Likelihood inference for a fractionally cointegrated vector autoregressive model* (revision v9)

Søren Johansen[†]
University of Copenhagen
and CREATES

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

February 22, 2011

Abstract

We consider model based inference in a fractionally cointegrated (or cofractional) vector autoregressive model with a restricted constant term, based on the conditional Gaussian likelihood. The model equations generate a process X_t which, under suitable conditions on the parameters, is 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 $d-b$, and no other fractionality order is possible. We consider the model defined by $0 < b \leq d$, and some submodels defined by the restrictions $d = b$, $d = d_0$ for some prespecified d_0, or $\rho = 0$. Our main technical contribution is the proof of consistency of the maximum likelihood estimators on a compact subset of $0 < b \leq d$. To this end, we consider the probability measure generated by the true values $d_0 - b_0 < 1/2$ for which $\beta_0' X_t + \rho_0'$ is stationary with mean zero and where ρ can be estimated consistently. We consider the conditional likelihood as a continuous 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. If the true value $b_0 > 1/2$ we 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 extended by $u^{-(d_0-b_0)}$. If $b_0 < 1/2$ all the estimators are asymptotically Gaussian and the rank test is asymptotically chi-squared distributed.

Keywords: Cofractional processes, cointegration rank, fractional cointegration, likelihood inference, vector autoregressive model.
JEL Classification: C32.

---

*We are grateful to Jim Stock, five referees, 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 275-05-0220), 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."

†Email: `Soren.Johansen@econ.ku.dk`

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

1 Introduction and motivation

The model equations and the models and submodels[1]

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

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

where ε_t are i.i.d.$(0, \Omega)$, Ω is positive definite, $0 < b \leq d$, and α and β are $p \times r,\ 0 \leq r \leq p$. The fractional difference operator is Δ^b and $L_b = 1 - \Delta^b = 1 - (1 - bL + \ldots) = bL + \ldots$ the fractional lag operator. The parameter space of $\mathcal{H}_r$ is given by the otherwise unrestricted parameters $(d, b, \alpha, \beta, \rho, \Gamma_1, \ldots, \Gamma_k, \Omega)$. In the special case $r = p$, the matrix $\Pi = \alpha\beta'$ is an unrestricted $p \times p$ matrix, and if $r = 0$ we have $\alpha = \beta = 0$.

This model was suggested by Johansen (2008) as a multivariate model for fractional processes as a generalization of the cointegrated vector autoregression (CVAR) model. It has the attractive features of a straightforward interpretation of β as the cointegrating parameters in the long-run stable relations, and of α as the parameters describing adjustment towards the long-run equilibria and (through the orthogonal complement) the common stochastic trends. If $d - b < 1/2$, $\beta' X_t + \rho'$ is asymptotically stationary with mean zero.

The model allows for a simple criteria for fractionality and cofractionality of X_t (or fractional cointegration; henceforth we use these terms synonymously), and at the same time the model is relatively easy to estimate, because for fixed (d, b) the model is estimated by reduced rank regression which reduces the numerical problem to an optimization of a function of just two variables.

The purpose of this paper is to conduct (quasi) Gaussian maximum likelihood inference in model (1) and show that the maximum likelihood estimator exists uniquely with large probability and is consistent, and to find the asymptotic distribution of maximum likelihood estimators and some likelihood ratio test statistics.

We are interested in testing the rank of the coefficient to $\Delta^{d-b} L_b X_t$ and in conducting inference on the parameters of model (1). We analyze the conditional likelihood function for $(X_1, \ldots, X_T)$ given initial values $X_{-n},\ n = 0, 1, \ldots$, 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.

Submodels

We also consider the model $\mathcal{H}_r(d = b)$ given by the parameter restriction $d = b$. Finally, we consider the model without deterministic terms, i.e. with $\rho = 0$. 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.

Granger historie

The inspiration for model (1) comes from Granger (1986), who noted the special role of

[1] SJ: Denne sub sub section inddeling skal fjernes, men tjener lige nu til at se hvad strukturen i indledningen er

the fractional lag operator $L_b = 1 - \Delta^b$ and suggested the model

$$A^*(L)\Delta^d X_t = \Delta^{d-b} L_b \alpha\beta' X_{t-1} + d(L)\varepsilon_t, \tag{2}$$

see also Davidson (2002). A simple way of deriving the main term of this model is to assume that we have linear combinations $\Delta^d \gamma' X_t$ and $\Delta^{d-b}\beta' X_t$ which are $I(0)$. Simple algebra shows that $\Delta^d X_t = \Delta^{d-b} L_b \alpha\beta' X_t + u_t$, $(\alpha = \gamma_\perp)$ where u_t is $I(0)$, see Johansen (2008 p. 652) for details. The added lag structure using the usual lag operator in (2) is difficult to analyze, due to the characteristic function being a transcendental function and no condition has been given for the solution to be fractional of order d. Model $\mathcal{H}_r$, however, can be expressed as $\Pi(L)X_t = \Psi(L_b)\Delta^{d-b}X_t = \Delta^{d-b}L_b\alpha\rho' + \varepsilon_t$, where the polynomial $\Psi(y)$ is given by

$$\Psi(y) = (1-y)I_p - \alpha\beta' y - \sum_{i=1}^{k} \Gamma_i (1-y) y^i = -\alpha\beta' + (1-y)\sum_{i=0}^{k} \Psi_i (1-y)^i \tag{3}$$

and the coefficients satisfy $\sum_{i=0}^k \Psi_i = I_p + \alpha\beta'$ and $\Psi_k = (-1)^{k+1}\Gamma_k$. That is, $\Delta^{d-b}X_t$ satisfies a VAR in the lag operator L_b rather than the standard lag operator $L = L_1$.

This structure means that the solution of (1) and the criteria for fractionality of order d and cofractionality $d-b$ can be found by analysing the polynomial Ψ, just as for a CVAR model.

Contribution

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 (on compact subsets) of product moments of processes that can be close to critical processes of the form $\Delta^{-1/2}\varepsilon_t$.

In our asymptotic distribution results we distinguish between "weak cointegration" (when the true value $b_0 < 1/2$) and "strong cointegration" ($b_0 > 1/2$), using the terminology of Hualde and Robinson (2010a). Specifically, we prove that for i.i.d. errors with sufficient moments finite, the estimated cointegrating vectors are asymptotically mixed Gaussian (LAMN) when $b_0 > 1/2$ and asymptotically Gaussian when $b_0 < 1/2$, so that in either case 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.

History and relation to other work

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 (2008, 2010), 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 (20010) analyzes a test for no cointegration and in Lasak (2008) she analyzes maximum likelihood estimation and inference; in both cases assuming "strong cointegration". In the same model as Lasak, but assuming "weak cointegration", 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).

DF tests

The analysis of the fractionally cointegrated 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, 1991) 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.

Structure of paper

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 5 we find the asymptotic properties of the likelihood ratio test for cointegration rank. Section 6 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 $\varepsilon_i, i = 1, \ldots, t$, and give properties of the solution in Theorems 2 and 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

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

and satisfy the evaluation $|\pi_n(a)| \leq c n^{a-1},\ n \geq 1$, see Lemma A.4. The fractional difference operator applied to a process $Z_t, t = \ldots, -1, 0, 1, \ldots, T$, is defined by

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

provided the right hand side exists. Note that $\Delta^{-a_1}\Delta^{-a_2} = \Delta^{-a_1-a_2}$ and the useful relation $\Delta^{-a_1}\pi_t(a_2) = \pi_t(a_1+a_2)$. We collect a few simple results in a lemma, where $\mathsf{D}^m \Delta^a Z_t$ denotes the m'th derivative with respect to a.

Lemma 1 *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 $\sum_{n=0}^{\infty} |\xi_n| < \infty$.*

(i) If the initial values Z_{-n}, $n \geq 0$ are bounded, then $\mathsf{D}^m \Delta^a Z_t$ exists for $a \geq 0$ and is continuous in a for $a > 0$.

We next consider fractional differences of Z_t without fixing initial values.

(ii) If $a \geq 0$ then $\mathsf{D}^m \Delta^a Z_t$ is a stationary process with absolutely summable coefficients which is almost surely continuous in $a \geq \eta > 0$.

(iii) If $a > -1/2$, then $\mathsf{D}^m \Delta^a Z_t$ is a stationary process with square summable coefficients.

Proof. The existence is a simple consequence of the evaluation $|\mathsf{D}^m \pi_n(-a)| \leq c(1 + \log n)^m n^{-a-1}$ for $n \geq 1$, see Lemma A.4, which implies that $\mathsf{D}^m \pi_n(-a)$ is absolutely summable and continuous in a for $a > 0$ and square summable for $a > -1/2$. For case (ii) the continuity follows because $|\mathsf{D}^m \Delta^a Z_t - \mathsf{D}^m \Delta^{\tilde{a}} Z_t| \leq c|a - \tilde{a}| \sum_{n=0}^{\infty} (1 + \log n)^{m+1} n^{-\eta_1 - 1} |Z_{t-n}|$ for $\min(a, \tilde{a}) \geq \eta_1 > 0$. This random variable has a finite mean and is hence finite except on a null set which depends on η_1 but not a and $\tilde{a}$. It follows that $|\mathsf{D}^m \Delta^a Z_t - \mathsf{D}^m \Delta^{\tilde{a}} Z_t| \overset{a.s}{\to} 0$, for $a \to \tilde{a}$. ∎

For $a < 1/2$, an example of these results is the stationary linear process

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

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

$$\Delta_+^{-a} \varepsilon_t = \sum_{n=0}^{t-1} \pi_n(a) \varepsilon_{t-n}, \ t = 1, \ldots, T.$$

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

The main result in Theorem 2 is the representation of the solution of the equations (1) in terms of a stationary process, and we introduce these processes in

Definition 1 *We define the class $\mathcal{Z}$ as the set of multivariate stochastic processes Z_t for which*

$$Z_t = \xi \varepsilon_t + \Delta^{b_0} \sum_{n=0}^{\infty} \xi_n^* \varepsilon_{t-n},$$

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

This is a fractional version of the usual Beveridge-Nelson decomposition, where $\sum_{n=0}^{\infty} \xi_n \varepsilon_{t-n} = (\sum_{n=0}^{\infty} \xi_n) \varepsilon_t + \Delta \sum_{n=0}^{\infty} \xi_n^* \varepsilon_{t-n}$.

For the asymptotic analysis we apply the result, that when $a > 1/2$ and $E|\varepsilon_t|^q < \infty$ for some $q > 1/(a - 1/2)$, then for $Z_t \in \mathcal{Z}$ we have

$$T^{-a+1/2} \Delta_+^{-a} Z_{[Tu]}^+ \Longrightarrow \xi W_{a-1}(u) = \xi \Gamma(a)^{-1} \int_0^u (u - s)^{a-1} dW(s) \text{ on } D^p[0, 1], \ a > 1/2 \quad (4)$$

where $\Gamma(a)$ is the gamma function and W denotes $p-$dimensional Brownian motion (BM) generated by ε_t. For $Z_t = \varepsilon_t$, see e.g. Taqqu (1975), and for $Z_t \in \mathcal{Z}$ the proof is given in JN(2010, Lemma D.2). The process W_{a-1} is the corresponding fractional Brownian motion (fBM) of type II, and $\Longrightarrow$ 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 under the same conditions on ε_t, see Jakubowski, Mémin, and Pages (1989) for the case $Z_t = \varepsilon_t$,

$$T^{-a}\sum_{t=1}^{T}\Delta_+^{-a}L_a Z_t^+\varepsilon_t' \xrightarrow{d} \xi\int_0^1 W_{a-1}dW',\ a > 1/2 \tag{5}$$

where $\xrightarrow{d}$ denotes convergence in distribution on $\mathbb{R}^{p\times p}$. This result is proved in JN (2010, p. 65) for the univariate model and the same proof can be applied.

2.2 Solution of fractional autoregressive equations

We consider equation (1) written as $\Pi(L)X_t = \Delta^{d-b}L_b\alpha\rho'\pi_t(1) + \varepsilon_t$, where we have replaced the constant one, $1_{\{t\geq 1\}}$, by $\pi_t(1) = 1_{\{t\geq 0\}}$, but note that $\Delta^{d-b}L_b\pi_t(1)|_{d=b=0} = 1_{\{t\geq 1\}}$.

If $d < 1/2$ we show in Theorem 2 that the solution of the equation is stationary around its mean, but for $d \geq 1/2$ the solution is nonstationary and we need a solution that takes into account initial values $X_{-n}, n = 0, 1, \ldots,$ and random shocks $\varepsilon_1, \ldots, \varepsilon_t$. A general solution can be found using the 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 expression $\Pi_-(L)X_t$ is defined if we assume initial values of X_t fixed and bounded. We then find

$$\varepsilon_t = \Pi(L)X_t + \Delta^{d-b}L_b\alpha\rho'\pi_t(1) = \Pi_+(L)X_t + \Pi_-(L)X_t + \Delta^{d-b}L_b\alpha\rho'\pi_t(1),$$

we find by applying $\Pi_+(L)^{-1}$ on both sides that for $t = 1, 2, \ldots.$

$$X_t = \Pi_+(L)^{-1}\varepsilon_t - \Pi_+(L)^{-1}\Pi_-(L)X_t + \Pi_+(L)^{-1}\Delta^{d-b}L_b\alpha\rho'\pi_t(1) = \Pi_+(L)^{-1}\varepsilon_t + \mu_t + \xi_t, \tag{6}$$

The first term is the stochastic component generated by $\varepsilon_1, \ldots, \varepsilon_t$, the second a deterministic component generated by initial values, and the last comes from the deterministic term in the equation. An example of this is the well known result that $X_t = vX_{t-1} + \rho + \varepsilon_t$ has the solution $X_t = \sum_{i=0}^{t-1} v^i\varepsilon_{t-i} + v^t X_0 + \sum_{i=0}^{t-1} v^i\rho$ for any v, whereas for $|v| < 1$ we have a stationary solution $X_t = \sum_{i=0}^{\infty} v^i\varepsilon_{t-i}$.

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.

Definition 2 *Let ε_t be i.i.d.$(0,\Omega)$ in p dimensions and consider $m\times p$ matrices ξ_n for which $\sum_{n=0}^{\infty}|\xi_n|^2<\infty$, and define $C(z)=\sum_{n=0}^{\infty}\xi_n z^n,\ |z|<1$. If $\sum_{n=0}^{\infty}|\xi_n|<\infty$ [2] then the process $Z_t=C(L)\varepsilon_t=\sum_{n=0}^{\infty}\xi_n\varepsilon_{t-n}$ is fractional of order 0 if $C(1)\neq 0$. A process Z_t is fractional of order $d>0$ (denoted $Z_t\in\mathcal{F}(d)$) 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

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

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 which guarantee that the process is fractional of order d and cofractional from d to $d-b$. These are given in terms of an explicit condition on 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), see Johansen (2008, Theorem 8 and 2009, Theorem 3). It is related to previous representation theorems of Engle and Granger (1987) and Johansen (1988, 1991) for the cointegrated VAR. Below we use the notation $\beta_\perp$ for a $p\times(p-r)$ matrix of full rank for which $\beta'\beta_\perp=0$, and note the orthogonal decomposition

$$I_p=\beta_0(\beta_0'\beta_0)^{-1}\beta_0'+\beta_{0\perp}(\beta_{0\perp}'\beta_{0\perp})^{-1}\beta_{0\perp}'=\beta_0\bar{\beta}_0'+\beta_{0\perp}\bar{\beta}_{0\perp}'. \tag{8}$$

Theorem 2 *Let $\Pi(z)=(1-z)^{d-b}\Psi(1-(1-z)^b)$ be given by (3) for any $0<b\leq d$. Assume that $\det(\Psi(y))=0$ implies that either $y=1$ or $y\notin\mathbb{C}_b$ and that α and β have rank $r<p$. Then*

$$(1-z)^d\Pi(z)^{-1}=C+(1-z)^bH(1-(1-z)^b), \tag{9}$$

where $H(y)$ is regular in a neighborhood of $\mathbb{C}_b$ if and only if $\det(\alpha_\perp'\Gamma\beta_\perp)\neq 0$ where $\Gamma=I_p-\sum_{i=1}^k\Gamma_i$. In this case $C_1=H(1)\neq 0$ because $\beta'H(1)\alpha=-I_r$, and

$$C=\beta_\perp(\alpha_\perp'\Gamma\beta_\perp)^{-1}\alpha_\perp'. \tag{10}$$

We define $F(z)=H(1-(1-z)^b)=\sum_{n=0}^{\infty}\tau_n z^n,\ |z|<1$, with $\sum_{n=0}^{\infty}\tau_n^2<\infty$.

For $d\geq 1/2$ we represent the solution of (1) as

$$X_t=C\Delta_+^{-d}\varepsilon_t+\Delta_+^{-(d-b)}Y_t^++\xi_t+\mu_t,\ t=1,\ldots,T, \tag{11}$$

where $\mu_t=-\Pi_+(L)^{-1}\Pi_-(L)X_t$, $\xi_t=\Pi_+(L)^{-1}\alpha\rho'\Delta^{d-b}L_b\pi_t(1)$, and $Y_t=\sum_{n=0}^{\infty}\tau_n\varepsilon_{t-n}$ and $Y_t^+=\sum_{n=0}^{t-1}\tau_n\varepsilon_{t-n}$ are fractional of order zero.

For $d<1/2$ we represent the solution of (1) as

$$X_t=C\Delta^{-d}\varepsilon_t+\Delta^{-(d-b)}Y_t+\xi_t,\ t=1,\ldots,T. \tag{12}$$

In both cases there is no γ for which $\gamma' X_t\in\mathcal{F}(c)$ for $c<d-b$.

[2] SJ: Jeg har erstattet "$C(z)$ can be extended to a continuous function on the boundary $|z|=1$ " med $\sum_{n=0}^{\infty}|\xi_n|<\infty$ som mange nok vil finde lidt mindre mystisk

Proof. The proof is given in Johansen (2008, Theorem 8 and 2009, Theorem 3). The condition (10) is necessary and sufficient for the representation (9) because if $\det(\alpha_\perp'\Gamma\beta_\perp) = 0$ then we get terms of the form $(1-z)^{-(d-ib)}$, $i \geq 2$, corresponding to models for $I(i)$ variables, $i \geq 2$ in the CVAR context, see Johansen (2008, Theorem 9). ∎

Thus, for $0 < r < p$, X_t is fractional of order d, and because $\beta' C = 0$, X_t is cofractional since $\beta' X_t = \Delta_+^{-(d-b)}\beta' Y_t^+ + \beta'\mu_t + \beta'\xi_t$ $(d \geq 1/2)$ is fractional of order $d-b$, and no linear combination gives other orders of fractionality, whereas for $d < 1/2$, $\beta' X_t = \Delta^{-(d-b)}\beta' Y_t + \beta'\xi_t$ is fractional of order $d-b$ and in fact stationary if $d-b < 1/2$.

If $r = 0$, we have $\alpha = \beta = 0$, $\alpha_\perp = \beta_\perp = I_p$, and $C = \Gamma^{-1}$ is assumed to have full rank, and thus X_t is fractional of order d and not cofractional.

Finally, if $r = p$ then $\alpha\beta'$ has full rank and $C = 0$ so that $X_t = \Delta_+^{-(d-b)} Y_t^+ + \mu_t + \xi_t$ is fractional of order $d-b$.

The stochastic properties of X_t are given in Theorem 2 in terms of the process $U_t = C_0\varepsilon_t + \Delta^{b_0} Y_t \in \mathcal{Z}$, see (1) it is seen from Theorem 3 that also $Y_t \in \mathcal{Z}$.

Theorem 3 *Under the assumptions of Theorem 2 and assuming that the roots of* $\det\Psi(y) = 0$ *are either* $y = 1$ *or* $|y| > 1$, *and hence outside* $\mathbb{C}_{b\vee 1}$, *we have*

$$Y_t = \sum_{n=0}^{\infty} \tau_n \varepsilon_{t-n} = C_1 \varepsilon_t + \Delta^b \sum_{n=0}^{\infty} \tau_n^* \varepsilon_{t-n}, \tag{13}$$

$$\sum_{n=0}^{\infty} |\tau_n| < \infty, \quad \sum_{n=0}^{\infty} |\tau_n^*| < \infty, \text{ and } \sum_{h=-\infty}^{\infty} |E(Y_t Y_{t-h}')| < \infty.$$

Note that for $\xi_t^* = \Delta^b[\sum_{n=0}^{\infty} \tau_n^* \alpha\rho' L_b \pi_{t-n}(1) - C_1\alpha\rho'\pi_t(1)] = O(t^{-b})$ *we have*

$$\xi_t = C_1\alpha\rho'\pi_t(1) + \xi_t^* \text{ and } \beta'\xi_t = -\rho'\pi_t(1) + O(t^{-b}). \tag{14}$$

Proof. *Proof of* (13): The extra assumption, $y \notin \mathbb{C}_{b\vee 1}$, implies that for any $b > 0$, $H(y) = \sum_{k=0}^{\infty} h_k y^k$ is regular for $|y| < 1 + \delta$ for some $\delta > 0$, so that h_k decrease exponentially and $\sum_{k=0}^{\infty} |h_k| < \infty$. From the expansion $1 - (1-z)^b = \sum_{n=1}^{\infty} b_n z^n$ we find that if $0 < b \leq 1$, then $b_n \geq 0$ and $\sum_{n=1}^{\infty} b_n = 1$. Therefore

$$H(1-(1-z)^b) = \sum_{k=0}^{\infty} h_k (\sum_{n=1}^{\infty} b_n z^n)^k = \sum_{n=0}^{\infty} \tau_n z^n,$$

satisfies $\sum_{n=0}^{\infty} |\tau_n| = \sum_{k=0}^{\infty} |h_k| (\sum_{n=0}^{\infty} b_n)^k = \sum_{k=0}^{\infty} |h_k| < \infty$.

