On an Estimation Method for an Alternative Fractionally Cointegrated Model

Federico Carlini [1]
Katarzyna Lasak [2]

[1] *CREATES, Aarhus University, Denmark;*
[2] *Faculty of Economics and Business Administration, VU University Amsterdam, and Tinbergen Institute, the Netherlands.*

On an Estimation Method for an Alternative Fractionally Cointegrated Model

Federico Carlini[†]

CREATES, Aarhus University, DK

E-mail: fcarlini@creates.au.dk

Katarzyna Łasak[*†]

VU Amsterdam & Tinbergen Institute, NL

E-mail: k.a.lasak@vu.nl

April 30, 2014

Abstract

In this paper we consider the Fractional Vector Error Correction model proposed in Avarucci (2007), which is characterized by a richer lag structure than models proposed in Granger (1986) and Johansen (2008, 2009). We discuss the identification issues of the model of Avarucci (2007), following the ideas in Carlini and Santucci de Magistris (2014) for the model of Johansen (2008, 2009). We propose a 4-step estimation procedure that is based on the switching algorithm employed in Carlini and Mosconi (2014) and the GLS procedure in Mosconi and Paruolo (2014). The proposed procedure provides estimates of the long run parameters of the fractional cointegrated system that are consistent and unbiased, which we demonstrate by a Monte Carlo experiment.

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 methodological 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 in the model, as well as of the common stochastic trends and the impulse response functions. Three different Fractional Vector Error Correction Models (FVECM) have been proposed in the literature in Granger (1986), Johansen (2008, 2009) and Avarucci (2007). They turn out to be identical in the simplest case without short run dynamics, but in general case they are characterized by different lag structures.

The 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, that have been mentioned in Johansen and Nielsen (2012) and further discussed in Carlini and Santucci de Magistris (2014).

In this paper we demonstrate that the identification problems also occur but are less severe in the model proposed by Avarucci (2007) and that statistical inference for this model is different. The estimation is slightly more complicated due to the multiplicative structure of the parameters involved, while testing would be less complicated due to the fact that the nesting structure follows the usual structure of the Vector AutoRegressive (VAR) models. We propose a 4 step algorithm to estimate the parameters of this model. The algorithm is based on the approach of Carlini and Mosconi (2014) that maximizes likelihood function using a switching algorithm and a GLS procedure proposed in Mosconi and Paruolo (2014).

In the Monte Carlo experiment we investigate the small sample properties of the estimates of all the parameters of this model and we check also the asymptotic properties.

The remainder of the paper is organized as follows. Section 2 presents the FVECM proposed in Avarucci (2007) and its properties. In Section 3 the identification issues are discussed. Section 4 introduce a 4 step algorithm to estimate the parameters of this model. Section 5 describes the small sample properties and the asymptotic properties of the parameter estimators by means of a Monte Carlo experiment. Section 6 concludes. The Appendix gives details on the GLS procedure in Mosconi and Paruolo (2014). Tables and Figures follow.

2 Avarucci model of Fractional Cointegration

The model of Avarucci (2007) is given by the following dynamics

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

where $B_j = -(\alpha\beta')A_j$. The vector of variables X_t is $p-$dimensional, the loadings α and the cointegrating relations β are $p \times r$ matrices where $0 \leq r \leq p$, A_j and $B_j, j = 1, \ldots, k$ are $p \times p$ matrices of the short run dynamics. The lag operator L is such that $LX_t = X_{t-1}$, the fractional difference is given by the binomial expansion $\Delta^d := (1-L)^d = \sum_{j=0}^{\infty}(-1)^j \binom{d}{j}L^j$, $\binom{d}{j}$ is the generalized binomial coefficient and $d \geq b > 0$.

