AARHUS UNIVERSITY

Coversheet

This is the accepted manuscript (post-print version) of the article.

Contentwise, the accepted manuscript version is identical to the final published version, but there may be differences in typography and layout.

How to cite this publication

Please cite the final published version:

Carlini, F., & Santucci de Magistris, P. (2017). On the identification of fractionally cointegrated VAR models with the F(d) condition. Journal of Business and Economic Statistics.

<https://doi.org/10.1080/07350015.2017.1294077>

Publication metadata

Title:	On the identification of fractionally cointegrated VAR models with the F(d) condition.
Author(s):	Federico Carlini & Paolo Santucci de Magistris
Journal:	Journal of Business and Economic Statistics
DOI/Link:	10.1080/07350015.2017.1294077
Document version:	Accepted manuscript (post-print)

General Rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

If the document is published under a Creative Commons license, this applies instead of the general rights.

On the Identification of Fractionally Cointegrated VAR Models with the $\mathcal{F}(d)$ Condition

Federico Carlini* Paolo Santucci de Magistris [†]

January 26, 2017

Abstract

This paper discusses identification problems in the fractionally cointegrated system of Johansen (2008) and Johansen and Nielsen (2012). It is shown that several equivalent re-parametrizations of the model associated with different fractional integration and cointegration parameters may exist for any choice of the lag-length when the true cointegration rank is known. The properties of these multiple non-identified models are studied and a necessary and sufficient condition for the identification of the fractional parameters of the system is provided. The condition is named $\mathcal{F}(d)$. This is a generalization of the well-known $I(1)$ condition to the fractional case. Imposing a proper restriction on the fractional integration parameter, d , is sufficient to guarantee identification of all model parameters and the validity of the $\mathcal{F}(d)$ condition. The paper also illustrates the indeterminacy between the cointegration rank and the lag-length. It is also proved that the model with rank zero and k lags may be an equivalent re-parametrization of the model with full rank and $k - 1$ lags. This precludes the possibility to test for the cointegration rank unless a proper restriction on the fractional integration parameter is imposed.

Keywords: Fractional Cointegration; Cofractional Model; Identification; Lag Selection.

JEL Classification: C18, C32, C52

*CREATES, Department of Economics and Business Economics, Aarhus University.

[†]**Corresponding author:** CREATES, Department of Economics and Business Economics, Aarhus University, Fuglesangs Alle 4, 8210 Aarhus V, Denmark. Tel.: +45 8716 5319. E-mail address: psantucci@econ.au.dk. The authors acknowledge support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. The authors are grateful to Niels Haldrup, Søren Johansen, Katarzyna Łasak, Bent Nielsen and Morten Ørregaard Nielsen for their suggestions that improved the quality of this work. The authors are also grateful to an anonymous referee for providing insightful comments. The authors would like to thank also James MacKinnon, Rocco Mosconi, Paolo Paruolo, the participants to the Third Long Memory Symposium (Aarhus 2013), the participants to the CFE'2013 conference (London 2013), and the seminar participants at Queen's University and at Bologna University for helpful comments. Finally, special thanks to Sølveig Nygaard Sørensen for her careful proofreading of the final version of this paper.

1 Introduction

The past decade has witnessed an increasing interest in the statistical definition of the concept of *fractional cointegration*, as a generalization of the idea of cointegration to processes with fractional degrees of integration. In the context of long-memory processes, fractional cointegration allows linear combinations of $I(d)$ processes to be $I(d - b)$, with $d, b \in \mathbb{R}_+$ with $0 < b \leq d$. More specifically, the concept of fractional cointegration implies the existence of common stochastic trends integrated of order d , with short-period departures from the long-run equilibrium integrated of order $d - b$. The coefficient b is the degree of fractional reduction obtained by the linear combination of $I(d)$ variables, namely the *cointegration gap*.

Notable methodological works in the field of fractional cointegration are Robinson and Marinucci (2003) and Christensen and Nielsen (2006) that develop regression-based semi-parametric methods to evaluate whether two fractional processes share common stochastic trends. Analogously, Hualde and Velasco (2008) propose to check for the absence of cointegration by comparing the estimates of the cointegration vector obtained with OLS and those obtained with a GLS type of estimator. Breitung and Hassler (2002) propose a multivariate score test statistic to determine the cointegration rank that is obtained by solving a generalized eigenvalue problem of the type proposed by Johansen (1988). Alternatively, Robinson and Yajima (2002) and Nielsen and Shimotsu (2007) suggest a testing procedure to evaluate the cointegration rank of the multivariate coherence matrix of two, or more, fractionally differenced series. Chen and Hurvich (2003, 2006) estimate cointegrated spaces and subspaces by the eigenvectors corresponding to the r smallest eigenvalues of an averaged periodogram matrix of tapered and differenced observations.

Despite the effort spent in defining testing procedures for the presence of fractional cointegration, for a long time the literature in this area lacked a fully parametric multivariate model explicitly characterizing the joint behavior of fractionally cointegrated processes. Interestingly, Granger (1986, p.222) already introduced the idea of common trends between $I(d)$ processes, but the subsequent theoretical works, see among many others Johansen (1988), have mostly been dedicated to cases with integer orders of integration. More recently, Johansen (2008) and Johansen and Nielsen (2012) have proposed the FCVAR $_{d,b}$ model, an extension of the well-known VECM to fractional processes, which is a tool for a direct modeling and testing of fractional cointegration. In particular, Johansen and Nielsen (2012) have extended the estimation method of Łasak (2010) to the FCVAR $_{d,b}$ model, deriving the asymptotic properties of the profile maximum likelihood estimator when $0 \leq d - b < 1/2$ and $b \neq 1/2$. Other contributions in the parametric framework for fractional cointegration are in Avarucci and Velasco (2009), Franchi (2010)

and Łasak and Velasco (2015).

This paper shows that the FCVAR _{d,b} model is not globally identified when the number of lags, k , is unknown. For a given number of lags, several sub-models with the same conditional densities but different values of the parameters may exist. Hence the parameters of the FCVAR _{d,b} model can not be uniquely identified and the multiplicity of not-identified sub-models can be determined for any FCVAR _{d,b} model with k lags. An analogous identification problem, for the FIVAR _{b} model is discussed in Tschernig et al. (2013a,b). It is proved that the $I(1)$ condition in the VECM of Johansen (1988) can be generalized to the fractional context. In analogy with the $I(1)$ condition for integer orders of integration, this condition is named $\mathcal{F}(d)$, and it is a necessary and sufficient condition for the identification of the parameters of the system. If the $\mathcal{F}(d)$ condition is not satisfied, the FCVAR _{d,b} parameters, including fractional and co-fractional parameters, d and b , can not be uniquely determined.

The identification problems in the FCVAR _{d,b} model are studied along the following lines. First, Proposition 1 extends the results in Theorem 3 of Johansen and Nielsen (2012), highlighting the close relationship between the lag structure and the lack of identification, and deriving a necessary and sufficient condition for identification associated to any lag-length. Proposition 1 also highlights the consequence of the indeterminacy of the lag-length on the fractional parameters d and b . Second, the paper shows the consequence of the lack of identification on the log-likelihood function, both asymptotically and in finite samples. Differently from the standard case, where the integration orders are fixed to integer values, the estimation of the FCVAR _{d,b} involves the maximization of the profile log-likelihood with respect to d and b , but the latter is affected by the indeterminacy generated by the over-specification of the lag-length. Moreover, we discuss a further identification issue that emerges when the cointegration rank is unknown. Under certain restrictions on d and b , the FCVAR _{d,b} with full rank and k lags is equivalent to the FCVAR _{d,b} with rank 0 and $k + 1$ lags. This last finding precludes the possibility to test for the absence of cointegration when the true number of lags is unknown based on the unrestricted FCVAR _{d,b} model. Finally, we prove that the FCVAR _{d,b} is identified for any lag $k > 1$, both in the known and unknown rank cases, if the fractional parameter d is restricted to be equal to the true fractional order, such that the $\mathcal{F}(d)$ condition is satisfied by construction. Building on this result, we show that it is sufficient to restrict the parameter set of d to solve the identification problem. In particular, d must belong to the sub-interval of $\mathbb{R}^+$ that includes the true fractional order, d_0 , but excludes other values of $d < d_0$ associated to equivalent models. The information about the true fractional order can be obtained by the exact local Whittle estimator of Shimotsu and Phillips (2005).

This paper is organized as follows. Section 2 discusses the identification problem from

a theoretical point of view. Section 3 discusses the consequences of the lack of identification on the inference on the parameters of the FCVAR_{*d,b*} model both asymptotically and in finite samples. Section 4 discusses the problems when the cointegration rank and the lag-length are both unknown. Section 5 concludes the paper. All proofs and additional numerical results are in the document with supplementary material.

2 The Identification Problem

This section provides a discussion of the identification problem related to the FCVAR_{*d,b*} model

$$\mathcal{H}_k : \quad \Delta^d X_t = \alpha \beta' \Delta^{d-b} L_b X_t + \sum_{i=1}^k \Gamma_i \Delta^d L_b^i X_t + \varepsilon_t \quad \varepsilon_t \sim iidN(0, \Omega), \quad (1)$$

where X_t is a p -dimensional vector, α and β are $p \times r$ matrices, and r defines the cointegration rank. Ω is the positive definite covariance matrix of the errors, and Γ_j , $j = 1, \dots, k$, are $p \times p$ matrices loading the short-run dynamics. The operator $L_b := 1 - \Delta^b$ is the so-called *fractional lag operator*, which, as noted by Johansen (2008), is necessary for characterizing the solutions of the system in (1) and obtaining the associated Granger representation in the fractional context. Following Definition 1 in Johansen and Nielsen (2012, p.2672), if X_t follows (1), then X_t is a fractional process of order d , denoted as $\mathcal{F}(d)$, and co-fractional of order $d - b$. The symbol $\mathcal{H}_k$ defines the model with k lags and $\theta = vec(d, b, \alpha, \beta, \Gamma_1, \dots, \Gamma_k, \Omega)$ is the parameter vector. The parameter space of model $\mathcal{H}_k$ is

$$\Theta_{\mathcal{H}_k} = \{\alpha \in \mathbb{R}^{p \times r_0}, \beta \in \mathbb{R}^{p \times r_0}, \Gamma_j \in \mathbb{R}^{p \times p}, j = 1, \dots, k, d \in \mathbb{R}^+, b \in \mathbb{R}^+, d \geq b > 0, \Omega > 0\},$$

where r_0 is the true cointegration rank and it is assumed known. The results in this section are obtained under the maintained assumption that the true cointegration rank is known and such that $0 < r_0 < p$. An extension to the case of unknown rank and number of lags is presented in Section 4.

