QED

Queen's Economics Department Working Paper No. 1237

Likelihood inference for a fractionally cointegrated vector autoregressive model

SÃÿren Johansen
University of Copenhagen and CREATES

Morten ÃŸrregaard Nielsen
Queen‘s University and CREATES

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

5-2010

Likelihood inference for a fractionally cointegrated vector autoregressive model*

Søren Johansen†
University of Copenhagen
and CREATES

Morten Ørregaard Nielsen‡
Queen's University
and CREATES

April 26, 2012

Abstract

We consider model based inference in a fractionally cointegrated (or cofractional) vector autoregressive model, based on the Gaussian likelihood conditional on initial values. We give conditions on the parameters such that the process X_t 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. For $b=1$, the model nests the I$(d-1)$ VAR model. We define the statistical model by $0 < b \leq d$, but conduct inference when the true values satisfy $0 \leq d_0 - b_0 < 1/2$ and $b_0 \neq 1/2$, for which $\beta_0' X_t$ is (asymptotically) a stationary process. Our main technical contribution is the proof of consistency of the maximum likelihood estimators. To this end we prove weak convergence of the conditional likelihood as a continuous stochastic process in the parameters when errors are i.i.d. with suitable moment conditions and initial values are bounded. Because the limit is deterministic this implies uniform convergence in probability of the conditional likelihood function. If the true value $b_0 > 1/2$, we prove that the limit distribution of $T^{b_0}(\hat{\beta} - \beta_0)$ is mixed Gaussian and for the remaining parameters it is Gaussian. The limit distribution of the likelihood ratio test for cointegration rank is a functional of fractional Brownian motion of type II. If $b_0 < 1/2$ all limit distributions are Gaussian or chi-squared. We derive similar results for the model with $d = b$ allowing for a constant term.

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

JEL Classification: C32.

*We are very grateful to Jim Stock and five referees for many useful and constructive comments that were above and beyond the call of duty and led to significant improvements to the paper. We are also grateful to Uwe Hassler, James MacKinnon, Ilya Molchanov, and seminar participants at various universities and conferences for comments, and to the Danish Social Sciences Research Council (FSE grant 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 cointegrated vector autoregressive (VAR) model for a p-dimensional nonstationary time series, X_t, is

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

where $\Delta X_{t-i} = X_{t-i} - X_{t-i-1}$. This model has been widely used for analyzing long-run economic relations given by the stationary combinations $\beta' X_t$ and for building empirical dynamic models in macroeconomics and finance, see for instance Juselius (2006).

Fractional processes are a useful tool for describing time series with slowly decaying autocorrelation functions and have played a prominent role in econometrics, see e.g. Henry and Zaffaroni (2003) and Gil-Alana and Hualde (2009) for reviews and examples, and it appears important to allow fractional orders of integration (fractionality) in time series models.

In this paper we analyze VAR models for fractional processes. The models allow X_t to be fractional of order d and $\beta' X_t$ to be fractional of order $d - b \geq 0$, in order to extend the usefulness of model (1) to fractional processes. We also consider a model with $d = b$ allowing for a constant term.

The model can be derived in two steps. First, in (1) we replace the usual lag operator $L = 1 - \Delta$ and difference operator Δ by the fractional lag and difference operators, $L_b = 1 - \Delta^b$ and $\Delta^b = (1 - L)^b$ defined by the binomial expansion $\Delta^b Z_t = \sum_{n=0}^{\infty} (-1)^n \binom{b}{n} Z_{t-n}$. Secondly, we apply the resulting model to $Z_t = \Delta^{d-b} X_t$. This defines the fractional VAR model, $\text{VAR}_{d,b}(k)$, see Johansen (2008),

$$\mathcal{H}_r : \Delta^d X_t = \Delta^{d-b} L_b \alpha \beta' X_t + \sum_{i=1}^{k} \Gamma_i \Delta^d L_b^i X_t + \varepsilon_t,\ t = 1, \dots, T, \tag{2}$$

where ε_t is p-dimensional i.i.d.$(0, \Omega)$, Ω is positive definite, and α and β are $p \times r$, $0 \leq r \leq p$. The parameter space of $\mathcal{H}_r$ is given by the otherwise unrestricted parameters $\lambda = (d, b, \alpha, \beta, \Gamma_1, \dots, \Gamma_k, \Omega)$. In the special case $r = p$, the $p \times p$ matrix $\Pi = \alpha\beta'$ is unrestricted, and if $r = 0$ the parameters α and β are not present, and finally if $k = r = 0$ the model is $\Delta^d X_t = \varepsilon_t$, so the parameters are (d, Ω). Note that the VARFIMA$(k + 1, d - 1, 0)$ is a special case for $b = 1$.

If we model data Y_t by $Y_t = \mu + X_t$, where X_t is given by (2), then $\Delta^a Y_t = \Delta^a (X_t + \mu) = \Delta^a X_t$ because $\Delta^a 1 = 0$ for $a > 0$, so that Y_t satisfies the same equations. For the same reason, when $d > b$ the model (2) is invariant to a restricted constant term, ρ, when included in a way similar to that in (1). Thus (2) is a model for the stochastic properties of the data and when they have been determined one can, for example, estimate the mean of the stationary linear combinations by the average.

Therefore, we also consider the model with $d = b$ and a constant term,

$$\mathcal{H}_r(d = b) : \Delta^d X_t = \alpha L_d(\beta' X_t + \rho') + \sum_{i=1}^{k} \Gamma_i \Delta^d L_d^i X_t + \varepsilon_t,\ t = 1, \dots, T, \tag{3}$$

with a similar interpretation of $\beta' X_t$ except now $\beta' X_t + \rho'$ is a mean zero process of fractional order zero. Note that $L_d \rho' = \rho'$ because $\Delta^d 1 = 0$.

We show that when $0 < r < p$, X_t is fractional of order d and cofractional of order $d - b$, that is, $\beta' X_t$ is fractional of order $d - b$. Moreover, if $d - b < 1/2$ then $\beta' X_t$ in model (2) is asymptotically a mean zero stationary process. The model has the attractive feature of a straightforward interpretation of β as the cointegrating parameters in the long-run relations, $\beta' X_t = 0$, which are stable in the sense that they are fractional of a lower order, and of α describing adjustment towards the long-run equilibria and (through the orthogonal complement) the common stochastic trends, which are fractional of order d.

The lag structure of models (2) and (3) admits simple criteria for fractionality and cofractionality of X_t (or fractional cointegration; henceforth we use these terms synonymously). 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. Finally, an appealing feature of the model is that it gives the possibility of the usual misspecification tests based on estimated residuals, although of course the theory for these would need to be developed in the current setting.

The purpose of this paper is to conduct (quasi) Gaussian maximum likelihood inference in models (2) and (3), to show that the maximum likelihood estimator exists uniquely and is consistent, and to find the asymptotic distributions of maximum likelihood estimators and some likelihood ratio test statistics. 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 calculations of the likelihood function and the maximum likelihood estimator, we need $\Delta^a X_t$ for $a > 0$. Because we do not know all initial values we assume that we have observations of $X_t, t = -N_0 + 1, \ldots, T$, and define initial values $\tilde{X}_{-n} = X_{-n}, n = 0, \ldots, N_0 - 1$ and $\tilde{X}_{-n} = 0, n \geq N_0$, and base the calculations on these. Thus we set aside N_0 observations for initial values. For the asymptotic analysis we represent X_t by its past values and we make suitable assumptions about their behaviour. Apart from that we assume only that ε_t is i.i.d.$(0, \Omega)$ with suitable moments.

We treat (d, b) as parameters to be estimated jointly with the other parameters. Another possibility is to impose the restriction $d = d_0$ for some prespecified d_0, e.g. $d_0 = 1$, and $b = b_0$, where $b_0 = 1$ yields the VARFIMA$(k + 1, d - 1, 0)$, or I$(d - 1)$ VAR, model. We note here that the models with $d = d_0$ and/or $b = b_0$ are submodels in $\mathcal{H}_r$, and results for these models can be derived by the methods developed for the general model $\mathcal{H}_r$. The same holds for the restriction $d = b$ in model $\mathcal{H}_r(d = b)$, see (3), even though a simple modification is needed due to the constant term. The univariate version of model (2) with a unit root was analyzed by Johansen and Nielsen (2010), henceforth JN (2010), and we refer to that paper for some technical results.

The inspiration for model (2) comes from Granger (1986), who noted the special role of the fractional lag operator $L_b = 1 - \Delta^b$ and suggested the model

$$A^*(L)\Delta^d X_t = \Delta^{d-b} L_b \alpha \beta' X_{t-1} + d(L)\varepsilon_t, \tag{4}$$

see also Davidson (2002). One way to derive the main term of this model is to assume that we have linear combinations (γ, β) of rank p for which $\Delta^d \gamma' X_t$ and $\Delta^{d-b} \beta' X_t$ are I(0). Simple algebra shows that $\Delta^d X_t = \Delta^{d-b} L_b \alpha \beta' X_t + u_t$, where α is a function of γ and u_t is I(0), see Johansen (2008, p. 652) for details.

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 in (d, b) 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 terminology of Hualde and Robinson (2010). Specifically, we prove that for i.i.d. errors with sufficient moments finite, the estimated cointegration vectors are locally asymptotically mixed normal (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. Note that the parameter value $b_0 = 1/2$ is a singular point in the sense that inference is different for $b_0 < 1/2$ and $b_0 > 1/2$. Close to $b_0 = 1/2$ we need many observations for the asymptotic results to be useful, and a similar situation occurs when the true value of either α or β is close to a matrix with lower rank, see Elliott (1998).

Although such LAMN 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 (2010). Specifically, in a vector autoregressive context, but in a model with $d = 1$ and a different lag structure from ours, Lasak (2010) 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 parameters (d and b).

The rest of the paper is laid out as follows. In the next section we describe the solution of the fractionally cointegrated vector autoregressive model and its properties. In Section 3 we derive the likelihood function and estimators and show consistency. In Section 4 we find the asymptotic distribution of estimators, and in Section 5 that of the likelihood ratio test for cointegration rank. Section 6 concludes and technical material is presented in appendices.

A word on notation. We let $\mathbb{C}^p(\mathbb{K})$ denote the space of continuous p-vector-valued functions on a compact set $\mathbb{K} \subseteq \mathbb{R}^q$, i.e. continuous functions $f : \mathbb{K} \to \mathbb{R}^p$, and let $\mathbb{D}^p(\mathbb{K})$ denote the corresponding space of cadlag functions. When $p = 1$ the superscript is omitted. 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 $|A| = (\text{tr}(A'A))^{1/2}$ 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 operator Δ^d, a truncated version Δ^d_+, and calculation of $\Delta^d X_t$. We show how equation (2) 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 Theorem 2. We then give assumptions for the asymptotic analysis and discuss identification of parameters, and finally we briefly discuss initial values.

2.1 The fractional difference operator

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

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

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

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

provided the right-hand side exists. Note that $\Delta^{-a_1}\Delta^{-a_2} = \Delta^{-a_1-a_2}$ and the useful relation $\Delta^{-a_1}\pi_t(a_2) = \pi_t(a_1 + a_2)$, using that $\pi_t(a) = 0$ for $t < 0$. 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 $s \times p$ 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 almost surely 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 and is almost surely continuous in $a > 0$.

(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.5, 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 $|\mathsf{D}^m\Delta^a Z_t - \mathsf{D}^m\Delta^{\tilde{a}} Z_t| \leq c|a - \tilde{a}| \sum_{n=1}^{\infty}(1+\log n)^{m+1} n^{-\eta_1-1}|Z_{t-n}|$ 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 or $\tilde{a}$. 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

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

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 $\Delta^{-a}\varepsilon_t 1_{\{t\geq 1\}}$, where $1_{\{A\}}$ denotes the indicator function for the event A, and call this a "type II" process.

The idea of conditioning on initial values is used in the analysis of autoregressive models for nonstationary processes, and we modify the definition of a fractional process to take initial values into account.

Definition 1 *Let ε_t be i.i.d.$(0, \Omega)$ in p dimensions and consider $s \times p$ matrices ξ_n for which $\sum_{n=0}^{\infty} |\xi_n| < \infty$, and define $C(z) = \sum_{n=0}^{\infty} \xi_n z^n,\ |z| < 1$. Then the linear process $C(L)\varepsilon_t = \sum_{n=0}^{\infty} \xi_n \varepsilon_{t-n}$ is fractional of order 0 if $C(1) \neq 0$. A process X_t is fractional of order $d > 0$ (denoted $X_t \in \mathcal{F}(d)$) if $\Delta^d X_t$ is fractional of order zero, and X_t is cofractional with cofractionality vector β if $\beta' X_t$ is fractional of order $d - b \geq 0$ for some $b > 0$.*

The same definitions hold for any $d \in \mathbb{R}$ and $b > 0$ for the truncated linear process

$$C_+(L)\varepsilon_t + \omega_t = 1_{\{t\geq 1\}} \sum_{n=0}^{t-1} \xi_n \varepsilon_{t-n} + \omega_t, \tag{5}$$

where ω_t is a deterministic term.

The main result in Theorem 2 in Section 2.3 is the representation of the solution of equation (2) in terms of certain stationary processes, which we introduce next.

Definition 2 *We define the class $\mathcal{Z}_b$ as the set of multivariate linear stationary processes Z_t which can be represented as*

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

where $b > 0$ and ε_t is i.i.d.$(0,\Omega)$ and the coefficient matrices satisfy $\sum_{n=0}^{\infty} |\xi_n^*| < \infty$.
We also define the corresponding truncated process $Z_t^+ = \xi\varepsilon_t + \Delta_+^b \sum_{n=0}^{t-1} \xi_n^* \varepsilon_{t-n}$.

Definition 2 is a fractional version of the usual Beveridge-Nelson decomposition, where $\sum_{n=0}^{\infty} \xi_n \varepsilon_{t-n} = (\sum_{n=0}^{\infty} \xi_n)\varepsilon_t + \Delta \sum_{n=0}^{\infty} \xi_n^* \varepsilon_{t-n} \in \mathcal{Z}_1$.

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 $Z_t \in \mathcal{Z}_b, b > 0$, we have

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

where $\Gamma(a)$ is the gamma function and W denotes p-dimensional Brownian motion (BM) generated by ε_t. 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 a function space ($\mathbb{C}^p$ or $\mathbb{D}^p$), see Billingsley (1968) or Kallenberg (2001). The proof of (6) is given in JN (2010, Lemma D.2) for $Z_t \in \mathcal{Z}_b, b > 0$, see also Taqqu (1975) for $Z_t = \varepsilon_t$.

We also have under the same conditions on ε_t and for $Z_t \in \mathcal{Z}_b, b > 0$, that

$$T^{-a} \sum_{t=1}^{T} \Delta_+^{-a} L_a Z_t^+ \varepsilon_t' \overset{D}{\to} \int_0^1 W_{a-1} dW',\ a > 1/2, \tag{7}$$

where $\overset{D}{\to}$ denotes convergence in distribution on $\mathbb{R}^{p\times p}$. This result is proved in JN (2010, p. 65) for univariate processes building on the result of Jakubowski, Mémin, and Pages (1989) for the case $Z_t = \varepsilon_t$ and $L_a = L_1$. The same proof can be applied for processes in $\mathcal{Z}_b$.

2.2 Solution of fractional autoregressive equations

The properties of the solution of (2) are given by the properties of the polynomial

$$\Psi(y) = (1-y)I_p - \alpha\beta' y - \sum_{i=1}^{k} \Gamma_i (1-y) y^i = -\alpha\beta' y + (1-y) \sum_{i=0}^{k} \Psi_i (1-y)^i, \tag{8}$$

where the coefficients satisfy $\sum_{i=0}^{k} \Psi_i = I_p$, $\Psi_0 = I_p - \sum_{i=1}^{k} \Gamma_i$, and $\Psi_k = (-1)^{k+1}\Gamma_k$. Equation (2) can be written as $\Pi(L)X_t = \Delta^{d-b}\Psi(L_b)X_t = \varepsilon_t$, so that

$$\Pi(z) = (1-z)^{d-b}\Psi(1-(1-z)^b). \tag{9}$$

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 (2) and the criteria for fractionality of order

d and cofractionality of order $d-b$ can be found by analyzing the polynomial $\Psi(y)$, just as for the cointegrated VAR model.

We want to solve X_t as a function of initial values $X_{-n}, n = 0, 1, \ldots,$ and random shocks $\varepsilon_1, \ldots, \varepsilon_t$. A solution can be found using the two operators, see Johansen (2008),

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

for which $\Pi(L)X_t = \Pi_+(L)X_t + \Pi_-(L)X_t$. Here the operator $\Pi_+(L)$ is defined for any sequence as 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 that the initial values of X_t are bounded. Then the equations in model (2) can be expressed as $\varepsilon_t = \Pi(L)X_t = \Pi_+(L)X_t + \Pi_-(L)X_t$, and by applying $\Pi_+(L)^{-1}$ on both sides we find, for $t = 1, 2, \ldots$, that

$$X_t = \Pi_+(L)^{-1}\varepsilon_t - \Pi_+(L)^{-1}\Pi_-(L)X_t = \Pi_+(L)^{-1}\varepsilon_t + \mu_t. \tag{10}$$

The first term is the stochastic component generated by $\varepsilon_1, \ldots, \varepsilon_t$, and the second is a deterministic component generated by initial values. An example of the solution (10) is the well known result that $y_t = vy_{t-1} + \varepsilon_t$ has the solution $y_t = \sum_{i=0}^{t-1} v^i \varepsilon_{t-i} + v^t y_0$ for any v and $t = 1, \ldots, T$. When $d < 1/2$ we use a representation of the solution which explicitly contains the stationarity of X_t. In the simple example $y_t = vy_{t-1} + \varepsilon_t$ with $|v| < 1$ this corresponds to using the solution $y_t = \sum_{i=0}^{\infty} v^i \varepsilon_{t-i}$ for $t = 1, \ldots, T$.

2.3 Properties of the solution: representation theorem

The solution (10) of equation (2) is valid without any assumptions on the parameters. We next give results which guarantee that X_t is fractional of order d and cofractional from d to $d-b$, that is $\Delta^d X_t$ and $\Delta^{d-b}\beta' X_t$ are fractional of order zero. These results are given in terms of an explicit condition on the roots of the polynomial $\det(\Psi(y))$ and the set C_b, which is the image of the unit disk under the mapping $y = 1 - (1-z)^b$, see Johansen (2008, p. 660). Note that C_1 is the unit disk and that C_b is increasing in b.

The following result is Granger's Representation Theorem for the cofractional VAR models (2) and (3), see also 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 model. 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, which defines $\bar{\beta}$ and $\bar{\beta}_\perp$,

$$I_p = \beta(\beta'\beta)^{-1}\beta' + \beta_\perp(\beta_\perp'\beta_\perp)^{-1}\beta_\perp' = \beta\bar{\beta}' + \beta_\perp\bar{\beta}_\perp'. \tag{11}$$

Theorem 2 *Let* $\Pi(z) = (1-z)^{d-b}\Psi(1-(1-z)^b)$ *be given by (8) and (9) for any* $0 < b \leq d$ *and let* $y = 1-(1-z)^b$. *We assume that* α *and* β *have rank* $r \leq p$ *and that* $\det(\Psi(y)) = 0$ *implies that either* $y = 1$ *or* $y \notin \mathsf{C}_{\max(b,1)}$, *and we define* $\Gamma = I_p - \sum_{i=1}^k \Gamma_i$. *Then:*

(i) It holds that

$$(1-z)^d\Pi(z)^{-1} = C + (1-z)^b C^* + (1-z)^{2b} H^*(1-(1-z)^b) = C + (1-z)^b H(1-(1-z)^b), \tag{12}$$

if and only if $\det(\alpha_\perp'\Gamma\beta_\perp) \neq 0$, *where* $H^*(y)$ *is regular in a neighborhood of* $\mathsf{C}_{\max(b,1)}$,

$$C = \beta_\perp(\alpha_\perp'\Gamma\beta_\perp)^{-1}\alpha_\perp', \text{ and } \beta' C^* \alpha = -I_r. \tag{13}$$

For $F^*(z) = H^*(1-(1-z)^b) = \sum_{n=0}^\infty \tau_n^* z^n$ *and* $F(z) = H(1-(1-z)^b) = \sum_{n=0}^\infty \tau_n z^n, |z| < 1$, *we have*

$$\sum_{n=0}^{\infty} |\tau_n| < \infty \text{ and } \sum_{n=0}^{\infty} |\tau_n^*| < \infty. \tag{14}$$

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

$$X_t = C\Delta_+^{-d}\varepsilon_t + \Delta_+^{-(d-b)}Y_t^+ + \mu_t,\ t = 1, \ldots, T, \tag{15}$$

where $\mu_t = -\Pi_+(L)^{-1}\Pi_-(L)X_t$ depends on initial values of X_t and $Y_t = \sum_{n=0}^{\infty}\tau_n\varepsilon_{t-n} \in \mathcal{Z}_b$ is fractional of order zero with $\sum_{h=-\infty}^{\infty}|E(Y_tY_{t-h}')| < \infty$. In this case $\beta' X_t$ is asymptotically stationary with mean zero. The solution of (3) with $d = b$ and a constant term is represented as

$$X_t = C\Delta_+^{-d}\varepsilon_t + Y_t^+ + \mu_t + C^*\alpha\rho',\ t = 1, \ldots, T, \tag{16}$$

and $\beta' X_t + \rho'$ is asymptotically stationary with mean zero.

(iii) For $d < 1/2$ we represent the solutions of (2) and (3) as

$$X_t = C\Delta^{-d}\varepsilon_t + \Delta^{-(d-b)}Y_t,\ t = 1, \ldots, T, \tag{17}$$

$$X_t = C\Delta^{-d}\varepsilon_t + \Delta^{-(d-b)}Y_t + C^*\alpha\rho',\ t = 1, \ldots, T. \tag{18}$$

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

Proof. *Proof of (i)*: The proofs of (12) and (13) are given in Johansen (2008, Theorem 8 and 2009, Theorem 3). The condition $\det(\alpha_\perp'\Gamma\beta_\perp) \neq 0$ is necessary and sufficient for the representation of X_t as an $\mathcal{F}(d)$ variable, 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 cointegrated VAR context, see Johansen (2008, Theorem 9).

To prove (14), it is enough to prove it for τ_n^* because $\tau_n = \tau_n^* - \tau_{n-1}^*, n \geq 1$. We note that because $H^*(y) = \sum_{n=0}^{\infty}\tau_n^* y^n$ is regular in a neighborhood of C_b we can extend $H^*(1-(1-z)^b)$ by continuity to $|z| = 1$, and define the transfer function

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

We then apply the proof in JN (2010, Lemma 1), which shows that because $|\partial\phi(e^{\mathrm{i}\lambda})/\partial\lambda|$ is square integrable when $b > 1/2$, we have $\sum_{n=0}^{\infty}(\tau_n^* n)^2 < \infty$ and hence $\sum_{n=0}^{\infty}|\tau_n^*| < \infty$.

For $b \leq 1/2$ we need another proof. The assumption $y \notin \mathsf{C}_1$ implies that $H^*(y) = \sum_{k=0}^{\infty} h_k^* y^k$ is regular for $|y| < 1 + \delta$ for some $\delta > 0$, so that h_k^* decrease exponentially. From the expansion $1-(1-z)^b = \sum_{m=1}^{\infty} b_m z^m$ with $b_m = -\pi_m(-b)$, we find that if $0 \leq b \leq 1/2$ then $b_m \geq 0$ and $\sum_{m=1}^{\infty} b_m = 1$. Therefore

$$H^*(1-(1-z)^b) = \sum_{k=0}^{\infty} h_k^* (\sum_{m=1}^{\infty} b_m z^m)^k = \sum_{n=0}^{\infty}\sum_{k=0}^{\infty} h_k^* (\sum_{m_1+\cdots+m_k=n}\prod_{i=1}^{k} b_{m_i}) z^n = \sum_{n=0}^{\infty}\tau_n^* z^n,$$

so that τ_n^* satisfies

$$\sum_{n=0}^{\infty}|\tau_n^*| \leq \sum_{n=0}^{\infty}\sum_{k=0}^{\infty}|h_k^*|(\sum_{m_1+\cdots+m_k=n}\prod_{i=1}^{k} b_{m_i}) \leq \sum_{k=0}^{\infty}|h_k^*|(\sum_{m=1}^{\infty} b_m)^k = \sum_{k=0}^{\infty}|h_k^*| < \infty.$$

Proof of (ii): For $d \geq 1/2$ we define $Y_t = \sum_{n=0}^{\infty}\tau_n\varepsilon_{t-n} = C^*\varepsilon_t + \Delta^b\sum_{n=0}^{\infty}\tau_n^*\varepsilon_{t-n} \in \mathcal{Z}_b$ which is fractional of order zero because $C^* \neq 0$, see (13), and has $\sum_{n=0}^{\infty}|\tau_n| < \infty$ which implies $\sum_{h=-\infty}^{\infty}|E(Y_tY_{t-h}')| < \infty$. Then (15) follows from (10), see also Johansen (2008, Theorem 8). For $\rho = 0$ and $d = b$ we find the solution X_t^0, say, from (15). Then $\Pi(L)(X_t^0 + C^*\alpha\rho') = \varepsilon_t + \Pi(L)C^*\alpha\rho' = \varepsilon_t - \alpha\beta' C^*\alpha\rho' = \varepsilon_t + \alpha\rho'$ so that $X_t = X_t^0 + C^*\alpha\rho'$ is a solution of (3). In this case we therefore find $\beta' X_t + \rho' = \beta' X_t^0 + \beta' C^*\alpha\rho' + \rho' = \beta' X_t^0$, which is asymptotically stationary with mean zero and fractional of order zero.