For $b > 1$ we need another argument. Because $H(y) = \sum_{n=0}^{\infty} \tau_n y^n$ is regular in a neighborhood of $\mathbb{C}_b$ and $\sum_{n=0}^{\infty} \tau_n^2 < \infty$ we can define the transfer function

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

We then apply the proof in JN (2010, Lemma 1), which shows that because $|\partial\phi(e^{i\lambda})/\partial\lambda|$ is square integrable when $b > 1/2$, we have $\sum_{n=0}^{\infty} (\tau_n n)^2 < \infty$ and hence $\sum_{n=0}^{\infty} |\tau_n| < \infty$. It follows that $\sum_{h=-\infty}^{\infty} |E(Y_t Y_{t-h}')| < \infty$.

Finally we have $H(y) = H(1) + (1-y)H^*(y)$, where $H^*(y) = \sum_{n=0}^{\infty} \tau_n^* y^n$ is regular in a neighborhood of $\mathbb{C}_{b\vee 1}$, which shows (13), and we can repeat the above arguments which show that $\sum_{n=0}^{\infty} |\tau_n^*| < \infty$.

Proof of (14): The expression for ξ_t is, noting that $L_b\pi_{t-n}(1) = 0, n \geq t$,

$$\xi_t = \sum_{n=0}^{t-1} \tau_n \alpha\rho L_b \pi_{t-n}(1) = \sum_{n=0}^{\infty} \tau_n \alpha\rho L_b \pi_{t-n}(1)$$

which by (13) is

$$\begin{aligned} \xi_t &= (C\Delta^{-d} + \Delta^{-d+b} \sum_{n=0}^{\infty} \tau_n L_n)\alpha\rho' \Delta^{d-b} L_b \pi_t(1) = \sum_{n=0}^{\infty} \tau_n \alpha\rho' L_b \pi_{t-n}(1) \\ &= C_1 \alpha\rho' L_b \pi_t(1) + \Delta^b \sum_{n=0}^{\infty} \tau_n^* \alpha\rho L_b \pi_{t-n}(1), \end{aligned}$$

so that $\beta' C_1 \alpha = -I_r$ implies that $\beta'\xi_t = \rho'\pi_t(1) + O(t^{-b})$. ■

Note that X_t has deterministic term generated from ρ which converges to the constant $C_1\alpha\rho'$ so that $\beta' X_t + \rho'$ has asymptotically mean zero when $d - b < 1/2$.

2.4 Assumptions for asymptotic analysis

We next formulate some assumptions needed for the asymptotic analysis of estimators and the likelihood function for model $\mathcal{H}_r$ with $0 < b \leq d$ and the two submodels $\mathcal{H}_r(d = b)$. We define the parameter set

$$\mathcal{N} = \{d, b : 0 < b \leq d \leq d_1\} \tag{15}$$

for some $d_1 > 0$, which can be arbitrarily large.

Assumption 1 *The process $X_t,\ t = 1, \ldots, T$, is generated by model $\mathcal{H}_r$ in (1) for some $k \geq 1$ and $r \geq 0$, or model $\mathcal{H}_r$ for $k = 0$ and $r > 0$, or model $\mathcal{H}_r(d = b)$ with $k \geq 0, r \geq 0$.*

Assumption 2 *The errors ε_t are i.i.d.$(0, \Omega_0)$ with $\Omega_0 > 0$ and $E|\varepsilon_t|^8 < \infty$.*

Assumption 3 *The initial values $X_{-n}, n \geq 0$, are bounded.*

Assumption 4 *The true parameter values $\lambda_0 = (\alpha_0, \beta_0, \rho_0, d_0, b_0, \Gamma_{01}, \ldots, \Gamma_{0k}, \Omega_0)$ satisfy $(d_0, b_0) \in \mathcal{N},\ d_0 - b_0 < 1/2,\ \Gamma_{0k} \neq 0$, α_0 and β_0 are $p \times r$ of rank r, and that $\det(\alpha_{0\perp}' \Gamma_0 \beta_{0\perp}) \neq 0$. Thus, if $r < p$, $\det(\Psi(y)) = 0$ has $p - r$ unit roots, and the remaining roots of $\det(\Psi(y))$ are outside $\mathbb{C}_{b_0 \vee 1}$. If $k = r = 0$ only $d_0 > 0$ is assumed.*

Importantly, in Assumption 2, 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. Assumption 3 about initial values is needed so that $\Delta^d X_t$ is defined for any $d \geq 0$, see Lemma 1. In Assumption 4 about the true values we include the condition that $d_0 - b_0 < 1/2$, which appears to be perhaps the most empirically relevant range of values for $d_0 - b_0$, see e.g. Henry and Zaffaroni (2003) and the references in the introduction, because in this case $\beta_0' X_t$ is (asymptotically) stationary so that $E(\beta_0' X_t + \rho_0) = 0$. Assumption 4 also includes the condition for cofractionality when $r > 0$, which ensures that X_t is fractional of order d_0 and $\beta_0' X_t$ is fractional of order $d_0 - b_0$. The condition $\Gamma_{0k} \neq 0$ 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, see Lemma 7.

2.5 Initial values

From (11) in Theorem 2 we find that for $a \geq 0$ that $\Delta^a X_t = \Delta_+^a X_t + \Delta_-^a X_t$ has the representation

$$\Delta^a X_t = \begin{cases} \Delta_+^{a-d_0}(C_0\varepsilon_t + \Delta_+^{b_0} Y_t^+) + \Delta_+^a \xi_t + \Delta_+^a \mu_t + \Delta_-^a X_t, & d_0 \geq 1/2 \\ \Delta^{a-d_0}(C_0\varepsilon_t + \Delta^{b_0} Y_t) + \Delta^a C_1 \pi_t(1) + \Delta^a \xi_t^*, & d_0 < 1/2 \end{cases} \tag{16}$$

for $t = 1, \ldots, T$, where μ_t, ξ_t and ξ_t^* are given in Theorem 2 and 3.

The theory in this paper will be developed for observations $X_1, \ldots, X_T$ generated by (1) assuming that, for $d_0 \geq 1/2$, all initial values are observed, that is, conditional on $X_{-n},\ n = 0, 1, \ldots$, and under the assumption that they are bounded, which seems a reasonable condition in practice. Thus, we follow the standard approach in the literature on inference for nonstationary autoregressive processes, where the initial values are observed but not modeled and inference is conditional on them. However, we do not set initial values equal to zero as is often done in the literature on fractional processes, but instead assume only that they are observed unmodelled bounded constants, which represents a significant generalization and makes the results more applicable.

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 $X_{-n}, n = 0, \ldots, N_0$, which are observed unmodelled 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 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.

For $d_0 < 1/2$ we use the representation (16) of X_t and $\Delta^a X_t$ as stationary processes around their mean.

2.6 Identification of parameters

Identification was discussed in JN (2010, Section 2.3). Let $\lambda = (d, b, \alpha\beta', \rho, \Gamma_1, \ldots, \Gamma_k, \Omega)$ denote all the parameters and let the corresponding probability measure be P_λ and the corresponding characteristic function be Π_λ, see (3). the parameter λ is identified if $P_\lambda = P_{\bar{\lambda}}$ implies that $\lambda = \bar{\lambda}$, or equivalently if $\Pi_\lambda(z) = \Pi_{\bar{\lambda}}(z)$ for all z. It was shown in JN (2010) that the parameters are identified if $k = \bar{k}$, $\Gamma_k \neq 0$, and $\bar{\Gamma}_k \neq 0$. See Lemma 7 for a proof that this also implies that the asymptotic variance is positive. JN (2010, Section 2.3) has a fuller discussion in the univariate case and an example of the indeterminacy between d, b, and k.

Note that if $k = r = 0$ the only parameters are d and Ω which are identified.

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 give the limit of the profile likelihood and the result on consistency of the MLE.

3.1 Calculation of MLE, profile likelihood function and its limit

In (3) we eliminate $\Psi_k = I_p + \alpha\beta' - \sum_{i=0}^{k-1} \Psi_i$ and define the regressors

$$X_{it} = (\Delta^{d+ib} - \Delta^{d+kb}) X_t,\ i = -1, \ldots, k-1,\ X_{kt} = \Delta^{d+kb} X_t, \tag{17}$$

and

$$\varepsilon_t(\lambda) = X_{kt} - \alpha\beta' X_{-1t} - \alpha\rho'\Delta^{d-b}L_b\pi_t(1) + \sum_{i=0}^{k-1}\Psi_i X_{it}. \tag{18}$$

The Gaussian likelihood function conditional on initial values $X_{-n}, n \geq 0$, is

$$-2T^{-1}\log L_T(\lambda) = \log\det(\Omega) + tr\{\Omega^{-1}T^{-1}\sum_{t=1}^{T}\varepsilon_t(\lambda)\varepsilon_t(\lambda)'\}, \tag{19}$$

where $\lambda = (d, b, \alpha, \beta, \rho, \Psi_*, \Omega)$ $(\Psi_* = (\Psi_0, \ldots, \Psi_{k-1}))$ are freely varying parameters.

For given values of $\psi = (d, b)$ we calculate the processes X_{kt}, $\{X_{it}\}_{i=-1}^{k-1}$, and $\Delta^{d-b}L_b\pi_t(1)$ for $d > b > 0$, when initial values are bounded, see Lemma 1. We calculate the MLEs $(\hat{\alpha}(\psi), \hat{\beta}(\psi), \hat{\rho}(\psi), \hat{\Psi}_*(\psi), \hat{\Omega}(\psi))$ for given $\psi = (d, b)$, and the partially maximized likelihood or likelihood profile,

$$\ell_{T,r}(\psi) = -2T^{-1}\log L_T(\psi, \hat{\alpha}(\psi), \hat{\beta}(\psi), \hat{\rho}(\psi), \hat{\Psi}_*(\psi), \hat{\Omega}(\psi)), \tag{20}$$

as continuous functions of ψ by reduced rank regression of X_{kt} on $(X_{-1t}', \Delta^{d-b}L_b\pi_t(1))'$ corrected for $\{X_{it}\}_{i=0}^{k-1}$, see Anderson (1951) and Johansen (1996). Finally the MLE and maximized likelihood can be calculated by optimizing $\ell_{T,r}(\psi)$ as a function of $\psi = (d, b)$ by a numerical procedure.

Note that for $r = p$ the likelihood profile $\ell_{T,p}(\psi)$ is found by regression of X_{kt} on $\{X_{it}\}_{i=-1}^{k-1}$ and $\Delta^{d-b}L_b\pi_t(1)$, i.e.

$$\ell_{T,p}(\psi) = \log\det(SSR_T(\psi)) = \log\det(T^{-1}\sum_{t=1}^{T}R_tR_t'), \tag{21}$$

where $R_t = (X_{kt}|\{X_{it}\}_{i=-1}^{k-1}, \Delta^{d-b}L_b\pi_t(1))$ denotes the regression residual.

We next want to define the probability limit, $\ell_p(\psi)$, of the profile likelihood function, $\ell_{T,p}(\psi)$. We note that Theorem 2 gives the properties of X_t at the true parameter point in terms of the stationary process $U_t = C_0 + \Delta^{b_0}Y_t$. Corresponding to X_{it} we define

$$\begin{aligned}
\beta_0'U_{jt} &= (\Delta^{d+jb} - \Delta^{d+kb})\Delta^{-d_0+b_0}\beta_0'Y_t, \ j = -1, \ldots, k-1 \\
\beta_{0\perp}'U_{it} &= (\Delta^{d+ib} - \Delta^{d+kb})\Delta^{-d_0}\beta_{0\perp}'U_t, \ \text{for } i \text{ so that } d + ib - d_0 > -1/2 \\
\beta_0'U_{kt} &= \Delta^{d+kb}\Delta^{-d_0+b_0}\beta_0'Y_t, \ \text{and } \beta_{0\perp}'U_{kt} = \Delta^{d+kb}\Delta^{-d_0}\beta_{0\perp}'U_t,
\end{aligned} \tag{22}$$

It is seen that the stochastic behavior of $\beta_0'X_{it}$ is determined by $\beta_0'U_{it}$, which is stationary because $d + jb - d_0 + b_0 \geq -d_0 + b_0 > -1/2$, whereas $\beta_{0\perp}'U_{it}$ is only stationary if $d + ib - d_0 > -1/2$. For a given $\psi = (d, b)$, we therefore define the class of stationary processes

$$\mathcal{F}(\psi) = \{\beta_0'U_{jt}, \beta_{0\perp}'U_{it} : i, j < k, \quad \text{and } d + ib - d_0 > -1/2\}.$$

The limit of $\log\det(SSR_T(\psi))$ is infinite if X_{kt} is nonstationary and finite if X_{kt} is (asymptotically) stationary, and we therefore define the subsets of $\mathcal{N}$,

$$\begin{aligned}
\mathcal{N}_{\text{div}}(\kappa) &= \mathcal{N} \cap \{d, b : d + kb - d_0 \leq -1/2 + \kappa\}, \kappa \geq 0, \\
\mathcal{N}_{\text{conv}}(\kappa) &= \mathcal{N} \cap \{d, b : d + kb - d_0 \geq -1/2 + \kappa\}, \kappa > 0, \\
\mathcal{N}_{\text{conv}}(0) &= \mathcal{N} \cap \{d, b : d + kb - d_0 > -1/2 + \kappa\},
\end{aligned}$$

and note that $\mathcal{N}_{\text{div}}(\kappa)$ is a family of sets decreasing (as $\kappa \to 0$) to the set $\mathcal{N}_{\text{div}}(0)$, where X_{kt} is nonstationary and $\log\det(SSR_T(\psi))$ diverges, and $\mathcal{N}_{\text{conv}}(\kappa)$ increase to the set $\mathcal{N}_{\text{conv}}(0)$ where X_{kt} is stationary and $\log\det(SSR_T(\psi))$ converges. We therefore define the limit likelihood function, $\ell_p(\psi)$, as

$$\ell_p(\psi) = \begin{cases} \infty & \text{if } \psi \in \mathcal{N}_{\text{div}}(0), \\ \log\det(Var(U_{kt}|\mathcal{F}_{stat}(\psi))) & \text{if } \psi \in \mathcal{N}_{\text{conv}}(0). \end{cases} \tag{23}$$

where we use the notation for any random vectors W and V with finite variance

$$Var(W|V) = Var(W) - Cov(W,V)Var(V)^{-1}Cov(V,W).$$

The above divergence and convergence results are proved in Theorem 4. It follows that the probability limit of $\ell_{T,p}(\psi)$ for a fixed $\psi = (d,b)$ is $\ell_p(\psi)$. [3]

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

We now show that if $\Delta^{b+kd}X_t$ is nonstationary, the likelihood profile function $\ell_{T,p}(\psi)$ is uniformly divergent on $\mathcal{N}_{\text{div}}(0)$, and if $\Delta^{b+kd}X_t$ is (asymptotically) stationary the likelihood converges in probability uniformly on compact sets of $\mathcal{N}_{\text{conv}}(0)$ to the deterministic limit $\ell_p(\psi)$ for $T \to \infty$. 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 model with $d = b$.

We define the compact set

$$\mathcal{K}(\eta,\eta_1) = \{d, b : \eta \leq b \leq d \leq d_1, \eta_1 \leq d - b\}$$

for $\eta > 0, \eta_1 \geq 0$, which is a family of sets such that $\mathcal{K}(\eta,\eta_1) \subset \mathcal{N}$ and $\mathcal{K}(\eta,0)$ increase to $\mathcal{N}$ as $\eta \to 0$.

Theorem 4 *Let Assumptions 1-4 hold and assume that* $(d_0,b_0) \in \mathcal{K}(\eta,\eta_1)$.

(i) Let $0 < \eta < \min(1/3, 1-2(d_0-b_0))$ *and* $\eta_1 > 0$ *and suppose* $E|\varepsilon_t|^q < \infty$ *for* $q > 2/\eta$. *The likelihood function for* $\mathcal{H}_p$ *satisfies*

$$\inf_{\psi\in\mathcal{N}_{div}(\kappa)\cap\mathcal{K}(\eta,\eta_1)} \ell_{T,p}(\psi) \overset{P}{\to} \infty \text{ as } (\kappa,T) \to (0,\infty), \tag{24}$$

$$\ell_{T,p}(\psi) \Longrightarrow \ell_p(\psi) \text{ on } C(\mathcal{N}_{conv}(\kappa)\cap\mathcal{K}(\eta,\eta_1)) \text{ as } T \to \infty \text{ for any } \kappa > 0. \tag{25}$$

The function $\ell_p(\psi)$, *has a strict minimum at* $\psi = \psi_0$, *that is*

$$\ell_p(\psi) \geq \ell_p(\psi_0) = \log\det(\Omega_0), \psi \in \mathcal{N} \tag{26}$$

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

(ii) Suppose $X_{-n} = 0, n > N_0$ *and* $E|\varepsilon_t|^q < \infty$ *for all* $q < \infty$.[4] *Then (24), (25), and (26) hold on the larger sets with* $\eta > 0$ *and* $\eta_1 = 0$.

[3] SJ: Husk kontinuiteten af $\ell_r(\psi)$
Her vi kun brug for ℓ_p?

[4] SJ: Er det nu rigtigt er det ikke nok med 8 momenter?

(iii) Suppose $\eta \le b_0 = d_0 \le d_1, b_0 \neq 1/2$, and $E|\varepsilon_t|^q < \infty$ for $q > 2/\eta$. Then, for the model $\mathcal{H}_r(b = d)$, the results (24), (25), and (26) hold on the sets $\mathcal{N}_{div}(\kappa) \cap \{b = d\}$ and $\mathcal{N}_{conv}(\kappa) \cap \{b = d\}$, respectively.

The proof is given in Appendix B. Note that the larger the compact set $\mathcal{K}(\eta, \eta_1)$, the more moments are needed. When consideration is restricted to the model $\mathcal{H}_r(b = d)$ and a parameter set defined by $\eta > 1/4$, i.e. in particular if consideration is restricted to the case of "strong cointegration" where $b > 1/2$, then the moment condition reduces to $E|\varepsilon_t|^8 < \infty$ [5](from Assumption 2).

The reason for the restriction in $\mathcal{K}(\eta, \eta_1)$ away from the boundary $\{d = b\}$ is that close to that boundary, the contribution from initial values does not vanish uniformly. This uniformity can be obtained by setting $X_{-n} = 0, n > N_0$, and if $d = b$ and $b_0 \neq 1/2$ the problem does not arise.

We now derive the important consequence of Theorem 4.

Theorem 5 *Let the assumptions of Theorem 4 be satisfied.*

(i) With probability converging to one, the maximum likelihood estimator in model $\mathcal{H}_r, r = 0, \ldots, p$, exists uniquely on $\mathcal{K}(\eta, \eta_1)$ for $\eta > 0, \eta_1 > 0$, and is consistent.

(ii) Suppose $X_{-n} = 0, n > N_0$. Then the results hold on $\mathcal{K}(\eta, 0) = \{d, b : \eta \le b \le d \le d_1\}$ for $\eta > 0$.

(iii) For the model $\mathcal{H}_r(b = d)$, existence, uniqueness, and consistency in model $\mathcal{H}_r, r = 0, \ldots, p$, hold on $\{d, b : \eta \le b = d \le d_1\}$ for $\eta > 0$ provided $b_0 \neq 1/2$.

Proof. Assume that (24) and (25) hold. We start with model $\mathcal{H}_p$, see (1), where α and β are $p \times p$. The result (24) shows that $P(\hat{\psi} \in \mathcal{N}_{\text{conv}}(\kappa) \cap \mathcal{K}(\eta, \eta_1)) \to 1$. On the set $\mathcal{N}_{\text{conv}}(\kappa) \cap \mathcal{K}(\eta, \eta_1)$ the convergence in distribution of the continuous process $\ell_{T,p}(\psi) = \log\det(SSR_T(\psi))$ in (25) shows that the probability limit $\ell_p(\psi)$ is continuous on $\mathcal{N}_{\text{conv}}(\kappa) \cap \mathcal{K}(\eta, \eta_1)$ and hence continuous on $\mathcal{N}_{\text{conv}}(0)$ if $E|\varepsilon_t|^q < \infty$ for all q.