Similarly to Johansen (2010), the concept of identification and equivalence between two models is formally introduced by the following definition.

Definition 1 *Let $\{\mathcal{P} = P_\theta, \theta \in \Theta\}$ be a family of probability measures, that is, a statistical model. We say that a parameter function $g(\theta)$ is identified if $g(\theta_1) \neq g(\theta_2)$ implies that $P_{\theta_1} \neq P_{\theta_2}$. On the other hand, if $P_{\theta_1} = P_{\theta_2}$ and $g(\theta_1) \neq g(\theta_2)$, the parameter function $g(\theta)$ is not identified. In this case, the statistical models P_{θ_1} and P_{θ_2} are equivalent.*

As noted in Johansen (1995a, p.177), the product $\alpha\beta'$ is identified but not the matrices α and β because if there was an $r \times r$ matrix ξ , the product $\alpha\beta'$ would be equal to $\alpha_\xi\beta'_\xi$, where $\alpha_\xi = \alpha\xi$ and $\beta_\xi = \beta(\xi')^{-1}$. In the following we do not discuss the identification of α and β , that is generally solved by a proper normalization of β . It can be shown instead that the other parameters of the FCVAR $_{d,b}$ model in (1) might not be identified, i.e. several *equivalent* sub-models associated with different values θ , can be found when k is overspecified. An illustration of the identification problem is provided by the following example.

Example 1: Consider the FCVAR $_{d,b}$ model with one lag,

$$\mathcal{H}_1 : \Delta^d X_t = \alpha\beta' \Delta^{d-b} L_b X_t + \Gamma_1 \Delta^d L_b X_t + \varepsilon_t, \quad (2)$$

which can be written as

$$\{\Delta^d [I_p + \alpha\beta' - \Gamma_1] + \Delta^{d-b} [-\alpha\beta'] + \Delta^{d+b} \Gamma_1\} X_t = \varepsilon_t.$$

First, examine the restriction, $\mathcal{H}_1^{(0)} : \Gamma_1^0 = 0$. Under $\mathcal{H}_1^{(0)}$, the model in equation (2) can be rewritten as

$$\{\Delta^{d_0} [I_p + \alpha\beta'] + \Delta^{d_0-b_0} [-\alpha\beta']\} X_t = \varepsilon_t.$$

Second, consider instead the restriction $\mathcal{H}_1^{(1)} : I_p + \alpha\beta' - \Gamma_1^1 = 0$. It follows that

$$\{\Delta^{d_1-b_1} [-\alpha\beta'] + \Delta^{d_1+b_1} [I_p + \alpha\beta']\} X_t = \varepsilon_t.$$

Given that the condition $\alpha\beta' \Delta^{d_0-b_0} = \alpha\beta' \Delta^{d_1-b_1}$ must hold in both sub-models, hence model (2) under $\mathcal{H}_1^{(0)}$ is equivalent to the model (2) under $\mathcal{H}_1^{(1)}$ if and only if

$$[I_p + \alpha\beta'] \Delta^{d_0} = [I_p + \alpha\beta'] \Delta^{d_1+b_1}.$$

This leads to the system of two equations in d_0 , b_0 , d_1 and b_1

$$\begin{cases} d_0 - b_0 = d_1 - b_1 \\ d_0 = d_1 + b_1 \end{cases} \quad (3)$$

which has a unique solution when $d_1 = d_0 - b_0/2$ and $b_1 = b_0/2$. Since the restrictions $\mathcal{H}_1^{(0)}$ and $\mathcal{H}_1^{(1)}$ lead to equivalent descriptions of the data, it follows that the fractional order of X_t implied by both models must be the same. However, in $\mathcal{H}_1^{(0)}$ the fractional order is represented by the parameter d_0 , i.e. $X_t \sim \mathcal{F}(d_0)$ since $\Delta^{d_0} X_t \sim \mathcal{F}(0)$, while in $\mathcal{H}_1^{(1)}$ the fractional order is given by the sum $d_1 + b_1$, i.e. $X_t \sim \mathcal{F}(d_1 + b_1)$. The identification

condition defined in 1 is clearly violated, as the conditional densities of $\mathcal{H}_1^{(0)}$ and $\mathcal{H}_1^{(1)}$ are such that

$$p_{\mathcal{H}_1^{(0)}}(X_1, \dots, X_T, \theta_0 | X_0, X_{-1}, \dots) = p_{\mathcal{H}_1^{(1)}}(X_1, \dots, X_T, \theta_1 | X_0, X_{-1}, \dots), \quad (4)$$

where $\theta_0 = \text{vec}(d_0, b_0, \alpha, \beta, \Omega)$ and $\theta_1 = \text{vec}(d_1, b_1, \alpha, \beta, \Gamma_1^1, \Omega)$ with $\Gamma_1^1 = I_p + \alpha\beta'$.

Example 1 can be extended to a generic lag-length $k_0 \geq 0$. Consider the model $\mathcal{H}_{k_0}$

$$\mathcal{H}_{k_0} : \quad \Delta^{d_0} X_t = \alpha_0 \beta_0' \Delta^{d_0 - b_0} L_{b_0} X_t + \sum_{i=1}^{k_0} \Gamma_i^0 \Delta^{d_0} L_{b_0}^i X_t + \varepsilon_t \quad \varepsilon_t \sim N(0, \Omega_0), \quad (5)$$

with $k_0 \geq 0$ lags, and $|\alpha'_{0,\perp} \Gamma^0 \beta_{0,\perp}| \neq 0$ with $\Gamma^0 = I_p - \sum_{i=1}^{k_0} \Gamma_i^0$. When a model $\mathcal{H}_k$ with $k > k_0$ is considered, then $\mathcal{H}_{k_0}$ is associated with the set of restrictions $\mathcal{H}_k^{(0)} : \Gamma_{k_0+1} = \Gamma_{k_0+2} = \dots = \Gamma_k = 0$ imposed on $\mathcal{H}_k$. However, there may be several alternative restrictions on $\Gamma_{k_0+1}, \Gamma_{k_0+2}, \dots, \Gamma_k$ leading to an equivalent sub-model as the one obtained under $\mathcal{H}_k^{(0)}$.

The following Proposition states the necessary and sufficient condition, called the $\mathcal{F}(d)$ condition, for identification of the parameters of the model $\mathcal{H}_k$.

Proposition 1 *Consider a FCVAR_{d,b} model with k lags,*

- i) Given $k > k_0 \geq 0$, the $\mathcal{F}(d)$ condition, defined as $|\alpha'_\perp \Gamma \beta_\perp| \neq 0$ with $\Gamma = I_p - \sum_{i=1}^k \Gamma_i$, is a necessary and sufficient condition for the identification of the set of parameters of $\mathcal{H}_k$ in equation (5).*
- ii) Given k_0 and k , with $k \geq k_0$, the number of equivalent sub-models that can be obtained from $\mathcal{H}_k$ is $m = \lfloor \frac{k+1}{k_0+1} \rfloor$, where $\lfloor x \rfloor$ denotes the greatest integer less or equal to x .*
- iii) For any $k \geq k_0$, all the equivalent sub-models are found for parameter values $d_j = d_0 - \frac{j}{j+1} b_0$ and $b_j = b_0/(j+1)$ for $j = 0, 1, \dots, m-1$.*

Proposition 1 has several important consequences that are worth being discussed in detail. First of all, the $\mathcal{F}(d)$ condition only holds for the sub-model of $\mathcal{H}_k$ for which $d = d_0$ and $b = b_0$, i.e. for the sub-model of $\mathcal{H}_k$ associated to the restriction $\mathcal{H}_k^{(0)} : \Gamma_{k_0+1} = \Gamma_{k_0+2} = \dots = \Gamma_k = 0$. In Example 1, the $\mathcal{F}(d)$ condition is only verified for $\mathcal{H}_1^{(0)}$, while for $\mathcal{H}_1^{(1)}$ we have that $|\alpha'_\perp \Gamma^1 \beta_\perp| = 0$, since $\Gamma^1 = I_p - (I_p + \alpha\beta') = -\alpha\beta'$. Note that the assumption $|\alpha'_{0,\perp} \Gamma^0 \beta_{0,\perp}| \neq 0$ imposed on model (5) guarantees that it is not possible to impose restrictions such that two or more sub-models of $\mathcal{H}_{k_0}$ are equivalent. In this sense, Proposition 1 generalizes Theorem 3 in Johansen and Nielsen (2012). Indeed, while

in Johansen and Nielsen (2012) the $\mathcal{F}(d)$ condition is only imposed on the $\mathcal{H}_{k_0}$ model with $k = k_0$ by assumption, Proposition 1.i) shows that a necessary and sufficient condition for the identification of the parameters of any $\mathcal{H}_k$ model, with $k > k_0$, is the validity of the $\mathcal{F}(d)$ condition. This has important consequences in practical applications when the true number of lags is unknown and it is potentially over-specified. Note that if the number of lags is under-specified, there is no identification problem, but this misspecification model leads to the invalidity of the results in Johansen and Nielsen (2012). In addition, Proposition 1.ii) characterizes the number of equivalent sub-models of $\mathcal{H}_k$ for a given k_0 , showing that their multiplicity depends on k and k_0 . Table 1 summarizes the number of equivalent sub-models for different values of k_0 and k . As a consequence of Proposition 1.ii), there are cases in which $k > k_0$ does not necessarily imply a lack of identification. For example, when $k = 2$ and $k_0 = 1$, there are no sets of restrictions on $\mathcal{H}_2$ leading to a sub-model equivalent to the one obtained under the restriction $d = d_0$, $b = b_0$, $\Gamma_1 = \Gamma_1^0$ and $\Gamma_2 = 0$. Hence, in this case, the multiplicity, m , of equivalent sub-models

$k_0 \downarrow k \rightarrow$	0	1	2	3	4	5	6	7	8	9	10	11	12
0	1	2	3	4	5	6	7	8	9	10	11	12	13
1	–	1	1	2	2	3	3	4	4	5	5	6	6
2	–	–	1	1	1	2	2	2	3	3	3	4	4
3	–	–	–	1	1	1	1	2	2	2	2	3	3
4	–	–	–	–	1	1	1	1	1	2	2	2	2
5	–	–	–	–	–	1	1	1	1	1	1	2	2

Table 1: Table reports the number of equivalent models (m) for different combinations of k and k_0 . When $k_0 > k$ the $\mathcal{H}_k$ is under-specified.

is 1. When k_0 is small there are several equivalent sub-models for small choices of k . As k_0 increases, multiple equivalent sub-models are only found for large values of k . For example, when $k_0 = 5$, then two equivalent sub-models can only be found for suitable restrictions of the $\mathcal{H}_{11}$ model. Moreover, Proposition 1.iii) shows that each sub-model of $\mathcal{H}_k$ equivalent to $\mathcal{H}_{k_0}$ with $|\alpha'_\perp \Gamma \beta_\perp| = 0$ has values of d and b that are fractions of d_0 and b_0 . Furthermore, when k is very large compared to k_0 , the $(m - 1)$ -th sub-model is associated with $d_{m-1} \approx d_0 - b_0$ and $b_{m-1} \approx 0$, i.e. located closely to the boundary of the parameter space.

