Carlini, Federico; Lasak, Katarzyna

Working Paper

Likelihood based inference for an Identifiable Fractional Vector Error Correction Model

Tinbergen Institute Discussion Paper, No. TI 2018-085/III

Provided in Cooperation with:
Tinbergen Institute, Amsterdam and Rotterdam

Suggested Citation: Carlini, Federico; Lasak, Katarzyna (2018) : Likelihood based inference for an Identifiable Fractional Vector Error Correction Model, Tinbergen Institute Discussion Paper, No. TI 2018-085/III, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:
https://hdl.handle.net/10419/185604

TI 2018-085/III
Tinbergen Institute Discussion Paper

Likelihood based inference for an Identifiable Fractional Vector Error Correction Model

Federico Carlini[1]
Katarzyna (K.A.) Lasak[2]

[1] USI, Lugano
[2] University of Amsterdam

Likelihood based inference for an Identifiable Fractional Vector Error Correction Model

Federico Carlini*, Katarzyna Łasak†

27 September 2018

Abstract

We consider the Fractional Vector Error Correction model proposed in Avarucci (2007), which is characterized by a richer lag structure than the models proposed in Granger (1986) and Johansen (2008, 2009). In particular, we discuss the properties of the model of Avarucci (2007) (FECM) in comparison to the model of Johansen (2008, 2009) (FCVAR). Both models generate the same class of processes, but the properties of the two models are different. First, opposed to the model of Johansen (2008, 2009), the model of Avarucci has a convenient nesting structure, which allows for testing the number of lags and the cointegration rank exactly in the same way as in the standard I(1) cointegration framework of Johansen (1995) and hence might be attractive for econometric practice. Second, we find that the model of Avarucci (2007) is almost free from identification problems, contrary to the model of Johansen (2008, 2009) and Johansen and Nielsen (2012), which identification problems are discussed in Carlini and Santucci de Magistris (2017). However, due to a larger number of parameters, the estimation of the FECM model of Avarucci (2007) turns out to be more complicated. Therefore, we propose a 4-step estimation procedure for this model that is based on the switching algorithm employed in Carlini and Mosconi (2014), together with the GLS procedure of Mosconi and Paruolo (2014). We check the performance of the proposed estimation procedure in finite samples by means of

*USI, Lugano, E-mail: federico.carlini@usi.ch

†Corresponding author. Tinbergen Institute and University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, Email: k.a.lasak@uva.nl.

a Monte Carlo experiment and we prove the asymptotic distribution of the estimators of all the parameters. The solution of the model has been previously derived in Avarucci (2007), while testing for the rank has been discussed in Łasak and Velasco (for cointegration strength >0.5) and Avarucci and Velasco (for cointegration strength <0.5). Therefore our paper fills in the gap for a complete inference based on Avarucci (2007) model.

Keywords: Error correction model, Gaussian VAR model, Fractional Cointegration, Estimation algorithm, Maximum likelihood estimation, Switching Algorithm, Reduced Rank Regression. **JEL:** C13, C32.

1 Introduction

The econometrics literature on fractional co-integration has developed rapidly in recent years. An empirically attractive modeling strategy is to use parametric inference, based on an econometric model that fully describes the system under consideration. It allows identification of the long-run and short-run structure of the model, as well as of the common stochastic trends and the impulse response functions summarizing the system dynamics. Three different Fractional Vector Error Correction Models (FVECM) have been proposed in the literature due to Granger (1986), Johansen (2008, 2009) and Avarucci (2007). These models turn out to be almost identical in the simplest case without short run dynamics, but more generally they are characterized by different lag structure specifications.

The FCVAR model proposed in Johansen (2008, 2009) has a convenient algebraic structure. The inference for this model has been developed in Johansen and Nielsen (2012). However, there exist identification problems in this model, as mentioned in Johansen and Nielsen (2012) and further discussed in Carlini and Santucci de Magistris (2017).

In this paper we demonstrate that the FECM model proposed by Avarucci (2007) is almost free from identification problems contrary to the FCVAR model of Johansen (2008, 2009). Also, designing testing procedures for the lag length and the cointegration rank is straightforward in FECM, due to the fact that the nesting structure follows the usual structure known for the I(1)

Cointegrated Vector AutoRegressive (CVAR) model. However, the estimation is more complicated in thie FECM model due to the multiplicative structure of the parameters involved.

We conclude that the model proposed by Avarucci might be more convenient for practitioners if we design an estimation procedure for this model, which we do in this paper.

We propose a 4 step algorithm, which is based on the approach of Carlini and Mosconi (2014) that maximizes the profile likelihood function using a switching algorithm and implements the GLS procedure proposed in Mosconi and Paruolo (2014). We check the performance of the proposed estimation procedure in finite samples by means of a Monte Carlo experiment and we prove the asymptotic distribution of the estimators of all the parameters. The solution of the model has been previously derived in Avarucci (2007), while testing for the rank has been discussed in Łasak and Velasco (for cointegration strength >0.5) and Avarucci and Velasco (for cointegration strength <0.5). Therefore our paper fills in the gap for a complete inference based on Avarucci (2007) model.

The remainder of the paper is organized as follows. Section 2 presents the FCVAR model proposed in Johansen (2008, 2009) and the FECM model suggested in Avarucci (2007). In particular we discuss derivation and we give the solutions of both models. We also present different representations of FECM. In Section 3 we discuss the identification and the convenient nesting structure of the FECM model. Section 4 introduces the profile likelihood and a 4 step switching algorithm to estimate the parameters of the FECM model together with the asymptotic distribution of the estimators of all the parameters. Section 5 illustrates the small sample properties of our estimation procedure by means of a Monte Carlo experiment. Section 6 concludes. Appendix A describes how to estimate a bilinear form with a GLS model. Appendix B presents Figures and Tables of the Monte Carlo experiment.

2 Model comparison

2.1 Johansen's FCVAR model

The model of Johansen (2008, 2009), which we denote in this paper as FCVAR, is given by the following dynamics

$$\Delta^d X_t = \alpha\beta'\Delta^{d-b}L_b X_t + \sum_{j=1}^{k}\Gamma_j\Delta^d L_b^j X_t + \varepsilon_t, \qquad \varepsilon_t \sim iid(0,\Omega), \tag{1}$$

where the vector of variables X_t is $p-$dimensional, the loadings α and the cointegrating relations β are $p\times r$ matrices with $0 \leq r \leq p$, Γ_j are $p\times p$ matrices of the short run dynamics and the fractional difference operator is given by the binomial expansion $\Delta^d := (1-L)^d = \sum_{j=0}^{\infty}(-1)^j\binom{d}{j}L^j$ and the fractional lag operator is defined $L_b = 1-\Delta^b$.

As shown in Johansen (2008, 2009), this model could be derived from the standard VAR model $\Delta Y_t = \alpha\beta' LY_t + \sum_{j=1}^{k}\Gamma_j\Delta L^j Y_t + \varepsilon_t$ analysed in Johansen (1995), where the lag operator L is such that $LX_t = X_{t-1}$, and the difference operator $\Delta = 1-L$, in the following way. First replace the difference operator Δ and the lag operator $L = 1-\Delta$ by fractional difference operator Δ^b and the fractional lag operator $L_b = 1-\Delta^b$, respectively, to obtain $\Delta^b Y_t = \alpha\beta'(1-\Delta^b)Y_t + \sum_{j=1}^{k}\Gamma_j\Delta^b L_b^j Y_t + \varepsilon_t$. Next define $Y_t = \Delta^{d-b}X_t$ to get the model (1).

The Granger representation of the model (1) is given in Johansen (2008, 2009):

$$X_t = C\Delta_+^{-d}\varepsilon_t + \Delta_+^{-(d-b)}Y_t^+ + \mu_t, \tag{2}$$

where μ_t is a deterministic component generated by initial values, $C = \beta_\perp(\alpha_\perp'\Gamma\beta_\perp)^{-1}\alpha_\perp'$ and $Y_t^+ = \sum_{n=0}^{t-1}\tau_n\varepsilon_{t-n}$, so Y_t^+ is fractional of order zero. Thus the solution of model (1) implies that X_t is a process integrated of order d ($I(d)$), while $\Delta^b X_t$ and $\beta' X_t$ are $I(d-b)$.

In general the cointegration rank r and number of lagged differences k is not known and needs to be determined. However, the nesting structure, as described in Carlini and Santucci di Magistris

(2017) turns out to be of the following form:

$$\begin{array}{ccccccccc}
\mathcal{H}_{0,0} & \subset & \mathcal{H}_{0,1} & \subset & \mathcal{H}_{0,2} & \subset & \dots & \subset & \mathcal{H}_{0,k} \\
\cap & & \cap & & \cap & & & & \cap \\
\mathcal{H}_{1,0} & \subset & \mathcal{H}_{1,1} & \subset & \mathcal{H}_{1,2} & \subset & \dots & \subset & \mathcal{H}_{1,k} \\
\cap & & \cap & & \cap & & & & \cap \\
\vdots & & \vdots & & \vdots & & \ddots & & \vdots \\
\cap & & \cap & & \cap & & & & \cap \\
\mathcal{H}_{p,0} & \subset & \mathcal{H}_{p,1} & \subset & \mathcal{H}_{p,2} & \subset & \dots & \subset & \mathcal{H}_{p,k}
\end{array}$$

with

$$\begin{array}{ccc}
\mathcal{H}_{0,1} & \subset & \mathcal{H}_{p,0} \\
\mathcal{H}_{0,2} & \subset & \mathcal{H}_{p,1} \\
\vdots & & \\
\vdots & & \\
\mathcal{H}_{0,k} & \subset & \mathcal{H}_{p,k-1}
\end{array},$$

where $\mathcal{H}_{p,k}$ denotes the hypothesis that the model (1) has cointegration rank $r = p$ and k lagged differences. Therefore, the joint identification of r and k, if both are unknown, becomes tricky.

It is also shown that there exists a number of equivalent FCVAR models, which causes problems with identification of fractional parameters d, b and lag length when the cointegration rank r is known, see Carlini and Santucci di Magistris (2017). They demonstrate that for any $k \geq k_0$, where k_0 denotes the number of lagged differences in the true DGP, the following holds:

- Given k_0 and k, with $k \geq k_0$, the number of equivalent sub-models that can be obtained is $m = [\frac{k+1}{k_0+1}]$, where $[x]$ denotes the greatest integer less than or equal to x.
- 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 $j = 0, 1, ..., m-1$.
- α, β are the same in these models.

Further, they give the number of equivalent sub-models in the following table

$k_0 \downarrow k \rightarrow$	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

Thus there are identification problems regarding the parameters $d, b, \Gamma_j, j = 1, \ldots, k$.