Proof of (iii): For $d < 1/2$, $C\Delta^{-d}\varepsilon_t + \Delta^{-(d-b)}Y_t$ is stationary and represents a solution of (2) and (3) for $\rho = 0$. We then add $C^*\alpha\rho'$ for $\rho \neq 0$.

Proof of (iv): We find from *(i)* and *(ii)* that if, for some $c < d - b$, $\gamma' X_t$ is fractional of order c then

$$\gamma' X_t = \gamma' C \Delta_+^{-d}\varepsilon_t + \gamma' C^* \Delta_+^{-d+b}\varepsilon_t + \Delta^{-d+2b}\sum_{n=0}^{t-1}\gamma'\tau_n^*\varepsilon_{t-n} + \gamma'\mu_t \in \mathcal{F}(c)$$

implies that $\gamma' C = 0$ and $\gamma' C^* = 0$. Hence $\gamma = \beta\xi$ and therefore $\gamma' C^*\alpha = \xi'\beta' C^*\alpha = -\xi' = 0$, so that $\gamma = 0$. ■

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

If $r = 0$ we have $\alpha = \beta = \rho = 0$, $\alpha_\perp = \beta_\perp = I_p$, 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 $X_t = \Delta_+^{-(d-b)}Y_t^+ + \mu_t$ (the $d \geq 1/2$ representation) is fractional of order $d - b$. Note, however, that the coefficients of Y_t^+ and μ_t depend on both d and b, so that (d, b) is identified, see Theorem 3.

The stochastic properties of X_t are given in Theorem 2 in terms of the process $U_t = C\varepsilon_t + \Delta^b Y_t \in \mathcal{Z}_b$, see Definition 2, and it follows from Theorem 2 that also $Y_t \in \mathcal{Z}_b$.

2.4 Assumptions for the data generating process

We here formulate assumptions on the true parameter $\lambda_0 = (d_0, b_0, \alpha_0, \beta_0, \Gamma_{01}, \ldots, \Gamma_{0k}, \Omega_0)$ needed for identification and for the asymptotic properties of the estimators and the likelihood function for model $\mathcal{H}_r$. For the model $\mathcal{H}_r(d = b)$ with $d = b$ and a constant term, i.e. (3), we replace b with ρ in the definition of λ. We define the parameter set

$$\mathcal{N} = \{d, b : 0 < b \leq d \leq d_1\} \tag{19}$$

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

Assumption 1 *For $k \geq 0$ and $0 \leq r \leq p$ the process X_t, $t = 1, \ldots, T$, is generated by model $\mathcal{H}_r$ in (2) or model $\mathcal{H}_r(d = b)$ in (3) with the parameter value λ_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 uniformly bounded, and $\tilde{X}_{-n} = X_{-n}$ for $n < N_0$ and $\tilde{X}_{-n} = 0$ for $n \geq N_0$.*

Assumption 4 *The true parameter value λ_0 satisfies $(d_0, b_0) \in \mathcal{N}$, $0 \leq d_0 - b_0 < 1/2$, $b_0 \neq 1/2$, and the identification conditions $\Gamma_{0k} \neq 0$ (if $k > 0$), α_0 and β_0 are $p \times r$ of rank r, $\alpha_0\beta_0' \neq -I_p$, and $\det(\alpha_{0\perp}'\Gamma_0\beta_{0\perp}) \neq 0$. Thus, if $r < p$, then $\det(\Psi(y)) = 0$ has $p - r$ unit roots and the remaining roots are outside $\mathsf{C}_{\max(b_0,1)}$. If $k = r = 0$ only $0 < d_0 \neq 1/2$ 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 8 moments, and we later specify the existence of further moments needed for the asymptotic properties of the maximum likelihood estimator. Assumption 3 about initial values is needed for nonstationary processes 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 $0 \leq d_0 - b_0 < 1/2$, which appears to be perhaps the most empirically relevant range of values for $d_0 - b_0$, see e.g. Henry and Zaffaroni (2003), Gil-Alana and Hualde (2009), and the references in the introduction, because in this case $\beta_0' X_t$ is (asymptotically) stationary with mean zero. 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 identification conditions in Assumption 4 guarantee that the lag length is well defined, that the parameters are identified, see Section 2.5, and that the asymptotic distribution of the maximum likelihood estimator is nonsingular, see Lemma 7.

2.5 Identification of parameters

In a statistical model with parameter λ we say that *the parameter value* λ_0 is identified if, for all λ for which $P_\lambda = P_{\lambda_0}$, it holds that $\lambda = \lambda_0$. We say that *the model* is generically identified if the set of unidentified parameter values has Lebesgue measure zero. In model (2) the parameters α and β enter, when $r > 0$, only through their product $\alpha\beta'$ so they are not individually identified. This is usually solved by normalizing β. We use the decomposition (11) and define $\tilde{\beta} = \beta(\bar{\beta}_0'\beta)^{-1}, \tilde{\alpha} = \alpha\beta'\bar{\beta}_0, \tilde{\rho} = \rho(\bar{\beta}_0'\beta)^{-1}$, so that $\alpha\beta' = \tilde{\alpha}\tilde{\beta}'$. We assume in the following that this normalization has been performed and use the notation α, β. Note that $\beta'\bar{\beta}_0 = I_r$. We define $\lambda = (d, b, \alpha, \beta, \Gamma_1, \ldots, \Gamma_k, \Omega)$ suitably modified if $r = p$, $r = 0$, or $k = 0$, see the discussion after (2), and apply the notation $\Pi_\lambda(L)$.

Theorem 3 *For any $k \geq 0$ and $0 \leq r \leq p$ we let λ denote all parameters of model $\mathcal{H}_r$ with k lags, see (2). We assume, see Assumption 4, that for λ and λ_0 it holds that $\Gamma_k \neq 0$ (if $k > 0$), α and β are $p \times r$ of rank r, $\alpha\beta' \neq -I_p$, and $\det(\alpha_\perp' \Gamma \beta_\perp) \neq 0$. Then $P_\lambda = P_{\lambda_0}$ implies $\lambda = \lambda_0$ so that λ_0 is identified. It follows that model $\mathcal{H}_r$ in (2) is generically identified. A similar result holds for model (3).*

Proof. If $P_\lambda = P_{\lambda_0}$ the mean and variance of X_t given the past are the same with respect to P_λ and P_{λ_0}, so that $\Omega = \Omega_0$, and, for all z,

$$\Pi_\lambda(z) = (1-z)^{d-b}\Psi_\lambda(1-(1-z)^b) = (1-z)^{d_0-b_0}\Psi_{\lambda_0}(1-(1-z)^{b_0}) = \Pi_{\lambda_0}(z). \tag{20}$$

If $k > 0$ and $r > 0$ then $\Psi_\lambda(1-(1-z)^b)$ is a polynomial in $(1-z)^b$, see (8), with highest order term $\Psi_k(1-z)^{(k+1)b}$ and lowest order term $-\alpha\beta'$. Hence (20) implies that $(1-z)^{d-b}\Psi_k(1-z)^{(k+1)b} = (1-z)^{d_0-b_0}\Psi_{0k}(1-z)^{(k+1)b_0} \neq 0$ and $(1-z)^{d-b}\alpha\beta' = (1-z)^{d_0-b_0}\alpha_0\beta_0' \neq 0$. This evidently implies that $(d, b) = (d_0, b_0)$ and therefore $\alpha\beta' = \alpha_0\beta_0'$ and that $\Psi_\lambda(y)$ and $\Psi_{\lambda_0}(y)$ have the same coefficients; that is $\lambda = \lambda_0$. If $k > 0$ and $r = 0$, then $\alpha = \beta = 0$ and $\alpha_\perp = \beta_\perp = I_p$ and $\Psi_0 = I_p - \sum_{i=1}^k \Gamma_i = \Gamma = \alpha_\perp'\Gamma\beta_\perp \neq 0$ and the same conclusion holds. In case $k = 0$ and $r > 0$, where the model is $\Delta^d X_t = \Delta^{d-b} L_b \alpha\beta' X_t + \varepsilon_t$, the conditions $\alpha\beta' \neq 0$ and $\alpha\beta' \neq -I_p$ for λ and λ_0 imply that λ_0 is identified. Finally, if $k = r = 0$ the model is $\Delta^d X_t = \varepsilon_t$ and $\lambda_0 = (d_0, \Omega_0)$ is identified.

Since the set of values of λ_0 that do not satisfy the given conditions has Lebesgue measure zero, it follows that model (2) is generically identified. ■

Identification was discussed in JN (2010, Section 2.3, Lemma 3 and Corollary 4) in the univariate case, and an example of an indeterminacy between d, b, and k was given. Theorem 3 shows that once the lag length has been determined the model is generically identified.

2.6 Initial values

In order for $\Delta^a X_t, a > 0$, to be well defined we assume that the initial values $X_{-n}, n \geq 0$, are uniformly bounded. The theory in this paper will be developed for observations $X_1, \ldots, X_T$ generated by (2) or (3) with fixed bounded initial values; that is, conditional on $X_{-n}, n \geq 0$, as developed in JN (2010), and we choose the representations given in Theorem 2.

The likelihood function depends on $\Delta^a X_t$ for different values of a and because we do not observe the infinitely many past values of X_t we choose initial values, $\tilde{X}_{-n}$, for the calculations and define $\tilde{\Delta}^a X_t = \Delta_+^a X_t + \Delta_-^a \tilde{X}_t$. The first term is a function of the observations $X_1, \ldots, X_T$, but the second is a function of initial values. A possible choice is $\tilde{X}_{-n} = 0, n \geq 0$, but we derive the theory for the choice $\tilde{X}_{-n} = X_{-n}, n < N_0$ and set $\tilde{X}_{-n} = 0$ for $n \geq N_0$. Thus we set aside N_0 observations for initial values, as is usually done in the analysis of an AR(k) model.

We prove consistency under the assumption that X_{-n} is uniformly bounded for $n \geq 0$, and derive the asymptotic distributions under the further assumption that $X_{-n} = 0$ for $n \geq T^{\upsilon}$ for a small υ.[1] In this way we allow the number of initial values in the representation of X_t to increase with T, thereby approximating the situation where the representation has infinitely many initial values.

The choice of N_0 entails a small sample bias/efficiency trade-off, with fewer initial values introducing bias, but also leaving more observations for parameter estimation. Simulations suggest that many initial values are needed if b_0 is close to 1/2, but for, say, $b_0 \geq 0.8$ about a handful of initial values are sufficient, which is also what is used in the (univariate) empirical application in Hualde and Robinson (2011, Section 5) who assume that both X_{-n} and $\tilde{X}_{-n}$ are zero in their theoretical analysis, but in their empirical application they actually condition on non-zero initial values. Such simulations and analytical results will be reported elsewhere.

For $d_0 \geq 1/2$ we use the representations (15) and (16) in terms of μ_{0t} which depends on the correct initial values, and approximate it as discussed above, and for $d_0 < 1/2$ we use the representations (17) and (18) of X_t as a stationary process around its mean. The initial values term μ_{0t} plays no role in that case because the initial values have been given their invariant distribution.

3 Likelihood function and maximum likelihood estimators

The log likelihood function $\log L_T(\lambda)$ is continuous in λ and we show that for the probability measure P determined by λ_0, $T^{-1} \log L_T(\lambda)$ converges as a continuous function on a compact set. Because the limit is deterministic we get uniform continuity in the parameter λ, and we use that to prove existence and uniqueness of the maximum likelihood estimator (MLE). We first discuss the calculation of the MLE and then find the likelihood and profile likelihood functions and their limits. We apply this to prove consistency of the MLE.

3.1 Calculation of MLE, profile likelihood function, and its limit

In (8) we eliminate $\Psi_k = I_p - \sum_{i=0}^{k-1} \Psi_i$ and define $\tilde{\Delta}^{d+ib} X_t = \Delta_+^{d+ib} X_t + \Delta_-^{d+ib} \tilde{X}_t$, the regressors

$$X_{-1,t} = (\tilde{\Delta}^{d-b} - \tilde{\Delta}^{d}) X_t, \; X_{kt} = \tilde{\Delta}^{d+kb} X_t, \; X_{it} = (\tilde{\Delta}^{d+ib} - \tilde{\Delta}^{d+kb}) X_t, \tag{21}$$

for $i = 0, \ldots, k-1$, and the residuals

$$\varepsilon_t(\lambda) = \Pi_+(L) X_t + \Pi_{0-}(L) \tilde{X}_t = X_{kt} - \alpha\beta' X_{-1,t} + \sum_{i=0}^{k-1} \Psi_i X_{it}, \tag{22}$$

where $\lambda = (d, b, \alpha, \beta, \Psi_*, \Omega)$ is freely varying and $\Psi_* = (\Psi_0, \ldots, \Psi_{k-1})$. The Gaussian likelihood function is now

$$-2T^{-1} \log L_T(\lambda) = \log\det(\Omega) + \mathrm{tr}(\Omega^{-1} T^{-1} \sum_{t=1}^{T} \varepsilon_t(\lambda)\varepsilon_t(\lambda)'). \tag{23}$$

[1] An alternative assumption is $\sum_{n=1}^{\infty} n^{-1/2} |X_{-n}| < \infty$, see Lemma A.8.

For the model with $d = b$ we define $X_{-1,t} = (1 - \tilde{\Delta}^d)(X_t - C_0^* \alpha_0 \rho_0')$ and $\theta_\rho' = \rho' + \beta' C_0^* \alpha_0 \rho_0'$ so that $(1 - \tilde{\Delta}^d)(\beta' X_t + \rho') = \beta' X_{-1,t} + \theta_\rho'$ and

$$\varepsilon_t(\lambda) = X_{kt} - \alpha\beta' X_{-1,t} - \alpha\theta_\rho' + \sum_{i=0}^{k-1} \Psi_i X_{it}. \tag{24}$$

Note that for $(\rho, \beta) = (\rho_0, \beta_0)$ we find $\theta_\rho = 0$ because $\beta_0' C_0^* \alpha_0 = -I_r$, see (13).

For fixed $\psi = (d, b)$ the MLE based on (23) is found by reduced rank regression of X_{kt} on $X_{-1,t}$ corrected for $\{X_{it}\}_{i=0}^{k-1}$, see Anderson (1951) or Johansen (1996). Note that this is equivalent to reduced rank regression of $\Delta^d X_t$ on $\Delta^{d-b} L_b X_t$ corrected for $\{\Delta^d L_b^i X_t\}_{i=1}^k$. The calculations are organized as follows. For fixed ψ in model $\mathcal{H}_r$ we define in analogy with the notation for the I(1) model, see Johansen (1996, pages 91-92), the residuals

$$R_{0t}(\psi) = (X_{kt} | X_{0t}, \ldots, X_{k-1,t}) \text{ and } R_{1t}(\psi) = (X_{-1,t} | X_{0t}, \ldots, X_{k-1,t})$$

from regressions of X_{kt} and $X_{-1,t}$ on $X_{0t}, \ldots, X_{k-1,t}$, respectively. We then define the product moments $S_{ij}(\psi) = T^{-1} \sum_{t=1}^T R_{it}(\psi) R_{jt}'(\psi)$ and the eigenvalue problem

$$0 = \det(\omega S_{11}(\psi) - S_{10}(\psi) S_{00}^{-1}(\psi) S_{01}(\psi)), \tag{25}$$

which gives eigenvalues $1 > \hat{\omega}_1(\psi) > \cdots > \hat{\omega}_p(\psi) > 0$ and the maximized profile likelihood function expressed as

$$\ell_{T,r}(\psi) = -2T^{-1} \log L_{\max}(\mathcal{H}_r) = \log\det(S_{00}(\psi)) + \sum_{i=1}^r \log(1 - \hat{\omega}_i(\psi)). \tag{26}$$

Finally the MLE and maximized likelihood can be calculated by minimizing $\ell_{T,r}(\psi)$ as a function of $\psi = (d, b)$ by a numerical optimization procedure.

For model (3) we assume $b = d$ and include $-\alpha\rho'$ in the definition of $\varepsilon_t(\lambda)$, see (24), and apply reduced rank regression of X_{kt} on $(X_{-1,t}', 1)$ corrected for $\{X_{it}\}_{i=0}^{k-1}$ to define the concentrated likelihood function $\ell_{T,r}(\psi)$. Below we focus on (2) and only include comments on (3) when the results or arguments are different.

A computer package for conducting statistical inference using the procedure described in this paper is available, see Nielsen and Morin (2012).

Using non- or semi-parametric estimates of d and b, followed by reduced rank regression estimation of the remaining parameters, would entail an efficiency loss for the asymptotically normal estimators, i.e. $(\hat{\alpha}, \hat{\Gamma}_1, \ldots, \hat{\Gamma}_k)$ when $b_0 > 1/2$ and all the estimators when $b_0 < 1/2$, because $\hat{d}$ and $\hat{b}$ are asymptotically correlated with those, but no efficiency loss for $\hat{\beta}$ when $b_0 > 1/2$. In addition, we have found that using $d = b = 1$ as starting values in the numerical iterations is a good choice, so there seems to be no advantage from initializing the search with preliminary estimates. The calculation of the fractional differences in $\{X_{it}\}_{i=-1}^k$ in each step of the numerical optimization algorithm can be time consuming for very large samples, but the actual optimization of $\ell_{T,r}(\psi)$ seems to be unproblematic.

Note that for $r = p$, $\ell_{T,p}(\psi)$ is found by regression of X_{kt} on $\{X_{it}\}_{i=-1}^{k-1}$

$$\ell_{T,p}(\psi) = \log\det(SSR_T(\psi)) = \log\det(T^{-1} \sum_{t=1}^T R_t R_t'), \tag{27}$$

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

The stochastic properties of X_t are given in Theorem 2 in terms of the stationary process $U_t = C_0 \varepsilon_t + \Delta^{b_0} Y_t$. We note that, for any $\psi = (d, b)$ for which $d + ib - d_0 > -1/2$, the process $\Delta^{d+ib}\Delta^{-d_0}U_t$ is stationary. On the other hand, $\Delta^{d+jb}\Delta^{-d_0}\beta_0'U_t = \Delta^{d+jb}\Delta^{-d_0+b_0}\beta_0'Y_t$ is stationary for all $j = -1, 0, \ldots, k$ because $d + jb - d_0 + b_0 \geq -d_0 + b_0 > -1/2$. Thus corresponding to X_{it}, see (21), we define

$$U_{-1,t} = \Delta^{d-b-d_0} L_b U_t,\ U_{kt} = \Delta^{d+kb-d_0} U_t,\ U_{it} = (\Delta^{d+ib} - \Delta^{d+kb})\Delta^{-d_0} U_t, \tag{28}$$

if they are stationary, and the class of stationary processes for a given ψ,

$$\mathcal{F}_{\text{stat}}(\psi) = \{\beta_0' U_{jt} \text{ for all } j, \text{ and } U_{it} \text{ for } d + ib - d_0 > -1/2\}.$$

For $d_0 < 1/2$, $d + ib - d_0 \geq -d_0 > -1/2$ so in that case $\mathcal{F}_{\text{stat}}(\psi)$ contains U_{it} for all i.

We next want to define the probability limit, $\ell_p(\psi)$, of the profile likelihood function $\ell_{T,p}(\psi)$ in (27). The limit of $\log\det(SSR_T(\psi))$ is infinite if X_{kt} is nonstationary and finite if X_{kt} is (asymptotically) stationary, see Theorem 4. We therefore define the subsets of $\mathcal{N}$,

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

and note that $\mathcal{N} = \mathcal{N}_{\text{div}}(\kappa) \cup \mathcal{N}_{\text{conv}}(\kappa)$ for all $\kappa \geq 0$. The family of sets $\mathcal{N}_{\text{div}}(\kappa)$ decreases (as $\kappa \to 0$) to the set $\mathcal{N}_{\text{div}}(0)$, which is exactly the set where X_{kt} is nonstationary and $\log\det(SSR_T(\psi))$ diverges. Similarly, $\mathcal{N}_{\text{conv}}(\kappa)$ is a family of sets increasing (as $\kappa \to 0$) to $\mathcal{N}_{\text{conv}}(0)$, which is the set where X_{kt} is stationary and $\log\det(SSR_T(\psi))$ converges pointwise in ψ in probability. We therefore define the limit likelihood function, $\ell_p(\psi)$, as

$$\ell_p(\psi) = \begin{cases} \infty & \text{if } \psi \in \mathcal{N}_{\text{div}}(0), \\ \log\det(Var(U_{kt}|\mathcal{F}_{\text{stat}}(\psi))) & \text{if } \psi \in \mathcal{N}_{\text{conv}}(0), \end{cases} \tag{29}$$

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

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

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

For $\eta > 0$ we define the family of compact sets,

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

which has the property that $\mathcal{K}(\eta) \subset \mathcal{N}$ increases to $\mathcal{N}$ as $\eta \to 0$.

We now show that for all $A > 0$ and all $\gamma > 0$ there exists a $\kappa_0 > 0$ and $T_0 > 0$ so that with probability larger than $1 - \gamma$, the profile likelihood $\ell_{T,p}(\psi)$ is uniformly larger than A on $\mathcal{K}(\eta) \cap \mathcal{N}_{\text{div}}(\kappa_0)$ for $T \geq T_0$. Thus the minimum of $\ell_{T,p}(\psi)$ cannot be attained on $\mathcal{K}(\eta) \cap \mathcal{N}_{\text{div}}(\kappa_0)$. On the rest of $\mathcal{K}(\eta)$, however, we show that $\ell_{T,p}(\psi)$ converges uniformly in probability as $T \to \infty$ to the deterministic limit $\ell_p(\psi)$ which has a strict minimum, $\log\det(\Omega_0)$, at ψ_0. We prove this by showing weak convergence, on a compact set, of the likelihood as a continuous process in the parameters. Because the limit is deterministic, weak convergence implies uniform convergence in probability, see Lemma A.4.

Theorem 4 *The function $\ell_p(\psi)$ has a strict minimum at $\psi = \psi_0$, that is*

$$\ell_p(\psi) \geq \ell_p(\psi_0) = \log\det(\Omega_0), \psi \in \mathcal{N}, \tag{30}$$

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

Let Assumptions 1-4 hold, so that in particular $E|\varepsilon_t|^8 < \infty$, and assume that $(d_0, b_0) \in \mathcal{K}(\eta)$. For $r = 0, \ldots, p$ it holds that

$$\ell_{T,r}(\psi_0) \overset{P}{\to} \log\det(\Omega_0), \tag{31}$$

and furthermore:

(i) *Suppose* $E|\varepsilon_t|^q < \infty$ *for some* $q > 1/\min(\eta/3, (1/2 - d_0 + b_0)/2)$*. Then the likelihood function for* $\mathcal{H}_p$ *satisfies that, for any* $A > 0$ *and* $\gamma > 0$*, there exists a* $\kappa_0 > 0$ *and a* $T_0 > 0$ *such that*

$$P(\inf_{\psi \in \mathcal{N}_{div}(\kappa_0) \cap \mathcal{K}(\eta)} \ell_{T,p}(\psi) \geq A) \geq 1 - \gamma \tag{32}$$

for all $T \geq T_0$*. It also holds that*

$$\ell_{T,p}(\psi) \Longrightarrow \ell_p(\psi) \text{ on } \mathbb{C}(\mathcal{N}_{conv}(\kappa_0) \cap \mathcal{K}(\eta)) \text{ as } T \to \infty. \tag{33}$$

(ii) *Suppose* $\eta \leq b_0 = d_0 \leq d_1$ *and* $E|\varepsilon_t|^q < \infty$ *for* $q > 3/\eta$*. Then, for model* $\mathcal{H}_p(d = b)$ *with a constant, the results (32) and (33) hold on the respective sets intersected with* $\{b = d\}$*.*

The proof is given in Appendix B. Note that, in general, the larger the compact set $\mathcal{K}(\eta)$ the more moments are needed. When consideration is restricted to the model $\mathcal{H}_r(d = b)$ and a parameter set defined by $\eta > 3/8$, i.e. in particular if consideration is restricted to the case of "strong cointegration" where $b_0 > 1/2$, then the moment condition reduces to $E|\varepsilon_t|^8 < \infty$ (from Assumption 2).

We now derive the important consequence of Theorem 4.

Theorem 5 *Let the assumptions of Theorem 4 be satisfied and let* $\hat{\lambda}$ *denote the MLE in model* $\mathcal{H}_r$ *respectively model* $\mathcal{H}_r(d = b)$*. Corresponding to Theorem 4(i)–(ii) we have:*

(i) *With probability converging to one,* $\hat{\lambda}$ *in model* $\mathcal{H}_r, r = 0, \ldots, p$*, exists uniquely for* $\psi \in \mathcal{K}(\eta), \eta > 0$*, and is consistent.*

(ii) *For model* $\mathcal{H}_r(d = b)$ *with a constant, existence, uniqueness, and consistency of* $\hat{\lambda}$ *hold for* $d \in \{d : 0 < \eta \leq d \leq d_1\}$*.*

Proof. To prove existence and consistency of the MLE we define the open neighborhood $\mathcal{N}(\psi_0, \epsilon) = \{\psi : |\psi - \psi_0| < \epsilon\}$, and want to find a set A_T with $P(A_T) \geq 1 - 2\gamma$ so that $\hat{\psi}$ exists on A_T and

$$P(\hat{\psi} \in A_T \cap \mathcal{N}(\psi_0, \epsilon)) \geq 1 - 3\gamma.$$

We first analyze model $\mathcal{H}_p$, see (2), where α and β are $p \times p$. For any $\gamma > 0$, (32) shows that we can find $\kappa_0 = \kappa_0(\gamma)$ and $T_0 = T_0(\gamma)$ and define $A_{1T} = \{\inf_{\psi \in \mathcal{N}_{\text{div}}(\kappa_0) \cap \mathcal{K}(\eta)} \ell_{T,p}(\psi) \geq 2 + \ell_p(\psi_0)\}$ so that $P(A_{1T}) \geq 1 - \gamma$ for all $T \geq T_0$.

