Queen's Economics Department Working Paper No. 1405

Nonstationary cointegration in the fractionally cointegrated VAR 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

5-2018

Nonstationary cointegration in the fractionally cointegrated VAR model*

Søren Johansen[†]
University of Copenhagen
and CREATES

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

May 1, 2018

Abstract

We consider the fractional cointegrated vector autoregressive (CVAR) model of Johansen and Nielsen (2012a) and make two distinct contributions. First, in their consistency proof, Johansen and Nielsen (2012a) imposed moment conditions on the errors that depend on the parameter space, such that when the parameter space is larger, stronger moment conditions are required. We show that these moment conditions can be relaxed, and for consistency we require just eight moments regardless of the parameter space. Second, Johansen and Nielsen (2012a) assumed that the cointegrating vectors are stationary, and we extend the analysis to include the possibility that the cointegrating vectors are nonstationary. Both contributions require new analysis and results for the asymptotic properties of the likelihood function of the fractional CVAR model, which we provide. Finally, our analysis follows recent research and applies a parameter space large enough that the usual (non-fractional) CVAR model constitutes an interior point and hence can be tested against the fractional model using a χ^2-test.

Keywords: Cointegration, fractional integration, likelihood inference, vector autoregressive model.

JEL Classification: C32.

1 Introduction

For a p-dimensional time series, X_t, the fractional cointegrated vector autoregressive (CVAR) model of Johansen (2008) and Johansen and Nielsen (2012a), hereafter JN(2012a), is

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

*We are grateful to the Canada Research Chairs program, the Social Sciences and Humanities Research Council of Canada (SSHRC), and the Center for Research in Econometric Analysis of Time Series (CREATES, funded by the Danish National Research Foundation, DNRF78) for research support.

†Email: `Soren.Johansen@econ.ku.dk`

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

where ε_t is p-dimensional independent and identically distributed with mean zero and covariance matrix Ω and Δ^b and $L_b = 1 - \Delta^b$ are the fractional difference and fractional lag operators, respectively.

The fractional difference is given by, for a generic p-dimensional time series Z_t,

$$\Delta^d Z_t = \sum_{n=0}^{\infty} \pi_n(-d) Z_{t-n}, \tag{2}$$

provided the sum is convergent, and the fractional coefficients $\pi_n(u)$ are defined in terms of the binomial expansion $(1-z)^{-u} = \sum_{n=0}^{\infty} \pi_n(u) z^n$, i.e.,

$$\pi_n(u) = \frac{u(u+1)\cdots(u+n-1)}{n!}.$$

With the definition of the fractional difference operator in (2), Z_t is said to be fractional of order d, denoted $Z_t \in I(d)$, if $\Delta^d Z_t$ is fractional of order zero, i.e., if $\Delta^d Z_t \in I(0)$. The latter property can be defined in the frequency domain as having spectral density matrix that is finite and non-zero near the origin or in terms of the linear representation coefficients if the sum of these is non-zero and finite, see, for example, JN(2012a, p. 2672). An example of a process that is fractional of order zero is the stationary and invertible ARMA model. Finally, then, if $Z_t \in I(d)$ and one or more linear combinations are fractional of a lower order, i.e., there exists a $p \times r$ matrix β such that $\beta' Z_t \in I(d-b)$ with $b > 0$, then Z_t is said to be (fractionally) cointegrated.

In this paper, we make two distinct contributions to the fractional CVAR literature. First, in their consistency proof, JN(2012a) imposed moment conditions on the errors that depend on the parameter space, such that when the parameter space is larger, stronger moment conditions are required. Specifically, with the lower bound for the parameter b being denoted by $\eta > 0$, the moment conditions in JN(2012a) include the requirement that $E|\varepsilon_t|^q < \infty$ for some $q > 3/\eta$, in addition to $E|\varepsilon_t|^8 < \infty$. That is, when the parameter space for b allows very small values, corresponding to very weak cointegration that would be difficult to detect in practice, the errors were required to have more moments. We show that the moment conditions can be relaxed, and we assume just $E|\varepsilon_t|^8 < \infty$ regardless of the parameter space. This requires new results on certain product moments of nonstationary processes, and specifically requires proving tightness of the inverse of such product moments.

Our second contribution is to extend the analysis of JN(2012a) to include the possibility that the cointegrating vectors are nonstationary. JN(2012a) assumed that the cointegrating vectors are stationary, i.e., that $d_0 - b_0 < 1/2$, and extending the results to allow $d_0 - b_0 > 1/2$ requires new analysis and results for the asymptotic properties of the likelihood function of the fractional CVAR model, which we provide. Such nonstationary cointegrating vectors have been found in many empirical studies; some examples in finance using the fractional CVAR model are Caporin *et al.* (2013, Table 2), Barunik and Dvorakova (2015, Table 6), and Dolatabadi *et al.* (2016, Tables 5–6).

Finally, following JN(2018), we also allow the parameter space to include the usual CVAR of Johansen (1996), which is obtained by the restriction $d = b = 1$, as an interior point in the parameter space. This, of course, allows testing the usual CVAR model as a restriction on the fractional CVAR model.

The remainder of the paper is laid out as follows. In the next section we give the assumptions and the main results. The results and their proofs rely on a series of bounds on product moments, which we give in Section 3. These bounds include the important tightness proof for the inverse of product moments of nonstationary processes, which is given in Lemma 2. Some concluding remarks are offered in Section 4. The proof of consistency of the estimators is quite involved, and is given in Section 5.

2 Assumptions and main results

In JN(2012a), asymptotic properties of maximum likelihood estimators and test statistics were derived for model (1) with the parameter space $\eta \leq b \leq d \leq d_1$ for some $d_1 > 0$, which can be arbitrarily large, and some η such that $0 < \eta \leq 1/2$. The parameter space was extended by Johansen and Nielsen (2018) to

$$\mathcal{N} = \mathcal{N}(\eta, \eta_1, d_1) = \{d, b : \eta \leq b \leq d + \eta_1, d \leq d_1\}; \tag{3}$$

again for an arbitrarily large $d_1 > 0$ and an arbitrarily small η such that $0 < \eta \leq 1/2$. While η is exactly the same as in JN(2012a), we have in (3) introduced the new constant $\eta_1 > 0$, which is zero in JN(2012a). We note that the parameter space $\mathcal{N}$ explicitly includes the line segment $\{d, b : \eta < d = b < d_1\}$ in the interior precisely because $\eta_1 > 0$. Although $\eta > 0$ can be arbitrarily small, a smaller η implies a stronger moment condition in both JN(2012a) and JN(2018). This moment condition is relaxed below.

We will assume that the data for $t \geq 1$ is generated by model (1). A standard approach for autoregressive models, which we follow, is to conduct inference using the conditional likelihood function of $X_1, \ldots, X_T$ given initial values $\{X_{-n}\}_{n\geq 0}$. That is, we interpret (1) as a model for $X_t, t = 1, \ldots, T$, given the past, and use the conditional density to build a conditional likelihood function. Thus, since our entire approach is conditional on the initial values $\{X_{-n}\}_{n\geq 0}$ we consider these non-random, as is standard for (especially nonstationary) autoregressive models.

However, it is difficult to imagine a situation where $\{X_s\}_{s=-\infty}^{T}$ is available, or perhaps even exists, so we assume that the data is only observed for $t = -N+1, \ldots, T$. JN(2016) argue in favor of the assumption that data was initialized in the finite past using two leading examples, political opinion poll data and financial volatility data, but we maintain the more general assumption from JN(2012a), where the data $\{X_{-n}\}_{n=N}^{\infty}$ may or may not exist, but in any case is not observed. However, although the initial values assumption is based on that of JN(2012a), our notation for initial values is closer to that of JN(2016) (in particular, our notation N and M_0 follows the notation in JN(2016), and is basically reversed from the notation in JN(2012a)). That is, given a sample of size $T_0 = T + N$, this is split into N initial values, $\{X_{-n}\}_{n=0}^{N-1}$, on which the estimation will be conditional, and T sample observations, $\{X_t\}_{t=1}^{T}$, to which the model is fitted. We summarize this in the following display:

$$\underbrace{\ldots, X_{-N}}_{\substack{\text{Data may or may not exist,} \\ \text{but is not observed}}} \quad , \quad \underbrace{X_{1-N}, \ldots, X_0}_{\substack{\text{Data is observed} \\ \text{(initial values)}}} \quad , \quad \underbrace{X_1, \ldots, X_T}_{\substack{\text{Data is observed} \\ \text{(estimation)}}} \tag{4}$$

The inclusion of initial values, i.e. letting $N \geq 1$, has the purpose of mitigating the effect of the unobserved part of the process from time $t \leq -N$. Note that the (both observed an unobserved) initial values, i.e. $\{X_{-n}\}_{n=0}^{\infty}$, are not assumed to be generated by the model (1), but will only be assumed to be bounded, non-random numbers, see Assumption 3 below. Also note that the statistical and econometric literature has almost universally assumed $N = 0$ and, in many cases, also assumed that data did not exist for $t \leq 0$ or was equal to zero for $t \leq 0$.

Because we do not observe data prior to time $t = 1-N$, it is necessary to impose $X_{-n} = 0$ for $n \geq N$ in the calculations, even if these (unobserved) initial values are not in fact zero. To obtain our results we will need different assumptions on the initial values, and we will discuss these below. Consequently, for calculation of the likelihood function, we will apply the truncated fractional difference operator defined by

$$\Delta_N^d X_t = \sum_{n=0}^{t-1+N} \pi_n(-d) X_{t-n},$$

and keep N fixed, but allow for more non-zero initial values in the data generating process (DGP); see Assumptions 3 and 5. Note that our Δ_0 corresponds to Δ_+ in, e.g., JN(2012a), and we will use the notations Δ_0 and Δ_+ synonymously. Efficient calculation of truncated fractional differences is discussed in Jensen and Nielsen (2014).

We therefore fit the model

$$\Delta_N^d X_t = \alpha\beta' \Delta_N^{d-b} L_b X_t + \sum_{i=1}^{k} \Gamma_i \Delta_N^d L_b^i X_t + \varepsilon_t, \quad t = 1, \ldots, T, \tag{5}$$

and consider maximum likelihood estimation of the parameters, conditional on only N initial values, $\{X_{-n}\}_{n=0}^{1-N}$. Define the residuals

$$\varepsilon_t(\lambda) = \Delta_N^d X_t - \alpha\beta' \Delta_N^{d-b} L_b X_t - \sum_{i=1}^{k} \Gamma_i \Delta_N^d L_b^i X_t, \tag{6}$$

where λ is the collection of parameters $\{d, b, \alpha, \beta, \Gamma_1, \ldots, \Gamma_k, \Omega\}$, which are freely varying; that is, λ is in a product space. The Gaussian log-likelihood function, conditional on N initial values, $\{X_{-n}\}_{n=0}^{1-N}$, is then

$$\log L_T(\lambda) = -\frac{T}{2} \log\det(\Omega) - \frac{T}{2} \operatorname{tr}(\Omega^{-1} T^{-1} \sum_{t=1}^{T} \varepsilon_t(\lambda)\varepsilon_t(\lambda)'), \tag{7}$$

and the maximum likelihood estimator, $\hat{\lambda}$, is defined as the argmax of (7) with respect to λ such that $(d, b) \in \mathcal{N}$. Specifically, for given values of (d, b), the log-likelihood function $\log L_T(\lambda)$ can be concentrated with respect to $\{\alpha, \beta, \Gamma_1, \ldots, \Gamma_k, \Omega\}$ by reduced rank regression, and the resulting concentrated log-likelihood function is then optimized numerically with respect to (d, b) over the parameter space $\mathcal{N}$ given in (3). Algorithms for optimizing the likelihood function (7) are discussed in more detail in JN(2012a, Section 3.1) and implemented in Nielsen and Popiel (2016); see also Section 5.3 below.

Before we impose further conditions on the DGP, we introduce the following notation. For any $n \times m$ matrix A, we define the norm $|A| = \operatorname{tr}(A'A)^{1/2}$ and use the notation $A_\perp$ for an $n \times (n-m)$ matrix of full rank for which $A'A_\perp = 0$. For symmetric positive definite matrices A and B we use $A > B$ to denote that $A - B$ is positive definite. We also let

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

denote the usual polynomial from the CVAR model. Then model (1) can be written as $\Pi(L)X_t = \Delta^{d-b}\Psi(L_b)X_t = \varepsilon_t$, so that

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

Finally, we let C_b denote the fractional unit circle, which is the image of the unit disk under the mapping $y = 1-(1-z)^b$, see (9) and Johansen (2008, p. 660), and we define $\Gamma = I_p - \sum_{i=1}^{k}\Gamma_i$.