2.2 Avarucci's FECM model

The model of Avarucci (2007), which we denote in this paper as FECM, is given by the following dynamics

$$\Delta^d X_t = \alpha\beta' \Delta^{d-b} L_b X_t + \sum_{j=1}^{k} B_j L^j \Delta^{d-b} L_b X_t + \sum_{j=1}^{k} A_j L^j \Delta^d X_t + \varepsilon_t \qquad \varepsilon_t \sim iid(0, \Omega) \qquad (3)$$

where $B_j = -A_j(\alpha\beta')$. Avarucci (2007) imposes the restriction $X_t = 0$ for $t < 1$. This model is similar to the model of Lobato and Velasco (2006) for testing for fractional unit root in the univariate framework.

The model of Avarucci (2007) has been derived using a standard assumption in a parametric framework (see Robinson and Hualde (2003), Dueker and Startz (1998) that the dynamics of the stationary process can be given by an autoregressive representation. Consider a fractionally

cointegrated system in a triangular form, i.e.

$$\begin{aligned} \xi' \Delta^d X_t &= u_{1t}, \\ \beta' \Delta^{d-b} X_t &= u_{2t}, \end{aligned} \quad \text{with } d - b \geq 0. \tag{4}$$

The triangular representation (4) can be shown to be equivalent to the FVECM without short run dynamics, i.e.

$$\Delta^d X_t = \alpha \beta' \Delta^{d-b} L_b X_t + V_t, \tag{5}$$

where $\alpha = -\xi_\perp (\beta' \xi_\perp)^{-1}$ and $\beta' \alpha = -I_r$, and r is the cointegration rank. The process u_t has the $VAR(k)$ representation $A(L)u_t = v_t$. Then V_t is also a $VAR(k)$ process, i.e.: $V_t = \sum_{j=1}^{k} A_j \xi_{t-j} + \varepsilon_t$. Consider the model (5), then

$$V_t = \Delta^d X_t - \alpha \beta' \Delta^{d-b} L_b X_t, \tag{6}$$

can be written as $\Delta^d X_t = \alpha \beta' \Delta^{d-b} L_b X_t + \sum_{j=1}^{k} A_j \xi_{t-j} + \varepsilon_t$ and further using (6) $\Delta^d X_t = \alpha \beta' \Delta^{d-b} L_b X_t + \sum_{j=1}^{k} A_j [\Delta^d X_{t-j} - \alpha \beta' \Delta^{d-b} L_b X_{t-j}] + \varepsilon_t$ to give finally $\Delta^d X_t = \alpha \beta' \Delta^{d-b} L_b X_t + \sum_{j=1}^{k} A_j \Delta^d X_{t-j} + \sum_{j=1}^{k} B_j \Delta^{d-b} L_b X_{t-j}] + \varepsilon_t$, where $B_j = -A_j(\alpha \beta')$.

The model (3) can also be written in another form. The representation proposed below is coherent with the representation in Johansen (2008). The model (3) can be reformulated as:

$$\Delta^{d-b} \left(I_p - \sum_{j=1}^{k} A_j L^j \right) \left(\Delta^b I_p - \alpha \beta' L_b \right) X_t = \varepsilon_t. \tag{7}$$

This representation emphasizes the nature of the process. In fact, the FECM model is a series created by connecting two systems: a VAR process identified by the lag polynomial $(I - \sum_{j=1}^{k} A_j L^j)$ and a FCVAR process identified by the lag polynomial $\Delta^{d-b}(\Delta^b I_p - \alpha \beta' L_b)$. The following scheme represents the FECM process:

$$\varepsilon_t \rightarrow \boxed{A(L)^{-1}} \rightarrow V_t \rightarrow \boxed{\Pi_{d,b}(L_b)^{-1}} \rightarrow X_t$$

The input of the system is the Gaussian error term ε_t transformed in a VAR process V_t through

the transfer function $A(L)^{-1}$. Finally the VAR process V_t is transformed into a Fractionally Cointegrated process by means of the transfer function $\Pi_{d,b}(L_b)^{-1} := \Delta^{b-d}(\Delta^b I_p - \alpha\beta' L_b)^{-1}$.

In linear system theory, the dynamics of two systems connected in a series can be analysed by checking the zeros and poles of their transfer functions contemporaneously. Hence, the dynamics of the FECM can be found by checking the characteristic roots of the polynomials $A(z)$ and $\Pi_{d,b}(y)$, where $y = 1 - (1-z)^b$. This means that we generate fractional cointegration if $det(\Pi_{d,b}(y)) = 0$ has some of the characteristic fractional roots equal to one and $\beta'\alpha$ is a full rank matrix.

The FECM model is characterized by a different (and more complicated) lag structure than the model proposed in Granger (1986), i.e.:

$$\Delta^d X_t = \alpha\beta' \Delta^{d-b} L_b X_{t-1} + \sum_{j=1}^{k} \Gamma_j L^j \Delta^d X_t + \varepsilon_t, \tag{8}$$

and the FCVAR model (1) discussed in the previous section. In fact, FECM model (3) contains both the usual lags based on a standard lag operator present in Granger's model (8) and lags using the fractional lag operator. The latter are different than those present in the FCVAR model (1) of Johansen. However, in the very particular case of $d = b = 1$ with $k = 0$ all three models reduce to the standard ECM. Besides, when $k = 0$, i.e. the short run dynamics components are not present, then FCVAR model (3) and FECM model (1) are equal apart for the initial values. The solution of Johansen's FCVAR model depends on the initial values, for $t < 0$, while the FECM model of Avarucci implicitly has the restriction for which the process starts in $t = 0$.

The moving average representation (MA) of the FECM model (3) is given in Avarucci (2007). Following his Theorem 2.2, X_t has the representation

$$X_t = C\Delta_+^{-d} V_t + C^* \Delta_+^{-d+b} V_t + \Delta_+^{-d+2b} \sum_{j=1}^{t-1} \Phi_j V_{t-j}, \tag{9}$$

where $\sum_{j=0}^{\infty} ||\Phi_j||^2 < \infty$, and $C = \beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\alpha'_\perp$, and $C^* = -[\bar{\beta}\bar{\alpha}' + C\bar{\beta}\bar{\alpha}' + \bar{\beta}\bar{\alpha}'C + C\bar{\beta}\bar{\alpha}'C]$, where $\Phi_j, j = 1, \ldots, t-1$ are $p \times p$ matrices, $V_t = A(L)^{-1}\varepsilon_t$ and if c is a generic $p \times r$ matrix then $\bar{c} := c(c'c)^{-1}$ and $c_\perp$ is a $p \times (p-r)$ matrix such that $c'_\perp c = c' c_\perp = 0$. Thus, X_t and $\beta' X_t$ are Type

II $I(d)$ and $I(d-b)$ processes respectively.

The proof of Theorem 2.2 is largely based on Theorem 8 in Johansen (2008) and the MA representation (9) is based on the solution (2) given in Johansen (2008, 2009).

Therefore both models generate the same class of processes. However, in the FECM model proposed by Avarucci (2007), cointegration always occurs if $b > 0$ unlike in the FCVAR model of Johansen (2008, 2009), where the system can not be cointegrated for $b > 0$ if $\alpha\beta'$ is a full rank matrix. Thus, the model of Avarucci (2007) has more natural interpretation of the parameter b.

3 Statistical identification of the Avarucci's FECM model

3.1 The nesting structure of the Avarucci's FECM model

The nesting structure of the Avarucci's FECM model differs from the presented in Section 2 the nesting structure of the Johansen's (2008, 2009) FCVAR model and it follows the simple VAR structure, which makes testing the cointegration rank r and the number of lagged differences k to be straightforward. If we define the model

$$\mathcal{H}_{r,k}: \quad \Delta^{d-b}(I_p - \sum_{i=1}^{k} A_i L^i)(\Delta^b I_p - \alpha\beta' L_b)X_t = \varepsilon_t, \quad r = 0, \ldots, p$$

then, the nesting structure of the Avarucci's FECM model is given by

$$\begin{array}{ccccccccc}
\mathcal{H}_{0,0} & \subset & \mathcal{H}_{0,1} & \subset & \mathcal{H}_{0,2} & \subset & \ldots & \subset & \mathcal{H}_{0,k} \\
\cap & & \cap & & \cap & & & & \cap \\
\mathcal{H}_{1,0} & \subset & \mathcal{H}_{1,1} & \subset & \mathcal{H}_{1,2} & \subset & \ldots & \subset & \mathcal{H}_{1,k} \\
\cap & & \cap & & \cap & & & & \cap \\
\vdots & & \vdots & & \vdots & & \ddots & & \vdots \\
\cap & & \cap & & \cap & & & & \cap \\
\mathcal{H}_{p,0} & \subset & \mathcal{H}_{p,1} & \subset & \mathcal{H}_{p,2} & \subset & \ldots & \subset & \mathcal{H}_{p,k}
\end{array},$$

For example, the inclusion $\mathcal{H}_{2,1} \subset \mathcal{H}_{2,2}$ can be tested by $A_2 = 0$ and the inclusion $\mathcal{H}_{1,1} \subset \mathcal{H}_{2,1}$ can be tested on a rank restriction on the matrix $\alpha\beta'$. Moreover, it is simple to prove that the model $\mathcal{H}_{0,1}$ is not nested in $\mathcal{H}_{2,0}$ because the term $\alpha\beta' L_b$ is zero in $\mathcal{H}_{0,1}$.

The nesting structure of the Johansen FCVAR model is more complicated as discussed in Section 2, see also Carlini and Santucci de Magistris (2017) for the details. The convenient nesting structure of Avarucci (2007) FECM model not only makes the testing procedures straightforward, but also assures identification of the model, contrary to Johansen (2008, 2009), which we discuss in the following sub-sections.

3.2 Identification of FECM when the lag length is unknown

Recall from Section 2 that in the Johansen's (2008, 2009) FCVAR model (1) there exists a number of equivalent models with overspecified lag length. In order to illustrate that this does not happen in the Avarucci's (2007) FECM model let us consider the model with just 2 lags:

$$\mathcal{H}_2 : \Delta^{d-b} \left(I_p - \sum_{j=1}^{2} A_j L^j \right) \left(\Delta^b I_p - \alpha\beta' L_b \right) X_t = \varepsilon_t. \tag{10}$$

where $\mathcal{H}_2$ indicates the model with $k = 2$ in (7).

Let us demonstrate under which restrictions the two sub-models of $\mathcal{H}_2$: the model with 2 lags $\mathcal{H}_2^{(0)}$ and the model with 1 lag $\mathcal{H}_2^{(1)}$

$$\mathcal{H}_2^{(0)} \quad : \quad \Delta^{d_0-b_0} \left(I_p - (I_p + \tilde{A}_1)L + \tilde{A}_1 L^2 \right) \left(\Delta^{b_0} I_p - \tilde{\alpha}\tilde{\beta}' L_{b_0} \right) X_t = \varepsilon_t. \tag{11}$$

$$\mathcal{H}_2^{(1)} \quad : \quad \Delta^{d_1-b_1} \left(I_p - \bar{A}_1 L \right) \left(\Delta^{b_1} I_p - \bar{\alpha}\bar{\beta}' L_{b_1} \right) X_t = \varepsilon_t. \tag{12}$$

can be reparameterized as in Carlini and Santucci de Magistris (2017), which would indicate identification problems.

