Queen's Economics Department Working Paper No. 1394

Testing the CVAR in the fractional CVAR 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

10-2017

Testing the CVAR in the fractional CVAR model*

Søren Johansen†
University of Copenhagen
and CREATES

Morten Ørregaard Nielsen‡
Queen's University
and CREATES

February 27, 2018

Abstract

We consider the fractional cointegrated vector autoregressive (CVAR) model of Johansen and Nielsen (2012a) and show that the likelihood ratio test statistic for the usual CVAR model is asymptotically chi-squared distributed. Because the usual CVAR model lies on the boundary of the parameter space for the fractional CVAR in Johansen and Nielsen (2012a), the analysis requires the study of the fractional CVAR model on a slightly larger parameter space so that the CVAR model lies in the interior. This in turn implies some further analysis of the asymptotic properties of the fractional CVAR model.

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

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

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.

*We are grateful to two referees for useful comments and suggestions, and 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. Parts of this paper were written while Johansen visited the Economics Department at Queen's University as a John Weatherall Distinguished Fellow. Their hospitality is gratefully acknowledged.

†Email: Soren.Johansen@econ.ku.dk

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

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.

When $d = b = 1$ in (1) the standard, non-fractional CVAR model, see Johansen (1996), is obtained as a very important special case. Given the importance of this model, it would be desirable to test the restriction $d = b = 1$ within the more general model (1), and, indeed, this test can be calculated straightforwardly using the software package of Nielsen and Popiel (2016). However, the asymptotic theory provided for model (1) in JN(2012a) is derived under the assumption that the parameter space is $\eta \leq b \leq d \leq d_1$ for some (arbitrarily small) $\eta > 0$ and some (arbitrarily large) $d_1 > 0$. Under this assumption, the standard CVAR model with $d = b = 1$ lies on the boundary of the parameter space, see Figure 1, and hence it does not follow under the assumptions in JN(2012a) that the test statistic for the standard model against the fractional model is asymptotically χ^2-distributed, see, e.g., Andrews (2001).

In this paper we show that it is possible to prove the main theorems in JN(2012a) for a larger parameter space, where, in particular, the line $d = b$ is no longer on the boundary. Hence, assuming $\eta < 1 < d_1$, the point $d = b = 1$ will be in the interior. The important implication is that the test statistic for the non-fractional model against the fractional model is asymptotically $\chi^2(2)$-distributed under our assumptions.

These new results allow, at least, two important applications. First, they allow testing the usual CVAR model against a model with more general fractional integration dynamics, as part of the model specification step in empirical analysis. This test has been calculated in empirical work, where it has been conjectured to be asymptotically χ^2-distributed as we verify in this paper. Second, it seems common to apply the model (1) with the restriction $d = b$ imposed, without testing this restriction against the unrestricted model. For examples of both these types of applications see, among others, Bollerslev, Osterrieder, Sizova, and Tauchen (2013), Dolatabadi, Nielsen, and Xu (2016), and Chen, Chiang, and Härdle (2016).

More generally, testing the usual CVAR model against the fractional CVAR model can be viewed as a model specification test for the CVAR model against a fractional alternative.

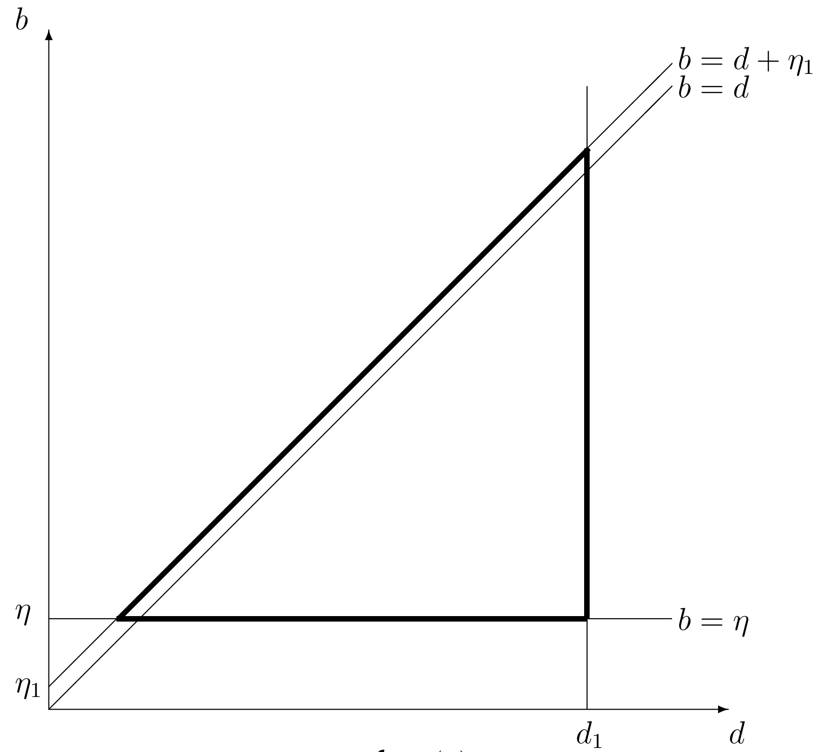

Figure 1: The parameter space $\mathcal{N}$ in (3) is the set bounded by bold lines.

There exists a large literature on testing univariate ARMA models against a fractional alternative, e.g., Robinson (1991), Agaiklogiou and Newbold (1994), Tanaka (1999), and Dolado, Gonzalo, and Mayoral (2002). Thus, the present paper contributes also to this literature by analyzing the test of the multivariate CVAR model against a fractional alternative.

The remainder of the paper is laid out as follows. In the next section we give the assumptions and the results. These results rely on an improved version of Lemma A.8 of JN(2012a), which is given in Section 3. Some implications of the results are discussed in Section 4.