The model can be written in another form. The form proposed is coherent with the representation in Johansen (2008) . The model in (1) can be reformulated as:

$$\Delta^d Y_t = \alpha\beta'\Delta^{d-b}L_b Y_t + \varepsilon_t,$$

where $L_b := 1 - \Delta^b$ is the fractional lag operator and with the restriction $Y_t = \left(I_p - \sum_{j=1}^{k} A_j L^j\right)X_t$, or more explicitly by

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

This model is characterized by a different (and more complicated) lag structure than the model proposed in Granger (1986)

$$\Delta^d X_t = \alpha\beta'\left(\Delta^{d-b} - \Delta^d\right)X_{t-1} + \sum_{j=1}^{k}\Gamma_j L^j \Delta^d X_t + \varepsilon_t, \tag{3}$$

and Johansen (2008, 2009),

$$\Delta^d X_t = \alpha\beta'(\Delta^{d-b} - \Delta^d)X_t + \sum_{j=1}^{k}\Gamma_j L^j(\Delta^{d-b} - \Delta^d)X_t + \varepsilon_t. \tag{4}$$

In fact model (1) contains both the usual lags based on a standard lag operator present in Granger's model (3) and fractional lags present in the model of (4) Johansen. In the particular case of $d = b = 1$ all 3 models reduce to the standard ECM. When the short run dynamics components

are made equal to zero (i.e. $k = 0$) the models are not present, apart for the initial values. The solution of the Johansen's model depends on the initial values, for $t < 0$, while the model of Avarucci implicitly assumes that the process starts in $t = 0$.

The moving average representation (MA) of the model (1) is given in Avarucci (2007). Following his Theorem 2.2, X_t has the representation

$$X_t = C\Delta_+^{-d}\varepsilon_t + C^*\Delta_+^{-d+b}\varepsilon_t + \Delta_+^{-d+2b}\sum_{j=1}^{t-1}\Phi_j\varepsilon_{t-j}, \tag{5}$$

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 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.

The proof of Theorem 2.2 is largely based on Theorem 8 in Johansen (2008) and the MA representation (5) is based on the solution given in Johansen (2008, 2009):

$$X_t = C\Delta_+^{-d}\varepsilon_t + \Delta_+^{-(d-b)}Y_t^+ + \mu_t, \tag{6}$$

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 $X_t \sim I(d)$, while $\Delta^b X_t$, $\beta'X_t \sim I(d-b)$.

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

The nesting structure of the Avarucci model also differs from the nesting structure of the Johansen (2008, 2009) model and it follows the VAR structure. If we define the model

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

4

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

$$
\begin{array}{ccccc}
\mathcal{H}_{0,0} & \subset & \mathcal{H}_{0,1} & \subset & \mathcal{H}_{0,2} \\
\cap & & \cap & & \cap \\
\mathcal{H}_{1,0} & \subset & \mathcal{H}_{1,1} & \subset & \mathcal{H}_{1,2} \\
\cap & & \cap & & \cap \\
\mathcal{H}_{2,0} & \subset & \mathcal{H}_{2,1} & \subset & \mathcal{H}_{2,2}
\end{array}
$$

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 model is more complicated, see Carlini and Santucci de Magistris (2014) for the details. Therefore, the inference for these two models should be different, despite of the fact that they both generate the same class of processes.

3 Identification issues

Let us consider the model:

$$
\mathcal{H}_2 : \Delta^{d-b} \left(\Delta^b I_p - \alpha\beta' L_b \right) \left(I_p - \sum_{j=1}^{2} A_j L^j \right) X_t = \varepsilon_t. \tag{7}
$$

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

The two sub-models

$$
\mathcal{H}_2^{(0)} \quad : \quad \Delta^{d_0 - b_0} \left(\Delta^{b_0} I_p - \tilde{\alpha}\tilde{\beta}' L_{b_0} \right) \left(I_p - (I_p + \tilde{A}_1)L + \tilde{A}_1 L^2 \right) X_t = \varepsilon_t. \tag{8}
$$

$$
\mathcal{H}_2^{(1)} \quad : \quad \Delta^{d_1 - b_1} \left(\Delta^{b_1} I_p - \bar{\alpha}\bar{\beta}' L_{b_1} \right) \left(I_p - \bar{A}_1 L \right) X_t = \varepsilon_t. \tag{9}
$$

can be reparameterized as in Carlini and Santucci de Magistris (2014).