First note, that in case of a unit root, the sub-model $\mathcal{H}_2^{(0)}$ in equation (11) can be written as:

$$\Delta^{d_0-b_0} \left(I_p - \tilde{A}_1 L \right) \left(\Delta^{b_0} I_p - \tilde{\alpha}\tilde{\beta}' L_{b_0} \right) (I_p - I_p L) X_t = \varepsilon_t.$$

or equivalently as

$$\Delta^{d_0-b_0}\left(I_p-\tilde{A}_1L\right)\left(\Delta^{b_0}I_p-\tilde{\alpha}\tilde{\beta}'L_{b_0}\right)\Delta X_t=\varepsilon_t$$

Therefore,

$$\mathcal{H}_2^{(0)}:\Delta^{d_0-b_0+1}\left(I_p-\tilde{A}_1L\right)\left(\Delta^{b_0}I_p-\tilde{\alpha}\tilde{\beta}'L_{b_0}\right)X_t=\varepsilon_t \tag{13}$$

Now, let us compare the sub-models (12) and (13). It is clear that the equations (13) and (12) reparametrize when $\tilde{\alpha}\tilde{\beta}'=\bar{\alpha}\bar{\beta}'$, $\tilde{A}_1=\bar{A}_1$, $b_1=b_0$ and $d_0-b_0+1=d_1-b_1$. Hence, $\mathcal{H}_2^{(0)}=\mathcal{H}_2^{(1)}$ if and only if $d_0+1=d_1$ and there is a unit root.

Furthermore, note that under the unit root the model $\mathcal{H}_2^{(0)}$ is a sub-model of the model $\mathcal{H}_2$, when we impose the restriction $A_2+A_1-I_p=0$. Instead, the sub-model $\mathcal{H}_2^{(1)}$ is the sub-model of the model $\mathcal{H}_2$ when we impose the restriction $A_2=0$.

Therefore, the parameter b is always identified. In order to rule out the identification problem for the parameter d and autoregressive parameters A_j, we only need to assume that the characteristic polynomial

$$\Pi(z)=\left(I_p-\sum_{j=1}^{k}A_jz^j\right)$$

has roots outside the unit circle, which is already assumed in Avarucci (2007). Therefore the identification problem for the parameter d and autoregressive parameters A_j is not present in the Avarucci's FECM model.

3.3 Lack of identification when $\alpha\beta'=0$

However, the problem of identification can arise when $\alpha\beta'=0$. In this situation, (7) is given by

$$\Delta^d\left(I_p-\sum_{j=1}^{k}A_jL^j\right)X_t=\varepsilon_t$$

and the parameter b is not identified. This particular feature of the model has been used in Lasak (2010) to propose a sup-test for no cointegration and is common for all fractionally cointegrated Vector Error Correction models. This identification issue is also relevant in the FCVAR model

when the number of lags is $k = 0$.

3.4 The proof of identification of the Avarucci's FECM model

In this sub-section we give the proof of identification of the FECM model (3) when the lag length k of the VAR is known. The proof of identification follows the same steps as in Johansen and Nielsen (2012).

Theorem 1. Suppose $b \neq 1$ and $\alpha\beta' \neq 0$, then the parameters $\theta = vec(d, b,\ A_1, \ldots,\ A_k, \alpha, \beta, \Omega)$ in the model

$$\Delta^{d-b}\left(I_p - \sum_{j=1}^{k} A_j L^j\right)\left(\Delta^b I_p - \alpha\beta' L_b\right) X_t = \varepsilon_t$$

are identifiable.

Proof

1. A parametric model is identified when $f_{\lambda_0}(x_t|I_{t-1}) = f_{\tilde{\lambda}}(x_t|I_{t-1})$ implies $\lambda_0 = \tilde{\lambda}$, where $f(x_t|I_{t-1})$ is the conditional density function. In the FECM model the parameter vector is given by $\lambda = vec(d, b, \alpha, \beta, A_1, \ldots, A_k, \Omega)$.

2. In the model (3) ε_t is assumed to be iid. We are interested in the first and the second moment of f. Hence we have to show that the conditions $E_{\lambda_0}(x_t|I_{t-1}) = E_{\tilde{\lambda}}(x_t|I_{t-1})$ and $\text{var}_{\lambda_0}(x_t|I_{t-1}) = \text{var}_{\tilde{\lambda}}(x_t|I_{t-1})$ imply $\lambda_0 = \tilde{\lambda}$. The equality for conditional variances requires that $\Omega_0 = \tilde{\Omega}$.

3. We use the decomposition $I_p = \beta\bar{\beta}' + \beta_\perp\bar{\beta}'_\perp$ where $\bar{\beta}' = (\beta'\beta)^{-1}\beta'$ and $\bar{\beta}'_\perp = (\beta'_\perp\beta_\perp)^{-1}\beta'_\perp$ to identify the parameters α and β defining $\tilde{\tilde{\alpha}} = \alpha\beta'\bar{\beta}_0$ and $\tilde{\tilde{\beta}} = \beta(\bar{\beta}'_0\beta)^{-1}$ so that $\alpha\beta' = \tilde{\tilde{\alpha}}\tilde{\tilde{\beta}}'$.

4. The equality for conditional means requires that

$$\Pi_{\lambda_0}(z) = (I_p - A_1^0 z - \cdots - A_k^0 z^k)(1-z)^{d_0-b_0}((1-z)^{b_0} I_p - \alpha_0\beta_0'(1-(1-z)^{b_0}) =$$

$$= (I_p - \tilde{A}_1 z - \cdots - \tilde{A}_k z^k)(1-z)^{\tilde{d}-\tilde{b}}((1-z)^{\tilde{b}} I_p - \tilde{\alpha}\tilde{\beta}'(1-(1-z)^{\tilde{b}}) = \Pi_{\tilde{\lambda}}(z)$$

If $k > 0$ and $r > 0$ then it is implied that $A_j^0 = \tilde{A}_j, j = 1, \ldots, k, d_0 = \tilde{d}, b_0 = \tilde{b}, \alpha_0 = \tilde{\alpha}$ and $\beta_0 = \tilde{\beta}$ when $b_0 = \tilde{b} \neq 1$.

5. If $b_0 = \tilde{b} = 1$ then $(I_p - A_1^0 z - \cdots - A_k^0 z^k)((1-z)I_p - \alpha_0\beta_0'(1-(1-z)) = (I_p - \tilde{A}_1 z - \cdots - \tilde{A}_k z^k)((1-z)I_p - \tilde{\alpha}\tilde{\beta}'(1-(1-z))$ is not generally solved by $A_j^0 = \tilde{A}_j, j = 1, \ldots, k, \alpha_0 = \tilde{\alpha}$ and $\beta_0 = \tilde{\beta}$. In fact, by the theory of matrix polynomials in Dennis at al. (1976), this problem can be easily explained. Suppose that you have a given matrix polynomial

$$B(z) = I_p - B_1 z - \ldots - B_k z^k$$

and $B_j, j = 1, \ldots, k$ are $p \times p$ fixed square matrices and we want to decompose it as

$$(I_p - D_1 z - \ldots D_{k-1} z^{k-1})(I_p - C_1 z)$$

where C_1 is called the right solvent of the matrix polynomial $B(z)$. We define the latent values as the values $z_1, \ldots, z_{pk}$ such that $|B(z_k)| = 0$ and the right latent vectors as the vectors $v_1, \ldots, v_{pk}$ such that $B(z_j)v_j = 0$ where z_j is a latent value. If $B(z)$ has p linearly independent right latent vectors $v_{1,}, \ldots, v_p$ corresponding to latent roots $z_1, \ldots, z_p$, then $C_1 := Q\Lambda Q^{-1}$ is a right solvent, where $Q = [v_1, \ldots, v_p]$ and $\Lambda = diag(z_1, \ldots, z_p)$, see Dennis et al. (1976). Therefore, in general the right solvent is not unique (because we can find many $z_1, \ldots, z_p$ that satisfy the requirement of the theorem) and there exist different matrices $D_j^{(l)}, j = 1, \ldots, k-1, C_1^{(l)}$ for $l = 1, \ldots, p$ that satisfy the decomposition of $B(z)$. For this reason, when $b_0 = \tilde{b} = 1$ the matrices $A_j, j = 1, \ldots, k$ (the $D_j, j = 1, \ldots, k-1$ matrices in the example) and α and β (the C_1 matrix in the example) are not identified.

6. Suppose now $r = 0$, then the parameters $(d, A_1, \ldots, A_k)$ are just identified and it follows the same argument as in Johansen and Nielsen (2012).

This proof shows that an identification problem occurs when the DGP value of the cointegration gap parameter b_0 is equal to 1. This identification issue can naturally affect the asymptotic distributions of the parameters of the model. We will discuss in Section 4.2 the consequences of the identification

issue on the estimation method proposed.

4 Estimation of the Avarucci's FECM model

In this section we use a profile likelihood approach to estimate the parameters of the FECM model. We concentrate the likelihood function on the parameters $\psi = (d, b)'$ as in Johansen and Nielsen (2012). The profile maximum likelihood estimator is

$$\arg\max_{\psi \in \mathcal{K}} \ell_T(\psi), \tag{14}$$

where $\mathcal{K}$ is a compact set defined as $\mathcal{K} = \left\{\eta \leq b \leq d \leq \bar{d}\right\}$ for some values $\eta > 0$ and $\bar{d} > 0$ and

$$\ell_T(\psi) = -\frac{1}{2T}\log\det(\Omega) + \frac{1}{pT}\log(2\pi).$$

Hence, the parameters α, β and $A_j, j = 1, \ldots, k$ are considered as nuisance parameters.

To maximize the likelihood function we use the same idea as in Johansen and Nielsen (2012). For any combination of ψ, we maximize the likelihood function with respect to the nuisance parameters α, β and $A_j, j = 1, \ldots, k$ with a numerical routine based on a switching algorithm described below.

4.1 The switching algorithm

For any given values of $\psi = (d, b)'$, we estimate

$$\Delta^d X_t - \alpha\beta'\Delta^{d-b}L_b X_t + A(I_k \otimes \alpha)(I_k \otimes \beta')\Delta^{d-b}L_b Z_t - A\Delta^d Z_t = \varepsilon_t.$$

Now we can use the switching algorithm to maximize the likelihood with respect to ψ.