We find from (33) that $\ell_{T,p}(\psi) = \log\det(SSR_T(\psi)) \Longrightarrow \ell_p(\psi)$ on the compact set $\mathcal{N}_0 = \mathcal{N}_{\text{conv}}(\kappa_0) \cap \mathcal{K}(\eta)$ so that $\ell_p(\psi)$ is continuous on $\mathcal{N}_0$. Because $\ell_p(\psi)$ is continuous and $> \ell_p(\psi_0)$ if $\psi \neq \psi_0$, see (30), and $\mathcal{N}_0 \backslash \mathcal{N}(\psi_0, \epsilon)$ is compact and does not contain ψ_0, we have $\min_{\psi \in \mathcal{N}_0 \backslash \mathcal{N}(\psi_0, \epsilon)} \ell_p(\psi) \geq \ell_p(\psi_0) + 3c_0$ for some $c_0 > 0$. By the uniform convergence of $\ell_{T,p}(\psi)$ to $\ell_p(\psi)$ on $\mathcal{N}_0$, see (33), we can find $T_1 = T_1(\gamma)$ and define $A_{2T} = \{\min_{\psi \in \mathcal{N}_0 \backslash \mathcal{N}(\psi_0, \epsilon)} |\ell_{T,p}(\psi) - \ell_p(\psi)| \leq c_0\}$ such that $P(A_{2T}) \geq 1 - \gamma$ for all $T \geq T_1$.

We now turn to the model $\mathcal{H}_r, r = 0, \ldots, p$. On the set A_{2T} we have for any $r \leq p$,

$$\min_{\psi \in \mathcal{N}_0 \backslash \mathcal{N}(\psi_0, \epsilon)} \ell_{T,r}(\psi) \geq \min_{\psi \in \mathcal{N}_0 \backslash \mathcal{N}(\psi_0, \epsilon)} \ell_{T,p}(\psi) \geq \min_{\psi \in \mathcal{N}_0 \backslash \mathcal{N}(\psi_0, \epsilon)} \ell_p(\psi) - c_0,$$

which is bounded below by $\ell_p(\psi_0) + 3c_0 - c_0 = \ell_r(\psi_0) + 2c_0$, recalling $\ell_r(\psi_0) = \log\det(\Omega_0) = \ell_p(\psi_0)$, see (30). On the set A_{1T} we have $\ell_{T,r}(\psi) \geq \ell_{T,p}(\psi) \geq 2 + \ell_p(\psi_0)$ and it follows that on $A_T = A_{1T} \cap A_{2T}$ with $P(A_T) \geq 1 - 2\gamma$,

$$\min_{\psi \in \mathcal{K}(\eta) \backslash \mathcal{N}(\psi_0, \epsilon)} \ell_{T,r}(\psi) \geq \ell_r(\psi_0) + 2\min(1, c_0).$$

On the other hand, at the point $\psi = \psi_0$ we have $\ell_{T,r}(\psi_0) \overset{P}{\to} \ell_r(\psi_0) = \log\det(\Omega_0)$, see (31), so that for all $T \geq T_2 = T_2(\gamma)$,

$$P(|\ell_{T,r}(\psi_0) - \ell_r(\psi_0)| \leq \min(c_0, 1)) \geq 1 - \gamma,$$

which implies that, on A_T, the minimum of $\ell_{T,r}(\psi)$ is attained inside $\mathcal{N}(\psi_0, \epsilon)$. Thus the MLE, $\hat{\psi}_r$, of ψ in model $\mathcal{H}_r$ exists on A_T and is contained in the set $\mathcal{N}(\psi_0, \epsilon)$, which proves consistency, see also van der Vaart (1998, Theorem 5.7).

The estimators $\hat{\alpha}(\psi), \hat{\beta}(\psi), \hat{\Psi}_*(\psi), \hat{\Omega}(\psi)$, see Section 3.1, are continuous functions of ψ and are therefore also consistent.

The second derivative of $-2T^{-1}\log L_T(\lambda)$ is positive definite in the limit almost surely at $\lambda = \lambda_0$, see Lemma 9. It is therefore also positive definite in a neighbourhood $\mathcal{N}(\lambda_0, \epsilon)$ for ϵ small. It follows from Theorem 6 and Lemma 9 that also the second derivative of $-2T^{-1}\log L_T(\lambda)$ is positive definite inside $\mathcal{N}(\lambda_0, \epsilon)$ with probability converging to one, but then $-2T^{-1}\log L_T(\lambda)$ is convex and the minimum is unique. ∎

The result in Theorem 5 on existence and consistency of the MLE involves analyzing the likelihood function on the set of admissible values $0 < b \leq d$. 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 (15), 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 (2011, pp. 3153-3154).[2]

The consistency results in our Theorem 5 apply to admissible parameter sets so large that they include values of (d, b) where $\Delta^{d+ib}X_t$ is asymptotically stationary, nonstationary, or critical. The inclusion of the near critical processes in the proof is made possible by a truncation argument, allowing us to show that when $v \in [-1/2 - \kappa_1, -1/2 + \kappa]$ for κ sufficiently small, then the appropriately normalized product moment of critical processes $\Delta_+^v\varepsilon_t$ is tight in v, and uniformly large for T sufficiently large, see (107) in Lemma A.9 below.

4 Asymptotic distribution of maximum likelihood estimators

In this section we exploit consistency of the MLE and expand the likelihood in a neighborhood of the true parameter to find the asymptotic distribution of the conditional MLE.

[2]In independent and concurrent work, Hualde and Robinson (2011) prove consistency for a large set of admissible values in a fractional model with one fractional parameter and initial values equal to zero, i.e. both $X_{-n} = 0$ and $\tilde{X}_{-n} = 0$ for $n \geq 0$. Also, their consistency proof applies only to the univariate case (see their discussion on pp. 3174-3176).

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

The likelihood function for model (2) in a neighborhood of the true value is expressed in terms of $\varepsilon_t(\lambda)$, see (22) and (23).

We have identified β by $\bar{\beta}_0'\beta = I_r$, see Section 2.5, and use (11) to write $\beta = \beta_0 + \beta_{0\perp}(\bar{\beta}_{0\perp}'\beta) = \beta_0 + \beta_{0\perp}\vartheta$, say. When $b_0 > 1/2$ we let $\mathcal{N}(\psi_0, \epsilon) = \{\psi : |\psi - \psi_0| \leq \epsilon\}$. Then for $(d, b) \in \mathcal{N}(\psi_0, \epsilon)$ and $\epsilon < 1/2$ sufficiently small we have that $\delta_{-1} = d - b - d_0 = (d - b - d_0 + b_0) - b_0 \leq -b_0 + 2\epsilon < -1/2$ and $d + ib - d_0 \geq -\epsilon$ for $i \geq 0$. Hence, $\beta_{0\perp}' X_{-1,t}$ is the only nonstationary process in $\varepsilon_t(\lambda)$, see (22), and this is only possible for $b_0 > 1/2$. The information for ϑ is proportional to $\sum_{t=1}^T (\beta_{0\perp}' X_{-1,t})(\beta_{0\perp}' X_{-1,t})' = O_P(T^{-2\delta_{-1}})$, and we therefore introduce the normalized parameter $\theta = \bar{\beta}_{0\perp}'(\beta - \beta_0)T^{-(\delta_{-1}+1/2)} = \vartheta T^{-(\delta_{-1}+1/2)}$ or $\beta = \beta_0 + \beta_{0\perp}\theta T^{\delta_{-1}+1/2}$, so the information for θ is proportional to T. We have $\beta' X_{-1,t} = \beta_0' X_{-1,t} + T^{\delta_{-1}+1/2}\theta'\beta_{0\perp}' X_{-1,t}$, see (21). Let $V_t = (X_{-1,t}'\beta_0, \{X_{it}'\}_{i=0}^{k-1}, X_{kt}')'$ and define as in (22), for $\phi = (d, b, \alpha, \Psi_*)$,

$$\varepsilon_t(\lambda) = \varepsilon_t(\phi, \theta) = -\alpha T^{\delta_{-1}+1/2}\theta'\beta_{0\perp}' X_{-1,t} + (-\alpha, \Psi_*, I_p)V_t. \tag{34}$$

For the model with $d = b$ and a constant and $d_0 > 1/2$ we change the definitions in this section and use $\theta_\beta = T^{d_0-1/2}\bar{\beta}_{0\perp}'(\beta - \beta_0)$, $\theta_\rho' = \rho' + \beta' C_0^* \alpha_0 \rho_0'$, and

$$\varepsilon_t(\lambda) = \varepsilon_t(\phi, \theta) = -\alpha(T^{-d_0+1/2}\theta_\beta', \theta_\rho') \begin{pmatrix} \beta_{0\perp}' X_{-1,t} \\ 1 \end{pmatrix} + (-\alpha, \Psi_*, I_p)V_t.$$

When $b_0 > 1/2$ the product moments needed to calculate the conditional likelihood function $-2T^{-1}\log L_T(\phi, \theta)$, see (23), are

$$\begin{pmatrix} \mathcal{A}_T(\psi) & \mathcal{C}_T(\psi) \\ \mathcal{C}_T(\psi)' & \mathcal{B}_T(\psi) \end{pmatrix} = T^{-1}\sum_{t=1}^T \begin{pmatrix} T^{\delta_{-1}+1/2}\beta_{0\perp}' X_{-1,t} \\ V_t \end{pmatrix} \begin{pmatrix} T^{\delta_{-1}+1/2}\beta_{0\perp}' X_{-1,t} \\ V_t \end{pmatrix}'. \tag{35}$$

We sometimes suppress the dependence on ψ in $\mathcal{A}_T(\psi), \mathcal{B}_T(\psi)$, and $\mathcal{C}_T(\psi)$. We indicate the values for $\psi = \psi_0$ by $\mathcal{A}_T^0, \mathcal{B}_T^0, \mathcal{C}_T^0$, and $X_{-1,t}^0$. Finally we define

$$\mathcal{C}_{\varepsilon T}^0 = T^{-1/2}\sum_{t=1}^T T^{1/2-b_0}\beta_{0\perp}' X_{-1,t}^0 \varepsilon_t'. \tag{36}$$

When $b_0 < 1/2$ all processes are (asymptotically) stationary and we replace $\delta_{-1} + 1/2$ by zero in the definitions of $\mathcal{A}_T$, $\mathcal{B}_T$, $\mathcal{C}_T$, and $\mathcal{C}_{\varepsilon T}$.

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

$$\log\det(\Omega) + \mathrm{tr}(\Omega^{-1}(\alpha\theta'\mathcal{A}_T\theta\alpha' + (-\alpha, \Psi_*, I_p)\mathcal{B}_T(-\alpha, \Psi_*, I_p)' - 2\alpha\theta'\mathcal{C}_T(-\alpha, \Psi_*, I_p)')). \tag{37}$$

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

$$\hat{\theta}(\psi, \alpha, \Psi_*, \Omega) = \mathcal{A}_T^{-1}\mathcal{C}_T(-\alpha, \Psi_*, I_p)'\Omega^{-1}\alpha(\alpha'\Omega^{-1}\alpha)^{-1}, \tag{38}$$

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

$$\begin{aligned} &\log\det(\Omega) + \mathrm{tr}(\Omega^{-1}(-\alpha, \Psi_*, I_p)\mathcal{B}_T(-\alpha, \Psi_*, I_p)') \\ &- \mathrm{tr}((-\alpha, \Psi_*, I_p)\mathcal{C}_T'\mathcal{A}_T^{-1}\mathcal{C}_T(-\alpha, \Psi_*, I_p)'\Omega^{-1}\alpha(\alpha'\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1}). \end{aligned} \tag{39}$$

For $(d, b) \in \mathcal{N}(\psi_0, \epsilon), \epsilon < 1/2$, and $i = 0, 1, \ldots, k$, U_{it} and $\beta_0' U_{-1,t}$ and their derivatives with respect to (d, b) are stationary because $d + ib - d_0 \geq d - d_0 \geq -\epsilon > -1/2$. Only $\beta_{0\perp}' X_{-1,t}$ is nonstationary and only 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 (6), so that on $\mathbb{D}^{p-r}([0,1])$,

$$T^{\delta_{-1}+1/2}\beta_{\perp 0}' X_{-1,[Tu]} \Longrightarrow \beta_{0\perp}' C_0 W_{d_0-d+b-1}(u) = F_\psi(u). \tag{40}$$

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

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

For $b_0 < 1/2$, all regressors X_{it} are stationary in a neighborhood of the true value. The various quantities $\mathcal{A}_T, \mathcal{B}_T, \mathcal{C}_T$, and $\mathcal{C}_{\varepsilon T}$ are defined as above without the factor $T^{-\delta_{-1}+1/2}$, but their asymptotic properties are now different. The estimator of θ and profile likelihood function are given by (38) and (39).

The next theorem summarizes the asymptotic results for the product moments and their derivatives with respect to ψ, denoted D^m, when $\psi \in \mathcal{N}(\psi_0, \epsilon)$.

Theorem 6 *Let Assumptions 1-4 be satisfied and let $\mathcal{N}(\psi_0, \epsilon) = \{\psi : |\psi - \psi_0| \le \epsilon\} \subset \mathcal{N}$.*

(i) Suppose $1/2 < b_0 < d_0$ and $|\varepsilon_t|^q < \infty$ for some $q > (b_0 - 1/2)^{-1}$, and let ϵ be chosen so small that $q > (b - d + d_0 - 1/2)^{-1}$ for all $\psi \in \mathcal{N}(\psi_0, \epsilon)$. Then, for $m \ge 0$ and with $n = (p-r)^2 + (r+kp+p)^2 + (p-r)(r+kp+p)$, the process $(\mathsf{D}^m\mathcal{A}_T(\psi), \mathsf{D}^m\mathcal{B}_T(\psi), \mathsf{D}^m\mathcal{C}_T(\psi))$ is tight on $\mathcal{N}(\psi_0, \epsilon)$, and on $\mathbb{C}^n(\mathcal{N}(\psi_0, \epsilon))$ we have, see (40),

$$(\mathcal{A}_T(\psi), \mathsf{D}^m\mathcal{B}_T(\psi), \mathsf{D}^m\mathcal{C}_T(\psi)) \Longrightarrow (\int_0^1 F_\psi(u)F_\psi(u)'du, \mathsf{D}^m\mathcal{B}(\psi), 0), \tag{41}$$

which holds jointly with

$$\mathcal{C}_{\varepsilon T}^0 \xrightarrow{D} \int_0^1 F_0(dW)', \quad F_0(u) = F_{\psi_0}(u). \tag{42}$$

(ii) Suppose $0 < b_0 < 1/2$ and $b_0 < d_0$. Then, for $m \ge 0$, the process $(\mathsf{D}^m\mathcal{A}_T(\psi), \mathsf{D}^m\mathcal{B}_T(\psi), \mathsf{D}^m\mathcal{C}_T(\psi))$ is tight on $\mathcal{N}(\psi_0, \epsilon)$, and on $\mathbb{C}^n(\mathcal{N}(\psi_0, \epsilon))$ we find

$$(\mathcal{A}_T(\psi), \mathsf{D}^m\mathcal{B}_T(\psi), \mathsf{D}^m\mathcal{C}_T(\psi)) \Longrightarrow (\mathcal{A}(\psi), \mathsf{D}^m\mathcal{B}(\psi), \mathsf{D}^m\mathcal{C}(\psi)),$$

which is deterministic, and the convergence holds jointly with

$$\mathcal{C}_{\varepsilon T}^0 \xrightarrow{D} N_{(p-r)\times p}(0, \Omega_0 \otimes \mathcal{A}^0). \tag{43}$$

(iii) For model $\mathcal{H}_r(d = b)$ with a constant the same results hold with the relevant restriction imposed, and the relevant modifications to the definitions, e.g. $F_\psi(u)$ is replaced by $(F_0(u)', 1)'$.

Proof. *Proof of (i)*: For $d_0 > 1/2$ it follows from Theorem 2 that for $U_t^+ = C_0\varepsilon_t + \Delta_+^{b_0}Y_t^+$,

$$\tilde{\Delta}^{d+ib}X_t = \Delta_+^{d+ib-d_0}U_t^+ + \Delta_+^{d+ib}\mu_{0t} + \Delta_-^{d+ib}\tilde{X}_t, \ t = 1, \dots, T, \tag{44}$$

and hence the regressors satisfy, see (21),

$$\begin{aligned} X_{-1,t} &= (\Delta_+^{d-b-d_0} - \Delta_+^{d-d_0})U_t^+ + (\Delta_+^{d-b} - \Delta_+^{d})\mu_{0t} + (\Delta_-^{d-b} - \Delta_-^{d})\tilde{X}_t = U_{-1,t}^+ + D_{-1,t}(\psi), \quad (45)\\ X_{kt} &= \Delta_+^{d+kb-d_0}U_t^+ + \Delta_+^{d+kb}\mu_{0t} + \Delta_-^{d+kb}\tilde{X}_t = U_{kt}^+ + D_{kt}(\psi),\\ X_{it} &= (\Delta_+^{d+ib-d_0} - \Delta_+^{d+kb-d_0})U_t^+ + (\Delta_+^{d+ib} - \Delta_+^{d+kb})\mu_{0t} + (\Delta_-^{d+ib} - \Delta_-^{d+kb})\tilde{X}_t = U_{it}^+ + D_{it}(\psi), \end{aligned}$$

for $i = 0, \dots, k-1$, where $D_{it}(\psi)$ is deterministic and generated by initial values, see (92). In model (3) with $d = b$ and a constant, we replace μ_{0t} by $\mu_{0t} + C_0^*\alpha_0\rho_0'$ in X_{it} for $i \ge 0$ and subtract $\Delta_+^d C_0^*\alpha_0\rho_0'$ from $X_{-1,t}$. When $d_0 < 1/2$ we use the stationary representations (17) and (18), in which case there is no initial values term involving μ_{0t} and U_t^+ in (45) is replaced by $U_t = C_0\varepsilon_t + \Delta^{b_0}Y_t$.

It follows from Lemma A.8 that $\mathsf{D}^m D_{it}(\psi)$ is uniformly small in ψ for $t \to \infty$, so that asymptotically we can replace the regressors $X_{it}, i \geq 0$, and $\beta_0' X_{-1,t}$ by the asymptotically stationary variables U_{it}^+ and $\beta_0' U_{-1,t}^+$, see (45), in the calculation of the product moments $\mathcal{A}_T, \mathcal{B}_T$, and $\mathcal{C}_T$. The nonstationary regressor $\beta_{0\perp}' X_{-1,t}$ is normalized by $T^{d-b-d_0+1/2}$ and it follows from (95) that $T^{d-b-d_0+1/2}\beta_{0\perp}' \mathsf{D}^m D_{-1,t}(\psi)$ converges uniformly in (t,ψ) to zero for $T \to \infty$. Thus we can replace this regressor by $\beta_{0\perp}' U_{-1,t}^+$.

By Theorem 2, $U_t = C_0\varepsilon_t + \Delta^{b_0} Y_t \in \mathcal{Z}_{b_0}$, where the class $\mathcal{Z}_{b_0}$ is given in Definition 2. Lemma A.9 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 $u = d + jb - d_0 + b_0 \geq -1/2 + (1/2 - 2\epsilon)$ and $\beta_{0\perp}' U_{it}^+, i \geq 0$, with indices $u = d + ib - d_0 \geq d - d_0 \geq -1/2 + (1/2 - \epsilon)$ so $\kappa_u = 1/2 - 2\epsilon$, and the nonstationary process $\beta_{0\perp}' U_{-1,t}^+$ with index $w = d - b - d_0 \leq -b_0 + 2\epsilon \leq -1/2 - (b_0 - 1/2 - 2\epsilon)$ so $\kappa_w = b_0 - 1/2 - 2\epsilon$ noting that we have chosen ϵ so small that $q > (b_0 - 1/2)^{-1}$ implies $q > 1/\kappa_w$. Tightness of $(\mathsf{D}^m\mathcal{A}_T(\psi), \mathsf{D}^m\mathcal{B}_T(\psi), \mathsf{D}^m\mathcal{C}_T(\psi))$ and convergence in distribution of $(\mathcal{A}_T(\psi), \mathsf{D}^m\mathcal{B}_T(\psi), \mathsf{D}^m\mathcal{C}_T(\psi))$ in $\mathbb{C}^n(\mathcal{N}(\psi_0, \epsilon))$ then follows from Lemma A.9.

The proof for $\mathcal{C}_{\varepsilon T}^0$ follows from (7).

Proof of (ii): For $b_0 < 1/2 < d_0$ the only difference in the above proof is that $\beta_{0\perp}' X_{-1,t}$ is stationary (apart from a deterministic term that converges uniformly to zero) and can be replaced by $\beta_{0\perp}' U_{-1,t}^+$. The limit of $(\mathcal{A}_T(\psi), \mathsf{D}^m\mathcal{B}_T(\psi), \mathsf{D}^m\mathcal{C}_T(\psi))$ then follows from (102) of Lemma A.9. In this case we find

$$\mathcal{A}(\psi) = E(\beta_{0\perp}' U_{-1,t} U_{-1,t}' \beta_{0\perp}). \tag{46}$$

Finally, $\beta_{0\perp}' U_{-1,t}^+ \varepsilon_t'$ is a martingale difference sequence and the Central Limit Theorem for martingales gives (43), see Hall and Heyde (1980, chp. 3).

If instead $b_0 < d_0 < 1/2$ we apply the representation (17) and find

$$\tilde{\Delta}^{d+ib} X_t = \Delta_+^{d+ib} \Delta^{-d_0}(C_0\varepsilon_t + \Delta^{b_0} Y_t) + \Delta_-^{d+ib} \tilde{X}_t, \ t = 1, \ldots, T.$$

In this case μ_{0t} plays 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. (34). For (d,b) close to (d_0, b_0) we define the corresponding stationary process

$$e_t(\phi) = U_{kt} - \alpha\beta_0' U_{-1,t} + \sum_{i=0}^{k-1} \Psi_i U_{it} = (-\alpha, \Psi_*, I_p)(U_{-1,t}'\beta_0, U_{*t}', U_{kt}')'. \tag{47}$$

In the following we use D_ϕ and $\mathsf{D}_{\phi\phi}^2$ to denote first- and second-order derivatives with respect to ϕ.

Lemma 7 *Let Assumptions 1-4 hold. We find for $\phi = \phi_0$ that $e_t(\phi_0) = \varepsilon_t(\phi_0) = \varepsilon_t$ and:*

(i) When $b_0 > 1/2$ we find

$$T^{-1}\sum_{t=1}^T \varepsilon_t(\phi)\varepsilon_t(\phi)' \overset{P}{\to} Ee_t(\phi)e_t(\phi)' = (-\alpha, \Psi_*, I_p)\mathcal{B}(\psi)(-\alpha, \Psi_*, I_p)', \tag{48}$$

$$\mathsf{D}_\phi Ee_t(\phi_0)'\Omega_0^{-1}e_t(\phi_0) = E(\mathsf{D}_\phi e_t(\phi_0)'\Omega_0^{-1}\varepsilon_t) + E(\varepsilon_t'\Omega_0^{-1}\mathsf{D}_\phi e_t(\phi_0)) = 0, \tag{49}$$

$$\mathsf{D}_{\phi\phi}^2 Ee_t(\phi_0)'\Omega_0^{-1}e_t(\phi_0) = E(\mathsf{D}_\phi e_t(\phi_0)'\Omega_0^{-1}\mathsf{D}_\phi e_t(\phi_0)) = \Sigma_0, \tag{50}$$

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

(ii) When $b_0 < 1/2$ we redefine $\phi = (\theta, d, b, \alpha, \Psi_*)$ and find the limits of the product moment matrices (35), see (46),

$$T^{-1} \sum_{t=1}^{T} \varepsilon_t(\phi)\varepsilon_t(\phi)' \xrightarrow{P} (-\alpha(\theta', I_r), \Psi_*, I_p) \begin{pmatrix} \mathcal{A}(\psi) & \mathcal{C}(\psi) \\ \mathcal{C}(\psi)' & \mathcal{B}(\psi) \end{pmatrix} (-\alpha(\theta', I_r), \Psi_*, I_p)', \quad (51)$$

where $\mathcal{C}(\psi) = Cov(\beta_{0\perp}' U_{-1,t}, (U_{-1,t}'\beta_0, U_{0t}', \ldots, U_{kt}')')$, and (49) and (50) hold with suitably redefined $e_t(\phi)$ and

$$\Sigma_0^{stat} = \mathsf{D}_{\phi\phi}^2 E e_t(\phi_0)'\Omega_0^{-1} e_t(\phi_0). \quad (52)$$

(iii) For model $\mathcal{H}_r(d = b)$ with a constant the same results hold with the relevant restriction imposed and the relevant modifications to the definitions.