2 Results and Methodology

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 $0 < \eta \leq 1/2$, which can be arbitrarily small (although a smaller η implies a stronger moment condition, see Theorem 1 below). We will instead consider the parameter space for d and b given by

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

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 $\eta_1 > 0$. The parameter space and the role of the constant $\eta_1 > 0$ is illustrated in Figure 1.

We will assume that the data for $t = 1, \ldots, T$ is generated by the model (1). Our approach is to conduct inference using the conditional likelihood function of $X_1, \ldots, X_T$ given initial values $\{X_{-n}\}_{n\geq 0}$. This approach is standard in (finite-order) autoregressive models, where conditional maximum likelihood leads to least squares estimation; see, e.g., Anderson (1971, pp. 183–184) or Hamilton (1994, pp. 122–123). That is, we interpret equation (1) as a model for $X_t, t = 1, \ldots, T$, given the past, which will allow us to apply the conditional density to build a conditional likelihood function; see (7) below. Thus, our entire approach is conditional on the initial values $\{X_{-n}\}_{n\geq 0}$, which are therefore considered 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 and 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 (an approximation to) 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 DGP; see Assumptions 3 and 5 and Footnote 1. Note that our Δ_0 corresponds to Δ_+ in, e.g., JN(2012a), and 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 conditional Gaussian log-likelihood function of $\{X_t\}_{t=1}^{T}$, given 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, 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, for given values of (d, b), 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).

Before we impose some further assumptions on the data generating process, 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 equation (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 $(d_0, b_0) \in \mathcal{N}$, $0 \leq d_0 - b_0 < 1/2$, $b_0 \neq 1/2$, 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 $\det\{\alpha_{0\perp}'\Gamma_0\beta_{0\perp}\} \neq 0$. If $r < p$, $\det\{\Psi(y)\} = 0$ 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.*

The conditions in Assumptions 1–4 are identical to those in JN(2012a). 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 some of our results. The conditions in Assumption 4 imply that the cointegrating relations $\beta_0' X_t$ are (asymptotically) stationary because $d_0 - b_0 < 1/2$, allow the important special case of $d_0 = b_0 (= 1)$, and also 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.

We are now ready to state our main results in the following two theorems. Both theorems require some strengthening of the assumptions, and these are identical to those in JN(2012a) and will be discussed subsequently. Note that the condition $q > (b_0 - 1/2)^{-1}$ when $b_0 > 1/2$ was used to prove (31) in JN(2012a), but was apparently overlooked in the statement of the consistency results in Theorems 4 and 5 in JN(2012a).

Theorem 1 *Let Assumptions 1–4 hold and assume, in addition, that $E|\varepsilon_t|^q < \infty$ for some $q > 1/\min\{\eta/3, (1/2 - d_0 + b_0)/2\}$, where $0 < \eta \leq 1/2$. If $b_0 > 1/2$ then assume also that $E|\varepsilon_t|^q < \infty$ for $q > (b_0 - 1/2)^{-1}$. Let the parameter space $\mathcal{N}(\eta, \eta_1, d_1)$ be given in (3), where η_1 is chosen such that $0 < \eta_1 < \min\{\eta/3, (1/2 - d_0 + b_0)/2\}$. 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.*

Proof of Theorem 1. Theorem 4 of JN(2012a) gives, for the smaller parameter space, the properties of the likelihood function which are needed to prove that the maximum likelihood estimator exists and is unique. If these results can be established under our assumptions (which are identical to those in JN, 2012a), but with the larger parameter space given in (3), then the proof of Theorem 5(i) of JN(2012a), showing that the maximum likelihood estimator exists uniquely with large probability for large T, can be used without any changes to prove our Theorem 1. The proof of Theorem 4 in JN(2012a) is given in their Appendix B. To avoid repeating their very lengthy proof, we only detail the differences.

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

$$X_t = C_0 \Delta_+^{-d_0} \varepsilon_t + \Delta_+^{-(d_0 - b_0)} Y_t^+ + \mu_{0t} \text{ for } d_0 \geq 1/2, \tag{10}$$

$$X_t = C_0 \Delta^{-d_0} \varepsilon_t + \Delta^{-(d_0 - b_0)} Y_t \text{ for } d_0 < 1/2. \tag{11}$$

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

$$C_0 = \beta_{0\perp}(\alpha_{0\perp}'\Gamma_0\beta_{0\perp})^{-1}\alpha_{0\perp}'. \tag{12}$$

Letting $\mathbb{I}\{\cdot\}$ denote the indicator function and using the definitions $\Psi_+(L)X_t = \mathbb{I}\{t \geq 1\}\sum_{i=0}^{t-1} \Psi_i X_{t-i}$ and $\Psi_-(L)X_t = \sum_{i=t}^{\infty} \Psi_i X_{t-i}$, the term μ_{0t} is

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

which expresses μ_{0t} as function of initial values. It is seen from (6) and (10), 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 of the form $\Delta_+^{d+ib}\mu_{0t}$, which are generated by the initial values, are considered deterministic in the analysis of the model.

We first analyze the deterministic terms and establish that these are uniformly small. In JN(2012a), this follows from their Lemma A.8. However, with our larger parameter space, this requires a new proof, and thus we give an improved version of Lemma A.8 of JN(2012a) in Lemma 1 in Section 3. It follows from Lemma 1(i) 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, so in the further analysis of the likelihood, we assume they are zero.