Note that

$$\underset{p\times pk}{A} = \begin{bmatrix} A_1 & A_2 & \cdots & A_k \end{bmatrix} \quad \underset{pk\times 1}{Z_t} = \begin{bmatrix} X_{t-1} \\ X_{t-2} \\ \vdots \\ X_{t-k} \end{bmatrix}.$$

We run the following switching algorithm:

Step 1. For given values of α, β, Ω, we estimate A in the following equation

$$\Delta^d X_t - \alpha\beta'\Delta^{d-b} L_b X_t = A[\Delta^d Z_t - (I_k \otimes \alpha)(I_k \otimes \beta')\Delta^{d-b} L_b Z_t] + \varepsilon_t$$

The parameters $\hat{A}$ are estimated with ordinary least squares.

Step 2. Given values of A, β, Ω, we estimate α in the following equation:

$$\Delta^d X_t - A\Delta^d Z_t = [I_p : -A](I_{k+1} \otimes \alpha)(I_{k+1} \otimes \beta')\Delta^{d-b} L_b \begin{bmatrix} X_t \\ Z_t \end{bmatrix} + \varepsilon_t$$

Using the Mosconi and Paruolo algorithm explained in the Appendix, we estimate α with generalized least squares by imposing $H_\alpha = [I_p : -A]$, $W_t = (I_{k+1} \otimes \beta')\Delta^{d-b} L_b \begin{bmatrix} X_t \\ Z_t \end{bmatrix}$ and K is the matrix such that $(I_{k+1} \otimes \alpha) = Kvec(\alpha)$.

Step 3. For given values of A, α, Ω, we estimate β in the following equation

$$\Delta^d X_t - A\Delta^d Z_t = [I_p : -A](I_{k+1} \otimes \alpha)(I_{k+1} \otimes \beta')\Delta^{d-b} L_b \begin{bmatrix} X_t \\ Z_t \end{bmatrix} + \varepsilon_t$$

Again, the Mosconi and Paruolo algorithm is needed to estimate β with generalized least squares after imposing $H_\beta = [I_p : -A](I_{k+1} \otimes \alpha)$, $W_t = \Delta^{d-b} L_b \begin{bmatrix} X_t \\ Z_t \end{bmatrix}$ where K is the matrix such that $(I_{k+1} \otimes \beta') = Kvec(\beta)$.

Step 4. For given values of A, α, β we estimate Ω as

$$\Omega = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t \varepsilon_t'$$

and then we evaluate the likelihood.

We iterate Step 1 - Step 2 - Step 3 and Step 4 until convergence. Finally, we optimize the likelihood function (14) with respect to $\psi = (d, b)'$ to calculate the ML estimator $\hat{\psi}$.

4.2 Estimation and Identification issues

As explained in Section 3, the identification issue in the FECM model arises when the DGP value $b_0 = 1$. In particular, the identification issue is relevant for the matrices α, β and $A_j, j = 1, \ldots, k$, because they are shown not to be unique when $b_0 = 1$. We maximize the profile likelihood function with respect to $\psi = (d, b)'$. Hence, the maximum likelihood estimator $\hat{\psi}$ is always identified, but the estimated nuisance parameters $\hat{\alpha}$, $\hat{\beta}$ and $\hat{A}_j, j = 1, \ldots, k$ are not identified.

By simple algebra, we note that if $b_0 = 1$ then the FECM model with k lags is a reparameterization of the FCVAR model where $b = 1$, $d \geq 1$ ($d = d_0 + 1$) and k lags. In fact, the characteristic polynomial of the FECM model when $b = 1$ is given by the following expression

$$\Delta^{d-1}(I_p - \sum_{j=1}^{k} A_j z^j)((1-z)I_p - \alpha\beta' z) = (1-z)^{d-1}\sum_{j=0}^{k+1} \Psi_j z^j$$

while the characteristic polynomial of the FCVAR model when $b = 1$ is given by the following expression:

$$(1-z)^d I_p = \alpha\beta'(1-z)^{d-1}z - \sum_{j=1}^{k} \Gamma_j (1-z)^d z^j = (1-z)^{d-1}\sum_{j=0}^{k+1} \Psi_j z^j.$$

Furthermore, the FCVAR model with $b = 1$ is an identified model. Hence, we could test in the FECM framework the hypothesis $\mathcal{H}_0 : b = 1$. The asymptotic distribution of this hypothesis has to be derived because this is a case of hypothesis testing in which the nuisance parameters are not

identified. A reference that describe in more detail this problem is Hansen (1996). If the hypothesis $\mathcal{H}_0 : b = 1$ is not rejected, then we study the FCVAR in which $d \geq 1$.

In the Monte Carlo experiment, we have generated a FECM model with $b = 1$, $b = 0.99$, $b = 1.01$ and $T = 100,000$ observations. When we estimate the model (setting the initial values of α, β and A to their true values), the switching algorithm converge very slowly and the number of iterations is approximately of an order of 10^9. These non-identified and almost-non-identified FECM models make the proposed estimation procedure very difficult to manage in the proximity of $b_0 = 1$.

4.3 Initial values

Using the switching algorithm it is important to have a good initial guess for parameters α, β and the matrix A. Hence we use as initial guess for these parameters the estimates from the equation

$$\Delta^d X_t - \alpha\beta' \Delta^{d-b} L_b X_t + \Xi \Delta^{d-b} L_b Z_t - A\Delta^d Z_t = \varepsilon_t$$

obtained with Conditional Sum of Squares (CSS) profile likelihood method. We get $\hat{d}, \hat{b}, \hat{\alpha}, \hat{\beta}, \hat{A}, \hat{\Xi}$ by maximizing the profile likelihood function (or profile CSS) with a reduced rank regression. In fact, this model can be estimated with a profile likelihood depending on the parameters d, b. Hence, we maximize

$$\ell(\psi) = -\log\det\left(\frac{1}{T}\sum_{i=1}^{T} \hat{\varepsilon}_t(\psi)\hat{\varepsilon}_t'(\psi)\right) - \frac{1}{pT}\log(2\pi)$$

over a compact parameter set $\mathcal{K} = \left\{\eta \leq b \leq d \leq \bar{d}\right\}$ for some values $\bar{d} > \eta > 0$.

The estimates $\hat{A}$, $\hat{\alpha}$ and $\hat{\beta}$ are used as initial guess to start up the switching algorithm. The estimates $\hat{d}$ and $\hat{b}$ are used as initial values in the optimization routine.

4.4 Asymptotic properties

The switching algorithm proposed in Section 4.1 gives us the estimates of all the parameters of the FECM model. We summarize the asymptotic properties of these estimates in Theorem 2 and Theorem 3 and prove them in the Apppendix C.

Theorem 2. The estimator $\hat{\beta}$ is consistent. The asymptotic distribution of $\hat{\beta}$ is mixed Gaussian with the optimal convegence rate, hence $\hat{\beta} - \beta \in O_P\left(T^{-b}\right)$, $b > 0.5$.

Theorem 3. The estimators $\hat{d},\ \hat{b},\ \hat{\alpha},\ \hat{A}_1, \ldots, \hat{A}_k$ are root-T consistent and have an asymptotic joint normal distribution.

Theorem 2 indicates that the asymptotic distribution of the estimator of the cointegrating relationship $\hat{\beta}$ in FECM model is equal to the distribution of the ML estimator of this paramater in Johansen and Nielsen (2012), Łasak (2008) and GLS in Robinson and Hualde (2003). It is also similar to Johansen (1995) for $b = 1$ fixed. Since $\hat{\beta}$ remains mixed normal, we can test for the values of cointegration vector using Wald test that will be χ^2 distributed. Theorem 3 implies that the standard inference rules for other estimators would be valid as well.

5 Simulation experiment

The Monte Carlo exercise is conducted with a simulation of the FECM model using the Jensen (2014)'s algorithm to generate the FCVAR model two times. The FECM model can be generated in two steps: in the first step invert a $\text{FCVAR}_{d,b}$ model with 0 lags to obtain Y_t, given by

$$Y_t = (I_p - A_{1,0}L - \ldots - A_{k,0}L^k)^{-1}\varepsilon_t.$$

In the second step generate the process

$$X_t = (\Delta^{d_0} I_p - \alpha_0\beta_0'\Delta^{d_0-b_0}L_{b_0})^{-1}Y_t$$

again with the Jensen (2014) algorithm. We developed a new routine that transforms the parameters of the VAR model into the parameters of the FCVAR model when $d = b = 1$. With these transformed parameters, we generate the process X_t.

We run three Monte Carlo experiments. In the first experiment we generate $N = 1500$ Monte

Carlo replications of the following data generating process

$$(I_p - A_{1,0}L - A_{2,0}L^2)(\Delta^{d_0} I_p - \alpha_0 \beta_0' \Delta^{d_0 - b_0} L_{b_0}) X_t = \varepsilon_t$$

where $\varepsilon_t \sim i.i.d.N(0, \Omega_0)$ and $t = 1, \ldots, T$. The numerical parameters of the DGP are

$$A_0 = [A_{1,0} : A_{2,0}] = \begin{bmatrix} -0.2 & 0.2 & 0.2 & 0 \\ 0 & 0.3 & -0.3 & -0.3 \end{bmatrix}$$

$$\alpha_0 = \begin{bmatrix} -0.3 \\ 0.3 \end{bmatrix} \quad \beta_0 = \begin{bmatrix} 1 \\ -0.4 \end{bmatrix} \quad \Omega_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and the values of d_0 and b_0 are chosen such that the inverse roots y of the determinant $|(1 - y)I_p - \alpha_0 \beta_0' y| = 0$ are outside the fractional circle $\mathbb{C}_{b_0}$ as described in Johansen (2008). The parameters in A_0 are chosen such that the inverse roots z of the lag polynomial calculated as $|I_p - A_{1,0}z - A_{2,0}z^2| = 0$ are outside the unit circle. Furthermore, we have chosen parameters for which no identification issues occur.

For each run of the Monte Carlo simulation we fit the FECM model with two lags, given by

$$(I_p - A_1 L - A_2 L^2)(\Delta^d I_p - \alpha \beta' \Delta^{d-b} L_b) X_t = \varepsilon_t \tag{15}$$

using the switching algorithm discussed in Section 4.1.

When we maximize the likelihood function, the first experiment is to use as initial values for the parameters d and b the true data generating process values d_0 and b_0, while the initial values in the switching algorithm for α, β and Ω are imposed to be α_0, β_0 and Ω_0. The maximization routine climbs the likelihood function within the values $d \in [0.01, 2]$ and $b \in [0.01, 2]$ in order to avoid negative - or close to zero - $\hat{d}$ and $\hat{b}$ estimates. Further, we impose in the maximization routine the restriction $d \geq b$ because in the FECM this inequality must be satisfied.

We introduce two parameters to control the convergence of the switching algorithm. The first

parameter is the maximum number of iterations of the switching algorithm N^{iter}. The program stops when a number of N^{iter} chosen is reached. The second parameter is a tolerance number Tol. The program stops when the absolute value of the likelihood at step $k+1$ minus the likelihood value at step k is less than Tol. In the Monte Carlo we have set these two parameters to be $N^{iter} = 20,000$ and $Tol = 10^{-8}$.