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 (8) and (9), 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

$$f_\phi(z) = (1-z)^{d-b-d_0+b_0}\Psi(1-(1-z)^b)|_{\beta=\beta_0,\rho=\rho_0}\Psi_0(y)^{-1}. \quad (53)$$

For $\phi = \phi_0$ we find $f_{\phi_0}(z) = 1$ so that $e_t(\phi_0) = \varepsilon_t$. The result (48) follows from (41) of Theorem 6. Differentiating the left-hand side of (48), 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 (49). Differentiating twice we find (50) 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 zero variance. 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 (8) and (53) the derivatives, where we use $y = 1-(1-z)^{b_0}$ and the relation $\Psi_k = I_p - \sum_{i=0}^{k-1}\Psi_i$,

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

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, post-multiplied by $\Psi_0(y)$, is

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

This should be zero for all y for Σ_0 to be singular. Because $\log(1-y)$ is not a polynomial we have $A\beta_0' y - \sum_{i=0}^{k-1} G_i((1-y)^{i+1} - (1-y)^{k+1}) = 0$ for all y, and hence $A = 0$ and $G_i = 0, i = 0, \ldots, k-1$. We therefore find that for all y the polynomial $(d_1-b_1)\Psi_0(y) - b_1[\mathsf{D}_y\Psi_0(y)](1-y)$ has only zero coefficients. In particular we find that the coefficients to $(1-y)^i, i = 0, 1, k+1$, are

$$0 = -(d_1 - b_1)\alpha_0\beta_0', \quad (54)$$

$$0 = d_1(\alpha_0\beta_0' + \Psi_{00}), \tag{55}$$
$$0 = (d_1 + b_1 k)\Psi_{0k},\ k > 0. \tag{56}$$

We want to show that $d_1 = b_1 = 0$. If $k > 0$, (56) and $\Psi_{0k} = (-1)^{k+1}\Gamma_{0k} \neq 0$ imply $d_1 + b_1 k = 0$. If $\alpha_0\beta_0' \neq 0$ we find from (54) that $d_1 - b_1 = 0$ and if $\alpha_0\beta_0' = 0$, (55) shows that $d_1\Psi_{00} = 0$. But in the latter case $\alpha_{0\perp} = \beta_{0\perp} = I_p$ and $\Psi_{00} = I_p - \sum_{i=1}^k \Gamma_{0i} = \Gamma_0 = \alpha_{0\perp}'\Gamma_0\beta_{0\perp} \neq 0$, so that in either case $d_1 = b_1 = 0$. If $k = 0$ and $r > 0$ then $\Psi_{00} = I_p$ and (55) shows that $d_1 = 0$ because $\alpha_0\beta_0' \neq -I_p$, and then (54) gives $d_1 = b_1 = 0$. Finally, if $k = r = 0$ the model is $\Delta^d X_t = \varepsilon_t$ and the condition for singularity is $d_1 d_0^{-1}\log(1-y)I_p = 0$ which implies $d_1 = 0$. Hence in all cases $d_1 = b_1 = 0$ and Σ_0 is positive definite.

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 the 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 the first- and second-order derivatives are tight on $\mathcal{N}(\psi_0, \epsilon)$ (by Theorem 6).

Lemma 8 *Let Assumptions 1-4 be satisfied and $(k, r) \neq (0, 0)$. We assume that $X_{-n} = 0$ for $n \geq T^{\upsilon}$ for some $\upsilon < 1/2$.*

(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 (2) at the true value is given by

$$\begin{pmatrix} T^{-1/2}\mathsf{D}_\phi \log L_T(\lambda_0) \\ T^{-1/2}\mathsf{D}_\theta \log L_T(\lambda_0) \end{pmatrix} \xrightarrow{D} \begin{pmatrix} N_{n_\phi}(0, \Sigma_0) \\ (\mathrm{vec}(\int_0^1 F_0(dG_0)'))' \end{pmatrix}, \tag{57}$$

where Σ_0 is given in (50), $n_\phi = 1 + 1 + pr + kp^2$ is the number of parameters in $\phi = (d, b, \alpha, \Psi_*)$, $F_0 = \beta_{0\perp}' C_0 W_{b_0-1}$, $G_0 = \alpha_0'\Omega_0^{-1}W$, and the two components in the limit in (57) are independent.

(ii) If $0 < b_0 < 1/2$ then the score with respect to all parameters is asymptotically Gaussian, $N_{n_\phi+(p-r)r}(0, \Sigma_0^{stat})$, see (52).

(iii) In model $\mathcal{H}_r(d = b)$ with a constant the same results hold with θ replaced by $(\theta_\beta', \theta_\rho')'$ and F_0 by $(W_{b_0-1}' C_0' \beta_{0\perp}, 1)'$.

Proof. For $\lambda = \lambda_0$ we find

$$\varepsilon_t(\lambda_0) = \varepsilon_t + \Pi_{0-}(L)(\tilde{X}_t - X_t) = \varepsilon_t + d_{0t},$$
$$\mathsf{D}\varepsilon_t(\lambda_0) = \mathsf{D}\Pi_{+0}(L)X_t + \mathsf{D}\Pi_{0-}(L)\tilde{X}_t = s_{1t} + d_{1t},$$

where d_{1t} is a linear combination of the deterministic terms $\mathsf{D}D_{it}(\psi)|_{\psi=\psi_0}$ and, if $b_0 > 1/2$, also $T^{1/2-b_0}\beta_{0\perp}'\mathsf{D}D_{-1,t}(\psi)|_{\psi=\psi_0}$, see (92) and (93). From Lemma A.8 we find that the terms of d_{1t} either tend to zero as $t \to \infty$ or satisfy $T^{1/2-b_0}\max_{1\leq t\leq T}|\beta_{0\perp}'\mathsf{D}D_{-1,t}(\psi)|_{\psi=\psi_0}| \to 0$ as $T \to \infty$, and that $T^{-1/2}\sum_{t=1}^T |d_{0t}| \to 0$ as $T \to \infty$. These properties are enough to show that $T^{-1/2}\sum_{t=1}^T |d_{0t}d_{1t}| \to 0$ as $T \to \infty$ and that in product moments where d_{1t} appears it can in fact be ignored.

Proof of (i): For $b_0 > 1/2$, $T^{-1/2}\mathsf{D}_\phi \log L_T(\lambda_0)$ for $\phi = (d, b, \alpha, \Psi_*)$ evaluated at λ_0 is

$$-T^{-1/2}\sum_{t=1}^T \varepsilon_t(\lambda_0)'\Omega_0^{-1}\mathsf{D}_\phi\varepsilon_t(\lambda_0) = -T^{-1/2}\sum_{t=1}^T \varepsilon_t'\Omega_0^{-1}(s_{1t} + d_{1t}) - T^{-1/2}\sum_{t=1}^T d_{0t}'\Omega_0^{-1}(s_{1t} + d_{1t}).$$

The first term is a martingale with sum of conditional variances $T^{-1}\sum_{t=1}^{T}(s_{1t}+d_{1t})'\Omega_0^{-1}(s_{1t}+d_{1t})\overset{P}{\to}\Sigma_0$, see Lemma 7, because d_{1t} can be ignored. In the second term we find that the second moment is bounded by $c(T^{-1/2}\sum_{t=1}^{T}|d_{0t}|)^2\to 0$. The result for the first block of (57) now follows from the Central Limit Theorem for martingales, see Hall and Heyde (1980, chp. 3).

The score function for θ evaluated at the true value is

$$
\begin{aligned}
&T^{-1/2}\mathsf{D}_\theta \log L_T(\lambda_0)\\
&=(\mathrm{vec}(T^{-1/2}\sum_{t=1}^{T}T^{-1/2-b_0}\beta_{0\perp}'X_{-1,t}^0\varepsilon_t'\Omega_0^{-1}\alpha_0))'+(\mathrm{vec}(T^{-1/2}\sum_{t=1}^{T}T^{1/2-b_0}\beta_{0\perp}'X_{-1,t}^0d_{0t}'\Omega_0^{-1}\alpha_0))',
\end{aligned}
$$

where $X_{-1,t}^0$ denotes $X_{-1,t}$ evaluated at $\psi=\psi_0$. The main term converges in distribution to $(\mathrm{vec}(\int_0^1 F_0(dG_0)'))'$, see (42) in Theorem 6, and the second term converges in probability to zero because $\max_{1\le t\le T}|T^{1/2-b_0}\beta_{0\perp}'X_{-1,t}^0|=O_P(1)$ by (6) and $T^{-1/2}\sum_{t=1}^{T}|d_{0t}|\to 0$ by (96). This proves the second block of (57). The independence of the two components in the limit of (57) follows exactly as in JN (2010, Lemma 10).

Proof of (ii): If $0<b_0<1/2$, all stochastic regressors are asymptotically stationary and we take $\beta=\beta_0+\beta_{0\perp}\theta$ and the score with respect to θ, evaluated at $\lambda=\lambda_0$, is

$$
T^{-1/2}\mathsf{D}_\theta \log L_T(\lambda_0)=T^{-1/2}\sum_{t=1}^{T}(\mathrm{vec}(\beta_{0\perp}'X_{-1,t}^0\varepsilon_t(\lambda_0)'\Omega_0^{-1}\alpha_0))'.
$$

The Central Limit Theorem for martingales gives the result.

Proof of (iii): The same methods can be used here, noting that the score with respect to ρ, evaluated at $\lambda=\lambda_0$, is $T^{-1/2}\sum_{t=1}^{T}\varepsilon_t(\lambda_0)'\Omega_0^{-1}\alpha_0$. ■

Lemma 9 *Let Assumptions 1-4 be satisfied and $(k,r)\neq(0,0)$.*

(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 (2) for $(\phi,\theta)=(\phi_0,0)$ converges in distribution to

$$
\begin{pmatrix}\Sigma_0 & 0\\ 0 & \alpha_0'\Omega_0^{-1}\alpha_0\otimes\int_0^1 F_0F_0'du\end{pmatrix}>0\ a.s.,\tag{58}
$$

where Σ_0 is given in (50) and $F_0(u)=\beta_{0\perp}'C_0W_{b_0-1}(u)$.

(ii) If $0<b_0<1/2$ the information per observation for all parameters is convergent in probability to the non-stochastic limit Σ_0^{stat} given in (52).

(iii) For the model $\mathcal{H}_r(d=b)$ with a constant the same results hold with F_0 replaced by $(F_0',1)'$.

Proof. *Proof of (i)*: The information matrices can be found from (37) and the deterministic terms can be ignored due to Lemma A.8. From (41) of Theorem 6 it holds that $\mathsf{D}^m\mathcal{C}_T^0\overset{P}{\to}0$. Using this and (50) we find using $\theta_0=0$ that

$$
\begin{aligned}
&-T^{-1}\mathsf{D}_{\phi\phi}^2\log L_T(\lambda_0)\overset{P}{\to}\Sigma_0,\\
&-T^{-1}\mathsf{D}_{\theta\theta}^2\log L_T(\lambda_0)=\alpha_0'\Omega_0^{-1}\alpha_0\otimes\mathcal{A}_T^0\overset{D}{\to}\alpha_0'\Omega_0^{-1}\alpha_0\otimes\int_0^1 F_0F_0'du,\\
&-T^{-1}\mathsf{D}_{\theta\phi}^2\log L_T(\lambda_0)=\mathsf{D}_{\theta\phi}^2\,\mathrm{tr}(\Omega^{-1}2\alpha\theta'\mathcal{C}_T(-\alpha,\Psi_*,I_p)')|_{\lambda=\lambda_0}\overset{P}{\to}0.
\end{aligned}
$$

Proof of (ii): If $0 < b_0 < 1/2$ we find the information for $\theta = \bar{\beta}_{0\perp}'(\beta - \beta_0)$ to be

$$-T^{-1}\mathsf{D}^2_{\theta\theta}\log L_T(\lambda_0) = \alpha_0'\Omega_0^{-1}\alpha_0 \otimes T^{-1}\sum_{t=1}^{T}(\beta_{0\perp}'X^0_{-1,t})(\beta_{0\perp}'X^0_{-1,t})',$$

and the cross term $-T^{-1}\mathsf{D}^2_{\theta\phi}\log L_T(\lambda_0)$ can be found similarly from (37). In this case the entire information matrix converges to a non-stochastic limit by the Law of Large Numbers because $X^0_{-1,t}$ is (asymptotically) stationary when $b_0 < 1/2$, see also (102).

Proof of (iii): The same methods can be applied in this case. ■

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 with $(k, r) \neq (0, 0)$ and suppose $(d_0, b_0) \in \text{int}(\mathcal{N})$. Assume also that $X_{-n} = 0$ for $n \geq T^{\upsilon}$ for some $\upsilon < 1/2$.*

(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 maximum likelihood estimators $\hat{\phi} = (\hat{d}, \hat{b}, \hat{\alpha}, \hat{\Psi}_*)$ and $\hat{\beta}$ for model (2) is given by

$$\begin{pmatrix} T^{1/2}\,\text{vec}(\hat{\phi} - \phi_0) \\ T^{b_0}\bar{\beta}_{0\perp}'(\hat{\beta} - \beta_0) \end{pmatrix} \overset{D}{\to} \begin{pmatrix} N_{n_\phi}\left(0, \Sigma_0^{-1}\right) \\ (\int_0^1 F_0F_0'du)^{-1}\int_0^1 F_0(dG_0)'(\alpha_0'\Omega_0^{-1}\alpha_0)^{-1} \end{pmatrix}, \tag{59}$$

where $F_0 = \beta_{0\perp}'C_0W_{b_0-1}$ and $G_0 = \alpha_0'\Omega_0^{-1}W$ are independent, and also the two components of (59) are independent. It follows that the asymptotic distribution of $\text{vec}(T^{b_0}\bar{\beta}_{0\perp}'(\hat{\beta} - \beta_0))$ is mixed Gaussian with conditional variance given by

$$(\alpha_0'\Omega_0^{-1}\alpha_0)^{-1} \otimes (\int_0^1 F_0F_0'du)^{-1}. \tag{60}$$

(ii) If $0 < b_0 < 1/2$ the estimators for $(d, b, \alpha, \beta, \Psi_*)$ are asymptotically Gaussian.

(iii) In the model $\mathcal{H}_r(d = b)$ with a constant the same results hold with the relevant restriction imposed and with F_0 replaced by $(F_0', 1)'$.

(iv) If $k = r = 0$ the model is $\Delta^d X_t = \varepsilon_t$ and $\hat{d}$ is asymptotically Gaussian.

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 $l_T = -2T^{-1}\log L_T$ that

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

Here asterisks indicate intermediate points between $(\hat{\phi}, \hat{\theta}, \hat{\Omega})$ and $(\phi_0, 0, \hat{\Omega})$, one for each row.

We have proved tightness of the product moments as functions of ψ in a compact set, see Theorem 6. Here we need tightness of the second derivatives in all parameters λ in a compact neighborhood of the true value, λ_0, which follows from Lemma A.2 because the second derivatives are continuously differentiable in the parameters $(\alpha, \beta, \Psi_*, \Omega)$ and the product moments. Because the second derivatives are tight and because $\lambda^* \overset{P}{\to} \lambda_0$ by Theorem 5, we apply Lemma A.3 to replace λ^* by λ_0. The limit of the information per observation is then given in Lemma 9.

The score functions normalized by $T^{1/2}$ are given in the proof of Lemma 8, and because Ω only acts as a scaling factor on a term that converges in distribution (and therefore is tight), tightness as a function of Ω follows. Hence we replace $(\phi_0, 0, \hat{\Omega})$ by λ_0 in the normalized score functions, and their weak limits are given in Lemma 8.

This yields the asymptotic distribution of $T^{1/2}((\mathrm{vec}(\hat{\phi}-\phi_0))', (\mathrm{vec}\,\hat{\theta})')'$. We then prove (59) using $T^{-d_0+b_0+\hat{d}-\hat{b}} = e^{O_P(T^{-1/2}\log T)} = 1 + o_P(1)$ and the relation

$$T^{b_0-1/2}\bar{\beta}_{0\perp}'(\hat{\beta}-\beta_0) = T^{-d_0+b_0+\hat{d}-\hat{b}}T^{1/2}\hat{\theta}.$$

The stochastic component of the process F_0 is a function of $\alpha_{0\perp}'W$, see (13) and (42), whereas $G_0 = \alpha_0'\Omega_0^{-1}W$, so that F_0 and G_0 are independent and the limit distribution of $T^{1/2}\hat{\theta}$ is mixed Gaussian. Finally, the independence of the two components of (59) follows from Lemma 8 and the block-diagonality in (58), see also Johansen (1991, p. 1573).

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 and the asymptotic variance is $(\Sigma_0^{stat})^{-1}$, see (52).

Proof of (iii): In the model $\mathcal{H}_r(d = b)$ with $b_0 > 1/2$ the same results hold by the same type of proof. For $0 < b_0 < 1/2$ we find the asymptotic distribution of $\hat{\beta}$ and $\hat{\rho}$ jointly with the other parameters from

$$\begin{pmatrix} T^{1/2}\bar{\beta}_{0\perp}'(\hat{\beta}-\beta_0) \\ T^{1/2}(\hat{\rho}-\rho_0) \end{pmatrix} = \begin{pmatrix} I_{p-r} & 0 \\ -\rho_0\alpha_0'C_0^{*\prime}\beta_{0\perp} & I_r \end{pmatrix} \begin{pmatrix} T^{1/2}\hat{\theta}_\beta \\ T^{1/2}\hat{\theta}_\rho \end{pmatrix},$$

which shows how the asymptotic variance can be calculated from Σ_0^{stat}, see (52).

Proof of (iv): Follows by the same methods. ■

The results in Theorem 10 show 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 $b_0 > 1/2$. 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 (i) inference on $\hat{\phi}$ to be conducted as if $\hat{\beta}$ were known and vice versa, and (ii) 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 fractionally cointegrated 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.

5 Likelihood ratio test for cofractional rank

We consider the model

$$\mathcal{H}_p : \Delta^d X_t = \Pi\Delta^{d-b}L_b X_t + \sum_{i=1}^{k}\Gamma_i\Delta^d L_b^i X_t + \varepsilon_t \tag{61}$$

and want to test the hypothesis $\mathcal{H}_r : \mathrm{rank}(\Pi) \leq r$ against the alternative $\mathcal{H}_p : \mathrm{rank}(\Pi) \leq p$. For model $\mathcal{H}_r, r = 0, 1, \ldots, p$, 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 (26) in Section 3.1, and let $\hat{\psi}_r$ be the MLE of ψ. The likelihood ratio (LR) statistic is

$$-2\log LR(\mathcal{H}_r|\mathcal{H}_p) = T\log\frac{\det(S_{00}(\hat{\psi}_r))\prod_{i=1}^{r}(1-\hat{\omega}_i(\hat{\psi}_r))}{\det(S_{00}(\hat{\psi}_p))\prod_{i=1}^{p}(1-\hat{\omega}_i(\hat{\psi}_p))} = T(\ell_{T,r}(\hat{\psi}_r) - \ell_{T,p}(\hat{\psi}_p)). \tag{62}$$

Theorem 11 *Let the assumptions of Theorem 10 hold with $(k,r)\neq(0,0)$.*

(i) If $b_0>1/2$ the likelihood ratio statistic for $\Pi=\alpha\beta'$, that is $\mathcal{H}_r$ in $\mathcal{H}_p$, has asymptotic distribution

$$-2\log LR(\mathcal{H}_r|\mathcal{H}_p)\overset{D}{\to}\operatorname{tr}\Big(\int_0^1(dB)B_{b_0-1}'\Big(\int_0^1 B_{b_0-1}B_{b_0-1}'du\Big)^{-1}\int_0^1 B_{b_0-1}(dB)'\Big),\tag{63}$$

where $B(u)$ is $(p-r)$-dimensional standard BM and $B_{b_0-1}(u)$ is the corresponding fBM. The limit distribution is continuous in b_0.

(ii) If $0<b_0<1/2$ then

$$-2\log LR(\mathcal{H}_r|\mathcal{H}_p)\overset{D}{\to}\chi^2((p-r)^2).\tag{64}$$

(iii) Let P_1 be the probability measure under the alternative $\Pi_1=\alpha_1\beta_1'=\alpha\beta'+\alpha^*\beta^{*\prime}$, where $\alpha_1=(\alpha,\alpha^*)$ and $\beta_1=(\beta,\beta^*)$ are $p\times(r+r^*)$ of rank $r_1=r+r^*>r$, and hence $\operatorname{rank}(\Pi_1)>r$. Assume that Assumption 1 is satisfied under the alternative. Then

$$-2\log LR(\mathcal{H}_r|\mathcal{H}_p)\overset{P_1}{\to}\infty.\tag{65}$$

(iv) In the model $\mathcal{H}_r(d=b)$ with a constant the results (i)–(iii) hold for $k\geq 0, r\geq 0$ and $B_{b_0-1}(u)$ replaced by $(B_{d_0-1}(u)',1)'$.

Proof. We give the proofs only for model (2) without the constant. The proofs for part *(iv)* are the same but with different notation and with the extended fBM replacing the fBM, reflecting the reduced rank regression of X_{kt} on $(X_{-1,t}',1)'$. For parts *(i)*–*(iii)* note that $(k,r)\neq(0,0)$ ensures that b is identified under the null, but for part *(iv)* this is not a problem because $b=d$ is identified also when $k=r=0$.

Proof of (i): We assume that $\operatorname{rank}(\Pi)=r$ and that $\Pi_0=\alpha_0\beta_0'$, 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 $\beta=\beta_0$, see Lawley (1956) and Johansen (2002) for an application to the cointegrated VAR model.

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

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

The statistic $LR(\mathcal{H}_r$ and $\beta=\beta_0|\mathcal{H}_p)$ is the test that $\Pi=\alpha\beta_0'$ (with rank r) against Π unrestricted, and $LR(\beta=\beta_0|\mathcal{H}_r)$ is the test that $\beta=\beta_0$ in the model with $\Pi=\alpha\beta'$ and $\operatorname{rank}(\Pi)=r$. We next find a first order approximation to each statistic and subtract them. For $T\to\infty$ 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

$$-2\log LR(\xi=\xi_0)=\mathsf{D}_\xi\log L_T(\xi_0,\eta_0)(\mathsf{D}^2_{\xi\xi}\log L_T(\xi_0,\eta_0))^{-1}\mathsf{D}_\xi\log L_T(\xi_0,\eta_0)'+o_P(1),\tag{66}$$

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

A first order approximation to $-2\log LR(\beta=\beta_0|\mathcal{H}_r)$: It follows from Lemma 9 that, for $\xi=\theta,\ \eta=(d,b,\alpha,\Psi_*,\Omega)$, 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

$$-2\log LR(\beta=\beta_0|\mathcal{H}_r)=(\operatorname{vec}(\mathcal{C}^0_{\varepsilon T}\Omega_0^{-1}\alpha_0))'(\alpha_0'\Omega_0^{-1}\alpha_0\otimes\mathcal{A}^0_T)^{-1}\operatorname{vec}(\mathcal{C}^0_{\varepsilon T}\Omega_0^{-1}\alpha_0)+o_P(1)\tag{67}$$

$$= \mathrm{tr}((\alpha_0'\Omega_0^{-1}\alpha_0)^{-1}\alpha_0'\Omega_0^{-1}\mathcal{C}_{\varepsilon T}^{0\prime}(\mathcal{A}_T^0)^{-1}\mathcal{C}_{\varepsilon T}^0\Omega_0^{-1}\alpha_0) + o_P(1),$$

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

A first order approximation to $-2\log LR(\mathcal{H}_r$ and $\beta = \beta_0|\mathcal{H}_p)$: In model (61) we introduce a convenient reparametrization by $\alpha = \Pi\bar{\beta}_0, \xi' = T^{-\delta_{-1}-1/2}\Pi\bar{\beta}_{0\perp}$, so that by (11) we have $\Pi = \alpha\beta_0' + T^{\delta_{-1}+1/2}\xi'\beta_{0\perp}'$. The equations are, see (34),

$$X_{kt} = \alpha\beta_0' X_{-1,t} + \xi' T^{d-b-d_0+1/2}\beta_{0\perp}' X_{-1,t} - \sum_{i=0}^{k-1}\Psi_i X_{it} + \varepsilon_t.$$

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

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

where $\eta = (d,b,\alpha,\Psi_*,\Omega)$. This expression is the same as the conditional likelihood (37) 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 the limit is block diagonal. It follows as above that

$$-2\log LR(\mathcal{H}_r \text{ and } \beta = \beta_0|\mathcal{H}_p) = \mathrm{tr}(\Omega_0^{-1}\mathcal{C}_{\varepsilon T}^{0\prime}(\mathcal{A}_T^0)^{-1}\mathcal{C}_{\varepsilon T}^0) + o_P(1). \quad (68)$$

A first order approximation to $-2\log LR(\mathcal{H}_r|\mathcal{H}_p)$: Subtracting (67) from (68) 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

$$\mathrm{tr}(\alpha_{0\perp}(\alpha_{0\perp}'\Omega_0\alpha_{0\perp})^{-1}\alpha_{0\perp}'\mathcal{C}_{\varepsilon T}^{0\prime}(\mathcal{A}_T^0)^{-1}\mathcal{C}_{\varepsilon T}^0) \quad (69)$$

$$\xrightarrow{D} \mathrm{tr}(\alpha_{0\perp}(\alpha_{0\perp}'\Omega_0\alpha_{0\perp})^{-1}\alpha_{0\perp}'\int_0^1 (dW)F_0'(\int_0^1 F_0F_0'du)^{-1}\int_0^1 F_0(dW)') = DF(\psi_0),$$

say, which is the desired result if we define $B = (\alpha_{0\perp}'\Omega_0\alpha_{0\perp})^{-1/2}\alpha_{0\perp}'W$ and note that B_{b_0-1} is a linear transformation of F_0.

The continuity of the limit distribution can be seen by noticing that the matrices $\int_0^1 F_\psi(dB)'$ and $\int_0^1 F_\psi F_\psi' du$, and hence also $DF(\psi)$, are continuous in $\mathbb{L}_2$ 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 (ii): In this case the result follows from the usual expansion of the LR test statistic and the asymptotic distribution in Theorem 10.