Then we analyze the stochastic terms in the likelihood function, which are expressed in terms of the stationary processes $C_0\varepsilon_t + \Delta^{b_0}Y_t$ and its differences $\Delta_+^{d+ib-d_0}(C_0\varepsilon_t + \Delta^{b_0}Y_t)$. The behavior of the stochastic terms depend on d and b, and therefore on the parameter space. More specifically, the stochastic terms are decomposed as $\beta_{0\perp}'\Delta_+^{d+ib-d_0}(C_0\varepsilon_t + \Delta^{b_0}Y_t)$ and $\beta_0'\Delta_+^{d+jb-d_0}(C_0\varepsilon_t + \Delta^{b_0}Y_t) = \beta_0'\Delta_+^{d+jb-d_0+b_0}Y_t$ for $i, j = -1, \ldots, k$. The former processes are $I(d_0 - d - ib)$, which can be either nonstationary, (asymptotically) stationary, or near critical in the sense that $d_0 - d - ib$ is close to $1/2$. On the other hand, the latter processes are (asymptotically) stationary for all $j \geq -1$ because $\beta_0'\Delta_+^{d+jb-d_0+b_0}Y_t \in I(d_0 - b_0 - d - jb)$ and $d_0 - b_0 - d - jb \leq d_0 - b_0 - d + b \leq d_0 - b_0 + \eta_1 < 1/2$ by choice of η_1. Thus, we have the same classification of processes into nonstationary, stationary, and near critical processes as in Appendix B.3 of JN(2012a).

Close to the critical value $d_0 - d - ib = 1/2$, the process $\beta_{0\perp}'\Delta_N^{d+ib}X_t$ is difficult to analyze because in a neighbourhood of this value it can be both stationary and nonstationary. The proof therefore considers a small neighborhood of the near critical processes of the form

$$-\kappa_1 \leq d_0 - d - ib - 1/2 \leq \kappa,$$

where the constant κ_1, in particular, plays an important role. The behaviour of the product moments of the stationary, the nonstationary, and the near critical processes is analyzed in Appendix B.3 of JN(2012a) in their Lemma A.9 and its corollaries. Those results can be applied in the present setting without change.

In the application of Lemma 1(i) and the results in Appendix B.3 of JN(2012a) dealing with the stochastic terms, we need to choose the constant κ_1 carefully. Specifically, on p. 2728 in Appendix B.3 of JN(2012a), κ_1 needs to be chosen such that $q^{-1} < \kappa_1 < \min\{\eta/3, (1/2 - d_0 + b_0)/2\}$ (in Appendix B.3 of JN(2012a) it is also required that $\kappa_1 < 1/6$, but this condition is redundant because we assume $\eta \leq 1/2$), while in the application of Lemma 1(i) we need to choose κ_1 such that $0 < \eta_1 < \kappa_1 < 1/4$. Choosing κ_1 to satisfy all these restrictions is possible because $q > 1/\min\{\eta/3, (1/2 - d_0 + b_0)/2\}$ and $\eta_1 < \min\{\eta/3, (1/2 - d_0 + b_0)/2\} < 1/4$. ■

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. Note that 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; see Lemma 1(ii). For example, Assumption 5(i) would be satisfied if $|X_{-n}| \leq cn^{-1/2-\epsilon}$ 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):

$$\underbrace{\ldots, X_{-M_0}}_{\text{Data does not exist}}, \quad \underbrace{X_{1-M_0}, \ldots, X_{-N}}_{\substack{\text{Data exists} \\ \text{but is not observed}}}, \quad \underbrace{X_{1-N}, \ldots, X_0}_{\substack{\text{Data is observed} \\ \text{(initial values)}}}, \quad \underbrace{X_1, \ldots, X_T}_{\substack{\text{Data is observed} \\ \text{(estimation)}}} \tag{13}$$

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* $0 < \eta_1 < \min\{\eta/3, (1/2 - d_0 + b_0)/2\}$. *Assume, in addition, that* $E|\varepsilon_t|^q < \infty$ *for some* $q > 1/\min\{\eta/3, (1/2 - d_0 + b_0)/2\}$. *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.*

[1]The proof of Lemma 1(ii), and especially equation (26), makes it clear that M_0 could be allowed to diverge. That is, we could allow $M_0 \to \infty$ as $T \to \infty$ as long as $M_0/\sqrt{T} \to 0$, as in Section 4.2 of JN(2012a). Allowing M_0 to diverge with T illustrates that if we use a large value of T for likelihood inference, then we can allow for more non-zero initial values, and still hope to make reliable inference on the parameters of interest. However, to avoid further notational complexity we do not consider this possibility in the remainder.

Table 1: Summary of moment conditions

Statement	Conclusion	Assumption on DGP	Assumption on q
Assns 2,4	assumed throughout	$0 \leq d_0 - b_0 < \frac{1}{2}, b_0 \neq \frac{1}{2}$	$q \geq 8$
Thm. 1	consistency	$b_0 < \frac{1}{2}$	$q > 1/\min\{\frac{\eta}{3}, \frac{1/2-d_0+b_0}{2}\}$
Thm. 1	consistency	$b_0 > \frac{1}{2}$	$q > 1/\min\{\frac{\eta}{3}, \frac{1/2-d_0+b_0}{2}, b_0 - \frac{1}{2}\}$
Thm. 2(i)	distn. params.	$b_0 < \frac{1}{2}$	$q > 1/\min\{\frac{\eta}{3}, \frac{1/2-d_0+b_0}{2}\}$
Thm. 2(ii)	distn. params.	$b_0 > \frac{1}{2}$	$q > 1/\min\{\frac{\eta}{3}, \frac{1/2-d_0+b_0}{2}, b_0 - \frac{1}{2}\}$
Cor. 1(i)	distn. $LR(d = b)$	$d_0 = b_0 < \frac{1}{2}$	$q > \frac{3}{\eta}$
	distn. $LR(d = b)$	$d_0 = b_0 > \frac{1}{2}$	$q > 1/\min\{\frac{\eta}{3}, b_0 - \frac{1}{2}\}$
Cor. 1(ii)	distn. $LR(d = b = 1)$	$d_0 = b_0 = 1$	$q > \frac{3}{\eta}$

Note: This table provides a summary of the different moment conditions and where they are applied.