A final remark on the relevance of the $\mathcal{F}(d)$ concerns the possibility of polynomial fractional cointegration that arises when $|\alpha'_\perp \Gamma \beta_\perp| = 0$. This means that models with polynomial fractional cointegration up to order $m = \lfloor \frac{k+1}{k_0+1} \rfloor$ can be obtained from the FCVAR $_{d,b}$ model for some combinations of k and k_0 . Consider again model $\mathcal{H}_1^{(1)}$ in (2).

After simple algebraical manipulations, model $\mathcal{H}_1^{(1)}$ can be formulated as

$$\Delta^{d_2} X_t = \Delta^{d_2-2b_1} (\alpha\beta' L_{b_1} X_t - \Gamma^1 \Delta^{b_1} L_{b_1} X_t) + \epsilon_t, \quad (6)$$

where $d_2 = d_1 + b_1$ and $\Gamma^1 = -\alpha\beta'$. Equation (6) defines a model for polynomial fractional cointegration as studied in Johansen (2008, p.667) and further extended in Franchi (2010). Polynomial fractional cointegration has analogies in the context of modeling I(2) variables in the standard VECM context, see Johansen (1995b), and the number of polynomial fractional trends depends on the rank of the matrix $\alpha'_\perp \Gamma \beta_\perp$. In particular, imposing the $\mathcal{F}(d)$ condition on the FCVAR $_{d,b}$ model does not only guarantee that the parameters d , b and $\Gamma_1, \dots, \Gamma_k$ are identified, but also rules out cases of polynomial fractional cointegration.

3 Identification and Inference

This section illustrates the problems in the estimation of the parameters of the FCVAR $_{d,b}$ that are induced by the lack of identification outlined in Section 2. As shown in Johansen and Nielsen (2012), the parameters of the FCVAR $_{d,b}$ can be estimated following a profile likelihood approach, where the estimates of the fractional parameters, $\hat{d}$ and $\hat{b}$, are obtained first by maximizing the profile log-likelihood

$$\hat{\psi} = \arg \max_{\psi} \ell_T(\psi), \quad (7)$$

where $\psi = (d, b)'$ and

$$\ell_T(\psi) = -\log |S_{00}(\psi)| - \sum_{l=1}^r \log(1 - \lambda_h(\psi)). \quad (8)$$

The quantities $\lambda(\psi)$ and $S_{00}(\psi)$ are obtained from the residuals, $R_{it}(\psi)$ of the reduced rank regression of $\Delta^d X_t$ ($i = 0$) and $\Delta^{d-b} L_b X_t$ ($i = 1$) on $\Delta^d L_b X_t, \dots, \Delta^d L_b^k X_t$. The product moment matrices $S_{ij}(\psi)$ for $i, j = 0, 1$ are $S_{ij}(\psi) = T^{-1} \sum_{t=1}^T R_{it}(\psi) R'_{jt}(\psi)$ and $\lambda_h(\psi)$ for $h = 1, \dots, p$ are the solutions, sorted in decreasing order, of the generalized eigenvalue problem

$$|\lambda(\psi) S_{11}(\psi) - S_{10}(\psi) S_{00}^{-1}(\psi) S_{01}(\psi)| = 0. \quad (9)$$

Given $\hat{d}$ and $\hat{b}$, the estimates $\hat{\alpha}$, $\hat{\beta}$, $\hat{\Gamma}_j$, $j = 1, \dots, k$, and $\hat{\Omega}$ are found by reduced rank regression as in Johansen (1988). Although the statistical model (5) is defined for all $0 < b_0 \leq d_0$, the asymptotic properties of the ML estimator are derived in Johansen and Nielsen (2012) when the true values satisfy $0 \leq d_0 - b_0 < 1/2$ and $b_0 \neq 1/2$, for which

$\beta'_0 X_t$ is (asymptotically) a stationary process. Therefore, the following analysis is carried out for combinations of d_0 and b_0 , which satisfy such constraint. Unfortunately, imposing the constraint $d - b < 1/2$ in the estimation of ψ does not provide sufficient information to achieve identification since, by construction, $d_0 - b_0 = d_i - b_i \forall i = 1, \dots, m$ with $m = \lfloor \frac{k+1}{k_0+1} \rfloor$, by Proposition 1.iii).

Since the values of ψ that maximize $\ell_T(\psi)$ must be found numerically, the consequences of the lack of identification of the FCVAR $_{d,b}$ model on the expected profile log-likelihood when $k > k_0$ are explored by means of numerical experiments based on the MATLAB package of Nielsen and Popiel (2016). In particular, since the asymptotic value of $\ell_T(\psi)$ is not available in closed-form as a function of the FCVAR $_{d,b}$ parameters, the asymptotic values of $\ell_T(\psi)$ are approximated by averaging over M simulations the value of $\ell_T(\psi)$ computed for different values of ψ and a large T . This provides a precise numerical approximation of the expected profile log-likelihood, $E[\ell_T(\psi)]$. Therefore, $M = 100$ simulated paths are generated from model (5) with $T = 50,000$ observations and $p = 2$ processes. The fractional parameters of the system are $d_0 = 0.8$ and $b_0 = d_0$. The restriction $b_0 = d_0$ simplifies the readability of the results without loss of generality, since the plots display $\bar{\ell}_T(d) = \frac{1}{M} \sum_{i=1}^M \ell_{i,T}(d)$ as a function of d in a two dimensional Cartesian system. The cointegration vector is $\beta_0 = [1, -1]'$, the vector of adjustment coefficients is $\alpha_0 = [0.5, -0.5]'$, and the matrices Γ_i^0 , $i = 1, \dots, k_0$, for different values of k_0 are chosen such that the roots of the characteristic polynomial are outside the fractional circle, see Johansen (2008). The average profile log-likelihood, $\bar{\ell}_T(d)$, and the average of the function $f(d) = |\hat{\alpha}'_{\perp}(d)\hat{\Gamma}(d)\hat{\beta}_{\perp}(d)|$ are computed with respect to a grid of alternative values for $d = [d_{\min}, \dots, d_{\max}]$. The average of $f(d)$ over the M simulations is an estimate of the value of the $\mathcal{F}(d)$ condition for different values of d . Hence $\bar{\mathcal{F}}(d) = \frac{1}{M} \sum_{i=1}^M f_i(d)$ for $d = [d_{\min}, \dots, d_{\max}]$ is plotted together with $\bar{\ell}_T(d)$. Due to space constraints, the results of the simulations can not be shown for a large number of parameter combinations. However, the results obtained with other parameters confirm the reported evidence.

Figure 1 reports the values of $\bar{\ell}_T(d)$ and $\bar{\mathcal{F}}(d)$ when $k = 1$ lags are chosen but $k_0 = 0$. It clearly emerges that the two global maxima of $\bar{\ell}_T(d)$ are associated to the pair of values $d_0 = 0.8$ and $d_1 = 0.4$, but in $d_1 = 0.4$ the $\bar{\mathcal{F}}(d)$ line is equal to zero. Similarly, as reported in Figure B.1 in the supplementary material, the expected log-likelihood function has three humps around $d_0 = 0.8$, $d_1 = 0.4$ and $d_2 = d_0 - 2/3d_0 = 0.2667$ when $k = 2$ and $k_0 = 0$. As in the previous case, the $\bar{\mathcal{F}}(d)$ line is approximately equal to zero around d_1 and d_2 . Consistently with the theoretical results presented in Section 2, the $\bar{\mathcal{F}}(d)$ line is instead far from zero in $d_0 = 0.8$ also in this case. Figure 2 reports the contour plot of the expected profile log-likelihood function in the 2-dimensional space of $(d, b) \in \mathbb{R}^2$, with $d \geq b$. The plot clearly highlights the presence of two equivalent

Figure 1: The figure reports simulated values of $\bar{\ell}(d)$ and $\bar{\mathcal{F}}(d)$ for different values of $d \in [0.3, 0.9]$ on the x-axis. The observations from the DGP are generated with $k_0 = 0$ lags and model $\mathcal{H}_k$ with $k = 1$ lags is estimated. The parameters of the DGP are $d_0 = b_0 = 0.8$, $\beta_0 = [1, -1]'$, $\alpha_0 = [-0.5, 0.5]'$.

peaks located inside the isolines with level -14.1928 that, as expected, are associated with the vectors $\psi_0 = [0.8, 0.8]'$ and $\psi_1 = [0.4, 0.4]'$. Notably, the function $\bar{\ell}(\psi)$ quickly decreases at the extremes of the parameter space, i.e. when $d > d_0$ and $b > b_0$ or when $d < d_0 - b_0/2$ and $b < b_0/2$. Instead, the function remains rather high and flat in the interval $b_0/2 < b \leq d < d_0$. This may induce further identification problems in finite samples as discussed in Section 3.1.

A slightly more complex evidence arises when $k_0 > 0$. Figure 3 reports $\bar{\ell}_T(d)$ and $\bar{\mathcal{F}}(d)$ when $k_0 = 1$ while $k = 2$ is chosen. The $\bar{\ell}_T(d)$ function is globally maximized in the region around $d_0 = 0.8$, thus supporting the theoretical results outlined in Proposition 1, i.e. when $k = 2$ and $k_0 = 1$ there is no lack of identification. However, another interesting evidence emerges. The $\bar{\ell}_T(d)$ function is flat and high in the region around $d = 0.5$, possibly inducing identification problems in finite samples, see Section 3.1. When $k = 3$, we expect $m = \frac{4}{2} = 2$ equivalent sub-models associated with $d_0 = 0.8$ and $d_1 = d_0 - 1/2d_0 = 0.4$. Indeed, by looking at Figure A.2 in the supplementary material, it emerges that the $\bar{\ell}_T(d)$ line has two global maxima around the values $d_0 = 0.8$ and $d_1 = 0.4$. As expected, in the region around $d_1 = 0.4$ the $\bar{\mathcal{F}}(d)$ line is close to zero.