Proof of (iii): The test for $\mathcal{H}_r$ in $\mathcal{H}_p$ is given in (62). We choose a small neighborhood $\mathcal{N}(\psi_0,\epsilon) = \{\psi : |\psi - \psi_0| \le \epsilon\}$ and find for fixed $\psi \in \mathcal{N}(\psi_0,\epsilon)$ that

$$\sum_{i=1}^p \log(1-\hat{\omega}_i(\psi)) = \sum_{i=1}^r \log(1-\hat{\omega}_i(\psi)) + \sum_{i=r+1}^p \log(1-\hat{\omega}_i(\psi))$$
$$\le \sum_{i=1}^r \log(1-\hat{\omega}_i(\psi)) + \log(1 - \min_{\psi\in\mathcal{N}(\psi_0,\epsilon)}\hat{\omega}_{r+1}(\psi)).$$

Adding $\log\det(S_{00}(\psi))$ on both sides and minimizing over ψ we find $\ell_{T,p}(\hat{\psi}_p) \le \ell_{T,r}(\hat{\psi}_r) + \log(1 - \min_{\psi\in\mathcal{N}(\psi_0,\epsilon)}\hat{\omega}_{r+1}(\psi))$, so that

$$-2\log LR(\mathcal{H}_r|\mathcal{H}_p) \ge -T\log(1 - \min_{\psi\in\mathcal{N}(\psi_0,\epsilon)}\hat{\omega}_{r+1}(\psi)). \quad (70)$$

We now show that the right hand side diverges to infinity under P_1, the probability measure described in *(iii)*, or equivalently that for some $\epsilon > 0, \delta > 0$, and any $\xi > 0$

there is a $T_0 = T(\epsilon, \xi, \delta)$ so that $P_1(\min_{\psi \in \mathcal{N}(\psi_0,\epsilon)} \hat{\omega}_{r+1}(\psi) \geq \delta) \geq 1 - \xi$ for all $T \geq T_0$. The eigenvalues are continuous functions of the product moments $\mathcal{A}_T(\psi), \mathcal{B}_T(\psi), \mathcal{C}_T(\psi)$, see (35). It therefore follows from Theorem 6 that, under P_1, $\hat{\omega}_{r+1}(\cdot)$ is tight on $\mathcal{N}(\psi_0, \epsilon)$ and $\hat{\omega}_{r+1}(\psi) \Longrightarrow \omega_{r+1}(\psi)$ on $\mathbb{C}(\mathcal{N}(\psi_0, \epsilon))$ as $T \to \infty$, see (41), where $\omega_{r+1}(\psi)$ is given by the solution of (138). This implies that $\omega_{r+1}(\psi) > 0$ is continuous in ψ. Therefore we can choose ϵ so small that $\min_{\psi \in \mathcal{N}(\psi_0,\epsilon)} \omega_{r+1}(\psi) > \delta$, say, for some small $\delta > 0$. Because the function $\omega_{r+1}(\cdot) \mapsto \min_{\psi \in \mathcal{N}(\psi_0,\epsilon)} \omega_{r+1}(\psi)$ is continuous in the uniform topology on $\mathcal{N}(\psi_0, \epsilon)$ we get that

$$\min_{\psi \in \mathcal{N}(\psi_0,\epsilon)} \hat{\omega}_{r+1}(\psi) \overset{P_1}{\to} \min_{\psi \in \mathcal{N}(\psi_0,\epsilon)} \omega_{r+1}(\psi) > \delta,$$

such that for any $\xi > 0$ we can find T_0 so that

$$P_1(\min_{\psi \in \mathcal{N}(\psi_0,\epsilon)} \hat{\omega}_{r+1}(\psi) > \delta) \geq 1 - \xi \text{ for all } T \geq T_0, \tag{71}$$

which completes the proof of *(iii)*. ■

We note that in model $\mathcal{H}_r$ with $k = 0$ we can test $r = 0$ by testing $\Pi = 0$, see (61), but then b is not identified under the null. For fixed b this LR statistic is denoted $LR(b) = -2 \log LR(\Pi = 0|b)$ and it is possible to consider a sup-type test, $\sup_b LR(b)$, where the supremum is taken either over stationary or non-stationary values of the index b, see Hansen (1996) for the general theory and Lasak (2010) for a cointegration test. Note that in model $\mathcal{H}_r(d = b)$ the parameter $b = d$ is identified and (63) applies also for $k = r = 0$.

The distribution (63) 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 only the parameter b_0, describing the "strength" of the cofractional relations, which determines the order of the fBMs in the limit distribution. For given hypothesized b_0 or estimated $\hat{b}_r$, the distribution (63) can be simulated to obtain critical values on a case-by-case basis. Alternatively, numerical CDFs have been simulated as functions of b_0 by MacKinnon and Nielsen (2011), and their computer programs can be used to immediately obtain critical values or P-values for the tests, including that in part *(iv)* for model $\mathcal{H}_r(d = b)$ with a constant. In either case, the continuity of the limit distribution (63) in b_0 ensures asymptotic validity of the approach.

We estimate the cofractional rank by conducting a sequence of tests, for a given size δ: test $\mathcal{H}_r$ for $r = 0, 1, \ldots$ until rejection, and the estimated rank $\hat{r}$ is the last non-rejected value of r. If the true rank is r_0, then consistency of the LR rank test in Theorem 11*(iii)* shows that any test of $r < r_0$ will reject with probability one as $T \to \infty$. Thus, $P_0(\hat{r} < r_0) \to 0$. Since the asymptotic size of the test for rank is δ we have that $P_0(\hat{r} > r_0) \to \delta$ and it follows that $P_0(\hat{r} = r_0) \to 1 - \delta$. 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

In this paper well known likelihood based inference results for the cointegrated VAR model (1) have been generalized to the cointegrated fractional $\text{VAR}_{d,b}(k)$ models,

$$\Delta^d X_t = \Delta^{d-b} L_b \alpha \beta' X_t + \sum_{i=1}^{k} \Gamma_i \Delta^d L_b^i X_t + \varepsilon_t, \; 0 < b \leq d, \tag{72}$$

$$\Delta^d X_t = L_d \alpha (\beta' X_t + \rho') + \sum_{i=1}^{k} \Gamma_i \Delta^d L_d^i X_t + \varepsilon_t. \tag{73}$$

For model (72) we have analyzed the conditional Gaussian likelihood given initial values. We have shown existence and consistency of the maximum likelihood estimators, 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. For the proof of consistency we assumed that initial values, $X_{-n}, n \geq 0$, are bounded, and for the asymptotic distribution theory we assumed that initial values are zero for $n \geq T^{\upsilon}$ for some $\upsilon < 1/2$. If $b_0 > 1/2$ 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 $B_{b_0-1}(u)$. If $b_0 < 1/2$ the estimators are all asymptotically Gaussian and the test for rank is asymptotically χ^2. The same type of results hold for the model with $d = d_0$, a prespecified value. For the model $\mathcal{H}_r(d = b)$ with a constant, i.e. (73), the same results hold except the test for rank involves $(B_{d_0-1}(u)', 1)'$.

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 $\mathbb{C}^p(\mathbb{K})$ and $\mathbb{D}^p(\mathbb{K})$.

A.1 Results on convergence in distribution

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

Lemma A.1 *If $X_T(s)$ is a sequence of p-dimensional continuous processes on a compact set $\mathbb{K} \subseteq \mathbb{R}^2$, i.e. $X_T(\cdot) \in \mathbb{C}^p(\mathbb{K})$, with*

$$||X_T(s)||_4 \leq c \text{ and } ||X_T(s_1) - X_T(s_2)||_4 \leq c|s_1 - s_2|, \quad s_1, s_2 \in \mathbb{K}, \tag{74}$$

for some constant $c > 0$, which does not depend on T, s_1, or s_2, then $X_T(s)$ is tight on $\mathbb{K}$.

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

Lemma A.2 *If the sequence of p-dimensional continuous processes $X_T(s)$ is tight on $\mathbb{K} \subseteq \mathbb{R}^q$, the vector $u \in \mathbb{R}^k$, and the function $F : \mathbb{R}^k \times \mathbb{R}^p \longmapsto \mathbb{R}^m$ is continuously differentiable, then $Z_T(u, s) = F(u, X_T(s))$ is tight on $\mathbb{R}^k \times \mathbb{K}$.*

Proof. JN (2010, Lemma A.2). ■

Lemma A.3 *Assume that $S_T \overset{P}{\to} s_0 \in \mathbb{K} \subseteq \mathbb{R}^q$ and that the $p \times p$ matrix-valued continuous process $X_T(s)$ is tight on $\mathbb{K}$. Then $X_T(S_T) - X_T(s_0) \overset{P}{\to} 0$.*

Proof. See JN (2010, Lemma A.3) for the vector-valued result. ■

Lemma A.4 *Let $X_T(s)$ be a sequence of p-dimensional continuous processes on a compact set $\mathbb{K} \subseteq \mathbb{R}^q$ and suppose $X_T(s) \Longrightarrow X(s)$ on $\mathbb{C}^p(\mathbb{K})$ as $T \to \infty$. If $X(s)$ is deterministic then $X_T(s) \overset{P}{\to} X(s)$ uniformly in $s \in \mathbb{K}$.*

Proof. If $X_T(s) \Longrightarrow X(s)$ and $X(s)$ is deterministic then $X_T(s)-X(s) \Longrightarrow 0$. By the Continuous Mapping Theorem it follows that $\sup_{s\in\mathbb{K}}|X_T(s)-X(s)| \overset{D}{\to} 0$ and therefore $\sup_{s\in\mathbb{K}}|X_T(s)-X(s)| \overset{P}{\to} 0$. ■

A.2 Bounds on product moments

We begin with some bounds on the fractional coefficients.

Lemma A.5 *(i) For $|u| \le u_0$ and all $j \ge 1$ it holds uniformly in u that*

$$|\mathsf{D}^m \pi_j(-u)| \le c(u_0)(1+\log j)^m j^{-u-1}, \tag{75}$$

$$|\mathsf{D}^m T^u \pi_j(-u)| \le c(u_0) T^u (1+|\log \frac{j}{T}|)^m j^{-u-1}. \tag{76}$$

(ii) Let $j \ge 1$ and let $\mathbb{K}$ denote any compact subset of $\mathbb{R}\backslash\mathbb{N}_0$. Then it holds that

$$\pi_j(-v) = \frac{1}{\Gamma(-v)} j^{-v-1}(1+\epsilon_j(v)), \tag{77}$$

where $\max_{v\in\mathbb{K}}|\epsilon_j(v)| \to 0$ as $j\to\infty$. Thus $\pi_j(-v) \ge cj^{-v-1}$ uniformly in $v\in\mathbb{K}$ and all $j \ge 1$.

Proof. For *(i)*, see JN (2010, Lemma B.3). To prove *(ii)* we apply Stirling's formula,

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

where $\max_{v\in\mathbb{K}}|\epsilon_j(v)| \to 0$ as $j\to\infty$. This proves the result and shows that the constant in the lower bound does not depend on v. ■

Our proof of tightness applies the result of Kallenberg (2001) in Lemma A.1 and involves evaluation of the fourth moment of linear processes and their product moments. For real coefficients $\zeta_{1n}, \zeta_{2n},\ n = 0, 1\ldots,$ we give evaluations of such moments in terms of the quantity

$$\xi_T(\zeta_1,\zeta_2) = \max_{0\le|n-m|\le T} \sum_{t=\max(n,m)}^{T+\min(n,m)} |\zeta_{1,t-n}\zeta_{2,t-m}|. \tag{78}$$

Lemma A.6 *For $i = 1, 2$, let ε_{it} be i.i.d.$(0,\sigma_i^2)$ with $E|\varepsilon_{it}|^8 < \infty$. Assume that $\{\zeta_{in}\}_{n=0}^\infty$ and $\{\xi_{in}\}_{n=0}^\infty$ are real coefficients satisfying $\sum_{n=0}^\infty |\xi_{in}| < \infty$. Define $Z_{it}^+ = \sum_{n=0}^{t-1} \xi_{in}\varepsilon_{i,t-n}$ and, for $t > N \ge 0$, also the processes $\overline{Z}_{it}^{(N)} = \sum_{n=N}^{t-1} \zeta_{in} Z_{i,t-n}^+$ and $\underline{Z}_{it}^{(N)} = \sum_{n=0}^{N-1} \zeta_{in} Z_{i,t-n}^+$. Then, for $0 \le N < T$,*

$$||T^{-1}\sum_{t=N+1}^T \overline{Z}_{1t}^{(N)}\overline{Z}_{2t}^{(N)}||_4 \le c\xi_T(\zeta_1,\zeta_2), \tag{79}$$

$$||T^{-1}\sum_{t=N+1}^T \underline{Z}_{1t}^{(N)}\underline{Z}_{2t}^{(N)}||_4 \le c\xi_N(\zeta_1,\zeta_2), \tag{80}$$

$$||T^{-1}\sum_{t=N+1}^T \underline{Z}_{1t}^{(N)}\underline{Z}_{2t}^{(N)} - E[\underline{Z}_{1t}^{(N)}\underline{Z}_{2t}^{(N)}]||_4 \le c(N/T)^{1/4}\xi_N(\zeta_1,\zeta_2), \tag{81}$$

$$||T^{-1}\sum_{t=N+1}^T \underline{Z}_{1t}^{(N)}\overline{Z}_{2t}^{(N)})||_4 \le c(N/T)^{1/4}\xi_N(\zeta_1,\zeta_1)^{1/2}\xi_T(\zeta_2,\zeta_2)^{1/2}. \tag{82}$$

Proof. We find $\sum_{n=0}^{t-1}\zeta_{in}Z_{i,t-n}^{+}=\sum_{h=0}^{t-1}(\zeta_i*\xi_i)_h\varepsilon_{i,t-h}$, where $(\zeta_i*\xi_i)_h=\sum_{n=0}^{h}\zeta_{i,h-n}\xi_{in}$, and

$$\xi_T((\zeta_1*\xi_1),(\zeta_2*\xi_2))\leq\max_{1\leq n_1,n_2\leq T}\sum_{h=\max(n_1,n_2)}^{T+\min(n_1,n_2)}\sum_{n=0}^{h-n_1}|\zeta_{1,h-n_1-n}\xi_{1n}|\sum_{m=0}^{h-n_2}|\zeta_{2,h-n_2-m}\xi_{2m}|$$
$$\leq c\sum_{m=0}^{\infty}|\xi_{2m}|\sum_{n=0}^{\infty}|\xi_{1n}|\xi_T(\zeta_1,\zeta_2)\leq c\xi_T(\zeta_1,\zeta_2) \tag{83}$$

because $\sum_{n=0}^{\infty}|\xi_{in}|<\infty$. Thus, it is enough to prove the results for $Z_{it}^{+}=\varepsilon_{it}$ or $\xi_{in}=1_{\{n=0\}}$.

Proof of (79): Using the notation $v_{si}=t_i-n_{si}$, where $s=1,2$ and $i=1,\dots,4$, we find

$$E(T^{-1}\sum_{t=N+1}^{T}\overline{Z}_{1t}^{(N)}\overline{Z}_{2t}^{(N)})^4=T^{-4}\sum_{(1)}(\prod_{i=1}^{4}\zeta_{1,t_i-v_{1i}}\zeta_{2,t_i-v_{2i}})E(\prod_{i=1}^{4}\varepsilon_{1,v_{1i}}\varepsilon_{2,v_{2i}}), \tag{84}$$

where $\sum_{(1)}$ is the sum over $N\leq n_{1i}=t_i-v_{1i},n_{2i}=t_i-v_{2i}<t_i\leq T,\ i=1,\dots,4$. We first sum over t_i for fixed (v_{1i},v_{2i}). Note that $t_i\geq N+v_{1i},t_i\geq N+v_{2i}$ and hence $t_i\geq N+\max(v_{1i},v_{2i})\geq\max(v_{1i},v_{2i})$. Similarly $t_i\leq T+\min(v_{1i},v_{2i})$. This gives the summation

$$\sum_{t_i=N+\max(v_{1i},v_{2i})}^{T+\min(v_{1i},v_{2i})}|\zeta_{1,t_i-v_{1i}}\zeta_{2,t_i-v_{2i}}|\leq\sum_{t_i=\max(v_{1i},v_{2i})}^{T+\min(v_{1i},v_{2i})}|\zeta_{1,t_i-v_{1i}}\zeta_{2,t_i-v_{2i}}|\leq\xi_T(\zeta_1,\zeta_2),$$

and summing over v_{1i},v_{2i} we get the bound

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

where $\sum_{(2)}$ is the summation over $1\leq v_{1i},v_{2i}\leq T-N,\ i=1,\dots,4$. The expectation is zero unless for each (l,i) there is a (k,j) for which $v_{li}=v_{kj}$ so the indices are equal in groups. The smallest number of restrictions, and hence the largest number of summations, occurs if the indices are equal in pairs. This leaves four summations from 1 to $T-N$ and hence a factor of $(T-N)^4$, and therefore the bound $c\xi_T(\zeta_1,\zeta_2)^4$.

Proof of (80): For $N=0$ we get from (79) that $||T^{-1}\sum_{t=1}^{T}Z_{1t}^{+}Z_{2t}^{+}||_4\leq c\xi_T(\zeta_1,\zeta_2)$. We apply this for coefficients for which $\zeta_{1n}=\zeta_{2n}=0,n\geq N$, so that $\xi_T(\zeta_1,\zeta_2)=\xi_N(\zeta_1,\zeta_2)$. We also have $Z_{it}^{+}=\sum_{n=0}^{t-1}\zeta_{it}\varepsilon_{t-n}=\sum_{n=0}^{N-1}\zeta_{it}\varepsilon_{t-n}=\underline{Z}_{it}^{(N)}$ for $t>N$. Thus from

$$T^{-1}\sum_{t=1}^{T}Z_{1t}^{+}Z_{2t}^{+}=T^{-1}\sum_{t=N+1}^{T}\underline{Z}_{1t}^{(N)}\underline{Z}_{2t}^{(N)}+(N/T)N^{-1}\sum_{t=1}^{N}Z_{1t}^{+}Z_{2t}^{+}$$

we find

$$||T^{-1}\sum_{t=N+1}^{T}\underline{Z}_{1t}^{(N)}\underline{Z}_{2t}^{(N)}||\leq c(\xi_N(\zeta_1,\zeta_2)+(N/T)\xi_N(\zeta_1,\zeta_2))\leq c\xi_N(\zeta_1,\zeta_2).$$

Proof of (81): The expression (84) now becomes

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

where $\sum_{(1)}$ is the sum over $0\leq n_{1i}=t_i-v_{1i},n_{2i}=t_i-v_{2i}\leq N<t_i\leq T,i=1,\dots,4$. In this case the bounds for t_i are $t_i\geq\max(v_{1i},v_{2i})$ and $t_i\leq N+\min(v_{1i},v_{2i})$ and $t_i\leq T$. Summing

over t_i we therefore get the factor

$$\sum_{t_i=\max(v_{1i},v_{2i})}^{\min(T,N+\min(v_{1i},v_{2i}))} |\zeta_{1,t_i-v_{1i}}\zeta_{2,t_i-v_{2i}}| \leq \sum_{t_i=\max(v_{1i},v_{2i})}^{N+\min(v_{1i},v_{2i})} |\zeta_{1,t_i-v_{1i}}\zeta_{2,t_i-v_{2i}}| = \xi_N(\zeta_1,\zeta_2),$$

that is, a factor $\xi_N(\zeta_1,\zeta_2)^4$ when summing over all t_i. For the contribution from the expectation we only consider v_{si} equal in pairs. Note that if $v_{1i} = v_{2i}$ for all i the contribution is zero because of the centering. Thus there exists i so that $v_{1i} \neq v_{2i}$ belong to different pairs and satisfy $0 < |v_{2i} - v_{1i}| = |-n_{2i} + n_{1j}| \leq N$. Hence we sum over $1 \leq v_{1i}, v_{2i} \leq T, i = 1, \ldots, 4$, in pairs with at least one restriction of the form $|v_{2i} - v_{1i}| \leq N$, so we get at most NT^3 terms. We therefore find the bound $(N/T)\xi_N(\zeta_1,\zeta_2)^4$ which proves (81).

Proof of (82): In this case we write (84) as

$$E(T^{-1}\sum_{t=N+1}^{T} \underline{Z}_{1t}^{(N)}\overline{Z}_{2t}^{(N)}))^4 = T^{-4}\sum_{(1)}(\prod_{i=1}^{4}\zeta_{1,n_{1i}}\zeta_{2,n_{2i}})E(\prod_{i=1}^{4}\varepsilon_{1,t_i-n_{1i}}\varepsilon_{2,t_i-n_{2i}}),$$

where the summation $\sum_{(1)}$ is over $0 \leq n_{1i} < N \leq n_{2i} \leq t_i \leq T, i = 1, \ldots, 4$.

We consider $t_i - n_{si}$ equal in pairs, which gives the fewest restrictions. Note, however, that $n_{1i} < N \leq n_{2i}$ implies that $t_i - n_{1i} > t_i - n_{2i}$ for all i, which means that there must exist a $j \neq i$ such that $t_i - n_{1i} = t_j - n_{1j}$ and therefore $|t_i - t_j| = |n_{1i} - n_{1j}| \leq N$, and another $k \neq l$ for which $t_k - n_{2k} = t_l - n_{2l}$ with no restriction on (t_k, t_l). We eliminate $n_{1j} = t_j - t_i + n_{1i}$ and $n_{2l} = t_l - t_k + n_{2k}$ and consider

$$|\sum_{n_{1i}=0}^{N-1}\zeta_{1,n_{1i}}\zeta_{1,t_j-t_i+n_{1i}}||\sum_{n_{2k}=0}^{T-1}\zeta_{2,n_{2k}}\zeta_{2,t_l-t_k+n_{2k}}| \leq \xi_N(\zeta_1,\zeta_1)\xi_T(\zeta_2,\zeta_2).$$

Summing over the two other pairs gives either the same factor or the mixed case,

$$(\sum_{n=0}^{N-1}\zeta_{1,n}\zeta_{2,t_p-t_q+n})^2 \leq \sum_{n=0}^{N-1}\zeta_{1,n}^2\sum_{n=0}^{N-1}\zeta_{2,t_p-t_q+n}^2 \leq \xi_N(\zeta_1,\zeta_1)\xi_T(\zeta_2,\zeta_2),$$

where the first inequality is Cauchy-Schwarz. Finally the summation over $t_i, i = 1, \ldots, 4$, with at least one restriction $|t_i - t_j| \leq N$ gives at most NT^3 terms and we find the bound (82). ■

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 (ζ_{1n},ζ_{2n}). We assume that ζ_1 and ζ_2 satisfy conditions of the type

$$|\zeta_{1,0}^{(a)}| \leq 1,\ |\zeta_{1n}^{(a)}| \leq c(1+\log n)^{m_1}n^{-a-1},\ n \geq 1, \tag{85}$$

$$|\zeta_{1,0}^{(a)*}| \leq 1,\ |\zeta_{1n}^{(a)*}| \leq cT^{a+1/2}(1+|\log\frac{n}{T}|)^{m_1}n^{-a-1},\ n \geq 1, \tag{86}$$

where c does not depend on a or n. We use superscript (a) to indicate the order of magnitude of the bound, but sometimes omit it when that should cause no confusion, and an asterisk to indicate the normalization by $T^{a+1/2}$. Note that (85) and (86) are satisfied by the fractional coefficients and their derivatives, see Lemma A.5.

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

$$\sum_{n=1}^{T} n^{-u-1} \leq 1 + \int_1^T x^{-u-1}dx = 1 + u^{-1}(1-T^{-u}) \leq 1 + \frac{1}{u} \leq 2\kappa^{-1},\ u \geq \kappa, \tag{87}$$

$$\kappa^{-1}(1-T^{-\kappa}) \leq u^{-1}(1-T^{-u}) = \int_1^T x^{-u-1}dx \leq \sum_{n=1}^{T} n^{-u-1},\ u \leq \kappa. \tag{88}$$

Lemma A.7 *For $i=1,2$, let $\zeta_{in}^{(a_i)}$ and $\zeta_{in}^{(a_i)*}$ satisfy (85) and (86) with $|a_i|\leq a_0$. Then:*

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

$$\xi_T(\zeta_1^{(a_1)},\zeta_2^{(a_2)})\leq c\begin{cases}(1+\log T)^{m_1+m_2+1}T^{-a}, & a\leq 0,\\ a^{-1}, & a>0.\end{cases} \tag{89}$$

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

$$\xi_T(\zeta_1^{(a_1)*},\zeta_2^{(a_2)*})\leq c\kappa^{-1}. \tag{90}$$

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

$$\xi_T(\zeta_1^{(a_1)},\zeta_2^{(a_2)*})\leq c(1+\log T)^{m_1+m_2+1}T^{-\min(a,\kappa)}. \tag{91}$$

Proof. In evaluating (78) we focus on terms with $t>\max(m,n)$, because the analysis with $t=m$ or $t=n$ is straightforward.