The sub-model $\mathcal{H}_2^{(0)}$ in equation (8) can be written as:

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

or equivalently as

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

Therefore,

$$\mathcal{H}_2^{(0)} : \Delta^{d_0 - b_0 + 1} \left(\Delta^{b_0} I_p - \tilde{\alpha} \tilde{\beta}' L_{b_0} \right) \left(I_p - \tilde{A}_1 L \right) X_t = \varepsilon_t \tag{10}$$

Now, let us compare the sub-models (9) and (10). It is clear that the equations (10) and (9) reparameterize 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, $\mathcal{H}_{2,0} = \mathcal{H}_{2,1}$ iff $d_0 + 1 = d_1$.

Furthermore, note that the sub-model $\mathcal{H}_2^{(0)}$ is a sub-model of the model $\mathcal{H}_2$, where 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 $\mathcal{H}_2$ where we impose the restriction $A_2 = 0$.

Therefore, the parameter b is always identified. It seems that an identification problem could occur as in Johansen (2008) and Johansen and Nielsen (2010) for certain values of d. If we want to avoid identification issues, we only need to assume that the characteristic polynomial

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

has roots outside of the unit circle, which is already assumed in Avarucci (2007). Therefore the identification problem for the parameter d is not present.

However, the problem of identification can arise when $\alpha \beta' = 0$. In this situation, the eq. (2) is given by

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

and the parameter b is not identified. This special feature of the model has been used in Łasak (2010) to propose a sup-test for no cointegration.

Furthermore, suppose that $b = 1$ and $\alpha \beta'$ is a full rank matrix, then we have

$$\Delta^{d-1}(\Delta I_p - \alpha \beta' L)(I_p - A_1 L) X_t = \varepsilon_t$$

This is equal to

$$\Delta^{d-1}(I_p - (\alpha \beta' + I_p) L)(I_p - A_1 L) X_t = \varepsilon_t$$

6

If $\alpha\beta' + I_p = A_1$, then an identification problem arises because the two factors of the last equation commute. Therefore the cases of no-cointegration and the cointegration of a full rank should be treated separately.

4   Estimation Algorithm

In this section we propose an estimation procedure for the model of Avarucci (2007). To keep the presentation simple, we consider the case $d = b$. Using the approach of Carlini and Mosconi (2014), the likelihood function can be maximized with a switching algorithm. The model is given by:

$$(\Delta^d I_p - \alpha_1\beta_1' L_d) \cdot A(L) \cdot X_t = \varepsilon_t \quad \varepsilon_t \sim N(0,\Omega)$$

where $A(L) = I_p - A_1 L - \cdots - A_k L^k$ and α_1 and β_1 are a $p \times r$ matrices.

The model can be written as:

$$\Delta^d X_t = A\Delta^d Z_t + \alpha_1\beta_1' L_d X_t - \alpha_1\beta_1' A L_d Z_t + \varepsilon_t \quad t = 1,\ldots,T$$

where

$$Z_t = \begin{bmatrix} LX_t \\ L^2 X_t \\ \vdots \\ L^k X_t \end{bmatrix}$$

and

$$A = [A_1 : A_2 : \cdots : A_k]$$

where A is the horizontal concatenation of the matrices $A_1,\ldots,A_k$.

The likelihood is maximized using the following switching procedure:

- A- step: given α_1, β_1 and Ω we estimate A. Defining:

$$Y_{At} = \Delta^d X_t - \alpha_1\beta_1' L_d X_t$$

$$W_{At} = \begin{bmatrix} \Delta^d Z_t \\ L_d Z_t \end{bmatrix}$$

and hence

$$Y_{At} = [I : -\alpha_1 \beta_1](I_2 \otimes A)W_{At} + \varepsilon_t$$

This can be solved by GLS on A (see Mosconi and Paruolo (2014) and Appendix). So, in this case, $H = [I : -\alpha_1 \beta_1]$ and $\theta' = (I_2 \otimes A)$.