Figure 3: Figure reports simulated values of $\bar{l}(d)$ and $\bar{\mathcal{F}}(d)$ for different values of $d \in [0.4, 1]$ on the x-axis. The observations from the DGP are generated with $k_0 = 1$ lags and model $\mathcal{H}_k$ with $k = 2$ lags is estimated. The parameters of the DGP are $d_0 = b_0 = 0.8$, $\beta_0 = [1, -1]'$, $\alpha_0 = [-0.5, 0.5]'$, and $\Gamma_1 = \begin{bmatrix} 0.3 & -0.2 \\ 0.4 & -0.5 \end{bmatrix}$.

$\hat{\Gamma}_1$ and $\hat{\Gamma}_2$ are such that $|\alpha'_\perp \Gamma \beta_\perp| \approx 0$. On the other hand, with another simulated path, the global maximum is found around $d = 0.8$, where the function $f(d)$ is far from zero, Panel b). As it emerges from this example, for any choice of $k > k_0$ there is the risk of obtaining estimates of the fractional parameters, d and b , that are far from the true ones. Figures B.3-B.5 in the supplementary material show that this situation is very likely to happen given the clear bimodality of the distribution of the Monte Carlo estimates in a number of alternative setups.

Tschernig et al. (2013a) discuss an analogous identification problem in the FIVAR_b model. The FIVAR_b extends the FIVAR model allowing the autoregressive structure to depend on the fractional lag operator, L_b , hence inducing more flexibility in the short-run terms. Tschernig et al. (2013a) show that an identification problem arises when the eigenvalues of the characteristic polynomial in the L_b operator are either close to 0 or to 1. Similarly to the FCVAR_{d,b}, the lack of identification leads to an high and flat log-likelihood function for a wide range of combinations of d and b . However, in the FCVAR_{d,b} model, the $\mathcal{F}(d)$ condition provides a necessary and sufficient condition for the identification. It is therefore crucial to develop a robust estimation procedure that guarantees that the estimated FCVAR_{d,b} parameters are correctly identified and satisfy the $\mathcal{F}(d)$ condition also when the lag-length is potentially overspecified.

Figure 4: The figure reports the values of the profile log-likelihood $l(d)$ and $\mathcal{F}(d)$ for different values of $d \in [0.3, 1]$ (x-axis) for two different simulated paths with $T = 1,000$ of the $\text{FCVAR}_{d,d}$ when $k_0 = 0$ and model $\mathcal{H}_1$ is estimated. The parameters of the DGP are $d_0 = b_0 = 0.8$, $\beta_0 = [1, -1]'$, $\alpha_0 = [-0.5, 0.5]'$.

3.2 Constrained Likelihood

In the previous sections, we have proved that the $\text{FCVAR}_{d,b}$ model suffers from identification problems when k is over-specified. In particular, a number of equivalent parametrization associated to fractions of the true d_0 and b_0 can be found for several choices of $k > k_0$. On the other hand, the fractional parameter d is equivalent to the true fractional order of the process X_t only in $d = d_0$. As illustrated above, this identification problem has clear consequences from a statistical point of view since a unique ML estimator of d and b does not exist. We therefore propose a new approach that is based on the idea of transforming the unrestricted maximum likelihood problem, whose properties have been studied in Johansen and Nielsen (2012) only for the case $k = k_0$, into a constrained maximum likelihood problem by imposing a very mild restriction on the parameter space of d . In particular, we suggest that $\hat{d}$ and $\hat{b}$ must be the solutions of the following constrained maximum likelihood problem

$$\begin{aligned} \hat{\psi} &= \arg \max_{\psi} \ell_T(\psi), \\ \text{s.t. } d &\geq \delta_{\min} \end{aligned} \tag{10}$$

where $\ell_T(\psi)$ is defined in (8) and $\delta_{\min}$ determines the lower bound on the parameter d . Restricting the parameter space of d is supported by the following lemma, which is a direct derivation of Proposition 1.

Figure 5: The figure reports the values of the profile log-likelihood $l(d)$ and $\mathcal{F}(d)$ for different values of $d \in [0.3, 1]$ (x-axis) for two different simulated paths with $T = 1,000$ of the $\text{FCVAR}_{d,d}$ when $k_0 = 1$ and model $\mathcal{H}_2$ is estimated in the data. The parameters of the DGP are $d_0 = b_0 = 0.8$, $\beta_0 = [1, -1]'$, $\alpha_0 = [-0.5, 0.5]'$, and $\Gamma_1 = \begin{bmatrix} 0.3 & -0.2 \\ 0.4 & -0.5 \end{bmatrix}$.

Lemma 1 *Let $\tilde{\Theta}_{\mathcal{H}_k} = \{d = d_0, b \in (0, d_0], \alpha \in \mathbb{R}^{p \times r}, \beta \in \mathbb{R}^{p \times r}, \Gamma_j \in \mathbb{R}^{p \times p}, j = 1, \dots, k; \Omega > 0\}$ be the restricted parameter space of model $\Theta_{\mathcal{H}_k}$ with $d = d_0 \in \mathbb{R}^+$, then the statistical model $\mathcal{P} = \{P_\theta : \theta \in \tilde{\Theta}_{\mathcal{H}_k}\}$ is identified, i.e. $P_{\theta_1} = P_{\theta_2}$ implies $\theta_1 = \theta_2$ for all $\theta_1, \theta_2 \in \tilde{\Theta}_{\mathcal{H}_k}$, and $|\alpha'_\perp \Gamma \beta_\perp| \neq 0 \forall \theta \in \tilde{\Theta}_{\mathcal{H}_k}$.*

It follows from Lemma 1 that once the parameter d is fixed to d_0 , then all the $\text{FCVAR}_{d,b}$ parameters are uniquely identified for any lag $k > k_0$. Under the constraint $d = d_0$, the profile log-likelihood function $\ell_T(\psi)$ only varies with respect to b and it has a unique maximum around b_0 . Interestingly, Lemma 1 provides theoretical support to the procedure, adopted in Bollerslev et al. (2013) and Caporin et al. (2013), of estimating the $\text{FCVAR}_{d,b}$ model by restricting the fractional parameter d to a constant value and by maximizing the profile log-likelihood function with respect to b only. Figure 6 reports the value of the *sliced* profile log-likelihood with respect to different values of b , when the parameter d is fixed at the true value $d_0 = 1$. It clearly emerges that, irrespectively of the choice of $k > k_0$, the profile log-likelihood function is uniquely maximized around b_0 . This is a direct consequence of Lemma 1. Figure B.6 in the supplementary material confirms this result also when $b_0 < d_0$. As expected the value of the *sliced* profile log-likelihood at the optimum is the highest for the model with $k = 4$ lags in both figures, since the model $\mathcal{H}_4$ nests all the other models with $k < 4$. However, the profile log-likelihood function becomes very flat when k increases. This is due to the efficiency loss caused by the inclusion in the model $\mathcal{H}_k$ of matrices of parameters, $\Gamma_j, j > k_0$, that should be theoretically

Figure 6: The figure reports the values of the expected profile log-likelihood, $\bar{l}(\psi)$, for different values of $b \in [0.51, 1.2]$ (x-axis) when $d = d_0 = 1$. The sample size is $T = 20,000$ and $k_0 = 0$, while $\mathcal{H}_k$ with $k = 1, 2, 3, 4$ is estimated. The parameters of the DGP are $d_0 = b_0 = 1$, $\beta_0 = [1, -1]'$, $\alpha_0 = [-0.5, 0.5]'$.

excluded. In small samples, this may generate a problem of weak identification analogous to the one discussed in Section 3.1.

Since there exists a unique maximum of the profile log-likelihood function when d is restricted to d_0 , then the asymptotic properties found in Johansen and Nielsen (2012) still hold. However, since d_0 is unknown in practice, we rely on a constrained optimization method, which sets to zero the probability of selecting a maximum outside a given interval for the parameter d . This means that a lower bound for d , namely $\delta_{\min}$, must be determined such that the optimization of the profile log-likelihood is performed in a region which should only contain one maximum. In the following, we illustrate a simple and direct way to select $\delta_{\min}$ in a data-driven fashion. In principle, any semi-parametric estimator of the fractional order of the series, e.g. the exact local Whittle estimator of Shimotsu and Phillips (2005), could be adopted to determine the fractional order of the system and a value for $\delta_{\min}$ could be easily determined by setting a lower bound based on the point estimate. Unfortunately, a multivariate version of the exact local Whittle in which all the processes share the same degree of fractional integration is not yet available in the literature. Indeed, under the assumption of fractional cointegration the multivariate log-likelihood of the model can not be determined due to the singularity of the coherence matrix at the origin, see the discussion in Nielsen and Shimotsu (2007) among others. Similarly to Nielsen and Shimotsu (2007), we therefore recommend to obtain a

semi-parametric estimate of d as

$$\tilde{d} = \frac{1}{p} \sum_{i=1}^p \tilde{d}_i, \quad (11)$$

where $\tilde{d}_i$ is the univariate exact local Whittle estimate of the parameter d on the i -th series. The exact local Whittle is defined as

$$\tilde{d}_i = \arg \min_{d \in \mathbb{D}} R(d_i, X_{t,i}) \quad i = 1, \dots, p \quad (12)$$

with

$$R(d_i, X_{t,i}) = \frac{1}{m} \sum_{j=1}^m \log(\lambda_j^{-2d_i}) + \log \left(\frac{1}{m} \sum_{j=1}^m I_{\Delta^{d_i} X_{t,i}}(\lambda_j) \right), \quad (13)$$

where $I_{\Delta^{d_i} X_{t,i}}(\lambda_j)$ is the periodogram of the fractional difference of the series $X_{t,i}$ evaluated at the Fourier frequency λ_j , where the number of frequencies used in the estimation is q_T and $\mathbb{D}$ is the admissible set of values of d , which according to Shimotsu and Phillips (2005) has length no larger than $\frac{9}{2}$. Under Assumptions 1-5 of Shimotsu and Phillips (2005), $\tilde{d}_i$ is a consistent estimator of d and asymptotically Gaussian with

$$\sqrt{q_T}(\tilde{d}_i - d_0) \xrightarrow{d} N\left(0, \frac{1}{4}\right), \quad (14)$$

where the asymptotic variance does not depend on any nuisance parameter and the rate of convergence depends on q_T . Therefore, once $\tilde{d}$ is estimated, then $\delta_{\min}$ can be determined as

$$\delta_{\min} = \tilde{d} - c \cdot \tilde{d}, \quad (15)$$

where $c \in [0, 1]$. Given this choice for $\delta_{\min}$, it is possible to evaluate the probability that the restriction $d \geq \delta_{\min}$ is associated to an identified system. Under identification, the ML estimator of Johansen and Nielsen (2012) can consistently estimate d_0 and b_0 for any $k \geq k_0$. Let's first define the set $\mathbb{D}^* = [\delta_{\min}, +\infty)$. If $d_0 \in \mathbb{D}^*$ and $d_0 - b_0/2 \notin \mathbb{D}^*$, then all the parameters of the FCVAR $_{d,b}$ model can be identified for all choices of k . Since $\delta_{\min}$ depends on the ELW estimator, which has the asymptotic Gaussian distribution in (14), we can compute the probability of the event $\delta_{\min} \in (d_0 - b_0/2, d_0)$, denoted as $P_{d_0-b_0/2, d_0} := Pr(d_0 - b_0/2 < \delta_{\min} < d_0)$, for $T \rightarrow \infty$ as

$$\lim_{T \rightarrow \infty} P_{d_0-b_0/2, d_0} = Pr\left(Z \leq \frac{2c}{1-c} d_0 \sqrt{q_T p}\right) - Pr\left(Z \leq \frac{2cd_0 - b_0}{1-c} \sqrt{q_T p}\right), \quad (16)$$

where Z follows a standard normal distribution. Note that the first probability in (16) tends to 1 as $T \rightarrow \infty$, while the second one tends to 0 only if $\frac{2cd_0 - b_0}{1-c} < 0$, that is if