Proof of Theorem 2. This follows from parts (i) and (ii) of Theorem 10 in JN(2012a). First, Lemma 1 shows that under Assumption 5, the deterministic terms generated by the initial values do not influence the asymptotic behavior of the score function and Hessian matrix; see also JN(2012a, p. 2694). Next, the proof of Theorem 10 in JN(2012a) relies on the usual Taylor expansion of the score function around the true values, and since we have made no changes to the assumptions on the data generating process or the true values, this proof applies to the current setting as well without any changes. Note that the moment condition $q > (b_0 - 1/2)^{-1}$ in part (ii) 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 $\hat{\beta}$. ■

The important implication of Theorem 2 is stated in the following corollary, where $LR(d = b)$ and $LR(d = b = 1)$ denotes the likelihood ratio test statistics for the hypotheses $H_{01} : d_0 = b_0$ and $H_{02} : d_0 = b_0 = 1$, respectively.

Corollary 1 *Let Assumptions 1–5 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 $0 < \eta_1 < \eta/3 \leq 1/6$. Assume, in addition, that $E|\varepsilon_t|^q < \infty$ for some $q > 3/\eta$. Then:*

(i) Let the null hypothesis $H_{01} : d_0 = b_0$ be true, and if $b_0 > 1/2$ assume also that $E|\varepsilon_t|^q < \infty$ for some $q > 1/\min\{\eta/3, b_0 - 1/2\}$. Then it holds that $LR(d = b) \overset{D}{\to} \chi^2(1)$.

(ii) Under the null hypothesis $H_{02} : d_0 = b_0 = 1$ it holds that $LR(d = b = 1) \overset{D}{\to} \chi^2(2)$.

Proof of Corollary 1. The corollary follows straightforwardly from Theorem 2 because $d_0 = b_0$ satisfies $(d_0, b_0) \in \text{int}(\mathcal{N})$ under (3). The conditions $q > 2/(1/2 - d_0 + b_0)$ and $\eta_1 < (1/2 - d_0 + b_0)/2$ from Theorem 2 are redundant when $d_0 = b_0$ because then $2/(1/2 - d_0 + b_0) = 4 < 3/\eta$ and $(1/2 - d_0 + b_0)/2 = 1/4 > \eta/3$ since $\eta/3 \leq 1/6$. ■

We note from the statements of Theorems 1 and 2, and in particular from their proofs, that the moment conditions and the conditions on the parameter space, i.e. on the user-chosen constants η and η_1, are closely linked. The different moment conditions that we apply are summarized in Table 1. Under the conditions of the hypotheses in Corollary 1, these simplify substantially. Specifically, under the null hypothesis $H_{02} : d_0 = b_0 = 1$ in

Corollary 1(ii), only $q > 3/\eta$ moments are required, in addition to $q \geq 8$ from Assumption 2, because all other moment restrictions from Theorems 1 and 2 are redundant when $d_0 = b_0 = 1$. For example, if η is chosen as $\eta > 3/8$ (i.e., in particular if consideration is restricted to the case of so-called "strong cointegration", where $b_0 > 1/2$), then the results in Corollary 1 follow under only the moment condition $q \geq 8$ in Assumption 2.

3 Improving Lemma A.8 of JN(2012a)

In this section we state and prove Lemma 1, which gives results for the impact of deterministic terms generated by initial values. It improves Lemma A.8(i) of JN(2012a) to accommodate the larger parameter space given in (3). Part (i) of Lemma 1 is used to prove convergence of the likelihood function, and hence in the consistency result in Theorem 1, while part (ii) is used in the analysis of the score function, and hence in the asymptotic distribution result in Theorem 2. Part (i) of the lemma requires a new proof and particularly requires careful choice of the constants η_1 in the parameter space (3) and κ_1 used in the lemma.

To state the lemma, we define the operators Δ_+ and Δ_- such that, for any a, $\Delta_+^a X_t = \Delta_0^a X_t$ and, for $a \geq 0$, $\Delta^a X_t = \Delta_+^a X_t + \Delta_-^a X_t$. We also define $\tilde{X}_t = X_t \mathbb{I}\{1 - N \leq t \leq 0\}$ as the initial values used in the calculations. Note that, for any a, $\Delta_-^a \tilde{X}_t = (\Delta_N^a - \Delta_0^a) X_t$, and that $\Delta_-^a X_t|_{a=0} = X_0 \mathbb{I}\{t = 0\}$ such that $\Delta_+^a \Delta_-^0 X_t = 0$.

When $d_0 \geq 1/2$, the deterministic terms in the likelihood function can be written as functions of

$$D_{it}(d,b) = \begin{cases} (\Delta_-^{d-b} - \Delta_-^{d})\tilde{X}_t + (\Delta_+^{d-b} - \Delta_+^{d})\mu_{0t}, & i = -1, \\ (\Delta_-^{d+ib} - \Delta_-^{d+kb})\tilde{X}_t + (\Delta_+^{d+ib} - \Delta_+^{d+kb})\mu_{0t}, & i = 0, \dots, k-1, \\ \Delta_-^{d+kb}\tilde{X}_t + \Delta_+^{d+kb}\mu_{0t}, & i = k. \end{cases} \tag{14}$$

where, see equations (8) and (97) in JN(2012a),

$$\mu_{0t} = F_+(L)\alpha_0\beta_0'\Delta_+^{-d_0+b_0}\Delta_-^{d_0-b_0}X_t - \sum_{j=0}^{k}(C_0\Psi_{0j}\Delta_+^{-d_0} + F_+(L)\Psi_{0j}\Delta_+^{-d_0+b_0})\Delta_-^{d_0+jb_0}X_t. \tag{15}$$

Here, C_0 is given in (12) and Ψ_{0j} are the coefficients in the polynomial $\Psi(y)$ in (8), both evaluated at the true values, and $F_+(L)Z_t = \sum_{n=0}^{t-1}\tau_{0n}Z_{t-n}$ where the coefficients τ_{0n} are given in (10) and (11). Note that μ_{0t} depends only on the true values of the parameters and on the initial values of X_t. When $d_0 < 1/2$, we use a different representation of the solution and hence leave out the terms involving $\Delta_+^{d+ib}\mu_{0t}$ in (14), see Theorem 2 in JN(2012a).

