg (2001) in Lemma A.1 and involves evaluation of the 4th moment of product moments of linear processes. We give a number of evaluations of such moments in terms of the quantity

$$\xi_T(\zeta_1, \zeta_2) = \max_{1 \leq n, m \leq T} \sum_{t=\max(n, m)}^T |\zeta_{1, t-n} \zeta_{2, t-m}|, \quad (51)$$

where $\zeta_{1n}, \zeta_{2n}, n = 0, 1, \dots$, are real coefficients.

Lemma A.4 *For $i = 1, 2$, let ε_{it} be i.i.d. $(0, \sigma_i^2)$ with $E|\varepsilon_{it}|^8 < \infty$. Assume that ξ_{in} are real coefficients satisfying $\sum_{n=0}^{\infty} |\xi_{in}| < \infty$, and define $Z_{it}^+ = \sum_{n=0}^{t-1} \xi_{in} \varepsilon_{it-n}$. Let ζ_{1n}, ζ_{2n} be real coefficients, then*

$$\|T^{-1} \sum_{t=1}^T \left(\sum_{n=0}^{t-1} \zeta_{1n} Z_{1t-n}^+ \right) \left(\sum_{m=0}^{t-1} \zeta_{2m} Z_{2t-m}^+ \right)\|_4 \leq c \xi_T(\zeta_1, \zeta_2). \quad (52)$$

Proof. We find

$$E \left[T^{-1} \sum_{t=1}^T \sum_{n=1}^t \sum_{m=1}^t \zeta_{1, t-n} Z_{1n}^+ \zeta_{2, t-m} Z_{2m}^+ \right]^4 = T^{-4} \sum_{(1)} \left[\prod_{k=1}^4 \zeta_{1, t_k - n_k} \zeta_{2, t_k - m_k} \right] E \left[\prod_{k=1}^4 Z_{1n_k}^+ Z_{2m_k}^+ \right],$$

where the summation $\sum_{(1)}$ is over $1 \leq n_k, m_k \leq t_k \leq T, k = 1, 2, 3, 4$. This on the other hand is bounded by

$$\begin{aligned} & \left(\prod_{k=1}^4 \max_{n_k, m_k} \sum_{t_k=\max(n_k, m_k)}^T |\zeta_{1, t_k - n_k} \zeta_{2, t_k - m_k}| \right) T^{-4} \sum_{(2)} |E \prod_{k=1}^4 Z_{1n_k}^+ Z_{2m_k}^+| \\ & \leq \xi_T^4(\zeta_1, \zeta_2) T^{-4} \sum_{(2)} |E \prod_{k=1}^4 Z_{1n_k}^+ Z_{2m_k}^+|, \end{aligned}$$

where the summation $\sum_{(2)}$ is over $1 \leq n_k, m_k \leq T, k = 1, 2, 3, 4$.

We want to show that the second factor is bounded, and find, using $Z_{it}^+ = \sum_{n=0}^{t-1} \xi_{in} \varepsilon_{it-n}$,

$$\sum_{(2)} E \left[\prod_{k=1}^4 Z_{1n_k}^+ Z_{2m_k}^+ \right] = \sum_{(3)} \left[\prod_{k=1}^4 \xi_{1i_k} \xi_{2j_k} \right] E \left[\prod_{k=1}^4 \varepsilon_{1n_k - i_k} \varepsilon_{2m_k - j_k} \right].$$

The sum $\sum_{(3)}$ extends over $0 \leq i_k < n_k \leq T$, $0 \leq j_k < m_k \leq T$, $k = 1, 2, 3, 4$. The number of terms is proportional to T^{16} , but most are zero because $E(\varepsilon_{it}) = 0$. We get a contribution if the eight subscripts $n_k - i_k, m_k - j_k$ are equal in pairs, triples, or more. If the indices are equal in four pairs (2, 2, 2, 2) there are four constraints, the combinations (3, 3, 2) and (4, 2, 2) give five constraints, whereas (4, 4) gives six constraints, and finally all equal, (8), gives seven constraints. Thus we get the fewest constraints with (2, 2, 2, 2) and hence the largest number of terms. The four constraints leave twelve summations. Next note that $\sum_{i=0}^{\infty} |\xi_{1i}| < \infty$, and $\sum_{i=0}^{\infty} |\xi_{2i}| < \infty$, so the eight summations over the indices $i_j, j_k, k = 1, 2, 3, 4$ are finite, which leaves four summations. However, with four summations we can get at most T^4 terms, which shows that $T^{-4} \sum_{(2)} |E \prod_{k=1}^4 Z_{1n_k}^+ Z_{2m_k}^+|$ is bounded by a coefficient that only depends on ξ_{1i} and ξ_{2i} , and not on T, ζ_{1n} , or ζ_{2m} , which proves (52). ■

Lemma A.5 For $|u| \leq u_0$ and all $j \geq 1$ it holds that

$$|\mathbf{D}^m \pi_j(-u)| \leq c(u_0)(1 + \log j)^m j^{-u-1}, \quad (53)$$

$$|\mathbf{D}^m T^u \pi_j(-u)| \leq c(u_0) T^u (1 + |\log \frac{j}{T}|)^m j^{-u-1}, \quad (54)$$

uniformly in u .

Proof. See JN (2010, Lemma B.3). ■

The next Lemma is the key result on the evaluation of $\xi_T(\zeta_1, \zeta_2)$ and hence the empirical moments for a class of processes defined by coefficients $(\zeta_{1n}(a_1), \zeta_{2n}(a_2))$ satisfying conditions of the type

$$|\zeta_{1,0}(a)| \leq 1, \quad |\zeta_{1n}(a)| \leq c(1 + \log n)^{m_1} n^{-a-1}, \quad n \geq 1, \quad (55)$$

$$|\zeta_{1,0}^*(a)| \leq 1, \quad |\zeta_{1n}^*(a)| \leq cT^{a+1/2} (1 + \log \frac{n}{T})^{m_1} n^{-a-1}, \quad n \geq 1, \quad (56)$$

where c does not depend on a or n . These inequalities are satisfied by the fractional coefficients and their derivatives, see Lemma A.5

We repeatedly use the elementary inequalities, for $\kappa \geq 0$,

$$\sum_{n=1}^T n^{-u-1} \leq c\kappa^{-1}, \quad u \geq \kappa, \quad (57)$$

$$c\kappa^{-1}(1 - T^{-\kappa}) \leq \sum_{n=1}^T n^{-u-1}, \quad u \leq \kappa. \quad (58)$$

Lemma A.6 Let $\zeta_{1n}(a_1), \zeta_{2n}(a_2), \zeta_{1n}^*(a_1)$, and $\zeta_{2n}^*(a_2)$ satisfy (55)–(56), and let $-1 \leq a_i \leq a_0, i = 1, 2$, and $0 < \kappa < 1/2$. Then, for any a ,

(i) Uniformly for $\min(a_1 + 1, a_2 + 1, a_1 + a_2 + 1) \geq a$ we have

$$\xi_T(\zeta_1(a_1), \zeta_2(a_2)) \leq c \begin{cases} (1 + \log T)^{m_1+m_2+1} T^{-a}, & a \leq 0, \\ a^{-1}, & a > 0. \end{cases} \quad (59)$$

(ii) Uniformly for $\max(a_1, a_2, a_1 + a_2 + 1) \leq -\kappa$ we have

$$\xi_T(\zeta_1^*(a_1), \zeta_2^*(a_2)) \leq c\kappa^{-1}. \quad (60)$$

(iii) Uniformly for $a_1 \geq -1/2 + a$ and $a_2 \leq -1/2 - \kappa$ we have

$$\xi_T(\zeta_1(a_1), \zeta_2^*(a_2)) \leq c(1 + \log T)^{m_1+m_2+1} T^{-\min(a, \kappa)} \quad (61)$$

Proof. In evaluating (51) we focus on terms with $t > \max(m, n)$, because the analysis with $t = m$ or $t = n$ is straightforward.

Proof of (59): For $t > \max(m, n)$ we first apply (55) and therefore bound the summation $\sum_{t=\max(n, m)+1}^T |\zeta_{1, t-n}(a_1) \zeta_{2, t-m}(a_2)|$ by

$$\sum_{t=\max(n, m)+1}^T c(1 + \log(t-n))^{m_1} (t-n)^{-a_1-1} c(1 + \log(t-m))^{m_2} (t-m)^{-a_2-1}.$$

For $a \leq 0$, we bound the log factors by $(1 + \log T)$ and $(t-n)^{-a_1-1} (t-m)^{-a_2-1} \leq (t - \max(n, m))^{-(a_1+a_2+1)-1}$. Then the bound for $\xi_T(\zeta_1(a_1), \zeta_2(a_2))$ follows because

$$\sum_{t=\max(n, m)+1}^T (t - \max(n, m))^{-a-1} \leq c(\log T) T^{-a} \text{ for } a \leq 0.$$

For $a > 0$ we bound $(1 + \log(t-n))^{m_1} (t-n)^{-a/3}$ and $(1 + \log(t-m))^{m_2} (t-m)^{-a/3}$ by a constant. Then $\xi_T(\zeta_1(a_1), \zeta_2(a_2))$ is by (57) bounded by

$$\max_{1 \leq n, m \leq T} \sum_{t=\max(n, m)+1}^T (t - \max(n, m))^{-a+2a/3-1} \leq ca^{-1}.$$

Proof of (60): We find that $\xi_T(\zeta_1^*(a_1), \zeta_2^*(a_2))$ is bounded by a constant times

$$\begin{aligned} & T^{-1} \max_{1 \leq n, m \leq T} \sum_{t=\max(n, m)+1}^T (1 + \log(\frac{t-n}{T}))^{m_1} (\frac{t-n}{T})^{-(a_1+1)} (1 + \log(\frac{t-m}{T}))^{m_2} (\frac{t-m}{T})^{-(a_2+1)} \\ \rightarrow & \max_{0 \leq x, y \leq 1} \int_{\max(x, y)}^1 (1 + \log(s-x))^{m_1} (s-x)^{-(a_1+1)} (1 + \log(s-y))^{m_2} (s-y)^{-(a_2+1)} ds \end{aligned}$$

for $T \rightarrow \infty$. This is uniformly bounded by $c\kappa^{-1}$ if $\max(a_1, a_2, a_1 + a_2 + 1) \leq -\kappa$.

Proof of (61): We evaluate the log factors by $(1 + \log T)$ and $T^{a_2+1/2} (t-m)^{-(a_2+1/2+\kappa)} \leq T^{a_2+1/2} T^{-(a_2+1/2+\kappa)} = T^{-\kappa}$. Because $a_1 + 1 \geq 0$ and $1/2 - \kappa > 0$ we find that the remaining terms in the summation are bounded as

$$(t-n)^{-a_1-1} (t-m)^{-1/2+\kappa} \leq (t - \max(n, m))^{-a_1-1-1/2+\kappa} \leq (t - \max(n, m))^{-a-1+\kappa},$$

where the last inequality follows from $-a_1 \leq 1/2 - a$. Summing over t gives the bound $T^{-\kappa} T^{\max(-a+\kappa, 0)} = T^{-\min(a, \kappa)}$. ■

Lemma A.7 *Let ε_t be i.i.d. $(0, \Omega)$ with $E|\varepsilon_t|^8 < \infty$ and define for $t > N$ the independent processes $U_{it}^+ = \sum_{n=0}^{N-1} \zeta_{in} \varepsilon_{t-n}$ and $V_{it}^+ = \sum_{m=N}^{t-1} \zeta_{im} \varepsilon_{t-m}$. Then*

$$\|T^{-1} \sum_{t=N+1}^T U_{1t}^+ V_{2t}^{+'}\|_4^4 \leq cNT^{-1} \xi_N(\zeta_1, \zeta_1) \xi_T(\zeta_1, \zeta_1) \xi_T(\zeta_2, \zeta_2)^2, \quad (62)$$

$$\|T^{-1} \sum_{t=N+1}^T U_{1t}^+ U_{2t}^{+'}\|_4^4 \leq cN^3 T^{-3} \max(\xi_N(\zeta_1, \zeta_1), \xi_N(\zeta_2, \zeta_2), \xi_N(\zeta_1, \zeta_2))^4, \quad (63)$$

$$|\text{Var}(T^{-1} \sum_{t=N+1}^T U_{1t}^+ U_{2t}^{+'})| \leq NT^{-1} \xi_N(\zeta_1, \zeta_1) \xi_N(\zeta_2, \zeta_2). \quad (64)$$

Proof. The coordinate processes in each row of U_{it}^+ and V_{it}^+ have the same structure, but with different ε , so we prove the result for univariate i.i.d. $(\varepsilon_{1t}, \varepsilon_{2t})$ with general variance matrix $\{\sigma_{ij}\}_{i,j=1,2}$.