$b_0 > 2cd_0$. This is illustrated in Figure 7, which displays the probability $P_{d_0-b_0/2, d_0}$ for different combinations of T , b_0 and c when $q_T = T^{0.5}$ and $d_0 = 0.8$. First, it emerges that when b_0 is very small, the probability of $\delta_{\min}$ being in the interval $(d_0 - b_0/2, d_0)$ approaches zero, since the size of the interval $(d_0 - b_0/2, d_0)$ shrinks to zero. It should be noted that the case $b_0 \approx 0$ is not empirically relevant. Indeed, when $b_0 \approx 0$ the strength of the cointegration relation is minimal and $L_{b_0} \approx 1$, thus excluding the practical implementation of the FCVAR $_{d,b}$. Moreover, Johansen and Nielsen (2012) have proved the asymptotic distribution of the ML estimator for values of $b_0 > d_0 - 1/2$, thus ruling out a number of cases for low values of b_0 . The probability $P_{d_0-b_0/2, d_0}$ increases with

Figure 7: The figure reports the probability that $\delta_{\min}$ is in the interval $(d_0 - b_0/2, d_0)$ for different values of $b_0 \in (0.3, 0.8)$ and $c \in [0.01, 0.3]$, with $q_T = T^{0.5}$ and $d_0 = 0.8$.

the strength of the cointegration relation, b_0 , and with T . For large values of b_0 , the probability of $\delta_{\min}$ being in the interval $(d_0 - b_0/2, d_0)$ approaches one. When $T = 50,000$, the probability $P_{d_0-b_0/2, d_0}$ is close to 1 already for intermediate values of b_0 and for almost all choices of c . In particular, choosing a small c in the range between 5% and 15%, that is setting the lower bound close to $\tilde{d}$, leads to the highest probability since the condition $b_0 > 2cd_0$ is more easily verified. On the other hand, in small samples, e.g. $T = 1,000$, the probability of $d_0 < \delta_{\min}$ might not be zero if c is too small, thus leading to a reduction in the probability $P_{d_0-b_0/2, d_0}$. This clearly defines a trade-off between asymptotic precision and finite sample robustness in the choice of c , which might be selected based on some knowledge or assumption on the DGP. Figure 8 shows that the probability of $P_{d_0-b_0/2, d_0}$ tends to 1 as $T \rightarrow \infty$ for different combinations of d_0 and b_0 when $c = 0.15$. In particular, the probability is already close to 85% for the case $d_0 = 0.6$ and $b_0 = 0.4$ when $T = 500$ and it is almost equal to 1 when $T = 5,000$. We therefore recommend the choice $c = 0.15$

in the empirical applications and we will adopt this value in the rest of the paper. In the

Figure 8: The figure reports the probability that $\delta_{\min}$ is in the interval $(d_0 - b_0/2, d_0)$ for different values combinations of d_0 and b_0 and for different sample sizes. The probability curves are related to a choice of $c = 0.15$. The x-axis reports the sample size T and the y-axis the probability $P_{d_0 - b_0/2, d_0}$.

next paragraph, we show that imposing the lower bound constraint in (10) is generally sufficient to solve the identification problem with a very mild restriction on the parameter space.

3.2.1 Monte Carlo Simulations

In this paragraph, we discuss the results of a number of Monte Carlo simulations to support the need for the approach based on the constrained log-likelihood outlined in (10) as opposed to the unconstrained one when the number of lags is unknown. Figure 9 reports the contour plot of the Monte Carlo estimates of the parameters d and b when a sample of $T = 2,500$ observations is generated by the following bivariate FCVAR $_{d,b}$ model

$$\Delta^{d_0} X_t = \alpha_0 \beta_0' \Delta^{d_0 - b_0} L_{b_0} X_t + \varepsilon_t \quad t = 1, \dots, T, \quad (17)$$

where $d_0 = 1$ and $b_0 = 0.8$. For each generated sample, the model $\mathcal{H}_2$ is estimated on the data. According to Proposition 1, three equivalent models can be found associated to different combinations of d and b , i.e. $\psi_0 = [1, 0.8]$, $\psi_1 = [0.6, 0.4]$ and $\psi_2 = [0.47, 0.27]$. From Panel a) of Figure 9 it clearly emerges that maximizing the constrained log-likelihood function (10) solves the identification problem discussed above.

Indeed, almost the entire probability mass of $\hat{\psi}$, based on $M = 1,000$ Monte Carlo estimates, is concentrated around ψ_0 . Only in a very limited number of cases the estimates are located around $[0.8, 0.5]$, and this could be attributed to the probability $P_{d_0-b_0/2, d_0}$ being slightly smaller than 1 in this setup. Instead, when the optimal parameters d and b are found by maximizing the unrestricted likelihood function, see Panel b), a large portion of the probability mass is located away from $\psi_0 = [1, 0.8]$. In particular, when the profile log-likelihood function is not constrained, the bivariate distribution of $\hat{\psi}$ is clearly multi-modal, as a consequence of the lack of identification as outlined in Proposition 1. For comparison, Figure 10 reports the distribution of $\hat{\psi}$ when the number of lags is

Figure 9: The figure reports the contour plot of $M = 1,000$ Monte Carlo estimates of the parameters d (x-axis) and b (y-axis) when a sample of $T = 2,500$ observations is generated by a FCVAR $_{d,b}$ model with $k_0 = 0$, $d_0 = 1$, $b_0 = 0.8$ and the cointegration vectors given by $\beta_0 = [1, -1]'$ and $\alpha_0 = [-0.5, 0.5]'$. Model $\mathcal{H}_2$ is estimated on the data. Panel a) is relative to the estimates based on the constrained log-likelihood (10) where $c = 0.15$ and $q_T = T^{0.5}$. Panel b) reports the contour plot for the unrestricted estimates.

correctly specified, i.e. $k = 0$. Not surprisingly, the distribution of $\hat{\psi}$ is well centered around ψ_0 , and the estimates are more efficient than those obtained with $k > 0$ since fewer FCVAR $_{d,b}$ parameters must be estimated under correct lag specification. However, k_0 is unknown in practice and is typically determined by a general-to-specific sequence of LR tests. In Section 4.1 we discuss the nesting structure of the FCVAR $_{d,b}$ model under unknown cointegration rank and lag-length and the optimal sequence of LR tests when the parameter space of d is properly restricted.

Figures B.7-B.11 in the supplementary material highlight the robustness of the constrained likelihood approach for different sample sizes and different combinations of k_0 and k . As T increases, the estimates based on the unconstrained likelihood still display

(a) Correct Number of Lags

Figure 10: The figure reports the contour plot of $M = 1,000$ Monte Carlo estimates of the parameters d (x-axis) and b (y-axis) when a sample of $T = 2,500$ observations is generated by a FCVAR $_{d,b}$ model with $k_0 = 0$, $d_0 = 1$, $b_0 = 0.8$ and the cointegration vectors given by $\beta_0 = [1, -1]'$ and $\alpha_0 = [-0.5, 0.5]'$. Model $\mathcal{H}_0$ is estimated on the data.

the identification problem, while the constrained estimates are all centered around d_0 and b_0 , see Figure B.7 which is based on $T = 10,000$. This confirms that the probability $P_{d_0-b_0/2, d_0}$ converges to 1 as $T \rightarrow \infty$ as illustrated in Figure 8. When $T = 1,000$, most of the probability mass is still concentrated around d_0 and b_0 , although the density is much more dispersed than the case with $T = 10,000$, see Figures B.8 and B.9. This is not surprising given the flatness of the log-likelihood function in small samples. Finally, the results do not qualitatively change when data is generated under $\mathcal{H}_1$ with $\Gamma_1 = \begin{bmatrix} 0.3 & -0.2 \\ -0.4 & 0.5 \end{bmatrix}$ and model $\mathcal{H}_3$ is estimated, see Figure B.10. As expected, the estimates based on the unconstrained likelihood are bimodal, since two equivalent sub-models of $\mathcal{H}_3$ can be found associated to $k_0 = 1$, see Table 1. Instead, the estimates based on the constrained likelihood are unimodal and centered around the true values d_0 and b_0 . Finally, when $d_0 - b_0 \approx 0.5$, see Figure B.11, the slow convergence rate makes the profile log-likelihood function extremely flat, although the sample size is moderately large, thus leading to rather dispersed estimates of ψ . However, compared to the unrestricted estimates which are found everywhere in the interval $0.3 < b < d < 1$, the constrained estimates are much more often located in the region around d_0 and b_0 .

4 Unknown Cointegration Rank

We now extend the previous results to allow both the cointegration rank and the lag-length to be unknown. This is the relevant case in the empirical applications, when testing for the presence of a cointegration relationship between two (or more) fractional processes but there is no preliminary information on the optimal choice of k . The unrestricted FCVAR $_{d,b}$ model is formulated as:

$$\mathcal{H}_{r,k}: \quad \Delta^d X_t = \Pi \Delta^{d-b} L_b X_t + \sum_{i=1}^k \Gamma_i \Delta^d L_b^i X_t + \varepsilon_t, \quad (18)$$

where $0 \leq r \leq p$ is the rank of the $p \times p$ matrix Π . The parameter space of model $\mathcal{H}_{r,k}$ is

$$\Theta_{\mathcal{H}_{r,k}} = \{\alpha \in \mathbb{R}^{p \times r}, \beta \in \mathbb{R}^{p \times r}, \Gamma_j \in \mathbb{R}^{p \times p}, j = 1, \dots, k, d \in \mathbb{R}^+, b \in \mathbb{R}^+, d \geq b > 0, \Omega > 0\}.$$

Compared to the parameter space of $\mathcal{H}_k$ in Section 2, the set $\Theta_{\mathcal{H}_{r,k}}$ also contains the cointegration rank, r , among the unknown parameters. Model $\mathcal{H}_{r,k}$ exhibits further identification issues than those illustrated in Section 2.