Assumption 1 *For $k \geq 0$ and $0 \leq r \leq p$ the process X_t, $t = 1, \ldots, T$, is generated by model (1) with the parameter value λ_0, using non-random initial values $\{X_{-n}\}_{n=0}^{\infty}$.*

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 $\{X_{-n}\}_{n=0}^{\infty}$ are uniformly bounded, i.e. $\sup_{n\geq 0}|X_{-n}| < \infty$.*

Assumption 4 *The true parameter value λ_0 satisfies $(d_0, b_0) \in \mathcal{N}$, $d_0 - b_0 \geq 0$, $b_0 \neq 1/2$, 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 $\det(\alpha_{0\perp}'\Gamma_0\beta_{0\perp}) \neq 0$. If $r < p$, $\det(\Psi_0(y)) = 0$ 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.*

The conditions in Assumptions 1–3 are identical to those in JN(2012a), while Assumption 4 is weaker than that in JN(2012a) since it does not impose $d_0 - b_0 < 1/2$. First, Assumption 1 implies that the data is only generated by model (1) starting at time $t = 1$. Specifically, the theory will be developed for observations $X_1, \ldots, X_T$, generated by model (1) with fixed, bounded initial values, $X_{-n}, n \geq 0$, that are not assumed to be generated by the model. That is, we conduct inference using the conditional likelihood function (7) and derive properties of estimators and tests using the conditional distribution of $X_1, \ldots, X_T$ given $X_{-n}, n \geq 0$, as developed by JN(2012a) and JN(2016).

Moreover, for $\Delta^a X_t, a > 0$, to be well-defined as an infinite sum, see (2), we assume that the initial values, $X_{-n}, n \geq 0$, are uniformly bounded, c.f. Assumption 3. Many of the intermediate results can be proved under just the boundedness assumption in Assumption 3, but to get the asymptotic distributions we need to impose the stronger Assumption 5 as discussed below. Assumption 2 importantly does not assume Gaussian errors for the asymptotic analysis, but only assumes ε_t is i.i.d. with eight moments, although the moment condition needs to be strengthened for the asymptotic distribution theory.

The conditions in Assumption 4 guarantee that the lag length is well defined and that the parameters are identified, see JN(2012a, Section 2.5) and Carlini and Santucci de Magistris (2017), who discuss identification of the parameters when the lag length is not fixed.

However, Assumption 4 does not impose that the cointegrating relations $\beta_0' X_t$ are (asymptotically) stationary, i.e. satisfy $0 \leq d_0 - b_0 < 1/2$, as in JN(2012a) and JN(2018), but instead only imposes $d_0 - b_0 \geq 0$, thus allowing both stationary and nonstationary cointegrating relations, and also allowing the important special case of $d_0 = b_0 (= 1)$.

We are now ready to state our main results in the following two theorems. The corresponding theorems in both JN(2012a) and JN(2018) required some strengthening of the moment assumptions, and these are avoided here.

Theorem 1 *Let Assumptions 1–4 hold and let the parameter space $\mathcal{N}(\eta, \eta_1, d_1)$ be given in (3), where η and η_1 are chosen such that $0 < \eta \leq 1/2$ and $0 < \eta_1 < 1/4$. Then, with probability converging to one, $\{\hat{d}, \hat{b}, \hat{\alpha}, \hat{\beta}, \hat{\Gamma}_1, \ldots, \hat{\Gamma}_k, \hat{\Omega}\}$ exists uniquely for $(d, b) \in \mathcal{N}$, and is consistent.*

The proof of Theorem 1 is given in Section 5. As discussed above, we note that Theorem 1 holds without any additional moment conditions beyond that in Assumption 2 and that it applies both when the cointegrating errors are stationary or nonstationary. In comparison, the moment conditions required in JN(2012a) and JN(2018) are summarized in Table 1 of JN(2018) and are much more involved than the simple condition in Assumption 2. For example, for consistency in the case with $b_0 > 1/2$, they require the additional moment condition $E|\varepsilon_t|^q < \infty$ for some $q > 1/\min\{\eta/3, (1/2 - d_0 + b_0)/2, b_0 - 1/2\}$.

The next theorem presents the asymptotic distributions of the estimators. For this result we will need to strengthen the condition in Assumption 3 on the initial values of the process and impose the following assumption, which was also made in JN(2012a) and JN(2018). The stochastic terms are not influenced by Assumption 5.

Assumption 5 *Either of the following conditions hold:*

(i) $\sup_{n \geq 0} |X_{-n}| < \infty$ *and the sum* $\sum_{n=1}^{\infty} n^{-1/2} |X_{-n}|$ *is finite,*

(ii) $\sup_{n \geq 0} |X_{-n}| < \infty$ *and* $X_{-n} = 0$ *for all* $n \geq M_0$ *for some* $M_0 \geq 0$.

The condition in Assumption 5(i) is that the (non-random) initial values satisfy the summability condition $\sum_{n=1}^{\infty} n^{-1/2} |X_{-n}| < \infty$. This allows the initial values to be non-zero back to the infinite past, but the summability condition implies that initial values do not influence the asymptotic distributions. For example, Assumption 5(i) would be satisfied if $|X_{-n}| \leq c n^{-1/2-\epsilon}$ for all $n \geq 1$ and a fixed $\epsilon > 0$.

Alternatively, under Assumption 5(ii), the initial values are assumed to be zero before some time in the past; that is, $X_{-n} = 0$ for all $n \geq M_0$, where $M_0 \geq 0$ is fixed.[1] Assumption 5(ii) is illustrated in the following display, see also (4):

$$\underbrace{\ldots, X_{-M_0}}_{\text{Data does not exist}}, \quad \underbrace{X_{1-M_0}, \ldots, X_{-N}}_{\substack{\text{Data exists}\\ \text{but is not observed}}}, \quad \underbrace{X_{1-N}, \ldots, X_0}_{\substack{\text{Data is observed}\\ \text{(initial values)}}}, \quad \underbrace{X_1, \ldots, X_T}_{\substack{\text{Data is observed}\\ \text{(estimation)}}} \tag{10}$$

[1] JN(2018) argue that M_0 could be allowed to diverge without altering the results as long as $M_0/\sqrt{T} \to 0$, as in Section 4.2 of JN(2012a). To avoid further notational complexity we do not consider this possibility.

Note that M_0 is a feature of the data generating process and is not related to N, which is chosen in the analysis of the data. The condition in Assumption 5(ii) was also imposed by JN(2016), and they provide some motivation for this assumption based on political polling data and financial volatility data.

Theorem 2 *Let Assumptions 1–5 hold with* $(d_0, b_0) \in \operatorname{int}(\mathcal{N})$ *and let the parameter space* $\mathcal{N}(\eta, \eta_1, d_1)$ *be given in (3), where* η *and* η_1 *are chosen such that* $0 < \eta \leq 1/2$ *and* $0 < \eta_1 < 1/4$. *Then the following hold.*

(i) If $b_0 < 1/2$ *the distribution of* $\{\hat{d}, \hat{b}, \hat{\alpha}, \hat{\beta}, \hat{\Gamma}_1, \ldots, \hat{\Gamma}_k\}$ *is asymptotically normal.*

(ii) If $b_0 > 1/2$ *we assume, in addition, that* $E|\varepsilon_t|^q < \infty$ *for some* $q > (b_0 - 1/2)^{-1}$. *Then the distribution of* $\{\hat{d}, \hat{b}, \hat{\alpha}, \hat{\Gamma}_1, \ldots, \hat{\Gamma}_k\}$ *is asymptotically normal and the distribution of* $\hat{\beta}$ *is asymptotically mixed Gaussian, and the two are independent.*

Proof of Theorem 2. This follows from parts (i) and (ii) of Theorem 10 in JN(2012a). Specifically, the proof of Theorem 10 in JN(2012a) relies on the usual Taylor expansion of the score function around the true values, and this applies to the current setting as well without any changes. ■

Note that the moment condition $q > (b_0 - 1/2)^{-1}$ in part (ii) of Theorem 2 is used in the proof of Theorem 10 in JN(2012a) to apply the functional CLT for processes that are fractional of order b_0 and obtain convergence to fractional Brownian motion, see also JN(2012b). This fractional Brownian motion appears in the mixed Gaussian asymptotic distribution of $\hat{\beta}$.

3 Bounds on product moments

We analyze product moments of processes that are either asymptotically stationary, near critical, or nonstationary, and we first define the corresponding fractional indices and the relevant class of processes. We use Definition A.1 from JN(2012a):

Definition 1 *We define* $\mathcal{S}(\kappa_w, \underline{\kappa}_v, \overline{\kappa}_v, \kappa_u)$ *as the set where the three fractional indices* $w, v,$ *and* u *are in the intervals*

$$[-w_0, -1/2 - \kappa_w], \; [-1/2 - \underline{\kappa}_v, -1/2 + \overline{\kappa}_v], \; [-1/2 + \kappa_u, u_0], \tag{11}$$

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

In the following we shall in fact always choose $\kappa_u = \kappa_w$, and $\overline{\kappa}_v = \underline{\kappa}_v$, where the last choice requires an argument, which we give when the results on the asymptotic behaviour of moments JN(2012a) are applied.

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 $\sum_{n=0}^{\infty}\xi_n\varepsilon_{t-n} = (\sum_{n=0}^{\infty}\xi_n)\varepsilon_t + \Delta\sum_{n=0}^{\infty}\xi_n^*\varepsilon_{t-n} \in \mathcal{Z}_1$. The main representation theorem in JN(2012a) shows that the solution of equation (1) is given in terms of processes in the classes $\mathcal{Z}_b$ or $\mathcal{Z}_b^+$. Thus for $Z_t \in \mathcal{Z}_b, b > 0$, and indices $(w, v, u) \in \mathcal{S}(\kappa_w, \underline{\kappa}_v, \overline{\kappa}_v, \kappa_u)$ as in Definition 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 $\xi\Delta_+^{-1/2}\varepsilon_t$.

We define product moments of fractional differences of processes in the class $\mathcal{Z}_{b_0}$, see Definition 2. For $m = m_1 + m_2$ we define the product moments

$$\mathsf{D}^m M_T(a_1, a_2) = T^{-1}\sum_{t=1}^{T}(\mathsf{D}^{m_1}\Delta_+^{a_1}Z_{1t}^+)(\mathsf{D}^{m_2}\Delta_+^{a_2}Z_{2t}^+)', \tag{12}$$

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

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

where a_1, a_2, a_3 can be u, w, and v in the intervals in Definition 1.

In the following, we will consider $M_T(a_1, a_2)$ as processes in the space of continuous functions, typically $\mathbb{C}^p(\mathcal{K})$, indexed by $(a_1, a_2) \in \mathcal{K}$, where $\mathcal{K}$ is a compact set in $\mathbb{R}^2$; see Billingsley (1968) or Kallenberg (2001) for the general theory. Let N_T be a normalizing sequence. We define $M_T(a_1, a_2) = \mathbf{O}_P(N_T)$ to mean that $N_T^{-1}M_T(a_1, a_2)$ is tight as a process in $\mathbb{C}^p(\mathcal{K})$ indexed by (a_1, a_2), and hence $\sup_{(a_1,a_2)\in\mathcal{K}}|N_T^{-1}M_T(a_1, a_2)|$ is tight. Similarly $M_T(a_1, a_2) = \mathbf{o}_P(N_T)$ means that $N_T^{-1}M_T(a_1, a_2)$ is tight as a process in $\mathbb{C}^p(\mathcal{K})$ indexed by (a_1, a_2) and that $\sup_{(a_1,a_2)\in\mathcal{K}}|N_T^{-1}M_T(a_1, a_2)| \overset{\text{P}}{\to} 0$. Finally, $\Longrightarrow$ is used for convergence in distribution as a process on a function space ($\mathbb{C}^p$ or $\mathbb{D}^p$). For example, $N_T^{-1}M_T(a_1, a_2) \Longrightarrow M(a_1, a_2)$ means that $N_T^{-1}M_T(a_1, a_2)$ converges in distribution as a process in $\mathbb{C}^p(\mathcal{K})$ to the limit $M(a_1, a_2)$.

When the product moments include nonstationary processes, these need further normalization. Therefore, we introduce the notation $M_T^{**}(w_1, w_2) = T^{w_1+w_2+1}M_T(w_1, w_2)$ and $M_T^*(w_1, a) = T^{w_1+1/2}M_T(w_1, a)$, where a can be u or v, to indicate that the nonstationary processes have been normalized by $T^{w_i+1/2}$.