Proof of (62): We find

$$E[T^{-1} \sum_{t=N+1}^T \sum_{n=0}^{N-1} \sum_{m=N}^{t-1} \zeta_{1n} \varepsilon_{1t-n} \zeta_{2m} \varepsilon_{2t-m}]^4 = T^{-4} \sum_{(1)} \left[\prod_{k=1}^4 \zeta_{1n_k} \zeta_{2m_k} \right] E \left[\prod_{k=1}^4 \varepsilon_{1t_k - n_k} \varepsilon_{2t_k - m_k} \right],$$

where the summation $\sum_{(1)}$ is over $0 \leq n_k < N \leq m_k < t_k \leq T$, $k = 1, 2, 3, 4$. However, $E[\prod_{k=1}^4 \varepsilon_{t_k - n_k} \varepsilon_{t_k - m_k}]$ is non-zero only if the subscripts are equal in pairs. Moreover, because $n_k < N \leq m_k$ we have $t_k - n_k > t_k - m_k$. This means that not all pairs can be of the form $t_i - n_i = t_j - m_j$ because $\sum_{i=1}^4 t_i - n_i > \sum_{j=1}^4 t_j - m_j$. Thus there is always at least one pair of the form $t_i - n_i = t_j - n_j$ and one of the form $t_k - m_k = t_l - m_l$. We then find $n_j = t_j - t_i + n_i$ and the summation over (n_i, n_j) gives $\sigma_{11} |\sum_{n_i=1}^{N-1} \zeta_{1, n_i+h} \zeta_{1n_i}| \leq \sigma_{11} \xi_N(\zeta_1, \zeta_1)$ and the restriction $|t_i - t_j| = |n_i - n_j| \leq N$. Similarly the pair $t_k - m_k = t_l - m_l$ gives rise to the contribution $\sigma_{22} |\sum_{m_k=N}^T \zeta_{2, m_k+h} \zeta_{2m_k}| \leq \sigma_{22} \xi_T(\zeta_2, \zeta_2)$ and the restriction $|t_k - t_l| = |m_k - m_l| \leq T - N$ from the summation over (m_k, m_l) .

If the two remaining pairs are matched similarly we get the factor $\xi_N(\zeta_1, \zeta_1)^2 \xi_T(\zeta_2, \zeta_2)^2$. What remains is the summation over (t_1, \dots, t_4) with two restrictions of the form $|t_i - t_j| \leq N$ and two of the form $|t_k - t_l| \leq T - N$. The number of such terms is bounded by $N^2 T^2$, and we find the total contribution from $\sum_{(1)}$ is

$$cN^2 T^2 \xi_N(\zeta_1, \zeta_1)^2 \xi_T(\zeta_2, \zeta_2)^2 \leq cT^3 N \xi_N(\zeta_1, \zeta_1) \xi_T(\zeta_1, \zeta_1) \xi_T(\zeta_2, \zeta_2)^2.$$

If the last two pairs have the form $t_i - n_i = t_j - m_j$ we find $n_i = t_i - t_j + m_j$ and the summation over (n_i, m_j) becomes

$$|\sigma_{12}| \left| \sum_{m_j=N}^T \zeta_{1, m_j+h} \zeta_{2m_j} \right| \leq |\sigma_{12}| \xi_T(\zeta_1, \zeta_2),$$

and the restriction on (t_1, \dots, t_4) becomes $|t_i - t_j| \leq T - N$. The contribution in this case is therefore at most

$$c \xi_N(\zeta_1, \zeta_1) \xi_T(\zeta_2, \zeta_2) \xi_T(\zeta_1, \zeta_2)^2 \leq c \xi_N(\zeta_1, \zeta_1) \xi_T(\zeta_1, \zeta_1) \xi_T(\zeta_2, \zeta_2)^2,$$

and the number of terms in the summation over (t_1, \dots, t_4) with only one restriction $|t_i - t_j| \leq N$ is NT^3 , which proves the result.

Proof of (63): The same calculation gives

$$E[T^{-1} \sum_{t=N+1}^T \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \zeta_{1n} \varepsilon_{1t-n} \zeta_{2m} \varepsilon_{2t-m}]^4 = T^{-4} \sum_{(2)} \prod_{k=1}^4 \zeta_{1n_k} \zeta_{2m_k} E[\prod_{k=1}^4 \varepsilon_{1t_k - n_k} \varepsilon_{2t_k - m_k}],$$

where the summation $\sum_{(2)}$ is over $0 \leq n_k, m_k < N < t_k \leq T$, $k = 1, 2, 3, 4$.

Again we have to match the subscripts $t_k - n_k, t_k - m_k$ in pairs and we find that any match will give a contribution of the form $\xi_N(\zeta_1, \zeta_1), \xi_N(\zeta_1, \zeta_2)$, or $\xi_N(\zeta_2, \zeta_2)$ but always with a restriction on (t_1, \dots, t_4) of the form $|t_i - t_j| \leq N$.

Thus we find the contribution is bounded by $\max(\xi_N(\zeta_1, \zeta_1), \xi_N(\zeta_1, \zeta_2), \xi_N(\zeta_2, \zeta_2))^4$ and the number of terms in the summation over (t_1, \dots, t_4) with four restrictions of the form $|t_i - t_j| \leq N$ is N^3T . This proves the result.

Proof of (64): We find as above

$$E[T^{-1} \sum_{t=N+1}^T \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \zeta_{1n} \varepsilon_{1t-n} \zeta_{2m} \varepsilon_{2t-m}]^2 = T^{-2} \sum_{(3)} \prod_{k=1}^2 \zeta_{1n_k} \zeta_{2m_k} E[\prod_{k=1}^2 \varepsilon_{1t_k - n_k} \varepsilon_{2t_k - m_k}],$$

where the summation $\sum_{(3)}$ is over $0 \leq n_k, m_k < N < t_k \leq T$, $k = 1, 2$. We get a contribution if the indices $t_k - n_k, t_k - m_k$ are equal in pairs and we find for $t_1 - n_1 = t_1 - m_1$ and $t_2 - n_2 = t_2 - m_2$ the contribution

$$\sigma_{12}^2 \left(\frac{T-N}{T} \right)^2 \prod_{k=1}^2 \sum_{n_k=0}^{N-1} \zeta_{1n_k} \zeta_{2n_k} = [E(T^{-1} \sum_{t=N+1}^T U_{1t}^+ U_{2t}^+)]^2,$$

which is the expectation squared which subtracted gives the variance.

The other terms we find if

$$\begin{aligned} t_1 - n_1 &= t_2 - n_2 \text{ and } t_1 - m_1 = t_2 - m_2, \\ t_1 - n_1 &= t_2 - m_2 \text{ and } t_1 - m_1 = t_2 - n_2. \end{aligned}$$

The first gives for $n_1 = t_1 - t_2 + n_2$ and $m_1 = t_1 - t_2 + m_2$

$$c \sum_{n_2=0}^{N-1} \zeta_{1n_2} \zeta_{1, n_2 + t_1 - t_2} \sum_{m_2=0}^{N-1} \zeta_{2m_2} \zeta_{2, m_2 + t_1 - t_2} \leq c \xi_N(\zeta_1, \zeta_1) \xi_N(\zeta_2, \zeta_2),$$

and summing over (t_1, t_2) with the restriction $|t_1 - t_2| \leq N$ we find that the contribution from $\sum_{(3)}$ is

$$cNT \xi_N(\zeta_1, \zeta_1) \xi_N(\zeta_2, \zeta_2).$$

In the other case we get

$$cNT \xi_N(\zeta_1, \zeta_2) \xi_N(\zeta_1, \zeta_2) \leq cNT \xi_N(\zeta_1, \zeta_1) \xi_N(\zeta_2, \zeta_2),$$

which was to be proved. ■

A.3 Limit theory for product moments of deterministic terms

The next lemma gives the results for the impact of the initial values $D_{it}(\psi) = \Delta_+^{d+ib}\mu_t + \Delta_-^{d+ib}X_t$, see (30), in the models considered, using the bounds in JN (2010, Lemma C.1). We let \mathbf{D}^m be the derivative with respect to $d + ib$.

Lemma A.8 *For $\kappa_0 > 0$ and $\kappa_1 > 0$ we have*

(i) *For $\delta_i = d + ib - d_0$, $\eta_i = d + ib - d_0 + b_0$, $D_{it}(\psi) = \Delta_+^{d+ib}\mu_t + \Delta_-^{d+ib}X_t$, and $d - b \geq \kappa_0$,*

$$\max_{-1/2 - \kappa_1 \leq \delta_i \leq u_1} |\mathbf{D}^m \beta'_{0\perp} D_{it}(\psi)| \rightarrow 0 \text{ as } t \rightarrow \infty, \quad (65)$$

$$\max_{-u_0 \leq \delta_i \leq -1/2 - \kappa_1} \max_{1 \leq t \leq T} |\mathbf{D}^m T^{\delta_i + 1/2} \beta'_{0\perp} D_{it}(\psi)| \rightarrow 0 \text{ as } T \rightarrow \infty, \quad (66)$$

$$\max_{-1/2 - \kappa_1 \leq \eta_i \leq u_1} |\mathbf{D}^m \beta'_0 D_{it}(\psi)| \rightarrow 0 \text{ as } t \rightarrow \infty, \quad (67)$$

$$\max_{-u_0 \leq \eta_i \leq -1/2 - \kappa_1} \max_{1 \leq t \leq T} |\mathbf{D}^m T^{\eta_i + 1/2} \beta'_0 D_{it}(\psi)| \rightarrow 0 \text{ as } T \rightarrow \infty. \quad (68)$$

(ii) *In the model with $d = d_0$, the results (65)–(68) hold.*

(iii) *If the initial values satisfy $X_{-n} = 0$, $n > N_0$, then (65)–(68) hold for $d - b \geq 0$.*

(iv) *In the model with $d = b$, (65)–(68) hold.*

Proof. From (6) and Theorem 3 we find

$$\begin{aligned} \mu_t &= -\Pi_+(L)^{-1} \Pi_-(L) X_t = -(C_0 \Delta_+^{-d_0} + \Delta_+^{-d_0 + b_0} F_+(L)) \sum_{j=-1}^k \Psi_{0i} \Delta_-^{d_0 + j b_0} X_t \\ &= -\sum_{j=0}^k C_0 \Psi_{0i} \Delta_+^{-d_0} \Delta_-^{d_0 + j b_0} X_t - \sum_{j=-1}^k F_+(L) \Psi_{0i} \Delta_+^{-d_0 + b_0} \Delta_-^{d_0 + j b_0} X_t \end{aligned} \quad (69)$$

because $C_0 \Psi_{0,-1} = -C_0 \alpha_0 \beta'_0 = 0$. Therefore, $D_{it}(\psi) = \Delta_+^{d+ib}\mu_t + \Delta_-^{d+ib}X_t$ is a linear combination of terms $G_+(L) \Delta_+^u \Delta_-^v X_t$ where u and v are defined by either $u = d + ib - \gamma_0$, $v = d_0 + j b_0 \geq \gamma_0$ ($\gamma_0 = d_0$ or $\gamma_0 = d_0 - b_0$) for $\Delta_+^{d+ib-\gamma_0} \Delta_-^{d_0 + j b_0} X_t$ or by $u = 0$, $v = d + ib \geq \kappa_0$ for $\Delta_-^{d+ib} X_t$.

From JN (2010, Lemma C.1) with $G(z) = \sum_{n=0}^{\infty} g_n z^n$ and $\sum_{n=0}^{\infty} |g_n| < \infty$ we have

$$\max_{\kappa_1 \leq \min(u+v, u+1, v) \leq u_1} |G_+(L) \mathbf{D}^m \Delta_+^u \Delta_-^v X_t| \rightarrow 0 \text{ as } t \rightarrow \infty, \quad (70)$$

$$\max_{\kappa_1 \leq \min(v-1/2, -u-1/2) \leq u_1} \max_{1 \leq t \leq T} |G_+(L) \mathbf{D}^m T^{u+1/2} \Delta_-^u \Delta_-^v X_t| \rightarrow 0 \text{ as } T \rightarrow \infty, \quad (71)$$

$$\max_{\kappa_0 \leq v \leq v_0} |\mathbf{D}^m \Delta_-^v X_t| \rightarrow 0 \text{ as } t \rightarrow \infty, \quad (72)$$

$$\max_{\kappa_0 \leq v \leq a - \kappa} \max_{1 \leq t \leq T} |\mathbf{D}^m T^{-a+v} \Delta_-^v X_t| \rightarrow 0 \text{ as } T \rightarrow \infty. \quad (73)$$

Proof of (65) and (66): The term $\Delta_+^{d+ib} \beta'_{0\perp} \mu_t$ contains terms of the form $G(L) \Delta_+^u \Delta_-^v X_t$ where $u = d + ib - \gamma_0$ and $v = d_0 + j b_0 \geq \gamma_0$ with $\gamma_0 = d_0$ or $\gamma_0 = d_0 - b_0$, see (69). If $d + ib - \gamma_0 \geq -1/2 - \kappa_1$ in (65), then for both choices of γ_0 we find $\min(u + v, u + 1, v) \geq \min(\kappa_0, 1/2 - \kappa_1, d_0 - b_0) > 0$, and the result (65) follows from (70).