Example 2: Consider the model with $k = 1$ lags and rank $0 \leq r \leq p$, given by

$$\mathcal{H}_{r,1}: \quad \Delta^d X_t = \Pi \Delta^{d-b} L_b X_t + \Gamma_1 \Delta^d L_b X_t + \varepsilon_t,$$

where the set of parameters is $\theta = \text{vec}(d, b, \Pi, \Gamma_1)$. Examine now the following two sub-models of $\mathcal{H}_{r,1}$. First, model $\mathcal{H}_{p,0}$ is

$$\mathcal{H}_{p,0}: \quad \Delta^{\tilde{d}} X_t = \tilde{\Pi} \Delta^{\tilde{d}-\tilde{b}} L_{\tilde{b}} X_t + \varepsilon_t,$$

with $\tilde{\theta} = \text{vec}(\tilde{d}, \tilde{b}, \tilde{\Pi})$ is the set of parameters. Second, model $\mathcal{H}_{0,1}$ is

$$\mathcal{H}_{0,1}: \quad \Delta^{d^*} X_t = \Gamma_1^* \Delta^{d^*} L_{b^*} X_t + \varepsilon_t,$$

where $\theta^* = \text{vec}(d^*, b^*, \Gamma_1^*)$ is the set of parameters. Both $\mathcal{H}_{p,0}$ and $\mathcal{H}_{0,1}$ can be written as

$$\left[\Delta^{\tilde{d}-\tilde{b}}(-\tilde{\Pi}) + \Delta^{\tilde{d}}(I_p + \tilde{\Pi}) \right] X_t = \varepsilon_t, \quad (19)$$

and

$$\left[\Delta^{d^*}(I - \Gamma_1^*) + \Delta^{d^*+b^*}(\Gamma_1^*) \right] X_t = \varepsilon_t. \quad (20)$$

Imposing the restrictions $\tilde{d} = d^* + b^*$, $\tilde{b} = b^*$ and $-\tilde{\Pi} = I_p - \Gamma_1^*$ on model $\mathcal{H}_{p,0}$ in (19)

leads to $\mathcal{H}_{p,0}$ and $\mathcal{H}_{0,1}$ being equivalent. Indeed, the probability densities are

$$p_{\mathcal{H}_{p,0}}(X_1, \dots, X_T; \tilde{\theta} | X_0, X_{-1}, \dots) = p_{\mathcal{H}_{0,1}}(X_1, \dots, X_T; \theta^* | X_0, X_{-1}, \dots), \quad (21)$$

when $\tilde{\theta} = \text{vec}(d^* + b^*, b^*, \Gamma_1^* - I_p, 0)$ and $\theta^* = \text{vec}(d^*, b^*, 0, \Gamma_1^*)$.

However, the sub-model $\mathcal{H}_{0,1}$ is not always a re-parametrization of $\mathcal{H}_{p,0}$. Indeed, by applying the restrictions $d^* = \tilde{d} - \tilde{b}$, $b^* = \tilde{b}$ and $\Gamma_1^* = I_p + \tilde{\Pi}$ on model $\mathcal{H}_{0,1}$ in (20), it follows that

$$p_{\mathcal{H}_{p,0}}(X_1, \dots, X_T; \tilde{\theta} | X_0, X_{-1}, \dots) = p_{\mathcal{H}_{0,1}}(X_1, \dots, X_T; \theta^* | X_0, X_{-1}, \dots), \quad (22)$$

where $\tilde{\theta} = \text{vec}(\tilde{d}, \tilde{b}, \tilde{\Pi}, 0)$ and $\theta^* = \text{vec}(\tilde{d} - \tilde{b}, \tilde{b}, 0, I_p + \tilde{\Pi})$. However, the equality (22) holds if and only if $\tilde{d} - \tilde{b} \geq \tilde{b} > 0$, i.e. $\tilde{d} \geq 2\tilde{b}$. This implies that $\mathcal{H}_{0,1} = \mathcal{H}_{p,0} \cap \{\tilde{d} \geq 2\tilde{b}\}$. Hence, $\mathcal{H}_{0,1} \subseteq \mathcal{H}_{p,0}$. The next proposition extends this example for any combination of k and r .

Proposition 2 *Consider an unrestricted FCVAR_{d,b} model*

$$\mathcal{H}_{r,k} : \quad \Delta^d X_t = \Pi \Delta^{d-b} L_b X_t + \sum_{j=1}^k \Gamma_j \Delta^{d-b} L_b X_t + \varepsilon_t, \quad (23)$$

where $0 \leq r \leq p$ is the rank of the matrix Π and k is the number of lags. Consider the following sub-models of $\mathcal{H}_{r,k}$: $\mathcal{H}_{p,k-1}$ with parameter set $\tilde{\theta} = \text{vec}(\tilde{d}, \tilde{b}, \tilde{\Pi}, \tilde{\Gamma}_1, \dots, \tilde{\Gamma}_{k-1}, \tilde{\Omega})$, and $\mathcal{H}_{0,k}$ with parameter set $\theta^* = \text{vec}(d^*, b^*, \Gamma_1^*, \dots, \Gamma_k^*, \Omega^*)$.

i) For any $k > 0$, model $\mathcal{H}_{0,k}$ is equivalent to $\mathcal{H}_{p,k-1}$ if the condition $\tilde{d} \geq 2\tilde{b}$ imposed on model $\mathcal{H}_{p,k-1}$ is satisfied. Hence $\mathcal{H}_{0,k} = \mathcal{H}_{p,k-1} \cap \{\tilde{d} \geq 2\tilde{b}\}$.

ii) The nesting structure of the FCVAR_{d,b} model is represented by the following scheme:

$$\begin{array}{ccccccccc} \mathcal{H}_{0,0} & \subset & \mathcal{H}_{0,1} & \subset & \mathcal{H}_{0,2} & \subset & \cdots & \subset & \mathcal{H}_{0,k} \\ \cap & & \cap & & \cap & & & & \cap & \mathcal{H}_{0,1} \subseteq \mathcal{H}_{p,0} \\ \mathcal{H}_{1,0} & \subset & \mathcal{H}_{1,1} & \subset & \mathcal{H}_{1,2} & \subset & \cdots & \subset & \mathcal{H}_{1,k} & \mathcal{H}_{0,2} \subseteq \mathcal{H}_{p,1} \\ \cap & & \cap & & \cap & & & & \cap & \vdots \\ \vdots & & \vdots & & \vdots & & \ddots & & \vdots & \vdots \\ \cap & & \cap & & \cap & & & & \cap & \mathcal{H}_{0,k} \subseteq \mathcal{H}_{p,k-1} \\ \mathcal{H}_{p,0} & \subset & \mathcal{H}_{p,1} & \subset & \mathcal{H}_{p,2} & \subset & \cdots & \subset & \mathcal{H}_{p,k} \end{array} \quad \text{with}$$

It follows from Proposition 2i) that model $\mathcal{H}_{0,k}$ can always be re-parametrized as model $\mathcal{H}_{p,k-1}$. On the other hand, model $\mathcal{H}_{p,k-1}$ can be formulated as $\mathcal{H}_{0,k}$ only when

the condition $\tilde{d} \geq 2\tilde{b}$ on model $\mathcal{H}_{p,k-1}$ holds. This leads to the peculiar nesting structure displayed in Proposition 2.ii). Notably the interpretation of the two models, $\mathcal{H}_{p,k-1}$ and $\mathcal{H}_{0,k}$, is different, although they are equivalent descriptions of the data. In model $\mathcal{H}_{p,k-1}$, the process X_t has p non-common stochastic trends fractional order $\tilde{d} - \tilde{b}$. Instead, in model $\mathcal{H}_{0,k}$, the process X_t has p non-common stochastic trends with fractional order d^* . A similar identification problem, due to indeterminacy between d , b and k , arises also in the univariate FAR(k) model studied in Johansen and Nielsen (2010)

$$\Delta^d Y_t = \pi \Delta^{d-b} L_b Y_t + \sum_{i=1}^k \gamma_i \Delta^d L_b^i Y_t + \varepsilon_t,$$

where Y_t is a univariate process and π is a scalar. Following the same procedure of the proof of Proposition 2, it follows that $\mathcal{M}_{0,k} = \mathcal{M}_{1,k-1} \cap \left\{ \tilde{d} \geq 2\tilde{b} \right\}$, where $\mathcal{M}_{0,k}$ defines the FAR model with $\pi = 0$ and k lags, while $\mathcal{M}_{1,k-1}$ defines the FAR model with $\pi \neq 0$ and $k-1$ lags. Therefore, the FAR(k) model has a similar circular nesting structure as in Proposition 23.ii). In Johansen and Nielsen (2010), the theoretical results are indeed obtained under the maintained assumption that the true number of lags k_0 is known.

The following Corollary shows that indeterminacy between cointegration rank and lag-length is not limited to $\mathcal{H}_{p,k-1}$ and $\mathcal{H}_{0,k}$, but it can be extended to any cointegration rank $0 < s < p$.

Corollary 1 *For any $k > 0$, model $\mathcal{H}_{s,k-1}$ with $0 < s < p$ and $\tilde{d} \geq 2\tilde{b}$ is equivalent to $\mathcal{H}_{0,k}$, if and only if the matrix $\Gamma^* = I_p - \sum_{j=1}^k \Gamma_j^*$ in model $\mathcal{H}_{0,k}$ has rank equal to s .*

In other words, if the matrix $\Gamma^* = I_p - \sum_{j=1}^k \Gamma_j^*$ in $\mathcal{H}_{0,k}$ has reduced rank of order $0 < s < p$, the models $\mathcal{H}_{s,k-1}$ and $\mathcal{H}_{0,k}$ are equivalent under $\tilde{d} \geq 2\tilde{b}$ in $\mathcal{H}_{s,k-1}$. This means that $\mathcal{H}_{0,k} \subseteq \mathcal{H}_{s,k-1}$ for any $0 < s \leq p$, if $\text{rank}(\Gamma) = s$.