Let $N(\psi_0, \epsilon) = \{\psi : |\psi - \psi_0| < \epsilon\}$ be a small neighborhood around ψ_0 and denote $\mathcal{N}_0 = \mathcal{N}_{\text{conv}}(\kappa) \cap \mathcal{K}(\eta, \eta_1)$. Because $\ell_p(\psi)$ is continuous and $> \ell_p(\psi_0)$ if $\psi \neq \psi_0$, see (26), and $\mathcal{N}_0 \backslash N(\psi_0, \epsilon)$ is compact and does not contain ψ_0, then $\min_{\psi \in \mathcal{N}_0 \backslash N(\psi_0, \epsilon)}(\ell_p(\psi) - \ell_p(\psi_0)) \ge c_0 > 0$. By the uniform convergence of $\ell_{T,p}(\psi)$ to $\ell_p(\psi)$ on $\mathcal{N}_0$, we can take for any $\xi > 0$ a $T_0(\xi, \epsilon)$ such that for all $T \ge T_0(\xi, \epsilon)$ we have

$$P(\min_{\psi \in \mathcal{N}_0 \backslash N(\psi_0, \epsilon)} |\ell_{T,p}(\psi) - \ell_p(\psi)| \le c_0/3) \ge 1 - \xi/2,$$

and therefore on this set we have

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

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

$$\min_{\psi \in \mathcal{N}_0 \backslash N(\psi_0, \epsilon)} (\ell_{T,r}(\psi) - \log\det(\Omega_0)) \ge \min_{\psi \in \mathcal{N}_0 \backslash N(\psi_0, \epsilon)} (\ell_{T,p}(\psi) - \log\det(\Omega_0)) \ge 2c_0/3.$$

[5]SJ: Det gør den da ikke. Man får at den reducerer til $E|\varepsilon|^q < \infty$ for et $q > 8$.

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

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

which implies that the minimum of $\ell_{T,r}(\psi)$ is attained at a point in $N(\psi_0, \epsilon)$. Thus the maximum likelihood estimator of ψ in model $\mathcal{H}_r$ exists with probability $1-\eta$ and is contained in the set $N(\psi_0, \epsilon)$, which proves consistency, see also van der Vaart (1998, Theorem 5.7)[6]. The second derivative of $\ell_p(\psi)$ is positive definite at $\psi = \psi_0$ and therefore in $N(\psi_0, \epsilon)$ for ϵ small. It follows from Theorem 9 and Lemma 9 that also the second derivative of $\ell_{T,p}(\psi)$ is positive definite in $N(\psi_0, \epsilon)$ with large probability, but then $\ell_{T,p}(\psi)$ is convex and the minimum is unique. The estimators $\hat{\alpha}(\psi), \hat{\beta}(\psi), \hat{\rho}(\psi), \hat{\Psi}_*(\psi), \hat{\Omega}(\psi)$, see (20) are continuous functions of ψ and therefore also consistent.

If $X_{-n} = 0, n > N_0$, or if $d = b$ and $d_0 \neq 1/2$, then the same proof can be used with $\mathcal{N}_{\text{conv}}(\kappa) \cap \mathcal{K}(\eta, 0)$ and $\mathcal{N}_{\text{conv}}(\kappa) \cap \{d, b : \eta \leq b = d \leq d_1\}$, respectively, instead of $\mathcal{N}_0$. ■

The result in Theorem 5 on existence and consistency of the MLE involves analyzing the likelihood function on the set of admissible values $0 < b < d$. 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 (11), 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).[7]

Unlike previous consistency results, our Theorem 5 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 $(T, \kappa) \to (\infty, 0)$, see (87) in Lemma A.8 below.

---

[6] SJ: Kan vi ikke droppe dette argument og bare henvise til van der Vaart. Det sparer næsten en side

[7] In 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.

4 Asymptotic distribution of maximum likelihood estimators

In this section we exploit that the maximum likelihood estimator is consistent and expand the likelihood in a neighborhood of the true parameter to find the asymptotic distribution of the conditional maximum likelihood estimator.

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

The likelihood function in a neighborhood of the true value is expressed in terms of $\varepsilon_t(\lambda)$ see (18).

The parameter ρ has information proportional to $\sum_{t=1}^T [\Delta^{d-b} L_b \pi_t(1)]^2 = O(T^{1-2(d-b)})$ for (d, b) close to (d_0, b_0) when $d_0 - b_0 < 1/2$. In order to get a $T^{1/2}$ consistent parameter we introduce $\theta_\rho = (\rho - \rho_0)T^{-(d-b)}$ or $\rho = \rho_0 + \theta_\rho T^{d-b}$, so that the information for θ_ρ is proportional to T.

The parameters α and β are not identified because $\alpha\beta' = (\alpha\beta'\bar{\beta}_0)(\beta'(\bar{\beta}_0'\beta)^{-1}) = \tilde{\alpha}\tilde{\beta}'$, but we can without loss of generality normalize β so that $\beta_0'\beta = I_r$. With this choice it follows from (8) that

$$\beta = \beta_0 + \beta_{0\perp}(\bar{\beta}_{0\perp}'\beta) = \beta_0 + \beta_{0\perp}\tilde{\theta}_\beta,$$

say. Finally we need the notation for the regressors

$$X_{-1t}^{ext} = \begin{pmatrix} \beta_{0\perp}' X_{-1t} \\ \Delta^{d-b} L_b \pi_t(1) \end{pmatrix}, \; X_{-1}^{cent} = \beta_0' X_{-1t} + \rho_0' \Delta^{d-b} L_b \pi_t(1). \tag{27}$$

For $b_0 > 1/2$ we let $N(\psi_0, \kappa_0) = \{|\psi - \psi_0| \leq \kappa_0\}$. Then for $(d, b) \in N(\psi_0, \kappa_0)$ we have the $\delta_{-1} = d - b - d_0 = (d - b - d_0 + b_0) - b_0 \leq -b_0 + 2\kappa_0 < -1/2$ and $d + ib - d_0 \geq -\kappa_0$, for $i \geq 0$, and hence $\beta_{0\perp}' X_{-1t}$ is the only nonstationary process in $\varepsilon_t(\lambda)$ and this is only possible for $b_0 > 1/2$. The information for $\tilde{\theta}_\beta$ is proportional to $\sum_{t=1}^T (\beta_{0\perp}' X_{-1t})(\beta_{0\perp}' X_{-1t})' = O_P(T^{-2\delta_{-1}})$. We therefore introduce the normalized parameter $\theta_\beta = \bar{\beta}_{0\perp}'(\beta - \beta_0)T^{-(\delta_{-1}+1/2)} = \tilde{\theta}_\beta T^{-(\delta_{-1}+1/2)}$ or $\beta = \beta_0 + \beta_{0\perp}\theta_\beta T^{\delta_{-1}+1/2}$, so the information for θ_β is proportional to T. We let $\theta = (\theta_\beta', \theta_\rho')'$ be $(p - r + 1) \times r$ and define $N_T = diag(T^{\delta_{-1}+1/2} I_{p-r}, T^{d-b})$ which implies that $\beta' X_{-1t} + \rho' \Delta^{d-b} L_b \pi_t(1)$ can be written as

$$\beta_0' X_{-1t} + \rho_0' \Delta^{d-b} L_b \pi_t(1) + T^{\delta_{-1}+1/2} \theta_\beta' \beta_{0\perp}' X_{-1t} + \theta_\rho' T^{d-b} = X_{-1}^{cent} + \theta' N_T X_{-1t}^{ext},$$

see (27). Let $V_t = (X_{-1}^{cent\prime}, \{X_{it}'\}_{i=0}^{k-1}, X_{kt}')'$ and define, for $\phi = (d, b, \alpha, \Psi_*)$,

$$\varepsilon_t(\lambda) = \varepsilon_t(\phi, \theta) = \alpha\theta' N_T X_{-1t}^{ext} + (-\alpha, \Psi_*, I_p)V_t, \tag{28}$$

see (18). The product moments needed to calculate the conditional likelihood function $-2T^{-1}\log L_T(\phi, \theta)$, see (19), are

$$\begin{pmatrix} \mathcal{A}_T & \mathcal{C}_T \\ \mathcal{C}_T' & \mathcal{B}_T \end{pmatrix} = T^{-1} \sum_{t=1}^T \begin{pmatrix} N_T X_{-1t}^{ext} \\ V_t \end{pmatrix} \begin{pmatrix} N_T X_{-1t}^{ext} \\ V_t \end{pmatrix}'. \tag{29}$$

Note that the matrices $N_T, \mathcal{A}_T, \mathcal{B}_T$, and $\mathcal{C}_T$ depend on $\psi = (d, b)$. We indicate the values for $\psi = \psi_0$ by $N_T^0, \mathcal{A}_T^0, \mathcal{B}_T^0$, and $\mathcal{C}_T^0$. Finally we define

$$\mathcal{C}_{\varepsilon T}^0 = T^{-1/2} \sum_{t=1}^T N_T^0 X_{-1t}^{ext} \varepsilon_t'. \tag{30}$$

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

$$\log\det(\Omega) + tr\{\Omega^{-1}(\alpha\theta'\mathcal{A}_T\theta\alpha' + (-\alpha,\Psi_*,I_p)\mathcal{B}_T(-\alpha,\Psi_*,I_p)' + 2\alpha\theta'\mathcal{C}_T(-\alpha,\Psi_*,I_p)')\}. \quad (31)$$

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

$$\hat\theta(\psi,\alpha,\Psi_*,\Omega) = -\mathcal{A}_T^{-1}\mathcal{C}_T(-\alpha,\Psi_*,I_p)'\Omega^{-1}\alpha(\alpha'\Omega^{-1}\alpha)^{-1}, \quad (32)$$

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

$$\begin{aligned}&\log\det(\Omega) + tr\{\Omega^{-1}(-\alpha,\Psi_*,I_p)\mathcal{B}_T(-\alpha,\Psi_*,I_p)'\} \\ &-tr\{(-\alpha,\Psi_*,I_p)\mathcal{C}_T'\mathcal{A}_T^{-1}\mathcal{C}_T(-\alpha,\Psi_*,I_p)'\Omega^{-1}\alpha(\alpha'\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1}\}.\end{aligned} \quad (33)$$

For $(d,b)\in N(\psi_0,\kappa_0)$ and $i=0,1,\ldots,k$, the processes U_{it} and $\beta_0'U_{-1,t}$, see (22), and their derivatives with respect to (d,b) are stationary because $d+ib-d_0\geq d-d_0\geq-\kappa_0>-1/2$. Only the process $\beta_{0\perp}'X_{-1t}$ is nonstationary when $b_0>1/2$. When normalized by $T^{\delta_{-1}+1/2}$ it will converge to fBM provided $E|\varepsilon_t|^q<\infty$ for some $q>1/(b_0-1/2)$, see (4), so that on $D^{p-r+1}[0,1]$,

$$N_T X_{-1[tu]}^{ext} = \begin{pmatrix} T^{\delta_{-1}+1/2}\beta_{\perp 0}'X_{-1[Tu]} \\ T^{d-b}\Delta^{d-b}L_b\pi_{[Tu]}(1)\end{pmatrix} \Longrightarrow \begin{pmatrix}\beta_{0\perp}'C_0W_{d_0-d+b-1}(u) \\ u^{-(d-b)}/\Gamma(1-d+b)\end{pmatrix} = F_\psi(u). \quad (34)$$

We show that the initial values can be neglected asymptotically and that the stationary processes $\{\beta_0'U_{-1t},U_{jt}\}_{j=-1}^k$ can replace the regressors $\{\beta_0'X_{-1t},X_{jt}\}_{j=-1}^k$. This means that the limit of $\mathcal{B}_T$ can be calculated as

$$\mathcal{B} = Var(U_{-1t}'\beta_0, U_{0t}',\ldots,U_{kt}')'.$$

For $b_0<1/2$, all regressors X_{it} are stationary in a neighborhood of the true value and we only re-scale $\rho-\rho_0=T^{d-b}\theta_\rho$, but define $\theta_\beta=\tilde\theta_\beta$. The various quantities $\mathcal{A}_T,\mathcal{B}_T,\mathcal{C}_T$, and $\mathcal{C}_{\varepsilon T}$ are defined as above, but their asymptotic properties are now different. The estimator of θ and profile likelihood function are given by (32) and (33).

The next theorem summarizes the asymptotic results for the product moments and their derivatives with respect to ψ, denoted D^m, when $\psi\in N(\kappa_0,\kappa_0)$.

Theorem 6 *Let Assumptions 1-4 be satisfied and let $N(\psi_0,\kappa_0)=\{|\psi-\psi_0|\leq\kappa_0\}\subset\mathcal{N}$ for κ_0 so small that $q>(b-d+d_0-1/2)^{-1}$ for $\psi\in N(\psi_0,\kappa_0)$.*

(i) If $1/2<b_0<d_0$ and $|\varepsilon_t|^q<\infty$ for some $q>(b_0-1/2)^{-1}$, then for $m\geq0$, it holds that $\mathsf{D}^m\mathcal{A}_T$, $\mathsf{D}^m\mathcal{B}_T$, and $\mathsf{D}^m\mathcal{C}_T$ are tight on $N(\psi_0,\kappa_0)$ and

$$(\mathcal{A}_T,\mathsf{D}^m\mathcal{B}_T,\mathsf{D}^m\mathcal{C}_T) \Longrightarrow (\int_0^1 F_\psi(u)F_\psi(u)'du, \mathsf{D}^m\mathcal{B}, 0) \quad (35)$$

as continuous processes on $N(\psi_0,\kappa_0)$. At the true value $\psi_0=(d_0,b_0)$ we find, with $F_0=F_{\psi_0}$,

$$\mathcal{C}_{\varepsilon T}^0 \xrightarrow{d} \int_0^1 F_0(dW)'. \quad (36)$$

The same results hold for the model $\mathcal{H}_r(d=b)$.

For $k = r = 0$, $\mathsf{D}^m\mathcal{B}_T = \mathsf{D}^m T^{-1}\sum_{t=1}^T \Delta^d X_t \Delta^d X_t' \Longrightarrow \mathsf{D}^m\mathcal{B}$.[8]

(ii) If $0 < b_0 < 1/2, b_0 < d_0$ $(\mathsf{D}^m\mathcal{A}_T, \mathsf{D}^m\mathcal{B}_T, \mathsf{D}^m\mathcal{C}_T)$ *is tight on* $N(\psi_0, \kappa_0)$ *and converge towards the deterministic mean, which we denote* $(\mathsf{D}^m\mathcal{A}, \mathsf{D}^m\mathcal{B}, \mathsf{D}^m\mathcal{C})$. *Finally*

$$\mathcal{C}^0_{\varepsilon T} \xrightarrow{d} N_{(p-r+1)\times p}(0, \mathcal{A}\otimes\Omega_0). \tag{37}$$

(iii) For the model with $d = b$ *the results (35) and (36) hold if* $b_0 > 1/2$ *and (37) hold if* $b_0 < 1/2$.

Proof. *Proof of (i)*: For $1/2 < b_0 < d_0$ it follows from Theorem 2, see also (16), that $\Delta^{d+ib}X_t$ has the representation

$$\Delta^{d+ib}X_t = \Delta_+^{d+ib}X_t + \Delta_-^{d+ib}X_t = \Delta_+^{d+ib-d_0}(C_0\varepsilon_t + \Delta_+^{b_0}Y_t^+) + D_{it}(\psi), t = 1,\dots,T, \tag{38}$$

where $D_{it}(\psi) = \Delta_+^{d+ib}\mu_t + \Delta_+^{d+ib}\xi_t + \Delta_-^{d+ib}X_t$ is the deterministic part generated by initial values and the constant term $\alpha_0\rho_0'\pi_t(1)$.

It follows from Lemma A.10 that $\mathsf{D}^m D_{it}(\psi)$, suitably normalized, is uniformly small so that asymptotically we can replace the regressors $X_{it},\ i \geq 0$ by the stationary variables U_{it}, see (22). For $i = -1$ we have the regressor

$$X_{-1t}^{cent} = (\Delta^{d-b} - \Delta^{d+kb})\beta_0' X_t + \Delta^{d-b}L_b\rho_0'\pi_t(1) = \beta_0' X_{-1t} + O(t^{-(d-b)}), \tag{39}$$

and we can replace the first term by $\beta_0' U_{-1t}$ and the other is uniformly small for $(d, b) \in N(\psi_0, \kappa_0)$ when $d_0 > b_0$.

The nonstationary regressor $(\Delta^{d-b} - \Delta^{d+kb})\beta_{0\perp}' X_t$ is normalized by $T^{\delta_{-1}+1/2}$ and it follows from (104) that $T^{\delta_{-1}+1/2}\beta_{0\perp}' D_{-1t}(\psi)$ goes uniformly to zero. Thus we can replace this regressor by $(\Delta_+^{d-b} - \Delta_+^{d+kb})\beta_{0\perp}' U_t$ in the evaluation of the limit of $\mathsf{D}^m\mathcal{A}_T, \mathsf{D}^m\mathcal{B}_T$, and $\mathsf{D}^m\mathcal{C}_t$.

By Theorem 2, $U_t = C_0\varepsilon_t + \Delta^{b_0}Y_t \in \mathcal{Z}$, where the class $\mathcal{Z}$ is given in Definition 1. Lemma A.8, therefore applies directly to product moments of $\Delta_+^{d+ib-d_0}U_t$, using the stationary processes $\beta_0' U_{jt}, j \geq -1$ with indices $u = d+jb-d_0+b_0 \geq -1/2+(1/2-2\kappa_0)$ and $\beta_{0\perp}' U_{it}', i \geq 0$ with index $u = d + ib - d_0 \geq d - d_0 \geq -1/2 + (1/2 - \kappa_0)$ so $\kappa_u = 1/2 - 2\kappa_0$, and the nonstationary process $\beta_{0\perp}' U_{-1t}$ with index $w = d-b-d_0 \leq -b_0+2\kappa_0 \leq -1/2-(b_0-1/2-2\kappa_0)$ so we choose $\kappa_w = b_0 - 1/2 - 2\kappa_0$ for small κ_0. We also take κ_0 so small that $q > (b_0 - 1/2)^{-1}$ implies that $q > 1/\kappa_w$. Tightness of $\mathsf{D}^m(\mathcal{A}_T, \mathcal{B}_T, \mathcal{C}_T)$ and the convergence in (35) then follow from (83), (82), and (84). Note that when $b_0 > 1/2$ there is no critical process in the neighborhood $N(\psi_0, \kappa_0)$. The proof of *(36)* follows from (5). [9]

Proof of (ii): For $b_0 < 1/2 < d_0$ we apply (16) and the only difference in the above proof is that $(\Delta^{d-b} - \Delta^{d+kb})\beta_{0\perp}' X_t$ is asymptotically stationary and can be replaced by $\beta_{0\perp}' U_{-1t}$. The limits of $\mathsf{D}^m(\mathcal{A}_T, \mathcal{B}_T, \mathcal{C}_T)$ then follow from Lemma A.8, (83).

We find

$$\mathcal{A} = \begin{pmatrix} E(\beta_{0\perp}' U_{-1t}U_{-1t}'\beta_{0\perp}) & 0 \\ 0 & \int_0^1 u^{-2(d_0-b_0)}du/\Gamma(1-d_0+b_0)^2 \end{pmatrix}.$$

[8] SJ: hvad skal vi med den?

[9] SJ: Her har jeg fjernet et bevis og henvist til *(5)*
Vi mangler produkt moment af $t^{-(d-b)}$ og $\Delta^u Z_t$ så vi kan få hele $\mathcal{C}$

Finally $\beta_{0\perp}' U_{-1t}\varepsilon_t'$ is a martingale difference sequence and the central limit theorem for martingales gives (37), see Hall and Heyde (1980, chp. 3).

If instead $b_0 < d_0 < 1/2$ we apply the representation, see (12) and find

$$\Delta^{d+ib} X_t = \Delta^{d+ib-d_0}(C_0\varepsilon_t + \Delta^{b_0} Y_t) + C_{10}\alpha_0\rho_0'\pi_t(1) + \xi_{0t}^*, t = 1, \ldots, T.$$

In this case the initial values play no role and the argument is as above.

Proof of (iii): The same proof as above works. ■

We next want to discuss the asymptotic variance of the stationary components and define for $b_0 > 1/2$ the parameter $\phi = (d, b, \alpha, \Psi_*)$ and the residual $\varepsilon_t(\phi) = \varepsilon_t(\phi, 0) = (-\alpha, \Psi_*, I_p)V_t$, c.f. (28). For (d, b) close to (d_0, b_0) we define the corresponding stationary process

$$e_t(\phi) = U_{kt} - \alpha W_t + \sum_{i=0}^{k-1} \Psi_i U_{it} = (-\alpha, \Psi_*, I_p)(\beta_0' U_{-1t}, U_{*t}', U_{kt}')' \tag{40}$$

where U_{it} is given in (22). In the following we use D_ϕ and $\mathsf{D}^2_{\phi\phi}$ to denote first- and second-order derivatives with respect to ϕ.

Lemma 7 *We find for $\phi = \phi_0$ that $e_t(\phi_0) = \varepsilon_t(\phi_0) = \varepsilon_t$, and furthermore*

(i) for $b_0 > 1/2$

$$T^{-1}\sum_{t=1}^{T} \varepsilon_t(\phi)\varepsilon_t(\phi)' \overset{P}{\to} Ee_t(\phi)e_t(\phi)' = (-\alpha, \Psi_*, I_p)\mathcal{B}(-\alpha, \Psi_*, I_p)', \tag{41}$$

$$\mathsf{D}_\phi E e_t(\phi_0)'\Omega_0^{-1} e_t(\phi_0) = E(\mathsf{D}_\phi e_t(\phi_0)'\Omega_0^{-1}\varepsilon_t) + E(\varepsilon_t'\Omega_0^{-1}\mathsf{D}_\phi e_t(\phi_0)) = 0, \tag{42}$$

$$\mathsf{D}^2_{\phi\phi} E e_t(\phi_0)'\Omega_0^{-1} e_t(\phi_0) = E(\mathsf{D}_\phi e_t(\phi_0)'\Omega_0^{-1}\mathsf{D}_\phi e_t(\phi_0)) = \Sigma_0, \tag{43}$$

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

(ii) for $b_0 < 1/2$ we find

$$T^{-1}\sum_{t=1}^{T} \varepsilon_t(\phi)\varepsilon_t(\phi)' \overset{P}{\to} \tag{44}$$

$$(-\alpha\theta_\beta', -\alpha\theta_\rho', -\alpha, \Psi_*, I_p)\begin{pmatrix} \mathcal{A} & \mathcal{C} \\ \mathcal{C}' & \mathcal{B} \end{pmatrix}(-\alpha\theta_\beta', -\alpha\theta_\rho', -\alpha, \Psi_*, I_p)'$$

and that (42) and (43) hold with suitably defined $e_t(\phi)$ with $\phi = (\theta_\beta, \theta_\rho, d, b, \alpha, \Psi_*)$ and we define

$$\Sigma_0^{stat} = \mathsf{D}^2_{\phi\phi} E e_t(\phi_0)'\Omega_0^{-1} e_t(\phi_0). \tag{45}$$

(iii) For $d = b$ the results (41), (42), and (43) hold if $b_0 > 1/2$ and (44), (42), and (43) hold if $b_0 < 1/2$.

