Economics
The Open-Access, Open-Assessment E-Journal

Vol. 10, 2016-26 | October 31, 2016 | http://dx.doi.org/10.5018/economics-ejournal.ja.2016-26

Automatic Identification of General Vector Error Correction Models

Ignacio Arbués, Ramiro Ledo, and Mariano Matilla-García

Abstract
There are a number of econometrics tools to deal with the different types of situations in which cointegration can appear: I(1), I(2), seasonal, polyno- mial, etc. There are also different kinds of Vector Error Correction models related to these situations. The authors propose a unified theoretical and practical framework to deal with many of these situations. To this aim: (i) they introduce a general class of models and (ii) provide an automatic method to identify models, based on estimating the Smith form of an autoregressive model. Their simulations suggest the power of the new proposed methodology. An empirical example illustrates the methodology.

(Published in Special Issue Recent Developments in Applied Economics)

JEL C01 C22 C32 C51 C52
Keywords Time series; unit root; cointegration; error correction; model identification; Smith form

Authors
Ignacio Arbués, Ministerio de Industria, Energía y Turismo and Instituto Complutense de Análisis Económico, Madrid, Spain
Ramiro Ledo, UNED and Universidad Complutense de Madrid, Facultad de Ciencias Económicas y Empresariales, Departamento de Economía Aplicada II, Spain
Mariano Matilla-García, ✉ UNED, Departamento de Economía Aplicada Cuantitativa, Paseo Senda del Rey, 11, 28040, Madrid, Spain,

Citation Ignacio Arbués, Ramiro Ledo, and Mariano Matilla-García (2016). Automatic Identification of General Vector Error Correction Models. *Economics: The Open-Access, Open-Assessment E-Journal*, 10 (2016-26): 1—41. http://dx.doi.org/10.5018/economics-ejournal.ja.2016-26

Received June 8, 2016 **Published as Economics Discussion Paper** July 5, 2016
Revised October 10, 2016 **Accepted** October 18, 2016 **Published** October 31, 2016

1 Introduction

The basis of the theory of cointegration was laid out in a surge of articles about the late eighties and early nineties, preceded by the seminal paper by Granger (1981). By the mid-nineties, there was a relatively complete theoretical framework for I(1) (Engle and Granger, 1987; Johansen, 1988; Johansen and Juselius, 1990), I(2) cointegration (Johansen, 1992; Paruolo, 1996), multicointegration (Granger and Lee, 1989) and seasonal cointegration (Hylleberg et al., 1990).

Since then, much of the attention in cointegration theory has been focused on panel data (Levin and Lin, 1993; Levin et al., 2002) or on fractional cointegration (Engle and Granger, 1987; Granger and Joyeux, 1980). Latest developments are focused on particular cases as cointegration in dynamic panels (Yu and Lee, 2010), or panel data multicointegration (Worthington and Higgs, 2010).

Also, some contributions have been made to the foundations and algebraic theory of multivariate integrated processes (Franchi, 2006, 2007). A concern that may explain the persistence of interest in the algebra of cointegration is the perceived necessity of a unifying theoretic framework for all situations.

In this article, we contribute in two ways to facilitate a more unified treatment of cointegration. First, in Section 2 we provide a general theoretical framework that covers any process that can be represented by an autoregressive model with unit roots. This framework has the following elements: (i) a classification of all such models based on the Smith form (SF); (ii) a class of General Vector Error Models (GVEC) that correspond to every unstable autoregressive model and (iii) a consistency result for the Least Squares estimator of the GVEC models. The first two points are in fact straightforward generalizations of the kind of reasoning that Hylleberg et al. (1990) used to find a seasonal vector error correction model by assuming that the process had a Wold representation with a Smith form of a particular shape. By removing the restriction on the SF, we find a general family of VEC models (General VEC or GVEC, hereinafter). The GVEC family covers the following cases: I(1), I(2), multicointegration (with a caveat that will be explained later), polynomial and seasonal cointegration. On the other hand, this framework does not include panel data or fractional integration. At this point, we do not know how difficult will be to generalize our results in those directions, but it seems that

fractional integration poses more difficulties from the theoretical point of view, whereas the algorithms we use are not well-suited to panel data.