To simplify the exposition, we introduce new notation for the elements of the matrices in model. The elements of the matrices are

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \quad \beta = \begin{bmatrix} 1 \\ \beta_1 \end{bmatrix} \quad \Omega = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix} \quad A = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & a_{11}^{(2)} & a_{12}^{(2)} \\ a_{21}^{(1)} & a_{22}^{(1)} & a_{21}^{(2)} & a_{22}^{(2)} \end{bmatrix}$$

where $\omega_{12} = \omega_{21}$. When we estimate the FECM model the vectors $\hat{\beta}^* = [\beta_1^* : \beta_2^*]'$ and $\hat{\alpha}^* = [\alpha_1^* : \alpha_2^*]'$ are normalized by calculating $\hat{\alpha} = \beta_1^* \hat{\alpha}^*$ and $\hat{\beta} = \frac{1}{\beta_1^*} \cdot \hat{\beta}^*$.

We present the Monte Carlo results when we simulate the process with $d_0 = 0.8$, $b_0 = 0.6$ and $T = 100,000$. The results of the sample statistics of the distributions of the estimated parameters are reported in Table 3.1. Figure 3.1 displays the plots of the densities of the estimates of the parameters in Eq. 15. These densities are calculated with a non-parametric method and they are smoothed by a Gaussian Kernel.

The biases of the Monte Carlo estimates are smaller than an order of magnitude of 10^{-3} and the standard deviations are smaller than 0.1. We analyzed if the Monte Carlo sample distributions were normal with a Jarque-Bera test and we do not reject the hypothesis for some of the elements in the matrix A and the fractional parameters d and b.

We tried to check if the switching algorithm is robust with respect to different initial values and the results are still the same when the sample size is $T = 100,000$. If the sample size is $T = 10,000$ then the initial values for α, β and Ω are crucial if we want to find the global maximum of the likelihood function.

The algorithm behaves differently depending on the choice of d_0 and b_0. In fact, if we simulate a model where $d_0 = b_0 = 0.9$, we need over 20,000 iterations of the switching algorithm to converge,

which is related to the identification of the model discussed in Section 3.

In the second Monte Carlo experiment we fix $d = d_0$ and $b = b_0$ and let the switching algorithm find the estimates of all other parameters. In Figure 3.2 and Table 3.2 the results are shown when the DGP takes values $d_0 = b_0 = 0.6$ and α_0, β_0, A_0 as before and $T = 100,000$. We notice that in this set up we could not reject the null hypothesis of normality for all the parameters but β.

In the third Monte Carlo experiment we fit the model

$$(I - A_1 L - A_2 L^2)(\Delta^d I_p - \Pi \Delta^{d-b} L_b) X_t = \varepsilon_t$$

where

$$\Pi = \begin{bmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{bmatrix},$$

i.e. matrix Π has a full column rank and we run the switching algorithm fixing $d = d_0$ and $b = b_0$. In Figure 3.3 and Table 3.3 the Monte Carlo results are shown when the DGP has $d_0 = b_0 = 0.6$ and α_0, β_0, A_0 as before and $T = 100,000$. We do not reject the null hypothesis of normality for all the estimated parameters with a Jarque-Bera test.

6 Conclusions

In this paper we discuss two fractionally cointegrated models: a FCVAR model proposed in Johansen (2008, 2009) and the FECM model proposed in Avarucci (2007) and Avarucci and Velasco (2010). They both generate the same class of processes, but due to different lag structures their properties differ significantly. The FECM model turns out to be characterised by a more convenient nesting structure, that allows a straightforward way for testing the cointegration rank and the number of lagged differences to be included as short run parameters. Further, the identification problems are far less severe in FECM model than in the FCVAR model. On the other hand, the estimation of FECM is more complicated due to the presence of two different parts that model the short run dynamics and the restriction that relates their parameters. Thus, we propose an estimation

procedure, which is based on the suggestion of Carlini and Mosconi (2014) that maximizes the likelihood function using a switching algorithm and the GLS procedure of Mosconi and Paruolo (2014). We prove the asymptotic distribution of the estimators of all the parameters and we illustrate by means of Monte Carlo experiment the performance of our procedure in finite samples. We find that close to the DGPs chosen in the Monte Carlo simulations, the estimated parameters $\hat{d}, \hat{b}, \hat{\alpha}, \hat{A}_j, j = 1, 2$ are normally distributed, whilst $\hat{\beta}$ has a fat-tailed distribution, which confirms the asymptotic theory developed. The solution of the model has been previously derived in Avarucci (2007), while testing for the rank has been discussed in Łasak and Velasco (for cointegration strength >0.5) and Avarucci and Velasco (for cointegration strength <0.5). Therefore our paper fills in the gap for a complete inference based on Avarucci (2007) model.

Appendix A

The following algorithm describes how to estimate a bilinear form with a GLS model. Further details can be found in Mosconi and Paruolo (2014).

Consider the following equation

$$Y_t = H\theta' W_t + \varepsilon_t \quad t = 1, \ldots, T$$

$$\varepsilon_t \sim iidN(0, \Omega), \quad vec(\theta) = K\psi$$

where Y_t and W_t are respectively $p_y \times 1$ and $p_w \times 1$ vectors, H and θ are respectively $p_y \times r$ and $p_w \times r$ matrices. Then, we can estimate ψ as

$$\hat{\psi} = (K'(H'\Omega^{-1}H \otimes S_{ww})K)^{-1}K'vec(S_{wy}\Omega^{-1}H)$$

where $S_{ww} = \sum_{t=1}^{T} W_t W_t'$ and $S_{wy} = \sum_{t=1}^{T} W_t Y_t'$.

Appendix B

Figure 1: Distributions of the parameters when the DGP parameters are $d_0 = 0.8$, $b_0 = 0.6$. $N = 1,500$ Monte Carlo replications and $Y_t, t = 1, \ldots, 100000$.

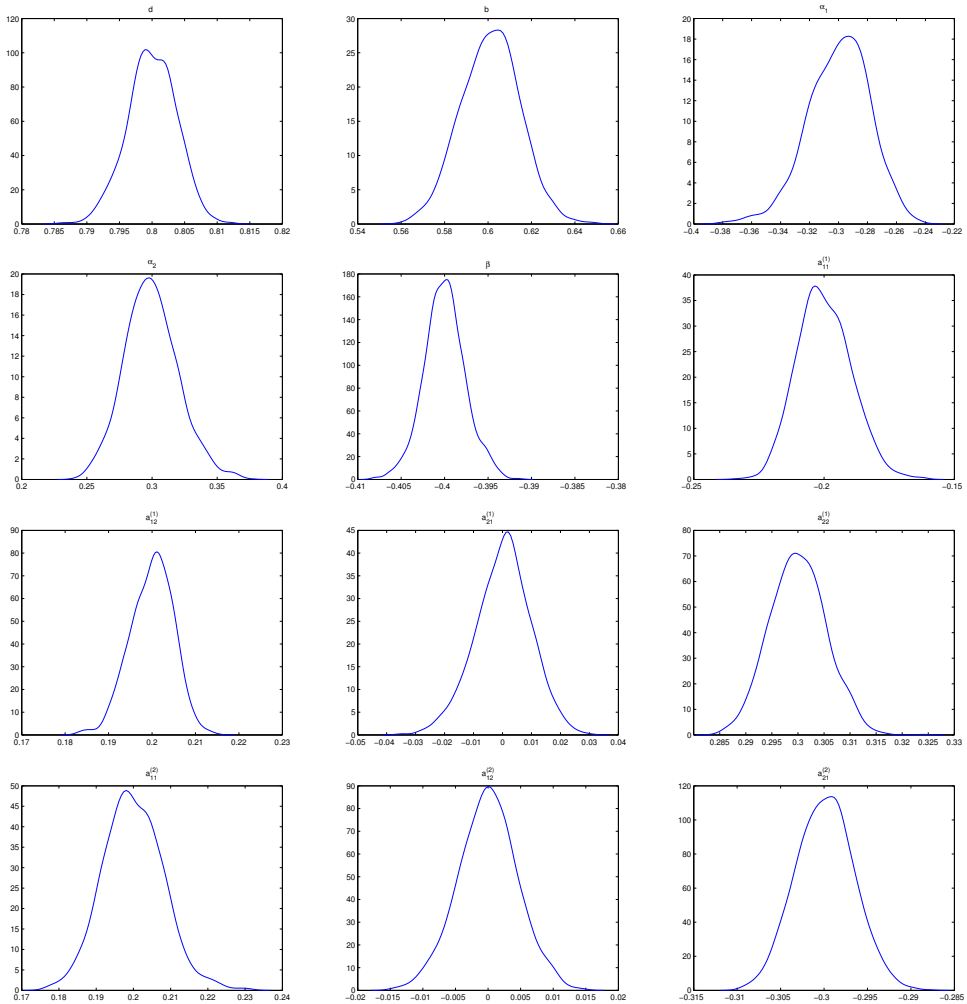

Table 1: Sample statistics of the Monte Carlo distributions when $d_0 = 0.8$ $b_0 = 0.6$, $T = 100000$ observations and $N = 1500$ Monte Carlo replications.

	d	b	α_1	α_2	β
Bias	0.0000	0.0012	0.0001	-0.0005	0.0000
Std.Dev	0.0038	0.0135	0.0214	0.0206	0.0024
Skew.	-0.1404	0.0117	-0.3912	0.3018	0.0546
Kurtosis	3.1239	3.0231	3.3295	3.2352	3.3915
p-value JB test	0.0514	>0.5000	0.0010	0.0010	0.0084
	$a_{11}^{(1)}$	$a_{12}^{(1)}$	$a_{21}^{(1)}$	$a_{22}^{(1)}$	
Bias	0.0001	0.0000	0.0002	0.0000	
Std.Dev	0.0103	0.0049	0.0096	0.0054	
Skew	0.2826	-0.2777	-0.1961	0.1228	
Kurtosis	3.1502	3.1095	3.1846	3.0136	
p-value JB test	0.0010	0.0010	0.0052	0.1434	
	$a_{11}^{(2)}$	$a_{12}^{(2)}$	$a_{21}^{(2)}$	$a_{22}^{(2)}$	
Bias	0.0000	0.0000	0.0000	-0.0002	
Std.Dev	0.0081	0.0046	0.0033	0.0030	
Skew	0.2360	-0.0388	0.0532	0.0482	
Kurtosis	3.3815	3.0897	2.9097	2.8717	
p-value JB test	0.0010	>0.5000	>0.5000	0.4295	

Figure 2: Distributions of the Monte Carlo simulations when $d_0 = b_0 = 0.6$ are fixed in the switching algorithm. The generated FECM paths have $T = 100000$ observations.