The terms $D_{it}(d,b)$ are functions of the variables $d+ib$, d, and $d+kb$. The m'th derivative of $D_{it}(d,b)$ with respect to d, say, is denoted $\mathsf{D}_d^m D_{it}(d,b)$ and the generic notation $\mathsf{D}^m D_{it}(d,b)$ is used for any m'th derivative involving d and/or b.

For the analysis of the deterministic terms in the score function we define the two operators, see Johansen (2008),

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

for which $\Pi(L)X_t = \Pi_+(L)X_t + \Pi_-(L)X_t$, see (9). Then the score function contains the deterministic terms

$$d_{0t} = \Pi_{0-}(L)(\tilde{X}_t - X_t) \text{ and } d_{1t} = \mathsf{D}\Pi_{0+}(L)\mu_{0t} + \mathsf{D}\Pi_{0-}(L)\tilde{X}_t, \tag{16}$$

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_{1t} is found as a linear combination of $\mathsf{D}D_{it}(\psi)|_{\psi=\psi_0}$, see (14), and also $T^{1/2-b_0}\beta_{0\perp}'\mathsf{D}D_{-1,t}(\psi)|_{\psi=\psi_0}$ if $b_0 > 1/2$, and hence is analyzed in part (i) of Lemma 1.

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*

$$\sup_{d+ib \in S_+} |\mathsf{D}^m D_{it}(d, b)| \to 0 \text{ as } t \to \infty, \tag{17}$$

$$\sup_{d+ib \in S_-} \max_{1 \le t \le T} |\mathsf{D}^m T^{d+ib-d_0+1/2} \beta_{0\perp}' D_{it}(d, b)| \to 0 \text{ as } T \to \infty. \tag{18}$$

(ii) Under Assumption 5 and $d_0 \ge b_0$ it holds that $T^{-1/2} \sum_{t=1}^T |d_{0t}| \to 0$.

Proof of Lemma 1(i). The following evaluations are taken from Lemmas B.3 and C.1 of JN(2010). For $|u| \le u_0$, $0 < v_0 \le v \le v_1$, $m \ge 0$, and $t \ge 1$, it holds that

$$|\mathsf{D}_u^m \pi_t(u)| \le c(1 + \log t)^m t^{u-1}, \tag{19}$$

$$|\mathsf{D}_u^m T^u \Delta_+^u \Delta_-^v X_t| \le c(1 + \log T)^{m+1} T^{\max\{-v, -1, u-v, u\}}, \tag{20}$$

where the constant c does not depend on u, v, m, t, or T. Our new Lemma 2 below shows that, for $u + v + 1 \ge a_1 > 0$ and $v \ge a_2 > 0$, it holds that

$$|\mathsf{D}_u^m \Delta_+^u \Delta_-^v X_t| \le c(1 + \log t)^{m+1} t^{\max\{-u-1, -v, -2u-v-1\}}, \tag{21}$$

where the constant c does not depend on u, v, m, or t. In each of the evaluations (19)–(21), the m'th derivative with respect to u gives rise to an extra logarithmic factor, which does not influence the convergences in (17) and (18), so in the following we assume $m = 0$.

Proof of (17) and (18) for the terms involving $\tilde{X}_t$: In both expressions, the initial values $\tilde{X}_t$ appear in terms of the form $\Delta_-^w \tilde{X}_t$ for $w = d + ib \ge -\eta_1$. We apply (19) to obtain the bound

$$|\Delta_-^w \tilde{X}_t| = |\sum_{j=0}^{N-1} \pi_{j+t}(-w) X_{-j}| \le c \sum_{j=0}^{N-1} (j + t)^{-w-1} \le cNt^{-w-1} \le cNt^{\eta_1 - 1}, \tag{22}$$

which tends to zero as $t \to \infty$, and hence proves (17).

To prove (18) we first note that if $d_0 \le 1/2$, then $S_- = \varnothing$ because $0 < \eta_1 < \kappa_1$, and consequently there is nothing to prove if $d_0 \le 1/2$. Thus, we prove the result for $d_0 > 1/2$. It follows from (22) that

$$\sup_{w \ge -\eta_1} \max_{1 \le t \le T} |\Delta_-^w \tilde{X}_t| \le c,$$

and for $w \in S_-$ it holds that

$$\sup_{w \le d_0 - 1/2 - \kappa_1} T^{w - d_0 + 1/2} \le T^{-\kappa_1} \to 0,$$

which proves (18).

Proof of (17) for the terms involving μ_{0t}: These terms are only present if $d_0 \geq 1/2$, which we therefore assume in the remainder of the proof. There are three types of terms; namely $\sup_{d+ib\in S_+}|\Delta_+^{d+ib}\mu_{0t}|$ for $i=-1,\ldots,k$, which are all equal, $\sup_{d+ib\in S_+}|\Delta_+^{d+kb}\mu_{0t}|$ for $i=0,\ldots,k$, which are dominated by $\sup_{d+kb\in S_+}|\Delta_+^{d+kb}\mu_{0t}|$, and $\sup_{d-b\in S_+}|\Delta_+^{d}\mu_{0t}|$, which is dominated by $\sup_{d\in S_+}|\Delta_+^{d}\mu_{0t}|$. Thus, for $w\in S_+$, we need only consider $\Delta_+^{w}\mu_{0t}$ given by