Proof of (89): For $t>\max(m,n)$ we first apply (85) and therefore bound the summation $\sum_{t=\max(n,m)+1}^{T+\min(n,m)}|\zeta_{1,t-n}^{(a_1)}\zeta_{2,t-m}^{(a_2)}|$ by

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

For $a\leq 0$, we bound the log factors by $(1+\log T)$. If $a_i\leq -1, i=1,2$, we bound $(t-n)^{-a_1-1}(t-m)^{-a_2-1}\leq T^{-a_1-a_2-2}\leq T^{-a-1}$ and the result follows. If $a_1\leq -1, a_2\geq -1$ we bound $(t-n)^{-a_1-1}\leq T^{-a_1-1}$ and find

$$\sum_{t=\max(n,m)+1}^{T+\min(n,m)}(t-n)^{-a_1-1}(t-m)^{-a_2-1}\leq T^{-a_1-1}\sum_{t=\max(n,m)+1}^{T+\min(n,m)}(t-m)^{-a_2-1}\leq c(\log T)T^{-a},$$

and similarly if $a_1\geq -1, a_2\leq -1$. If $a_i\geq -1, i=1,2$, then $(t-n)^{-a_1-1}(t-m)^{-a_2-1}\leq (t-\max(n,m))^{-(a_1+a_2+1)-1}$ and the bound for $\xi_T(\zeta_1^{(a_1)},\zeta_2^{(a_2)})$ follows because

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

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 (87) bounded by

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

Proof of (90): We find that $\xi_T(\zeta_1^{(a_1)*},\zeta_2^{(a_2)*})$ is bounded by a constant times the maximum (over $0\leq |n-m|\leq T$) of

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

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

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

Proof of (91): 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 $T^{-\kappa}T^{\max(-a+\kappa,0)} = T^{-\min(a,\kappa)}$. ■

A.3 Limit theory for product moments of deterministic terms

The next lemma gives results for the impact of deterministic terms generated by initial values and the constant term, see (44), in the models considered, using the bounds in JN (2010, Lemma C.1). For the product moments in the proof of consistency we define, for $d_0 \geq 1/2$,

$$D_{it}(\psi) = \begin{cases} (\Delta_-^{d-b} - \Delta_-^{d})\tilde{X}_t + (\Delta_+^{d-b} - \Delta_+^{d})\mu_{0t}, & i=-1, \\ (\Delta_-^{d+ib} - \Delta_-^{d+kb})\tilde{X}_t + (\Delta_+^{d+ib} - \Delta_+^{d+kb})\mu_{0t}, & i=0,\ldots,k-1, \\ \Delta_-^{d+kb}\tilde{X}_t + \Delta_+^{d+kb}\mu_{0t}, & i=k. \end{cases} \tag{92}$$

In model (3) with $d=b$ and a constant, we replace μ_{0t} by $\mu_{0t} + C_0^*\alpha_0\rho_0'$ in $D_{it}(\psi)$ for $i \geq 0$ and subtract $\Delta_+^d C_0^*\alpha_0\rho_0'$ from $D_{-1,t}(\psi)$. For $d_0 < 1/2$ we leave out the terms involving $\Delta_+^{d+ib}\mu_{0t}$ because we use the representations (17) and (18). For the analysis of the score function we define the deterministic terms

$$d_{0t} = \Pi_{0-}(L)(\tilde{X}_t - X_t) \text{ and } d_{1t} = \mathsf{D}\Pi_{0+}(L)\mu_{0t} + \mathsf{D}\Pi_{0-}(L)\tilde{X}_t, \tag{93}$$

where D^m denotes derivatives with respect to $d+ib$ and $\mathsf{D}\Pi_{0-}(L)$ denotes the derivative of $\Pi_-(L)$ evaluated at the true value. Note that the expression for d_{0t} is the same for models (2) and (3) because for the latter model we find from $\varepsilon_t = \Pi_0(L)X_t + \alpha_0\rho_0'$ that

$$\varepsilon_t(\lambda_0) = \Pi_{0+}(L)X_t + \Pi_{0-}(L)\tilde{X}_t + \alpha_0\rho_0' = \Pi_0(L)X_t + \alpha_0\rho_0' + \Pi_{0-}(L)(\tilde{X}_t - X_t) = \varepsilon_t + \Pi_{0-}(L)(\tilde{X}_t - X_t).$$

The expression for d_{1t} is found as a linear combination of $\mathsf{D}D_{it}(\psi)|_{\psi=\psi_0}$, see (92), and also $T^{1/2-b_0}\beta_{0\perp}'\mathsf{D}D_{-1,t}(\psi)|_{\psi=\psi_0}$ if $b_0 > 1/2$.

Lemma A.8 *We let $\eta > 0$ and $\kappa_1 > 0$, where $\kappa_1 < 1/2$ if $d_0 < 1/2$ and $\kappa_1 < \min(1/2, d_0 - 1/2)$ if $d_0 > 1/2$. It then holds that:*

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

$$\max_{-1/2-\kappa_1 \leq \delta_i \leq u_1} |\mathsf{D}^m D_{it}(\psi)| \to 0 \text{ as } t \to \infty, \tag{94}$$

$$\max_{-u_0 \leq \delta_i \leq -1/2-\kappa_1} \max_{1 \leq t \leq T} |\mathsf{D}^m T^{\delta_i+1/2}\beta_{0\perp}' D_{it}(\psi)| \to 0 \text{ as } T \to \infty. \tag{95}$$

(ii) In model $\mathcal{H}_r(d=b)$ with a constant the same results hold.

(iii) If $X_{-n} = 0, n \geq T^{\upsilon}$, then for both $\mathcal{H}_r$ and for $\mathcal{H}_r(d=b)$ with a constant,

$$T^{-1/2}\sum_{t=1}^{T}|d_{0t}| \to 0 \text{ for } \upsilon < 1/2. \tag{96}$$

Proof. From (10) we have that $\mu_{0t} = -\Pi_{0+}(L)^{-1}\Pi_{0-}(L)X_t$, and from Theorem 2 and $C_0\alpha_0\beta_0' = 0$ we get

$$\mu_{0t} = -(C_0\Delta_+^{-d_0} + \Delta_+^{-d_0+b_0}F_+(L))(-\alpha_0\beta_0'\Delta_-^{d_0-b_0} + \sum_{j=0}^{k}\Psi_{0j}\Delta_-^{d_0+jb_0})X_t \tag{97}$$

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

From JN (2010, Lemma C.1) we have the evaluations

$$|\frac{\partial^m}{\partial u^m}\Delta_+^u\Delta_-^v X_t| \leq c(1+\log t)^{m+1}t^{-\min(u+v,u+1,v)}, \tag{98}$$

$$|\frac{\partial^m}{\partial u^m}T^u\Delta_+^u\Delta_-^v X_t| \leq c(1+\log T)^{m+1}T^{-\min(v,1,v-u,-u)}. \tag{99}$$

We see that differentiating the fractional coefficients gives 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 continue setting $m = 0$.

Proof for $\Delta_-^{d+ib}\tilde{X}_t$ in (i) and (ii): To prove (94) we find from (75) of Lemma A.5 that because $d+ib \geq d-b \geq 0$ we have

$$|\Delta_-^{d+ib}\tilde{X}_t| = |\sum_{j=0}^{N_0-1}\pi_{j+t}(-d-ib)X_{-n}| \leq c|\sum_{j=0}^{N_0-1}(j+t)^{-(d+ib)-1}| \leq cN_0t^{-(d-b)-1}, \tag{100}$$

which proves *(i)* and *(ii)* because $\max_{0\leq d+ib\leq d_1}|\Delta_-^{d+ib}\tilde{X}_t| \leq ct^{-1}$. The proof of (95) for $\Delta_-^{d+ib}\tilde{X}_t$ follows from (100) because $\max_{-u_0\leq\delta_i\leq-1/2-\kappa_1}T^{\delta_i+1/2} \to 0$.

Proof for $\Delta_+^{(i+1)d}C_0^*\alpha_0\rho_0'$ for $i \geq 0$ in (i) and (ii): We find $\Delta_+^{(i+1)d}1 = \Delta^{(i+1)d}1 - \Delta_-^{(i+1)d}1 = -\Delta_-^{(i+1)d}1$, which is bounded by $c\sum_{n=t}^{\infty}n^{-\eta-1} \leq ct^{-\eta}$ uniformly for $(i+1)d \geq d \geq \eta > 0$ which proves both (94) and (95).

Proof for $\Delta_+^{d+ib}\mu_{0t}$ in (i) and (ii): This term is only present if $d_0 > 1/2$ and we only apply the condition $d-b \geq 0$. We first prove (94). The term $\Delta_+^{d+ib}\mu_{0t}$ contains terms of the form $G_+(L)\Delta_+^u\Delta_-^v X_t$ with $G(z) = \sum_{n=0}^{\infty}g_nz^n$ and $\sum_{n=0}^{\infty}|g_n| < \infty$, and where $u = d+ib-\gamma_0$ and $v = d_0+jb_0 \geq \gamma_0$ with $\gamma_0 = d_0$ or $\gamma_0 = d_0-b_0$, see (97). Because $\delta_i = d+ib-d_0 \geq -1/2-\kappa_1$ in (94), then for both choices of γ_0 we find $u+v \geq d+ib \geq d_0-1/2-\kappa_1, v \geq \gamma_0$, and $u+1 \geq d_0+1/2-\kappa_1-\gamma_0 \geq 1/2-\kappa_1$ so that from (98) we get for $d_0 > b_0$ that

$$|\Delta_+^{d+ib-\gamma_0}\Delta_-^{d_0+jb_0}X_t| \leq c(1+\log t)t^{-\min(d_0-1/2-\kappa_1,1/2-\kappa_1,d_0-b_0)} \to 0.$$

The Dominated Convergence Theorem shows the same result for $G_+(L)\Delta_+^{d+ib-\gamma_0}\Delta_-^{d_0+jb_0}X_t$, and (94) follows for $\Delta_+^{d+ib}\mu_{0t}$ when $d_0 > b_0$.

If $d_0 = b_0$ then $\Delta_+^0\Delta_-^0 X_t = 0$ and (97) implies

$$\Delta_+^{d+ib}\mu_{0t} = -\sum_{j=0}^{k}(C_0\Psi_{0j}\Delta_+^{d+ib-d_0} + F_+(L)\Psi_{0j}\Delta_+^{d+ib})\Delta_-^{d_0+jb_0}X_t.$$

To prove (94) with $d_0 = b_0$ we take $u = d+ib-\gamma_0$ where $\gamma_0 = d_0$ or 0 and $v = d_0+jb_0 \geq d_0$ and find from (98) for $d+ib \geq d_0-1/2-\kappa_1 > 0$ that

$$|\Delta_+^{d+ib-\gamma_0}\Delta_-^{d_0+jb_0}X_t| \leq c(1+\log t)t^{-\min(d_0-1/2-\kappa_1-\gamma_0+d_0,d_0+1/2-\kappa_1-\gamma_0,d_0)} \to 0.$$

To prove (95) we take $l \geq i$ and apply (99) with $u = d+lb-\gamma_0$ and $v = d_0+jb_0 \geq \gamma_0 \geq 0$. Because $u+v \geq d-b \geq 0$ and $v \geq 0$ imply $v \geq -u$ and $v-u \geq -u$ we have $\min(v,1,v-u,-u) = \min(1,-u)$ and thus

$$|T^{d+ib-d_0+1/2}\Delta_+^{d+lb-\gamma_0}\Delta_-^{d_0+jb_0}X_t| = T^{(i-l)b+1/2-d_0+\gamma_0}|T^{d+lb-\gamma_0}\Delta_+^{d+lb-\gamma_0}\Delta_-^{d_0+jb_0}X_t|$$
$$\leq c(1+\log T)T^{\max(-1/2+(i-l)b-d_0+\gamma_0,d+ib+1/2-d_0)}$$

$$\leq c(1+\log T)T^{\max(-1/2,-\kappa_1)} \to 0$$

using $d+ib-d_0+1/2 \leq -\kappa_1$. If we apply this for $l=i=-1$ and $i=-1, l=0$ then we find the result for $(\Delta_+^{d-b}-\Delta_+^{d})\mu_{0t}$. With $l=i, l=k$ and $l=i=k$ we find the result for $(\Delta_+^{d+ib}-\Delta_+^{d+kb})\mu_{0t}$ and $\Delta_+^{d+kb}\mu_{0t}$, respectively.

Proof of (iii): The deterministic term $d_{0t}=\Pi_{0-}(L)(\tilde{X}_t-X_t)$ only depends on $X_{-n}, n \geq N_0$, because $\tilde{X}_{-n}=X_{-n}, n<N_0$. We find the terms $\Delta_-^{d_0+ib_0}X_t, i \geq -1$, which are bounded by $cT^{v}t^{-1-(d_0-b_0)}$, see (100). It follows that $T^{-1/2}\sum_{t=1}^{T}|d_{0t}| \to 0$ for $v<1/2$.[3] ■

A.4 Limit theory for product moments of stochastic terms

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 define $\mathcal{S}(\kappa_w, \underline{\kappa}_v, \bar{\kappa}_v, \kappa_u)$ as the set where the three fractional indices w, v, and u are in the intervals*

$$[-w_0, -1/2-\kappa_w],\ [-1/2-\underline{\kappa}_v, -1/2+\bar{\kappa}_v],\ [-1/2+\kappa_u, u_0], \tag{101}$$

respectively, and where we assume $0 \leq \bar{\kappa}_v < \underline{\kappa}_v$ and $0 < \underline{\kappa}_v < \min(b_0/3, \kappa_w/2, \kappa_u/2, 1/6)$.

In the following we assume these bounds on (u, v, w). Thus for $Z_t \in \mathcal{Z}_b, b>0$, see Definition 2, 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 applications we always choose fixed values of $\underline{\kappa}_v$, κ_u, and κ_w, but we shall sometimes choose small values $(\to 0)$ of $\bar{\kappa}_v$.

In the subsequent lemmas we derive results for product moments of fractional differences of processes in the class $\mathcal{Z}_{b_0}$, see Definition 2, or the deterministic term. For $m=m_1+m_2$ we define the product moments

$$\mathsf{D}^m M_T(a_1,a_2) = T^{-1}\sum_{t=1}^{T}(\mathsf{D}^{m_1}\Delta_+^{a_1}Z_{1t}^+)(\mathsf{D}^{m_2}\Delta_+^{a_2}Z_{2t}^+)',$$

$$M_T((a_1,a_2),(a_1,a_2)) = T^{-1}\sum_{t=1}^{T}\begin{pmatrix}\Delta_+^{a_1}Z_{1t}^+\\ \Delta_+^{a_2}Z_{2t}^+\end{pmatrix}\begin{pmatrix}\Delta_+^{a_1}Z_{1t}^+\\ \Delta_+^{a_2}Z_{2t}^+\end{pmatrix}',$$

$$M_T(a_1,a_2|a_3) = M_T(a_1,a_2) - M_T(a_1,a_3)M_T^{-1}(a_3,a_3)M_T(a_3,a_2),$$

where a_1, a_2, a_3 can be u, w, and v in the intervals in Definition A.1, or they can be the constant one, in which case the notation $M_T(1,a_2)$ means that $\Delta_+^{a_1}Z_{1t}^+$ has been replaced by 1. 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 $\mathcal{K}$. Finally, we introduce the notation $M_T^{**}(w_1,w_2)=T^{w_1+w_2+1}M_T(w_1,w_2)$ and $M_T^{*}(w,a)=T^{w+1/2}M_T(w,a)$, where a can be u, v, or 1, to indicate that the nonstationary processes have been normalized by $T^{w_i+1/2}$.

[3] Under the alternative assumption $\sum_{n=1}^{\infty}n^{-1/2}|X_{-n}|<\infty$ (replacing $X_{-n}=0$ for $n \geq T^{v}$) the argument is

$$|\Delta_-^{d_0+ib_0}X_t| \leq c\sum_{n=0}^{\infty}(n+t)^{-1-(d_0-b_0)}|X_{-n}| \leq ct^{-1/2-(d_0-b_0)}\sum_{n=0}^{\infty}n^{-1/2}|X_{-n}| \leq ct^{-1/2-(d_0-b_0)},$$

such that $T^{-1/2}\sum_{t=1}^{T}t^{-1/2-(d_0-b_0)} \leq cT^{-(d_0-b_0)} \to 0$ for $d_0>b_0$. If $d_0=b_0$ then $\Delta_-^{d_0-b_0}X_t=\Delta_-^{0}X_t=0$ for $t \geq 1$ and the dominating term becomes $T^{-1/2}\sum_{t=1}^{T}|\Delta_-^{d_0}X_t| \leq cT^{\max(-1/2,-d_0)} \to 0$.

Lemma A.9 *Let* $Z_{it} = \xi_i \varepsilon_t + \Delta^{b_0} \sum_{n=0}^{\infty} \xi_{in}^* \varepsilon_{t-n} \in \mathcal{Z}_{b_0}, i = 1, 2,$ *and define* $M_T(a_1, a_2)$ *as above and assume that* $E|\varepsilon_t|^q < \infty$ *for some* $q > \kappa_w^{-1}$ *and* $q \geq 8$. *Then it holds jointly that:*

(i) Uniformly for $-w_0 \leq w \leq -1/2 - \kappa_w$ *and* $-1/2 + \kappa_u \leq u \leq u_0$ *we find that*

$$\mathsf{D}^m M_T(u_1, u_2) \Longrightarrow \mathsf{D}^m E(\Delta^{u_1} Z_{1t})(\Delta^{u_2} Z_{2t})', \tag{102}$$

$\mathsf{D}^m M_T^{**}(w_1, w_2)$ *is tight, and*

$$M_T^{**}(w_1, w_2) \Longrightarrow \xi_1 \int_0^1 W_{-w_1-1}(s) W_{-w_2-1}(s)' ds \xi_2', \tag{103}$$

$$\mathsf{D}^m M_T^*(w, u) = \mathbf{O}_P((1 + \log T)^{2+m} T^{-\min(\kappa_u, \kappa_w)}). \tag{104}$$

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

$$M_T^*(w, v) = \boldsymbol{O}_P((1 + \log T)^2 T^{\underline{\kappa}_v}), \tag{105}$$

$$M_T(v, u) = \boldsymbol{O}_P(1). \tag{106}$$

(ii) If we choose $N = T^\alpha$ *with* $0 < \alpha < 1/4$, *and* (ξ_1', ξ_2') *has full rank, then for* $-1/2 - \underline{\kappa}_v \leq v_i \leq -1/2 + \overline{\kappa}_v$ *we find*

$$M_T((v_1, v_2), (v_1, v_2)) \geq c \frac{1 - N^{-2\overline{\kappa}_v}}{2\overline{\kappa}_v} (\xi_1', \xi_2')' \Omega_0 (\xi_1', \xi_2') + R_T, \tag{107}$$

where $R_T = \mathbf{o}_P(1)$ *uniformly for* $|v_i + 1/2| \leq \underline{\kappa}_v$.

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 (74) 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 Lemma A.6, for suitable coefficients satisfying (85) and (86). 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 (102): We define the coefficients $\zeta_{i,t-n} = \pi_{t-n}(-u_i)$, which satisfy condition (85). 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 (79) with $N = 0$ and (89) 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

$$||T^{-1} \sum_{t=1}^{T} (\Delta_+^{u_1} Z_{1t}^+ - \Delta_+^{\tilde{u}_1} Z_{1t}^+)(\Delta_+^{u_2} Z_{2t}^+)'||_4 + ||T^{-1} \sum_{t=1}^{T} (\Delta_+^{\tilde{u}_1} Z_{1t}^+)(\Delta_+^{u_2} Z_{2t}^+ - \Delta_+^{\tilde{u}_2} Z_{2t}^+)'||_4. \tag{108}$$

We apply (79) with $N = 0$ 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 (85), see also JN (2010, Lemma B.3), and it follows from (89) with $a = 2\kappa_u$ that the first term of (108) is bounded by $c|u_1 - \tilde{u}_1|$. A similar proof works for the other term of (108), and tightness then follows from (74).

Notice that the second condition of (74) follows in the same way as the first using the inequalities in Lemma A.7. 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 $\Delta_+^{u_i} Z_{it}^+ = \sum_{h=0}^{t-1} (\pi(-u_i) * \xi_i)_h \varepsilon_{t-h}$ and $\Delta^{u_i} Z_{it} = \sum_{h=0}^{\infty} (\pi(-u_i) * \xi_i)_h \varepsilon_{t-h}$ we see that it is enough that the variance of the difference converges uniformly to zero,

$$Var(\Delta^{u_i} Z_{it} - \Delta_+^{u_i} Z_{it}^+) = \sum_{h=t}^{\infty} (\pi(-u_i) * \xi_i)_h \Omega (\pi(-u_i) * \xi_i)_h \to 0 \text{ as } t \to \infty.$$

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 (103): We define $\zeta^*_{i,t-n}(w_i) = T^{w_i+1/2}\pi_{t-n}(-w_i)$ for $w_i \le -1/2 - \kappa_w$ so that $\max(w_1, w_2, w_1 + w_2 + 1) \le -2\kappa_w < 0$. We then apply (79) with $N = 0$ and (90) with $\kappa = 2\kappa_w$ and find that (74) holds and $M^{**}_T(w_1, w_2)$ is tight. Because $-1/(w_i + 1/2) \le \kappa_w^{-1} < q$ we get the limit

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

see (6) and also JN (2010, Lemma D.2) for a few more details. The Continuous Mapping Theorem gives the result (103).

Proof of (104): We apply (79) with $N = 0$ and (91) 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 $w \le -1/2 - \kappa_w, u \ge -1/2 + \kappa_u, a = \kappa_u$, and $\kappa = \kappa_w$ that

$$||M^*_T(w, u)||_4 \le c(1 + \log T)T^{-\min(\kappa_u,\kappa_w)},$$
$$||M^*_T(w, u) - M^*_T(\tilde{w}, \tilde{u})||_4 \le c|(w, u) - (\tilde{w}, \tilde{u})|(1 + \log T)^2 T^{-\min(\kappa_u,\kappa_w)},$$

and (74) implies that $M^*_T(w, u) = \mathbf{O}_P((1 + \log T)^2 T^{-\min(\kappa_u,\kappa_w)})$. The extra $(1 + \log T)$ in the increment is due to JN (2010, Lemma B.3, eqn (56)).

Proof of (105): We first apply (79) with $N = 0$, $\zeta_{1,t-n} = \pi_{t-n}(-v)$, and $\zeta^*_{2,t-n} = T^{w+1/2}\pi_{t-n}(-w)$ and find from (91) with $a = -\underline{\kappa}_v, \kappa = \kappa_w$ that for $v \ge -1/2 - \underline{\kappa}_v, w \le -1/2 - \kappa_w$ we get

$$||M^*_T(w, v)||_4 \le c(1 + \log T)T^{\underline{\kappa}_v}, \tag{109}$$
$$||M^*_T(w, v) - M^*_T(\tilde{w}, \tilde{v})||_4 \le c|(w, v) - (\tilde{w}, \tilde{v})|(1 + \log T)^2 T^{\underline{\kappa}_v},$$

and (74) then shows that $M^*_T(w, v) = \mathbf{O}_P((1 + \log T)^2 T^{\underline{\kappa}_v})$.

Proof of (106): We define $\zeta_{1,t-n} = \pi_{t-n}(-u)$ and $\zeta_{2,t-n} = \pi_{t-n}(-v)$ where $v \ge -1/2 - \underline{\kappa}_v$ and $u \ge -1/2 + \kappa_u$, so that $\min(u + 1, v + 1, u + v + 1) \ge \min(\kappa_u, 1/2) - \underline{\kappa}_v > 0$, see Definition A.1. It then follows from (79) with $N = 0$ and (89) that (74) is satisfied and hence that $M_T(u, v)$ is tight.

Proof of (107): Because we need to decompose the processes we use the notation

$$P_T(U_1, U_2) = T^{-1}\sum_{t=1}^{T} U^+_{1t}U^{+\prime}_{2t} \text{ and } P_{T,N}(U_1, U_2) = T^{-1}\sum_{t=N+1}^{T} U^+_{1t}U^{+\prime}_{2t}$$

for product moments of any processes U_{1t} and U_{2t}. We define $\tilde{Z}^+_{it}$ by $Z^+_{it} = \xi_i\varepsilon_t + \Delta_+^{b_0}\tilde{Z}^+_{it}, i = 1, 2$, $\xi = blockdiag(\xi_1, \xi_2)$, $\Delta_+^v Z^+_t = (\Delta_+^{v_1} Z^{+\prime}_{1t}, \Delta_+^{v_2} Z^{+\prime}_{2t})'$, $\Delta_+^v \tilde{Z}^+_t = (\Delta_+^{v_1}\tilde{Z}^{+\prime}_{1t}, \Delta_+^{v_2}\tilde{Z}^{+\prime}_{2t})'$, and $\Delta_+^v\varepsilon_t = (\Delta_+^{v_1}\varepsilon'_t, \Delta_+^{v_2}\varepsilon'_t)'$, and find the evaluation

$$P_T(\Delta_+^v Z, \Delta_+^v Z) \ge \xi P_T(\Delta_+^v\varepsilon, \Delta_+^v\varepsilon)\xi' + P_T(\Delta_+^{b_0+v}\tilde{Z}, \Delta_+^v\varepsilon)\xi' + \xi P_T(\Delta_+^v\varepsilon, \Delta_+^{b_0+v}\tilde{Z}), \tag{110}$$

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

We define the index $u_i = v_i + b_0 \ge -1/2 + (b_0 - \underline{\kappa}_v)$ 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 (106) for the components of $P_T(\Delta_+^{b_0+v}\tilde{Z}, \Delta_+^v\varepsilon)$ and its transposed which are therefore $\mathbf{O}_P(1)$.