The second contribution (Section 3) is an automatic method to identify GVEC models based on an estimate of the SF of the autoregressive model. We show by means of Monte Carlo simulations (Section 4) that our method works better than the Johansen test in the restricted cases in which the latter applies. But of course, our method has also the advantage that it can detect other situations. This is the case in the practical example of Section 5, where our algorithm detects both seasonal cointegration and higher order of integration. Therefore, in series that have a marked seasonal behavior, with our method it is no longer necessary to do a cointegration analysis with the seasonally adjusted series, as is sometimes done.

2 General Vector Error Correction Models

In this section, we generalize some fundamental results of the representation of cointegrated variables by means of VEC models. We will denote polynomials and power series using the inderminate z. When we use them to define autoregressive or ARMA models, we substitute the backshift operator B for z, so for example $\phi(z) = \phi_0 + \ldots + \phi_p z^p$, but $\phi(B)y_t = \phi_0 y_t + \ldots + \phi_p y_{t-p}$. Sometimes, when the context makes clear that ϕ is a polynomial, we can drop the indeterminate for ease of notation. The same applies for vectors and matrices whose entries are polynomials or power series.

Assumption 1. *Let y_t be a $n \times 1$ random vector, s be a positive integer and $d(z)$ a real polynomial such that all its roots belong to $\{\omega_k\}_{k=0}^{s-1}$ where $\omega_k = \exp 2\pi ik/s$ and $d(B)y_t$ is stationary and purely nondeterministic.*

The fact that we limit the roots of d to that set is just for convenience, as the most common cause of unit roots other than the unity is seasonal integration. Hence, s can be interpreted as the number of observations per year (or per week in the case of daily data). All the subsequent developments can easily be adapted to the general case, although most applications do not require that.

Under assumption 1, if the multiplicity of ω_k is d_k we can say that y_t is integrated of order d_k with respect to ω_k and denote this property as $y_t \sim I_k(d_k)$.

We say that y_t is polynomially cointegrated of order j at ω_k when there exists a polynomial vector $a(z)$ such that $a(B)'y_t$ is $I_k(d_k - j)$. There can be several cointegration relationships for y_t. For each k and j, let us say there is a set $\mathscr{A}$ of such vectors with exactly $r_{k,j}$ elements and $\{a(\omega_k)\}_{a\in\mathscr{A}}$ is linearly independent. For the sake of generality, we do not require that all the components of y_t have the same order of integration, so there may be trivial cointegration relationships. Then, if $y_t = (y_{1,t}, y_{2,t})'$, where $y_{1,t} \sim I(1)$ and $y_{2,t} \sim I(0)$, a trivial cointegrating vector would be $(0,1)'$. For convenience, we can say that y_t is cointegrated of order 0 with rank n, so $r_{k,0} = n$. Clearly $r_{k,j} \geq r_{k,j+1}$ and there is some j from which $r_{k,j} = 0$ onwards. We call $r_{k,j}$ cointegration ranks.

Assumption 2. *The Wold representation of $d(B)y_t$ is $\Psi(B)\varepsilon_t$, where $\Psi(z)$ is rational, that is, its entries are polynomial fractions.*

Now, we will present our generalization of the Granger Representation Theorem (GRT) as in Engle and Granger (1987). First, we adapt statements (1)–(3).

Proposition 1. *If assumptions 1 and 2 hold and y_t has cointegration ranks $r_{k,j}$, then*

(a) $\Psi(z) = U(z)D(z)V(z)$, *where* $\det U(z)$ *and* $\det V(z)$ *have no roots in the closed unit circle,* $D(z) = D_0(z)\cdot\ldots\cdot D_{s-1}(z)$ *and*

$$D_k(z) = \begin{pmatrix} I_{s_{k,0}} & 0 & 0 & \ldots \\ 0 & \left(1-\omega_k^{-1}z\right)I_{s_{k,1}} & 0 & \ldots \\ 0 & 0 & \left(1-\omega_k^{-1}z\right)^2 I_{s_{k,2}} & \ldots \\ \ldots & \ldots & \ldots & \ldots \end{pmatrix}$$

where $s_{k,j} = r_{k,j} - r_{k,j+1}$. $D(z)$ is called the Smith form[1] *of $\Psi(x)$. Conversely, if the SF is as above, then y_t has cointegration ranks $r_{k,j}$.*

(b) There is a VARMA representation $A(B)y_t = m(B)\varepsilon_t$ such that $m(z)$ is a polynomial. If D divides d, then the SF of $A(z)$ is $d(z)D(z)^{-1}$ and $m(z)$ has no unit roots. In this case, there is also the infinite VAR representation $\Phi(B)y_t = m(B)^{-1}A(B)y_t = \varepsilon_t$.