$$F_+(L)\alpha_0\beta_0'\Delta_+^{w-d_0+b_0}\Delta_-^{d_0-b_0}X_t-\sum_{j=0}^{k}(C_0\Psi_{0j}\Delta_+^{w-d_0}+F_+(L)\Psi_{0j}\Delta_+^{w-d_0+b_0})\Delta_-^{d_0+jb_0}X_t, \quad (23)$$

see (15). We note that the terms in (23) are of the form $\sum_{i=0}^{t-1}A_{ni}R_{n,t-i}(w)$, $n=1,2,3$, for suitable matrices A_{ni}, which satisfy $\sum_{i=0}^{\infty}|A_{ni}|<\infty$, and

$$R_{1t}(w)=\Delta_+^{w-d_0+b_0}\Delta_-^{d_0-b_0}X_t, R_{2t}(w)=\Delta_+^{w-d_0}\Delta_-^{d_0+jb_0}X_t, R_{3t}(w)=\Delta_+^{w-d_0+b_0}\Delta_-^{d_0+jb_0}X_t.$$

Thus, we show that $\sup_{w\in S_+}|R_{n,t-i}(w)|\to 0$ as $t\to\infty$ for i fixed and $n=1,2,3$, such that, in particular, $\sup_{w\in S_+}|R_{n,t-i}(w)|\leq c_n$ for some c_n that does not depend on i. For each n, the coefficients $|A_{ni}|, i=0,\ldots,$ are summable in i, and therefore the sequence $|A_{ni}|\sup_{w\in S_+}|R_{n,t-i}(w)|\mathbb{I}\{i\leq t\}$ is dominated by the summable series $c_n|A_{ni}|$. It then follows from the Dominated Convergence Theorem that $\sup_{w\in S_+}|\sum_{i=0}^{t-1}A_{ni}R_{n,t-i}(w)|\to 0$ for $n=1,2,3$, which proves (17) in view of (23).

We apply (21) for each $R_{nt}(w)$, giving the proofs for $i=0$ to simplify notation. For $R_{1t}(w)$ we first note that $\Delta_-^{0}X_t=X_0\mathbb{I}\{t=0\}$, such that $\Delta_+^{u}\Delta_-^{0}X_t=0$, to see that $R_{1t}(w)$ is in fact zero when $d_0=b_0$, i.e.,

$$R_{1t}(w)|_{d_0=b_0}=\Delta_+^{w-d_0+b_0}\Delta_-^{d_0-b_0}X_t|_{d_0=b_0}=\Delta_+^{w}\Delta_+^{0}\Delta_-^{0}X_t=0.$$

We therefore assume $d_0>b_0$ in the proof for $R_{1t}(w)$. Let $u=w-d_0+b_0$, $v=d_0-b_0$ such that for $w\geq d_0-1/2-\kappa_1$ we find $u+v+1=w+1\geq d_0+1/2-\kappa_1=a_1>0$ and $v=d_0-b_0=a_2>0$. Then (21) shows that

$$|\Delta_+^{u}\Delta_-^{v}X_t|\leq c(1+\log t)t^{\max\{-u-1,-v,-2u-v-1\}}\leq c(1+\log t)t^{\max\{-u-1,-a_2,-a_1-u\}}.$$

Moreover, $u+a_1\geq d_0+b_0-2\kappa_1\geq 1/2-2\kappa_1>0$ and $u+1\geq 1/2+b_0-\kappa_1\geq 1/2-\kappa_1>0$, such that $\sup_{w\in S_+}|R_{1t}(w)|\to 0$.

The proof for $R_{2t}(w)$ is the same as that for $R_{1t}(w)$, setting $u=w-d_0$ and $v=d_0+jb_0\geq d_0>0$. Finally, for $R_{3t}(w)$ we let $u=w-d_0+b_0$, $v=d_0+jb_0\geq d_0>0$ and apply the same proof as for $R_{1t}(w)$.

Proof of (18) for the terms involving μ_{0t}: Again only the case $d_0>1/2$ needs to be considered (because $S_-=\varnothing$ when $d_0=1/2$) and there are three types of terms to be analyzed: (i) The terms $\sup_{d+ib\in S_-}|T^{d+ib-d_0+1/2}\Delta_+^{d+ib}\mu_{0t}|$ for $i=-1,\ldots,k$, which are all equal, (ii) the terms $\sup_{d+ib\in S_-}|T^{d+ib-d_0+1/2}\Delta_+^{d+kb}\mu_{0t}|$ for $i=0,\ldots,k$, and (iii) the term $\sup_{d-b\in S_-}T^{d-b-d_0+1/2}|\Delta_+^{d}\mu_{0t}|$. Thus, for $w=d+ib\in S_-$, we analyze

$$T^{w-d_0+1/2}\beta_{0\perp}'F_+(L)\alpha_0\beta_0'\Delta_+^{w+hb-d_0+b_0}\Delta_-^{d_0-b_0}X_t \quad (24)$$
$$-T^{w-d_0+1/2}\sum_{j=0}^{k}(C_0\Psi_{0j}\Delta_+^{w+hb-d_0}+F_+(L)\Psi_{0j}\Delta_+^{w+hb-d_0+b_0})\Delta_-^{d_0+jb_0}X_t,$$

where $h = 0$ (for terms of type (i)), $h = (k - i)b$ (for terms of type (ii)), or $h = b$ (for the term of type (iii)), see (15). By application of the Dominated Convergence Theorem we only need to prove that $\sup_{w \in S_-} |Q_{nT}(w)| \to 0$ as $T \to \infty$ for $n = 1, 2, 3$, where

$$\begin{aligned}
Q_{1T}(w) &= T^{w-d_0+1/2} \max_{1 \le t \le T} |\Delta_+^{w+hb-d_0+b_0} \Delta_-^{d_0-b_0} X_t|, \\
Q_{2T}(w) &= T^{w-d_0+1/2} \max_{1 \le t \le T} |\Delta_+^{w+hb-d_0} \Delta_-^{d_0+jb_0} X_t|, \\
Q_{3T}(w) &= T^{w-d_0+1/2} \max_{1 \le t \le T} |\Delta_+^{w+hb-d_0+b_0} \Delta_-^{d_0+jb_0} X_t|.
\end{aligned}$$