- β_1-step: given A and Ω and α_1, we estimate β_1. Defining

$$Y_{bt} = \Delta^d X_t - A\Delta^d Z_t$$

and

$$W_{bt} = L_d X_t - AL_d Z_t,$$

we solve the following problem:

$$Y_{bt} = \alpha_1 \beta_1' W_{bt} + \varepsilon_t$$

This can be solved by GLS as in Mosconi and Paruolo (2014), where $H = \alpha_1$ and $\theta' = \beta_1'$.

- α_1−step: given A, Ω and β_1, we estimate α_1. Defining

$$Y_{at} := Y_{bt}$$

and

$$W_{at} = \beta_1' W_{bt}$$

we solve the following problem:

$$Y_{at} = \alpha_1 W_{at} + \varepsilon_t.$$

Again, this can be solved by GLS as in Mosconi and Paruolo (2014), where $H = I_2$ and $\theta' = \alpha_1$.

- Ω−step: given a_1, β_1, A we estimate Ω. The likelihood is maximized with respect to Ω by

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

defining $\varepsilon_t = \Delta^d X_t - A\Delta^d Z_t - a_1 \beta_1 L_d X_t + a_1 \beta_1 AL_d Z_t$.

Iterating these steps and changing the parameter d in a maximization algorithm, the likelihood is maximized and we get $\hat{d}, \hat{a}_1, \hat{b}_1, \hat{A}, \hat{\Omega}$.

5   Simulation experiment.

The Monte Carlo works as follows. We generate $N = 1,000$ Monte Carlo replication from the following DGP

$$(\Delta^{d_0} I_p - \alpha_0 \beta_0' L_{d_0})(I_p - A_{1,0}L - A_{2,0}L^2)X_t = \varepsilon_t \quad \varepsilon_t \sim i.i.d.N(0, \Omega_0), \quad t = 1, \ldots, T$$

where

$$A_0 = [A_{1,0} : A_{2,0}] = \begin{bmatrix} 0.2 & -0.6 & -0.2 & 0.1 \\ -0.3 & 0.5 & 0.2 & -0.5 \end{bmatrix},$$

$$\alpha_0 = \begin{bmatrix} -0.5 \\ 0.5 \end{bmatrix}, \quad \beta_0 = \begin{bmatrix} 1 \\ -0.2 \end{bmatrix} \quad \text{and} \quad \Omega_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

and $T = 100, 250, 10000$. The values chosen for d_0 are 0.3, 0.6 and 0.9. The data are generated with a modified version of the program provided in Jensen (2014).

For each run of the Monte Carlo, we fit the model

$$(\Delta^d I_p - \alpha \beta' L_d)(I_p - A_1 L - A_2 L^2)X_t = \varepsilon_t, \quad \varepsilon_t \sim iidN(0, \Omega) \tag{11}$$

using the switching algorithm described in Section 4. When we maximize the likelihood, the initial value for the parameter d is chosen in a neighbourhood of the true value d_0, while the initial values for α, β and Ω are imposed to be α_0, β_0 and Ω_0. The maximization routine climbs the likelihood over values of d in $[0.1, 2]$ in order to avoid negative - or close to zero - $\hat{d}$ estimates. Hence, we collect the estimates $\hat{d}, \hat{\alpha}, \hat{\beta}, \hat{A}_1, \hat{A}_2$ and $\hat{\Omega}$ for each iteration of the Monte Carlo.

To simplify the exposition, we introduce new notation for the elements of the matrices in the model (11). We call the elements of the matrices

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

where $\omega_{21} = \omega_{12}$.