4.1 Model Selection under Unknown Rank and Lag-length

The peculiar nesting structure of the FCVAR $_{d,b}$ obviously impacts the joint selection of the number of lags and the cointegration rank. Indeed, the likelihood ratio statistic for cointegration rank r , denoted as $LR_{r,k} := -2 \log LR(\mathcal{H}_{r,k} | \mathcal{H}_{p,k})$, see Johansen and Nielsen (2012, p.2698), is given by

$$-2 \log LR(\mathcal{H}_{r,k} | \mathcal{H}_{p,k}) = T(\ell_T^{(r,k)}(\hat{d}_{r,k}, \hat{b}_{r,k}) - \ell_T^{(p,k)}(\hat{d}_{p,k}, \hat{b}_{p,k})), \quad (24)$$

where $\ell_T^{(r,k)}$ is the profile log-likelihood of the FCVAR $_{d,b}$ model with rank r and k lags. Analogously, $\hat{d}_{r,k}$ and $\hat{b}_{r,k}$ are the arguments that maximize $\ell_T^{(r,k)}$. The asymptotic properties of the $LR_{r,k}$ test, under the maintained assumption of correct specification of the

lag-length, i.e. $k = k_0$, are provided in Johansen and Nielsen (2012). Unfortunately, the values of the profile log-likelihoods $\ell_T^{(0,k)}(\hat{d}_{0,k}, \hat{b}_{0,k})$ and $\ell_T^{(p,k-1)}(\hat{d}_{p,k-1}, \hat{b}_{p,k-1})$ are equal when $\tilde{d} \geq 2\tilde{b}$ in model $\mathcal{H}_{p,k-1}$, and the number of parameters of model $\mathcal{H}_{p,k-1}$ is the same as in $\mathcal{H}_{0,k}$. Hence, the equality of $\ell_T^{(0,k)}(\hat{d}_{0,k}, \hat{b}_{0,k})$ and $\ell_T^{(p,k-1)}(\hat{d}_{p,k-1}, \hat{b}_{p,k-1})$ influences the general-to-specific sequence of tests for the joint selection of the cointegration rank and the lag-length. Indeed, assuming that the general-to-specific procedure for the optimal lag selection terminates in $\mathcal{H}_{p,k-1}$, then it would be impossible to know whether the optimal model is $\mathcal{H}_{p,k-1}$ or $\mathcal{H}_{0,k}$ if the estimates $\hat{d}_{p,k-1}$ and $\hat{b}_{p,k-1}$ are such that $\hat{d}_{p,k-1} \geq 2\hat{b}_{p,k-1}$.

Therefore, a problem of joint selection of k and r arises in the FCVAR $_{d,b}$ when the cointegration rank is unknown and potentially equal to 0 or p . Moreover, under $\mathcal{H}_{0,k}$ with $k > 0$, the parameter b is defined but it does not have the usual interpretation as *cointegration gap*, since it only appears in the generalized lag operator. A test for the null hypothesis that $r = 0$ has been proposed by Lasak (2010) and extended in Lasak and Velasco (2015) to allow for multiple degrees of fractional cointegration. Alternatively, a solution to the indeterminacy in the FCVAR $_{d,b}$ framework is to rely on a preliminary estimate of the cointegration rank based on a frequency domain procedure, following for example the testing procedure of Nielsen and Shimotsu (2007). Instead, we show that it is sufficient to impose a constraint on the fractional parameter d to solve the problem of indeterminacy of cointegration rank and lag-length and develop a consistent testing procedure for model selection.

4.1.1 Model Selection with an Identification Restriction

Unfortunately, a solution to the joint indeterminacy of cointegration rank and lag-length is not available within the unrestricted FCVAR $_{d,b}$ framework. However, a simple solution to the identification problem caused by the indeterminacy of cointegration rank and lag-length can be achieved by a suitable restriction of the parameter space of d . Consider the model with unknown rank and unknown lag structure. The model can be expressed by the parameter set $\Theta_{r,k} = \{d_0 \in \mathbb{R}^+, b \in (0, d_0], \Gamma_j \in \mathbb{R}^{p \times p}, j = 1, \dots, k, \alpha \in \mathbb{R}^{p \times r}, \beta \in \mathbb{R}^{p \times r}, \Omega > 0\}$ where $0 \leq r \leq p$ and $k \geq 0$ are unknown. The following lemma holds.

Lemma 2 *Let $\tilde{\Theta}_{\mathcal{H}_{r,k}} = \{d = d_0, b \in [0, d_0], \alpha \in \mathbb{R}^{p \times r}, \beta \in \mathbb{R}^{p \times r}, \Gamma_j \in \mathbb{R}^{p \times p}, j = 1, \dots, k; \Omega > 0\}$ be the restricted parameter space of model $\Theta_{\mathcal{H}_{r,k}}$ with $d = d_0 \in \mathbb{R}^+$ for $0 \leq r \leq p$ and $k \geq 0$, then the nesting structure for the statistical models $\mathcal{P} = \{P_\theta :$*

$\theta \in \Theta_{r,k} \}_{k=0,1,\dots}^{r=0,\dots,p}$ can be written as

$$\begin{array}{ccccccc}
\mathcal{H}_{0,0} & \subset & \mathcal{H}_{0,1} & \subset & \cdots & \subset & \mathcal{H}_{0,k} \\
\cap & & \cap & & & & \cap \\
\mathcal{H}_{1,0} & \subset & \mathcal{H}_{1,1} & \subset & \cdots & \subset & \mathcal{H}_{1,k} \\
\vdots & & \vdots & & & & \vdots \\
\cap & & \cap & & & & \cap \\
\mathcal{H}_{p,0} & \subset & \mathcal{H}_{p,1} & \subset & \cdots & \subset & \mathcal{H}_{p,k}
\end{array}$$

When $d = d_0$ is fixed, Lemma 2 proves that the $\text{FCVAR}_{d,b}$ has a nesting structure that does not exhibit the problem outlined in Proposition 2, since $\mathcal{H}_{p,k-1}$ and $\mathcal{H}_{0,k}$ are two distinct models. Therefore, under the assumption that $d = d_0$, one could develop a general-to-specific sequence of LR tests which consists of iterating the tests $LR_{p,k-1} := -2 \log LR(\mathcal{H}_{p,k-1} | \mathcal{H}_{p,k})$ over k with fixed p (full rank) until the null hypothesis is rejected in k^* . Under the restriction $d = d_0$, the distribution of these tests is a classical $\chi^2(p^2)$, since, under full rank, the test statistic reduces to the one in Nielsen (2006, eq. 3.1). Therefore, the results in Nielsen (2006) can also be applied in the fractional context and the likelihood ratio test for $\Gamma_k = 0$ is asymptotically $\chi^2(p^2)$ even if a fractional unit root is present in the DGP. Therefore, the asymptotic power of each $LR_{p,k-1}$ test of rejecting an underspecified model is maximal, although the true cointegration rank is smaller than p , and the asymptotic size of each test is α . Once the lag is selected, then the sequence of cointegration rank tests $LR_{r,k^*} = -2 \log LR(\mathcal{H}_{r,k^*} | \mathcal{H}_{p,k^*})$ can be performed over $r \in [0, p-1]$ with k fixed to k^* . Following the results of Johansen and Nielsen (2012), the asymptotic distribution of LR_{r,k^*} is a functional of the type II fractional Brownian motion for $r = 0, \dots, p-1$ if $b_{r,k^*} > 0.5$ and it is a $\chi^2(q^2)$ if $b_{r,k^*} < 0.5$ where $q = p-r$. The asymptotic distribution for the case $b_{r,k^*} > 0.5$ is tabulated in MacKinnon and Nielsen (2014).

In practice, since d_0 is unknown, we suggest to set a mild constraint on the parameter set of d to enforce identification at each step in the selection procedure. Analogously to the discussion in Section 3.2, the estimates of $d_{r,k}$ and $b_{r,k}$, for any $0 \leq r \leq p$ and $k \geq 0$, must be the solutions of the following constrained maximum likelihood problem

$$\begin{aligned}
\hat{\psi}_{r,k} &= \arg \max_{\psi} \ell_T^{(r,k)}(\psi_{r,k}), \\
\text{s.t. } d_{r,k} &\geq \delta_{\min}
\end{aligned} \tag{25}$$

where the lower bound on the parameter $d_{r,k}$, $\delta_{\min}$, can be determined by a preliminary estimate of the fractional order of the process as in (15). Therefore, under the constraint

$d_{r,k} \geq \delta_{\min}$, we can test $\mathcal{H}_{p,k}$ against $\mathcal{H}_{p,k-1}$, without the risk of having an equivalent parametrization in $\mathcal{H}_{0,k}$ under the null hypothesis. It is important to stress that $\delta_{\min}$ does not depend on r and k so that it can be determined *a priori*. If the asymptotic probability $\lim_{T \rightarrow \infty} P_{d_0-b_0/2, d_0} = 1$, then the standard asymptotic results discussed above are unchanged since $\lim_{T \rightarrow \infty} P(LR_{p,k-1} > C_{p^2}(a) | d_0 - b_0/2 < \delta_{\min} < d_0) = P(LR_{p,k-1} > C_{p^2}(a))$ where $C_{p^2}(a)$ is the critical value of a $\chi^2(p^2)$ distribution at the a significance level. Otherwise, if $\lim_{T \rightarrow \infty} P_{d_0-b_0/2, d_0} < 1$, we could experience asymptotic loss of power and/or size distortions in the sequence of LR tests.

$\mathcal{H}_{0,0}$	$\mathcal{H}_{0,1}$	$\mathcal{H}_{0,2}$	$\mathcal{H}_{0,3}$	$\mathcal{H}_{0,4}$	$\mathcal{H}_{0,0}$	$\mathcal{H}_{0,1}$	$\mathcal{H}_{0,2}$	$\mathcal{H}_{0,3}$	$\mathcal{H}_{0,4}$
0	0	0	0	0	0	0	0	0	0
$\mathcal{H}_{1,0}^*$	$\mathcal{H}_{1,1}$	$\mathcal{H}_{1,2}$	$\mathcal{H}_{1,3}$	$\mathcal{H}_{1,4}$	$\mathcal{H}_{1,0}$	$\mathcal{H}_{1,1}^*$	$\mathcal{H}_{1,2}$	$\mathcal{H}_{1,3}$	$\mathcal{H}_{1,4}$
92.5	1.6	1.3	1.5	1.3	0	93.8	1.2	1.9	1.6
$\mathcal{H}_{2,0}$	$\mathcal{H}_{2,1}$	$\mathcal{H}_{2,2}$	$\mathcal{H}_{2,3}$	$\mathcal{H}_{2,4}$	$\mathcal{H}_{2,0}$	$\mathcal{H}_{2,1}$	$\mathcal{H}_{2,2}$	$\mathcal{H}_{2,3}$	$\mathcal{H}_{2,4}$
1.4	0	0.1	0.1	0.3	0	1.4	0	0.1	0

Table 2: The two tables report the percentage of selected models when the DGP is a FCVAR $_{d,b}$ model with $k_0 = 0$ (left) and $k_0 = 1$ (right) lags, $r_0 = 1$ cointegration rank and fractional parameters $d_0 = b_0 = 0.8$. The results are based on $M = 1,000$ Monte Carlo simulations and $T = 2,500$ observations. The sequence of LR test is performed with a theoretical significance level of $a = 1\%$ for both the selection of the cointegration rank and the optimal number of lags. The estimations of d and b for each model are carried out under the restriction $d > \delta_{\min}$, with $\delta_{\min} = \tilde{d} - c \cdot \tilde{d}$, with $c = 0.15$.