For the asymptotic analysis we sometimes apply the result that, when $w < -1/2$ and $E|\varepsilon_t|^q < \infty$ for some $q > \max\{2, -1/(w + 1/2)\}$, then for $Z_t \in \mathcal{Z}_b$, $b > 0$, we have

$$T^{w+1/2}\Delta_+^w Z_{[Tu]}^+ \Longrightarrow \xi W_{-w-1}(u) = \xi\Gamma(-w)^{-1}\int_0^u (u - s)^{-w-1}(dW) \text{ on } \mathbb{D}^p([0, 1]), \tag{13}$$

where $\Gamma(\cdot)$ is the gamma function and W denotes p-dimensional Brownian motion (BM) generated by ε_t. The process W_{-w-1} is the corresponding fractional Brownian motion (fBM) of type II. The proof of (13) is given in JN(2010, Lemma D.2) for $Z_t \in \mathcal{Z}_b, b > 0$, see also Taqqu (1975) for $Z_t = \varepsilon_t$. Note that the moment condition $q > \max\{2, 1/(-w - 1/2)\}$ is in fact necessary; see JN(2012b).

The next lemma is Lemma A.9 of JN(2012a) and is reproduced here for ease of reference, although the results are presented in a different order. It contains the key results on the asymptotic behavior of product moments of processes that can be stationary, nonstationary, or critical, in the sense of the intervals in Definition 1.

Lemma 1 (JN(2012a) Lemma A.9) *Let* $Z_{it} = \xi_i\varepsilon_t + \Delta^{b_0}\sum_{n=0}^{\infty}\xi_{in}^*\varepsilon_{t-n} \in \mathcal{Z}_{b_0}, i = 1,2,$ *define* $M_T(a_1, a_2)$ *in (12), and assume that* $E|\varepsilon_t|^8 < \infty$. *Then it holds jointly that:*

(i) Uniformly for $(w_i, v, u_j) \in \mathcal{S}(\kappa_w, \underline{\kappa}_v, \overline{\kappa}_v, \kappa_u)$, $i = 1,2, j = 1,2,$

$$\mathsf{D}^m M_T(u_1, u_2) \Longrightarrow \mathsf{D}^m E(\Delta^{u_1}Z_{1t})(\Delta^{u_2}Z_{2t})', \tag{14}$$

$$\mathsf{D}^m M_T^{**}(w_1, w_2) = \mathbf{O}_P(1), \tag{15}$$

$$\mathsf{D}^m M_T^{*}(w, u) = \mathbf{O}_P((1+\log T)^{2+m}T^{-\min(\kappa_u,\kappa_w)}), \tag{16}$$

$$M_T^{*}(w, v) = \boldsymbol{O}_P((1+\log T)^{2}T^{\underline{\kappa}_v}), \tag{17}$$

$$M_T(v, u) = \boldsymbol{O}_P(1). \tag{18}$$

(ii) If we choose $N = T^{\alpha}$ *with* $0 < \alpha < 1/4$, *then for* $-1/2 - \underline{\kappa}_v \le v_i \le -1/2 + \overline{\kappa}_v$, $i = 1,2$, *we find*

$$M_T((v_1, v_2), (v_1, v_2)) \ge c\frac{1 - N^{-2\overline{\kappa}_v}}{2\overline{\kappa}_v}[\xi_1', \xi_2']'\Omega_0[\xi_1', \xi_2'] + R_T, \tag{19}$$

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

(iii) Assume, in addition, that $E|\varepsilon_t|^q < \infty$ *for some* $q > \kappa_w^{-1}$. *Then, uniformly for* $-w_0 \le w_i \le -1/2 - \kappa_w$, $i = 1,2$,

$$M_T^{**}(w_1, w_2) \Longrightarrow \xi_1\int_0^1 W_{-w_1-1}(s)W_{-w_2-1}(s)'ds\xi_2'. \tag{20}$$

We note that (19) is proved in JN(2012a) for$\overline{\kappa}_v < \underline{\kappa}_v$, but the inequality

$$\min_{|v_i+1/2|\le\overline{\kappa}_v} M_T((v_1, v_2), (v_1, v_2)) \ge \min_{|v_i+1/2|\le\underline{\kappa}_v} M_T((v_1, v_2), (v_1, v_2)) \tag{21}$$

shows that the result also holds for $\overline{\kappa}_v = \underline{\kappa}_v$.

The main problem in our model, which was not a problem in JN(2012a) because of their assumption that $d_0 - b_0 < 1/2$, is that the moment condition required in (20) becomes very strong. Indeed, we will require this result for κ_w arbitrarily small, and so the moment condition in (20) would require existence of all moments of ε_t. For example, we will need (20) with arbitrarily small κ_w to conclude that

$$M_T^{**}(w, w)^{-1} = \mathbf{O}_P(1) \text{ and } M_T^{**}(w_1, w_1|w_2)^{-1} = \mathbf{O}_P(1). \tag{22}$$

Thus, as an alternative to (20), we next prove the result (22) without the additional assumption that $E|\varepsilon_t|^q < \infty$ for some $q > \kappa_w^{-1}$. That is, we apply only the simpler assumption that $E|\varepsilon_t|^8 < \infty$.

3.1 Obtaining the bound (22) without additional moment conditions

The important realization here is that when $w < -1/2$ is arbitrarily close to $-1/2$, the moment condition $q > -1/(w + 1/2)$ required to obtain convergence of $T^{w+1/2}\Delta_+^w\varepsilon_t$ to fractional Brownian motion (see JN(2012b)) requires existence of all moments of ε_t. The basic idea in proving (22) is to derive a lower bound for $M_T^{**}(w_1, w_2)$ in which the relevant processes have fractional index that is sufficiently far away from the critical point $-1/2$ that the moment condition is implied by the existence of $q = 8$ moments.[2]

The next lemma provides the required result.

Lemma 2 *Let $Z_{it} = \xi_i \varepsilon_t + \Delta^{b_0} \sum_{n=0}^{\infty} \xi_{in}^* \varepsilon_{t-n} \in \mathcal{Z}_{b_0}$, where ξ_i is $m \times p$ and $\varepsilon_t \in \mathbb{R}^p$ with $E|\varepsilon_t|^8 < \infty$. Then, as $T \to \infty$, it holds uniformly for $-w_0 \leq w_i \leq -1/2 - \kappa_w$, $i = 1, 2$, that*

$$\det(M_T^{**}(w_1, w_1)) \geq (\frac{\pi^{2m}}{4^m} + O(T^{-1})) \det(M_T^{**}(w_1 - 1, w_1 - 1)), \tag{23}$$

$$\det(M_T^{**}((w_1, w_2), (w_1, w_2)) \geq (\frac{\pi^{4m}}{4^{2m}} + O(T^{-1})) \det(M_T^{**}((w_1 - 1, w_2 - 1), (w_1 - 1, w_2 - 1)), \tag{24}$$

where the $O(T^{-1})$ terms do not depend on w_i.

Proof. Define the $T \times m$ matrix $Z = [Z_1, \ldots, Z_T]'$ and the $T \times T$ fractional integration matrix

$$\Phi(w) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \pi_1(w) & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \pi_{T-1}(w) & \cdots & \pi_1(w) & 1 \end{bmatrix}$$

such that $\left[\Delta_+^w Z_1, \ldots, \Delta_+^w Z_T\right]' = \Phi(-w) Z$ and $M_T^{**}(w, w) = T^{2w} Z' \Phi(-w)' \Phi(-w) Z$. We note the following properties of $\Phi(w)$:

$$\det(\Phi(w)) = 1, \ \Phi(w_1)\Phi(w_2) = \Phi(w_1 + w_2), \text{ and } \Phi(w)^{-1} = \Phi(-w). \tag{25}$$

The first property in (25) is trivial, the second follows because $\Delta_+^{w_1} \Delta_+^{w_2} Z_t = \Delta_+^{w_1 + w_2} Z_t$ (see Lemma A.4 in JN(2016)), and the third property is a consequence of the second property using $\Phi(0) = I_T$.

Proof of (23): With this notation we find, using $X = \Phi(1 - w_1) Z$ and the properties in (25), that

$$\Phi(-w_1) Z = \Phi(-1)\Phi(1 - w_1) Z = \Phi(-1) X$$

and

$$\begin{aligned} \frac{\det(M_T^{**}(w_1, w_1))}{\det(M_T^{**}(w_1 - 1, w_1 - 1))} &= \frac{T^{2w_1 m}}{T^{2(w_1 - 1)m}} \frac{\det(Z'\Phi(-w_1)'\Phi(-w_1)Z)}{\det(Z'\Phi(1 - w_1)'\Phi(1 - w_1)Z)} \\ &= T^{2m} \frac{\det(X'\Phi(-1)'\Phi(-1)X}{\det(X'X)} \\ &\geq T^{2m} \lambda_{\min}^m \left(\Phi(-1)'\Phi(-1)\right), \end{aligned} \tag{26}$$

[2] Hualde and Robinson (2011, eqn. (2.36), p. 3163) faced a similar problem for a univariate process. By the Cauchy-Schwarz inequality they reduced the problem to showing that, for suitable $w_1 < w_2 < -1/2$, the random variable $M = \inf_{w_1 \leq w \leq w_2} W_{-w}(1)^2$ is positive almost surely. This, however, cannot be correct. The two-dimensional random variable $(W_{-w_1}(1), W_{-w_2}(1))$ has a nonsingular zero-mean Gaussian distribution and the set $A = \{W_{-w_1}(1) < 0 < W_{-w_2}(1)\}$ satisfies $P(A) > 0$. Because $W_{-w}(1)$ is continuous in w for $w < -1/2$, there exists on A a $w^* \in [w_1, w_2]$ for which $W_{-w^*}(1) = M = 0$, and therefore $P(M = 0) \geq P(A) > 0$.

where the inequality follows from, e.g., Horn and Johnson (2013, p. 258). From Rutherford (1948), see also Tanaka (1996, eqn. (1.4)), we find the eigenvalues

$$\lambda_t\left(\Phi(-1)'\Phi(-1)\right) = 4\sin^2\left(\frac{\pi}{2}\frac{2t-1}{2T+1}\right), \quad t = 1, \ldots, T,$$

such that, in particular,

$$\lambda_{\min}\left(\Phi(-1)'\Phi(-1)\right) = 4\sin^2\left(\frac{\pi}{2}\frac{1}{2T+1}\right) = \frac{\pi^2}{4}T^{-2} + O(T^{-3}). \tag{27}$$

The bound (23) follows by combining (26) and (27).

Proof of (24): Define $Z_i = [Z_{1i}, \ldots, Z_{Ti}]'$ for $i = 1, 2$ and the block matrices $\tilde{Z} = \operatorname{block\,diag}\{Z_1, Z_2\}$, $\tilde{\Phi}(w_1, w_2) = \operatorname{block\,diag}\{\Phi(w_1), \Phi(w_2)\}$. Then, as in (26), we find

$$\begin{aligned}
&\frac{\det(M_T^{**}((w_1, w_2), (w_1, w_2)))}{\det(M_T^{**}((w_1-1, w_2-1), (w_1-1, w_2-1)))} \\
&= T^{4m}\frac{\det(\tilde{Z}'\tilde{\Phi}(-w_1, -w_2)'\tilde{\Phi}(-w_1, -w_2)\tilde{Z})}{\det(\tilde{Z}'\tilde{\Phi}(1-w_1, 1-w_2)'\tilde{\Phi}(1-w_1, 1-w_2)\tilde{Z})} \geq T^{4m}\lambda_{\min}^{2m}(\tilde{\Phi}(-1, -1)),
\end{aligned}$$

and the result follows by (27). ■

The lower bound obtained in Lemma 2 is used in the next lemma to provide a justification for the tightness of the inverse in (22).