If $d+ib-\gamma_0 \leq -1/2-\kappa_1$ in (66), then $v \geq -u$ so that $\min(v-1/2, -u-1/2) = -u-1/2 \geq \kappa_1 > 0$, and the result follows from (71) because the normalization $T^{d+ib-d_0+1/2}$ of $\Delta_+^{d+ib}\beta'_{0\perp}\mu_t$ is enough for both choices of γ_0 to ensure $\max_{1 \leq t \leq T} |T^{d+ib-d_0+1/2}\Delta_+^{d+ib}\beta'_{0\perp}\mu_t| \rightarrow 0$. Note that the condition $d-b \geq \kappa_0$ is not used in this case.

The term $\Delta_-^{d+ib}\beta'_{0\perp}X_t$ has $d+ib \geq d-b \geq \kappa_0$, so that (65) follows from (72) and (66) follows from (73).

Proof of (67) and (68): The term $\Delta_+^{d+ib}\beta'_0\mu_t$ contains terms of the form $G(L)\Delta_+^u\Delta_-^vX_t$ but only from the second sum in (69) because $\beta'_0C_0 = 0$. Thus, $u = d+ib-d_0+b_0$ and $v = d_0+jb_0$. If $u \geq -1/2-\kappa_1$, then $u+v \geq d-b \geq \kappa_0$ so that $\min(u+v, u+1, v) \geq \min(\kappa_0, 1/2-\kappa_1, d_0-b_0) > 0$, and (67) follows from (70).

If instead $u \leq -1/2-\kappa_1$ then from $v+u \geq 0$ we find $\min(v-1/2, -u-1/2) = -u-1/2 \geq \kappa_1 > 0$, and (68) follows from (71). Note again that $d-b \geq \kappa_0$ was not used in this case.

The term $\Delta_-^{d+ib}\beta'_0X_t$. For $d+ib \geq d-b \geq \kappa_0$ the result follows from (72) and for $T^{d+ib-d_0+b_0+1/2}\Delta_-^{d+ib}\beta'_0X_t$ we have $d+ib \leq d_0-b_0-1/2-\kappa_1$ and the result follows from (73).

Proof of (ii): This follows because the model with $d = d_0$ is a submodel of \mathcal{H}_r .

Proof of (iii): It is seen from the proof of (i) that the condition $d-b \geq \kappa_0$ is only used for terms with $i = -1$. We thus have to prove that if only finitely many initial values are nonzero, (65)–(68) hold for $i = -1$. For simplicity we set $m = 0$. From JN (2010, Lemma C.1) we get the evaluations for $v \geq 0$

$$|\Delta_-^vX_t| = \left| \sum_{n=0}^{N_0} \pi_{n+t}(-v)X_{-n} \right| \leq cN_0t^{-1}, \quad (74)$$

$$|T^v\Delta_-^vX_t| = |T^v \sum_{n=0}^{N_0} \pi_{n+t}(-v)X_{-n}| \leq cT^v, \quad (75)$$

$$|\Delta_+^u\Delta_-^vX_t| = \left| \sum_{n=0}^{N_0} \left(\sum_{j=0}^{t-1} \pi_j(-u)\pi_{n+t-j}(-v) \right) X_{-n} \right| \leq cN_0 \sum_{j=0}^{t-1} j^{-u-1}(t-j)^{-v-1}, \quad (76)$$

We start with the term $\Delta_-^{d-b}X_t$. It follows from (74) that (65) and (67) hold. Similarly from (75) that (66) and (68) hold for the term $\Delta_-^{d-b}X_t$.

Next we analyze $\Delta_+^{d-b}\mu_t$, which is composed of terms of the form $\Delta_+^u\Delta_-^vX_t$ for $u = d-b-\gamma_0$ and $v = d_0+jb_0 \geq \gamma_0$. For such a term we apply (76) and find from JN (2010, Lemma B.4, equation (62)) that it is bounded by $t^{-\min(u+v+1, u+1, v+1)}$ which tends to zero because $\min(u+v+1, u+1, v+1) \geq \min(1, 1/2-\kappa_1, 1+\gamma_0) > 0$. This proves (65) and (67) for the term $\Delta_+^{d-b}\mu_t$. For (66) and (68) we need not prove anything because the condition $d-b \geq \kappa_0$ was not used in the proof of these in case (i).

Proof of (iv): In case $d = b$ we only have to analyze the terms with $i = -1$. We find $\Delta_-^{d-b}X_t = \Delta_0^0X_t = 0$ which leaves only the term $\Delta_+^{d-b}\mu_t = \mu_t$. Because the condition $d-b \geq \kappa_0$ was not used in the proofs of (66) and (68) for the terms $T^{\delta_{-1}+1/2}\beta'_{0\perp}\mu_t$ and $T^{\eta_{-1}+1/2}\beta'_0\mu_t$ we only have to consider (65) and (67).

In (65) we assume $\delta_{-1} = d-b-d_0 = -d_0 \geq -1/2-\kappa_1$, which is not possible for κ_1 sufficiently small, so there is nothing to prove.

For (67), when $d_0 = b_0$, we find from (69) that

$$\beta'_0 \mu_t = - \sum_{j=-1}^k \beta'_0 F_+(L) \Psi_{0i} \Delta_-^{(1+j)d_0} X_t,$$

where the term with $j = -1$ is zero because $\Delta_-^0 X_t = 0$. The remaining terms $\Delta_-^{(1+j)d_0} X_t$ tend to zero by (72), which proves (67). ■

A.4 Limit theory for product moments of stochastic terms

In this section we analyze product moments of processes that are either asymptotically stationary, near critical, or nonstationary and we first define the corresponding fractional indices.

Definition A.9 *We take three fractional indices w, v , and u in the intervals*

$$[-w_0, -1/2 - \kappa_w], [-1/2 - \underline{\kappa}_v, -1/2 + \bar{\kappa}_v], \text{ and } [-1/2 + \kappa_u, u_0], \quad (77)$$

respectively, where we assume $0 \leq \bar{\kappa}_v$ and $\underline{\kappa}_v < \min(\kappa_w, \kappa_u) < 1/2$.

In the following we assume these bounds on (u, v, w) . In the applications we always choose fixed values of κ_u and κ_w , but we shall sometimes choose small values ($\rightarrow 0$) of $\underline{\kappa}_v$ or $\bar{\kappa}_v$.

Definition A.10 *We define the class \mathcal{Z} as the set of multivariate stochastic processes Z_t for which*

$$Z_t = \xi \varepsilon_t + \Delta^{1/2} \sum_{n=0}^{\infty} \xi_n^* \varepsilon_{t-n},$$

where ε_t is i.i.d. $(0, \Omega)$ and the coefficient matrices satisfy $\sum_{n=0}^{\infty} |\xi_n^| < \infty$ and $\text{Var}(\xi \varepsilon_t) > 0$.*

This is a fractional version of the usual Beveridge-Nelson decomposition, where $\sum_{n=0}^{\infty} \xi_n \varepsilon_{t-n} = (\sum_{n=0}^{\infty} \xi_n) \varepsilon_t + \Delta \sum_{n=0}^{\infty} \xi_n^* \varepsilon_{t-n}$. It follows from Theorem 3 and (10) that $\beta'_0 (C_0 \varepsilon_t + \Delta^{b_0} Y_t) = \Delta^{b_0} \beta'_0 Y_t \in \mathcal{Z}$ and that $Y_t \in \mathcal{Z}$. Thus for $Z_t \in \mathcal{Z}$ and indices (w, v, u) as in Definition A.9, $\Delta_+^w Z_t^+$ is nonstationary, $\Delta_+^u Z_t^+$ is asymptotically stationary, and $\Delta_+^v Z_t^+$ is close to a critical process of the form $\Delta_+^{-1/2} \varepsilon_t$. We derive in Lemma A.11 and Corollary A.12 results for product moments of fractional differences of processes in \mathcal{Z} .

For $m = m_1 + m_2$ we define the product moments

$$\begin{aligned} \mathbf{D}^m M_T(a_1, a_2) &= T^{-1} \sum_{t=1}^T (\mathbf{D}^{m_1} \Delta_+^{a_1} Z_{1t}^+) (\mathbf{D}^{m_2} \Delta_+^{a_2} Z_{2t}^+)', \\ M_T((a_1, a_2), (a_1, a_2)) &= T^{-1} \sum_{t=1}^T \begin{pmatrix} \Delta_+^{a_1} Z_{1t}^+ \\ \Delta_+^{a_2} Z_{2t}^+ \end{pmatrix} \begin{pmatrix} \Delta_+^{a_1} Z_{1t}^+ \\ \Delta_+^{a_2} Z_{2t}^+ \end{pmatrix}', \end{aligned}$$

etc. Let N_T be a normalizing sequence and define $M_T(a_1, a_2) = \mathbf{O}_P(N_T)$ on a compact set K to mean that $N_T^{-1} M_T(a_1, a_2)$ is tight on K and $M_T(a_1, a_2) = \mathbf{o}_P(N_T)$ to mean that $N_T^{-1} M_T(a_1, a_2) \implies 0$ on K .

Lemma A.11 *Let $Z_{it} = \xi_i \varepsilon_t + \Delta^{1/2} \sum_{n=0}^{\infty} \xi_{in}^* \varepsilon_{t-n} \in \mathcal{Z}, i = 1, 2$, and define $M_T(a_1, a_2)$ as above and assume that $E|\varepsilon_t|^q < \infty$ for $q > \kappa_w^{-1}$. Then uniformly for (w, v, u) in the sets defined in (77) with $0 \leq \underline{\kappa}_v \leq 1/7$ we find*

$$\mathbf{D}^m M_T(u_1, u_2) \implies \mathbf{D}^m E(\Delta^{u_1} Z_{1t})(\Delta^{u_2} Z_{2t})', \quad (78)$$

$\mathbf{D}^m M_T(w_1, w_2) T^{w_1+w_2+1}$ is tight and

$$M_T(w_1, w_2) T^{w_1+w_2+1} \implies \int_0^1 W_{-w_1-1}(s) W_{-w_2-1}(s)' ds, \quad (79)$$

$$\mathbf{D}^m M_T(w, u) T^{w+1/2} = \mathbf{O}_P((1 + \log T)^{2+m} T^{-\min(\kappa_u, \kappa_w)}), \quad (80)$$

$$M_T(v, u) = \mathbf{O}_P(1) \text{ when } \underline{\kappa}_v \leq \kappa_u/2, \quad (81)$$

$$M_T(w, v) T^{w+1/2} = \mathbf{O}_P((1 + \log T)^2 T^{\underline{\kappa}_v}). \quad (82)$$

Finally if $N = T^\alpha, \alpha < 1/9$, and (ξ'_1, ξ'_2) has full rank, then

$$M_T((v_1, v_2), (v_1, v_2))^{-1} = \mathbf{O}_P\left(\frac{2\bar{\kappa}_v}{1 - N^{-2\bar{\kappa}_v}}\right). \quad (83)$$

Proof. A matrix valued process $\mathbf{D}^m M_T(a_1, a_2)$ is tight if the coordinate processes are tight, and the (i, j) 'th coordinate is a finite sum of univariate processes constructed the same way, so it is enough to prove the result for univariate processes. We prove tightness by checking condition (50) of Lemma A.1 for $\mathbf{D}^m M_T(a_1, a_2)$. The moments are evaluated by $\xi_T(\zeta_1, \zeta_2)$, see (52), for suitable coefficients satisfying (55) and (56).

We introduce the notation $M_T^{**}(w_1, w_2) = T^{w_1+w_2+1} M_T(w_1, w_2)$ to indicate that the non-stationary processes have been normalized by $T^{w_i+1/2}$. We give the proofs for $m_1 = m_2 = 0$, as the extra factors of $(1 + \log T)^{m_i}$ do not change the evaluations.

Proof of (78): We define the coefficients $\zeta_{i,t-n} = \pi_{t-n}(-u_i)$, which satisfy condition (55). The assumption that $u_i \geq -1/2 + \kappa_u$ implies $\min(u_1 + u_2 + 1, u_1 + 1, u_2 + 1) \geq 2\kappa_u$, so we can apply (52) and (59) which shows that $\|M_T(u_1, u_2)\|_4 \leq c$.