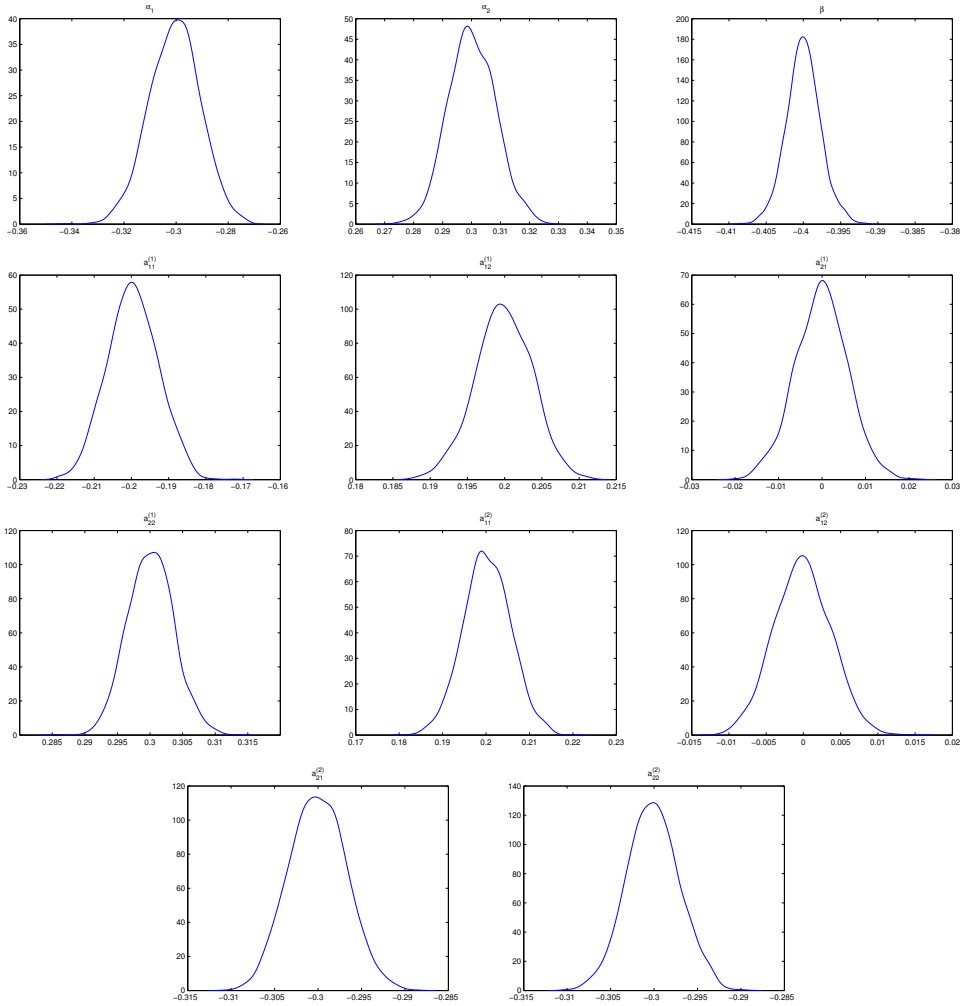

Table 2: Sample statistics of the Monte Carlo distributions $d_0 = b_0 = 0.6$ kept fixed in the switching algorithm with $T = 100000$ and $N = 1750$ Monte Carlo replications.

	d	α_1	α_2	β_1
Bias	-	-0.0005	0.0002	0.0000
Std.Dev	-	0.0097	0.0081	0.0023
Skew.	-	-0.1066	0.0750	0.0655
Kurtosis	-	3.0495	2.9494	3.4033
p-value JB test	-	0.1671	0.3867	0.0031
	$a_{11}^{(1)}$	$a_{12}^{(1)}$	$a_{21}^{(1)}$	$a_{22}^{(1)}$
Bias	0.0004	-0.0002	-0.0002	0.0001
Std.Dev	0.0068	0.0038	0.0060	0.0035
Skew	0.0432	-0.0834	-0.0087	0.0745
Kurtosis	2.8909	2.9927	3.0914	3.0576
p-value JB test	0.4818	0.3513	>0.5000	0.3810
	$a_{11}^{(2)}$	$a_{12}^{(2)}$	$a_{21}^{(2)}$	$a_{22}^{(2)}$
Bias	0.0002	-0.0001	-0.0000	-0.0001
Std.Dev	0.0054	0.0038	0.0032	0.0030
Skew	0.0200	0.0748	0.0591	0.0374
Kurtosis	2.9412	2.9225	2.8213	2.9423
p-value JB test	>0.5000	0.3455	0.1802	>0.5000

Figure 3: Distributions of the Monte Carlo simulations when $d_0 = b_0 = 0.6$ are fixed in the switching algorithm. The generated FECM paths have $T = 100,000$ observations.

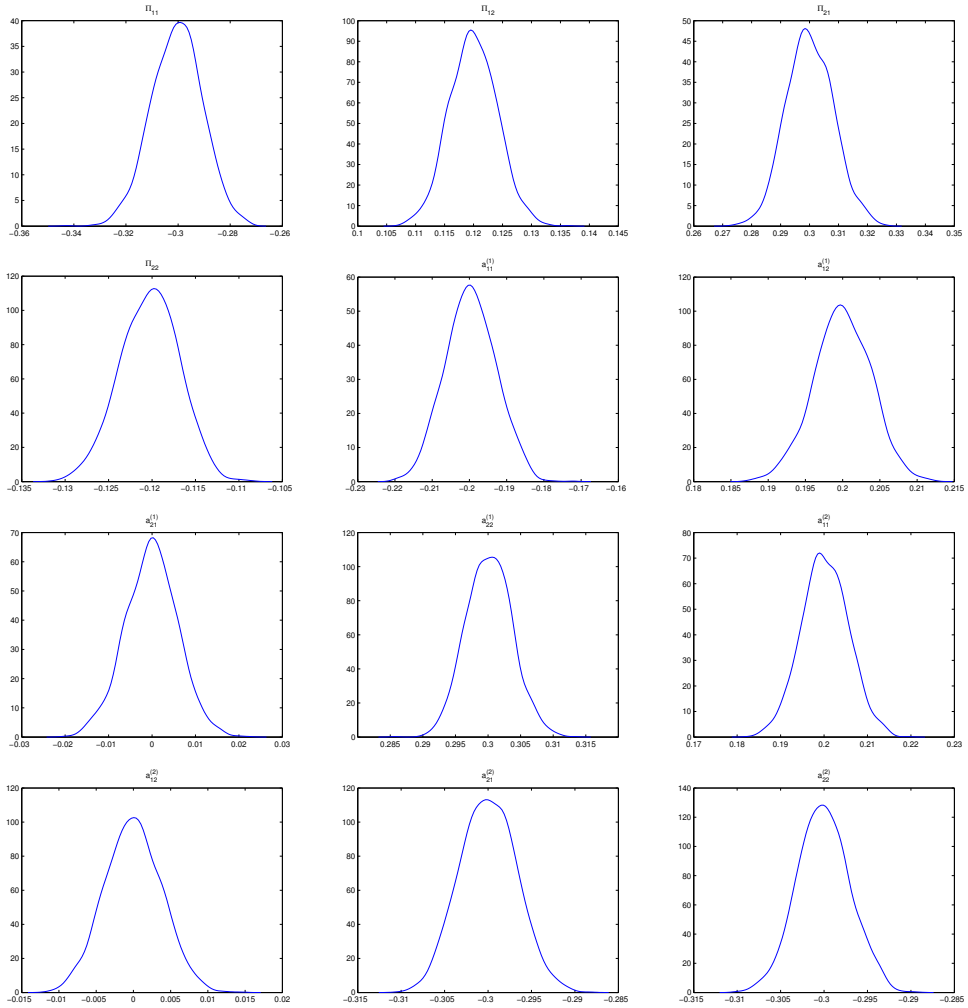

Table 3: Sample statistics of the Monte Carlo distributions $d_0 = b_0 = 0.6$ kept fixed and $T = 100000$, $N = 1750$ Monte Carlo replications and estimation of the matrix Π.

	π_{11}	π_{12}	π_{21}	π_{22}
Bias	- 0.0005	0.0000	0.0002	0.0003
Std.Dev	0.0098	0.0041	0.0081	0.0034
Skew.	-0.1072	0.0678	0.0755	-0.0877
Kurtosis	3.0452	3.0158	2.9507	2.9101
p-value JB test	0.1663	0.4993	0.3846	0.2338
	$a_{11}^{(1)}$	$a_{12}^{(1)}$	$a_{21}^{(1)}$	$a_{22}^{(1)}$
Bias	0.0004	0.0000	-0.0001	0.0002
Std.Dev	0.0068	0.0038	0.0060	0.0035
Skew	0.0428	-0.0699	-0.0070	0.0829
Kurtosis	2.8821	3.0308	3.0953	3.0458
p-value JB test	0.4444	0.4578	>0.5000	0.3315
	$a_{11}^{(2)}$	$a_{12}^{(2)}$	$a_{21}^{(2)}$	$a_{22}^{(2)}$
Bias	0.0002	0.0000	0.0000	-0.0001
Std.Dev	0.0054	0.0038	0.0033	0.0030
Skew	0.0202	0.0642	0.0589	0.0379
Kurtosis	2.9403	2.9068	2.8339	2.9455
p-value JB test	>0.5000	0.3859	0.2123	>0.5000

Appendix C

Here we provide proofs of the asymptotic distribution of the estimators of all the parameters obtained from the procedure proposed in Section 4.1.

C.1 Model. Consider the model with one lag

$$\begin{aligned} \Delta^d X_t &= \alpha\beta'\Delta^{d-b}L_b X_t + V_t \\ V_t &= AV_{t-1} + \varepsilon_t. \end{aligned}$$

Multiplying by $I_p - AL$ and using $(I_p - AL)V_t = \varepsilon_t$ we find

$$\Delta^d X_t = \alpha\beta'\Delta^{d-b}L_b X_t - A\alpha\beta'\Delta^{d-b}L_b X_{t-1} + \Delta^d A X_{t-1} + \varepsilon_t. \tag{16}$$

C.2 Representation. For the model (16) the representation is given by equation (9) in Section 2.2 and can be written as

$$X_t = \Delta^{-d_0}(C_0 V_t + \Delta^{b_0} Z_{1t}),$$

where Z_{1t} is stationary. Inserting

$$V_t = (I_p - A_0 L)^{-1}\varepsilon_t = (I_p - A_0)^{-1}\varepsilon_t + \Delta\sum_{i=0}^{\infty} A_i^*\varepsilon_{t-i} = (I_p - A_0)^{-1}\varepsilon_t + \Delta Z_{2t}$$

where Z_{2t} is stationary, we find

$$X_t = \Delta^{-d_0}\{C_0(I_p - A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t} + \Delta^{b_0} Z_{1t}\}, \tag{17}$$

Note that $\beta_0' X_t = \beta_0'\Delta^{-d_0+b_0} Z_{1t}$ is fractional of order $d_0 - b_0$.