Proof. *Proof of (i):* 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

$$f_\phi(z) = (1-z)^{d-b-d_0+b_0}\Psi(1-(1-z)^b)|_{\beta=\beta_0}\Psi_0(y)^{-1}. \tag{46}$$

For $\phi = \phi_0$ we find $f_{\phi_0}(z) = 1$ so that $e_t(\phi_0) = \varepsilon_t$. The result (41) follows from (35) of Theorem 6. Differentiating the left hand side of (41), we find the limit

$$E(\mathsf{D}_\phi e_t(\phi_0)'\Omega_0^{-1}e_t(\phi_0)) = 2E(\mathsf{D}_\phi e_t(\phi_0)'\Omega_0^{-1}\varepsilon_t) = 0,$$

because $\mathsf{D}_\phi e_t(\phi_0)$ is measurable with respect to $\varepsilon_1, \ldots, \varepsilon_{t-1}$. Therefore

$$\mathsf{D}_\phi E(e_t(\phi_0)e_t(\phi_0)') = E(\mathsf{D}_\phi e_t(\phi_0)e_t(\phi_0)') = 0$$

which proves (42). Differentiating twice we find (43) the same way.

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 $\mathsf{D}_\phi e_t(\phi_0)$ 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 $\mathsf{D}_s f_{\phi_0+sh}(z)|_{s=0} = 0$. We find from (3) and (46) the derivatives, where we use $y = 1 - (1-z)^{b_0}$,

$$\begin{aligned}
\mathsf{D}_d f_{\phi_0}(z) &= \log(1-z)I_p = b_0^{-1}\log(1-y)I_p, \\
\mathsf{D}_b f_{\phi_0}(z) &= -b_0^{-1}\log(1-y)(I_p + \mathsf{D}_y\Psi_0(y)(1-y)\Psi_0(y)^{-1}), \\
\mathsf{D}_{\Psi_i} f_{\phi_0}(z) &= (1-y)^i, i = 0, \ldots, k-1, \\
\mathsf{D}_\alpha f_{\phi_0}(z) &= -\beta_0' y.
\end{aligned}$$

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

$$b_0^{-1}\log(1-y)\{(d_1 - b_1)\Psi_0(y) - b_1\mathsf{D}_y\Psi_0(y)(1-y)\} - \{A\beta_0' y\Psi_0(y) + \sum_{i=0}^{k-1} G_i(1-y)^i\Psi_0(y)\}.$$

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 $G_i = 0, i = 0, \ldots, k-1$. We then find that the coefficient to $b_0^{-1}\log(1-y)$ should be zero, so that

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

For $y = 1$ we find from (3) 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 $b_1\mathsf{D}_y\Psi_0(y)$ 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 (3) $\Psi_{0k} \neq 0$ is the same as $\Gamma_{0k} \neq 0$.

Proof of (ii) and (iii): The same proof can be used as for *(i)* by a change of notation. ■

4.2 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 5) and first and second derivatives are tight on $N(\psi_0, \kappa_0)$ (by Theorem 6).

Lemma 8 *Let Assumptions 1-4 be satisfied.*

(i) If $b_0 > 1/2$ 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

$$\begin{pmatrix} T^{-1/2}\mathsf{D}_\phi \log L_T(\lambda_0) \\ T^{-1/2}\mathsf{D}_\theta \log L_T(\lambda_0) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} N_{n_\phi}(0, \Sigma_0) \\ (\operatorname{vec} \int_0^1 F_0(dG_0)')' \end{pmatrix}, \tag{47}$$

where Σ_0 is given in (43), $n_\phi = 1+1+pr+kp^2$ is the number of parameters in $\phi = (d, b, \alpha, \Psi_*)$, $F_0 = (W_{b_0-1}(u)'C_0'\beta_{0\perp}, u^{-(d_0-b_0)}/\Gamma(1-(d_0-b_0))'$, and $G_0 = \alpha_0'\Omega_0^{-1}W$.

(ii) If $0 < b_0 < 1/2$ then the score with respect to all parameters is asymptotically Gaussian, $N_{n_\phi+(p-r)r+r}(0, \Sigma_0^{stat})$, see (45).

Proof. *Proof of (i)*: For $b_0 > 1/2$ the score function for $\phi = (d, b, \alpha, \Psi_*)$ evaluated at the true value is

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

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

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

see Lemma 7. The result for the first block of (47) now follows from the central limit theorem for martingales, see Hall and Heyde (1980, chp. 3).

The score function for θ_β and θ_ρ evaluated at the true value is

$$\begin{aligned} T^{-1/2}\mathsf{D}_{\theta_\beta} \log L_T(\lambda_0) &= T^{-1/2}\sum_{t=1}^{T} \varepsilon_t'\Omega_0^{-1}(\alpha_0 \otimes T^{-(b_0-1/2)}\beta_{0\perp}'L_{b_0}\Delta^{d_0-b_0}X_t) \\ &\xrightarrow{d} (\operatorname{vec}\int_0^1 (\beta_{0\perp}'C_0W_{b_0-1}(dG_0)')', \\ T^{-1/2}\mathsf{D}_{\theta_\rho} \log L_T(\lambda_0) &= T^{-1/2}\sum_{t=1}^{T} (\operatorname{vec}(T^{d_0-b_0}L_{b_0}\pi_t(1-d_0+b_0)\varepsilon_t'\Omega_0^{-1}\alpha_0))' \\ &\xrightarrow{d} (\operatorname{vec}\int_0^1 \frac{u^{-d_0+b_0}}{\Gamma(1-d_0+b_0)}(dG_0)')' \end{aligned}$$

see (36) of Theorem 6, which proves the second block of (47).

Proof of (ii): If $0 < b_0 < 1/2$, all stochastic regressors are asymptotically stationary and we take $\beta = \beta_0 + \beta_{0\perp}\theta_\beta$ and the score with respect to θ_β is

$$T^{-1/2}\mathsf{D}_{\theta_\beta} \log L_T(\lambda_0) = T^{-1/2}\sum_{t=1}^{T} (\operatorname{vec}(\beta_{0\perp}'\Delta^{d_0-b_0}L_{b_0}X_t\varepsilon_t'\Omega_0^{-1}\alpha_0))',$$

which is a martingale difference sequence as is the score for θ_ρ given above. The central limit theorem for martingales gives the result. ∎

Lemma 9 *Let Assumptions 1-4 be satisfied.*

(i) If $b_0 > 1/2$ 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

$$\begin{pmatrix} \Sigma_0 & 0 \\ 0 & \alpha_0'\Omega_0^{-1}\alpha_0 \otimes \int_0^1 F_0 F_0' du \end{pmatrix} > 0, \tag{48}$$

where Σ_0 is given in (43) and $F_0 = (W_{b_0-1}(u)'C_0'\beta_{0\perp}, u^{-(d_0-b_0)}/\Gamma(1-(d_0-b_0))'$.

(ii) If $0 < b_0 < 1/2$ the information for all parameters is convergent in probability to a non-stochastic limit given in (45).

Proof. *Proof of (i):* The information matrices for the different parameters can be found from (31). From (35) of Theorem 6 it holds that $\mathsf{D}^m\mathcal{C}_T^0 \xrightarrow{P} 0$. Using this and (43) we find for $\theta_0 = 0$ that

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

Proof of (ii): If $0 < b_0 < 1/2$ we find the information for $\beta = \beta_0 + \beta_{0\perp}\theta_\beta$ and $\rho = \rho_0 + T^{-d_0+b_0}\theta_\rho$ to be

$$\begin{aligned} -T^{-1}\mathsf{D}^2_{\theta_\beta\theta_\beta}\log L_T(\lambda_0) &= \alpha_0'\Omega_0^{-1}\alpha_0 \otimes T^{-1}\sum_{t=1}^T (\Delta^{d_0-b_0}L_{b_0}\beta_{0\perp}'X_t)(\Delta^{d_0-b_0}L_{b_0}\beta_{0\perp}'X_t)', \\ -T^{-1}\mathsf{D}^2_{\theta_\rho\theta_\rho}\log L_T(\lambda_0) &= \alpha_0'\Omega_0^{-1}\alpha_0 \otimes T^{-1}\sum_{t=1}^T [T^{d_0-b_0}L_{b_0}\pi_t(1-d_0+b_0)]^2, \\ -T^{-1}\mathsf{D}^2_{\theta_\beta\theta_\rho}\log L_T(\lambda_0) &= \alpha_0'\Omega_0^{-1}\alpha_0 \otimes T^{-1}\sum_{t=1}^T \Delta^{d_0-b_0}L_{b_0}\beta_{0\perp}'X_t T^{d_0-b_0}L_{b_0}\pi_t(1-d_0+b_0), \end{aligned}$$

which converges to a non-stochastic limit by the law of large numbers because $\Delta^{d_0-b_0}L_{b_0}X_t$ is stationary when $b_0 < 1/2$. ■

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 10 *Let the assumptions of Theorems 4 and 5 be satisfied and suppose $(d_0, b_0) \in int(\mathcal{N})$.*

(i) If $b_0 > 1/2$ and $E|\varepsilon_t|^q < \infty$ for some $q > (b_0 - 1/2)^{-1}$, the asymptotic distribution of the Gaussian maximum likelihood estimators $\hat\phi = (\hat d, \hat b, \hat\alpha, \hat\Psi_*)$, $\hat\beta$, and $\hat\rho$ for model (1) is given by

$$\begin{pmatrix} T^{1/2}\,\mathrm{vec}(\hat\phi - \phi_0) \\ \begin{pmatrix} T^{b_0}\bar\beta_{0\perp}'(\hat\beta - \beta_0) \\ T^{1/2-d_0+b_0}(\hat\rho - \rho_0) \end{pmatrix} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} N_{n_\phi}(0, \Sigma_0^{-1}) \\ (\int_0^1 F_0 F_0' du)^{-1}\int_0^1 F_0 (dG_0)'(\alpha_0'\Omega_0^{-1}\alpha_0)^{-1} \end{pmatrix}, \tag{49}$$

where $F_0 = (W_{b_0-1}(u)'C_0'\beta_{0\perp}, u^{-(d_0-b_0)}/\Gamma(1-(d_0-b_0))'$ *and* $G_0 = \alpha_0'\Omega_0^{-1}W$ *are independent. It follows that the asymptotic distribution of* $\mathrm{vec}(T^{b_0}\bar{\beta}_{0\perp}'(\hat{\beta}-\beta_0), T^{1/2-d_0+b_0}(\hat{\rho}-\rho_0))$ *is mixed Gaussian with conditional variance given by*

$$(\alpha_0'\Omega_0^{-1}\alpha_0)^{-1} \otimes (\int_0^1 F_0F_0'du)^{-1}. \tag{50}$$

In the model $\mathcal{H}_r(d=b)$ *the same results hold with the relevant restriction imposed.*

(ii) If $0 < b_0 < 1/2$ *the estimators are asymptotically Gaussian with variance* Σ_0^{stat}*.*

Proof. *Proof of (i):* For $b_0 > 1/2$ we find limit distributions of $T^{1/2}(\hat{\phi}-\phi_0)$ and $T^{1/2}\hat{\theta}$, by applying the usual expansion of the score function around $\phi = \phi_0$, $\theta = 0$, and $\Omega = \hat{\Omega}$. Using Taylor's formula with remainder term we find for $l_T = T^{-1}\log L_T$ that

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

Here the asterisks indicate intermediate points between $(\hat{\phi}, \hat{\theta}, \hat{\Omega})$ and $(\phi_0, 0, \hat{\Omega})$, one for each score function, which therefore converge to $(\phi_0, 0, \Omega_0)$ in probability by Theorem 5.

Because the first and second derivatives are tight, see Theorem 6 and Lemma A.2, and $\lambda^* \xrightarrow{P} \lambda_0$, see Theorem 5, 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 8 and the limit of the information per observation in Lemma 9, see (48). Pre-multiplying by its inverse we find (49). The stochastic component of the process F_0 is a function of $\alpha_{0\perp}'W$, see (10) and (36), whereas $G_0 = \alpha_0'\Omega_0^{-1}W$, so that F_0 and G_0 are independent and the limit distribution of $T^{b_0}\bar{\beta}_{\perp 0}'(\hat{\beta}-\beta_0)$ and $T^{1/2-d_0+b_0}(\hat{\rho}-\rho_0)$ is mixed Gaussian.

Proof of (ii): If $0 < b_0 < 1/2$ the result follows from the results about score and information by the same type of proof. ■

In the model $\mathcal{H}_r(d=b)$, the same expansions can be made and similar results derived. The results in Theorem 10 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) when $1/2 < b_0$. 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 the 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.[10]

[10]SJ: Jeg mener vi gør så lidt som muligt ved modellen med $\rho = 0$, specielt her hvor der blot står at hvis en parameter estimator er asymptotisk Gaussian kan man teste parameteren ved et χ^2 test

5 Likelihood ratio test for cofractional rank

We consider the model

$$\mathcal{H}_p : \Delta^d X_t = \Pi \Delta^{d-b} L_b X_t + \mu \Delta^{d-b} L_b \pi_t(1) + \sum_{i=0}^{k} \Gamma_i \Delta^d L_b^i X_t + \varepsilon_t \tag{51}$$

and want to test the hypothesis $\mathcal{H}_r : rank(\Pi, \mu) = r$, against the alternative $\mathcal{H}_p : rank(\Pi, \mu) = p$. 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 $\mathcal{H}_r, r = 0, 1, \ldots, p$. The likelihood ratio (LR) statistic is

$$-2 \log LR(\mathcal{H}_r | \mathcal{H}_p) = \ell_{T,p}(\hat{\psi}_p) - \ell_{T,r}(\hat{\psi}_r). \tag{52}$$

Theorem 11 *Under the assumptions of Theorem 10 and $1/2 < b_0$, the likelihood ratio statistic for $(\Pi, \mu) = \alpha(\beta', \rho')$ that is $\mathcal{H}_r$ in $\mathcal{H}_p$ has asymptotic distributions*

$$-2 \log LR(\mathcal{H}_r | \mathcal{H}_p) \xrightarrow{d} tr\{\int_0^1 (dB) B_{b_0-1}^{ext\prime} (\int_0^1 B_{b_0-1}^{ext} B_{b_0-1}^{ext\prime} du)^{-1} \int_0^1 B_{b_0-1}^{ext} (dB)'\}, \tag{53}$$

where B is $(p-r)$-dimensional standard BM, B_{b_0-1} is the corresponding fBM and $B_{b_0-1}^{ext} = (B_{b_0-1}'(u), u^{-(d_0-b_0)})'$. The limit distributions are continuous in (d_0, b_0).

If $0 < b_0 < 1/2$ then

$$-2 \log LR(\mathcal{H}_r | \mathcal{H}_p) \xrightarrow{d} \chi^2((p-r)^2). \tag{54}$$

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 $rank(\Pi) > r$, and assume that Assumption 1 is satisfied under the alternative, then

$$-2 \log LR(\mathcal{H}_r | \mathcal{H}_p) \xrightarrow{P} \infty. \tag{55}$$

In the model $d = b$ the same results hold, and in the model with $\rho = 0$ the results holds with $B_{b_0-1}^{ext}$ replaced with B_{b_0-1}.

If $k = 0$ and $r = 0$ then (53) holds with b_0 replaced by d_0.

Proof. We give the proofs only for the most general model, the other proofs being the same but with different notation.

Proof of (53): We assume that $rank(\Pi, \mu) = r$, and that $\Pi_0 = \alpha_0\beta_0', \mu_0 = \alpha_0\rho_0'$ where α_0 and β_0 are $p \times r$ of rank r. It is convenient to introduce the extra hypothesis that $\Pi = \alpha\beta'$ and $\beta = \beta_0, \rho = \rho_0$, or $(\Pi, \mu) = \alpha(\beta_0', \rho_0')$, see Lawley (1956), and Johansen (2002) for an application to the cointegrated VAR model.

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

$$\frac{\max_{(\Pi,\mu)=\alpha(\beta',\rho')} L}{\max L} = \frac{\max_{(\Pi,\mu)=\alpha(\beta_0',\rho_0')} L}{\max L} / \frac{\max_{(\Pi,\mu)=\alpha(\beta_0',\rho_0')} L}{\max_{\Pi=\alpha\beta'} L} = \frac{LR(\mathcal{H}_r \text{ and } \beta = \beta_0, \rho = \rho_0 | \mathcal{H}_p)}{LR(\beta = \beta_0, \rho = \rho_0 | \mathcal{H}_r)}$$

The statistic $LR(\mathcal{H}_r$ and $\beta = \beta_0, \rho = \rho_0 | \mathcal{H}_p)$ is the test that $(\Pi, \mu) = \alpha(\beta_0', \rho_0')$ (with rank r) against (Π, μ) unrestricted, and $LR(\beta = \beta_0, \rho = \rho_0 | \mathcal{H}_r)$ is the test that $(\beta', \rho') = (\beta_0', \rho_0')$ in the model with $(\Pi, \mu) = \alpha(\beta', \rho')$ and $rank(\Pi) = r$. We next find a first order approximation to each statistic and subtracting them and passing to the limit 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

$$-2\log LR(\xi=\xi_0) = \mathsf{D}_\xi \log L_T(\xi_0,\eta_0)(\mathsf{D}^2_{\xi\xi}\log L_T(\xi_0,\eta_0))^{-1}\mathsf{D}_\xi \log L_T(\xi_0,\eta_0)' + o_P(1), \quad (56)$$

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

A first order approximation to $-2\log LR(\beta=\beta_0, \rho=\rho_0|\mathcal{H}_r)$: It follows from Lemma 9 that, for $\xi = \theta = (\theta_\beta, \theta_\rho),\ \eta = (d, b, \alpha, \Psi_*, \Omega)$, the asymptotic information per observation is block diagonal at the true value, and Theorem 6 and Lemma A.2 show that the information is tight as a process in the parameters. Thus we have that $-2\log LR(\beta=\beta_0, \rho=\rho_0|\mathcal{H}_r)$ is

$$\begin{aligned} & (\operatorname{vec}\mathcal{C}^0_{\varepsilon T}\Omega_0^{-1}\alpha_0)'(\alpha_0'\Omega_0^{-1}\alpha_0 \otimes \mathcal{A}^0_T)^{-1}\operatorname{vec}\mathcal{C}^0_{\varepsilon T}\Omega_0^{-1}\alpha_0 + o_P(1) \\ = \;& tr\{(\alpha_0'\Omega_0^{-1}\alpha_0)^{-1}\alpha_0'\Omega_0^{-1}\mathcal{C}^{0\prime}_{\varepsilon T}\mathcal{A}^{0-1}_T\mathcal{C}^0_{\varepsilon T}\Omega_0^{-1}\alpha_0\} + o_P(1), \end{aligned} \quad (57)$$

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

A first order approximation to $-2\log LR(\mathcal{H}_r$ and $\beta=\beta_0, \rho=\rho_0|\mathcal{H}_p)$: In model (51) we introduce a convenient reparametrization by $\alpha = \Pi\bar{\beta}_0, \xi_\beta = T^{-\delta_{-1}-1/2}\Pi\bar{\beta}_{0\perp}, \xi_\rho = T^{-(d-b)}(\mu - \alpha\rho_0')$, so that $\Pi = \alpha\beta_0' + T^{\delta_{-1}+1/2}\xi'\beta_{0\perp}'$ and $\mu = \alpha\rho_0' + T^{d-b}\xi_\rho$. The equations are, see (27), with $N_T = (T^{d-b-d_0+1/2}I_{p-r}, T^{-(d-b)})$

$$X_{kt} = \alpha X^{cent}_{-1t} + \xi' N_T X^{ext}_{-1t} + \sum_{i=1}^k \Psi_i X_{it} + \varepsilon_t.$$

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

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

where $\eta = (d, b, \alpha, \Psi_*, \Omega)$. This expression is the same as the conditional likelihood (31) 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

$$-2\log LR(\mathcal{H}_r \text{ and } \beta=\beta_0, \rho=\rho_0|\mathcal{H}_p) = tr\{\Omega_0^{-1}\mathcal{C}^{0\prime}_{\varepsilon T}\mathcal{A}^{0-1}_T\mathcal{C}^0_{\varepsilon T}\} + o_P(1). \quad (58)$$

A first order approximation to $-2\log LR(\mathcal{H}_r|\mathcal{H}_p)$: Subtracting (57) from (58) 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

$$\begin{aligned} & tr\{\alpha_{0\perp}(\alpha_{0\perp}'\Omega_0\alpha_{0\perp})^{-1}\alpha_{0\perp}'\mathcal{C}^{0\prime}_{\varepsilon T}\mathcal{A}^{0-1}_T\mathcal{C}^0_{\varepsilon T}\} \\ & \xrightarrow{d} tr\{\alpha_{0\perp}(\alpha_{0\perp}'\Omega_0\alpha_{0\perp})^{-1}\alpha_{0\perp}'\int_0^1 (dW)F_0'(\int_0^1 F_0F_0'du)^{-1}\int_0^1 F_0(dW)'\} = DF(\psi_0), \end{aligned} \quad (59)$$

say, which is the desired result if we define $B = (\alpha'_{0\perp}\Omega_0\alpha_{0\perp})^{-1/2}\alpha'_{0\perp}W$.