The results reported in Table 2 confirm the reliability of the constrained sequential selection procedure that selects the true model in almost 93% of cases when $k_0 = 0$ and 94% when $k_0 = 1$. Interestingly, a model smaller than the true one is never selected suggesting a power close to 100% for both the lag and rank tests. Concerning the empirical size of the tests for the optimal number of lags, it emerges that the probability of selecting each of the models with $k > k_0$ is close to the theoretical significance level that is set to 1%, thus signaling a good finite-sample approximation of the asymptotic $\chi^2(p^2)$ distribution. Moreover, the probability associated to the selection of model $\mathcal{H}_{p,k_0}$, when the true model is $\mathcal{H}_{1,k_0}$, is 1.4% in both cases, that is again close to the theoretical significance level of $a = 1\%$. The results for different setups for the DGP are reported in the supplementary material and they all confirm this evidence. Notably, the contour plots in Figure 11 show that the distribution of $\hat{d}_{r^*,k^*}$ and $\hat{b}_{r^*,k^*}$ is unimodal and centered around the true values in all cases. Finally, Table 3 compares the selection frequencies based on the constrained log-likelihood approach (left panel) and those obtained when estimating unrestrictedly (right panel). Again, imposing the constraint $d > \delta_{\min}$ leads to the a selection frequency of

the true model that is in line with the size of each individual sets (again we set $a = 1\%$). Instead, the true model (with $r_0 = 0$ and $k_0 = 1$) is never selected when relying on the unrestricted estimator, but, in line with Theorem 2, the model with full rank and $k_0 - 1 = 0$ lags is selected most of the times.

Figure 11: The figure reports the contour plot of $M = 1,000$ Monte Carlo estimates of the parameters d (x-axis) and b (y-axis) when a sample of $T = 2,500$ observations is generated by a $\text{FCVAR}_{d,b}$ model with $k_0 = 0$ (left) and $k_0 = 1$ (right), $d_0 = 1$, $b_0 = 0.8$ and the cointegration vectors given by $\beta_0 = [1, -1]'$ and $\alpha_0 = [-0.5, 0.5]'$. The estimates are associated to the optimal selected model, $\mathcal{H}_{r^*, k^*}$.

$\mathcal{H}_{0,0}$	$\mathcal{H}_{0,1}^*$	$\mathcal{H}_{0,2}$	$\mathcal{H}_{0,3}$	$\mathcal{H}_{0,4}$	$\mathcal{H}_{0,0}$	$\mathcal{H}_{0,1}^*$	$\mathcal{H}_{0,2}$	$\mathcal{H}_{0,3}$	$\mathcal{H}_{0,4}$
0	91.9	1.1	1.8	0.8	0	0	0.7	0.9	1.9
$\mathcal{H}_{1,0}$	$\mathcal{H}_{1,1}$	$\mathcal{H}_{1,2}$	$\mathcal{H}_{1,3}$	$\mathcal{H}_{1,4}$	$\mathcal{H}_{1,0}$	$\mathcal{H}_{1,1}$	$\mathcal{H}_{1,2}$	$\mathcal{H}_{1,3}$	$\mathcal{H}_{1,4}$
0	0.3	0.2	0	0.2	23	1.1	1.6	2.0	1.8
$\mathcal{H}_{2,0}$	$\mathcal{H}_{2,1}$	$\mathcal{H}_{2,2}$	$\mathcal{H}_{2,3}$	$\mathcal{H}_{2,4}$	$\mathcal{H}_{2,0}$	$\mathcal{H}_{2,1}$	$\mathcal{H}_{2,2}$	$\mathcal{H}_{2,3}$	$\mathcal{H}_{2,4}$
0	0.2	0.1	0	0	62.9	0.4	1.0	1.4	1.3

Table 3: The two tables report the percentage of selected models when the DGP is a FCVAR_{d,b_0} model with $k_0 = 1$ lags, $r_0 = 0$ cointegrating rank and fractional parameters $d_0 = 0.8$ $b_0 = 0.4$. The results are based on $M = 1,000$ Monte Carlo simulations and $T = 2,500$ observations. The sequence of LR test is performed with a theoretical significance level of $a = 1\%$ for both the selection of the cointegration rank and the optimal number of lags. The estimations of d and b are carried out under the restriction $d > \delta_{\min}$ with $\delta_{\min} = \tilde{d} - c \cdot \tilde{d}$, with $c = 0.15$ on the left panel and unrestrictedly on the right panel.

5 Conclusion

This paper discusses in detail the identification problems that affect the FCVAR $_{d,b}$ model of Johansen (2008). The main finding is that the fractional parameters of the system can not be uniquely determined when the lag structure is over-specified. In particular, the multiplicity of equivalent sub-models is provided in closed form given k and k_0 . It is also shown that a necessary and sufficient condition for the identification is that the $\mathcal{F}(d)$ condition, i.e. $|\alpha'_\perp \Gamma \beta_\perp| \neq 0$, is fulfilled. A simulation study highlights the practical problem of multiple humps in the expected profile log-likelihood function as a consequence of the identification problem and the over-specification of the lag structure. Furthermore, the simulations reveal a problem of weak identification, characterized by the presence of local and global maxima of the profile likelihood function in finite samples. We also prove that it is sufficient to restrict d to d_0 to solve the identification problem. However, since d_0 is unknown, we impose a lower-bound restriction on d , where the lower bound is determined on the basis of a preliminary semiparametric estimate of d_0 . This imposes the mildest restriction on the parameter space of the FCVAR $_{d,b}$ model. The Monte Carlo simulations show that the restricted estimates of the model parameters are unimodal and centered around the true values in most cases. It is also proved that model $\mathcal{H}_{0,k}$ is equivalent to model $\mathcal{H}_{p,k-1}$ under certain conditions on d and b . Unfortunately, the $\mathcal{F}(d)$ condition does not provide any information for the identification in this case, but it is again sufficient to impose a suitable lower bound restriction on the parameter space of d to solve this identification problem and retrieve a nesting structure of FCVAR $_{d,b}$ model that allows testing for the unknown lag-length and cointegration rank in the standard general-to-specific fashion.

References

- Avarucci, M. and Velasco, C. (2009). A Wald test for the cointegration rank in nonstationary fractional systems. *Journal of Econometrics*, 151(2):178–189.
- 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(2):409–424.
- Breitung, J. and Hassler, U. (2002). Inference on the cointegration rank in fractionally integrated processes. *Journal of Econometrics*, 110(2):167–185.
- Caporin, M., Rinaldo, A., and Santucci de Magistris, P. (2013). On the predictabil-

- ity of stock prices: A case for high and low prices. *Journal of Banking & Finance*, 37(12):5132–5146.
- Chen, W. and Hurvich, C. (2003). Semiparametric estimation of multivariate fractional cointegration. *Journal of the American Statistical Association*, 98:629–642.
- Chen, W. and Hurvich, C. (2006). Semiparametric estimation of fractional cointegrating subspaces. *Annals of Statistics*, 34:2939–2979.
- Christensen, B. J. and Nielsen, M. Ø. (2006). Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting. *Journal of Econometrics*, 133(1):343–371.
- Franchi, M. (2010). A representation theory for polynomial cofractionality in vector autoregressive models. *Econometric Theory*, 26(04):1201–1217.
- Granger, C. W. J. (1986). Developments in the study of cointegrated economic variables. *Oxford Bulletin of Economics and Statistics*, 48(3):213–28.
- Hualde, J. and Velasco, C. (2008). Distribution-free tests of fractional cointegration. *Econometric Theory*, 24:216–255.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12:231–254.
- Johansen, S. (1995a). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford.
- Johansen, S. (1995b). A statistical analysis of cointegration for I(2) variables. *Econometric Theory*, 11(01):25–59.
- Johansen, S. (2008). A representation theory for a class of vector autoregressive models for fractional processes. *Econometric Theory*, Vol 24, 3:651–676.
- Johansen, S. (2010). Some identification problems in the cointegrated vector autoregressive model. *Journal of Econometrics*, 158(2):262–273.
- Johansen, S. and Nielsen, M. Ø. (2010). Likelihood inference for a nonstationary fractional autoregressive model. *Journal of Econometrics*, 158(1):51–66.
- Johansen, S. and Nielsen, M. Ø. (2012). Likelihood inference for a fractionally cointegrated vector autoregressive model. *Econometrica*, 80(6):2667–2732.

- Lasak, K. (2010). Likelihood based testing for no fractional cointegration. *Journal of Econometrics*, 158(1):67–77.
- Lasak, K. and Velasco, C. (2015). Fractional cointegration rank estimation. *Journal of Business & Economic Statistics*, 33(2):241–254.
- MacKinnon, J. G. and Nielsen, M. Ø. (2014). Numerical Distribution Functions Of Fractional Unit Root And Cointegration Tests. *Journal of Applied Econometrics*, 29(1):161–171.
- Nielsen, B. (2006). *Order determination in general vector autoregressions*, volume 52 of *Lecture Notes–Monograph Series*, pages 93–112. Institute of Mathematical Statistics.
- Nielsen, M. Ø. and Popiel, M.K. (2016). A MATLAB program and user’s guide for the fractionally cointegrated VAR model. Working Paper, Queen’s University.
- Nielsen, M. Ø. and Shimotsu, K. (2007). Determining the cointegration rank in nonstationary fractional system by the exact local whittle approach. *Journal of Econometrics*, 141:574–596.
- Robinson, P. M. and Marinucci, D. (2003). Semiparametric frequency domain analysis of fractional cointegration. In Robinson, P. M., editor, *Time Series with Long Memory*, pages 334–373. Oxford University Press.
- Robinson, P. M. and Yajima, Y. (2002). Determination of cointegrating rank in fractional systems. *Journal of Econometrics*, 106:217–241.
- Shimotsu, K. and Phillips, P. C. (2005). Exact local whittle estimation of fractional integration. *Annals of Statistics*, 33(4):1890–1933.
- Tschernig, R., Weber, E., and Weigand, R. (2013a). Fractionally integrated VAR models with a fractional lag operator and deterministic trends: Finite sample identification and two-step estimation. Working Paper, University of Regensburg.
- Tschernig, R., Weber, E., and Weigand, R. (2013b). Long-run identification in a fractionally integrated system. *Journal of Business & Economic Statistics*, 31(4):438–450.