C.3 Likelihood. We let λ be a notation for all parameters, and $\lambda\backslash d$ are all the parameters except d. We want to find expressions for $D_d\varepsilon_t(\lambda\backslash d)$ say, and use the notation

$$D_d\varepsilon_t(\lambda_0) = D_d\varepsilon_t(\lambda_0\backslash d)|_{d=d_0}.$$

We first define $\varepsilon_t(\lambda)$ and the (conditional given initial values) likelihood function, setting $\Omega = \Omega_0$:

$$\begin{aligned} \varepsilon_t(\lambda) &= \Delta^d X_t - \alpha\beta'\Delta^{d-b} L_b X_t + A\alpha\beta'\Delta^{d-b} L_b L X_t - \Delta^d A X_{t-1} \\ -2\log L(\lambda) &= tr\{\Omega_0^{-1}\sum_{t=1}^{T} \varepsilon_t(\lambda)\varepsilon_t(\lambda)'\} \end{aligned}$$

We next insert the X_t from (17):

$$\begin{aligned} \varepsilon_t(\lambda) &= \Delta^{d-d_0}\{C_0(I_p - A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t} + \Delta^{b_0} Z_{1t}\} \\ &\quad -\alpha\beta'\Delta^{d-d_0-b} L_b\{C_0(I_p - A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t} + \Delta^{b_0} Z_{1t}\} \\ &\quad +A\alpha\beta'\Delta^{d-d_0-b} L_b L\{C_0(I_p - A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t} + \Delta^{b_0} Z_{1t}\} \\ &\quad -A\Delta^{d-d_0} L\{C_0(I_p - A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t} + \Delta^{b_0} Z_{1t}\}. \end{aligned} \tag{18}$$

C.4 The asymptotic distribution of $\hat{\beta}$. We normalize the usual way as $\beta = \beta_0 + \beta_{0\perp}\theta$, and first set all other the parameters equal to their true value. We find

$$\begin{aligned} \varepsilon_t(\lambda_0\backslash\theta) &= \{C_0(I_p - A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t} + \Delta^{b_0} Z_{1t}\} \\ &\quad -\alpha_0(\beta_0' + \theta'\beta_{\perp 0}')\Delta^{-b_0} L_{b_0}\{C_0(I_p - A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t} + \Delta^{b_0} Z_{1t}\} \\ &\quad +A_0\alpha_0(\beta_0' + \theta'\beta_{\perp 0}')\Delta^{-b_0} L_{b_0} L\{C_0(I_p - A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t} + \Delta^{b_0} Z_{1t}\} \\ &\quad -A_0 L\{C_0(I_p - A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t} + \Delta^{b_0} Z_{1t}\}. \end{aligned}$$

Differentiating with respect to θ we find two terms

$$\begin{aligned} D_\theta\varepsilon_t(\lambda_0\backslash\theta)|_{\theta=\theta_0} &= -\alpha_0(d\theta)'\beta_{\perp 0}'\Delta^{-b_0} L_{b_0}\{C_0(I_p - A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t} + \Delta^{b_0} Z_{1t}\} \\ &\quad +A_0\alpha_0(d\theta)'\beta_{\perp 0}'\Delta^{-b_0} L_{b_0} L\{C_0(I_p - A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t} + \Delta^{b_0} Z_{1t}\} \end{aligned}$$

In this expression we keep the most nonstationary terms, which determine the asymptotic behaviour

of the score function, and find, using $D_\theta \varepsilon_t(\lambda_0) = D_\theta \varepsilon_t(\lambda_0 \backslash \theta)|_{\theta=\theta_0}$

$$\begin{aligned} D_\theta \varepsilon_t(\lambda_0) &\simeq -\alpha_0(d\theta)'\beta_{\perp 0}'\{C_0(I_p - A_0)^{-1}(\Delta^{-b_0} - 1)\varepsilon_t\} \\ &\quad + A_0\alpha_0(d\theta)'\beta_{\perp 0}'\{C_0(I_p - A_0)^{-1}(\Delta^{-b_0} - 1)\varepsilon_t \\ &= -(I_p - A_0)\alpha_0(d\theta)'\beta_{\perp 0}'C_0(I_p - A_0)^{-1}(\Delta^{-b_0} - 1)\varepsilon_t. \end{aligned}$$

Note that the Avarucci model has the adjustment coefficients $\tilde{\alpha}_0 = (I_p - A_0)\alpha_0$ such that $\tilde{\alpha}_{0\perp}' = \alpha_{0\perp}'(I_p - A_0)^{-1}$ and we could define $\tilde{C}_0 = C_0(I_p - A_0)^{-1}$. The score function then becomes

$$\begin{aligned} & -2T^{-b_0-1/2} D_\theta \log L(\lambda_0) \\ = \; & tr\{(d\theta)'\beta_{\perp 0}'C_0(I_p - A_0)^{-1}T^{-b_0-1/2}\sum_{t=1}^{T}(\Delta^{-b_0} - 1)\varepsilon_t\varepsilon_t'\Omega_0^{-1}(I_p - A_0)\alpha_0\} \\ & \xrightarrow{D} tr\{(d\theta)'\beta_{\perp 0}'\tilde{C}_0\int_0^1 W_{b_0-1}(dW)'\Omega_0^{-1}\tilde{\alpha}_0\}. \end{aligned}$$

where

$$\begin{aligned} S_{T,t} &= T^{-b_0+1/2}(\Delta^{-b_0} - 1)\varepsilon_t \xrightarrow{D} W_{b_0-1}(u), \\ T^{-1}\sum_{t=1}^{T} S_{T,t}\varepsilon_t' &= T^{-b_0-1/2}\sum_{t=1}^{T}(\Delta^{-b_0} - 1)\varepsilon_t\varepsilon_t' \xrightarrow{D} \int_0^1 W_{b_0-1}(dW)' \\ T^{-1}\sum_{t=1}^{T} S_{T,t}S_{T,t} &= T^{-2b_0}\sum_{t=1}^{T}\{(\Delta^{-b_0} - 1)\varepsilon_t\}\{(\Delta^{-b_0} - 1)\varepsilon_t\}' \\ &\xrightarrow{D} \int_0^1 W_{b_0-1}W_{b_0-1}'du \end{aligned}$$

Note that $\tilde{C}_0 W_{b_0-1}$ depends on $\tilde{\alpha}_{\perp 0}'W$ and is independent of $\tilde{\alpha}_0'\Omega_0^{-1}W$, so the limit of the score function is mixed Gaussian. The information is found as the limit

$$\begin{aligned} & T^{-2b_0} tr\{\Omega_0^{-1}\sum_{t=1}^{T} D_\theta \varepsilon_t(\lambda_0) D_\theta \varepsilon_t(\lambda_0)'\} \\ & \xrightarrow{D} tr\{\Omega_0^{-1}\tilde{\alpha}_0(d\theta)'\beta_{\perp 0}'\tilde{C}_0\int_0^1 W_{b_0-1}W_{b_0-1}'du\tilde{C}_0'\beta_{\perp 0}(d\theta)\tilde{\alpha}_0'\}. \end{aligned}$$

A Taylors expansion shows that (provided the estimator is consistent) we find for all matrices $d\theta$

$$\begin{aligned} & tr\{(d\theta)'\beta'_{\perp 0}\tilde{C}_0 T^{-1}\sum_{t=1}^{T} S_{T,t}\varepsilon'_t\Omega_0^{-1}\tilde{\alpha}_0\} \\ \simeq\ & -tr\{(d\theta)'\beta'_{\perp 0}\tilde{C}_0 T^{-1}\sum_{t=1}^{T} S_{T,t}S'_{T,t}\tilde{C}'_0\beta_{\perp 0}(\hat{\theta}-\theta_0)\} \end{aligned}$$

hence

$$\begin{aligned} T^{b_0}(\hat{\theta}-\theta)' \ &\simeq\ [\beta'_{\perp 0}\tilde{C}_0 T^{-1}\sum_{t=1}^{T} S_{T,t}S'_{T,t}\tilde{C}'_0\beta_{\perp 0}]^{-1}\beta'_{\perp 0}\tilde{C}_0 T^{-1}\sum_{t=1}^{T} S_{T,t}\varepsilon'_t\Omega_0^{-1}\tilde{\alpha}_0 \\ &\xrightarrow{D} [\beta'_{\perp 0}\tilde{C}_0\int_0^1 W_{b_0-1}W'_{b_0-1}du\tilde{C}'_0\beta_{\perp 0}]^{-1}\beta'_{\perp 0}\tilde{C}_0\int_0^1 W_{b_0-1}(dW)'\Omega_0^{-1}\tilde{\alpha}_0 \end{aligned}$$

C.5 The asymptotic distribution of $\hat{\mathbf{d}}$. First all parameters are set to their null value except d, and we note that all terms have the difference Δ^{d-d_0} :

$$\begin{aligned} \varepsilon_t(\lambda_0\backslash d) \ =\ & \Delta^{d-d_0}\{C_0(I_p-A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t} + \Delta^{b_0}Z_{1t}\} \\ & -\Delta^{d-d_0}\alpha_0\beta'_0\Delta^{-b_0}L_{b_0}\{C_0(I_p-A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t} + \Delta^{b_0}Z_{1t}\} \\ & +\Delta^{d-d_0}A_0\alpha_0\beta'_0\Delta^{-b_0}L_{b_0}L\{C_0(I_p-A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t} + \Delta^{b_0}Z_{1t}\} \\ & -\Delta^{d-d_0}A_0L\{C_0(I_p-A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t} + \Delta^{b_0}Z_{1t}\}. \end{aligned}$$

We also note that the in the second and third terms, we can use $\beta'_0 C_0 = 0$, so we are left with

$$\begin{aligned} \varepsilon_t(\lambda_0\backslash d) \ =\ & \Delta^{d-d_0}\{C_0(I_p-A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t} + \Delta^{b_0}Z_{1t}\} \\ & -\Delta^{d-d_0}\alpha_0\beta'_0 L_{b_0}Z_{1t} \\ & +\Delta^{d-d_0}A_0\alpha_0\beta'_0 L_{b_0}LZ_{1t} \\ & -\Delta^{d-d_0}A_0L\{C_0(I_p-A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t} + \Delta^{b_0}Z_{1t}\}. \end{aligned}$$

All nonstationary term vanishes and $D_d\varepsilon_t(\lambda_0)$ is stationary. It follows that the score

$$T^{-1/2}D_d \log L = T^{-1/2}tr\{\sum_{t=1}^{T} D_d\varepsilon_t(\lambda_0)\varepsilon_t'\Omega_0^{-1}\}$$

asymptotically Gaussian, and the information is found as the limit of

$$tr\{T^{-1}\sum_{t=1}^{T} D_d\varepsilon_t(\lambda_0)D_d\varepsilon_t(\lambda_0)'\Omega_0^{-1}\}.$$

Thus the asymptotic distribution of $T^{1/2}(\hat{d}-d_0)$ is Gaussian.