Lemma 3 *Let $Z_{it} = \xi_i\varepsilon_t + \Delta^{b_0}\sum_{n=0}^{\infty}\xi_{in}^*\varepsilon_{t-n} \in \mathcal{Z}_{b_0}$, where $\xi_i, i = 1, 2$, is $m \times p$, $[\xi_1', \xi_2']$ has full rank, and $\varepsilon_t \in \mathbb{R}^p$ with $E|\varepsilon_t|^8 < \infty$. Then, as $T \to \infty$, it holds uniformly for $-w_0 \leq w_i \leq -1/2 - \kappa_w, i = 1, 2$, that*

$$M_T^{**}(w_1, w_1)^{-1} = \mathbf{O}_P(1), \tag{28}$$

$$M_T^{**}(w_1, w_1|w_2)^{-1} = \mathbf{O}_P(1). \tag{29}$$

Proof. *Proof of (28)*: We want to show that

$$P(\inf_{-w_0 \leq w_1 \leq -1/2-\kappa_w} \det(M_T^{**}(w_1, w_1)) = 0) \to 0.$$

We apply the bound (23) in Lemma 2, so that we have to analyze $\det(M_T^{**}(w_1-1, w_1-1))$. Because $w_1 - 1 \leq -3/2 - \kappa_w$, the additional moment condition for weak convergence in (13) and (20) becomes $q > (1+\kappa_w)^{-1}$ and is not binding for $M_T^{**}(w_1-1, w_1-1)$ since $q = 8$ moments are assumed. Thus, we can apply Lemma 1(iii) to obtain

$$\inf_{-w_0 \leq w_1 \leq -1/2-\kappa_w} \det(M_T^{**}(w_1-1, w_1-1)) \overset{\mathrm{D}}{\to} \inf_{-w_0 \leq w_1 \leq -1/2-\kappa_w} \det(\xi_1\int_0^1 W_{-w_1}(u)W_{-w_1}(u)'du\xi_1'), \tag{30}$$

where the right-hand side is positive almost surely. To see this, assume that there were a point $w^* \in [-w_0, -1/2-\kappa_w]$, for which

$$\det(\xi_1\int_0^1 W_{-w^*}(u)W_{-w^*}(u)'du\xi_1') = 0,$$

which implies that for some non-zero vector $\lambda \in \mathbb{R}^p$,

$$\int_0^1 \lambda' W_{-w^*}(u) W_{-w^*}(u)' \lambda du = 0,$$

and hence $\lambda' W_{-w^*}(u) = 0$ for almost all $u \in [0,1]$. Using partial integration and the recurrence relation for the Gamma function,

$$\lambda' W_{-w^*}(u) = \frac{1}{\Gamma(-w^*)} \int_0^u \lambda' W(s)(u-s)^{-w^*-1} ds,$$

so that $\lambda' W_{-w^*}(u) = 0$ implies that the Brownian path $\lambda' W(u)$ is zero (for almost all u), which clearly has probability zero.

It follows that, because 0 is a continuity point for the limit distribution,

$$\begin{aligned}
&P\left(\inf_{-w_0 \le w_1 \le -1/2-\kappa_w} \det(M_T^{**}(w_1,w_1)) = 0\right) \\
&\le P\left(\left(\frac{\pi^{2m}}{4^m} + O(T^{-1})\right) \inf_{-w_0 \le w_1 \le -1/2-\kappa_w} \det(M_T^{**}(w_1-1,w_1-1)) = 0\right) \\
&\to P\left(\frac{\pi^{2m}}{4^m} \inf_{-w_0 \le w_1 \le -1/2-\kappa_w} \det(\xi_1 \int_0^1 W_{-w_1}(u) W_{-w_1}(u)' du \xi_1') = 0\right) = 0.
\end{aligned}$$

Proof of (29): For notational simplicity we use $M_T = M_T^{**}((w_1-1,w_2-1),(w_1-1,w_2-1))$ in this proof. From (24) we have, for T sufficiently large,

$$\det(M_T^{**}(w_1,w_1|w_2)) = \frac{\det(M_T^{**}((w_1,w_2),(w_1,w_2)))}{\det(M_T^{**}(w_2,w_2))} \ge c\frac{\det(M_T)}{\det(M_T^{**}(w_2,w_2))}$$

for some finite constant $c > 0$.

By (15) of Lemma 1 we find that $\sup_{-w_0 \le w_2 \le -1/2-\kappa_w} \det(M_T^{**}(w_2,w_2)) = O_P(1)$, so that

$$P\left(\inf_{-w_0 \le w_i \le -1/2-\kappa_w} \det(M_T^{**}(w_1,w_1|w_2)) = 0\right) \le P\left(\inf_{-w_0 \le w_i \le -1/2-\kappa_w} \det(M_T) = 0\right). \quad (31)$$

Again the additional moment condition for weak convergence in (13) and (20) is not binding for M_T because $q = 8$ is assumed, so by (20) of Lemma 1 we find

$$\inf_{-w_0 \le w_i \le -1/2-\kappa_w} \det(M_T) \xrightarrow{D} \inf_{-w_0 \le w_i \le -1/2-\kappa_w} \det(\int_0^1 X(u)X(u)' du),$$

where $X(u) = \operatorname{diag}\{\xi_1, \xi_2\} \operatorname{diag}\{W_{-w_1}(u), W_{-w_2}(u)\}$ such that the right-hand side is positive almost surely. Hence,

$$P\left(\inf_{-w_0 \le w_i \le -1/2-\kappa_w} \det(M_T) = 0\right) \to P\left(\inf_{-w_0 \le w_i \le -1/2-\kappa_w} \det(\int_0^1 X(u)X(u)' du) = 0\right) = 0, \quad (32)$$

and the result follows by (31) and (32). ■

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 the following corollary. This corollary is identical to Corollary A.10 in JN(2012a), and the proof is given there, with the exception that we apply our Lemma 2 to avoid the additional moment condition in Lemma 1(iii).

Corollary 1 (JN(2012a) Corollary A.10) *Let the assumptions of Lemma 1 be satisfied. Then the following hold uniformly in* $(w, v, u) \in \mathcal{S}(\kappa_w, \underline{\kappa}_v, \overline{\kappa}_v, \kappa_u)$*, see (11) of Definition 1:*

(i) It holds that

$$M_T^{**}(w_1, w_2|w_3, u) = M_T^{**}(w_1, w_2|w_3) + \boldsymbol{o}_P(1), \tag{33}$$
$$M_T(u_1, u_2|w, u_3) \Longrightarrow Var(\Delta^{u_1} Z_{1t}, \Delta^{u_2} Z_{2t}|\Delta^{u_3} Z_{3t}), \tag{34}$$
$$M_T(v, u_1|w, u_2) = \boldsymbol{O}_P(1). \tag{35}$$

(ii) If $N = T^\alpha$ *with* $0 < \alpha < 1/4$*, then*

$$M_T((v_1, v_2), (v_1, v_2)|w, u) \geq c \frac{1 - N^{-2\overline{\kappa}_v}}{2\overline{\kappa}_v} (\xi_1', \xi_2')' \Omega_0 (\xi_1', \xi_2') + R_T, \tag{36}$$

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

Note that the result (36) is valid also for $\underline{\kappa}_v = \overline{\kappa}_v$, see (21). We apply the results of Lemmas 1 and 2 and Corollary 1 in the analysis of $\ell_{T,p}(\psi)$ and $\ell_{T,r}(\psi)$ to show that they converge uniformly in ψ, which is the key ingredient in the proof of consistency of the MLE. The results for $m = 0, 1, 2$ in Lemmas 1 and 2 are used to show that the information matrix is tight in a neighborhood of the true value.

4 Conclusions and discussion

In this paper we have analyzed the fractional cointegrated VAR model of Johansen and Nielsen (2012a) and made two distinct contributions. First, in their consistency proof, Johansen and Nielsen (2012a) imposed moment conditions on the errors that depend on the parameter space, such that when the parameter space is larger, stronger moment conditions are required. We have shown that these moment conditions can be relaxed, and for consistency we require just eight moments regardless of the parameter space. In light of the complicated moment conditions of Johansen and Nielsen (2012a, 2018) summarized in Table 1 of Johansen and Nielsen (2018), our contribution provides a substantial simplification of the assumptions.

Second, Johansen and Nielsen (2012a) assumed that the cointegrating vectors are stationary. However, nonstationary cointegrating relations have been found in much empirical work; see, e.g., the references cited in the introduction. In this paper, we have therefore extended the analysis to allow the cointegrating vectors in the fractionally cointegrated VAR model to also be nonstationary.

Finally, our analysis has followed recent research in Johansen and Nielsen (2018) and applied a parameter space large enough that the usual (non-fractional) cointegrated VAR model constitutes an interior point and hence can be tested against the fractional model using a χ^2-test.

The main technical contribution that has allowed these extensions of the theory is the proof of a strictly positive lower bound for product moments of nonstationary processes, assuming only a relatively weak moment condition. This bound is proved in Lemmas 2 and 3.

5 Proof of Theorem 1

Theorem 4 of JN(2012a) gives, under their assumptions, the properties of the likelihood function. These are used in Theorem 5(i) of JN(2012a) to show that the maximum likelihood estimator exists uniquely with large probability for large T, i.e. to prove the result in our Theorem 1 for the smaller parameter set, with $\eta_1 = 0$ in (3). Thus, if the results of Theorem 4 of JN(2012a) can be established under our assumptions, which are weaker than those in JN(2012a) as discussed above, then we can apply the proof of Theorem 5(i) of JN(2012a) with minor changes to prove our Theorem 1 for the larger parameter set with $\eta_1 > 0$ in (3).

The strategy of the proof is as follows. First, we discuss the solution of the autoregressive equations, i.e., the representation theory. Then we show that the contribution of the initial values $\{X_{-n}\}_{n\geq 0}$ to the likelihood function is negligible, such that we only need to analyze the stochastic terms. Next, we analyze the likelihood function and discuss the convergence and divergence of the likelihood on different parts of the parameter space. This establishes the notation necessary to present a version of Theorem 4 of JN(2012a), stated as Theorem 3, which we subsequently prove.

5.1 Solution of the equations

The solution, $X_t, t \geq 1$, of the equations (1) for the DGP is found in Theorem 2 of JN(2012a) under Assumptions 1–4 as

$$X_t = C_0 \Delta_+^{-d_0} \varepsilon_t + \Delta_+^{-(d_0-b_0)} Y_t^+ + \mu_{0t} \text{ for } d_0 \geq 1/2, \tag{37}$$

$$X_t = C_0 \Delta^{-d_0} \varepsilon_t + \Delta^{-(d_0-b_0)} Y_t \text{ for } d_0 < 1/2. \tag{38}$$

Here, $Y_t = \sum_{n=0}^{\infty} \tau_{0n} \varepsilon_{t-n}$ is a stationary process and $Y_t^+ = \sum_{n=0}^{t-1} \tau_{0n} \varepsilon_{t-n}$, for some matrix coefficients τ_{0n} depending only on the true values and satisfying $\sum_{n=0}^{\infty} |\tau_{0n}| < \infty$. The matrix C_0 is given by

$$C_0 = \beta_{0\perp}(\alpha_{0\perp}' \Gamma_0 \beta_{0\perp})^{-1} \alpha_{0\perp}'. \tag{39}$$

Moreover, by a fractional version of the Beveridge-Nelson decomposition, see eqn. (12) of JN(2012a), the stationary process Y_t can be written as $Y_t = C_0^* \varepsilon_t + \Delta^{b_0} \sum_{n=0}^{\infty} \tau_{0n}^* \varepsilon_{t-n}$ for some matrix coefficients τ_{0n}^* depending only on the true values and satisfying $\sum_{n=0}^{\infty} |\tau_{0n}^*| < \infty$, where $C_0^* = \sum_{n=0}^{\infty} \tau_{0n}$ satisfies

$$\beta_0' C_0^* \alpha_0 = -I_r. \tag{40}$$

We let $\mathbb{I}\{\cdot\}$ denote the indicator function and define

$$\Psi_+(L) X_t = \mathbb{I}\{t \geq 1\} \sum_{i=0}^{t-1} \Psi_i X_{t-i} \text{ and } \Psi_-(L) X_t = \sum_{i=t}^{\infty} \Psi_i X_{t-i},$$

see (8). Then the term μ_{0t} is

$$\mu_{0t} = -\Psi_{0+}(L)^{-1} \Psi_{0-}(L) X_t,$$

which expresses μ_{0t} as function of the fixed initial values $\{X_{-n}\}_{n=0}^{\infty}$. It is seen from (6) and (37), say, that the likelihood function contains terms of the form $\Delta_N^{d+ib} X_t$ from which there are both stochastic and deterministic contributions. The latter arise because the likelihood is analyzed conditional on the initial values. Therefore, the terms in the likelihood function that are generated by the initial values are considered deterministic in the analysis of the model.

We define $\tilde{X}_t = X_t \mathbb{I}\{1-N \le t \le 0\}$ as the initial values used in the calculations, see (6) and (7). We also define the operator Δ_- such that, for any Z_t and any $a \ge 0$ it holds that $\Delta^a Z_t = \Delta_+^a Z_t + \Delta_-^a Z_t$. When $d_0 \ge 1/2$, the deterministic terms in the likelihood are simple functions of

$$D_{it}(d,b) = \begin{cases} (\Delta_-^{d-b} - \Delta_-^{d})\tilde{X}_t + (\Delta_+^{d-b} - \Delta_+^{d})\mu_{0t}, & i=-1, \\ (\Delta_-^{d+ib} - \Delta_-^{d+kb})\tilde{X}_t + (\Delta_+^{d+ib} - \Delta_+^{d+kb})\mu_{0t}, & i=0,\ldots,k-1, \\ \Delta_-^{d+kb}\tilde{X}_t + \Delta_+^{d+kb}\mu_{0t}, & i=k, \end{cases} \tag{41}$$

see eqn. (14) in JN(2018). When $d_0 < 1/2$, we use a different representation of the solution and hence leave out the terms involving μ_{0t} in (41), see Theorem 2 in JN(2012a).