Next we consider $\|M_T(u_1, u_2) - M_T(\tilde{u}_1, \tilde{u}_2)\|_4$ which we bound by

$$\|T^{-1} \sum_{t=1}^T (\Delta_+^{u_1} Z_{1t}^+ - \Delta_+^{\tilde{u}_1} Z_{1t}^+) (\Delta_+^{u_2} Z_{2t}^+)'\|_4 + \|T^{-1} \sum_{t=1}^T (\Delta_+^{\tilde{u}_1} Z_{1t}^+) (\Delta_+^{u_2} Z_{2t}^+ - \Delta_+^{\tilde{u}_2} Z_{2t}^+)'\|_4. \quad (84)$$

We apply (52) to the first term with $\zeta_{1,t-n} = (\pi_{t-n}(-u_1) - \pi_{t-n}(-\tilde{u}_1))$ and $\zeta_{2,t-n} = \pi_{t-n}(-u_2)$ bounded by (55), see also JN (2010, Lemma B.3), and it follows from (59) with $a = 2\kappa_u$ that the first term of (84) is bounded by $c(u_1 - \tilde{u}_1)$. A similar proof works for the other term of (84), and tightness then follows from (50).

Notice that the second condition of (50) follows in the same way as the first using the inequalities in Lemma A.6. The only difference is an extra log factor and the factor $(u_1 - \tilde{u}_1)$.

We next apply the law of large numbers to identify the limit as an expectation. From $\Delta^{u_i} Z_{it} = \Delta_+^{u_i} Z_{it} + \Delta_-^{u_i} Z_{it} = \sum_{j=0}^{\infty} \varsigma_{ij}(-u_i) \varepsilon_{t-j}$ we find

$$\begin{aligned} M_T(u_1, u_2) &= T^{-1} \sum_{t=1}^T \Delta^{u_1} Z_{1t} \Delta^{u_2} Z'_{2t} + T^{-1} \sum_{t=1}^T \Delta_-^{u_1} Z_{1t} \Delta_-^{u_2} Z'_{2t} \\ &\quad - T^{-1} \sum_{t=1}^T \Delta^{u_1} Z_{1t} \Delta_-^{u_2} Z'_{2t} - T^{-1} \sum_{t=1}^T \Delta_-^{u_1} Z_{1t} \Delta^{u_2} Z'_{2t}. \end{aligned}$$

The first term converges in probability to $E(\Delta^{u_1} Z_{1t})(\Delta^{u_2} Z_{2t})'$ by a LLN for stationary ergodic processes. By the Cauchy-Schwarz inequality the remaining terms tend to zero because

$$E\left(T^{-1} \sum_{t=1}^T \Delta_-^{u_i} Z_{it} \Delta_-^{u_i} Z'_{it}\right) = T^{-1} \sum_{t=1}^T \sum_{k=t}^{\infty} \varsigma_{ik}(-u_i) \Omega \varsigma_{ik}(-u_i)' \rightarrow 0.$$

We proved above that $M_T(u_1, u_2)$ is tight and therefore $M_T(u_1, u_2) \implies E(\Delta^{u_1} Z_{1t})(\Delta^{u_2} Z_{2t})'$.

Proof of (79): We define $\zeta_{i,t-n}^*(w_1) = T^{w_i+1/2} \pi_{t-n}(-w_i)$ for $w_i \leq -1/2 - \kappa_w$ so that $\max(w_1, w_2, w_1 + w_2 + 1) \leq -2\kappa_w < 0$. We then apply (52) and (60) with $\kappa = 2\kappa_w$, and find that (50) holds so that $M_T^*(w_1, w_2)$ is tight. Because $-1/(w + 1/2) \leq \kappa_w^{-1} < q$ we obtain the limit

$$T^{w_i+1/2} \Delta_+^{w_i} Z_{i[Ts]}^+ \implies W_{-w_i-1}(s), \quad i = 1, 2, \quad \text{on } D[0, 1],$$

see (4). The continuous mapping theorem gives the result (79).

Proof of (80): We apply (52) and (61) for $\zeta_{1,t-n}(u) = \pi_{t-n}(-u)$ and $\zeta_{2,t-n}^*(w) = T^{w+1/2} \pi_{t-n}(-w)$ and find for $w \leq -1/2 - \kappa_w$ and $u \geq -1/2 + \kappa_u$ that with $a = \kappa_u, \kappa = \kappa_w$,

$$\begin{aligned} \|M_T^*(w, u)\|_4 &\leq c(1 + \log T) T^{-\min(\kappa_u, \kappa_w)}, \\ \|M_T^*(w, u) - M_T^*(\tilde{w}, \tilde{u})\|_4 &\leq c|(w, u) - (\tilde{w}, \tilde{u})|(1 + \log T)^2 T^{-\min(\kappa_u, \kappa_w)}, \end{aligned}$$

and (50) implies that $M_T^*(w, u) = \mathbf{O}_P((1 + \log T)^2 T^{-\min(\kappa_u, \kappa_w)})$.

Proof of (81): We define $\zeta_{1,t-n} = \pi_{t-n}(-u)$ and $\zeta_{2,t-n} = \pi_{t-n}(-v)$ where $v \geq -1/2 - \underline{\kappa}_v$ and $u \geq -1/2 + \kappa_u$, so that $\min(u + 1, v + 1, u + v + 1) \geq \kappa_u - \underline{\kappa}_v \geq \kappa_u/2 > 0$. It then follows from (52) and (59) that (50) is satisfied and hence that $M_T(u, v)$ is tight.

Proof of (82): We first apply (52) with $\zeta_{1,t-n} = \pi_{t-n}(-v)$ and $\zeta_{2,t-n}^* = T^{w+1/2} \pi_{t-n}(-w)$ and find from (61) with $a = -\underline{\kappa}_v$ and $\kappa = \kappa_w$ that for $v \geq -1/2 - \underline{\kappa}_v$ and $w \leq -1/2 - \kappa_w$ we get

$$\begin{aligned} \|M_T^*(w, v)\|_4 &\leq c(1 + \log T) T^{\underline{\kappa}_v}, \\ \|M_T^*(w, v) - M_T^*(\tilde{w}, \tilde{v})\|_4 &\leq c|(w, v) - (\tilde{w}, \tilde{v})|(1 + \log T)^2 T^{\underline{\kappa}_v}, \end{aligned} \tag{85}$$

and (50) shows that $M_T^*(w, v) = \mathbf{O}_P((1 + \log T)^2 T^{\underline{\kappa}_v})$.

Proof of (83): Define \tilde{Z}_{it}^+ by $Z_{it}^+ = \xi_i \varepsilon_t + \Delta_+^{1/2} \tilde{Z}_{it}^+, i = 1, 2$, and because we need to decompose the processes we use the notation

$$M_T(U, V) = T^{-1} \sum_{t=1}^T U_t^+ V_t^{+'}$$

for product moments. We define $\xi = \text{blockdiag}(\xi_1, \xi_2)$, $\Delta_+^v Z_t = (\Delta_+^{v_1} Z'_{1t}, \Delta_+^{v_2} Z'_{2t})'$, $\Delta_+^v \tilde{Z}_t = (\Delta_+^{v_1} \tilde{Z}'_{1t}, \Delta_+^{v_2} \tilde{Z}'_{2t})'$, and $\Delta_+^v \varepsilon_t = (\Delta_+^{v_1} \varepsilon'_t, \Delta_+^{v_2} \varepsilon'_t)'$ and find the evaluation

$$M_T(\Delta_+^v Z, \Delta_+^v Z) \geq \xi M_T(\Delta_+^v \varepsilon, \Delta_+^v \varepsilon) \xi' + M_T(\Delta_+^{1/2+v} \tilde{Z}, \Delta_+^v \varepsilon) \xi' + \xi M_T(\Delta_+^v \varepsilon, \Delta_+^{1/2+v} \tilde{Z}),$$

where the inequality means that the difference is positive semi-definite.

We define the index $u_i = v_i + 1/2 \geq -1/2 + (1/2 - \underline{\kappa}_v)$ for $\Delta_+^{1/2+v_i} \tilde{Z}_{it}^+$ so that $\kappa_u = 1/2 - \underline{\kappa}_v \geq 5/14$ for $\underline{\kappa}_v \leq 1/7$. It follows that we can use (81) for the components of $M_T(\Delta_+^{1/2+v} \tilde{Z}, \Delta_+^v \varepsilon)$ and its transposed which are therefore $\mathbf{O}_P(1)$.

We next consider $M_T(\Delta_+^v \varepsilon, \Delta_+^v \varepsilon)$ and decompose $\Delta_+^{v_i} \varepsilon_t$ for $t > N = T^\alpha$:

$$\Delta_+^{v_i} \varepsilon_t = \sum_{n=0}^{N-1} \pi_n(-v_i) \varepsilon_{t-n} + \sum_{n=N}^{t-1} \pi_n(-v_i) \varepsilon_{t-n} = U_{it}^+ + V_{it}^+. \quad (86)$$

We define $U_t^+ = (U_{1t}^+, U_{2t}^+)'$ and $V_t^+ = (V_{1t}^+, V_{2t}^+)'$ and evaluate the product moment as

$$M_T(\Delta_+^v \varepsilon, \Delta_+^v \varepsilon) \geq M_T(U, U) + M_T(U, V) + M_T(V, U).$$

We next show that $M_T(U, V) + M_T(V, U) = \mathbf{o}_P(1)$. We apply Lemma A.7 for U_{it}^+ and V_{jt}^+ with coefficients $\zeta_{in} = \pi_n(-v_i)$ and $\zeta_{jn} = \pi_n(-v_j)$ and find from (62) that

$$\|T^{-1} \sum_{t=N}^T U_{it}^+ V_{jt}^+\|_4^4 \leq T^{-1} N \xi_N(\zeta_i, \zeta_i) \xi_T(\zeta_i, \zeta_i) \xi_T(\zeta_j, \zeta_j)^2.$$

From (59) in Lemma A.6 with $a = -2\underline{\kappa}_v$, we find that $\xi_T(\zeta_j, \zeta_j) \leq c(1 + \log T)T^{2\underline{\kappa}_v}$ uniformly for $-1/2 - \underline{\kappa}_v \leq v_j \leq -1/2 + \bar{\kappa}_v$. Similarly $\xi_N(\zeta_i, \zeta_i) \leq c(1 + \log N)N^{2\underline{\kappa}_v}$, so that for $N = T^\alpha$ we find that $\|T^{-1} \sum_{t=N}^T U_{it}^+ V_{jt}^+\|_4^4$ is bounded by

$$c(1 + \log N)(1 + \log T)^3 N^{1+2\underline{\kappa}_v} T^{-1+6\underline{\kappa}_v} = c(1 + \log T)^4 T^{-1+6\underline{\kappa}_v+\alpha(1+2\underline{\kappa}_v)} = o(1),$$

for $\alpha < 1/9$ and $\underline{\kappa}_v \leq 1/7$. The second condition of (50) is checked the same way, and the $\log T$ factors do not matter, so that

$$M_T(U, V) + M_T(V, U) = \mathbf{o}_P(1) \text{ for } \alpha < 1/9 \text{ and } \underline{\kappa}_v \leq 1/7. \quad (87)$$

What remains is the term $M_T(U, U)$. We define for integer N and $-1/2 - \underline{\kappa}_v \leq v_i \leq -1/2 + \bar{\kappa}_v$ the coefficient

$$F_{Nij} = \sum_{n=0}^{N-1} \pi_n(-v_i) \pi_n(-v_j) \geq 1 + c \frac{N^{-(v_1+v_2+1)} - 1}{-(v_1 + v_2 + 1)} \geq 1 + c \frac{1 - N^{-2\bar{\kappa}_v}}{2\bar{\kappa}_v},$$

see (58). Note that $F_{Nij} \rightarrow \infty$ as $(\bar{\kappa}_v, N) \rightarrow (0, \infty)$. We find that the mean of $M_T(U, U)$ is

$$E(T^{-1} \sum_{t=N+1}^T U_t^+ U_t^{+'}) = T^{-1}(T - N) \begin{pmatrix} F_{N11} & F_{N12} \\ F_{N12} & F_{N22} \end{pmatrix} \otimes \Omega_0,$$

and from (64) and (59) with $a = -2\underline{\kappa}_v$ we have that the variance of $M_T(U, U)$ is bounded by $T^{-1} N \xi_N(\zeta_1, \zeta_1) \xi_N(\zeta_2, \zeta_2) \leq T^{-1+\alpha(1+4\underline{\kappa}_v)}$ which tends to zero because $\alpha < 1/9$ and $\underline{\kappa}_v \leq 1/7$. Thus,

$$M_T(U, U) = T^{-1}(T - N) \begin{pmatrix} F_{N11} & F_{N12} \\ F_{N12} & F_{N22} \end{pmatrix} \otimes \Omega_0 + \mathbf{o}_P(1),$$

so that $M_T((v_1, v_2), (v_1, v_2))$ is bounded below by

$$\xi M_T(U, U) \xi' + \mathbf{o}_P(1) \geq c \frac{1 - N^{-2\bar{\kappa}_v}}{2\bar{\kappa}_v} (\xi'_1, \xi'_2)' \Omega_0 (\xi'_1, \xi'_2) + \mathbf{o}_P(1),$$

which proves (83). ■

For the proof of existence and consistency of the MLE we need the product moments that enter the likelihood function $\ell_{p,T}(\psi)$ and therefore define

$$\begin{aligned} M_T((a_1, a_2), a_3) &= T^{-1} \sum_{t=1}^T \begin{pmatrix} \mathbf{D}^{m_1} \Delta_+^{a_1} Z_{1t}^+ \\ \mathbf{D}^{m_2} \Delta_+^{a_2} Z_{2t}^+ \end{pmatrix} (\mathbf{D}^{m_3} \Delta_+^{a_3} Z_{3t}^+)', \\ M_T(a_1, a_2 | a_3) &= M_T(a_1, a_2) - M_T(a_1, a_3) M_T^{-1}(a_3, a_3) M_T(a_3, a_2). \end{aligned}$$