The continuity of the limit distribution can seen by noticing that the matrices $\int_0^1 (dB)F'_\psi$ and $\int_0^1 F_\psi F'_\psi du$, and hence also $DF(\psi)$, are continuous in $\mathcal{L}_2$ as functions of ψ and that is enough for convergence in distribution so that if $\psi_n \to \psi$ then $DF(\psi_n) \xrightarrow{d} DF(\psi)$.

Proof of (54): In this case the result follows from the usual expansion of the LR test statistic and the asymptotic distribution in Theorem 10.

Proof of (55): We want to analyze the alternative that $\Pi = \alpha\beta' + \alpha_1\beta'_1 = (\alpha, \alpha_1)(\beta, \beta_1)'$, where $rank(\Pi) > r$, and apply the same methods as in the proof of (53). Under the alternative there are more parameters and therefore the information matrix is larger, but still asymptotically block diagonal. The information for the parameters $(d, b, \alpha, \beta, \rho, \Psi_*, \Omega)$ is therefore also asymptotically block diagonal so that (56) 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

$$\varepsilon_t^{alt}(\phi_0) = \varepsilon_t + \alpha_1(\zeta'_0, 0)X_{-1t}^{ext} = \varepsilon_t + \alpha_1\zeta'_0\beta'_{0\perp}X_{-1t}, \tag{60}$$

where $\beta'_1X_{-1t} = \zeta'_0\beta'_{0\perp}X_{-1t} = \zeta'_0\beta'_{0\perp}(\Delta^{d_0-b_0} - \Delta^{d_0+kb_0})X_t$ is an asymptotically stationary $\mathcal{F}(0)$ process $\zeta'_0\beta'_{0\perp}(1 - \Delta^{(1+k)b_0})Y_t^{alt} = \zeta'_0\beta'_{0\perp}U_{-1t}^{alt}$. To analyze the approximation (59) we define $\mathcal{C}_{\varepsilon T}^{alt} = T^{-b_0}\sum_{t=1}^T \beta'_{0\perp}X_{-1t}\varepsilon_t^{alt}(\phi_0)'$ and $\mathcal{A}_T^{alt} = T^{-2b_0}\sum_{t=1}^T \beta'_{0\perp}X_{-1t}X'_{-1t}\beta_{0\perp}$ and find the inequality

$$\mathcal{C}_{\varepsilon T}^{alt\prime}\mathcal{A}_T^{0-1}\mathcal{C}_{\varepsilon T}^{alt} \geq (\boldsymbol{\zeta}'_0\mathcal{C}_{\varepsilon T}^{alt})'(\boldsymbol{\zeta}'_0\mathcal{A}_T^{alt}\boldsymbol{\zeta}_0)^{-1}(\boldsymbol{\zeta}'_0\mathcal{C}_{\varepsilon T}^{alt}) \tag{61}$$

and want to show that the right hand side tends to infinity in probability. We find from (83) of Lemma A.8 and (60) that

$$T^{b_0-1}\zeta'_0\mathcal{C}_{\varepsilon T}^{alt} = T^{-1}\sum_{t=1}^T \zeta'_0\beta'_{0\perp}X_{-1t}\varepsilon'_t + T^{-1}\sum_{t=1}^T \zeta'_0\beta'_{0\perp}X_{-1t}X'_{-1t}\beta_{0\perp}\zeta_0\alpha'_1,$$

which converges in probability to $E(\zeta'_0\beta'_{0\perp}U_{-1t}^{alt}U_{-1t}^{alt\prime}\beta_{0\perp}\zeta_0\alpha'_1) = Var(\zeta'_0\beta'_{0\perp}U_{-1t}^{alt})\alpha'_1$. We also find that

$$T^{2b_0-1}\zeta'_0\mathcal{A}_T^{alt}\zeta_0 = T^{-1}\sum_{t=1}^T \zeta'_0\beta'_{0\perp}X_{-1t}X'_{-1t}\beta_{0\perp}\zeta_0 \xrightarrow{P} E(\zeta'_0\beta'_{0\perp}U_{-1t}^{alt}U_{-1t}^{alt\prime}\beta_{0\perp}\zeta_0),$$

because under the alternative $\zeta'_0\beta'_{0\perp}U_{-1t}^{alt}$ is an asymptotically stationary $\mathcal{F}(0)$ process. Inserting both these expressions into (61) we see that the right hand side multiplied by T^{-1} converges in probability to the deterministic limit $\alpha_1 Var(\zeta'_0S_{z,-1,t}^0)\alpha'_1 > 0$, which proves (55). ■

The distribution (53) 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, 1991). 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, and the parameter d_0 appears only in the part of the distribution originating from the deterministic term. For given hypothesized (d_0, b_0) or estimated $(\hat{d}_r, \hat{b}_r)$, the distribution (53) can be simulated to obtain critical values on a case-by-case basis. The continuity of the limit distribution $DF(\psi)$ in $\psi = (d, b)$ ensures consistent estimates of $p-$values. Alternatively, numerical CDFs have been simulated as functions of (d, b) by MacKinnon and Nielsen (2010), and their computer programs can be used to obtain critical values or P-values for the test.

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 $r = 0, 1, \ldots$ 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 11 shows that any test of $r < r_0$ will reject with probability one as $T \to \infty$. Thus, $P_{r_0}(\hat{r} < r_0) \to 0$. Since the asymptotic size of the test for rank is δ we also have that $P_{r_0}(\hat{r} = r_0) \to 1 - \delta$ and it follows that $P_{r_0}(\hat{r} > r_0) \to \delta$. This shows that $\hat{r}$ is almost consistent, in the sense that it attains the true value with probability $1 - \delta$ as $T \to \infty$.

6 Conclusion

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

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

to the cointegrated fractional VAR model,

$$\Delta^d X_t = \Delta^{d-b} L_b \alpha(\beta' X_t + \rho') + \sum_{i=1}^{k} \Gamma_i \Delta^d L_b^i X_t + \varepsilon_t, \ 0 < b \le d.$$

We have analyzed the conditional Gaussian likelihood given initial values, which we assumed bounded. Under the assumption that $d_0 - b_0 < 1/2$ (and $b_0 \neq 1/2$) we have shown existence and consistency on compact subsets of the parameter space and derived the asymptotic distribution of the maximum likelihood estimator as well as the asymptotic distribution of the LR test for the rank of $\alpha\beta'$. In the asymptotic analysis we assumed i.i.d. errors with suitable moment conditions. If $1/2 < b_0$ inference on β is asymptotically mixed Gaussian while the estimators of the remaining parameters are asymptotically Gaussian, and the LR test for rank is expressed in terms of fractional Brownian motion B_{b_0-1} extended by $u^{-d_0+b_0}$. If $b_0 < 1/2$ asymptotic distributions are Gaussian and the test for rank is asymptotically χ^2.

The same type of results hold for the models with $d = b$ and $d = d_0$ a prespecified value. For the model with $\rho = 0$ the same result hold except the test for rank involves B_{b_0-1} instead of $B_{b_0-1}^{ext}$.

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*

$$||X_n(s)||_4 \le c, \text{ and } ||X_n(s) - X_n(t)||_4 \le c|s-t| \tag{62}$$

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 $F : \mathbb{R}^k \times \mathbb{R}^p \longmapsto \mathbb{R}^q$ 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 $S_n \overset{P}{\to} s_0 \in [0,1]^m$ and that the $p \times p$ matrix-valued continuous process $X_n(s)$ is tight on $[0,1]^m$. Then $X_n(S_n) - X_n(s_0) \overset{P}{\to} 0$.*

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

A.2 Bounds on product moments

We begin with some bounds on the fractional coefficients.

Lemma A.4 *For $|u| \le u_0$ and all $j \ge 1$ it holds that*

$$|\mathsf{D}^m \pi_j(-u)| \le c(u_0)(1 + \log j)^m j^{-u-1}, \tag{63}$$

$$|\mathsf{D}^m T^u \pi_j(-u)| \le c(u_0) T^u (1 + |\log \frac{j}{T}|)^m j^{-u-1}, \tag{64}$$

uniformly in u. For $|v + 1/2| \le \delta_0$ and all $j \ge 1$ it holds that

$$\pi_j(-v) \ge c(\delta_0) j^{-v-1}, \tag{65}$$

uniformly in v.

Proof. For (63) and (64), see JN (2010, Lemma B.3). For (65), let $u = -v \in [1/2 - \delta_0, 1/2 + \delta_0]$. We apply Stirling's formula,

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

where $\max_{|u-1/2| \le \delta_0} |\epsilon(u,j)| \to 0$ as $j \to \infty$. This proves the result and shows that the constant can be chosen to depend only on δ_0. ■

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

$$\xi_T(\zeta_1, \zeta_2) = \max_{1 \le n,m \le T} \sum_{t=\max(n,m)}^{T} |\zeta_{1,t-n} \zeta_{2,t-m}|, \tag{66}$$

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

Lemma A.5 *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 $Z_{it}^+ = \sum_{n=0}^{t-1} \xi_{in}\varepsilon_{i,t-n}$. Let ζ_{1n}, ζ_{2n} be real coefficients, then*

$$||T^{-1} \sum_{t=1}^{T} (\sum_{n=0}^{t-1} \zeta_{1n} Z_{1,t-n}^+)(\sum_{m=0}^{t-1} \zeta_{2m} Z_{2,t-m}^+)||_4 \leq c\xi_T(\zeta_1, \zeta_2) \tag{67}$$

Proof. *Proof of (??):* We find $\sum_{n=0}^{t-1} \zeta_{1n} Z_{1,t-n}^+ = \sum_{h=0}^{t-1} (\zeta_1 * \xi_1)_h \varepsilon_{1,t-h}$, where $(\zeta_1 * \xi_1)_h = \sum_{n=0}^{h} \zeta_{1,h-n}\xi_{1n}$, and

$$\begin{aligned} \xi_T((\zeta_1 * \xi_1), (\zeta_2 * \xi_2)) &\leq \sum_{h=\max(n_1,n_2)}^{T} \sum_{n=0}^{h-n_1} |\zeta_{1,h-n_1-n}||\xi_{1n}| \sum_{m=0}^{h-n_2} |\zeta_{2,h-n_2-m}||\xi_{2m}| \\ &\leq c\sum_{m=0}^{\infty} |\xi_{2m}| \sum_{m=0}^{\infty} |\xi_{1m}|\xi_T(\zeta_1, \zeta_2) \leq c\xi_T(\zeta_1, \zeta_2) \end{aligned}$$

because $\sum_{n=0}^{\infty} |\xi_{in}| < \infty$. Thus, it is enough to prove (**??**) for $Z_{it}^+ = \varepsilon_{it}$.

We find the expression

$$E(T^{-1} \sum_{t=1}^{T} (\sum_{n_1=0}^{t-1} \zeta_{1n_1} Z_{1,t-n_1}^+)(\sum_{n_2=0}^{t-1} \zeta_{2n_2} Z_{2,t-n_2}^+))^4 = T^{-4} \sum_{(1)} (\prod_{i=1}^{4} \zeta_{1,t_i-n_{1i}}\zeta_{2,t_i-n_{2i}}) E(\prod_{i=1}^{4} \varepsilon_{1,n_{1i}}\varepsilon_{2,n_{2i}}), \tag{68}$$

where $\sum_{(1)}$ is the sum over $1 \leq n_{1i}, n_{2i} \leq t_i \leq T,\ i = 1, \ldots, 4$. We first sum over $\{t_i\}_{i=1}^k$, for fixed (n_{1i}, n_{2i}) and find the bound

$$\xi_T^4(\zeta_1, \zeta_2) T^{-4} \sum_{(2)} E(\prod_{i=1}^{4} \varepsilon_{1,n_{1i}}\varepsilon_{2,n_{2i}}),$$

where $\sum_{(2)}$ is the summation over $1 \leq n_{1i}, n_{2i} \leq T,\ i = 1, \ldots, 4$. The expectation is zero unless for each (l, i) there is a (k, j) for which $n_{li} = n_{kj}$ giving at least four restrictions, which we get for the indices being equal in four pairs. There are eight summations, and with at least four restrictions that gives at most four summations which gives the bound $\xi_T^4(\zeta_1, \zeta_2)$. ■

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

$$|\zeta_{1,0}(a)| \leq 1,\ |\zeta_{1n}(a)| \leq c(1 + \log n)^{m_1} n^{-a-1},\ n \geq 1, \tag{69}$$

$$|\zeta_{1,0}^*(a)| \leq 1,\ |\zeta_{1n}^*(a)| \leq cT^{a+1/2}(1 + |\log \frac{n}{T}|)^{m_1} n^{-a-1},\ n \geq 1, \tag{70}$$

where c does not depend on a or n. These inequalities are satisfied by the fractional coefficients and their derivatives, see Lemma A.4.

Proof of (ii): Proof of (ii):

We repeatedly use the elementary inequalities, for $0 < \kappa < 1$,

$$\sum_{n=1}^{T} n^{-u-1} \leq 1 + \int_1^T x^{-u-1} dx = 1 + u^{-1}(1 - T^{-u}) \leq 1 + \frac{1}{u} \leq 2\kappa^{-1}, \ u \geq \kappa, \quad (71)$$

$$\kappa^{-1}(1 - T^{-\kappa}) \leq u^{-1}(1 - T^{-u}) = \int_1^T x^{-u-1} dx \leq \sum_{n=1}^{T} n^{-u-1}, \ u \leq \kappa. \quad (72)$$

Lemma A.6 *Let* $\zeta_{1n}(a_1), \zeta_{2n}(a_2), \zeta_{1n}^*(a_1),$ *and* $\zeta_{2n}^*(a_2)$ *satisfy (69)–(70), and let* $-1 \leq a_i \leq a_0, i = 1, 2$. *Then:*

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

$$\xi_T(\zeta_1(a_1), \zeta_2(a_2)) \leq c \begin{cases} (1+\log T)^{m_1+m_2+1} T^{-a}, & a \leq 0, \\ a^{-1}, & a > 0. \end{cases} \tag{73}$$

(ii) Uniformly for $\max(a_1, a_2, a_1+a_2+1) \leq -\kappa$ *for some* $\kappa > 0$, *we have*

$$\xi_T(\zeta_1^*(a_1), \zeta_2^*(a_2)) \leq c\kappa^{-1}. \tag{74}$$

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

$$\xi_T(\zeta_1(a_1), \zeta_2^*(a_2)) \leq c(1+\log T)^{m_1+m_2+1} T^{-\min(a,\kappa)} \tag{75}$$

Proof. In evaluating (66) we focus on terms with $t > \max(m, n)$, because the analysis with $t = m$ or $t = n$ is straightforward.

Proof of (73): For $t > \max(m, n)$ we first apply (69) 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 (71) 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 (74): We find that $\xi_T(\zeta_1^*(a_1), \zeta_2^*(a_2))$ is bounded by a constant times times the maximum of

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

$$\to \int_{\max(x,y)}^{1} (1+|\log(s-x)|)^{m_1}(s-x)^{-(a_1+1)}(1+|\log(s-y)|)^{m_2}(s-y)^{-(a_2+1)} ds$$

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

Proof of (75): 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 $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})$ be i.i.d. in two dimensions with $E(\varepsilon_t) = 0$, $Cov(\varepsilon_{it}, \varepsilon_{jt}) = \sigma_{ij}$, and $E|\varepsilon_t|^8 < \infty$, and let $\zeta_{in} = \zeta_{in}(a_i)$ satisfy (69) with $m_1 + m_2 = m$ and $|a_i + 1/2| \leq \kappa$ for some $0 < \kappa < 1/6$. We define $U_{it} = \sum_{n=0}^{N} \zeta_{in}\varepsilon_{it-n}$ and $V_{it} = \sum_{n=N+1}^{t-1} \zeta_{in}\varepsilon_{it-n}$ and the product moments*

$$M_T(a_1, a_2) = T^{-1} \sum_{t=N+1}^{T} (U_{1t}U_{2t} - E(U_{1t}U_{2t})), \text{ and } Q_T(a_1, a_2) = T^{-1} \sum_{t=N+1}^{T} U_{1t}V_{2t}.$$

Then

$$\begin{aligned} ||M_T(a_1, a_2)||_4 &\leq c(1 + \log N)^m T^{-1/4} N^{1/4+2\kappa}, & (76)\\ ||Q_T(a_1, a_2)||_4 &\leq c(1 + \log T)^m T^{-1/4+3\kappa/2} N^{1/4+\kappa/2}. & (77) \end{aligned}$$

If in particular $\zeta_{in}(v_i) = \mathsf{D}^{m_i}\pi_n(-v_i)$ where $|v_i + 1/2| \leq \kappa$, then if $\alpha < (1 - 6\kappa)/(1 + 2\kappa)$

$$(M_T(v_1, v_2), Q_T(v_1, v_2)) \Longrightarrow 0 \text{ on } C([-1/2 - \kappa, -1/2 + \kappa] \times [-1/2 - \kappa, -1/2 + \kappa]). \quad (78)$$

Proof. *Proof of (76)*: Raising the summation to the fourth power we find, see (68)

$$\sum_{(1)} (\prod_{i=1}^{4} \zeta_{1n_{1i}}\zeta_{2n_{2i}}) E \prod_{i=1}^{4} (\varepsilon_{1,t_i-n_{1i}}\varepsilon_{2,t_i-n_{2i}} - \sigma_{12}1_{\{t_i-n_{1i}=t_i-n_{2i}\}}) \quad (79)$$

where $\sum_{(1)}$ is the sum over $1 \leq n_{1i}, n_{2i} \leq N < t_i \leq T$, $i = 1, \ldots, k$. The expectation in (79) is only nonzero if for each (i, l) there is a (j, k) such that $t_i - n_{li} = t_j - n_{kj}$. A special case of this is if $t_i - n_{1i} = t_i - n_{2i}$ for all i, but $t_i - n_{1i} \neq t_j - n_{1j}$ for $j \neq i$, so that $E \prod_{i=1}^{4} \varepsilon_{1,t_i-n_{1i}}\varepsilon_{2,t_i-n_{2i}} = \prod_{i=1}^{4} E\varepsilon_{1,t_i-n_{1i}}\varepsilon_{2,t_i-n_{2i}} = \sigma_{12}^4$. In this case there is no contribution due to the centering in $M_T(a_1, a_2)$. This means we only get a contribution if for each (i, l) there is (s, j) $(i \neq j)$ so that $t_i - n_{li} = t_j - n_{sj}$, but then $t_i - t_j = n_{li} - n_{sj}$ which is bounded in absolute value by N. Hence, when we sum over $\{t_i\}_{i=1}^4$ we get at most $2N(T - N)$ terms from the summation over t_i and t_j and at most T^2 when summing over the rest.

Next we evaluate the summation over $\{n_{1i}, n_{2i}\}_{i=1}^4$ for fixed $\{t_i\}_{i=1}^4$. For given (l, i) and t_i, n_{li} we define the sets of indices:

$$G_{li} = \{(s, j) : t_j - n_{sj} = t_i - n_{li}\}.$$

For fixed $\{t_i\}_{i=1}^4$ and (l, i) we have for $(s, j) \in G_{il}$ that $n_{sj} = n_{li} + t_j - t_i$ and we use this to eliminate all n_{sj} but the smallest, which we call n, and find $n_{sj} = n + r_{sj}$ where $r_{sj} \geq 0$. Let $r = \max_{s,j} r_{sj}$, then

$$\sum_{n_{sj}:s,j\in G_{li}} \prod_{s,j\in G_{li}} \zeta_{s,n_{sj}} = \sum_{n=0}^{N-1-r} \prod_{s,j\in G_{li}} \zeta_{s,n+r_{sj}}.$$

We first take $k \geq 3$ and evaluate

$$\zeta_{s,n+r_{sj}} \leq c(1 + \log(n + r_{sj}))^{m_s}(n + r_{sj})^{-1/2+\kappa} \leq cn^{-1/2+\kappa/3},$$

say, so that the sum of more than 3 products is bounded by $c\sum_{n=0}^{N-1-r} n^{(-1/2+\kappa/3)k} \leq c$. For $k=2$ use the bound

$$(1+\log(n+r_{s_ij_i}))^{m_{s_i}}(n+r_{s_ij_i})^{-1/2+\kappa} \leq (1+\log N)^{m_{s_i}} n^{-1/2+\kappa}$$

and find

$$c\sum_{n=0}^{N-1-r}\prod_{i=1}^{2}(1+\log(n+r_{s_ij_i}))^{m_{s_i}} n^{(-1/2+\kappa)} \leq c(1+\log N)^{m_{s_1}+m_{s_2}} N^{2\kappa}. \tag{80}$$

Thus the summation $\sum_{(1)}\prod_{i=1}^{4}\zeta_{1,n_{1i}}\zeta_{2,n_{2i}}$ for fixed $\{t_i\}_{i=1}^{4}$ depends on the groups of identical indices, but is clearly maximized for 4 pairs and the upper bound is $c(1+\log N)^{4(m_1+m_2)}N^{8\kappa}$. Therefore, the fourth moment is bounded by

$$cT^{-4}T^2N(T-N)(1+\log N)^{4m}N^{8\kappa} = c(1+\log N)^{4m}T^{-1}N^{1+8\kappa},$$

which proves (76).