5.2 Negligibility of initial values

We now establish that the deterministic terms generated by the initial values are uniformly negligible. This follows from results in JN(2018), which generalizes JN(2012a) to apply to the larger parameter space $\mathcal{N}$. In particular, we will apply the results in Lemma 1(i) of JN(2018) to conclude that deterministic terms from initial values do not influence the limit behavior of product moments, and hence do not influence the limit behavior of the likelihood function. For ease of reference, we quote Lemma 1(i) of JN(2018), where the terms $D_{it}(d,b)$ are the initial values contributions to the likelihood function given in (41) and D^m denotes m'th order derivatives with respect to d and/or b.

Lemma 4 (JN(2018) Lemma 1(i)) *Let Assumption 3 be satisfied. Choose κ_1 and η_1 such that $0 < \eta_1 < \kappa_1 < 1/4$ and define the intervals $S_+ = [d_0 - 1/2 - \kappa_1, \infty[$ and $S_- = [-\eta_1, d_0 - 1/2 - \kappa_1]$. Then the functions $\mathsf{D}^m D_{it}(d,b)$ are continuous in $(d,b) \in \mathcal{N}(\eta, \eta_1, d_1)$ and satisfy*

$$\sup_{d+ib \in S_+} |\mathsf{D}^m D_{it}(d,b)| \to 0 \text{ as } t \to \infty, \tag{42}$$

$$\sup_{d+ib \in S_-} \max_{1 \le t \le T} |\mathsf{D}^m T^{d+ib-d_0+1/2} \beta_{0\perp}' D_{it}(d,b)| \to 0 \text{ as } T \to \infty. \tag{43}$$

We first note that the result in Lemma 4(i) does not depend on the assumption that $d_0 - b_0 < 1/2$, which was made in both JN(2012a) and JN(2018) to deal with the stochastic terms.

However, the results in eqns. (42) and (43) depend on the parameter $\kappa_1 > 0$, which in JN(2018) is the parameter that separates the nonstationary processes (with fractional index $\le -1/2 - \kappa_1$) from the stationary and near-critical processes (with fractional index $\ge -1/2 - \kappa_1$). This is important because the initial values contribution is normalized by $T^{w+1/2} = T^{d+ib-d_0+1/2}$ in the nonstationary case in (43) and not in the stationary and near-critical case in (42). It is further relevant because, as we will see below, in our case we will need to choose our version of κ_1 arbitrarily small, so because Lemma 4 requires setting κ_1 such that $\eta_1 < \kappa_1 < 1/4$, this suggests that we would need to let η_1 be arbitrarily small. To avoid this, additional arguments are needed, which we now provide.

In the present notation, the nonstationary processes have fractional index $w \leq -1/2 - \kappa_w$, and the stationary and near-critical processes have fractional index $\geq -1/2 - \kappa_w$, see Definition 1. Let $\eta_1 \in (0, 1/4)$ be fixed and choose any κ_1 such that $\eta_1 < \kappa_1 < 1/4$. Thus, in the present notation, κ_w is the parameter that separates the nonstationary processes from the stationary and near-critical processes, while κ_1 will be used for the application of Lemma 4. Because $\kappa_w > 0$ will be chosen arbitrarily small in the following subsections, we assume without loss of generality that $\kappa_w < \kappa_1$.

For the stationary and near-critical processes, which have fractional index $\geq -1/2 - \kappa_w \geq -1/2 - \kappa_1$, we can apply (42) directly to conclude that the initial values contribution is uniformly negligible. Similarly, for the (normalized) nonstationary processes with fractional index $\leq -1/2 - \kappa_1$, we can apply (43) to conclude that the contribution from the initial values is uniformly negligible. We are then left with the (normalized) nonstationary processes with fractional index in the interval $[-1/2 - \kappa_1, -1/2 - \kappa_w]$, for some arbitrarily small $\kappa_w > 0$. For this interval, we can apply the non-normalized result in (42) together with the evaluation $T^{w+1/2} \leq T^{-\kappa_w} \leq 1$, which shows that the normalized initial values contribution is smaller than the non-normalized contribution.

Thus, with these additional arguments, it follows from Lemma 4, that the initial values do not influence the limit behavior of product moments, and hence do not influence the limit behavior of the likelihood function. In the subsequent analysis of the likelihood function, we can therefore assume that the deterministic terms generated by the initial values are zero.

5.3 Convergence of the profile likelihood function

Because the deterministic terms generated by the initial values can be assumed to be zero, we can rewrite $\varepsilon_t(\lambda)$ in (6), and hence the likelihood (7), as

$$\begin{aligned}
\varepsilon_t(\lambda) &= \Delta_+^{d+kb} X_t - \alpha\beta'(\Delta_+^{d-b} - \Delta_+^{d}) X_t - \sum_{i=0}^{k-1} \Psi_{i+1} (\Delta_+^{d+ib} - \Delta_+^{d+kb}) X_t \\
&= X_{kt} - \alpha\beta' X_{-1,t} - \sum_{i=0}^{k-1} \Psi_{i+1} X_{it}, \qquad (44)
\end{aligned}$$

where

$$\begin{aligned}
X_{it} &= (\Delta_+^{d+ib} - \Delta_+^{d+kb}) X_t, \quad i = 0, \ldots, k-1, \\
X_{kt} &= \Delta_+^{d+kb} X_t, \text{ and } X_{-1t} = (\Delta_+^{d-b} - \Delta_+^{d}) X_t.
\end{aligned}$$

Moreover, the process X_t can be expressed in terms of the stationary process $C_0\varepsilon_t + \Delta^{b_0} Y_t$, see (37) and (38), so that X_{it} is expressed in terms of the differences $\Delta_+^{d+ib-d_0}(C_0\varepsilon_t + \Delta^{b_0} Y_t)$. We define δ_m as the fractional index for $\beta_{0\perp}' X_{mt}$ and ϕ_n as the fractional index for $\beta_0' X_{nt}$, and find for $m, n = -1, \ldots, k$ that

$$\delta_m = d + mb - d_0 \text{ and } \phi_n = d + nb - d_0 + b_0. \qquad (45)$$

For notational reasons we define $\phi_{-2} = \delta_{-2} = -\infty$ and $\phi_{k+1} = \delta_{k+1} = \infty$. Then $\beta_{0\perp}' X_{mt} \in I(-\delta_m)$ and $\beta_0' X_{nt} \in I(-\phi_n)$.

For fixed ψ, the conditional MLE is found from (7) and (44) by reduced rank regression of $X_{k,t}$ on $X_{-1,t}$ corrected for the regressors $X_{0t},\ldots,X_{k-1,t}$. We define the corresponding residuals

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

and their sums of squares

$$S_{ij}(\psi) = T^{-1}\sum_{t=1}^{T} R_{it}(\psi)R_{jt}(\psi)' \text{ for } i,j=0,1.$$

Then we solve the generalized eigenvalue problem

$$\det(\omega S_{11}(\psi) - S_{10}(\psi)S_{00}(\psi)^{-1}S_{01}(\psi)) = 0, \tag{46}$$

which gives eigenvalues $1 \geq \hat{\omega}_1(\psi) \geq \cdots \geq \hat{\omega}_p(\psi) > 0$ that all depend on ψ. The maximized likelihood (scaled by $-2T^{-1}$), for fixed ψ, is given by

$$\ell_{T,r}(\psi) = \log\det(S_{00}(\psi)) + \sum_{i=1}^{r}\log(1-\hat{\omega}_i(\psi)) = \log\det(\hat{\Omega}(\psi)), \tag{47}$$

see Johansen (1996) for the details in the $I(1)$ model or JN(2012a) and Nielsen and Popiel (2016) for details in the fractional CVAR. Finally, the MLE of ψ is found as the argmin of the profile likelihood in (47).

In the full rank case with $r = p$, the profile likelihood $\ell_{T,p}(\psi)$ is found by regression of X_{kt} on $\{X_{it}\}_{i=-1}^{k-1}$, i.e.,

$$\ell_{T,p}(\psi) = \log\det(SSR_T(\psi)) = \log\det(T^{-1}\sum_{t=1}^{T} R_t(\psi)R_t(\psi)'), \tag{48}$$

where $R_t(\psi) = (X_{kt}|\{X_{it}\}_{i=-1}^{k-1})$ denotes the regression residual. Equivalently, this is found by regressing $\beta_{0\perp}' X_{kt}$ and $\beta_0' X_{kt}$ on the regressors $\beta_{0\perp}' X_{mt}$ and $\beta_0' X_{mt}$ for $-1 \leq n, m \leq k-1$, where these can be either asymptotically stationary, near critical or nonstationary. We define $\mathcal{F}_{\text{stat}}(\psi)$ as the set of stationary regressors for a given ψ, and if $\Delta_+^{d+kb}X_t$ is stationary, we let $\Omega(\psi)$ be the variance of $\Delta_+^{d+kb}X_t$ conditional on the variables in $\mathcal{F}_{\text{stat}}(\psi)$. That is,

$$\mathcal{F}_{\text{stat}}(\psi) = \{\Delta_+^{d+mb}\beta_{0\perp}' X_t\colon\ \delta_m > -1/2, m < k\} \cup \{\Delta_+^{d+nb}\beta_0' X_t\colon\ \phi_n > -1/2, n < k\}, \tag{49}$$

$$\Omega(\psi) = Var(\Delta_+^{d+kb}X_t|\mathcal{F}_{\text{stat}}(\psi)) \text{ if } \Delta_+^{d+kb}X_t \text{ is stationary.} \tag{50}$$

We next define the probability limit, $\ell_p(\psi)$, of the profile likelihood function $\ell_{T,p}(\psi)$ in (48). The limit of $\log\det(SSR_T(\psi))$ is finite if X_{kt} is (asymptotically) stationary and infinite if X_{kt} is near critical or nonstationary, see Theorem 3. We therefore define the subsets of $\mathcal{N}$,

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

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

$$\ell_p(\psi) = \begin{cases} \infty & \text{if } \psi \in \mathcal{N}_{\text{div}}(0), \\ \log\det(\Omega(\psi)) & \text{if } \psi \in \mathcal{N}_{\text{conv}}(0). \end{cases} \tag{51}$$

We are now ready to state and prove a version of Theorem 4 of JN(2012a). We 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 $\mathcal{N}_{\text{div}}(\kappa_0)$ for $T \geq T_0$. Thus, the minimum of $\ell_{T,p}(\psi)$ cannot be attained on $\mathcal{N}_{\text{div}}(\kappa_0)$. On the rest of $\mathcal{N}$, 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 function $\ell_{T,p}(\psi)$ as a continuous process in the parameters.

Theorem 3 *Let Assumptions 1-4 hold, so that in particular $E|\varepsilon_t|^8 < \infty$ and $d_0 \geq b_0$. Then the following hold:*

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

$$\ell_p(\psi) \geq \ell_p(\psi_0) = \log\det(\Omega_0) \textit{ for } \psi \in \mathcal{N}, \tag{52}$$

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

(ii) For $r = 0, \ldots, p$ it holds that

$$\ell_{T,r}(\psi_0) \xrightarrow{\text{P}} \log\det(\Omega_0). \tag{53}$$

(iii) 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

$$P(\inf_{\psi \in \mathcal{N}_{div}(\kappa_0)} \ell_{T,p}(\psi) \geq A) \geq 1 - \gamma \textit{ for all } T \geq T_0. \tag{54}$$

It also holds that

$$\ell_{T,p}(\psi) \Longrightarrow \ell_p(\psi) \textit{ on } \mathbb{C}(\mathcal{N}_{conv}(\kappa_0)) \textit{ as } T \to \infty. \tag{55}$$

The remainder of this section is devoted to the proof of Theorem 3, and is divided into several subsections.

5.4 Proof of Theorem 3(i)

The proof of Theorem 3(i), i.e. of (52), in Appendix B.1 of JN(2012a) applies without change under our assumptions. In the next two subsections we give the proofs of Theorem 3(ii) and (iii), respectively.