Each term has a factor with a bound from (20) for suitable choices of u and v. Note that all three cases have either $v = d_0 - b_0 \ge 0$ or $v = d_0 + jb_0 \ge d_0 > 0$, which implies that $u - v \le u$ and $u + v \ge w \ge -\eta_1$, so that $-v \le u + \eta_1$. This shows that $\max\{-v, -1, u - v, u\} = \max\{u + \eta_1, -1\}$, so that the bound in (20), multiplied by T^z, becomes

$$|T^{u+z} \Delta_+^u \Delta_-^v X_t| \le c(1 + \log T) T^{\max\{u+z+\eta_1, z-1\}}. \tag{25}$$

For $n = 1, 2, 3$ we apply (25) with the choices

$$\begin{aligned}
n = 1 &: u = w + hb - d_0 + b_0, z = 1/2 - hb - b_0, \\
n = 2 &: u = w + hb - d_0, z = 1/2 - hb, \\
n = 3 &: u = w + hb - d_0 + b_0, z = 1/2 - hb - b_0,
\end{aligned}$$

respectively. For all three cases we find that $u + z + \eta_1 = w - d_0 + 1/2 + \eta_1 \le -\kappa_1 + \eta_1 < 0$ by choice of $\eta_1 < \kappa_1$, and for all three cases we find that $z - 1 \le 1/2 - hb - 1 \le -1/2 < 0$, so it follows from (25) that $\sup_{w \in S_-} |Q_{nT}(w)| \to 0$ as $T \to \infty$ for $n = 1, 2, 3$. ∎

Proof of Lemma 1(ii). The deterministic term $d_{0t} = \Pi_{0-}(L)(\tilde{X}_t - X_t)$ depends on the terms $\Delta_-^{d_0+ib_0} X_t$ for $i \ge -1$.

Suppose first that Assumption 5(i) is satisfied. We then apply (19) to obtain the bound

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

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

Next, suppose Assumption 5(ii) is satisfied. We again apply (19) and find

$$|\Delta_-^{d_0+ib_0} X_t| = |\sum_{n=0}^{M_0-1} \pi_{n+t}(-d_0 - ib_0) X_{-n,T}| \le c \sum_{n=0}^{M_0-1} (n + t)^{-(d_0-b_0)-1} \le c M_0 t^{-(d_0-b_0)-1}. \tag{26}$$

It follows that $T^{-1/2} \sum_{t=1}^T |d_{0t}| \to 0$. ∎

The final result, given as Lemma 2, presents a new bound on $|\mathsf{D}_u^m \Delta_+^u \Delta_-^v X_t|$ and thus improves Lemma C.1 in JN(2010). This bound is critical to the analysis of the initial values on the larger parameter space compared with Lemma A.8 in JN(2012a). Furthermore, it is also this new bound that allows us to include the case $d_0 = 1/2$ in the proof of Lemma 1, which was missing in the proof of Lemma A.8 in JN(2012a).

Furthermore, the bound in Lemma 2 allows us to avoid the condition that $\kappa_1 < d_0 - 1/2$ when $d_0 > 1/2$ in Lemma 1, which was assumed in Lemma A.8 in JN(2012a), but apparently was overlooked in the statement of the main theorems and assumptions in JN(2012a). Specifically, this would have required the existence of $q > (d_0 - 1/2)^{-1}$ when $d_0 > 1/2$, in addition to other conditions on q, so that κ_1 can be chosen to satisfy $q^{-1} < \kappa_1 < d_0 - 1/2$. The use of our new Lemma 2 allows us to avoid this condition in the proof of Lemma 1 and hence avoid strengthening the moment condition on q.

Lemma 2 *Let Assumption 3 be satisfied. Then, uniformly for $u + v + 1 \geq a_1 > 0$ and $v \geq a_2 > 0$, it holds that*

$$|\mathsf{D}_u^m \Delta_+^u \Delta_-^v X_t| \leq c(1 + \log t)^{m+1} t^{\max\{-u-1,-v,-2u-v-1\}},$$

where the constant c does not depend on u, v, m, or t.

Proof of Lemma 2. We prove the result for $m = 0$ and find that

$$|\Delta_+^u \Delta_-^v X_t| \leq c \sum_{j=0}^{t-1} \sum_{k=t-j}^{\infty} |\pi_j(-u)||\pi_k(-v)| = A_t + B_t,$$

where the inner summation is split in two at $k = \max\{j, t - j\}$ to define

$$A_t = c \sum_{j=0}^{t-1} \sum_{k=t-j}^{\max\{j,t-j\}-1} |\pi_j(-u)||\pi_k(-v)| \text{ and } B_t = c \sum_{j=0}^{t-1} \sum_{k=\max\{j,t-j\}}^{\infty} |\pi_j(-u)||\pi_k(-v)|.$$

Next, for $r > s$ we use the decomposition and evaluations

$$\pi_r(w) = \pi_s(w) \prod_{i=s+1}^{r} (1 + (w-1)/i) = \pi_s(w)\alpha_{s,r}(w), \tag{27}$$

$$|\pi_s(w)| \leq cs^{w-1}, \text{ and } |\alpha_{s,r}(w)| \leq cr^{w-1},$$

where the constant c does not depend on w, r, or s; see Lemma A.3 of JN(2016). We will also need Lemma B.4 from JN(2010), which shows that

$$\sum_{j=1}^{t-1} j^{u-1}(t-j)^{v-1} \leq c(1 + \log t) t^{\max\{u-1,v-1,u+v-1\}}, \tag{28}$$

where c does not depend on t, u, v for $|u| \leq u_0, |v| \leq v_0$.