Corollary A.12 *If $\bar{\kappa}_v > 0$, $0 < \underline{\kappa}_v \leq \min(1/7, \kappa_u/2, \kappa_w/2)$, and the assumptions of Lemma A.11 hold, then*

$$T^{w_1+w_2+1} M_T(w_1, w_2 | u) = T^{w_1+w_2+1} M_T(w_1, w_2) + \mathbf{o}_P(1), \quad (88)$$

$$M_T(u_1, u_2 | w, u_3) \implies \text{Var}(\Delta^{u_1} Z_{1t}, \Delta^{u_2} Z_{2t} | \Delta^{u_3} Z_{3t}), \quad (89)$$

$$M_T(v, u_1 | w, u_2) = \mathbf{O}_P(1), \quad (90)$$

and for $N = T^\alpha$, $\alpha < (\kappa_w - \underline{\kappa}_v)/(1/2 + \underline{\kappa}_v)$, we have

$$M_T((v_1, v_2), (v_1, v_2) | w, u)^{-1} = \mathbf{O}_P\left(\frac{2\bar{\kappa}_v}{1 - N^{-2\bar{\kappa}_v}}\right). \quad (91)$$

Proof. *Proof of (88):* We introduce the notation $M_T^{**}(w_1, w_2)$ and decompose

$$M_T^{**}(w_1, w_2 | u) - M_T^{**}(w_1, w_2) = -M_T^*(w_1, u) M_T(u, u)^{-1} M_T^*(u, w_2),$$

and find from (80) that $M_T^*(w_i, u) \implies 0$, which together with (78) shows the result.

Proof of (89): We decompose

$$M_T(u_1, u_2 | w, u_3) - M_T(u_1, u_2 | u_3) = -M_T^*(u_1, w | u_3) M_T^{**}(w, w | u_3)^{-1} M_T^*(w, u_2 | u_3),$$

and find from (88) and Lemma A.11 that the right hand side is $\mathbf{o}_P(1)$ as $T \rightarrow \infty$, because $M_T^*(u_i, w | u_3) = \mathbf{O}_P((1 + \log T)^2 T^{-\min(\kappa_u, \kappa_w)})$, see (80). The result then follows from (78).

Proof of (90): We decompose $M_T(v, u_1 | w, u_2)$ as

$$M_T(v, u_1) - \begin{pmatrix} M_T^*(w, v) \\ M_T(u_2, v) \end{pmatrix}' \begin{pmatrix} M_T^{**}(w, w) & M_T^*(w, u_2) \\ M_T^*(u_2, w) & M_T(u_2, u_2) \end{pmatrix}^{-1} \begin{pmatrix} M_T^*(w, u_1) \\ M_T(u_2, u_1) \end{pmatrix}.$$

Because $M_T^*(w, u_2) \implies 0$ by (80), we first note that the second term is

$$M_T^*(v, w) M_T^{**}(w, w)^{-1} M_T^*(w, u_1) + M_T(v, u_2) M_T(u_2, u_2)^{-1} M_T(u_2, u_1) + \mathbf{o}_P(1).$$

Now $M_T^*(w, v) = \mathbf{O}_P((1 + \log T)^2 T^{\underline{\kappa}_v})$ and $M_T^*(w, u_1) = \mathbf{O}_P((1 + \log T)^2 T^{-\min(\kappa_u, \kappa_w)})$, so that by (79) and because $\underline{\kappa}_v \leq \frac{1}{2} \min(\kappa_u, \kappa_w)$, $M_T^*(v, w) M_T^{**}(w, w)^{-1} M_T^*(w, u_1) \implies 0$. Using (78) and (81) the result follows.

Proof of (91): The proof is similar to that of (83) except for conditioning on a stationary and a nonstationary variable. We start by eliminating the stationary variable. We find

$$\begin{aligned} &M_T((v_1, v_2), (v_1, v_2) | w, u) - M_T((v_1, v_2), (v_1, v_2) | w) \\ &= -M_T((v_1, v_2), u | w) M_T(u, u | w)^{-1} M_T(u, (v_1, v_2) | w), \end{aligned}$$

where $M_T(u, u|w)^{-1} = \mathbf{O}_P(1)$, see (89), and $M_T((v_1, v_2), u|w) = \mathbf{O}_P(1)$, see (90). Thus for $\Delta_+^{v_i} Z_{it}^+ = \Delta_+^{v_i} \varepsilon_t + \Delta_+^{v_i+1/2} \tilde{Z}_{it}^+$, $i = 1, 2$, $Z_t^+ = (Z_{1t}^+, Z_{2t}^+)'$, and $\Delta_+^w Z_{3t}^+ = \Delta_+^w \varepsilon_t + \Delta_+^{w+1/2} \tilde{Z}_{3t}^+$ it is enough to consider $M_T((v_1, v_2), (v_1, v_2)|w) = M_T(\Delta^v Z, \Delta^v Z | \Delta^w Z_3)$ which is bounded below by

$$M_T(\Delta^v \varepsilon, \Delta^v \varepsilon | \Delta^w Z_3) + M_T(\Delta^v \varepsilon, \Delta^{1/2+v} \tilde{Z} | \Delta^w Z_3) + M_T(\Delta^{1/2+v} \tilde{Z}, \Delta^v \varepsilon | \Delta^w Z_3).$$

It follows from (90) for $u_i = v_i + 1/2 \geq -\underline{\kappa}_v$, and $w \leq -1/2 - \kappa_w$, that the last two terms are $\mathbf{O}_P(1)$.

We next decompose the first term as $\Delta_+^{v_i} \varepsilon_t = U_{it}^+ + V_{it}^+$, see (86), and evaluate

$$M_T(\Delta^v \varepsilon, \Delta^v \varepsilon | \Delta^w Z_3) \geq M_T(U, U | \Delta^w Z_3) + M_T(U, V | \Delta^w Z_3) + M_T(V, U | \Delta^w Z_3).$$

The last two terms are evaluated as

$$M_T(U, V | \Delta^w Z_3) = M_T(U, V) - M_T^*(U, \Delta^w Z_3) M_T^{**}(\Delta^w Z_3, \Delta^w Z_3)^{-1} M_T^*(\Delta^w Z_3, V).$$

It follows from (87) that $M_T(U, V) = T^{-1} \sum_{t=N+1}^T U_t^+ V_t^{+'} = \mathbf{o}_P(1)$, from (79) that because $q > \kappa_w^{-1}$ we have $M_T^{**}(\Delta^w Z_3, \Delta^w Z_3)^{-1} = \mathbf{O}_P(1)$ for $w \leq -1/2 - \kappa_w$, and (82) shows that $M_T^*(\Delta^w Z_3, V) = O_P((1 + \log T)^2 T^{\underline{\kappa}_v})$. For the term

$$M_T^*(U, \Delta^w Z_3) = T^{-1} \sum_{t=N+1}^T U_t^+ \Delta_+^w Z_{3t}^{+'} T^{w+1/2} = \sum_{n=0}^{N-1} \pi_n(-v_1) [T^{-1} \sum_{t=N+1}^T \varepsilon_{t-n} \Delta_+^w Z_{3t}^{+'} T^{w+1/2}],$$

we apply (80) with $u = 0 = -1/2 + 1/2$ ($\kappa_u = 1/2$) so that $T^{-1} \sum_{t=N+1}^T \varepsilon_{t-n} \Delta_+^w Z_{3t}^{+'} T^{w+1/2} = \mathbf{O}_P((1 + \log T)^2 T^{-\kappa_w})$ for $\kappa_w < 1/2$. It follows that

$$\|M_T^*(U, \Delta^w Z_3)\|_2 \leq c(1 + \log T)^2 T^{-\kappa_w} \sum_{n=1}^{N-1} n^{-v_1-1} \leq c(1 + \log T)^2 T^{-\kappa_w + (1/2 + \underline{\kappa}_v)\alpha}. \quad (92)$$

Combining these results we find $M_T(U, V | \Delta^w Z_3) = \mathbf{O}_P((1 + \log T)^4 T^{-\kappa_w + \underline{\kappa}_v + (1/2 + \underline{\kappa}_v)\alpha}) = \mathbf{O}_P(1)$ for $\alpha < (\kappa_w - \underline{\kappa}_v)/(1/2 + \underline{\kappa}_v)$.

Finally we need to analyze

$$M_T(U, U | \Delta^w Z_3) = M_T(U, U) - M_T^*(U, \Delta^w Z_3) M_T^{**}(\Delta^w Z_3, \Delta^w Z_3)^{-1} M_T^*(\Delta^w Z_3, U)$$

The first term is $\mathbf{O}_P(2\bar{\kappa}_v/(1 - N^{-2\bar{\kappa}_v}))$, see (83), and (92) shows that $M_T^*(U, \Delta^w Z_3) = \mathbf{o}_P(1)$, which together with (79) proves the result. ■

Appendix B Proof of Theorem 5

We give the proof for case (i). The changes needed for cases (ii)–(iv) are outlined immediate following the statement of Theorem 5 in the main text.

By Lemma A.8 the deterministic terms generated by initial values are uniformly small. Note that (65) and (67) are formulated for index $\geq -1/2 - \kappa_1$, which covers not only the asymptotically stationary Z_{it}, W_{jt} but also those which are nearly critical, whereas (66) and (68) deals with the nonstationary Z_{it}, W_{jt} . Hence initial values do not influence the limit behavior of product moments, and in the remainder of the proof of Theorem 5, both under

(20) and (21), we therefore assume that the deterministic terms generated by initial values are zero.

In the following we want to use the result that if we regress a stationary variable on stationary and nonstationary variables, the limit of the normalized residual sum of squares is the same as if we leave out the nonstationary variables in the regression. Similarly if we regress a nonstationary variable on stationary and nonstationary variables, the limit of the normalized residual sums of squares is the same as if we leave out the stationary variables in the regression. These results are made precise in Lemma A.11 and Corollary A.12, which we apply repeatedly below to prove uniform convergence. Special problems arise if the regression contains processes that are nearly critical, but again the necessary results have been proved in Lemma A.11 and Corollary A.12.

The behavior of the processes depends on d and b . Note that $\beta'_{0\perp} \Delta^{d+mb} X_t \in \mathcal{F}(d_0 - d - mb)$ and $\beta'_0 \Delta^{d+nb} X_t \in \mathcal{F}(d_0 - b_0 - d - nb)$, and it is convenient to define the fractional indices $\delta_m = d - d_0 + mb$ and $\eta_m = d - d_0 + b_0 + nb$. Thus the fractional order is the negative fractional index. For notational reasons in Definition B.1 below we define $\eta_{-2} = \delta_{-2} = -\infty$ and $\eta_{k+1} = \delta_{k+1} = \infty$.

B.1 Proof of Theorem 5 under Assumption 1 with $d_0 - b_0 < 1/2$ and $E|\varepsilon_t|^8 < \infty$

If $d_0 - b_0 < 1/2$ then $d + nb - d_0 + b_0 \geq b_0 - d_0 > -1/2$, so that $\{W_{it}\}_{i=-1}^k$ is asymptotically stationary. The process $\Delta^{d+mb} \beta'_{0\perp} X_t$ is critical if $\delta_m = d + mb - d_0 = -1/2$, see Figure 1. The parameter space has to be divided into a set of “interiors” and “boundaries” as given in the next definition.

Definition B.1 We define the disjoint covering of $\mathcal{N} = \cup_{m=-1}^{k+1} \mathcal{N}_m$ by the sets

$$\mathcal{N}_m = \{\psi \in \mathcal{N} : \delta_{m-1} \leq -1/2 < \delta_m\}. \quad (93)$$

We take $0 < \kappa < \kappa_1 \leq 1/7$ and define the (κ_1, κ) -interior, $\mathcal{N}_m^{int}(\kappa_1, \kappa) = \cup_{-1 \leq m \leq k+1} \mathcal{N}_m^{int}(\kappa_1, \kappa)$, where

$$\mathcal{N}_m^{int}(\kappa_1, \kappa) = \{\psi \in \mathcal{N} : \delta_{m-1} \leq -1/2 - \kappa_1 \text{ and } -1/2 + \kappa \leq \delta_m\}, \quad (94)$$

and the (κ_1, κ) -boundary $\mathcal{N}_m^{bd}(\kappa_1, \kappa) = \cup_{-1 \leq m \leq k} \mathcal{N}_m^{bd}(\kappa_1, \kappa)$, where

$$\mathcal{N}_m^{bd}(\kappa_1, \kappa) = \{\psi \in \mathcal{N} : -1/2 - \kappa_1 \leq \delta_m \leq -1/2 + \kappa\}. \quad (95)$$

We have defined in (94) the (κ_1, κ) -interior $\mathcal{N}_m^{int}(\kappa_1, \kappa)$ as the set of ψ for which all processes are either clearly stationary or clearly nonstationary in the sense that their fractional index is either $\geq -1/2 + \kappa$ or $\leq -1/2 - \kappa_1$. The (κ_1, κ) -boundary $\mathcal{N}_m^{bd}(\kappa_1, \kappa)$ is the set where the process $\Delta^{d+mb} \beta'_{0\perp} X_t$ has an index which is close to the critical value of $-1/2$, see Figure 1 for an illustration.