5.5 Proof of Theorem 3(ii)

The proof of this result in Appendix B.2 of JN(2012a) applies (20) and the moment condition $E|\varepsilon_t|^q < \infty$ for some $q > (b_0 - 1/2)^{-1}$. With our new Lemma 3, we now give the proof without this condition and to the larger parameter set in (3). Throughout this proof all moment matrices (S_{ij} etc.) are evaluated at the true value $\psi = \psi_0$.

Case 1: $b_0 > 1/2$. For $\psi = \psi_0$, the regressand $\beta_{0\perp}' X_{-1,t}$ is nonstationary with index $\delta_{-1}^0 = -b_0 < -1/2$ and $\beta_0' X_{-1,t} = \beta_0' Y_t$ is stationary with index $\phi_{-1}^0 = 0$. Pre- and post-multiplying by the matrix $[T^{1/2-b_0}\beta_{0\perp}, \beta_0]'$ and its transposed in (46) we get

$$\det\left(\omega\begin{bmatrix}\beta_{0\perp}' S_{11}^{**}\beta_{0\perp} & \beta_{0\perp}' S_{11}^{*}\beta_0\\ \beta_0' S_{11}^{*}\beta_{0\perp} & \beta_0' S_{11}\beta_0\end{bmatrix} - \begin{bmatrix}\beta_{0\perp}' S_{10}^{*}\\ \beta_0' S_{10}\end{bmatrix} S_{00}^{-1}\begin{bmatrix}\beta_{0\perp}' S_{10}^{*}\\ \beta_0' S_{10}\end{bmatrix}'\right) = 0.$$

Here, $T^{1-2b_0}\beta_{0\perp}' S_{11}\beta_{0\perp}$ and its inverse are $O_P(1)$ by (15) of Lemma 1 and Lemma 3 (for fixed $\psi = \psi_0$), while $\beta_{0\perp}' S_{11}^{*}\beta_0$ and $\beta_{0\perp}' S_{10}^{*}$ are $O_P((1+\log T)^2 T^{-\min(1/2, b_0-1/2)})$ by (16) of Lemma 1. Furthermore, at the true value $\psi = \psi_0$, the following limits exist by the law of large numbers,

$$(\beta_0' S_{11}\beta_0, \beta_0' S_{10}, S_{00}) \overset{\mathrm{P}}{\to} (\Sigma_{\beta_0\beta_0}, \Sigma_{\beta_0 0}, \Sigma_{00}). \tag{56}$$

Thus, the solutions of the eigenvalue problem (46) are asymptotically equivalent to those of the eigenvalue problem

$$\begin{aligned}&\det\left(\begin{bmatrix}\omega\beta_{0\perp}' S_{11}^{**}\beta_{0\perp} & 0\\ 0 & \omega\beta_0' S_{11}\beta_0 - \beta_0' S_{10}S_{00}^{-1}S_{01}\beta_0\end{bmatrix}\right)\\ &= \det(\omega\beta_{0\perp}' S_{11}^{**}\beta_{0\perp})\det(\omega\beta_0' S_{11}\beta_0 - \beta_0' S_{10}S_{00}^{-1}S_{01}\beta_0) = 0.\end{aligned}$$

This equation has $p-r$ zero roots and r positive roots, where the latter are the roots of the equation $\det(\omega\beta_0' S_{11}\beta_0 - \beta_0' S_{10}S_{00}^{-1}S_{01}\beta_0) = 0$. Moreover, by (56), these roots converge to the solutions of the equation

$$\det(\omega\Sigma_{\beta_0\beta_0} - \Sigma_{\beta_0 0}\Sigma_{00}^{-1}\Sigma_{0\beta_0}) = 0,$$

which we denote $\omega_1^0, \ldots, \omega_r^0$. Therefore, see (47),

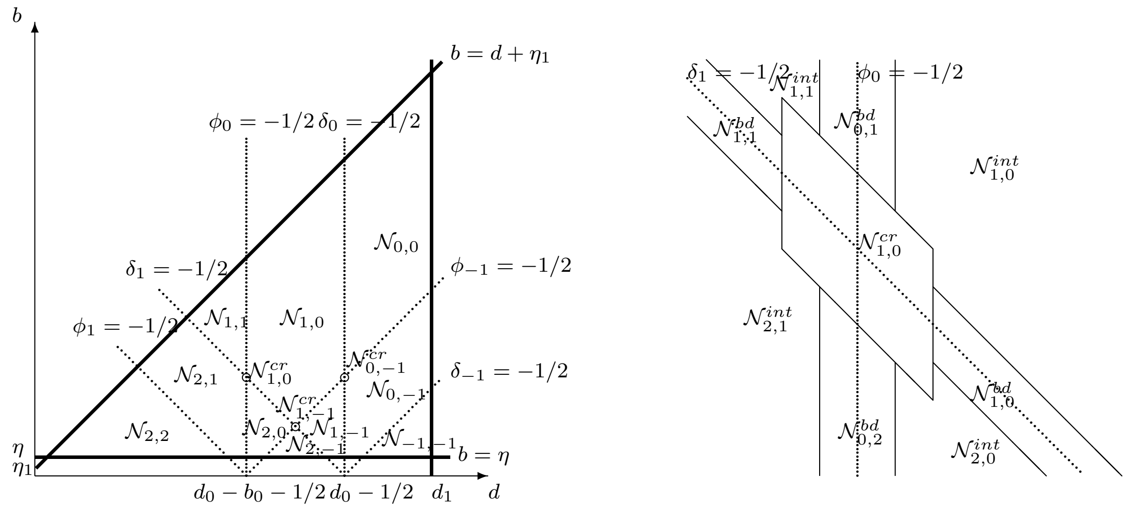
$$\ell_{T,r}(\psi_0) = \log\det(S_{00}) + \sum_{i=1}^r \log(1-\hat\omega_i) \overset{\mathrm{P}}{\to} \log\det(\Sigma_{00}) + \sum_{i=1}^r \log(1-\omega_i^0) = \log\det(\Omega_0). \tag{57}$$

Case 2: $b_0 < 1/2$. In this case $\beta_{0\perp}' X_{-1,t}$ is stationary with index $\delta_{-1}^0 = -b_0 > -1/2$. The profile likelihood is still given by (47), but (56) becomes $S_{ij} \overset{\mathrm{P}}{\to} \Sigma_{ij}^0 = E(S_{ij})$, say, and the limit (57) still holds.

5.6 Proof of Theorem 3(iii)

To analyze the properties of the likelihood function, the parameter space is partitioned as in Figure 1, using two sets of lines $\delta_m = -1/2$ and $\phi_n = -1/2$, where $n, m = -1, \ldots, k$. These lines may intersect, and close to these intersection points there are two nearly critical processes. Specifically, we partition the parameter space into "interiors", "critical points", and "boundaries", which depend on two parameters $0 < \kappa < \kappa_2$.

Figure 1: Illustration of parameter space with $k = 1$



(a) The parameter space $\mathcal{N}$ is bounded by the bold lines $b = \eta$, $b = d + \eta_1$, and $d = d_1$. The sets $\mathcal{N}^{cr}_{mn} = \mathcal{N}^{cr}_{mn}(\kappa_2)$, where two processes are close to being critical, and the sets $\mathcal{N}_{mn}$ are also shown (if $k > 1$ there would be more dotted lines).

(b) An intersection point: For $k \geq 1$ the set $\mathcal{N}^{cr}_{1,0}$ covers the intersection between the lines $\delta_1 = -1/2$ and $\phi_0 = -1/2$, where Z_{1t} and W_{0t} are nearly critical. The sets $\mathcal{N}^{int}_{mn} = \mathcal{N}^{int}_{mn}(\kappa)$ and $\mathcal{N}^{bd}_{mn} = \mathcal{N}^{bd}_{mn}(\kappa_2, \kappa)$ are also indicated.

Definition 3 *We define for $-1 \leq n \leq m \leq k+1$ a covering of the parameter set $\mathcal{N} = \cup_{-1 \leq n \leq m \leq k+1} \mathcal{N}_{mn}$ where*

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

as well as the corresponding interiors

$$\mathcal{N}^{int}_{mn}(\kappa) = \{\psi \in \mathcal{N} : \max(\delta_{m-1}, \phi_{n-1}) \leq -1/2 - \kappa \text{ and } -1/2 + \kappa \leq \min(\delta_m, \phi_n)\}. \quad (58)$$

We next define for $-1 \leq n < m \leq k$ and each pair of near critical indices $\delta_m = \phi_n = -1/2$, the critical sets around the point where $\delta_m = \phi_n = -1/2$:

$$\mathcal{N}^{cr}_{mn}(\kappa_2) = \{\psi : |\phi_n + 1/2| \leq \kappa_2, \text{ and } |\delta_m + 1/2| \leq \kappa_2\}, \quad (59)$$

Similarly we define for $-1 \leq n \leq m \leq k$ around each line where $\delta_n = -1/2$, the boundary sets

$$\mathcal{N}^{bd}_{mn}(\kappa_2, \kappa) = \{\psi : |\delta_m + 1/2| \leq \kappa, \text{ and } \phi_{n-1} \leq -1/2 - \kappa_2 \text{ and } -1/2 + \kappa_2 \leq \phi_n\}, \quad (60)$$

and for $-1 \leq m < n \leq k+1$ around the line where $\phi_m = -1/2$, the boundary sets

$$\mathcal{N}^{bd}_{mn}(\kappa_2, \kappa) = \{\psi : |\phi_m + 1/2| \leq \kappa, \text{ and } \delta_{n-1} \leq -1/2 - \kappa_2 \text{ and } -1/2 + \kappa_2 \leq \delta_n\}. \quad (61)$$

The interpretation of the sets $\mathcal{N}_{mn}$ is that, for $\psi \in \mathcal{N}_{mn}$, the processes $\beta'_{0\perp}\Delta^{d+mb}_{+}X_t$ and $\beta'_0\Delta^{d+nb}_{+}X_t$ are asymptotically stationary, whereas $\beta'_{0\perp}\Delta^{d+(m-1)b}_{+}X_t$ and $\beta'_0\Delta^{d+(n-1)b}_{+}X_t$ are either nonstationary or (nearly) critical. The true value ψ_0 is contained in $\mathcal{N}_{0,-1}$ if $b_0 > 1/2$ and in $\mathcal{N}_{-1,-1}$ if $b_0 < 1/2$.

Note that $\phi_m = \delta_m + b_0$, so that $\phi_m > \delta_m$, see (45). Hence, if $\beta_{0\perp}'\Delta^{d+mb}X_t$ of index δ_m is asymptotically stationary then $\beta_0'\Delta^{d+mb}X_t$ of index ϕ_m is also asymptotically stationary. Likewise, if $\beta_0'\Delta^{d+mb}X_t$ is nonstationary then $\beta_{0\perp}'\Delta^{d+mb}X_t$ is also nonstationary. Furthermore, note that if $\eta_1 < 1/2 - d_0 + b_0$, there are no crossing points and the sets $\mathcal{N}_{mn}^{cr}(\kappa_2)$ are empty, in which case the proof is easily simplified accordingly.

The set $\mathcal{N}_{mn}^{cr}(\kappa_2)$ contains the crossing point between the lines given by $\delta_m = -1/2$ and $\phi_n = -1/2$, where $\beta_{0\perp}'X_{mt}$ and $\beta_0'X_{nt}$ are both critical. To the left they are nonstationary and to the right stationary. The set $\mathcal{N}_{mn}^{bd}(\kappa_2,\kappa)$ covers the line segment $\phi_n = -1/2$ between $\mathcal{N}_{m,n-1}^{cr}(\kappa_2)$ and $\mathcal{N}_{mn}^{cr}(\kappa_2)$, and $\mathcal{N}_{nm}^{bd}(\kappa_2,\kappa)$ covers the line segment $\delta_m = -1/2$ between $\mathcal{N}_{mn}^{cr}(\kappa_2)$ and $\mathcal{N}_{m+1,n}^{cr}(\kappa_2)$. In these sets either $\beta_{0\perp}'X_{mt}$ and $\beta_0'X_{nt}$ is nearly critical, but not both. See Figure 1 for illustrations in the case $k = 1$.

The following theorem proves (54) and (55) of Theorem 3. Here we use the notation $\ell_{T,p}(\psi) \Longrightarrow \infty$ on $\mathbb{C}(\mathcal{N}_{kn}^{cr}(\kappa_2))$, for example, as shorthand for the more precise statement in (54) on the space $\mathcal{N}_{kn}^{cr}(\kappa_2)$.