When the data are generated with fractional parameters $d_0 = 0.3$ and $d_0 = 0.6$, the bias and the standard deviations of the Monte Carlo estimates are given in the Tables 1-12 in Appendix. From these Tables emerges that if T increases the bias and the standard error of α, β, Ω decrease, while if T increases just the standard error of A_1 and A_2 decrease.

The plots in the appendix are the Monte Carlo distributions of some of the parameters estimates in model (11), smoothed by a Gaussian Kernel when $d_0 = 0.6$ and $T = 100$ and $T = 10,000$.

The top-left plot in Figure 1 shows that the distribution of $\hat{d}$ when $T = 250$ has the estimated values centred in two zones. This is due to the fact that the maximization algorithm converged at values of d on the boundary of the interval $d = [0.1, 2]$, giving as a result a bimodal distribution where the first mode is in 0.1 and the second one centred close to $d = 0.6$. This happens when the sample is small because the estimates of $\hat{d}$, sometimes, tend to assume values close to zero or negative. This also explains why the distributions of the parameters α, β, A_1 and A_2 are skewed.

Instead the plot in Figure 2 shows that the distribution of $\hat{d}$ when $T = 10,000$ is uni-modal and symmetric as the distributions of all the other parameters.

Another characteristic of the algorithm is that if we generate data with $d_0 = 0.9$ and a large T, the maximization routine rarely is not able to converge, stopping at $\hat{d} = 0.1$ with an allert of non-convergence. We noted that the number of iterations for the convergence of the switching algorithm when d is close to 0.9 becomes so big that we could need more than 10,000 iterations even if the starting values of α, β and Ω are imposed on their true values. This phenomenon becomes cumbersome in terms of speed and in terms of reliability of the estimates of the parameters. Hence, further research has to be done to improve the convergence properties of this algorithm when the data generating process is generated by a fractional parameter d_0 close to 1.

6   Conclusions

In this paper we consider estimation of an alternative fractionally cointegrated model that has been proposed in Avarucci (2007) and Avarucci and Velasco (2010). This model is characterized by a richer lag structure than models previously proposed by Granger (1986) and Johansen (2008, 2009). We present a 4 step algorithm to estimate this model. The algorithm is based on the approach of Carlini and Mosconi (2014) that maximizes likelihood function using a switching algorithm and the

GLS procedure as in Mosconi and Paruolo (2014). We study small sample properties of the estimates of all the parameters of fractionally cointegrated model obtained using the proposed algorithm. We also demonstrate by a Monte Carlo experiment that the proposed procedure provides estimates of the long run parameters of the fractionally cointegrated system that are consistent and unbiased.

Appendix

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 X_t are respectively $p_y \times 1$ and $p_x \times 1$ vectors, H and θ are respectively $p_y \times r$ and $p_x \times r$ matrices. Then, we can estimate ψ as

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

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

Tables

	d	α_1	α_2	β_1	ω_{11}	ω_{12}	ω_{22}
Bias	0.029	-0.584	2.110	-4.143	-0.065	0.004	-0.070
St. Error	0.292	3.834	6.245	88.426	0.142	0.098	0.135

Table 1: Bias and Standard Error when the DGP is generated with $d_0 = 0.3$ and $T = 100$

	$a_{11}^{(1)}$	$a_{12}^{(1)}$	$a_{21}^{(1)}$	$a_{22}^{(1)}$	$a_{11}^{(2)}$	$a_{12}^{(2)}$	$a_{21}^{(2)}$	$a_{22}^{(2)}$
Bias	0.480	1.200	0.415	-0.445	0.169	-0.158	-0.089	0.661
St. Error	0.618	0.760	0.938	0.691	0.557	0.370	0.644	0.589

Table 2: Bias and Standard Error when the DGP is generated with $d_0 = 0.3$ and $T = 100$

	d	α_1	α_2	β_1	ω_{11}	ω_{12}	ω_{22}
Bias	0.023	-0.398	0.635	-2.252	-0.027	0.002	-0.024
St. Error	0.217	1.447	2.013	108.134	0.089	0.063	0.088