[1] We go back to this in Section 3.1.

(c) *There are full rank $n \times r_{k,j}$ polynomial matrices $\alpha(z)$ and $\gamma(z)$ such that $\alpha(z)'\Psi(z)$ and $\Psi(z)\gamma(z)$ have jth-order zeros at ω_k.*

Proof. See Appendix A. □

Statements (a) and (b) are of paramount importance for our results, because they mean that the cointegration structure of y_t can be entirely obtained from the SF of $\Psi(z)$ or $\Phi(B)$. In order to distinguish both diagonal matrices we can use $D^{\Psi}(z)$ and $D^{\Phi}(z)$. Let $\delta_1(z),\ldots,\delta_h(z)$ be the distinct diagonal elements of $D^{\Phi}(z)$. For $j=1,\ldots,h-1$, we denote by m_{jk} the multiplicity of ω_k in δ_{j+1}/δ_j. If we define

$$\Delta_{jk\ell}(z) = \begin{cases} \frac{\delta_{j+1}(z)}{(1-\omega_k^{-1}z)^{\ell}} & \mathrm{Im}\omega_k = 0 \\ \frac{\delta_{j+1}(z)}{(1-2\mathrm{Re}\omega_k z+z^2)^{\ell}} & \mathrm{Im}\omega_{jk} > 0 \end{cases}$$

and $\Delta_h = \delta_h$. Now, we can generalize statement (4) of GRT.

Proposition 2. *In the conditions of Proposition 1, y_t satisfies the model*

$$\Gamma(B)\Delta_h(B)y_t = \sum_{j=1}^{h-1}\Big\{\sum_{\mathrm{Im}\omega_{jk}=0}\sum_{\ell=1}^{m_{jk}}\Pi_{k\ell}^{(j)}\Delta_{jk\ell}(B) + \sum_{\mathrm{Im}\omega_{jk}>0}\sum_{\ell=1}^{m_{jk}}(\Pi_{k\ell}^{(j,-)}+\Pi_{k\ell}^{(j,+)}B)\Delta_{jk\ell}(B)\Big\}y_{t-1} + m(B)\varepsilon_t, \tag{1}$$

where $\Gamma(z)$ is a polynomial matrix. If $D^{\Psi}(z)|d(z)$, then there is a representation with $\Gamma(z)$ an infinite power series and $m(B)=1$.

Proof. See appendix A. □

We will see now some examples of well-known models that are particular cases of GVEC. When y_t is $I(1)$, the simplest case occurs when $D^{\Phi}(z) = \mathrm{diag}(1-z,\ldots,1-z)$. Then, the components of y_t are not actually cointegrated. A stable VAR model can be fitted for their differences $(\nabla y_t^1,\ldots,\nabla y_t^n)$. In a similar fashion, if y_t is seasonally integrated and $D^{\Phi}(z) = \mathrm{diag}(1-z^s,\ldots,1-z^s)$, a stable VAR is appropriate for the seasonal differences.

We encounter classical $I(1)$ cointegration when

$$D^{\Phi}(z) = \begin{pmatrix} I_r & 0 \\ 0 & (1-z)I_{n-r} \end{pmatrix}$$

and $0 < r < n$. Then (1) is the usual VEC model $\Gamma(B)\nabla y_t = \Pi y_{t-1} + \varepsilon_t$, where $\text{rank}\Pi = r$. When the series y_t is $I(2)$, we find

$$D^{\Phi}(z) = \begin{pmatrix} I_r & 0 & 0 \\ 0 & (1-z)I_{r'} & 0 \\ 0 & 0 & (1-z)^2 I_{r''} \end{pmatrix},$$

and then associate GVEC model is $\Gamma(B)\nabla^2 y_t = \Pi^{(1)}_{1,1} y_{t-1} + \Pi^{(1)}_{1,2}\nabla y_{t-1} + \varepsilon_t$.