Proof for A_t: Note that the condition $t - j \leq k \leq \max\{j, t - j\} - 1$ in the summation in A_t implies that $k \leq j - 1$ because $t - j \leq k \leq t - j - 1$ is not possible. Thus, for $k < j$ we find from (27) that

$$|\pi_j(-u)||\pi_k(-v)| = |\pi_k(-u)||\alpha_{k,j}(-u)||\pi_k(-v)| \leq cj^{-u-1}k^{-(u+v+1)-1}$$

and it follows that

$$A_t \leq c \sum_{j=t/2}^{t-1} j^{-u-1} \sum_{k=t-j}^{j-1} k^{-(u+v+1)-1} \leq c \sum_{j=t/2}^{t-1} j^{-u-1}(t-j)^{-(u+v+1)}$$
$$\leq c(1+\log t)t^{\max\{-u-1,-u-v-1,-2u-v-1\}},$$

where we used (28) and that $\sum_{k=t-j}^{j-1} k^{-(u+v+1)-1} \leq \sum_{k=t-j}^{\infty} k^{-(u+v+1)-1} \leq c(t-j)^{-(u+v+1)}$ for $u+v+1 \geq a_1 > 0$.

Proof for B_t: In the summation in B_t the condition $k \geq \max\{j, t-j\}$ implies that $k \geq j$ and $k \geq t/2$. Then, from (27) we find that

$$|\pi_j(-u)||\pi_k(-v)| = |\pi_j(-u)||\pi_j(-v)||\alpha_{j,k}(-v)| \leq cj^{-(u+v+1)-1}k^{-v-1},$$

and it follows that

$$B_t \leq c\sum_{j=0}^{t-1} j^{-(u+v+1)-1} \sum_{k=t/2}^{\infty} k^{-v-1} \leq ct^{-v} \sum_{j=0}^{t-1} j^{-(u+v+1)-1} \leq ct^{-v},$$

where we used that $\sum_{k\geq t/2}^{\infty} k^{-v-1} \leq ct^{-v}$ for $v \geq a_2 > 0$ and $\sum_{j=0}^{t-1} j^{-(u+v+1)-1} \leq c$ for $u+v+1 \geq a_1 > 0$. ■

4 Conclusions and discussion

In this paper, we have shown that the test statistic for the usual CVAR model in the more general fractional CVAR model is asymptotically chi-squared distributed. In the analysis of the fractional CVAR in Johansen and Nielsen (2012a), the usual CVAR was on the boundary of the parameter space, so in this article we studied the fractional CVAR model on a slightly larger parameter space for which the CVAR model lies in the interior. This analysis required improving several related results in Johansen and Nielsen (2012a); in particular regarding the negligibility of the contribution of the initial values of the process to the likelihood function.

Our main results, presented in Corollary 1, show that the likelihood ratio test of the usual CVAR is asymptotically $\chi^2(2)$ and that the likelihood ratio test of the less restrictive hypothesis that $d = b$ in the fractional model is asymptotically $\chi^2(1)$. Thus, the tests are very easy to implement and can be calculated straightforwardly using the software package of Nielsen and Popiel (2016). Both tests are important in empirical analysis as part of model determination to test a more simple and parsimonious formulation of the empirical model. As mentioned in the introduction, these tests have been calculated in empirical work (see references in the introduction) with a conjectured χ^2-distribution, which we have now verified.

References

1. Agiakloglou, C., and Newbold, P. (1994). Lagrange multiplier tests for fractional difference. *Journal of Time Series Analysis* **14**, 253–262.
2. Anderson, T.W. (1971). *The Statistical Analysis of Time Series*, John Wiley & Sons, New York.
3. Andrews, D.W.K. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica* **69**, 683–734.

4. Bollerslev, T., Osterrieder, D., Sizova, N., and Tauchen, G. (2013). Risk and return: long-run relations, fractional cointegration and return predictability. *Journal of Financial Economics* **108**, 409–424.
5. 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.
6. Chen, C.Y., Chiang, T.C., and Härdle, W.K. (2016). Downside risk and stock returns: an empirical analysis of the long-run and short-run dynamics from the G-7 countries. SFB Discussion Paper 649, Humboldt-Universität zu Berlin.
7. Dolado, J.J., Gonzalo, J., and Mayoral, L. (2002). A fractional Dickey-Fuller test for unit roots. *Econometrica* **70**, 1963–2006.
8. 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* **38**, 623–639.
9. Hamilton, J.D. (1994). *Time Series Analysis*, Princeton University Press, Princeton.
10. Jensen, A.N., and Nielsen, M.Ø. (2014). A fast fractional difference algorithm. *Journal of Time Series Analysis* **35**, 428–436.
11. Johansen, S. (1996). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, 2nd ed., Oxford University Press, Oxford.
12. Johansen, S. (2008). A representation theory for a class of vector autoregressive models for fractional processes. *Econometric Theory* **24**, 651–676.
13. Johansen, S., and Nielsen, M.Ø. (2010). Likelihood inference for a nonstationary fractional autoregressive model. *Journal of Econometrics* **158**, 51–66.
14. Johansen, S., and Nielsen, M.Ø. (2012a). Likelihood inference for a fractionally cointegrated vector autoregressive model. *Econometrica* **80**, 2667–2732.
15. Johansen, S., and Nielsen, M.Ø. (2012b). A necessary moment condition for the fractional functional central limit theorem. *Econometric Theory* **28**, 671–679.
16. 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.
17. 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.
18. Robinson, P.M. (1991). Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics* **47**, 67–84.
19. Tanaka, K. (1999). The nonstationary fractional unit root. *Econometric Theory* **15**, 549–582.