The proof of Theorem 5 under (20) requires that we prove:

$$(i) \text{ for fixed } \kappa_1 \leq 1/7 : \sup_{\psi \in \mathcal{N}_m^{bd}(\kappa_1, \kappa)} |\ell_{T,p}(\psi) - \ell_p(\psi)| \xrightarrow{P} 0 \text{ as } (\kappa, T) \rightarrow (0, \infty), \quad (96)$$

$$(ii) \text{ for fixed } \kappa < \kappa_1 = 1/7 : \sup_{\psi \in \mathcal{N}_m^{int}(\kappa_1, \kappa)} |\ell_{T,p}(\psi) - \ell_p(\psi)| \xrightarrow{P} 0 \text{ as } T \rightarrow \infty. \quad (97)$$

B.1.1 Proof of (96): convergence on $\mathcal{N}_m^{bd}(\kappa_1, \kappa)$

The profile likelihood $\ell_{T,p}(\psi) = \log \det(SSR_T(\psi))$, see (16), is derived by regressing $\Delta^{d+kb}X_t$ on the other variables, which can be either asymptotically stationary or not. For $\psi \in \mathcal{N}_m^{bd}(\kappa_1, \kappa)$, we collect all asymptotically stationary regressors $\{Z_{it}\}_{i=m+1}^{k-1}$ and $\{W_{jt}\}_{j=-1}^{k-1}$ in a vector where the lowest fractional index for Z_{it} is $\delta_{m+1} = \delta_m + b \geq b - 1/2 - \kappa_1 \geq -1/2 + 5/14$, and the lowest for W_{jt} is $\eta_{-1} = d - b - d_0 + b_0 \geq -d_0 + b_0 \geq -1/2 + (b_0 - d_0 + 1/2)$, so we choose $\kappa_u = \min(5/14, b_0 - d_0 + 1/2)$. The nonstationary processes $\{Z_{it}\}_{i=-1}^{m-1}$ are collected in a vector with largest fractional index $w = \delta_{m-1} \leq -1/2 - (1/2 - \kappa) \leq -1/2 - 5/14$ for $\kappa \leq 1/7$, so we set $\kappa_w = 5/14$. This implies that $-(w + 1/2)^{-1} \leq 14/5$, so that $q > 14/5$ are enough moments to get weak convergence of $T^{w+1/2}\Delta_+^w\varepsilon_t$ to fBM.

The near critical index of $\Delta^{d+mb}\beta'_{0\perp}X_t$ is $v = d + mb - d_0 \in [-1/2 - \kappa_1, -1/2 + \kappa]$. We define $\underline{\kappa}_v = \kappa_1$ and $\bar{\kappa}_v = \kappa$, see Figure 1.

We consider two cases.

a: If Z_{kt} and W_{kt} are asymptotically stationary (indices $u_1 = d + kb - d_0$ and $u_2 = d + kb - d_0 + b_0$) then $\psi \in \mathcal{N}_m^{bd}(\kappa_1, \kappa)$ for some $m = -1, \dots, k-1$, and for $\Delta^{d+kb}X_t = \bar{\beta}_0 W_{kt} + \bar{\beta}_{0\perp} Z_{kt} = B_0(W'_{kt}, Z'_{kt})'$ we find $SSR_T(\psi) = B_0 M_T((u_1, u_2), (u_1, u_2)|v, w, u) B'_0$ where

$$\begin{aligned} M_T((u_1, u_2), (u_1, u_2)|v, w, u) &= M_T((u_1, u_2), (u_1, u_2)|w, u) \\ &\quad - M_T((u_1, u_2), v|w, u) M_T(v, v|w, u)^{-1} M_T(v, (u_1, u_2)|w, u). \end{aligned}$$

We find from (89) that for $w \leq -1/2 - \kappa_w$ it holds that $B_0 M_T((u_1, u_2), (u_1, u_2)|w, u) B'_0 \implies \text{Var}(S_{kt}|\mathcal{F}_{stat}(\psi)) = \ell_p(\psi)$. From (90) and (91) we find similarly that

$$M_T(u_i, v|w, u) = \mathbf{O}_P(1) \text{ and } M_T(v, v|w, u)^{-1} = \mathbf{O}_P\left(\frac{2\kappa}{1 - N^{-2\kappa}}\right),$$

where the latter tends to zero for $(\kappa, T) \rightarrow (0, \infty)$ when $N = T^\alpha$ for some $\alpha < (\kappa_w - \underline{\kappa}_v)/(1/2 + \underline{\kappa}_v) = (5/14 - \kappa_1)/(1/2 + \kappa_1)$. This proves (96) on $\mathcal{N}_m^{bd}(\kappa_1, \kappa)$, $m = -1, \dots, k-1$.

b: If $\psi \in \mathcal{N}_k^{bd}(\kappa_1, \kappa)$, then Z_{kt} is near critical with index $v_1 = d + kb - d_0 \in [-1/2 - \kappa_1, -1/2 + \kappa]$, and the remaining processes $Z_{it}, i < k$, are nonstationary with index $w \leq -1/2 - (1/2 - \kappa_1) \leq -1/2 - 5/14$. The stationary processes W_{jt} have index $u \geq -1/2 + (b_0 - d_0 + 1/2) = -1/2 + \kappa_u$, so that $\kappa_w = 5/14$ and $\kappa_u = b_0 - d_0 + 1/2$. We decompose $\det(SSR_T(\psi))$ as

$$\det(B_0 M_T((v_1, u_2), (v_1, u_2)|u, w) B'_0) = \det(M_T(u_2, u_2|u, w)) \det(M_T(v_1, v_1|u, u_2, w)) \det(B_0)^2,$$

where the first factor converges in distribution by (89), and the second factor increases to infinity for κ_1 fixed and $(\kappa, T) \rightarrow (0, \infty)$, see (91), so that $\ell_{T,p}(\psi) \implies \infty = \ell_p(\psi)$ on $\mathcal{N}_k^{bd}(\kappa_1, \kappa)$.

B.1.2 Proof of (97): convergence on $\mathcal{N}_m^{int}(\kappa_1, \kappa)$

For $\psi \in \mathcal{N}_m^{int}(\kappa_1, \kappa)$, we collect all asymptotically stationary regressors $\{Z_{it}\}_{i=m}^{k-1}$ and $\{W_{jt}\}_{j=-1}^{k-1}$ in a vector where the lowest fractional index for Z_{it} is $\delta_m \geq -1/2 + \kappa$ and the lowest for W_{jt} is $\eta_{-1} = d - b - d_0 + b_0 \geq -d_0 + b_0 \geq -1/2 + (b_0 - d_0 + 1/2)$, so we choose $\kappa_u = \min(\kappa, b_0 - d_0 + 1/2)$. The nonstationary processes $\{Z_{it}\}_{i=-1}^{m-1}$ are collected in a vector with largest fractional index $w = \delta_{m-1} \leq -1/2 - \kappa_1$, so $\kappa_w = \kappa_1$, and $-(w + 1/2)^{-1} \leq \kappa_1^{-1} = 7$. This means that $q > 7$ moments are enough for weak convergence of $T^{w+1/2}\Delta_+^w\varepsilon_t$ to fBM.

We want to prove (97) and consider two cases:

a: Consider first Z_{kt} and W_{kt} asymptotically stationary, that is $\psi \in \mathcal{N}_m^{int}(\kappa_1, \kappa)$ for some $m = -1, \dots, k$. Then $\Delta^{d+kb}\beta'_{0\perp}X_t$ and $\Delta^{d+kb}\beta'_0X_t$ are asymptotically stationary with indices $u_1 = d + kb - d_0 \geq -1/2 + \kappa$ and $u_2 = d + kb - d_0 + b_0 \geq -1/2 + \kappa$. With the notation from Lemma A.11 and Corollary A.12 we have that $SSR_T(\psi) = B_0M_T((u_1, u_2), (u_1, u_2)|w, u)B'_0$ and it follows from (89), see also (19), that for fixed (κ_1, κ) and $T \rightarrow \infty$,

$$\ell_{T,p}(\psi) = \log \det(SSR_T(\psi)) \implies \log \det(\text{Var}(S_{kt}|\mathcal{F}_{stat}(\psi))) = \ell_p(\psi), \quad \psi \in \mathcal{N}_m^{int}(\kappa_1, \kappa).$$

b: Suppose $\Delta^{d+kb}\beta'_{0\perp}X_t$ is nonstationary (index $w_1 \leq -1/2 - \kappa_1$), that is $\psi \in \mathcal{N}_{k+1}^{int}(\kappa_1, \kappa)$. Then $\det(SSR_T(\psi))$ is decomposed as

$$\det(B_0M_T((w_1, u_2), (w_1, u_2)|u, w)B'_0) = \det(M_T(w_1, w_1|w, u)) \det(M_T(u_2, u_2|w_1, u, w)) \det(B_0)^2,$$

where the first factor is $\mathbf{O}_P(T^{2\kappa_1})$, see (79) and (88), and the second factor converges in distribution by (89). It follows that $\ell_{T,p}(\psi) \implies \infty$ for $\psi \in \mathcal{N}_{k+1}^{int}(\kappa_1, \kappa)$.

This completes the proof of Theorem 5 under Assumption 1 and (20).

B.2 Proof of Theorem 5 under Assumption 1 with $E|\varepsilon_t|^q < \infty$ for all q

If b_0 is not greater than $d_0 - 1/2$, then W_{jt} is not necessarily asymptotically stationary and the parameter set has to be cut up as in Figure 2, using two sets of lines $\delta_m = d + mb - d_0 = -1/2$ and $\eta_n = d + nb - d_0 + b_0 = -1/2$. These lines may intersect and close to these intersection points there are two almost critical processes. This requires a new proof, and we need to assume moments of all orders.

Definition B.2 We define the disjoint covering $\mathcal{N} = \cup_{-1 \leq n \leq m \leq k+1} \mathcal{N}_{mn}$, where

$$\mathcal{N}_{mn} = \{\psi \in \mathcal{N} : \max(\delta_{m-1}, \eta_{n-1}) \leq -1/2 < \min(\delta_m, \eta_n)\}, \quad (98)$$

as well as the interiors

$$\mathcal{N}_{mn}^{int}(\kappa) = \{\psi \in \mathcal{N} : \max(\delta_{m-1}, \eta_{n-1}) \leq -1/2 - \kappa, -1/2 + \kappa \leq \min(\delta_m, \eta_n)\}$$

and the cross points and boundaries

$$\begin{aligned} \mathcal{N}_{mn}^{cr}(\kappa_2) &= \{\psi : |\eta_n + 1/2| \leq \kappa_2, |\delta_m + 1/2| \leq \kappa_2\}, -1 \leq n < m \leq k, \\ \mathcal{N}_{mn}^{bd}(\kappa_2, \kappa) &= \{\psi : |\delta_m + 1/2| \leq \kappa, -1/2 + \kappa_2 \leq \eta_n, \eta_{n-1} \leq -1/2 - \kappa_2\}, -1 \leq n \leq m \leq k, \\ \mathcal{N}_{mn}^{bd}(\kappa_2, \kappa) &= \{\psi : |\eta_m + 1/2| \leq \kappa, -1/2 + \kappa_2 \leq \delta_n, \delta_{n-1} \leq -1/2 - \kappa_2\}, -1 \leq m < n \leq k + 1, \end{aligned} \quad (99)$$

The interpretation of \mathcal{N}_{mn} is that for $\psi \in \mathcal{N}_{mn}$, $\beta'_{0\perp}\Delta_+^{d+mb}X_t$ and $\beta'_0\Delta_+^{d+nb}X_t$ are asymptotically stationary, whereas $\beta'_{0\perp}\Delta_+^{d+(m-1)b}X_t$ and $\beta'_0\Delta_+^{d+(n-1)b}X_t$ are nonstationary. The true value ψ_0 is contained in $\mathcal{N}_{0,-1}$.