Table 3: Bias and Standard Error when the DGP is generated with $d_0 = 0.3$ and $T = 250$

	$a_{11}^{(1)}$	$a_{12}^{(1)}$	$a_{21}^{(1)}$	$a_{22}^{(1)}$	$a_{11}^{(2)}$	$a_{12}^{(2)}$	$a_{21}^{(2)}$	$a_{22}^{(2)}$
Bias	0.459	1.238	0.683	-0.467	0.169	-0.138	-0.181	0.538
St. Error	0.314	0.326	0.351	0.336	0.259	0.190	0.267	0.232

Table 4: Bias and Standard Error when the DGP is generated with $d_0 = 0.3$ and $T = 250$

	d	α_1	α_2	β_1	ω_{11}	ω_{12}	ω_{22}
Bias	-0.001	-0.002	0.007	-0.002	-0.001	-0.001	-0.001
St. Error	0.026	0.058	0.060	0.105	0.014	0.010	0.014

Table 5: Bias and Standard Error when the DGP is generated with $d_0 = 0.3$ and $T = 10^5$

	$a_{11}^{(1)}$	$a_{12}^{(1)}$	$a_{21}^{(1)}$	$a_{22}^{(1)}$	$a_{11}^{(2)}$	$a_{12}^{(2)}$	$a_{21}^{(2)}$	$a_{22}^{(2)}$
Bias	0.401	1.299	0.798	-0.500	0.200	-0.101	-0.200	0.501
St. Error	0.022	0.020	0.020	0.024	0.018	0.018	0.018	0.016

Table 6: Bias and Standard Error when the DGP is generated with $d_0 = 0.3$ and $T = 10^5$

	d	α_1	α_2	β_1	ω_{11}	ω_{12}	ω_{22}
Bias	-0.039	-0.331	2.002	-0.649	-0.063	0.004	-0.073
St. Error	0.404	2.732	5.298	21.586	0.142	0.098	0.134

Table 7: Bias and Standard Error when the DGP is generated with $d_0 = 0.6$ and $T = 100$

	$a_{11}^{(1)}$	$a_{12}^{(1)}$	$a_{21}^{(1)}$	$a_{22}^{(1)}$	$a_{11}^{(2)}$	$a_{12}^{(2)}$	$a_{21}^{(2)}$	$a_{22}^{(2)}$
Bias	0.521	1.279	0.423	-0.445	0.117	-0.168	-0.114	0.630
St. Error	0.535	0.545	0.811	0.676	0.466	0.342	0.569	0.545

Table 8: Bias and Standard Error when the DGP is generated with $d_0 = 0.6$ and $T = 100$

	d	α_1	α_2	β_1	ω_{11}	ω_{12}	ω_{22}
Bias	-0.027	-0.202	0.684	0.082	-0.026	0.002	-0.025
St. Error	0.281	0.997	2.036	7.082	0.089	0.063	0.087

Table 9: Bias and Standard Error when the DGP is generated with $d_0 = 0.6$ and $T = 250$.

	$a_{11}^{(1)}$	$a_{12}^{(1)}$	$a_{21}^{(1)}$	$a_{22}^{(1)}$	$a_{11}^{(2)}$	$a_{12}^{(2)}$	$a_{21}^{(2)}$	$a_{22}^{(2)}$
Bias	0.492	1.291	0.657	-0.455	0.136	-0.150	-0.178	0.538
St. Error	0.287	0.274	0.312	0.394	0.250	0.193	0.254	0.215

Table 10: Bias and Standard Error when the DGP is generated with $d_0 = 0.6$ and $T = 250$.

	d	α_1	α_2	β_1	ω_{11}	ω_{12}	ω_{22}
Bias	0.000	-0.003	0.003	-0.000	0.000	-0.000	-0.000
St. Error	0.024	0.031	0.028	0.047	0.014	0.010	0.014