Proof of (77): Note that because $n_1 < N \leq n_2$ the expectation of Q_T is zero. If we raise the sum to the fourth power we get (79), but now the summation $\sum_{(1)}$ is over $1 \leq n_{1i} < N \leq n_{2i} < t_i \leq T,\ i=1,\ldots,k$.

Because sums of products of three or more ζ terms are bounded as in (80), the largest contribution is from the case where the subscripts are equal in pairs, but $n_{1i} < N < n_{2i}$ implies that $t_i - n_{1i} > t_i - n_{2i}$. This means that $t_i - n_{1i}$ must equal $t_j - n_{1j}$ for some $j \neq i$, implying $|t_i - t_j| < N$. For fixed $\{t_i\}_{i=1}^{4}$ we find that summing over $n_{1j} = t_j - t_i + n_{1i}$ such a pair will give a contribution, apart from the log-factors, of $cN^{2\kappa}$, whereas summation over the other three pairs give at most $c(T^{2\kappa})^3$ because $N < n_{2i}, n_{2j} < T$. Similarly, summation over t_i, t_j give at most $N(T-N)^3$ terms, but then $E(Q_T(a_1,a_2)^4)$ is bounded by

$$T^{-4}c(1+\log T)^{4m}T^{6\kappa}(T-N)^3NN^{2\kappa} = c(1+\log T)^{4m}T^{-1+6\kappa}N^{1+2\kappa},$$

which proves (77).

Proof of (78): In order to prove tightness we check condition (**??**). It follows from (76) and (77) that for $\alpha < 1/(1+8\kappa) < (1-6\kappa)/(1+2\kappa)$ we have

$$||M_T(v_1,v_2)||_4 \leq c_T \to 0 \text{ and } ||Q_T(v_1,v_2)||_4 \leq c_T \to 0.$$

Next define $\delta_{t-n_1,t-n_2} = \varepsilon_{t-n_1}\varepsilon_{t-n_2} - \sigma_{12}1_{\{t-n_1=t-n_2\}}$ and consider $M_T(v_1,v_2) - M_T(\tilde{v}_1,\tilde{v}_2)$ which contains the difference

$$\begin{aligned} & \pi_{n_1}(-v_1)\pi_{n_2}(-v_2) - \pi_{n_1}(-\tilde{v}_1)\pi_{n_2}(-\tilde{v}_2) \\ = \ & (\pi_{n_1}(-v_1) - \pi_{n_1}(-\tilde{v}_1))\pi_{n_2}(-\tilde{v}_2) + \pi_{n_1}(-v_1)(\pi_{n_2}(-v_2) - \pi_{n_2}(-\tilde{v}_2)), \end{aligned}$$

where the first term is, by the meanvalue theorem,

$$(\pi_{n_1}(-v_1) - \pi_{n_1}(-\tilde{v}_1))\pi_{n_2}(-\tilde{v}_2) = (v_1 - \tilde{v}_1)\mathsf{D}\pi_{n_1}(-v_1^*)\pi_{n_2}(-\tilde{v}_2) = (v_1-\tilde{v}_1)\zeta_{1n_1}\zeta_{2n_2}.$$

Here ζ_{1n_1} and ζ_{2n_2} satisfy (69) with $a_i = v_i \geq -1/2-\kappa$ and $m_1 = 1, m_2 = 0$. Therefore we have from (76) that

$$||\sum_{t=N+1}^{T}\sum_{1\leq n_1,n_2<N}(\pi_{n_1}(-v_1) - \pi_{n_1}(-\tilde{v}_1))\pi_{n_2}(-\tilde{v}_2)\delta_{t-n_1,t-n_2}||_4 \leq c_T|v_1-\tilde{v}_1|,$$

where $c_T \to 0$, and a similar expression for the other term. This shows that

$$||M_T(v_1, v_2) - M_T(\tilde{v}_1, \tilde{v}_2)||_4 \leq c_T|v - \tilde{v}| \leq c|v - \tilde{v}|,$$

and hence that $M_T(v_1, v_2)$ is tight.

Next consider $||Q_T(v_1, v_2) - Q_T(\tilde{v}_1, \tilde{v}_2)||_4$ which is bounded by

$$||T^{-1}\sum_{t=N+1}^{T}(U_{1t} - \tilde{U}_{1t})V_t||_4 + ||T^{-1}\sum_{t=N+1}^{T}\tilde{U}_{1t}(V_t - \tilde{V}_t)||_4,$$

where $U_{1t} - \tilde{U}_{1t} = \sum_{n=0}^{N-1}(\pi_n(-v_1) - \pi_n(-\tilde{v}_1))\varepsilon_{t-n} = |v_1 - \tilde{v}_1|\sum_{n=0}^{N-1}\zeta_{1n}\varepsilon_{1,t-n}$ and $V_t - \tilde{V}_t = \sum_{n=N}^{t-1}(\pi_n(-v_2) - \pi_n(-\tilde{v}_2))\varepsilon_{2,t-n} = |v_2 - \tilde{v}_2|\sum_{n=N}^{t-1}\zeta_{2n}\varepsilon_{2,t-n}$. Now apply the same proof as above. ■

A.3 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.1 *We take three fractional indices $w, v,$ and u in the intervals*

$$[-w_0, -1/2 - \kappa_w], \ [-1/2 - \underline{\kappa}_v, -1/2 + \bar{\kappa}_v], \ \textit{and}\ [-1/2 + \kappa_u, u_0], \tag{81}$$

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

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 ($\to 0$) of $\bar{\kappa}_v$.

Thus for $Z_t \in \mathcal{Z}$, see Definition 1 and indices (w, v, u) as in Definition A.1, $\Delta_+^w Z_t^+$ is nonstationary, $\Delta_+^u Z_t^+$ is asymptotically stationary, and $\Delta_+^v Z_t^+$ is close to a critical process of the form $\Delta_+^{-1/2}\varepsilon_t$. In the subsequent lemmas we derive results for product moments of fractional differences of processes in the class $\mathcal{Z}$.

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

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

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 $\mathcal{K}$ to mean that $N_T^{-1}M_T(a_1, a_2)$ is tight on $\mathcal{K}$ and $M_T(a_1, a_2) = \mathbf{o}_P(N_T)$ to mean that $N_T^{-1}M_T(a_1, a_2) \Longrightarrow 0$ on $\mathcal{K}$.

Lemma A.8 *Let $Z_{it} = \xi_i\varepsilon_t + \Delta^{b_0}\sum_{n=0}^{\infty}\xi_{in}^*\varepsilon_{t-n} \in \mathcal{Z}, i = 1, 2,$ and define $M_T(a_1, a_2)$ as above and assume that $E|\varepsilon_t|^q < \infty$ for $q > \kappa_w^{-1}$ and $q \geq 8$. Then:*

(i) Uniformly for $-w_0 \leq w \leq -1/2 - \kappa_w$ and $-1/2 + \kappa_u \leq u \leq u_0$, see Definition A.1, we find

$$
\begin{aligned}
\mathsf{D}^m M_T(u_1, u_2) &\implies \mathsf{D}^m E(\Delta^{u_1} Z_{1t})(\Delta^{u_2} Z_{2t})', && (82)\\
M_T(w_1, w_2)T^{w_1+w_2+1} &\implies \xi_1 \int_0^1 W_{-w_1-1}(s)W_{-w_2-1}(s)'ds\xi_2', && (83)\\
\mathsf{D}^m M_T(w, u)T^{w+1/2} &= \mathbf{O}_P((1 + \log T)^{2+m}T^{-\min(\kappa_u,\kappa_w)}). && (84)
\end{aligned}
$$

Uniformly for $-w_0 \leq w \leq -1/2 - \kappa_w$, $-1/2 - \underline{\kappa}_v \leq v \leq v_0$, and $-1/2 + \kappa_u \leq u \leq u_0$ we find

$$
\begin{aligned}
M_T(w, v)T^{w+1/2} &= \boldsymbol{O}_P((1 + \log T)^2 T^{\underline{\kappa}_v}), && (85)\\
M_T(v, u) &= \boldsymbol{O}_P(1). && (86)
\end{aligned}
$$

(ii) If we assume further that $\underline{\kappa}_v < \min(b_0/2, 1/6)$ and choose $N = T^{\alpha}$, $\alpha < (1-6\underline{\kappa}_v)/(1+2\underline{\kappa}_v)$, and (ξ_1', ξ_2') has full rank, then for $-1/2 - \underline{\kappa}_v \leq v_i \leq -1/2 + \bar{\kappa}_v$ we find

$$
M_T((v_1, v_2), (v_1, v_2)) \geq c\frac{1 - N^{-2\bar{\kappa}_v}}{2\bar{\kappa}_v} + R_T, \quad (87)
$$

where $R_T = \mathbf{o}_P(1)$ uniformly for $|v_i + 1/2| \leq \underline{\kappa}_v$. This implies that for $(\bar{\kappa}_v, T) \to (0, \infty)$ we have

$$
\begin{aligned}
&\min_{-1/2-\underline{\kappa}_v \leq v_i \leq -1/2+\bar{\kappa}_v} \det(M_T((v_1, v_2), (v_1, v_2))) \xrightarrow{P} \infty, && (88)\\
&\max_{-1/2-\underline{\kappa}_v \leq v_i \leq -1/2+\bar{\kappa}_v} M_T((v_1, v_2), (v_1, v_2))^{-1} \xrightarrow{P} 0. && (89)
\end{aligned}
$$

Proof. A matrix valued process $\mathsf{D}^m M_T(a_1, a_2)$ 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 (62) of Lemma A.1 for $\mathsf{D}^m M_T(a_1, a_2)$. The moments are evaluated by $\xi_T(\zeta_1, \zeta_2)$, see (**??**), for suitable coefficients satisfying (69) and (70).

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 nonstationary 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 (83): We define the coefficients $\zeta_{i,t-n} = \pi_{t-n}(-u_i)$, which satisfy condition (69). 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 (**??**) and (73) 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

$$
||T^{-1}\sum_{t=1}^{T}(\Delta_+^{u_1}Z_{1t}^+ - \Delta_+^{\tilde{u}_1}Z_{1t}^+)(\Delta_+^{u_2}Z_{2t}^+)'||_4 + ||T^{-1}\sum_{t=1}^{T}(\Delta_+^{\tilde{u}_1}Z_{1t}^+)(\Delta_+^{u_2}Z_{2t}^+ - \Delta_+^{\tilde{u}_2}Z_{2t}^+)'||_4. \quad (90)
$$

We apply (**??**) 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 (69), see also JN (2010, Lemma B.3), and it follows from (73) with $a = 2\kappa_u$ that the first term of (90) is bounded by $c(u_1 - \tilde{u}_1)$. A similar proof works for the other term of (90), and tightness then follows from (62).

Notice that the second condition of (62) 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

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

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

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

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

Proof of (82): We define $\zeta_{i,t-n}^*(w_i) = T^{w_i+1/2}\pi_{t-n}(-w_i)$ 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 (**??**) and (74) with $\kappa = 2\kappa_w$, and find that (62) holds so that $M_T^{**}(w_1,w_2)$ is tight. Because $-1/(w_i+1/2) \leq \kappa_w^{-1} < q$ we obtain the limit

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

see (4) and JN (2010, Lemma D.2) for a few more details. The continuous mapping theorem gives the result (82).

Proof of (84): We apply (**??**) and (75) 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 (62) implies that $M_T^*(w,u) = \mathbf{O}_P((1+\log T)^2 T^{-\min(\kappa_u,\kappa_w)})$.

Proof of (85): We first apply (**??**) 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 (75) 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

$$\begin{aligned} ||M_T^*(w,v)||_4 &\leq c(1+\log T)T^{\underline{\kappa}_v}, \\ ||M_T^*(w,v) - M_T^*(\tilde{w},\tilde{v})||_4 &\leq c|(w,v)-(\tilde{w},\tilde{v})|(1+\log T)^2 T^{\underline{\kappa}_v}, \end{aligned} \tag{91}$$

and (62) shows that $M_T^*(w,v) = \mathbf{O}_P((1+\log T)^2 T^{\underline{\kappa}_v})$.

Proof of (86): 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 $\min(u+1, v+1, u+v+1) \geq \kappa_u - \underline{\kappa}_v > 0$, see Definition A.1. It then follows from (**??**) and (73) that (62) is satisfied and hence that $M_T(u,v)$ is tight.

Proof of (87), (88), and (89): Define $\tilde{Z}_{it}^{+}$ by $Z_{it}^{+} = \xi_i \varepsilon_t + \Delta_{+}^{b_0} \tilde{Z}_{it}^{+}$, $i = 1, 2$, and because we need to decompose the processes we use the notation

$$M_T(U, V) = T^{-1} \sum_{t=1}^{T} U_t^{+} V_t^{+\prime}$$

for product moments. We define $\xi = blockdiag(\xi_1, \xi_2)$, $\Delta_{+}^{v} Z_t = (\Delta_{+}^{v_1} Z_{1t}^{\prime}, \Delta_{+}^{v_2} Z_{2t}^{\prime})^{\prime}$, $\Delta_{+}^{v} \tilde{Z}_t = (\Delta_{+}^{v_1} \tilde{Z}_{1t}^{\prime}, \Delta_{+}^{v_2} \tilde{Z}_{2t}^{\prime})^{\prime}$, and $\Delta_{+}^{v} \varepsilon_t = (\Delta_{+}^{v_1} \varepsilon_t^{\prime}, \Delta_{+}^{v_2} \varepsilon_t^{\prime})^{\prime}$ and find the evaluation

$$M_T(\Delta_{+}^{v} Z, \Delta_{+}^{v} Z) \geq \xi M_T(\Delta_{+}^{v} \varepsilon, \Delta_{+}^{v} \varepsilon) \xi^{\prime} + M_T(\Delta_{+}^{b_0+v} \tilde{Z}, \Delta_{+}^{v} \varepsilon) \xi^{\prime} + \xi M_T(\Delta_{+}^{v} \varepsilon, \Delta_{+}^{b_0+v} \tilde{Z}),$$

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

We define the index $u_i = v_i + b_0 \geq -1/2 + (b_0 - \underline{\kappa}_v)$ for $\Delta_{+}^{b_0+v_i} \tilde{Z}_{it}^{+}$ so that $\kappa_u - \underline{\kappa}_v = b_0 - 2\underline{\kappa}_v > 0$. It follows that we can use (86) for the components of $M_T(\Delta_{+}^{b_0+v} \tilde{Z}, \Delta_{+}^{v} \varepsilon)$ 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 $t > N = T^{\alpha}$:

$$\Delta_{+}^{v_i} \varepsilon_t = \sum_{n=0}^{N-1} \pi_n(-v_i) \varepsilon_{t-n} + \sum_{n=N}^{t-1} \pi_n(-v_i) \varepsilon_{t-n} = U_{it}^{+} + V_{it}^{+}. \tag{92}$$

We define $U_t^{+} = (U_{1t}^{+\prime}, U_{2t}^{+\prime})^{\prime}$ and $V_t^{+} = (V_{1t}^{+\prime}, V_{2t}^{+\prime})^{\prime}$ 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 (77) in Lemma A.7 and find for $Q_T = M_T(U_{it}^{+}, V_{jt}^{+}) \Longrightarrow 0$ for $\alpha < (1 - 6\underline{\kappa}_v)/(1 + 2\underline{\kappa}_v)$. Thus,

$$M_T(U, V) + M_T(V, U) = \mathbf{o}_P(1). \tag{93}$$

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

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

see (65) and (72). Note that $F_{Nij} \to \infty$ as $(\bar{\kappa}_v, N) \to (0, \infty)$. The mean of $M_T(U, U)$ is

$$E(T^{-1} \sum_{t=N+1}^{T} U_t^{+} U_t^{+\prime}) = T^{-1}(T - N) \begin{pmatrix} F_{N11} & F_{N12} \\ F_{N12} & F_{N22} \end{pmatrix} \otimes \Omega_0,$$

and the difference $R_T(v_1, v_2) = M_T(U, U) - E(M_T(U, U)) \Longrightarrow 0$ uniformly for $|v_i + 1/2| \leq \underline{\kappa}_v$. It follows that $\xi M_T((v_1, v_2), (v_1, v_2)) \xi^{\prime}$ is bounded below by

$$\xi M_T(U, U) \xi^{\prime} \geq c \frac{1 - N^{-2\bar{\kappa}_v}}{2\bar{\kappa}_v} (\xi_1^{\prime}, \xi_2^{\prime})^{\prime} \Omega_0 (\xi_1^{\prime}, \xi_2^{\prime}) + \mathbf{o}_P(1),$$

where the remainder term is uniformly small for $|v_i + 1/2| \leq \underline{\kappa}_v$ independently of $\bar{\kappa}_v$. This proves (87), (88) and (89). ■

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

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

and so on.

Corollary A.9 *If the assumptions of Lemma A.8 (i) hold, then*

$$\begin{aligned} T^{w_1+w_2+1}M_T(w_1,w_2|u) &= T^{w_1+w_2+1}M_T(w_1,w_2) + \boldsymbol{o}_P(1), && (94)\\ M_T(u_1,u_2|w,u_3) &\Longrightarrow Var(\Delta^{u_1}Z_{1t},\Delta^{u_2}Z_{2t}|\Delta^{u_3}Z_{3t}), && (95)\\ M_T(v,u_1|w,u_2) &= \boldsymbol{O}_P(1). && (96) \end{aligned}$$

If the assumptions of Lemma A.8 (ii) hold and also $\alpha < (\kappa_w - \underline{\kappa}_v)/(1/2 + \underline{\kappa}_v)$, then

$$M_T((v_1,v_2),(v_1,v_2)|w,u) \geq c\frac{1-N^{-2\bar{\kappa}_v}}{2\bar{\kappa}_v} + R_T, \quad (97)$$

where $R_T = \mathbf{O}_P(1)$ uniformly for $|v_i + 1/2| \leq \underline{\kappa}_v$. This implies that, if $\mathcal{S}$ denotes the set defined by $\kappa_w, \kappa_u, \underline{\kappa}_v, \bar{\kappa}_v$ in (81), and $\underline{\kappa}_v < \min(b_0/2, 1/6)$, we have for $(\bar{\kappa}_v, T) \to (0,\infty)$,

$$\min_{\mathcal{S}} \det(M_T((v_1,v_2),(v_1,v_2)|w,u)) \overset{P}{\to} \infty, \quad (98)$$

$$\max_{\mathcal{S}} M_T((v_1,v_2),(v_1,v_2)|w,u)^{-1} \overset{P}{\to} 0. \quad (99)$$

Proof. *Proof of (94)*: We 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 (84) that $M_T^*(w_i,u) \Longrightarrow 0$, which together with (83) shows the result.

Proof of (95): 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 (94) and Lemma A.8 that the right hand side is $\mathbf{o}_P(1)$ as $T\to\infty$, because $M_T^*(u_i,w|u_3) = \mathbf{O}_P((1+\log T)^2 T^{-\min(\kappa_u,\kappa_w)})$, see (84). The result then follows from (83).

Proof of (96): 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 $M_T^*(w,u_2) \Longrightarrow 0$ by (84), 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 by (85), $M_T^*(w,v) = \mathbf{O}_P((1+\log T)^2 T^{\underline{\kappa}_v})$ and $M_T^*(w,u_1) = \mathbf{O}_P((1+\log T)^2 T^{-\min(\kappa_u,\kappa_w)})$, so that by (82) and because $\underline{\kappa}_v < \min(\kappa_u,\kappa_w)$, $M_T^*(v,w)M_T^{**}(w,w)^{-1}M_T^*(w,u_1) \Longrightarrow 0$. Using (83) and (86) the result follows.

Proof of (97), (98), and (99): The proof is similar to that of (87) 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 (95), and $M_T((v_1,v_2),u|w) = \mathbf{O}_P(1)$, see (96). Thus, for $\Delta_+^{v_i} Z_{it}^+ = \Delta_+^{v_i}\varepsilon_t + \Delta_+^{v_i+b_0}\tilde{Z}_{it}^+, i = 1,2, Z_t^+ = (Z_{1t}^{+\prime}, Z_{2t}^{+\prime})'$, and $\Delta_+^w Z_{3t}^+ = \Delta_+^w \varepsilon_t + \Delta_+^{w+b_0}\tilde{Z}_{3t}^+$ 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

$$M_T(\Delta^v\varepsilon, \Delta^v\varepsilon|\Delta^w Z_3) + M_T(\Delta^v\varepsilon, \Delta^{b_0+v}\tilde{Z}|\Delta^w Z_3) + M_T(\Delta^{b_0+v}\tilde{Z}, \Delta^v\varepsilon|\Delta^w Z_3).$$

It follows from (96) for $u_i = v_i + b_0 \geq -1/2 + b_0 - \underline{\kappa}_v$ (i.e., $\kappa_u = b_0 - \underline{\kappa}_v > \underline{\kappa}_v$), and $w \leq -1/2 - \kappa_w$, that the last two terms are $\boldsymbol{O}_P(1)$.