Note that $\eta_m = \delta_m + b_0$, so that $\eta_m > \delta_m$. Hence $\beta'_{0\perp}\Delta_+^{d+mb}X_t$ asymptotically stationary implies that $\beta'_0\Delta_+^{d+mb}X_t$ is asymptotically stationary, and $\beta'_0\Delta_+^{d+mb}X_t$ nonstationary implies that $\beta'_{0\perp}\Delta_+^{d+mb}X_t$ is nonstationary.

The set $\mathcal{N}_{mn}^{cr}(\kappa_2)$ contains the crossing point between the lines $\delta_m = -1/2$ and $\eta_n = -1/2$, where Z_{mt} and W_{nt} are nearly critical. To the left they are nonstationary and to the right

stationary. In the remaining two wedges there is one of each. The set $\mathcal{N}_{mn}^{bd}(\kappa_2, \kappa)$ covers the line segment between $\mathcal{N}_{m,n-1}^{cr}(\kappa_2)$ and $\mathcal{N}_{mn}^{cr}(\kappa_2)$ and $\mathcal{N}_{nm}^{bd}(\kappa_2, \kappa)$ covers the line segment between $\mathcal{N}_{mn}^{cr}(\kappa_2)$ and $\mathcal{N}_{m+1,n}^{cr}(\kappa_2)$. In these sets either Z_{mt} or W_{nt} is nearly critical, but not both. See Figures 2 and 3 for illustrations.

The proof of Theorem 5 under (21) requires that we prove:

$$(i) : \sup_{\psi \in \mathcal{N}_{mn}^{cr}(\kappa_2)} |\ell_{T,p}(\psi) - \ell_p(\psi)| \xrightarrow{P} 0 \text{ as } (\kappa_2, T) \rightarrow (0, \infty), \quad (100)$$

$$(ii) \text{ for fixed } \kappa_2 : \sup_{\psi \in \mathcal{N}_{mn}^{bd}(\kappa_2, \kappa)} |\ell_{T,p}(\psi) - \ell_p(\psi)| \xrightarrow{P} 0 \text{ as } (\kappa, T) \rightarrow (0, \infty), \quad (101)$$

$$(iii) \text{ for fixed } \kappa : \sup_{\psi \in \mathcal{N}_{mn}^{int}(\kappa)} |\ell_{T,p}(\psi) - \ell_p(\psi)| \xrightarrow{P} 0 \text{ as } T \rightarrow \infty. \quad (102)$$

Note the difference to the setup with $d_0 - b_0 < 1/2$. We have to let κ_2 tend to zero to evaluate the neighborhood $\mathcal{N}_{mn}^{cr}(\kappa_2)$ of the intersection points. This means that when we fix κ_2 in the analysis of $\mathcal{N}_{mn}^{bd}(\kappa_2, \kappa)$, the nonstationary processes have an index $\leq -1/2 - \kappa_2$ which means that we need moments of order at least κ_2^{-1} , which can be very large. This is the reason that we need to assume moments of all orders. Note also that the strips defined in $\mathcal{N}_{mn}^{bd}(\kappa_2, \kappa)$ have width 2κ , whereas the strips defined when $d_0 - b_0 < 1/2$ are asymmetric and go from $-1/2 - \kappa_1$ on the nonstationary side to $-1/2 + \kappa$ on the stationary side. The large and fixed value of κ_1 is there to guarantee that not too many moments are needed in that situation.

B.2.1 Proof of (100): convergence on $\mathcal{N}_{mn}^{cr}(\kappa_2)$

We want to show (100) where $(\kappa_2, T) \rightarrow (0, \infty)$ and consider two cases.

a: We first consider the case where Z_{kt} and W_{kt} are asymptotically stationary (indices u_1 for Z_{kt} and u_2 for W_{kt}), and $\psi \in \mathcal{N}_{mn}^{cr}(\kappa_2)$, $-1 \leq n < m \leq k - 1$. In Figure 3 an example is the set $\mathcal{N}_{1,0}^{cr}(\kappa_2)$. We now have two near critical processes with indices $v_1 = \delta_m$ and $v_2 = \eta_n$ in $[-1/2 - \kappa_2, -1/2 + \kappa_2]$. The stationary processes have an index $u \geq \min(\delta_{m+1}, \eta_{n+1}) = \min(\delta_m, \eta_n) + b \geq -1/2 + (1/2 - \kappa_2) \geq -1/2 + 5/14$ for $\kappa_2 \leq 1/7$ and the nonstationary processes have an index $w \leq \max(\delta_{m-1}, \eta_{n-1}) = \max(\delta_m, \eta_n) - b \leq -1/2 - (1/2 - \kappa_2) \leq -1/2 - 5/14$, so that when we apply the evaluations of product moments in Lemma A.11 and Corollary A.12, we can take $\kappa_w = \kappa_u = 5/14$ and $\underline{\kappa}_v = \bar{\kappa}_v = \kappa_2$. This means that $-(w + 1/2)^{-1} \leq 14/5$ and $q > 14/5$ are sufficient moments to apply weak convergence of $T^{w+1/2} \Delta_{\pm}^w \varepsilon_t$ to fBM.

We decompose the main factor of $SSR_T(\psi) = B_0 M_T((u_1, u_2), (u_1, u_2) | v_1, v_2, w, u) B_0'$ as

$$\begin{aligned} & M_T((u_1, u_2), (u_1, u_2) | w, u) \\ & - M_T((u_1, u_2), (v_1, v_2) | w, u) M_T^{-1}((v_1, v_2), (v_1, v_2) | w, u) M_T((v_1, v_2), (u_1, u_2) | w, u) \end{aligned} \quad (103)$$

and apply Corollary A.12. We find from (89) that because $\min(u_1, u_2, u) \geq -1/2 + 5/14$ and $w \leq -1/2 - 5/14$ we have for $T \rightarrow \infty$,

$$\log \det(B_0 M_T((u_1, u_2), (u_1, u_2) | w, u) B_0') \implies \log \det(\text{Var}(S_{kt} | \mathcal{F}_{stat}(\psi))) = \ell_p(\psi). \quad (104)$$

We next show that $\log \det(SSR_T(\psi)) \implies \ell_p(\psi)$ by showing that the second term of (103) is $\mathbf{o}_P(1)$.

The critical processes are $Z_{mt} = \Delta_+^{d+mb}\beta'_{0\perp}X_t$ and $W_{nt} = \Delta_+^{d+nb}\beta'_0X_t$ with stochastic parts

$$\begin{aligned} Z_{mt} &: \Delta_+^{v_1}(\beta'_{0\perp}C_0\varepsilon_t + \Delta_+^{b_0}\beta'_{0\perp}Y_t^+), \\ W_{nt} &: \Delta_+^{v_2}\beta'_0Y_t^+ = \Delta_+^{v_2}(\beta'_0\tau\varepsilon_t + \Delta^{1/2}\beta'_0\sum_{n=0}^{\infty}\tilde{\tau}_n\varepsilon_{t-n}), \end{aligned}$$

see (10), which are fractional differences of processes in \mathcal{Z} . The leading coefficients are $\xi_1 = \beta'_{0\perp}C_0$ and $\xi_2 = \beta'_0\tau = \beta'_0H(1)$, respectively, for which (ξ'_1, ξ'_2) has full column rank, because

$$(\alpha_{0\perp}, \alpha_0)'(\xi'_1, \xi'_2) = \begin{pmatrix} \alpha'_{0\perp}C'_0\beta_{0\perp} & \alpha'_{0\perp}H(1)'\beta_0 \\ 0 & \alpha'_0H(1)'\beta_0 \end{pmatrix} = \begin{pmatrix} \alpha'_{0\perp}C'_0\beta_{0\perp} & \alpha'_{0\perp}H(1)'\beta_0 \\ 0 & -I_r \end{pmatrix}$$

has full rank, see Theorem 3. It follows from (91) that $M_T((v_1, v_2), (v_1, v_2)|w, u)^{-1} \implies 0$ as $(\kappa_2, T) \rightarrow (0, \infty)$, and from (90) that $M_T((u_1, u_2), (v_1, v_2)|w, u) = \mathbf{O}_P(1)$.

b: Next consider the situation where Z_{kt} is nearly critical (index v_1) and W_{kt} is stationary (index u_2), that is $\psi \in \mathcal{N}_{kn}^{cr}(\kappa_2)$ for some $n = -1, \dots, k-1$. In Figure 2 this corresponds to the sets $\mathcal{N}_{1,0}^{cr}(\kappa_2)$ and $\mathcal{N}_{1,-1}^{cr}(\kappa_2)$. In this case we define $v_1 = \delta_k \in [-1/2 - \kappa_2, -1/2 + \kappa_2]$, $v = \eta_n \in [-1/2 - \kappa_2, -1/2 + \kappa_2]$, $u_2 = \eta_k \geq -1/2 + (1/2 - \kappa_2) \geq -1/2 + 5/14$, $w \leq \max(\delta_{k-1}, \eta_{n-1}) \leq -1/2 - (1/2 - \kappa_2) \leq -1/2 - 5/14$ for $\kappa_2 \leq 1/7$, and if $n < k-1$ then also $u \geq \eta_{n+1} \geq -1/2 + 5/14$, so we can choose $\kappa_u = \kappa_w = 5/14$. The determinant $\det(SSR_T(\psi))$ has, apart from the factor $\det(B_0)^2$, the form

$$\det(M_T((v_1, u_2), (v_1, u_2)|v, w, u)) = \det(M_T(u_2, u_2|v, w, u)) \det(M_T(v_1, v_1|u_2, v, w, u)).$$

The first factor converges in distribution to $\det(E(S_{wkt}S'_{wkt}|\mathcal{F}_{stat}(\psi)))$ as $(\kappa_2, T) \rightarrow (0, \infty)$, see (89), (90), and (91). Thus we investigate $M_T(v_1, v_1|u_2, v, w, u)$ which tends to infinity by an argument similar to the proof of (83). Thus

$$\log \det(SSR_T(\psi)) \implies \infty = \ell_p(\psi) \text{ on } \mathcal{N}_{kn}^{cr}(\kappa_2), n = -1, \dots, k-1, \text{ as } (\kappa_2, T) \rightarrow (0, \infty).$$

B.2.2 Proof of (101): convergence on $\mathcal{N}_{mn}^{bd}(\kappa_2, \kappa)$

We have defined the sets $\mathcal{N}_{mn}^{bd}(\kappa_2, \kappa)$ in (99) in such a way that interchanging η with δ and n with m we transform $\mathcal{N}_{mn}^{bd}(\kappa_2, \kappa)$ into $\mathcal{N}_{nm}^{bd}(\kappa_2, \kappa)$, so the analysis is to a large extent the same. In the proof of (101) we fix κ_2 and let $(\kappa, T) \rightarrow (0, \infty)$.

For $\psi \in \mathcal{N}_{mn}^{bd}(\kappa_2, \kappa)$, $n \leq m$, the near critical index is $v = \delta_m \in [-1/2 - \kappa, -1/2 + \kappa]$, the asymptotically stationary processes have an index $u \geq \min(\eta_n, \delta_{m+1}) \geq \min(-1/2 + \kappa_2, -1/2 - \kappa + b) \geq -1/2 + \kappa_2$, and the nonstationary processes have an index $w \leq \max(\eta_{n-1}, \delta_{m-1}) \leq \max(-1/2 - \kappa_2, -1/2 + \kappa - b) \leq -1/2 - \kappa_2$, see Figures 2 and 3. For $\psi \in \mathcal{N}_{mn}^{bd}(\kappa_2, \kappa)$, $m < n$, the near critical index is $v = \eta_m \in [-1/2 - \kappa, -1/2 + \kappa]$ and also $u \geq -1/2 + \kappa_2$, $w \leq -1/2 - \kappa_2$. Thus, in both cases we can take $\kappa_u = \kappa_w = \kappa_2$ fixed and $\bar{\kappa}_v = \bar{\kappa}_v = \kappa$ tending to zero. Note that no matter how small κ_2 has been chosen it is fixed, and we can use that $M_T^{**}(w, w)^{-1} = \mathbf{O}_P(1)$ because $-(w + 1/2)^{-1} \leq \kappa_2^{-1} < \infty$ and we have assumed moments of all orders.