Table 11: Bias and Standard Error when the DGP is generated with $d_0 = 0.6$ and $T = 10^5$

	$a_{11}^{(1)}$	$a_{12}^{(1)}$	$a_{21}^{(1)}$	$a_{22}^{(1)}$	$a_{11}^{(2)}$	$a_{12}^{(2)}$	$a_{21}^{(2)}$	$a_{22}^{(2)}$
Bias	0.401	1.300	0.799	-0.500	0.199	-0.101	-0.199	0.500
St. Error	0.023	0.020	0.019	0.027	0.019	0.020	0.019	0.016

Table 12: Bias and Standard Error when the DGP is generated with $d_0 = 0.6$ and $T = 10^5$
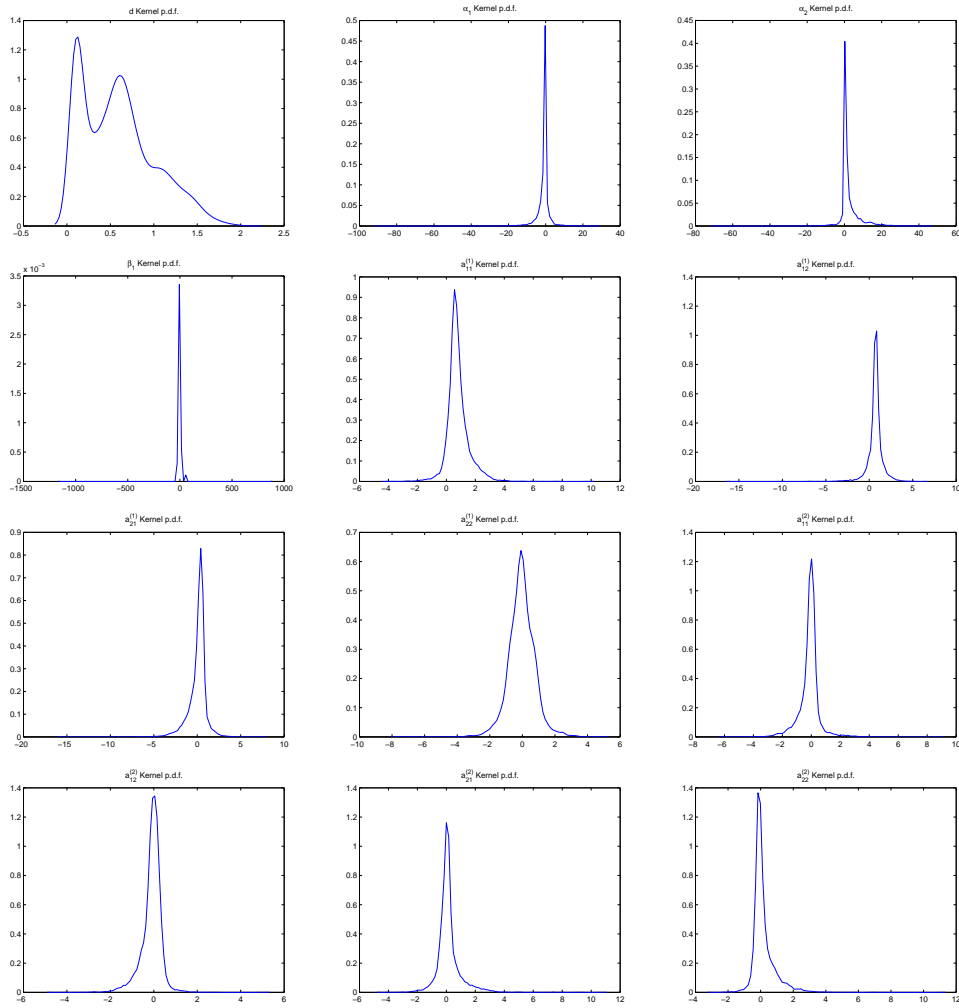
Figures