If we allow unit roots other than unity, in particular seasonal roots, the picture is much more complicated. However, there is a particular case that has been already described in the literature. Let us assume that y_t represents quarterly data, so $s = 4$. If

$$D^{\Phi}(z) = \begin{pmatrix} I_r & 0 \\ 0 & (1-z^4)I_{n-r} \end{pmatrix},$$

then (1) boils down to the model in section 4 of Hylleberg et al. (1990). With our notation,

$$\Gamma(B)(1-B^4)y_t = \Pi^{(1)}_{0,1}(1+B+B^2+B^3)y_{t-1} +$$
$$\left(\Pi^{(1,+)}_{1,1} + \Pi^{(1,-)}_{1,1}B\right)(1-B^2)y_{t-1} + \Pi^{(1)}_{2,1}(1-B-B^2+B^3)y_{t-1} + \varepsilon_t.$$

As we mentioned above, multicointegration also fits in our framework. Following Granger and Lee (1989), consider the case of the production (p_t) and sales (s_t). They are cointegrated I(1) variables with cointegrating vector $(1,-1)'$, so $p_t - s_t$ is stationary. In addition, $\nabla^{-1}(s_t - p_t)$ is cointegrated with s_t (again, with $(1,-1)'$). This relation holds, for example, when $\nabla(p_t, s_t)' = \Psi(B)\varepsilon_t$ and

$$\Psi(B) = \begin{pmatrix} \nabla & \nabla+1 \\ \nabla & 1 \end{pmatrix}.$$

If we calculate the SF of $\Psi(B)$ in $\mathbb{R}[z]$, we get the representation

$$\nabla \begin{pmatrix} p_t \\ s_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ B & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \nabla^2 \end{pmatrix} \begin{pmatrix} \nabla & \nabla+1 \\ 1 & 1 \end{pmatrix} \varepsilon_t. \tag{2}$$

Note that the alternative representation by Haldrup and Salmon (1998),

$$\nabla \begin{pmatrix} p_t \\ s_t \end{pmatrix} = \begin{pmatrix} \nabla & \nabla^{-1}+\nabla^{-2} \\ \nabla & \nabla^{-2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \nabla^2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \varepsilon_t$$

is in a greater ring, that contains ∇^{-1}. The condition $D^{\Psi}|d$ is violated because the second element ∇^2 in the diagonal of the SF of $\Psi(B)$ does not divide the operator ∇ that we use to make (p_t, s_t) stationary. Nevertheless, in order to get a representation without a moving average part this problem can be circumvented by modeling the integrated vector $\nabla(P_t, S_t)' = (p_t, s_t)'$ ($\nabla = \operatorname{lcm}(\nabla, \nabla^2)\nabla^{-1}$). From (2), we get

$$\begin{pmatrix} -\nabla+1 & 1 \\ -1 & \nabla \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \nabla^2 \end{pmatrix} \begin{pmatrix} -B & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} P_t \\ S_t \end{pmatrix} = \varepsilon_t.$$

This is the I(2) cointegration case with $r_2 = 0$. In fact, we can follow the constructive proof of Proposition 2 and arrive to the model

$$\nabla^2 \begin{pmatrix} P_t \\ S_t \end{pmatrix} = -\nabla y_{t-1} + \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} y_{t-1} + \varepsilon_t.$$

Here we can see the two cointegration relations. The second row means that sales and inventory are cointegrated, and the sum of the two rows gives the relation between sales and production.

We will get asymptotic properties of of the least squares estimation of models of the form of (1) in the case that $\Gamma(B)$ is of finite order, but first we need an additional assumption.

Assumption 3. *Let $\mathscr{F}_t$ be the σ–field generated by $\{\varepsilon_s : s \le t\}$. Then, $\mathbb{E}[\varepsilon_t|\mathscr{F}_{t-1}] = 0$, $\sum_{t,i} \mathbb{E}[|\varepsilon_{it}|^{2+\delta}|\mathscr{F}_{t-1}] < \infty$ almost surely for $\delta > 0$ and $\mathbb{V}[\varepsilon_t|\mathscr{F}_{t-1}] = \Sigma$, where Σ is positive definite.*