We want to prove (101) and consider four cases.

a: If Z_{kt} and W_{kt} are asymptotically stationary and $\psi \in \mathcal{N}_{mn}^{bd}(\kappa_2, \kappa)$, $-1 \leq m \leq k-1$, $-1 \leq n \leq k$, then the proof of (101) is the same as given in Section B.1.1a.

b: If Z_{kt} is near critical and $\psi \in \mathcal{N}_{kn}^{bd}(\kappa_2, \kappa)$, $n = -1, \dots, k$, then the proof is the same as given in Section B.1.1b.

c: If $\psi \in \mathcal{N}_{m, k+1}^{bd}(\kappa_2, \kappa)$, $-1 \leq m < k$, then W_{mt} is critical (index $v \in [-1/2 - \kappa, -1/2 + \kappa]$), Z_{kt} nonstationary (index $w_1 \leq -1/2 - \kappa_2$), and W_{kt} asymptotically stationary (index $u_2 \geq -1/2 + (1/2 - \kappa_2)$). In this case $SSR_T(\psi) = B_0 M_T((w_1, u_2), (w_1, u_2)|v, w, u) B_0'$ and

$$M_T((w_1, u_2), (w_1, u_2)|v, w, u) = A_{22} - A_{21} A_{11}^{-1} A_{12} = A_{22.1},$$

where we use the notation

$$A_{11} = M_T(v, v|w, u), \quad A_{12} = M_T(v, (w_1, u_2)|w, u), \quad A_{22} = M_T((w_1, u_2), (w_1, u_2)|w, u)$$

and the representation, see Magnus and Neudecker (1999, p. 11, equation (7)),

$$A_{22.1}^{-1} = A_{22}^{-1} [A_{22} + A_{21} A_{11.2}^{-1} A_{12}] A_{22}^{-1}.$$

We find from Lemma A.11 and Corollary A.12 that

$$\begin{aligned} A_{22}^{**} &= M_T^{**}((w_1, u_2), (w_1, u_2)|w, u) = \mathbf{O}_P(1), \\ A_{22}^{**^{-1}} &= \mathbf{O}_P(1), \\ A_{12}^* &= M_T^*(v, (w_1, u_2)|w, u) = \mathbf{O}_P((1 + \log T)^2 T^\kappa), \\ A_{11.2}^{-1} &= \mathbf{O}_P\left(\frac{2\kappa}{1 - N^{-2\kappa}}\right). \end{aligned}$$

Therefore we have for $2w + 1 \leq -2\kappa_2$, that

$$A_{22.1}^{-1} = T^{2w+1} A_{22.1}^{**^{-1}} = \mathbf{O}_P(T^{-2\kappa_2} (1 + \frac{2\kappa}{1 - N^{-2\kappa}} (1 + \log T)^4 T^{2\kappa})) = \mathbf{o}_P(1).$$

This implies that $\det(SSR_T(\psi)^{-1}) \implies 0$, and hence that $\ell_{T,p}(\psi) = \log \det(SSR_T(\psi)) \implies \infty$ for $\psi \in \mathcal{N}_{m, k+1}^{bd}(\kappa_2, \kappa)$, $m = -1, \dots, k-1$, for κ_2 fixed and $(\kappa, T) \rightarrow (0, \infty)$.

d: The last case is when $\psi \in \mathcal{N}_{k, k+1}^{bd}(\kappa_2, \kappa)$. In this case W_{kt} is nearly critical with index $v_2 \in [-1/2 - \kappa, -1/2 + \kappa]$ and Z_{kt} is nonstationary with index $w_1 \leq -1/2 - \kappa_2$. The remaining processes are all nonstationary with index $w \leq -1/2 - \kappa_2$ and $\det(SSR_T(\psi))$ is

$$\det(B_0 M_T(w_1, v_2|w) B_0') = \det(M_T(v_2, v_2|w_1, w)) \det(M_T(w_1, w_1|w)) \det(B_0)^2.$$

Here both factors tend to infinity for κ_2 fixed and $(\kappa, T) \rightarrow (0, \infty)$, and therefore $\ell_{T,p}(\psi) \implies \infty = \ell_p(\psi)$ for $\psi \in \mathcal{N}_{k, k+1}^{bd}(\kappa_2, \kappa)$.

B.2.3 Proof of (102): convergence on $\mathcal{N}_{mn}^{int}(\kappa)$

We want to prove (102) for κ fixed and $T \rightarrow \infty$, and consider three cases.

a: Consider first Z_{kt} and W_{kt} asymptotically stationary, that is $\psi \in \mathcal{N}_{mn}^{int}(\kappa)$, $-1 \leq n \leq m \leq k$. The proof can be taken from Section B.1.2a.

b: Suppose Z_{kt} is nonstationary (index w_1) and W_{kt} is stationary (index u_2), then $\psi \in \mathcal{N}_{k+1, n}^{int}(\kappa)$, $n = -1, \dots, k$. The proof can be taken from Section B.1.2b.

c: Finally if both Z_{kt} and W_{kt} are nonstationary, that is $\psi \in \mathcal{N}_{k+1, k+1}^{int}(\kappa)$, then

$$SSR_T(\psi) = B_0 M_T((w_1, w_2), (w_1, w_2)|w) B_0',$$

which converges to infinity, see (79), so that $\ell_{T,p}(\psi) \implies \infty$ for $T \rightarrow \infty$ and $\psi \in \mathcal{N}_{k+1, k+1}^{int}(\kappa)$.

References

1. Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics* **22**, pp. 327–351.
2. Avarucci, M. and Velasco, C. (2009). A Wald test for the cointegration rank in non-stationary fractional systems. *Journal of Econometrics* **151**, pp. 178–189.
3. Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.
4. Davidson, J. (2002). A model of fractional cointegration, and tests for cointegration using the bootstrap. *Journal of Econometrics* **110**, pp. 187–212.
5. Davydov, Yu. A. (1970). The invariance principle for stationary processes. *Theory of Probability and Its Applications* **15**, pp. 487–498.
6. Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: representation, estimating and testing. *Econometrica* **55**, pp. 251–276.
7. Granger, C. W. J. (1986). Developments in the study of cointegrated economic variables. *Oxford Bulletin of Economic and Statistics* **48**, pp. 213–228.
8. Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and its Application*, Academic Press, New York.
9. Henry, M. and Zaffaroni, P. (2003). The long range dependence paradigm for macroeconomics and finance. In Doukhan, P., Oppenheim, G., and Taqqu, M. S., eds, *Theory and Applications of Long-Range Dependence*, Birkhäuser, Boston, pp. 417–438.
10. Hualde, J. and Robinson, P. M. (2010a). Semiparametric inference in multivariate fractionally cointegrated systems. Forthcoming in *Journal of Econometrics*.
11. Hualde, J. and Robinson, P. M. (2010b). Gaussian pseudo-maximum likelihood estimation of fractional time series models. Working paper, Universidad de Navarra.
12. 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.
13. Jeganathan, P. (1999). On asymptotic inference in cointegrated time series with fractionally integrated errors. *Econometric Theory* **15**, pp. 583–621.
14. Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* **12**, pp. 231–254.
15. Johansen, S. (2002). A small sample correction of the test for cointegrating rank in the vector autoregressive model. *Econometrica* **70**, pp. 1929–1961.
16. Johansen, S. (2008). A representation theory for a class of vector autoregressive models for fractional processes. *Econometric Theory* **24**, pp. 651–676.
17. Johansen, S. and Nielsen, M. Ø. (2010). Likelihood inference for a nonstationary fractional autoregressive model. Forthcoming in *Journal of Econometrics*.
18. Kallenberg, O. (2001). *Foundations of Modern Probability Theory*, 2nd edition, Springer, New York.
19. Lasak, K. (2008a). Likelihood based testing for no fractional cointegration. Forthcoming in *Journal of Econometrics*.
20. Lasak, K. (2008b). Maximum likelihood estimation of fractionally cointegrated systems. CREATES Research Paper 2008–53.
21. Lawley, D. N. (1956). A general method for approximating to the distribution of

- likelihood ratio criteria. *Biometrika* **43**, pp. 296–303.
22. Lobato, I. N. and Velasco, C. (2007). Efficient Wald tests for fractional unit roots. *Econometrica* **75**, pp. 575–589.
 23. Magnus, J. R. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, revised edition, John Wiley and Sons, New York.
 24. Marinucci, D. and Robinson, P. M. (2000). Weak convergence of multivariate fractional processes. *Stochastic Processes and their Applications* **86**, pp. 103–120.
 25. Marmol, F. and Velasco, C. (2004). Consistent testing of cointegrating relationships. *Econometrica* **72**, pp. 1809–1844.
 26. Phillips, P. C. B. (1991). Optimal inference in cointegrated systems. *Econometrica* **59**, pp. 283–306.
 27. Phillips, P. C. B. and Hansen, B. E. (1990). Statistical inference in instrumental variables regression with I(1) variables. *Review of Economic Studies* **57**, pp. 99–125.
 28. Robinson, P. M. and Hualde, J. (2003). Cointegration in fractional systems with unknown integration orders. *Econometrica* **71**, pp. 1727–1766.
 29. Saikkonen, P. (1991). Asymptotically efficient estimation of cointegration regressions. *Econometric Theory* **7**, pp. 1–21.
 30. van der Vaart, A. W. (1998). *Asymptotic Statistics*, Cambridge University Press, Cambridge.

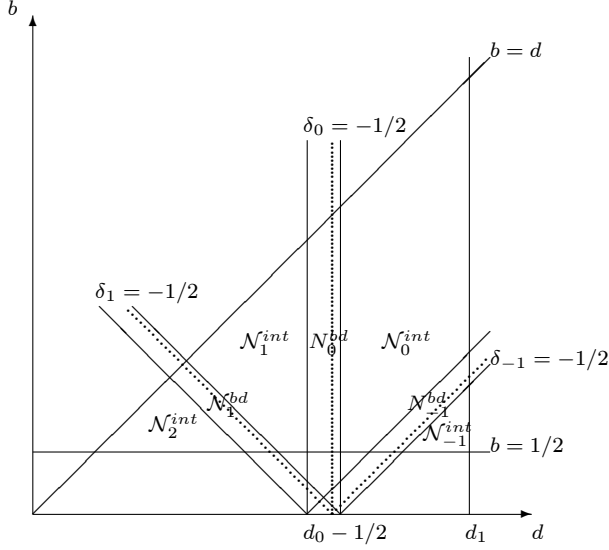


Figure 1: The case $d_0 - b_0 < 1/2$. The parameter space \mathcal{N} is the set bounded by the lines $b = 1/2$, $b = d$, and $d = d_1$. The sets $\mathcal{N}_m^{bd} = \mathcal{N}_m^{bd}(\kappa_1, \kappa)$, where a process is close to being critical, and the sets $\mathcal{N}_m^{int} = \mathcal{N}_m^{int}(\kappa_1, \kappa)$ are illustrated assuming $k = 1$. If $k \geq 2$ there would be more lines.

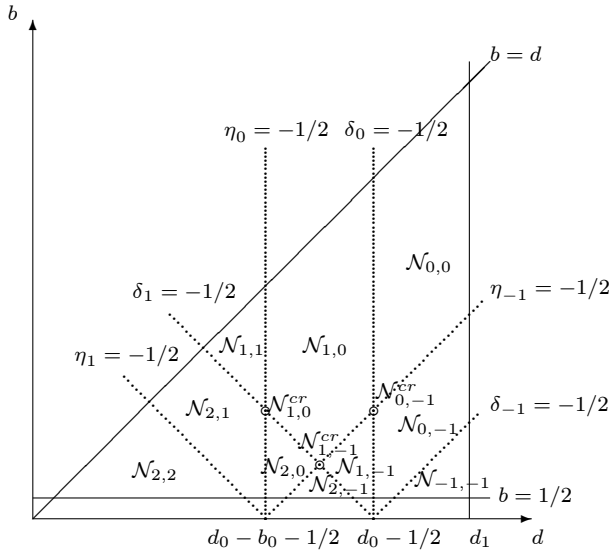


Figure 2: The general case. The parameter space \mathcal{N} is bounded by the lines $b = 1/2$, $b = d$, and $d = d_1$. The sets $\mathcal{N}_{mn}^{cr} = \mathcal{N}_{mn}^{cr}(\kappa_2)$, where two processes are close to being critical, and the sets \mathcal{N}_{mn} are illustrated assuming $k = 1$. If $k \geq 2$ there would be more lines.

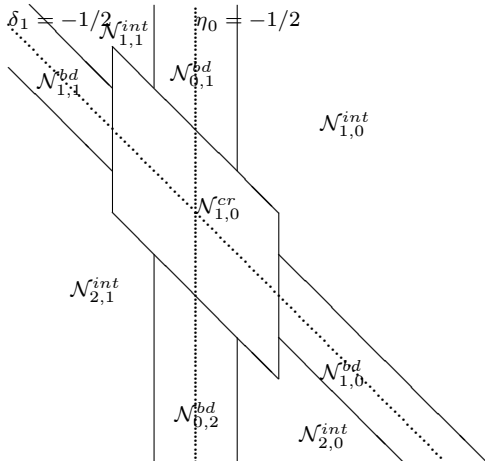


Figure 3: The set $\mathcal{N}_{1,0}^{cr} = \mathcal{N}_{1,0}^{cr}(\kappa_2)$. For $k \geq 1$ the set $\mathcal{N}_{1,0}^{cr}$ covers the intersection between the lines $\delta_1 = -1/2$ and $\eta_0 = -1/2$, where Z_{1t} and W_{0t} are nearly critical. The sets $\mathcal{N}_{mn}^{int} = \mathcal{N}_{mn}^{int}(\kappa)$ and $\mathcal{N}_{mn}^{bd} = \mathcal{N}_{mn}^{bd}(\kappa_2, \kappa)$ are also indicated.