Figure 1: Smoothed Kernel distribution of the estimators of the model (11) when $d_0 = 0.6$ and $T = 100$
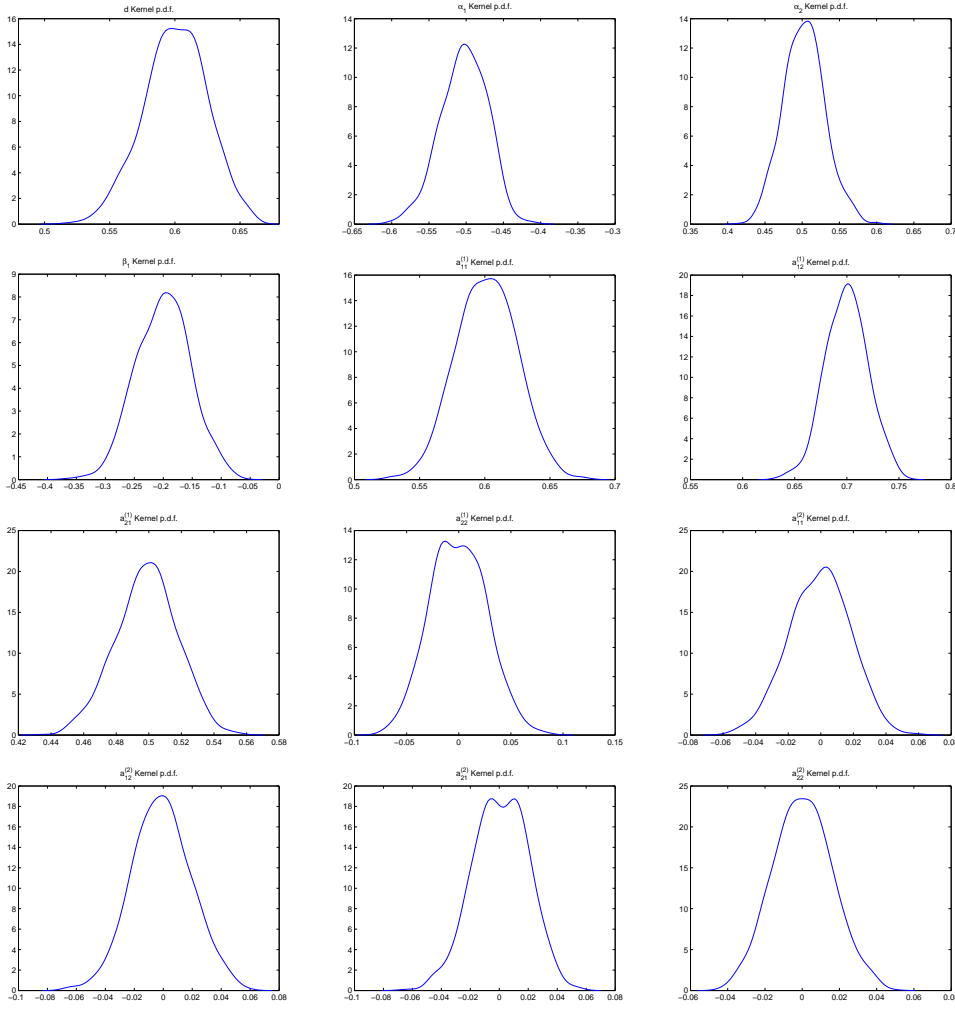
14

Figure 2: Smoothed Kernel distribution of the estimators of the model (11) when $d_0 = 0.6$ and $T = 10^5$

References

[1] Avarucci, M. (2007), Three Essays on Fractional Cointegration, PhD 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. (2014), On the identification of fractionally cointegrated VAR models with the F(d) condition, CREATES Research Paper 2013-44.

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

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

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

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

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

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

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

[12] 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.