C.6 The asymptotic distribution of $\hat{\mathbf{b}}$. The derivatives are found as follows. First all parameters are set to their null value except b,:

$$\begin{aligned}
\varepsilon_t(\lambda_0\backslash b) &= \{C_0(I_p-A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t} + \Delta^{b_0}Z_{1t}\} \\
&\quad -\alpha_0\beta_0'\Delta^{-b}L_b\{C_0(I_p-A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t} + \Delta^{b_0}Z_{1t}\} \\
&\quad +A_0\alpha_0\beta_0'\Delta^{-b}L_bL\{C_0(I_p-A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t} + \Delta^{b_0}Z_{1t}\} \\
&\quad -A_0L\{C_0(I_p-A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t} + \Delta^{b_0}Z_{1t}\}.
\end{aligned}$$

Again we employ $\beta_0'C_0=0$ and find

$$\begin{aligned}
\varepsilon_t(\lambda_0\backslash b) &= \{C_0(I_p-A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t} + \Delta^{b_0}Z_{1t}\} \\
&\quad -\alpha_0\beta_0'\Delta^{-b}L_b\Delta^{b_0}Z_{1t} \\
&\quad +A_0\alpha_0\beta_0'\Delta^{-b}L_bL\Delta^{b_0}Z_{1t} \\
&\quad -A_0L\{C_0(I_p-A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t} + \Delta^{b_0}Z_{1t}\}.
\end{aligned}$$

Taking derivative with respect to b we find

$$D_b\varepsilon_t(\lambda_0\backslash b)|_{b=b_0} = (I_p-A_0)\alpha_0\beta_0'D_b\Delta^{-(b-b_0)}|_{b=0}Z_{1t}$$

Thus $D_b\varepsilon_t(\lambda_0)$ is stationary and the asymptotic distribution of $\hat{b}$ is Gaussian. The information is found as

$$tr\{T^{-1}\sum_{t=1}^{T} D_b\varepsilon_t(\lambda_0)D_b\varepsilon_t(\lambda_0)'\Omega_0^{-1}\}.$$

and the off diagonal elements are

$$tr\{T^{-1}\sum_{t=1}^{T} D_b\varepsilon_t(\lambda_0)D_d\varepsilon_t(\lambda_0)'\Omega_0^{-1}\}.$$

This shows that we get a joint Gaussian limit distribution of $T^{1/2}(\hat{d}-d_0, \hat{b}-b_0)$.

A small comment in the end: we know that

$$\begin{aligned}
\pi_n(d) &= \frac{d(d+1)\dots(d+n-1)}{n!} \\
\pi_n(0) &= \begin{cases} 1 & n=0 \\ 0 & n=1,2,... \end{cases}
\end{aligned}$$

for small d we have

$$\pi_n(d) \approx d\frac{(n-1)!}{n!},$$

such that $D^d\pi_n(d)|_{d=0} = 0$ for $n = 0$, and

$$D^{d-d_0}\pi_n(d-d_0)|_{d=d_0} = D^d\pi_n(d)|_{d=0} = 1/n, n = 1, 2, \dots$$

Thus and Z_t we find

$$D^{d-d_0}\pi_n(d-d_0)|_{d=d_0}Z_t = Z_{t-1} + Z_{t-2}/2 + Z_{t-3}/3 + \dots$$

Thus if Z_t is stationary, then also $D^{d-d_0}\pi_n(d-d_0)|_{d=d_0}Z_t$ is stationary.

C.7 The asymptotic distribution of $\hat{\alpha}$ and $\hat{A}$. We fix all the parameters but α to their true values and we get

$$\begin{aligned}\varepsilon_t(\lambda_0\backslash\alpha) &= \Delta^{d-d_0}\{C_0(I_p-A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t} + \Delta^{b_0}Z_{1t}\} \\ &\quad -\alpha\beta_0' L_{b_0}\{Z_{1t}\} \\ &\quad +A_0\alpha\beta_0' L_{b_0}L\{Z_{1t}\} \\ &\quad -A_0 L\{\Delta^{b_0}Z_{1t} + C_0(I_p-A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t}\}.\end{aligned} \tag{19}$$

Taking the derivative with respect to α we get

$$D_\alpha\varepsilon_t(\lambda_0\backslash\alpha) = -d\alpha\beta_0' L_{b_0}Z_{1t} + A_0 d\alpha\beta_0' L_{b_0}L(Z_{1t}) \tag{20}$$

Thus $D_\alpha\varepsilon_t(\lambda_0)$ is stationary and the asymptotic distribution of $\hat{\alpha}$ is Gaussian. In fact, it follows that the score

$$T^{-1/2}D_\alpha \log L = T^{-1/2}tr\{\sum_{t=1}^{T} D_\alpha\varepsilon_t(\lambda_0)\varepsilon_t'\Omega_0^{-1}\}$$

asymptotically Gaussian, and the information is found as the limit of

$$tr\{T^{-1}\sum_{t=1}^{T} D_\alpha\varepsilon_t(\lambda_0)D_\alpha\varepsilon_t(\lambda_0)'\Omega_0^{-1}\}.$$

We fix the parameters to their true values but A and we get

$$\begin{aligned}\varepsilon_t(\lambda_0\backslash\alpha) &= \Delta^{d-d_0}\{C_0(I_p-A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t} + \Delta^{b_0}Z_{1t}\} \\ &\quad -\alpha_0\beta_0' L_{b_0}\{Z_{1t}\} \\ &\quad +A\alpha_0\beta_0' L_{b_0}L\{Z_{1t}\} \\ &\quad -AL\{\Delta^{b_0}Z_{1t} + C_0(I_p-A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t}\}.\end{aligned}$$

Taking the derivative with respect to A we get

$$D_A\varepsilon_t(\lambda_0\backslash\alpha) = -dA\alpha_0\beta_0' L_{b_0}LZ_{1t} - dAL\{\Delta^{b_0}Z_{1t} + C_0(I_p-A_0)^{-1}\varepsilon_t + C_0\Delta Z_{2t}\}$$

Thus $D_A\varepsilon_t(\lambda_0)$ is stationary and the asymptotic distribution of $\hat{A}$ is Gaussian. In fact, it follows that the the score

$$T^{-1/2}D_A \log L = T^{-1/2}tr\{\sum_{t=1}^{T} D_A\varepsilon_t(\lambda_0)\varepsilon_t'\Omega_0^{-1}\}$$

asymptotically Gaussian, and the information is found as the limit of

$$tr\{T^{-1}\sum_{t=1}^{T} D_A\varepsilon_t(\lambda_0)D_A\varepsilon_t(\lambda_0)'\Omega_0^{-1}\}.$$

The off diagonal elements

$$tr\{T^{-1}\sum_{t=1}^{T} D_A\varepsilon_t(\lambda_0)D_\alpha\varepsilon_t(\lambda_0)'\Omega_0^{-1}\}, tr\{T^{-1}\sum_{t=1}^{T} D_A\varepsilon_t(\lambda_0)D_d\varepsilon_t(\lambda_0)'\Omega_0^{-1}\}$$

$$tr\{T^{-1}\sum_{t=1}^{T} D_A\varepsilon_t(\lambda_0)D_b\varepsilon_t(\lambda_0)'\Omega_0^{-1}\}, tr\{T^{-1}\sum_{t=1}^{T} D_\alpha\varepsilon_t(\lambda_0)D_d\varepsilon_t(\lambda_0)'\Omega_0^{-1}\}$$

$$tr\{T^{-1}\sum_{t=1}^{T} D_\alpha\varepsilon_t(\lambda_0)D_b\varepsilon_t(\lambda_0)'\Omega_0^{-1}\}$$

are product of stationary components.

Hence the asymptotic distribution of $T^{1/2}vec(\hat{d}-d_0, \hat{b}-b_0, \hat{\alpha}-\alpha_0, \hat{A}-A_0)$ is Gaussian.

References

[1] Avarucci, M. (2007), Three Essays on Fractional Cointegration, Ph.D. Thesis, University of Rome Tor Vergata.

[2] Avarucci, M., Velasco C. (2009), A Wald Test for the Cointegration Rank in Nonstationary Fractional Systems, Journal of Econometrics, 151, 178-189.

[3] Carlini, F., Mosconi, R. (2014), Twice Integrated Models, Manuscript.

[4] Carlini, F., Santucci de Magistris, P. (2017), On the identification of fractionally cointegrated VAR models with the F(d) condition, Forthcoming in Journal of Business and Economics Statistics.

[5] Dennis, J-E., Traub, J.F. and R.P. Weber (1976), The Algebraic Theory of Matrix Polynomials, SIAM Journal on Numerical Analysis, Vol. 13, No.6, pp. 831-845

[6] Dueker, M., Startz, R. (1998), Maximum-likelihood estimation of fractional cointegration with an application to U.S. and Canadian bond rates. Review of Economics and Statistics, 83.

[7] Granger, C. W. J. (1986), Developments in the Study of Cointegrated Economic Variables, Oxford Bulletin of Economics and Statistics, 48, 213-28.

[8] Hansen, B.E. (1996), Inference when a nuisance parameter is not identified under the null hypothesis, Econometrica 64, 413-430.

[9] Jensen, A. N. (2014), Efficient simulation of the Johansen-Nielsen model, University of Copenhagen, working paper.

[10] Johansen, S. (2008), A representation theory for a class of vector autoregressive models for fractional processes, Econometric Theory, 24, 651-676.

[11] Johansen, S. (2009), Representation of cointegrated autoregressive processes with application to fractional cointegration, Econometric Reviews, 28, 121-145.

[12] Johansen, S., Nielsen, M. O. (2012), Likelihood inference for a fractionally cointegrated vector autoregressive model, Econometrica, 80, 2667-2732.

[13] Łasak, K. (2008), Maximum likelihood estimation of fractionally cointegrated systems, CREATES Research Paper 2008-53.

[14] Łasak, K. (2010), Likelihood based testing for no fractional cointegration, Journal of Econometrics, 158, 67-77.

[15] Lobato, I., Velasco, C. (2006) ,Optimal Fractional Dickey-Fuller tests, Econometrics Journal, 9, 492-510.

[16] Mosconi, R., Paruolo, P. (2014), Rank and order conditions for identification in simultaneous system of cointegrating equations with integrated variables of order two, MPRA Paper 53589,

[17] Robinson, P. M., Hualde, J. (2003), Cointegration in Fractional Systems with Unknown Integration Orders, Econometrica, 71, 1727-1766.