We next decompose the first term as $\Delta_+^{v_i}\varepsilon_t = U_{it}^+ + V_{it}^+$, see (92), 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 (93) that $M_T(U,V) = T^{-1}\sum_{t=N+1}^T U_t^+ V_t^{+\prime} = \boldsymbol{o}_P(1)$, because $\alpha < (1 - 6\underline{\kappa}_v)/(1+2\underline{\kappa}_v)$. From (82) we find 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 (85) shows that $M_T^*(\Delta^w Z_3, V) = \mathbf{O}_P((1+\log T)^2 T^{\underline{\kappa}_v})$. For the term

$$M_T^*(U,\Delta^w Z_3) = T^{-1}\sum_{t=N+1}^T U_t^+ \Delta_+^w Z_{3t}^{+\prime} T^{w+1/2} = \sum_{n=0}^{N-1}\pi_n(-v_1)(T^{-1}\sum_{t=N+1}^T \varepsilon_{t-n}\Delta_+^w Z_{3t}^{+\prime}T^{w+1/2}),$$

we apply (84) with $u = 0 = -1/2 + 1/2$ ($\kappa_u = 1/2$) so that $T^{-1}\sum_{t=N+1}^T \varepsilon_{t-n}\Delta_+^w Z_{3t}^{+\prime}T^{w+1/2} = \boldsymbol{O}_P((1+\log T)^2 T^{-\kappa_w})$. It follows that

$$M_T^*(U,\Delta^w Z_3) = \mathbf{O}_P((1+\log T)^2 T^{-\kappa_w}\sum_{n=1}^{N-1} n^{-v_1-1}) = \mathbf{O}_P((1+\log T)^2 T^{-\kappa_w + \alpha(1/2+\underline{\kappa}_v)}). \quad (100)$$

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

Finally we need to analyze

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

where the dependence on $\bar{\kappa}_v$ enters. Here $M_T(U,U) \geq c(1 - N^{-2\bar{\kappa}_v})/2\bar{\kappa}_v + \boldsymbol{o}_P(1)$, where the $\boldsymbol{o}_P(1)$ term is uniform in $|v_i + 1/2| \leq \underline{\kappa}_v$, see (87). Furthermore, (100) shows that $M_T^*(U,\Delta^w Z_3) = \boldsymbol{o}_P(1)$ because $\alpha < (\kappa_w - \underline{\kappa}_v)/(1/2 - \underline{\kappa}_v)$[11]. Together with (82) this proves (97), (98), and (99). ∎

[11] SJ: Hvad skal du bruge $< \kappa_w/(1/2+\underline{\kappa}_v)$ til?

A.4 Limit theory for product moments of deterministic terms

The next lemma gives results for the impact of initial values and deterministic terms, see (38), in the models considered, using the bounds in JN (2010, Lemma C.1). We define

$$d > b : D_{it}(\psi) = \begin{cases} \Delta_+^{d+ib}\mu_{0t} + \Delta_-^{d+ib}X_t + \Delta_+^{d+ib}\xi_{0t}, & d_0 \geq 1/2 \\ \Delta_+^{d+ib}\xi_{0t}, & d_0 < 1/2 \end{cases} \quad (101)$$

$$d = b, D_{it}(d) = \begin{cases} \Delta_+^{(1+i)d}\mu_{0t} + \Delta_-^{(1+i)d}X_t + \Delta_+^{(1+i)d}\xi_{0t}^*, & d_0 \geq 1/2 \\ \Delta_+^{(1+i)d}\xi_{0t}^* & d_0 < 1/2 \end{cases} \quad (102)$$

and let D^m be the derivative with respect to $d + ib$. Note that when $d = b$ the term $C_1\alpha_0\rho_0'\pi_t(1)$ is not included in $D_{it}(\psi)$ but is part of ??.

Lemma A.10 *For $\eta_1 > 0$, and $0 < \kappa_1 < 1/2$ we have*

(i) For $\delta_i = d + ib - d_0$ and $d - b \geq \eta_1$, the functions $\mathsf{D}^m D_{it}(\psi)$ are continuous in ψ and

$$\max_{-1/2-\kappa_1 \leq \delta_i \leq u_1} |\mathsf{D}^m D_{it}(\psi)| \rightarrow 0 \text{ as } t \rightarrow \infty, \quad (103)$$

$$\max_{-u_0 \leq \delta_i \leq -1/2-\kappa_1} \max_{1 \leq t \leq T} |\mathsf{D}^m T^{\delta_i+1/2}\beta_{0\perp}' D_{it}(\psi)| \rightarrow 0 \text{ as } T \rightarrow \infty. \quad (104)$$

(ii) If the initial values satisfy $X_{-n} = 0, n > N_0$, then (103)–(104) hold for $d - b \geq 0$.

(iii) If $d = b$ $(\eta \leq d \leq d_1)$ the results (103)–(104) hold if $\eta \leq b_0 \neq 1/2$.

Proof. Theorem 3 and (6) show that because $C_0\alpha_0\beta_0' = 0$, we have that $\mu_{0t} = -\Pi_+(L)^{-1}\Pi_-(L)X_t$ is

$$-(C_0\Delta_+^{-d_0} + \Delta_+^{-d_0+b_0}F_+(L))(-\alpha_0\beta_0'\Delta_-^{d_0-b_0} + \sum_{j=0}^k \Psi_{0j}\Delta_-^{d_0+jb_0})X_t$$

$$= -\sum_{j=0}^k [C_0\Psi_{0j}\Delta_+^{-d_0} + F_+(L)\Psi_{0j}\Delta_+^{-d_0+b_0}]\Delta_-^{d_0+jb_0}X_t + F_+(L)\alpha_0\beta_0'\Delta_+^{-d_0+b_0}\Delta_-^{d_0-b_0}X_t. \quad (105)$$

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

$$\max_{\kappa_1 \leq \min(u+v,u+1,v) \leq u_1} |G_+(L)\mathsf{D}^m\Delta_+^u\Delta_-^v X_t| \rightarrow 0 \text{ as } t \rightarrow \infty, \quad (106)$$

$$\max_{\kappa_1 \leq \min(v-1/2,-u-1/2,1/2) \leq u_1} \max_{1 \leq t \leq T} |G_+(L)\mathsf{D}^m T^{u+1/2}\Delta_-^u\Delta_-^v X_t| \rightarrow 0 \text{ as } T \rightarrow \infty. \quad (107)$$

It is seen from (63) that differentiating the fractional coefficients give an extra factor of the order $(1 + \log t)$ and it is seen from the proof, that such a factor does not change the results, so we shall give the proof for $m = 0$.

Proof of (i):

Proof of (103) and (104) for $\Delta_-^{d+ib}X_t, d_0 \geq 1/2$: We find, uniformly for $d + ib \geq \eta_1$ that for $t \rightarrow \infty$,

$$|\Delta_-^{d+ib}X_t| = |\sum_{n=0}^\infty \pi_{t+n}(-d-ib)X_{-n}| \leq c\sum_{n=0}^\infty (t+n)^{-d-ib-1} \leq c\sum_{n=0}^\infty (t+n)^{-\eta_1-1} \rightarrow 0, \quad (108)$$

which proves (103). Multiplying by $T^{\delta_i+1/2} \to 0$ when $\delta_i + 1/2 \le -\kappa_1$ shows (104).

Proof of (103) for $\Delta_+^{d+ib}\mu_{0t}, d_0 \ge 1/2$: The term $\Delta_+^{d+ib}\mu_{0t}$ contains terms of the form $G(L)\Delta_+^u\Delta_-^v X_t$ where $u = d + ib - \gamma_0$ and $v = d_0 + jb_0 \ge \gamma_0$ with $\gamma_0 = d_0$ or $\gamma_0 = d_0 - b_0$, see (105). If $d + ib - \gamma_0 \ge -1/2 - \kappa_1$ in (103), then for both choices of γ_0 we find $\min(u+v, u+1, v) \ge \min(\eta_1, 1/2 - \kappa_1, d_0 - b_0) > 0$, and the result (103) follows from (106).

Proof of (104) for $\Delta_+^{d+ib}\mu_{0t}, d_0 \ge 1/2$: If $d + ib - \gamma_0 \le -1/2 - \kappa_1$ in (104), then $v \ge -u$ so that $\min(v - 1/2, -u - 1/2, 1/2) = \min(-u - 1/2, 1/2) \ge \kappa_1 > 0$, and the result follows from (107) because the normalization $T^{d+ib-d_0+1/2}$ of $\Delta_+^{d+ib}\beta_{0\perp}'\mu_{0t}$ is enough for both choices of γ_0 to ensure $\max_{1\le t\le T}|T^{d+ib-d_0+1/2}\Delta_+^{d+ib}\beta_{0\perp}'\mu_{0t}| \to 0$. Note that the condition $d - b \ge \eta_1$ is not used in this case.

Proof of (103) and (104) for $\Delta_+^{d+ib}\xi_{0t}$: We find from Theorem 3 that $|\Delta_+^{d+ib}\xi_{0t}| \le ct^{-\eta_1}$ when $d-b \ge \eta_1$, which proves both (103) and (104) because $T^{\delta_i+1/2} \to 0$ when $\delta_i+1/2 \le -\kappa_1$.

Proof of (ii): We next consider the case $X_{-n} = 0, n \ge N_0$.

Proof of (103) and (104) for $\Delta_-^{d+ib}X_t$: From (108) we find that $|\Delta_-^{d+ib}X_t| \le c\sum_{n=0}^{N_0}(t+n)^{-(d+ib)-1} \le ct^{-1} \to 0$ for $d+ib \ge 0$, so that (103) holds. In (104) we assume $\delta_i \le -1/2-\kappa_1$ so that $T^{\delta_i+1/2} \to 0$, and that proves (104).

Proof of (103) and (104) for $\Delta_+^{d+ib}\mu_{0t}$: This contains terms of the form $G(L)\Delta_+^u\Delta_-^v X_t$. For $u = d + ib - \gamma_0 \ge -1/2 - \kappa_1$ and $v = d_0 + jb_0 \ge \gamma_0$ we have $u + v \ge 0$. For such a term we apply the evaluation from JN (2010, Lemma B.4, equation (62))

$$|\Delta_+^u\Delta_-^v X_t| = |\sum_{n=0}^{N_0}(\sum_{j=0}^{t-1}\pi_j(-u)\pi_{n+t-j}(-v))X_{-n}| \le ct^{-\min(u+v+1,u+1,v+1)}, \tag{109}$$

which tends to zero because $\min(u + v + 1, u + 1, v + 1) \ge \min(1, 1/2 - \kappa_1, 1 + \gamma_0) > 0$. This proves (103) for the term $\Delta_+^{d+ib}\mu_{0t}$. For (104) we need not prove anything because the condition $d - b \ge \eta_1$ was not used in the proof in case (i).

Proof of (iii): The model with $d = b$ is not covered by the previous results because they were proved under the assumption that $d - b \ge \eta_1 > 0$.

Proof of (103) and (104) for $\Delta_-^{(1+i)d}X_t$: For $i \ge 0$ we find from (108) that $|\Delta_-^{(1+i)d}X_t| \le ct^{-d} \le ct^{-\eta} \to 0$ whereas for $i = -1$ we have $\Delta_-^0 X_t = 0$.

Proof of (103) and (104) for $\Delta_+^{(1+i)d}\mu_{0t}$: Because the condition $(1 + i)d \ge \eta_1$ was not used in the proof of (104) for the term $T^{\delta_i+1/2}\beta_{0\perp}'\mu_{0t}$ we only have to consider (103). As in the proof of (i) we get terms of the form $G(L)\Delta_+^u\Delta_-^v X_t$ where $u = d(1 + i) - \gamma_0$ is assumed $\ge -1/2 - \kappa_1$, and $v = d_0(1 + j) \ge d_0$, for $j \ge 0$ with $\gamma_0 = d_0$ or $\gamma_0 = 0$, see (105), and the term with $\Delta_-^{d_0-d_0}X_t = \Delta_-^0 X_t = 0$. Then for $i \ge 0$ we find $u + v \ge d(i+1) - \gamma_0 + d_0(1+j) \ge d - \gamma_0 + d_0 \ge \eta$ so that $\min(u + v, u + 1, v) \ge \min(\eta, 1/2 - \kappa_1, d_0) > 0$ which proves (103) for $i \ge 0$. For $i = -1$ we get $u = -\gamma_0 \ge -1/2 - \kappa_1, v \ge d_0$. If $\gamma_0 = 0$, then $u = 0$ and $|\Delta_+^u\Delta_-^v X_t| = |\Delta_-^v X_t| \le ct^{-d_0} \to 0$, and if $\gamma_0 = d_0$ then we are left with the term $\Delta_+^{-d_0}\Delta_-^{d_0}X_t$, where $u = -d_0 \ge -1/2 - \kappa_1$ implies $d_0 \le 1/2 + \kappa_1$. If $d_0 > 1/2$ we choose $\kappa_1 < d_0 - 1/2$ and there is nothing to prove. If $d_0 = b_0 < 1/2$ we instead represent X_t by the stationary solution, $X_t = C_0\Delta^{-d_0}\varepsilon_t + Y_t + \xi_{0t}$ then μ_{0t} is not present and there is nothing to prove.

Proof of (103) and (104) for $\Delta_+^{d+ib}\xi_{0t}^*$: Finally, see (102), we have $|\Delta_+^{d+ib}\xi_{0t}^*| = |\Delta_+^{d(1+i)}\sum_{t=0}^{t-1}\tau_{n0}\alpha_0\rho_0'\pi_{t-n}($ $d_0)| \to 0$ uniformly in $d \in [\eta, d_1]$. ■

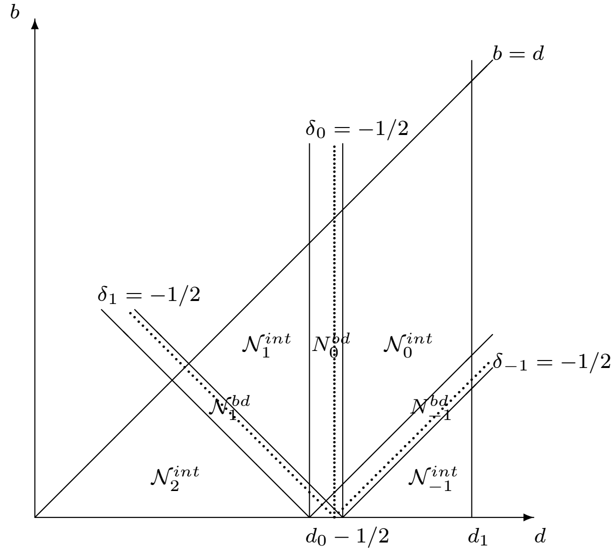

Figure 1: The parameter space $\mathcal{N}$ is the set bounded by $b > 0$, $b \leq d$, and $d \leq d_1$. The sets $\mathcal{N}_m^{bd} = \mathcal{N}_m^{bd}(\kappa_1, \kappa_2)$, where a process is close to being critical, and the sets $\mathcal{N}_m^{int} = \mathcal{N}_m^{int}(\kappa_1, \kappa_2)$ are illustrated assuming $k = 1$. If $k \geq 2$ there would be more lines.

Appendix B Proof of Theorem 4

We give the proof for case (i).

By Lemma A.10 the deterministic terms generated by initial values are uniformly small. Note that (103) is formulated for index $\geq -1/2-\kappa_1$, which covers not only the asymptotically stationary $\beta_0' X_{jt}$ and $\beta_{0\perp}' X_{it}$ but also those which are nearly critical, whereas (104) deals with the nonstationary $\beta_0' X_{jt}$ and $\beta_{0\perp}' X_{it}$. Hence initial values do not influence the limit behavior of product moments, and in the remainder of the proof of Theorem 4 we therefore assume that the deterministic terms generated by initial values are zero.

In the following we 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 from 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 from the regression. Special problems arise if the regression contains processes that are nearly critical. These results are made precise in Lemma A.8 and Corollary A.9, which we apply repeatedly below to prove uniform convergence.

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 $\delta_m = d - d_0 + mb$. Thus the fractional order is the negative fractional index. For notational reasons in Definition B.2 below we define $\delta_{-2} = -\infty$ and $\delta_{k+1} = \infty$.

The process $\Delta^{d+mb} \beta_{0\perp}' X_t$ is critical if $\delta_m = d + mb - d_0 = -1/2$, see Figure 1, and we partition the parameter space into "interiors" and "boundaries" as given in the next definition.

Definition B.2 *We take $0 < \kappa_2 < \kappa_1$ and define the $(\kappa_1, \kappa_2)-$interiors,*

$$\mathcal{N}_m^{int}(\kappa_1, \kappa_2) = \{\psi \in \mathcal{N} : \delta_{m-1} \leq -1/2 - \kappa_1 \text{ and } -1/2 + \kappa_2 \leq \delta_m\},\ -1 \leq m \leq k+1, \tag{110}$$

and the $(\kappa_1, \kappa_2)-$boundaries,

$$\mathcal{N}_m^{bd}(\kappa_1, \kappa_2) = \{\psi \in \mathcal{N} : -1/2 - \kappa_1 \leq \delta_m \leq -1/2 + \kappa_2\},\ -1 \leq m \leq k. \tag{111}$$

Note (recalling $\delta_{k+1} = \infty$) that $\mathcal{N}_{k+1}^{int}(\kappa_1, \kappa_2) = \mathcal{N}_{k+1}^{int}(\kappa_1)$ does not depend on κ_2 and

$$\begin{aligned} \mathcal{N}_{\text{conv}}(\kappa) &= \cup_{m=-1}^{k-1}(\mathcal{N}_m^{int}(\kappa_1, \kappa_2) \cup \mathcal{N}_m^{bd}(\kappa_1, \kappa_2)) \cup \mathcal{N}_k^{int}(\kappa_1, \kappa) = \{\psi \in \mathcal{N} : \delta_k \geq -1/2 + \kappa\}, \\ \mathcal{N}_{\text{div}}(\kappa) &= \mathcal{N}_{k+1}^{int}(\kappa_1) \cup \mathcal{N}_k^{bd}(\kappa_1, \kappa) = \{\psi \in \mathcal{N} : \delta_k \leq -1/2 + \kappa\}. \end{aligned}$$

We have defined in (110) the $(\kappa_1, \kappa_2)-$interior $\mathcal{N}_m^{int}(\kappa_1, \kappa_2)$ 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 $\geq -1/2 + \kappa_2$ or $\leq -1/2 - \kappa_1$. The $(\kappa_1, \kappa_2)-$boundary $\mathcal{N}_m^{bd}(\kappa_1, \kappa_2)$ is the set where the process $\beta_{0\perp}' X_{mt}$ has an index which is close to the critical value of $-1/2$, see Figure 1 for an illustration.

We apply Lemma A.8 and Corollary A.9 as well as Definition A.1 of the notation $\kappa_w, \underline{\kappa}_v, \overline{\kappa}_v$, and κ_u. We note that for $(d, b) \in \mathcal{N}$ all indices are bounded.

B.1 Proof of (24): divergence of $\ell_{T,p}(\psi)$ on $\mathcal{N}_{\mathbf{div}}(\kappa) \cap \mathcal{K}(\eta, \eta_1)$

In this proof we first show that, for a suitable $\kappa_1 > 0$,

$$\inf_{\psi \in \mathcal{N}_k^{bd}(\kappa_1, \kappa) \cap \mathcal{K}(\eta, \eta_1)} \ell_{T,p}(\psi) \xrightarrow{P} \infty \text{ as } (\kappa, T) \to (0, \infty). \tag{112}$$

Then we show that, for the above choice of κ_1 and for fixed $\kappa > 0$,

$$\inf_{\psi \in \mathcal{N}_{k+1}^{int}(\kappa_1) \cap \mathcal{K}(\eta, \eta_1)} \ell_{T,p}(\psi) \xrightarrow{P} \infty \text{ as } T \to \infty. \tag{113}$$

Proof of (112): In this set, $\beta_{0\perp}' X_{kt}$ is near critical with index $v_1 = d + kb - d_0 \in [-1/2 - \kappa_1, -1/2 + \kappa]$, so we take $\underline{\kappa}_v = \kappa_1$ and $\overline{\kappa}_v = \kappa$ and choose $\kappa < \eta/2$. The processes $\beta_{0\perp}' X_{it}, i < k$, are nonstationary with index $d + ib - d_0 \leq d + (k-1)b - d_0 \leq -1/2 - (\eta - \kappa) \leq -1/2 - \eta/2$, which determines $\kappa_w = \eta/2$. Then $q > 2/\eta$ are enough moments for weak convergence of the nonstationary processes to fBM, c.f. (82) of Lemma A.8. The stationary processes $\beta_0' X_{jt}$ have index $d + jb - d_0 + b_0 \geq -d_0 + b_0 = -1/2 + (1/2 - d_0 + b_0)$ such that $\kappa_u = 1/2 - d_0 + b_0 > 0$.

We want to apply Corollary A.9, so the choice of $\kappa_1 = \underline{\kappa}_v$ needs to be such that the condition from Corollary A.9, $\underline{\kappa}_v < \min(b_0/2, 1/6)$, is satisfied along with $\underline{\kappa}_v < \min(\kappa_u, \kappa_w)$ from Definition A.1. The choice $\kappa_1 = \underline{\kappa}_v < \eta/2$ satisfies these inequalities if we take $\eta/2 < \min(1/6, 1/2 - 2(d_0 - b_0))$. We further take κ_1 so close to $\eta/2$ that $q > 1/\kappa_1 > 2/\eta$.