We next consider $P_T(\Delta_+^v\varepsilon, \Delta_+^v\varepsilon) \ge P_{T,N}(\Delta_+^v\varepsilon, \Delta_+^v\varepsilon)$ and decompose, for $t > N = T^\alpha$,

$$V^+_{it} = \Delta_+^{v_i}\varepsilon_t = \underline{V}^{(N)}_{it} + \overline{V}^{(N)}_{it} = \sum_{n=0}^{N-1}\pi_n(-v_i)\varepsilon_{t-n} + \sum_{n=N}^{t-1}\pi_n(-v_i)\varepsilon_{t-n}. \tag{111}$$

We define $\underline{V}^{(N)}_t = (\underline{V}^{(N)\prime}_{1t}, \underline{V}^{(N)\prime}_{2t})'$, $\overline{V}^{(N)}_t = (\overline{V}^{(N)\prime}_{1t}, \overline{V}^{(N)\prime}_{2t})'$ and evaluate the product moment

$$P_{T,N}(\Delta_+^v\varepsilon, \Delta_+^v\varepsilon) \ge P_{T,N}(\underline{V}^{(N)}, \underline{V}^{(N)}) + P_{T,N}(\underline{V}^{(N)}, \overline{V}^{(N)}) + P_{T,N}(\overline{V}^{(N)}, \underline{V}^{(N)}). \tag{112}$$

Analysis of $P_{T,N}(\underline{V}^{(N)},\overline{V}^{(N)})$: It follows from (89) for $a_i = v_i \geq -1/2-\underline{\kappa}_v$ that $\xi_T(\zeta_2,\zeta_2) \leq c(1+\log T)T^{\underline{\kappa}_v}$ and $\xi_N(\zeta_1,\zeta_1) \leq c(1+\log N)N^{\underline{\kappa}_v}$, so that (82) implies

$$||P_{T,N}(\underline{V}^{(N)},\overline{V}^{(N)})||_4 \leq c(1+\log T)T^{-(1-2\underline{\kappa}_v)/4+\alpha(1+2\underline{\kappa}_v)/4}, \tag{113}$$

which converges to zero for $\alpha < 1/4$ and $\underline{\kappa}_v < 1/6$ because $-(1-2\underline{\kappa}_v)/4+\alpha(1+2\underline{\kappa}_v)/4 < 0$.

To prove tightness we check condition (74). We take two points (v_1,v_2) and $(\tilde{v}_1,\tilde{v}_2)$. For convenience we introduce the notation $M_{T,N}(v_1,v_2) = P_{T,N}(\underline{V}^{(N)},\overline{V}^{(N)})$ to emphasize the dependence on (v_1,v_2). Then the difference $M_{T,N}(v_1,v_2) - M_{T,N}(\tilde{v}_1,\tilde{v}_2)$ contains differences like $\pi_{n_1}(-v_1)\pi_{n_2}(-v_2) - \pi_{n_1}(-\tilde{v}_1)\pi_{n_2}(-\tilde{v}_2)$, which we can write as

$$(\pi_{n_1}(-v_1) - \pi_{n_1}(-\tilde{v}_1))\pi_{n_2}(-\tilde{v}_2) + \pi_{n_1}(-v_1)(\pi_{n_2}(-v_2) - \pi_{n_2}(-\tilde{v}_2)),$$

where the first term is, by the Mean Value Theorem,

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

for some intermediate value v_1^*. Here ζ_{1n_1} and ζ_{2n_2} satisfy (85) with $a_i = v_i \geq -1/2-\underline{\kappa}_v$ and $m_1 = 1, m_2 = 0$. Therefore we have from (82) and (89) that

$$||\sum_{t=N+1}^{T}\sum_{1\leq n_1,n_2<N}(\pi_{n_1}(-v_1) - \pi_{n_1}(-\tilde{v}_1))\pi_{n_2}(-\tilde{v}_2)\varepsilon_{1,t-n_1}\varepsilon_{2,t-n_2}||_4 \leq c_T|v_1-\tilde{v}_1|,$$

where $c_T \to 0$, see (113), and a similar expression for the other term. This shows that

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

and hence that $M_{T,N}(v_1,v_2) = P_{T,N}(\underline{V}^{(N)},\overline{V}^{(N)})$ is tight and therefore $\mathbf{o}_P(1)$ by (113).

Analysis of $P_{T,N}(\underline{V}^{(N)},\underline{V}^{(N)})$: We define for $-1/2-\underline{\kappa}_v \leq v_i \leq -1/2+\overline{\kappa}_v$ the coefficient

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

where the dependence on $\overline{\kappa}_v$ appears for the first time, see Lemma A.5*(ii)* and (88). Note that $F_{Nij} \to \infty$ as $(\overline{\kappa}_v, N) \to (0,\infty)$. We find that

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

The difference $R_T(v_1,v_2) = P_{T,N}(\underline{V}^{(N)},\underline{V}^{(N)}) - E(P_{T,N}(\underline{V}^{(N)},\underline{V}^{(N)})) \Longrightarrow 0$ uniformly for $|v_i + 1/2| \leq \underline{\kappa}_v$ by (81) of Lemma A.6 and (89) because

$$||R_T(v_1,v_2)||_4 \leq c(N/T)^{1/4}\xi_N(\zeta_1,\zeta_2) \leq cT^{-1/4}(1+\log N)N^{1/4+\underline{\kappa}_v} \to 0$$

for $\alpha < 1/4$ and $\underline{\kappa}_v < 1/6$ because $-1/4+\alpha(1/4+\underline{\kappa}_v) < 0$. Tightness follows as for $M_{T,N}(v_1,v_2)$ in the analysis of $P_{T,N}(\underline{V}^{(N)},\underline{V}^{(N)})$. Hence

$$\xi P_{T,N}(\underline{V}^{(N)},\underline{V}^{(N)})\xi' \geq c\frac{1-N^{-2\overline{\kappa}_v}}{2\overline{\kappa}_v}(\xi_1',\xi_2')'\Omega_0(\xi_1',\xi_2') + \mathbf{o}_P(1),$$

where the remainder term is uniformly small for $|v_i+1/2| \leq \underline{\kappa}_v$ independently of $\overline{\kappa}_v$. From (110) and multiplying (112) by ξ and ξ' we find (107). ■

We apply the results of Lemma A.9 and Corollary A.10 in the analysis of $\ell_{T,p}(\psi)$ and $\ell_{T,r}(\psi)$ to show that they converge, which is the key ingredient in the proof of consistency of the MLE and in the test for rank. The results for $m = 0,1,2$ in Lemma A.9 are used to show that the information matrix is tight in a neighborhood of the true value and the results are summarized for $\mathcal{A}_T(\psi), \mathcal{B}_T(\psi)$, and $\mathcal{C}_T(\psi)$ in Theorem 6.

For the proof of existence and consistency of the MLE we need the product moments that enter the likelihood function $\ell_{T,p}(\psi)$, which are analyzed in Corollaries A.10 and A.12 to follow.

Corollary A.10 *If the assumptions of Lemma A.9 hold, then uniformly for* $(w, v, u) \in \mathcal{S}(\kappa_w, \underline{\kappa}_v, \bar{\kappa}_v, \kappa_u)$, *see (101) of Definition A.1:*

(i) It holds that

$$M_T^{**}(w_1, w_2|w_3, u) = M_T^{**}(w_1, w_2|w_3) + \boldsymbol{o}_P(1), \tag{114}$$

$$M_T(u_1, u_2|w, u_3) \Longrightarrow Var(\Delta^{u_1} Z_{1t}, \Delta^{u_2} Z_{2t}|\Delta^{u_3} Z_{3t}), \tag{115}$$

$$M_T(v, u_1|w, u_2) = \mathbf{O}_P(1). \tag{116}$$

(ii) If $N = T^\alpha$ *with* $0 < \alpha < 1/4$, *and* (ξ_1', ξ_2') *has full rank, then*

$$M_T((v_1, v_2), (v_1, v_2)|w, u) \geq c\frac{1 - N^{-2\overline{\kappa}_v}}{2\overline{\kappa}_v}(\xi_1', \xi_2')'\Omega_0(\xi_1', \xi_2') + R_T, \tag{117}$$

where $R_T = \mathbf{O}_P(1)$ *uniformly for* $|v_i + 1/2| \leq \underline{\kappa}_v$.

Proof. *Proof of (i)*: The proofs of (114), (115), and (116) are the same, so we give only the latter. We decompose $M_T(v, u_1|w, u_2)$ as

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

where the second term is

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

because $M_T^*(w, u_2) \Longrightarrow 0$ by (104). The result follows by application of Lemma A.9.

Proof of (ii): The proof is similar to that of (107) except for conditioning on a stationary and a nonstationary variable. We start by eliminating the stationary variable and find that $M_T((v_1, v_2), (v_1, v_2)|w, u)$ is

$$M_T((v_1, v_2), (v_1, v_2)|w) - M_T((v_1, v_2), u|w) M_T(u, u|w)^{-1} M_T(u, (v_1, v_2)|w),$$

where $M_T(u, u|w)^{-1} = \mathbf{O}_P(1)$ by (115) and $M_T((v_1, v_2), u|w) = \mathbf{O}_P(1)$ by (116), and we therefore continue with $M_T((v_1, v_2), (v_1, v_2)|w)$.

We decompose $\Delta_+^{v_i} Z_{it}^+ = \xi_i \Delta_+^{v_i} \varepsilon_t + \Delta_+^{v_i + b_0} \tilde{Z}_{it}^+$, $i = 1, 2$, and define $Z_t^+ = (Z_{1t}^{+\prime}, Z_{2t}^{+\prime})'$ and $\tilde{Z}_t^+ = (\tilde{Z}_{1t}^{+\prime}, \tilde{Z}_{2t}^{+\prime})'$. Then we have the evaluation

$$\begin{aligned} M_T((v_1, v_2), (v_1, v_2)|w) &\geq \xi P_T(\Delta^v \varepsilon, \Delta^v \varepsilon|\Delta^w Z_3)\xi' \\ &\quad + \xi P_T(\Delta^v \varepsilon, \Delta^{b_0 + v} \tilde{Z}|\Delta^w Z_3)\xi' + \xi P_T(\Delta^{b_0 + v} \tilde{Z}, \Delta^v \varepsilon|\Delta^w Z_3)\xi'. \end{aligned}$$

It follows from (116) for $u_i = v_i + b_0 \geq -1/2 + b_0 - \underline{\kappa}_v$ (i.e., $\kappa_u = b_0 - \underline{\kappa}_v > 2\underline{\kappa}_v$) and $w \leq -1/2 - \kappa_w$ that the last two terms are $\mathbf{O}_P(1)$.

We next evaluate $P_T(\Delta^v \varepsilon, \Delta^v \varepsilon|\Delta^w Z_3) \geq P_{T,N}(\Delta^v \varepsilon, \Delta^v \varepsilon|\Delta^w Z_3)$ and decompose $\Delta_+^{v_i} \varepsilon_t = \underline{V}_{it}^{(N)} + \overline{V}_{it}^{(N)}$, see (111), and stack them into $\underline{V}^{(N)}$ and $\overline{V}^{(N)}$. We bound $P_{T,N}(\Delta^v \varepsilon, \Delta^v \varepsilon|\Delta^w Z_3)$ from below by

$$\begin{aligned} &P_{T,N}(\underline{V}^{(N)}, \underline{V}^{(N)}|\Delta^w Z_3) + P_{T,N}(\underline{V}^{(N)}, \overline{V}^{(N)}|\Delta^w Z_3) + P_{T,N}(\overline{V}^{(N)}, \underline{V}^{(N)}|\Delta^w Z_3) \\ &= P_{T,N}(\underline{V}^{(N)}, \underline{V}^{(N)}) - R_{1T} + R_{2T} + R_{2T}', \end{aligned}$$

where

$$R_{1T} = P^*_{T,N}(\underline{V}^{(N)}, \Delta^w Z_3) P^{**}_T(\Delta^w Z_3, \Delta^w Z_3)^{-1} P^*_{T,N}(\Delta^w Z_3, \underline{V}^{(N)}),$$
$$R_{2T} = P_{T,N}(\underline{V}^{(N)}, \overline{V}^{(N)}) - P^*_{T,N}(\underline{V}^{(N)}, \Delta^w Z_3) P^{**}_T(\Delta^w Z_3, \Delta^w Z_3)^{-1} P^*_{T,N}(\Delta^w Z_3, \overline{V}^{(N)}),$$

and asterisks denote that nonstationary processes have been normalized as for M^*_T and M^{**}_T. We next show that, for $N = T^\alpha$,

$$P^*_{T,N}(\underline{V}^{(N)}, \Delta^w Z_3) = \mathbf{O}_P((1+\log T)^2 (T^{3\alpha\underline{\kappa}_v - 2\underline{\kappa}_v} + T^{-1/4+\alpha(1+2\underline{\kappa}_v)/4}), \tag{118}$$
$$P^*_{T,N}(\overline{V}^{(N)}, \Delta^w Z_3) = \mathbf{O}_P((1+\log T)^2 T^{\underline{\kappa}_v}). \tag{119}$$

If these were proved and $\alpha < 1/4$ and $\underline{\kappa}_v < 1/6$, it follows that R_{1T} and R_{2T} are $\mathbf{o}_P(1)$, see also (113). Thus, proving (118) and (119) completes the proof of (ii), see (107) for the main term $P_{T,N}(\underline{V}^{(N)}, \underline{V}^{(N)})$.

Proof of (118): We decompose $\Delta^w_+ Z_3 = \underline{W}^{(N)}_t + \overline{W}^{(N)}_t = \sum_{n=0}^{N-1} \pi_n(-w) Z^+_{3,t-n} + \sum_{n=N}^{t-1} \pi_n(-w) Z^+_{3,t-n}$, and evaluate

$$||P^*_{T,N}(\underline{V}^{(N)}, \Delta^w Z_3)||_4 \le ||P^*_{T,N}(\underline{V}^{(N)}, \underline{W}^{(N)})||_4 + ||P^*_{T,N}(\underline{V}^{(N)}, \overline{W}^{(N)})||_4.$$

From (80) and (91) with $a = -\underline{\kappa}_v$ and $\kappa = \kappa_w$ we find that $||P^*_{T,N}(\underline{V}^{(N)}, \underline{W}^{(N)})||_4$ is bounded by

$$\begin{aligned} c\xi_N(\zeta_1^{(v_1)}, \zeta_2^{(w)}) T^{w+1/2} &\le c(N/T)^{\kappa_w} \xi_N(\zeta_1^{(v_1)}, \zeta_2^{(w)*}) \\ &\le c(1+\log T) T^{-\kappa_w + \alpha(\kappa_w + \underline{\kappa}_v)} \le c(1+\log T) T^{-2\underline{\kappa}_v + 3\alpha\underline{\kappa}_v}, \end{aligned}$$

using $\kappa_w > 2\underline{\kappa}_v$ and where $\xi_N(\zeta_1^{(v_1)}, \zeta_2^{(w)*})$ denotes $\xi_N(\zeta_1^{(v_1)}, \zeta_2^{(w)})$ normalized by $N^{w+1/2}$. Similarly $||P^*_{T,N}(\underline{V}^{(N)}, \overline{W}^{(N)})||_4$ is, by (82), (90), and (91), bounded by

$$c(N/T)^{1/4} \xi_N(\zeta_1^{(v_1)}, \zeta_1^{(v_1)})^{1/2} \xi_T(\zeta_2^{(w)*}, \zeta_2^{(w)*})^{1/2} \le c(1+\log T) T^{-1/4+\alpha(1+2\underline{\kappa}_v)/4}.$$

Proof of (119): Because $\Delta^v_+ \varepsilon_t = \underline{V}^{(N)}_t + \overline{V}^{(N)}_t$ for $t > N = T^\alpha$ we have

$$P^*_{T,N}(\overline{V}^{(N)}, \Delta^w Z_3) = P^*_T(\Delta^v \varepsilon, \Delta^w Z_3) - P^*_{T,N}(\underline{V}^{(N)}, \Delta^w Z_3) - (N/T)^{1/2-w} P^*_N(\Delta^v \varepsilon, \Delta^w Z_3).$$

The first term is $\mathbf{O}_P((1+\log T)^2 T^{\underline{\kappa}_v})$ by (105), the second is $\mathbf{O}_P((1+\log T)^2 (T^{3\alpha\underline{\kappa}_v - 2\underline{\kappa}_v} + T^{-1/4+\alpha(1+2\underline{\kappa}_v)/4}))$ by (118), and the last term is $\mathbf{O}_P((1+\log N)^2 N^{\underline{\kappa}_v + 1 + \kappa_w} T^{-1-\kappa_w})$ by (105). The first term dominates which proves the result. ■

Lemma A.11 *If the assumptions of Lemma A.9 hold, then:*

(i) Uniformly for $-1/2 + \kappa_u \le u \le u_0$ it holds that

$$\mathsf{D}^m M_T(1,u) = \mathbf{O}_P((1+\log T)^{2+m} T^{-\kappa_u}). \tag{120}$$

Uniformly for $-w_0 \le w \le -1/2 - \kappa_w$ it holds that $\mathsf{D}^m M^*_T(1,w) = \mathbf{O}_P(1)$ and

$$M^*_T(1,w) \Longrightarrow \xi \int_0^1 W_{-w-1}(s) ds. \tag{121}$$

Uniformly in $-1/2 - \underline{\kappa}_v \le v \le v_0$ it holds that

$$M_T(1,v) = \mathbf{O}_P((1+\log T)^2 T^{\underline{\kappa}_v}). \tag{122}$$

(ii) Uniformly for $(w,v,u) \in \mathcal{S}(\kappa_w, \underline{\kappa}_v, \bar{\kappa}_v, \kappa_u)$, see (101) of Definition A.1, it holds that

$$M_T(u_1,u_2|1) \Longrightarrow Var(\Delta^{u_1} Z_{1t}, \Delta^{u_2} Z_{2t}), \tag{123}$$

$$M_T^*(w_1,u_2|1) = \mathbf{O}_P((1+\log T)^2 T^{-\min(\kappa_u,\kappa_w)}) + \mathbf{o}_P(1), \tag{124}$$

$$M_T^{**}(w_1,w_2|1) \Longrightarrow \xi_1 \int_0^1 (W_{-w_1-1}(s)|1)(W_{-w_2-1}(s)|1)' ds \xi_2', \tag{125}$$

$$M_T(v,u|1) = \mathbf{O}_P(1), \tag{126}$$

$$M_T^*(v,w|1) = \mathbf{O}_P((1+\log T)^2 T^{\underline{\kappa}_v}), \tag{127}$$

$$M_T((v_1,v_2),(v_1,v_2)|1) \geq M_T((v_1,v_2),(v_1,v_2)) + \mathbf{o}_P(1), \tag{128}$$

where $(W_{-w-1}(s)|1) = W_{-w-1}(s) - \int_0^1 W_{-w-1}(s)ds$.

Proof. *Proof of (i)*: The variable $M_T(1,a) = T^{-1}\sum_{t=1}^T \Delta_+^a Z_{2t}^+$ is a linear process in ε_{2t} with mean zero, so that it follows from JN (2010, Lemma B.1) that $||M_T(1,a)||_4 \leq c||M_T(1,a)||_2$. As in the proof of Lemma A.6 it is enough to prove the result for $Z_{2t} = \varepsilon_{2t}$, and as in Lemma A.9 we give only the proof for $m=0$ because the additional $(\log T)$-factors do not change the proof. We find because $|\pi_n(-a)| \leq cn^{-a-1}$ that

$$||M_T(1,a)||_2^2 \leq cT^{-2}\sum_{t=1}^T [\sum_{n=1}^{t-1} n^{-a-1}]^2 \leq c(1+\log T)^2 T^{-1+2\max(-a,0)}. \tag{129}$$

For $a = u \geq -1/2 + \kappa_u$ we find $||M_T(1,u)||_4 \leq c(1+\log T)T^{-\kappa_u}$ and for $a = v \geq -1/2 - \underline{\kappa}_v$ we get $||M_T(1,v)||_4 \leq c(1+\log T)T^{\underline{\kappa}_v}$. For $a = w$ we get $||M_T^*(1,w)||_4 \leq c$ by the same method as in the proof of (90). We also find from (6) that

$$T^{-1}\sum_{t=1}^T T^{w+1/2}\Delta_+^w Z_t \overset{D}{\to} \xi \int_0^1 W_{-w-1}(s)ds.$$

Proof of (ii): To prove (123)-(127) we use decompositions like $M_T(u_1,u_2|1) = M_T(u_1,u_2) - M_T(u_1,1)M_T(1,1)^{-1}M_T(1,u_2)$ and apply Lemmas A.9 and A.11*(i)*, and note that $M_T(1,1)^{-1} = 1$.

To prove (128) we follow the proof of (107) and write $Z_{it}^+ = \xi_i\varepsilon_t + \Delta_+^{b_0}\tilde{Z}_{it}^+$ and the same argument shows that we only need to consider $\xi_i\varepsilon_t$, and it is then enough to prove the result for $Z_{it}^+ = \varepsilon_t$. We decompose $\Delta_+^v\varepsilon_t = \underline{V}_t^{(N)} + \overline{V}_t^{(N)}$ as in (111) and find, as in (112), that

$$\begin{aligned} M_T((v_1,v_2),(v_1,v_2)|1) &\geq P_{T,N}(\underline{V}^{(N)},\underline{V}^{(N)}|1) + P_{T,N}(\underline{V}^{(N)},\overline{V}^{(N)}|1) + P_{T,N}(\overline{V}^{(N)},\underline{V}^{(N)}|1) \\ &= P_{T,N}(\underline{V}^{(N)},\underline{V}^{(N)}) - R_{1T} + R_{2T} + R_{2T}', \end{aligned}$$

where

$$R_{1T} = P_{T,N}(\underline{V}^{(N)},1)M_T(1,1)^{-1}P_{T,N}(1,\underline{V}^{(N)}),$$

$$R_{2T} = P_{T,N}(\underline{V}^{(N)},\overline{V}^{(N)}) - P_{T,N}(\underline{V}^{(N)},1)M_T(1,1)^{-1}P_{T,N}(1,\overline{V}^{(N)}).$$

For $N = T^\alpha$ and $a = v_2 \geq -1/2 - \underline{\kappa}_v$ we get from (129) that

$$||P_{T,N}(1,\underline{V}^{(N)})||_4 \leq c(1+\log T)T^{-1/2+\alpha(1/2+\underline{\kappa}_v)}, \tag{130}$$

$$||P_{T,N}(1,\overline{V}^{(N)})||_4 \leq c(1+\log T)T^{\underline{\kappa}_v}. \tag{131}$$

This shows that $R_{1T} = \mathbf{o}_P(1)$ for $\alpha < 1/4$ and $\underline{\kappa}_v < 1/6$ because $-1/2 + \alpha(1/2+\underline{\kappa}_v) < 0$, and also that $R_{2T} = \mathbf{o}_P(1)$ because $-1/2 + \alpha(1/2+\underline{\kappa}_v) + \underline{\kappa}_v < 0$ and because (113) shows that $P_{T,N}(\underline{V}^{(N)},\overline{V}^{(N)}) = \mathbf{o}_P(1)$ for $\alpha < 1/4$ and $\underline{\kappa}_v < 1/6$. ∎

Corollary A.12 *If the assumptions of Lemma A.9 hold, then uniformly for* $(w, v, u) \in \mathcal{S}(\kappa_w, \underline{\kappa}_v, \overline{\kappa}_v, \kappa_u)$, *see (101) of Definition A.1:*

(i) It holds that

$$M_T^{**}(w_1, w_2|w_3, u, 1) = M_T^{**}(w_1, w_2|w_3, 1) + \boldsymbol{o}_P(1), \tag{132}$$

$$M_T(u_1, u_2|w, u_3, 1) \Longrightarrow Var(\Delta^{u_1} Z_{1t}, \Delta^{u_2} Z_{2t}|\Delta^{u_3} Z_{3t}), \tag{133}$$

$$M_T(v, u_1|w, u_2, 1) = \mathbf{O}_P(1). \tag{134}$$

(ii) If $N = T^\alpha$ *with* $\alpha < 1/4$, *and* (ξ_1', ξ_2') *has full rank, then for* $-1/2 - \underline{\kappa}_v \leq v_i \leq -1/2 + \overline{\kappa}_v$ *we find*

$$M_T((v_1, v_2), (v_1, v_2)|w, u, 1) \geq c\frac{1 - N^{-2\overline{\kappa}_v}}{2\overline{\kappa}_v}(\xi_1', \xi_2')'\Omega_0(\xi_1', \xi_2') + R_T, \tag{135}$$

where $R_T = \mathbf{O}_P(1)$ *uniformly for* $|v_i + 1/2| \leq \underline{\kappa}_v$.

Proof. *Proof of (i)*: The proofs are identical to those of (114), (115), and (116) except $M_T(a_1, a_2)$ are replaced by $M_T(a_1, a_2|1)$ and the results follow by application of Lemma A.11*(ii)*.

Proof of (ii): The proof is identical to that of Corollary A.10*(ii)* except all product moments are also conditional on a constant, 1, such that the remainder terms are now

$$R_{1T} = P_{T,N}^*(\underline{V}^{(N)}, \Delta^w Z_3|1)P_T^{**}(\Delta^w Z_3, \Delta^w Z_3|1)^{-1}P_{T,N}^*(\Delta^w Z_3, \underline{V}^{(N)}|1),$$

$$R_{2T} = P_{T,N}(\underline{V}^{(N)}, \overline{V}^{(N)}|1) - P_{T,N}^*(\underline{V}^{(N)}, \Delta^w Z_3|1)P_T^{**}(\Delta^w Z_3, \Delta^w Z_3|1)^{-1}P_{T,N}^*(\Delta^w Z_3, \overline{V}^{(N)}|1).$$

We thus need to show that, for $N = T^\alpha$,

$$P_{T,N}^*(\underline{V}^{(N)}, \Delta^w Z_3|1) = \mathbf{O}_P((1 + \log T)^2(T^{3\alpha\underline{\kappa}_v - 2\underline{\kappa}_v} + T^{-1/4 + \alpha(1+2\underline{\kappa}_v)/4}), \tag{136}$$

$$P_{T,N}^*(\overline{V}^{(N)}, \Delta^w Z_3|1) = \mathbf{O}_P((1 + \log T)^2 T^{\underline{\kappa}_v}). \tag{137}$$

If these were proved and $\alpha < 1/4$ and $\underline{\kappa}_v < 1/6$, it follows that R_{1T} and R_{2T} are $\mathbf{o}_P(1)$, see also (113), (130), and (131). Thus, proving (136) and (137) completes the proof of part (ii), see (107) and (128) for the main term $P_{T,N}(\underline{V}^{(N)}, \underline{V}^{(N)}|1)$.