Theorem 4 *Let Assumptions 1–4 hold, so that in particular $E|\varepsilon_t|^8 < \infty$ and $d_0 \geq b_0$.*

(i) For $(\kappa_2, T) \to (0,\infty)$ it holds that

$$\ell_{T,p}(\psi) \Longrightarrow \ell_p(\psi) \text{ on } \mathbb{C}(\mathcal{N}_{mn}^{cr}(\kappa_2)) \text{ for } -1 \leq n < m \leq k-1, \tag{62}$$
$$\ell_{T,p}(\psi) \Longrightarrow \infty \text{ on } \mathbb{C}(\mathcal{N}_{kn}^{cr}(\kappa_2)) \text{ for } -1 \leq n < k. \tag{63}$$

(ii) For fixed κ_2 and $(\kappa, T) \to (0,\infty)$ it holds that

$$\ell_{T,p}(\psi) \Longrightarrow \ell_p(\psi) \text{ on } \mathbb{C}(\mathcal{N}_{mn}^{bd}(\kappa_2,\kappa)) \text{ for } n \leq m < k, \tag{64}$$
$$\ell_{T,p}(\psi) \Longrightarrow \infty \text{ on } \mathbb{C}(\mathcal{N}_{mn}^{bd}(\kappa_2,\kappa)) \text{ for } n \leq m = k, \tag{65}$$
$$\ell_{T,p}(\psi) \Longrightarrow \infty \text{ on } \mathbb{C}(\mathcal{N}_{k,k+1}^{bd}(\kappa_2,\kappa)), \tag{66}$$
$$\ell_{T,p}(\psi) \Longrightarrow \ell_p(\psi) \text{ on } \mathbb{C}(\mathcal{N}_{mn}^{bd}(\kappa_2,\kappa)) \text{ for } m < n \leq k, \tag{67}$$
$$\ell_{T,p}(\psi) \Longrightarrow \infty \text{ on } \mathbb{C}(\mathcal{N}_{m,k+1}^{bd}(\kappa_2,\kappa)) \text{ for } m < k. \tag{68}$$

(iii) For fixed κ and $T \to \infty$ it holds that

$$\ell_{T,p}(\psi) \Longrightarrow \ell_p(\psi) \text{ on } \mathbb{C}(\mathcal{N}_{mn}^{int}(\kappa)) \text{ for } -1 \leq n \leq m \leq k, \tag{69}$$
$$\ell_{T,p}(\psi) \Longrightarrow \infty \text{ on } \mathbb{C}(\mathcal{N}_{k+1,n}^{int}(\kappa)) \text{ for } -1 \leq n \leq k, \tag{70}$$
$$\ell_{T,p}(\psi) \Longrightarrow \infty \text{ on } \mathbb{C}(\mathcal{N}_{k+1,k+1}^{int}(\kappa)). \tag{71}$$

In the proof of Theorem 4, given in the next three subsections, we apply the following notation. We use u_1, v_1, or w_1 to indicate the index of $\beta_{0\perp}'X_{kt}$, depending on whether it is asymptotically stationary, critical, or nonstationary, and similarly we use u_2, v_2, or w_2 to indicate the index of $\beta_0'X_{kt}$. We collect all asymptotically stationary regressors in a vector with indices greater than or equal to u, and the nonstationary regressors are collected in a vector with indices smaller than or equal to w. Finally, we use the notation $\kappa_u, \kappa_w, \underline{\kappa}_v, \overline{\kappa}_v$ as in Definition 1 to describe the intervals for the indices in order to apply Corollary 1.

5.7 Proof of Theorem 4(i): the critical sets $\mathcal{N}_{mn}^{cr}(\kappa_2)$

We assume $\kappa_2 < \eta$ and let $(\kappa_2, T) \to (0, \infty)$. The set $\mathcal{N}_{mn}^{cr}(\kappa_2)$ is defined by the inequalities

$$|\phi_n + 1/2| \le \kappa_2 \text{ and } |\delta_m + 1/2| \le \kappa_2$$

for $-1 \le n < m \le k$. We note that $v_1 = \delta_m$ and $v_2 = \phi_n$ are the near critical indices and let $\underline{\kappa}_v = \overline{\kappa}_v = \kappa_2$, see (21). The nonstationary regressors are given by the indices $\phi_{-1}, \ldots, \phi_{n-1}, \delta_{-1}, \ldots, \delta_{m-1}$ and the maximal index is

$$w = \max(\phi_{n-1}, \delta_{m-1}) = \max(\phi_n, \delta_m) - b \le -1/2 + \kappa_2 - \eta.$$

The stationary regressors are defined by the indices $\phi_{n+1}, \ldots, \phi_{k-1}, \delta_{m+1}, \ldots, \delta_{k-1}$, and the minimal index is

$$u = \min(\phi_{n+1}, \delta_{m+1}) = \min(\phi_n, \delta_m) + b \ge -1/2 - \kappa_2 + \eta.$$

Thus we can take $\kappa_u = \kappa_w = \eta - \kappa_2$.

The notation (u, v, w) for the indices for the regressands $\beta_{0\perp}' X_{kt}$ and $\beta_0' X_{kt}$ differ depending on the values of m, n. We consider two cases: $n < m < k$ and $n < m = k$.

Proof of (62) on the sets $\mathcal{N}_{mn}^{cr}(\kappa_2)$ for $n < m < k$: When $m < k$ both $\beta_{0\perp}' X_{kt}$ and $\beta_0' X_{kt}$ are asymptotically stationary with indices $u_1 = \delta_k$ and $u_2 = \phi_k$, such that $SSR_T(\psi) = B_0 M_T((u_1, u_2), (u_1, u_2)|v_1, v_2, w, u) B_0'$. We decompose the matrix as follows

$$\begin{aligned}&M_T((u_1, u_2), (u_1, u_2)|w, u) \\ &- M_T((u_1, u_2), (v_1, v_2)|w, u) M_T((v_1, v_2), (v_1, v_2)|w, u)^{-1} M_T((v_1, v_2), (u_1, u_2)|w, u),\end{aligned} \tag{72}$$

and we apply Corollary 1. From (34) we see that, for $T \to \infty$,

$$\log\det(B_0 M_T((u_1, u_2), (u_1, u_2)|w, u) B_0') \Longrightarrow \log\det(\Omega(\psi)) = \ell_p(\psi), \tag{73}$$

see (50) and (51). We next show that the second term of (72) is $\mathbf{o}_P(1)$.

The critical processes are $\Delta_+^{d+mb} \beta_{0\perp}' X_t$ and $\Delta_+^{d+nb} \beta_0' X_t$ with indices v_1, v_2 and stochastic components

$$\Delta_+^{d+mb} \beta_{0\perp}' X_t : \Delta_+^{v_1}(\xi_1 \varepsilon_t + \Delta_+^{b_0} \beta_{0\perp}' Y_t^+), \tag{74}$$

$$\Delta_+^{d+nb} \beta_0' X_t : \Delta_+^{v_2} \beta_0' Y_t^+ = \Delta_+^{v_2}(\xi_2 \varepsilon_t + \Delta^{b_0} \beta_0' \sum_{n=0}^{\infty} \tau_{0n} \varepsilon_{t-n}), \tag{75}$$

see (37)–(40). These are fractional differences of processes in $\mathcal{Z}_{b_0}$ with leading coefficients $\xi_1 = \beta_{0\perp}' C_0$ and $\xi_2 = \beta_0' C_0^*$. Here, $[\xi_1', \xi_2']$ has full column rank because

$$[\alpha_{0\perp}, \alpha_0]'[\xi_1', \xi_2'] = \begin{bmatrix} \alpha_{0\perp}' C_0' \beta_{0\perp} & \alpha_{0\perp}' C_0^{*\prime} \beta_0 \\ 0 & \alpha_0' C_0^{*\prime} \beta_0 \end{bmatrix} = \begin{bmatrix} \alpha_{0\perp}' C_0' \beta_{0\perp} & \alpha_{0\perp}' C_0^{*\prime} \beta_0 \\ 0 & -I_r \end{bmatrix} \tag{76}$$

has full rank, see (39), (40), and Assumption 1. It follows from (35) and (36) that, for $(\kappa_2, T) \to (0, \infty)$,

$$\begin{aligned} M_T((v_1, v_2), (v_1, v_2)|w, u)^{-1} &= \mathbf{o}_P(1), \\ M_T((u_1, u_2), (v_1, v_2)|w, u) &= \mathbf{O}_P(1), \end{aligned}$$

and from (73) we find $\log\det(SSR_T(\psi)) \Longrightarrow \ell_p(\psi)$.

Proof of (63) on the sets $\mathcal{N}^{cr}_{kn}(\kappa_2)$ for $n < m = k$: In this case $\beta'_{0\perp}X_{kt}$ is near critical with index $v_1 = \delta_k \in [-1/2-\kappa_2, -1/2+\kappa_2]$, setting $\overline{\kappa}_v = \underline{\kappa}_v = \kappa_2$, and $\beta'_0 X_{kt}$ with index $u_2 = \phi_k \geq -1/2+\kappa_u$ is stationary. The other near critical process is the regressor $\beta'_0 X_{nt}$ with $v = \phi_n$, $n < k$. The determinant $\det(SSR_T(\psi))$ has, apart from the factor $(\det B_0)^2$, the form

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

The first factor satisfies $\det(M_T(u_2,u_2|v_1,w,u)) \Longrightarrow \det(E(M_T(u_2,u_2|u)) > 0$ as $(\kappa_2,T)\to(0,\infty)$, see Corollary 1.

The second factor is analyzed as follows. Let $M = M_T((v_1,v),(v_1,v)|w,u)$. For $\xi_1 = \beta'_{0\perp}C_0$ and $\xi_2 = \beta'_0 C^*_0$, the matrix $[\xi'_1,\xi'_2]$ has full rank, see (76), such that the two critical processes are given in (74) and (75). From (36) we have for $\underline{\kappa}_v = \overline{\kappa}_v = \kappa_2$ and $N = T^\alpha$, that uniformly in $(v_1,v,w,u)\in\mathcal{S}(\kappa_w,\kappa_u,\underline{\kappa}_v,\overline{\kappa}_v)$ it holds that

$$M^{-1} = \begin{bmatrix} M^{11} & M^{12} \\ M^{21} & M^{22}\end{bmatrix} \leq \left(c\frac{1-T^{-2\alpha\kappa_2}}{2\kappa_2}\begin{bmatrix}\xi_1\Omega_0\xi'_1 & \xi_1\Omega_0\xi'_2 \\ \xi_2\Omega_0\xi'_1 & \xi_2\Omega_0\xi'_2\end{bmatrix} + R_T\right)^{-1},$$

where $R_T = \mathbf{O}_P(1)$. Because $(1-T^{-2\alpha\kappa_2})/(2\kappa_2)\to\infty$ as $(\kappa_2,T)\to(0,\infty)$, it follows that the factor $M_T(v_1,v_1|u_2,v,w,u) = M_{11} - M_{12}M_{22}^{-1}M_{21} = M_{11.2}$ satisfies

$$M_T(v_1,v_1|u_2,v,w,u)^{-1} = M_{11.2}^{-1} = M^{11} \Longrightarrow 0 \text{ as } (\kappa_2,T)\to(0,\infty),$$

and therefore $\log\det(SSR_T(\psi)) \Longrightarrow \infty$ if $m = k$.

5.8 Proof of Theorem 4(ii): the boundary sets $\mathcal{N}^{bd}_{mn}(\kappa_2,\kappa)$ for $n \leq m$

We fix $\kappa_2 < \eta$, assume $\kappa < \eta - \kappa_2$, and let $(\kappa,T)\to(0,\infty)$. The set $N^{bd}_{mn}(\kappa_2,\kappa)$ with $-1\leq n\leq m\leq k$ is defined by the inequalities

$$|\delta_m + 1/2| \leq \kappa,\ \phi_{n-1}\leq -1/2-\kappa_2, \text{ and } -1/2+\kappa_2\leq\phi_n.$$

There is only one near critical process with index $v = \delta_m$ so we let $\overline{\kappa}_v = \underline{\kappa}_v = \kappa$. The stationary regressors are given by the indices $\delta_{m+1},\ldots,\delta_{k-1}$ and $\phi_n,\ldots,\phi_{k-1}$ with minimal value

$$u = \min(\phi_n,\delta_{m+1}) = \min(\phi_n,\delta_m+b) \geq \min(-1/2+\kappa_2,-1/2-\kappa+\eta) = -1/2+\kappa_2,$$

and the nonstationary processes given by indices $\delta_{-1},\ldots,\delta_{m-1},\phi_{-1},\ldots,\phi_{n-1}$ with a maximal value

$$w = \max(\delta_{m-1},\phi_{n-1}) \leq \max(\delta_m - b, -1/2-\kappa_2) \leq \max(-1/2+\kappa-\eta,-1/2+\kappa_2) = -1/2-\kappa_2,$$

so we let $\kappa_u = \kappa_w = \kappa_2$. We consider two cases defined by $n\leq m<k$ and $n\leq m=k$.