For $X_{kt} = \bar{\beta}_0 \beta_0' X_{kt} + \bar{\beta}_{0\perp} \beta_{0\perp}' X_{kt} = B_0(X_{kt}' \beta_0, X_{kt}' \beta_{0\perp})'$ where $B_0 = (\bar{\beta}_0, \bar{\beta}_{0\perp})$, see (8), we decompose the determinant $\det(SSR_T(\psi)) = \det(B_0 M_T((v_1, u_2), (v_1, u_2)|u, w) B_0')$ as

$$\det(M_T(u_2, u_2|u, w)) \det(M_T(v_1, v_1|u, u_2, w))(\det(B_0))^2,$$

where the first factor converges in distribution by (95) uniformly in $\psi \in \mathcal{N}_k^{bd}(\kappa_1, \kappa)$, and the second factor diverges to infinity uniformly in $\psi \in \mathcal{N}_k^{bd}(\kappa_1, \kappa)$ for $(\kappa, T) \to (0, \infty)$, see (99). This proves (112).

Proof of (113): On $\mathcal{N}_{k+1}^{int}(\kappa_1)$, all $\beta_{0\perp}' X_{it}$ are nonstationary with index $d + ib - d_0 \leq -1/2 - \kappa_1$ so we set $\kappa_w = \kappa_1$. Thus, we need $q > \kappa_1^{-1}$ moments for weak convergence of the nonstationary processes to fBM, which is satisfied because $1/q > 1/\kappa_1 > 2/\eta$. As above, $\beta_0' X_{jt}$ are stationary with $\kappa_u = 1/2 - d_0 + b_0$.

We decompose $\det(SSR_T(\psi)) = \det(B_0 M_T((w_1, u_2), (w_1, u_2)|u, w) B_0')$ as

$$\det(M_T(w_1, w_1|w, u)) \det(M_T(u_2, u_2|w_1, u, w)) \det(B_0)^2.$$

The second factor is $\mathbf{O}_P(1)$ uniformly in $\psi \in \mathcal{N}_{k+1}^{int}(\kappa_1) \cap \mathcal{K}(\eta, \eta_1)$ by (95). In the first factor we normalize $T^{w_1+w_2-1} M_T(w_1, w_1|w, u)$ to convergence, see (94), so that the first factor is $\mathbf{O}_P(T^{-(w_1+w_2-1)}) = \mathbf{O}_P(T^{2\kappa_1})$, and

$$\min_{\psi \in \mathcal{N}_{k+1}^{int}(\kappa_1) \cap \mathcal{K}(\eta,\eta_1)} \ell_{p,T}(\psi) \xrightarrow{P} \infty \text{ as } T \to \infty.$$

This proves (113) and completes the proof of (24).

B.2 Proof of (25): weak convergence of $\ell_{T,p}(\psi)$ to $\ell_p(\psi)$ on $C(\mathcal{N}_{\mathbf{conv}}(\kappa) \cap \mathcal{K}(\eta, \eta_1))$

For this proof, we take $\kappa > 0$ as fixed and examine the subsets of $\mathcal{N}_{\text{conv}}(\kappa)$ in turn. The idea in the proof below is that we first choose a suitable fixed $\kappa_1 > 0$ and then show, for $m \leq k-1$,

$$\sup_{\psi \in \mathcal{N}_m^{bd}(\kappa_1,\kappa_2) \cap \mathcal{K}(\eta,\eta_1)} |\ell_{T,p}(\psi) - \ell_p(\psi)| \xrightarrow{P} 0 \text{ as } (\kappa_2, T) \to (0, \infty). \tag{114}$$

Next we fix $\kappa_1 > 0$ (at the above choice) and also fix $\kappa_2 > 0$ and show that, for $m \leq k$,

$$\sup_{\psi \in \mathcal{N}_m^{int}(\kappa_1,\kappa_2) \cap \mathcal{K}(\eta,\eta_1)} |\ell_{T,p}(\psi) - \ell_p(\psi)| \xrightarrow{P} 0 \text{ as } T \to \infty, \tag{115}$$

noting that in the case $m = k$ we have $\kappa_2 = \kappa$.

Proof of (114): We want to show that the profile likelihood $\ell_{T,p}(\psi) = \log\det(SSR_T(\psi))$, see (21), converges as a continuous process to $\ell_p(\psi)$ by choosing a suitable κ_1 and letting $(\kappa_2, T) \to (0, \infty)$ in the application of Corollary A.9.

On $\mathcal{N}_m^{bd}(\kappa_1, \kappa_2)$ the near critical index of $\beta_{0\perp}' X_{mt}$ is $v = d + mb - d_0 \in [-1/2 - \kappa_1, -1/2 + \kappa_2]$, and we choose $\underline{\kappa}_v = \kappa_1 < \eta/2$ as above so that $q > 1/\kappa_1 > 1/\kappa_w$ and $\overline{\kappa}_v = \kappa_2 < \eta/2$.

The profile likelihood is derived by regressing $\Delta^{d+kb} X_t$ on the other variables, which can be either asymptotically stationary or not. Again, the processes $\Delta^{d+jb} \beta_0' X_t$ is stationary and we choose $\kappa_u = 1/2 - d_0 + b_0$. We collect all asymptotically stationary regressors $\{\beta_{0\perp}' X_{it}\}_{i=m+1}^{k-1}$ and $\{\beta_0' X_{jt}\}_{j=-1}^{k-1}$ in a vector where the lowest fractional index for $\beta_{0\perp}' X_{it}$ is $\delta_{m+1} = \delta_m + b \geq -1/2 + (\eta - \kappa_1) > -1/2 + \eta/2$, so we choose $\kappa_u = \eta/2$. The nonstationary processes $\{\beta_{0\perp}' X_{it}\}_{i=-1}^{m-1}$ are collected in a vector with largest fractional index $w = \delta_{m-1} = d + (m-1)b - d_0 \leq -1/2 + \kappa_2 - b \leq -1/2 - \eta/2$, so we set $\kappa_w = \eta/2$. This implies that $q > 2/\eta = \kappa_w^{-1}$ are enough moments to get weak convergence of the nonstationary processes to fBM.

Since $\psi \in \mathcal{N}_m^{bd}(\kappa_1, \kappa_2)$ and $m \leq k-1$, $\beta_{0\perp}' X_{kt}$ and $\beta_0' X_{kt}$ are asymptotically stationary (indices $u_1 = d + kb - d_0 = d + mb - d_0 + (k-m)b \geq -1/2 - \kappa_1 + b \geq -1/2 + \eta/2$ and $u_2 = d + kb - d_0 + b_0 \geq -1/2 + (1/2 - d_0 + b_0) \geq -1/2 + \eta/2$.

Thus, for the application of Corollary A.9, we have chosen

$$\kappa_u = \kappa_w = \eta/2, \overline{\kappa}_v = \kappa_2 < \eta/2, \underline{\kappa}_v = \kappa_1 < \eta/2,$$

so that κ_1 satisfies the conditions for the results in Corollary A.9.

We find, see Corollary A.9, that $SSR_T(\psi) = B_0 M_T((u_1, u_2), (u_1, u_2)|v, w, u) B_0'$, where

$$\begin{aligned}
& B_0 M_T((u_1, u_2), (u_1, u_2)|v, w, u) B_0' - Var(U_{kt}|\mathcal{F}_{stat}(\psi)) \\
= \; & B_0 M_T((u_1, u_2), (u_1, u_2)|w, u) B_0' - Var(U_{kt}|\mathcal{F}_{stat}(\psi)) \\
& - B_0 M_T((u_1, u_2), v|w, u) M_T(v, v|w, u)^{-1} M_T(v, (u_1, u_2)|w, u) B_0'.
\end{aligned}$$

For fixed $\kappa_2 > 0$, we find from (96) the uniform convergence on $C(\mathcal{N}_m^{bd}(\kappa_1, \kappa_2) \cap \mathcal{K}(\eta, \eta_1))$:

$$B_0 M_T((u_1, u_2), (u_1, u_2)|w, u)B_0' - Var(S_{kt}|\mathcal{F}_{stat}(\psi)) \Longrightarrow 0 \text{ as } T \to \infty.$$

From (97) we find similarly that for $\underline{\kappa}_v < \min(\kappa_u, \kappa_w)$ it follows that on $C(\mathcal{N}_m^{bd}(\kappa_1, \kappa_2) \cap \mathcal{K}(\eta, \eta_1))$

$$M_T(u_i, v|w, u) = \mathbf{O}_P(1) \text{ as } T \to \infty.$$

Finally we find from (99) that if $\mathcal{S}$ denotes the set defined by the choices of $(\kappa_u, \kappa_w, \underline{\kappa}_v, \bar{\kappa}_v)$ we have

$$\max_{\mathcal{S}} M_T(v, v|w, u)^{-1} \overset{P}{\to} 0 \text{ as } (\kappa_2, T) \to (0, \infty)$$

with $N = T^{\alpha}$ for some $\alpha < (\kappa_w - \underline{\kappa}_v)/(1/2 + \underline{\kappa}_v)$,which is a positive number because $(\eta - \kappa_2 - \kappa_1)/(1/2 + \kappa_1) > 0$.

This proves (114) on $\mathcal{N}_m^{bd}(\kappa_1, \kappa_2) \cap \mathcal{K}(\eta, \eta_1)$, $m = -1, \ldots, k-1$.

Proof of (115): For $\psi \in \mathcal{N}_m^{int}(\kappa_1, \kappa_2)$, we collect all asymptotically stationary regressors $\{\beta_{0\perp}' X_{it}\}_{i=m}^{k-1}$ and $\{\beta_0' X_{jt}\}_{j=-1}^{k-1}$ in a vector where the lowest fractional index for $\beta_{0\perp}' X_{it}$ is $\delta_m \geq -1/2 + \kappa_2$ and the lowest for $\beta_0' X_{jt}$ is $-1/2 + 1/2 - d_0 + b_0$, so we choose $\kappa_u = \kappa_2$. The nonstationary processes $\{\beta_{0\perp}' X_{it}\}_{i=-1}^{m-1}$ are collected in a vector with largest fractional index $w = \delta_{m-1} \leq -1/2 - \kappa_1$, so $\kappa_w = \kappa_1$. Because $m \leq k$, $\beta_{0\perp}' X_{kt}$ and $\beta_0' X_{kt}$ are asymptotically stationary with indices $u_1 = d + kb - d_0 \geq -1/2 + \kappa_2$ and $u_2 = d + kb - d_0 + b_0 \geq u_1 \geq -1/2 + \kappa_2$.

Here, κ_1 and κ_2 are fixed, and we want to apply (95) of Corollary A.9. With $\kappa_w = \kappa_1$ chosen as $1/q > \kappa_1^{-1} > 2/\eta$ we need $q > 2/\eta$ moments for weak convergence of the nonstationary processes to fBM.

With the notation from Lemma A.8 and Corollary A.9 we have that $SSR_T(\psi) = B_0 M_T((u_1, u_2), (u_1, u_2)|$ and it follows from (95), see also (23), that for fixed κ_1, κ_2 and $T \to \infty$, (115) follows.

B.3 Proof of (26): Unique minimum

On $\mathcal{N}_{\text{div}}(0)$ the inequality is trivially satisfied and on $\mathcal{N}_{conv}(0)$ we have that $U_{kt} = \Delta^{d+kb-d_0}(C_0 \varepsilon_t + \Delta^{b_0} Y_t)$ is stationary. The transfer function for $U_t = C_0 \varepsilon_t + \Delta^{b_0} Y_t$ is $f_0(z)^{-1}$, where $f_0(z) = (1-z)^{-d_0} \Pi_0(z) = (1-z)^{-b_0} \Psi_0(z)$ for $|z| < 1$, see (3). For given ψ let us assume that $\{\beta_{0\perp}' U_{it}\}_{i=m}^{k}$ are stationary and $\{\beta_{0\perp}' U_{it}\}_{i=-1}^{m-1}$ are nonstationary. We define

$$S_t = U_{kt} - \sum_{i=m}^{k-1} \Psi_i \bar{\beta}_{0\perp} \beta_{0\perp}' U_{it} - \sum_{j=-1}^{k-1} \Psi_j \bar{\beta}_0 \beta_0' U_{jt} = f_m(L)(C_0 \varepsilon_t + \Delta^{b_0} Y_t),$$

and

$$f_m(L) = \Delta^{d-d_0}[\Delta^{kb} I_p - \sum_{i=m}^{k-1} \Psi_i \bar{\beta}_{0\perp}(\Delta^{ib} - \Delta^{kb})\beta_{0\perp}' - \sum_{j=-1}^{k-1} \Psi_j \bar{\beta}_0 (\Delta^{jb} - \Delta^{kb})\beta_0'].$$

The transfer function of the stationary linear process S_t is $f_m(z) f_0(z)^{-1}$, which has $f_m(0) f_0(0)^{-1} = I_p$, so that S_t is of the form $S_t = \varepsilon_t + \tau_1 \varepsilon_{t-1} + \ldots$. It follows that $Var(S_t) \geq \Omega_0$ and equality holds only for $S_t = \varepsilon_t$ or $f_m(z) = f_0(z)$, which implies that $(d, b) = (d_0, b_0)$. Note that $Var(S_t)$ is quadratic in the parameters $\{\Psi_i \bar{\beta}_{0\perp}\}_{i=m}^{k-1}, \{\Psi_j \bar{\beta}_0\}_{j=-1}^{k-1}$, and that minimizing over these, the residual variance satisfies the same inequality

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

where equality holds only for $\psi = \psi_0$.

This completes the proof of Theorem 4.

References

1. Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics* **22**, 327–351.
2. Avarucci, M. and Velasco, C. (2009). A Wald test for the cointegration rank in nonstationary fractional systems. *Journal of Econometrics* **151**, 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**, 187–212.
5. Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: representation, estimating and testing. *Econometrica* **55**, 251–276.
6. Granger, C. W. J. (1986). Developments in the study of cointegrated economic variables. *Oxford Bulletin of Economic and Statistics* **48**, 213–228.
7. Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and its Application*, Academic Press, New York.
8. 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, 417–438.
9. Hualde, J. and Robinson, P. M. (2010*a*). Semiparametric inference in multivariate fractionally cointegrated systems. *Journal of Econometrics* **157**, 492–511.
10. Hualde, J. and Robinson, P. M. (2010*b*). Gaussian pseudo-maximum likelihood estimation of fractional time series models. Working paper, Universidad de Navarra.
11. 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**, 111–137.
12. Jeganathan, P. (1999). On asymptotic inference in cointegrated time series with fractionally integrated errors. *Econometric Theory* **15**, 583–621.
13. Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* **12**, 231–254.
14. Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* **59**, 1551–1580.
15. Johansen, S. (1996). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, 2nd edition, Oxford University Press, Oxford.
16. Johansen, S. (2002). A small sample correction of the test for cointegrating rank in the vector autoregressive model. *Econometrica* **70**, 1929–1961.
17. Johansen, S. (2008). A representation theory for a class of vector autoregressive models for fractional processes. *Econometric Theory* **24**, 651–676.
18. Johansen, S. (2009). Representation of cointegrated autoregressive processes with application to fractional processes. *Econometric Reviews* **28**, 121–145.
19. Johansen, S. and Nielsen, M. Ø. (2010). Likelihood inference for a nonstationary fractional autoregressive mode. *Journal of Econometrics* **158**, 51–66.
20. Kallenberg, O. (2001). *Foundations of Modern Probability Theory*, 2nd edition, Springer, New York.
21. Lasak, K. (2008). Maximum likelihood estimation of fractionally cointegrated systems.

CREATES Research Paper 2008–53.

22. Lasak, K. (2010). Likelihood based testing for no fractional cointegration. *Journal of Econometrics* **158**, 67–77.
23. Lawley, D. N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* **43**, 296–303.
24. Lobato, I. N. and Velasco, C. (2007). Efficient Wald tests for fractional unit roots. *Econometrica* **75**, 575–589.
25. MacKinnon, J. G. and Nielsen, M. Ø. (2010). Numerical distribution functions of fractional unit root and cointegration tests. QED working paper 1240, Queen's University.
26. Magnus, J. R. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, revised edition, John Wiley and Sons, New York.
27. Marinucci, D. and Robinson, P. M. (2000). Weak convergence of multivariate fractional processes. *Stochastic Processes and their Applications* **86**, 103–120.
28. Marmol, F. and Velasco, C. (2004). Consistent testing of cointegrating relationships. *Econometrica* **72**, 1809–1844.
29. Phillips, P. C. B. (1991). Optimal inference in cointegrated systems. *Econometrica* **59**, 283–306.
30. Phillips, P. C. B. and Hansen, B. E. (1990). Statistical inference in instrumental variables regression with I(1) variables. *Review of Economic Studies* **57**, 99–125.
31. Robinson, P. M. and Hualde, J. (2003). Cointegration in fractional systems with unkown integration orders. *Econometrica* **71**, 1727–1766.
32. Saikkonen, P. (1991). Asymptotically efficient estimation of cointegration regressions. *Econometric Theory* **7**, 1–21.
33. Taqqu, M. S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **31**, 287–302.
34. van der Vaart, A. W. (1998). *Asymptotic Statistics*, Cambridge University Press, Cambridge.

Kommentarer

5. MN: Jeg har delt den meget lange Assumption 1 fra foer op i flere dele, ala artikel 1. Det goer det noget mere overskueligt. Det koster kun et par linjer, men hvis det bliver et problem, saa kan vi aendre det tilbage til sidst.

SJ: Du kan vel et eller andet med linie afstand så det ikke ser så diominerende ud?

9. MN: Jeg har aendret antagelsen saa roedderne skal vaere udenfor $\mathbb{C}_{b_0 \vee 1}$.

SJ: OK se formuleringen i Theorem 3

8. MN: Nu hvor vores assumptions er lidt mere overskuelige har jeg alligevel medtaget $d_0 - b_0 < 1/2$ her. Vi diskuterer den jo ogsaa i teksten umiddelbart efter antagelserne, saa man kan vel ikke paastaa at vi proever at gemme den vaek.

SJ: OK

7. MN: Vi skal bruge 8 momenter hele vejen igennem, saa det har jeg tilfoejet her.

SJ: OK

6. MN: Modellen $\mathcal{H}_r(d = d_0)$ er ikke naevnt i Assumption 1. Burde vi naevne den ogsaa? Vi naevner $\mathcal{H}_r(d = b)$ eksplicit, for da er antagelserne paa k, r lidt anderledes end for $\mathcal{H}_r$. Antagelserne for $\mathcal{H}_r(d = d_0)$ er praecis som for $\mathcal{H}_r$.

SJ: Jeg synes faktisk at hypotesen fylder lidt for meget. Kan vi ikke bare nævne i introduktionen at man selvfølgelig kan antage parameteren d kendt og får tilsvarende resultater, samt at man kan sætte konstanten til nul og får samme resultater undtagen at fBM ikke skal forlænges med $u^{-(d-b)}$. Jeg skal formulere et forslag.

Hypotesen $d = b$ kræver special behandling så den er anderledes.

11. MN: Skal vi have $\Gamma_k \neq 0, \bar{\Gamma}_k \neq 0$, som jeg har skrevet nu, eller er det ok med $(\Gamma_k, \bar{\Gamma}_k) \neq (0,0)$, som der stod foer? Eller er det det samme?

SJ: OBS Dette afsnit skal måske skrives om med henblik på bedømmer bemærkningerne. $(a,b) \neq (0,0)$ betyder geometrisk hele $\mathbb{R}^2$ uden $(0,0)$ men $a \neq 0$ og $b \neq 0$ betyder hele $\mathbb{R}^2$ uden de to koordinatakser.

13. MN: I beviset for saetningen har jeg ikke kunnet komme udenom at kraeve $q > 1/\eta_1$ momenter (udover $q > 2/\eta$). Det er lidt aergeligt, men jeg har proevet at reparere. Nu er resultatet (ii) selvfoelgelig ikke laengere saa staerkt som foer. Vi har stadig at det gaelder for en lidt stoerre maengde, men vi faar ikke laengere en paenere momentbetingelse der. Faktisk bliver den grimmere idet vi skal have $q > (1/2 - d_0 + b_0)^{-1}$, og dvs alle momenter.

SJ: Det har vi diskuteret, men jeg mener at $\eta_1 > 0$ er nødvendig for initial værdierne, på den måde det nu gøres, hvorimod η skal bruges til momenter.

14. MN: Her skal vaere et argument for at de andre parametre (fundet ved regression for $\mathcal{H}_p$) er konsistente. Det maa vaere noget med at de er kontinuerte funktioner af $\hat{\psi}$ og derfor er konsistente fordi $\hat{\psi}$ er konsistent.

SJ: Er indføjet

30. MN: Bemaerk at begge Lemmaer om M_T og Q_T kun har en-dimensionale ε_{it}. Skal der tilfoejes en kort kommentar om at de samme resultater holder multivariat, eller skal vi aendre dem til multivariate ε_{it}?

SJ: Vi har i begyndelsen af beviset for Lemma A. 10 en bemærkning om at det er nok med eendimensionale.

MN: Her har jeg forsoegt ved hjaelp af en kort kommentar at komme udenom at skulle bevise det hele for hver eneste delmodel. Er det godt nok?

SJ: Se andre kommentarer

MN: Blev det for kort?

SJ: Nej