Proof of (136): We find

$$P_{T,N}^*(\underline{V}^{(N)}, \Delta^w Z_3|1) = P_{T,N}^*(\underline{V}^{(N)}, \Delta^w Z_3) - P_{T,N}(\underline{V}^{(N)}, 1)M_T(1, 1)^{-1}P_{T,N}^*(\Delta^w Z_3, 1).$$

The first term is considered in (118). Next, $P_{T,N}^*(\Delta^w Z_3, 1) = M_T^*(w, 1) - (N/T)^{1/2-w}M_N^*(w, 1)$ such that

$$P_{T,N}(\underline{V}^{(N)}, 1)M_T(1, 1)^{-1}P_{T,N}^*(\Delta^w Z_3, 1) = \mathbf{O}_P((1 + \log T)^2 T^{-1/2+\alpha(1/2+\underline{\kappa}_v)})$$

by (121) and (130). The right-hand side is dominated by $\mathbf{O}_P((1 + \log T)^2 T^{-2\underline{\kappa}_v + 3\alpha\underline{\kappa}_v})$ for $\kappa < 1/6$, and summing up we thus find (136).

Proof of (137): The proof is identical to that of (119) except we refer to (127) and (136) instead of (105) and (118). ■

Appendix B Proof of Theorem 4

By Lemma A.8 deterministic terms generated by initial values are uniformly small. Note that (94) 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 the nearly critical ones, whereas (95) deals with the nonstationary $\beta_{0\perp}' X_{it}$. Hence deterministic terms in the processes do not influence the limit behavior of product moments, and in the remainder of the proof of Theorem 4 we therefore assume that they are zero and replace the regressors X_{it} by their stochastic component U_{it}^+, see (45).

B.1 Proof of (30): unique minimum of $\ell_p(\psi)$

On $\mathcal{N}_{\text{div}}(0)$ the inequality is trivially satisfied and on $\mathcal{N}_{\text{conv}}(0)$ 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 $f_0(z) = (1-z)^{-d_0}\Pi_0(z) = (1-z)^{-b_0}\Psi_0(1-(1-z)^b)$ for $|z| < 1$, see (8).

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, so that $\mathcal{F}_{\text{stat}}(\psi) = \sigma(\{U_{it}\}_{i=m}^{k-1}, \{\beta_0' U_{jt}\}_{j=-1}^{m-1})$. We define, see also (22),

$$S_t^{(m)} = U_{kt} + \sum_{i=m}^{k-1} \Psi_i U_{it} + \sum_{j=0}^{m-1} \Psi_j \bar{\beta}_0 \beta_0' U_{jt} - \Pi \bar{\beta}_0 \beta_0' U_{-1,t} = g^{(m)}(L)(C_0\varepsilon_t + \Delta^{b_0}Y_t),$$

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

The transfer function of the stationary linear process $S_t^{(m)}$ is $g^{(m)}(z)f_0(z)^{-1}$, which has $g^{(m)}(0)f_0(0)^{-1} = I_p$, so that $S_t^{(m)}$ is of the form $S_t^{(m)} = \varepsilon_t + \xi_1\varepsilon_{t-1} + \dots$. It follows that $Var(S_t^{(m)}) \geq \Omega_0$ and equality holds only for $S_t^{(m)} = \varepsilon_t$ or $g^{(m)}(z) = f_0(z)$ for all $|z| < 1$, which implies that $(d, b) = (d_0, b_0)$, $m = 0$, and that $\Psi_j = \Psi_{j0}$ and $\Pi\bar{\beta}_0 = \alpha_0$.

Note that $Var(S_t^{(m)})$ is quadratic in the parameters $\{\Psi_i\}_{i=m}^{k-1}, \{\Psi_j\bar{\beta}_0\}_{j=0}^{k-1}, \Pi\bar{\beta}_0$, and that minimizing over these, the residual variance satisfies the same inequality,

$$Var(U_{kt}|\mathcal{F}_{\text{stat}}(\psi)) = Var(S_t^{(m)}|\mathcal{F}_{\text{stat}}(\psi)) \geq \Omega_0 \text{ for all } \psi.$$

Equality holds only for $\psi = \psi_0$ so this completes the proof of (30).

B.2 Proof of (31): convergence in probability of $\ell_{T,r}(\psi_0)$

We find from (34) that the matrices in the reduced rank regression can be expressed in terms of $\mathcal{A}_T, \mathcal{B}_T$, and $\mathcal{C}_T$, see (35). The eigenvalues in (25) are continuous functions of the product moment matrices, so that (41) shows that $\{\hat{\omega}_i(\psi)\}_{i=1}^r \Longrightarrow \{\omega_i(\psi)\}_{i=1}^r$ on $\mathbb{C}^r(\mathcal{N}(\psi_0, \epsilon))$ as $T \to \infty$. It follows that $\{\omega_i(\psi)\}_{i=1}^r$ are continuous in ψ and given as solutions of

$$\det(\omega\Sigma_{\beta\beta} - \Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta}) = 0, \tag{138}$$

where $\Sigma_{00} = Var(U_{kt}|\mathcal{F}_t), \Sigma_{0\beta} = Cov(U_{kt}, \beta_0' U_{-1t}|\mathcal{F}_t)$, and $\Sigma_{\beta\beta} = Var(\beta_0' U_{-1t}|\mathcal{F}_t)$, and where $\mathcal{F}_t = \sigma(U_{0t}, \dots, U_{k-1,t})$, see Johansen (1996, chapter 11) for the detailed proof for the I(1) model. For $\psi = \psi_0$, $\ell_{T,r}(\psi_0)$ is given by

$$\log\det(S_{00}(\psi_0)) + \sum_{i=1}^r \log(1 - \hat{\omega}_i(\psi_0)) \overset{P}{\to} \log\det(Var(U_{kt}|\mathcal{F}_t)) + \sum_{i=1}^r \log(1 - \omega_i(\psi_0))$$
$$= \log\det(\Sigma_{00} - \Sigma_{0\beta}\Sigma_{\beta\beta}^{-1}\Sigma_{0\beta}) = \log\det(\Sigma_{00|\beta}) = \log\det(\Omega_0).$$

This completes the proof of (31).

B.3 Proof of (i): model $\mathcal{H}_p$

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 Appendix A.4 and especially Lemma A.9 and Corollary

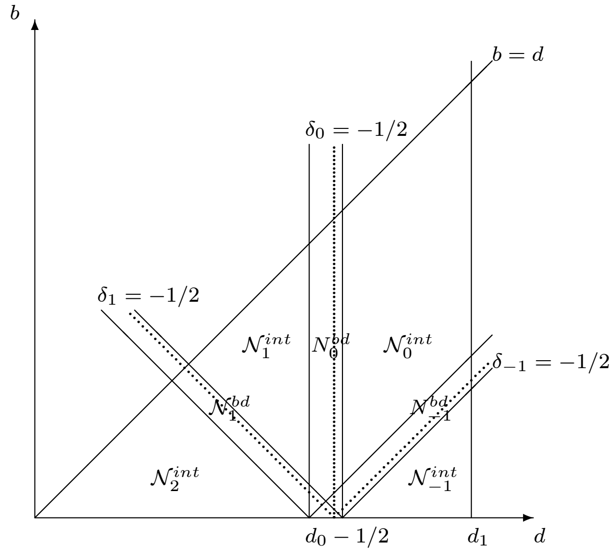

Figure 1: The parameter space $\mathcal{N}$ is the set bounded by $b > 0$, $b \leq d$, and $d \leq d_1$. The sets $\mathcal{N}_m^{bd} = \mathcal{N}_m^{bd}(\kappa_1, \kappa)$, where a process is close to being critical, and the sets $\mathcal{N}_m^{int} = \mathcal{N}_m^{int}(\kappa_1, \kappa)$ are illustrated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

A.10, which we apply repeatedly below to show weak convergence of the profile likelihood as a process indexed by the parameters d and b.

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" given as follows.

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

$$\mathcal{N}_m^{int}(\kappa_1, \kappa) = \{\psi \in \mathcal{N} : \delta_{m-1} \leq -1/2 - \kappa_1 \text{ and } -1/2 + \kappa \leq \delta_m\}, \ -1 \leq m \leq k+1, \quad (139)$$

and the $(\kappa_1, \kappa)-$*boundaries,*

$$\mathcal{N}_m^{bd}(\kappa_1, \kappa) = \{\psi \in \mathcal{N} : -1/2 - \kappa_1 \leq \delta_m \leq -1/2 + \kappa\}, \ -1 \leq m \leq k. \quad (140)$$

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

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

In (139) we define the $(\kappa_1, \kappa)-$interior $\mathcal{N}_m^{int}(\kappa_1, \kappa)$ as the set of ψ for which all processes are either clearly stationary or clearly nonstationary in the sense that their fractional index is either $\geq -1/2 + \kappa$ or $\leq -1/2 - \kappa_1$. The $(\kappa_1, \kappa)-$boundary $\mathcal{N}_m^{bd}(\kappa_1, \kappa)$ 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.

The profile likelihood for model $\mathcal{H}_p$ is derived by regressing $X_{kt} = \Delta^{d+kb} X_t$ on the other variables, which can be either asymptotically stationary, nonstationary, or near critical. We apply the expression $\ell_{T,p}(\psi) = \log\det(SSR_T(\psi))$, see (27), and Lemma A.9 and Corollary A.10 to find the asymptotic properties of $\det(SSR_T(\psi))$. We use the notation $\kappa_w, \underline{\kappa}_v, \overline{\kappa}_v$, and κ_u, see (101) in Definition A.1, and note that for $(d, b) \in \mathcal{N}$ all indices are bounded. The assumptions in Theorem 4 imply that $q^{-1} < \min(\eta/3, (1/2 - (d_0 - b_0))/2)$ and $q^{-1} \leq 1/8$, so $q^{-1} < \min(1/6, \eta/3, (1/2 - (d_0 - b_0))/2)$. We can therefore choose a κ_1 in the interval $q^{-1} < \kappa_1 < \min(1/6, \eta/3, (1/2 - (d_0 - b_0))/2)$, and apply this fixed κ_1 in the proof below.

B.3.1 Analysis of $\mathcal{N}_m^{bd}(\kappa_1, \kappa)$

In order to apply Corollary A.10 we need to define the indices $\kappa_w, \underline{\kappa}_v, \overline{\kappa}_v$, and κ_u. For $\psi \in \mathcal{N}_m^{bd}(\kappa_1, \kappa)$ the process $\beta_{0\perp}' X_{mt}$ is near critical with index $v = \delta_m \in [-1/2 - \kappa_1, -1/2 + \kappa]$, so we define $\underline{\kappa}_v = \kappa_1$ and $\overline{\kappa}_v = \kappa$. The nonstationary processes $\{\beta_{0\perp}' X_{it}\}_{i=-1}^{m-1}$ are collected in a vector with largest fractional index $w = \delta_{m-1} = \delta_m - b \le -1/2 + \kappa - b \le -1/2 - 2\eta/3$, because $b \ge \eta$ and $\kappa < \kappa_1 < \eta/3$, so we define $\kappa_w = 2\eta/3 > q^{-1}$, so we have enough moments for weak convergence of the nonstationary processes to fBM, c.f. the moment condition needed for (103) of Lemma A.9. Finally the asymptotically stationary processes $\{\beta_{0\perp}' X_{it}\}_{i=m+1}^{k}$ have smallest index $\delta_{m+1} = \delta_m + b \ge -1/2 - \kappa_1 + \eta \ge -1/2 + 2\eta/3$ because $b \ge \eta$ and $\kappa_1 < \eta/3$, and $\{\beta_0' X_{jt}\}_{j=-1}^{k}$ have smallest index $\delta_{-1} + b_0 \ge -d_0 + b_0 = -1/2 + (1/2 - (d_0 - b_0))$, so we choose $\kappa_u = \min(2\eta/3, 1/2 - (d_0 - b_0))$.

With these choices $\underline{\kappa}_v$ satisfies the conditions in Definition A.1 for the application of Corollary A.10, because $b_0 \ge \eta$ implies that

$$\underline{\kappa}_v = \kappa_1 < \min(1/6, \eta/3, (1/2 - (d_0 - b_0))/2) \le \min(b_0/3, \kappa_w/2, \kappa_u/2, 1/6).$$

We can now prove that, for $m = k$ and any $A > 0$ and $\gamma > 0$, there is a $\kappa_0 > 0$ and $T_0 > 0$ so that for $T \ge T_0$,

$$P(\inf_{\psi \in \mathcal{N}_k^{bd}(\kappa_1, \kappa_0) \cap \mathcal{K}(\eta)} \ell_{T,p}(\psi) \ge A) \ge 1 - \gamma. \tag{141}$$

For the rest of the proof we let κ_0 be fixed at this value. Furthermore, for $m < k$, we can prove that for this fixed value of κ_0,

$$\sup_{\psi \in \mathcal{N}_m^{bd}(\kappa_1, \kappa_0) \cap \mathcal{K}(\eta)} |\ell_{T,p}(\psi) - \ell_p(\psi)| \overset{P}{\to} 0 \text{ as } T \to \infty. \tag{142}$$

Proof of (141): For $\psi \in \mathcal{N}_k^{bd}(\kappa_1, \kappa)$, $\beta_0' X_{kt}$ is stationary with index $u_1 = \delta_k + b_0$ and $\beta_{0\perp}' X_{kt}$ is near critical with index $v_1 = \delta_k$. Applying the decomposition $X_{kt} = \bar{\beta}_0 \beta_0' X_{kt} + \bar{\beta}_{0\perp} \beta_{0\perp}' X_t = B_0(X_{kt}'\beta_0, X_{kt}'\beta_{0\perp})'$ where $B_0 = (\bar{\beta}_0, \bar{\beta}_{0\perp})$, see (11), we decompose the determinant

$$\begin{aligned} \det(SSR_T(\psi)) &= \det(B_0 M_T((u_1, v_1), (u_1, v_1)|w, u) B_0') \\ &= \det(M_T(u_1, u_1|w, u)) \det(M_T(v_1, v_1|w, u, u_1))(\det(B_0))^2. \end{aligned}$$

Uniformly in $\psi \in \mathcal{N}_k^{bd}(\kappa_1, \kappa)$ the first factor converges in distribution by (115).

For the second factor we apply (117) for $N = T^\alpha$:

$$M_T(v_1, v_1|w, u, u_1) \ge c \frac{1 - T^{-2\overline{\kappa}_v \alpha}}{2\overline{\kappa}_v} (\xi_1' \xi_2')' \Omega_0^{-1} (\xi_1' \xi_2') + R_T, \tag{143}$$

where $\max_{|v_1 + 1/2| \le \underline{\kappa}_v} |R_T|$ is bounded with probability $\ge 1 - \gamma$ for $T \ge T_0$. Thus, the smallest eigenvalue of $M_T(v_1, v_1|w, u, u_1)$ is bounded below by a constant times $(1 - T^{-2\overline{\kappa}_v \alpha})/(2\overline{\kappa}_v)$. This factor is increasing in T from zero to $1/(2\overline{\kappa}_v)$ and decreasing in $2\overline{\kappa}_v$ from $\alpha \log T$ to zero. It follows that for any $A > 0$ we can find (κ_0, T_0) so that for $\overline{\kappa}_v \le \kappa_0$ and $T \ge T_0$ it holds that $c(1 - T^{-2\overline{\kappa}_v \alpha})/2\overline{\kappa}_v \ge A$. Using $\overline{\kappa}_v = \kappa = \kappa_0$ we then find that $\inf_{\psi \in \mathcal{N}_k^{bd}(\kappa_1, \kappa_0) \cap \mathcal{K}(\eta)} \ell_{T,p}(\psi)$ is large with probability $\ge 1 - \gamma$ for $T \ge T_0$. This proves (141).

Proof of (142): For $\psi \in \mathcal{N}_m^{bd}(\kappa_1, \kappa_0)$ with $m < k$, $\beta_0' X_{kt}$ is stationary with index u_1 and $\beta_{0\perp}' X_t$ is stationary with index u_2. Then $SSR_T(\psi) = B_0 M_T((u_1, u_2), (u_1, u_2)|w, v, u) B_0'$, and

$$\begin{aligned} &SSR_T(\psi) - Var(U_{kt}|\mathcal{F}_{\text{stat}}(\psi)) \\ &= B_0 M_T((u_1, u_2), (u_1, u_2)|w, u) B_0' - Var(U_{kt}|\mathcal{F}_{\text{stat}}(\psi)) \\ &- B_0 M_T((u_1, u_2), v|w, u) M_T(v, v|w, u)^{-1} M_T(v, (u_1, u_2)|w, u) B_0'. \end{aligned} \tag{144}$$

For fixed $\kappa_0 > 0$, we find from (115) that on $\mathbb{C}(\mathcal{N}_m^{bd}(\kappa_1, \kappa_0) \cap \mathcal{K}(\eta))$,

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

We then apply Lemma A.4 which shows that weak convergence to a deterministic limit implies uniform convergence in probability.

For the last term of (144) we apply (116) to see that on $\mathbb{C}(\mathcal{N}_m^{bd}(\kappa_1, \kappa_0) \cap \mathcal{K}(\eta))$,

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

and (143) shows that the factor $(1 - T^{-2\kappa_0\alpha})/2\kappa_0$ can be chosen so large that the smallest eigenvalue of $M_T(v, v|w, u)$ is large with probability $\geq 1 - \gamma$ for $T \geq T_0$. This implies that $M_T(v, v|w, u)^{-1}$ is small uniformly on $\mathcal{N}_m^{bd}(\kappa_1, \kappa_0) \cap \mathcal{K}(\eta)$, which proves (142).

B.3.2 Analysis of $\mathcal{N}_m^{int}(\kappa_1, \kappa_0)$

For $\psi \in \mathcal{N}_m^{int}(\kappa_1, \kappa_0)$ the asymptotically stationary processes $\{\beta_{0\perp}' X_{it}\}_{i=m}^{k-1}$ and $\{\beta_0' X_{jt}\}_{j=-1}^{k}$ have indices greater than $-1/2 + \kappa_0$ and $-1/2 + (1/2 - (d_0 - b_0))$, respectively, so we collect them in a vector with lowest index $u \geq -1/2 + \kappa_u$ for $\kappa_u = \min(\kappa_0, 1/2 - d_0 + b_0)$. 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 that $\kappa_w = \kappa_1$.

We can then prove that for $m = k+1$, where $\mathcal{N}_{k+1}^{int}(\kappa_1, \kappa_0) = \mathcal{N}_{k+1}^{int}(\kappa_1)$, and any $A > 0, \gamma > 0$ there is a $T_0 > 0$ so that for $T \geq T_0$,

$$P(\inf_{\psi \in \mathcal{N}_{k+1}^{int}(\kappa_1) \cap \mathcal{K}(\eta)} \ell_{T,p}(\psi) \geq A) \geq 1 - \gamma. \tag{145}$$

For $m \leq k$ we can prove that (for $\kappa_0 > 0$ fixed at the value determined in (141))

$$\sup_{\psi \in \mathcal{N}_m^{int}(\kappa_1, \kappa_0) \cap \mathcal{K}(\eta)} |\ell_{T,p}(\psi) - \ell_p(\psi)| \overset{P}{\to} 0 \text{ as } T \to \infty. \tag{146}$$

Proof of (145): For $\psi \in \mathcal{N}_{k+1}^{int}(\kappa_1)$, $\beta_0' X_{kt}$ is stationary with index u_1 and $\beta_{0\perp}' X_{kt}$ is nonstationary with index $w_1 \leq -1/2 - \kappa_1$. We decompose

$$\begin{aligned} \det(SSR_T(\psi)) &= \det(B_0 M_T((u_1, w_1), (u_1, w_1)|w, u) B_0') \\ &= \det(M_T(w_1, w_1|w, u)) \det(M_T(u_1, u_1|w_1, w, u)) \det(B_0)^2. \end{aligned}$$

The second factor is $\mathbf{O}_P(1)$ uniformly in $\psi \in \mathcal{N}_{k+1}^{int}(\kappa_1) \cap \mathcal{K}(\eta)$ by (115). In the first factor we normalize $T^{2w_1+1} M_T(w_1, w_1|w, u)$ to convergence to an almost surely positive limit, see (114), so that the first factor is proportional to $T^{-(2w_1+1)} \geq T^{2\kappa_1} \to \infty$, which proves (145).

Proof of (146): For $\psi \in \mathcal{N}_m^{int}(\kappa_1)$ and $m \leq k$, $\beta_0' X_{kt}$ is stationary with index u_1 and $\beta_{0\perp}' X_{kt}$ is stationary with index u_2, and $SSR_T(\psi) = B_0 M_T((u_1, u_2), (u_1, u_2)|w, u) B_0'$. It follows from (115) and Lemma A.4, see also (29), that for fixed κ_1, κ_0, (146) holds.

Finally, (32) follows from (141) and (145), and (33) follows from (142) and (146). This completes the proof of Theorem 4*(i)*.

B.4 Proof of (ii): model $\mathcal{H}_p(d = b)$

The proof for model (3) in part *(ii)* is identical to that for model (2) given in part *(i)* with two modifications. First, the definitions of "interiors" and "boundaries" in Definition B.2 need to be simplified to take into account the restriction $d = b$, that is the 45° line in Figure 1. Second, all references to results in Lemma A.9 and Corollary A.10 need to be replaced with references to Lemma A.11 and Corollary A.12.

References

Anderson, T.W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics* **22**, 327–351.
Avarucci, M. and Velasco, C. (2009). A Wald test for the cointegration rank in nonstationary fractional systems. *Journal of Econometrics* **151**, 178–189.
Billingsley, P. (1968). *Convergence of Probability Measures,* Wiley, New York.
Davidson, J. (2002). A model of fractional cointegration, and tests for cointegration using the bootstrap. *Journal of Econometrics* **110**, 187–212.
Elliott, G. (1998). On the robustness of cointegration methods when regressors almost have unit roots. *Econometrica* **66**, 149–158.
Engle, R.F. and Granger, C.W.J. (1987). Co-integration and error correction: representation, estimating and testing. *Econometrica* **55**, 251–276.
Gil-Alana, L.A. and Hualde, J. (2009). Fractional integration and cointegration: an overview and an empirical application. In Patterson, K. and Mills, T.C., eds., *Palgrave Handbook of Econometrics*, vol. II, Palgrave MacMillan, New York, 434–469.
Granger, C.W.J. (1986). Developments in the study of cointegrated economic variables. *Oxford Bulletin of Economic and Statistics* **48**, 213–228.
Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and its Application*, Academic Press, New York.
Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* **64**, 413-430.
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.
Hualde, J. and Robinson, P.M. (2010). Semiparametric inference in multivariate fractionally cointegrated systems. *Journal of Econometrics* **157**, 492–511.
Hualde, J. and Robinson, P.M. (2011). Gaussian pseudo-maximum likelihood estimation of fractional time series models. *Annals of Statistics* **39**, 3152–3181.
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.
Jeganathan, P. (1999). On asymptotic inference in cointegrated time series with fractionally integrated errors. *Econometric Theory* **15**, 583–621.
Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* **12**, 231–254.
Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* **59**, 1551–1580.
Johansen, S. (1996). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, 2nd ed., Oxford University Press, Oxford.
Johansen, S. (2002). A small sample correction of the test for cointegrating rank in the vector autoregressive model. *Econometrica* **70**, 1929–1961.
Johansen, S. (2008). A representation theory for a class of vector autoregressive models for fractional processes. *Econometric Theory* **24**, 651–676.
Johansen, S. (2009). Representation of cointegrated autoregressive processes with application to fractional processes. *Econometric Reviews* **28**, 121–145.
Johansen, S. and Nielsen, M.Ø. (2010). Likelihood inference for a nonstationary fractional autoregressive model. *Journal of Econometrics* **158**, 51–66.

Juselius, K. (2006). *The Cointegrated VAR Model: Methodology and Applications*, Oxford University Press, Oxford.

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

Lasak, K. (2008). Maximum likelihood estimation of fractionally cointegrated systems. CREATES Research Paper 2008–53.

Lasak, K. (2010). Likelihood based testing for no fractional cointegration. *Journal of Econometrics* **158**, 67–77.

Lawley, D.N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* **43**, 296–303.

Lobato, I.N. and Velasco, C. (2007). Efficient Wald tests for fractional unit roots. *Econometrica* **75**, 575–589.

MacKinnon, J.G. and Nielsen, M.Ø. (2011). Numerical distribution functions of fractional unit root and cointegration tests. QED working paper 1240, Queen's University.

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

Marmol, F. and Velasco, C. (2004). Consistent testing of cointegrating relationships. *Econometrica* **72**, 1809–1844.

Nielsen, M.Ø. and Morin L. (2012). FCVARmodel.m: A Matlab software package for estimation and testing in the fractionally cointegrated VAR model. QED working paper 1273, Queen's University.

Phillips, P.C.B. (1991). Optimal inference in cointegrated systems. *Econometrica* **59**, 283–306.

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.

Robinson, P.M. and Hualde, J. (2003). Cointegration in fractional systems with unknown integration orders. *Econometrica* **71**, 1727–1766.

Saikkonen, P. (1991). Asymptotically efficient estimation of cointegration regressions. *Econometric Theory* **7**, 1–21.

Taqqu, M.S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **31**, 287–302.

van der Vaart, A.W. (1998). *Asymptotic Statistics*, Cambridge University Press, Cambridge.