Proof of (64) on the sets $\mathcal{N}^{bd}_{mn}(\kappa_2,\kappa)$ for $n\leq m<k$: In this case $\beta'_{0\perp}X_{kt}$ is stationary with index $u_1 = \delta_k > \delta_m$ and $\beta'_0 X_{kt}$ is stationary with index $u_2 = \phi_k > \phi_n$. Then $\det(SSR_T(\psi))$ is, apart from the factor $(\det B_0)^2$, the determinant of a matrix of the form

$$\begin{aligned} M_T((u_1,u_2),(u_1,u_2)|v,w,u) &= M_T((u_1,u_2),(u_1,u_2)|w,u) \\ &\quad - M_T((u_1,u_2),v|w,u)M_T(v,v|w,u)^{-1}M_T(v,(u_1,u_2)|w,u) \end{aligned} \tag{77}$$

Here $\log\det(B_0' M_T((u_1,u_2),(u_1,u_2)|w,u)B_0) \Longrightarrow \ell_p(\psi)$ by (34); see also (73).

Next, $M_T((u_1,u_2),v|w,u) = \mathbf{O}_P(1)$ by (35) and $M_T(v,v|w,u)^{-1} \Longrightarrow 0$ by (36). Thus, the decomposition (77) shows that

$$\ell_{T,p}(\psi) = \log\det(SSR_T(\psi)) \Longrightarrow \ell_p(\psi).$$

Proof of (65) on the sets $\mathcal{N}^{bd}_{mn}(\kappa_2,\kappa)$ for $n \le m = k$: Here $\beta_{0\perp}' X_{kt}$ is near critical with index $v_1 = \delta_k$ and $\beta_0' X_{kt}$ is stationary with index $u_2 = \phi_k > \phi_n$, such that, apart from a factor $(\det B_0)^2$, $\det(SSR_T(\psi))$ is of the form

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

The first factor diverges due to (36) and the second satisfies

$$\det(M_T(u_2,u_2|v_1,w,u)) \Longrightarrow \det(E(M_T(u_2,u_2|u))) > 0$$

by Corollary 1. Thus, $\log\det(SSR_T(\psi)) \Longrightarrow \infty$.

5.9 Proof of Theorem 4(ii): the boundary sets $\mathcal{N}^{bd}_{mn}(\kappa_2,\kappa)$ for $m < n$

Again, we fix $\kappa_2 < \eta$, assume $\kappa < \eta - \kappa_2$, and let $(\kappa,T) \to (0,\infty)$. The set $N^{bd}_{mn}(\kappa_2,\kappa)$ with $-1 \le m < n \le k+1$ is given by

$$|\phi_m + 1/2| \le \kappa,\ \delta_{n-1} \le -1/2 - \kappa_2, \text{ and } -1/2 + \kappa_2 \le \delta_n.$$

There is one near critical process with index $v = \phi_m$ and we let $\underline{\kappa}_v = \overline{\kappa}_v = \kappa$. The nonstationary regressors have indices $\phi_{-1},\ldots,\phi_{m-1},\delta_{-1},\ldots,\delta_{n-1}$ with maximal index

$$w = \max(\phi_{m-1},\delta_{n-1}) = \max(\phi_m - b, -1/2-\kappa_2) \le \max(-1/2+\kappa-\eta, -1/2-\kappa_2) = -1/2-\kappa_2,$$

and the stationary processes have indices $\phi_{m+1},\ldots,\phi_{k-1},\delta_n,\ldots,\delta_{k-1}$ with minimal index

$$u = \min(\phi_{m+1},\delta_n) \ge \min(\phi_m + b, -1/2+\kappa_2) \ge \min(-1/2-\kappa+\eta, -1/2+\kappa_2) = -1/2+\kappa_2,$$

so $\kappa_u = \kappa_w = \kappa_2$. We consider the three cases: $(m,n) = (k,k+1)$, $m < n \le k$, and $n = k+1, m < k$.

Proof of (66) on the set $\mathcal{N}^{bd}_{k,k+1}(\kappa_2,\kappa)$: In this case $\beta_{0\perp}' X_{kt}$ is nonstationary with index $w_1 = \delta_k$ and $\beta_0' X_{kt}$ is near critical with index $v_2 = \phi_k$. There are no stationary processes. Then $\det(SSR_T(\psi))$ is, apart from the factor $(\det B_0)^2$,

$$\det(M_T((w_1,v_2),(w_1,v_2)|w)) = \det(M_T(v_2,v_2|w_1,w))\det(M_T(w_1,w_1|w)).$$

The first factor diverges to infinity as $(\kappa,T) \to (0,\infty)$ by (36). Because $2w_1 + 1 \le -2\kappa_2$, the second factor satisfies $\det(M_T(w_1,w_1|w)) \ge \det(T^{2\kappa_2} M_T^{**}(w_1,w_1|w))$, which diverges to infinity by (29) of Lemma 3.

Proof of (67) on the sets $\mathcal{N}^{bd}_{mn}(\kappa_2,\kappa)$ for $m < n \le k$: In this case both $\beta_{0\perp}' X_{kt}$ and $\beta_0' X_{kt}$ are stationary, and the proof is identical to that of (64), see (77).

Proof of (68) on the sets $\mathcal{N}^{bd}_{m,k+1}(\kappa_2,\kappa)$ for $m < k$: Here $\beta_{0\perp}' X_{kt}$ is nonstationary with index $w_1 = \delta_k$, while $\beta_0' X_{kt}$ is stationary with index $u_2 = \phi_k$. Therefore $SSR_T(\psi) = B_0 M_T((w_1,u_2),(w_1,u_2)|v,w,u)B_0'$ and

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

where we use the notation

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

The Woodbury matrix identity, see Magnus and Neudecker (1999, p. 11, eqn. (7)), then gives

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

We find from Lemmas 1 and 3 and Corollary 1 that

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

Because $2w + 1 \leq -2\kappa_2$, we then find that

$$A_{22.1}^{-1} = T^{2w+1}A_{22.1}^{**-1} = \mathbf{O}_P(T^{-2\kappa_2}(1 + \frac{2\kappa}{1 - N^{-2\kappa}}(1 + \log T)^4 T^{2\kappa})) = \mathbf{o}_P(1) \text{ as } (\kappa, T) \to (0, \infty).$$

This implies that $\det(SSR_T(\psi)^{-1}) \Longrightarrow 0$, and hence $\ell_{T,p}(\psi) = \log\det(SSR_T(\psi)) \Longrightarrow \infty$ as $(\kappa, T) \to (0, \infty)$.

5.10 Proof of Theorem 4(iii): the interior sets $\mathcal{N}_{mn}^{int}(\kappa)$

We now fix κ and κ_2 and let $T \to \infty$. The set $N_{mn}^{int}(\kappa)$ is defined by

$$\max(\delta_{m-1}, \phi_{n-1}) \leq -1/2 - \kappa \text{ and } -1/2 + \kappa \leq \min(\delta_m, \phi_n).$$

The nonstationary regressors (indices $\delta_{-1}, \ldots, \delta_{m-1}, \phi_{-1}, \ldots, \phi_{n-1}$) have maximal index $w = \max(\delta_{m-1}, \phi_{n-1}) \leq -1/2-\kappa$, and the stationary regressors (indices $\delta_m, \ldots, \delta_{k-1}, \phi_n, \ldots, \phi_{k-1}$) have smallest index $u = \min(\delta_m, \phi_n) \geq -1/2 + \kappa$. Thus, $\kappa_u = \kappa_w = \kappa$ and there are no near critical processes.

We consider three cases: $n \leq m \leq k$, $n < m = k + 1$, and $n = m = k + 1$.

Proof of (69) on the sets $\mathcal{N}_{mn}^{int}(\kappa)$ for $n \leq m \leq k$: In this case $\beta_{0\perp}' X_{kt}$ and $\beta_0' X_{kt}$ are both stationary with indices $u_1 = \delta_k$ and $u_2 = \phi_k$. Therefore

$$\det(SSR_T(\psi)) = \det(B_0 M_T((u_1, u_2), (u_1, u_2)|w, u)B_0') \Longrightarrow \ell_p(\psi)$$

as $T \to \infty$, see (34).

Proof of (70) on the sets $\mathcal{N}_{k+1,n}^{int}(\kappa)$ for $n < k + 1$: Here $\beta_{0\perp}' X_{kt}$ is nonstationary with index $w_1 = \delta_k$ and $\beta_0' X_{kt}$ is stationary with index $u_2 = \phi_k$. Then, apart from $(\det B_0)^2$, $\det(SSR_T(\psi))$ is given by

$$\det(M_T((w_1, u_2), (w_1, u_2)|w, u)) = \det(M_T(u_2, u_2)|w, u)) \det(M_T((w_1, w_1)|w, u_2)).$$

The first factor converges to $\ell_p(\psi)$, see (34), and the last one diverges due to lack of normalization, see (29) of Lemma 3. Thus, $\log\det(SSR_T(\psi)) \Longrightarrow \infty$ as $T \to \infty$.

Proof of (71) on the set $\mathcal{N}_{k+1,k+1}^{int}(\kappa)$: In this case all variables are nonstationary, so that

$$\log\det(SSR_T(\psi)) = \log\det(B_0 M_T((w_1, w_2), (w_1, w_2)|w)B_0') \Longrightarrow \infty$$

as $T \to \infty$, due to lack of normalization, see Lemma 3.

References

1. Barunik, J., and Dvorakova, S. (2015). An empirical model of fractionally cointegrated daily high and low stock market prices. *Economic Modelling* **45**, 193–206.
2. Caporin, M., Ranaldo, A., and Santucci de Magistris, P. (2013). On the predictability of stock prices: a case for high and low prices. *Journal of Banking & Finance* **37**, 5132–5146.
3. Carlini, F., and Santucci de Magistris, P. (2017). On the identification of fractionally cointegrated VAR models with the $\mathcal{F}(d)$ condition. *Journal of Business & Economic Statistics*, forthcoming. DOI: 10.1080/07350015.2017.1294077.
4. Dolatabadi, S., Nielsen, M.Ø., and Xu, K. (2016). A fractionally cointegrated VAR model with deterministic trends and application to commodity futures markets. *Journal of Empirical Finance* **38B**, 623–639.
5. Horn, R.A., and Johnson, C.R. (2013). *Matrix Analysis*, second ed., Cambridge University Press, Cambridge.
6. Hualde, J., and Robinson, P.M. (2011). Gaussian pseudo-maximum likelihood estimation of fractional time series models. *Annals of Statistics* **39**, 3152–3181.
7. Jensen, A.N., and Nielsen, M.Ø. (2014). A fast fractional difference algorithm. *Journal of Time Series Analysis* **35**, 428–436.
8. Johansen, S. (1996). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, 2nd ed., Oxford University Press, Oxford.
9. Johansen, S. (2008). A representation theory for a class of vector autoregressive models for fractional processes. *Econometric Theory* **24**, 651–676.
10. Johansen, S., and Nielsen, M.Ø. (2010). Likelihood inference for a nonstationary fractional autoregressive model. *Journal of Econometrics* **158**, 51–66.
11. Johansen, S., and Nielsen, M.Ø. (2012a). Likelihood inference for a fractionally cointegrated vector autoregressive model. *Econometrica* **80**, 2667–2732.
12. Johansen, S., and Nielsen, M.Ø. (2012b). A necessary moment condition for the fractional functional central limit theorem. *Econometric Theory* **28**, 671–679.
13. Johansen, S., and Nielsen, M.Ø. (2016). The role of initial values in conditional sum-of-squares estimation of nonstationary fractional time series models. *Econometric Theory* **32**, 1095–1139.
14. Johansen, S., and Nielsen, M.Ø. (2018). Testing the CVAR in the fractional CVAR model. QED working paper 1394, Queen's University. *Journal of Time Series Analysis*, forthcoming. DOI: 10.1111/jtsa.12300.
15. Magnus, J.R., and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, revised ed., John Wiley and Sons, New York.
16. Nielsen, M.Ø., and Popiel, M.K. (2016). A Matlab program and user's guide for the fractionally cointegrated VAR model, QED working paper 1330, Queen's University.
17. Rutherford, D.E. (1948). Some continuant determinants arising in physics and chemistry. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics* **62**, 229–236.
18. Tanaka, K. (1996). *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*, John Wiley and Sons, New York.

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