Proposition 3. *Let y_t satisfy model (1), where $\Gamma(z)$ is a polynomial matrix and ε_t fulfils assumption 3. In addition, for simplicity we assume that $d|D$, so $m(z) = 1$. Let us stack the coefficients of model (1) as $\beta = [\Gamma : \Pi]$ with $\Gamma = [\Gamma_1, \ldots]$, $\Pi = [\Pi^{(1)}, \ldots, \Pi^{(h)}]$ and $\Pi^{(j)} = [\Pi_{11}^{(j)}, \ldots]$ and let $\hat{\beta} = [\hat{\Gamma} : \hat{\Pi}]$ be the least squares estimator of β. Then, $\hat{\beta} \xrightarrow{p} \beta$ and*

$$T^{1/2}(\hat{\Gamma} - \Gamma) \xrightarrow{d} N(0, \Xi) \tag{3}$$

$$L_T(\hat{\Pi} - \Pi) = O_p(1) \tag{4}$$

where $L_T = \operatorname{diag}(T^{q_{jk\ell}} I_n)_{jk\ell}$, $q_{jk\ell}$ is the minimum multiplicity of the unit roots of $\Delta_{jk\ell}(z)$.

Proof. See Appendix A. □

It is likely that consistency could be proved in the case that the true $\Gamma(B)$ has infinite order. That would require the order of the estimate $\hat{\Gamma}(B)$ to diverge to infinity at a certain rate, as in Lewis and Reinsel (1985).

This result should be extended to allow the presence of moving average terms in the model. The results of Barrio Castro and Osborn (2011), suggest that significant improvements could be achieved with this generalization, that could be attempted following the lines of Yap and Reinsel (1995).

3 Identification

We have built an R package to automatically identify and estimate GVEC models. This package can be obtained from the corresponding author until it is uploaded to a public repository. The main steps of the procedure are: (i) estimate a VAR representation $\Phi(B)y_t = \varepsilon_t$; (ii) estimating the SF of $\Phi(z)$ and (iii) applying Proposition 2 to obtain the GVEC model.

This procedure resembles the unit root determination method of the program TRAMO[2], albeit with the additional complication that polynomial matrices cannot

[2] Current versions of this program, together with SEATS have been developed by Agustín Maravall and his team at the Bank of Spain, upon the programs originally developed by A. Maravall and V. Gómez.

be factorized as simply as polynomials. The method of TRAMO is described in the introductory notes by Maravall (2008).

Steps (i) and (iii) are straightforward. What we need now is a way to estimate the SF of a polynomial or rational matrix. In fact, we will describe a method to estimate the SF in any ring in which we can perform the Euclidean division (Euclidean ring). In Section 3.1 we will describe the algorithm to do that. In Section 3.2 we adapt the algorithm to identify GVEC models.

3.1 Smith form

The existence of the SF of a matrix with elements in a Principal Ideal Domain is guaranteed by Proposition 2.11 in Hungerford (1980).

Let A be a matrix with entries in a ring R and $D = \text{diag}(d_1, \ldots, d_r, 0, ..., 0)$ its SF. Although the SF is not strictly unique, the ideal (d_i) generated by the ith element of the main diagonal of D is unique, so d_i is unique up to multiplication by an invertible element of R. Hence, to achieve uniqueness of D, it is necessary to impose additional constraints. For example, in the ring $\mathbb{R}[z]$ of polynomials over $\mathbb{R}$ we may set the coefficient of the highest order term equal to unity, that is, to force d_i to be monic.

In general, we will denote by R_1 a subset of R such that for any $a \in R$ there is a unique $a_1 \in R_1$ with $a = ua_1$ and u invertible (R_1 always exists because by the axiom of choice, we can pick one element from each equivalence class with respect to the relation $a \sim b \Leftrightarrow \exists u \in R, a = u^{-1}b$). Thus, when $R = \mathbb{R}[z]$, R_1 is the set of the monic polynomials. We define a function $a \mapsto u(a)$ such that $u(a)$ is invertible and $u(a)a \in R_1$.

The proof of the existence of the SF consists of showing that there is an algorithm that by means of elementary operations transforms A into D. We will call that the 'exact algorithm' for reasons that will be obvious later. In the next subsection, we will present a stylized description of the algorithm.

We say that R is a Euclidean Ring with degree function $\varphi : R - \{0\} \mapsto \mathbb{N}$ when: (i) $pq \neq 0$ implies $\varphi(pq) \geq \varphi(p)$ and (ii) for any $p, q \in R$, there are some $m, r \in R$ such that $p = mq + r$ and $r = 0$ or $\varphi(r) < \varphi(r)$.

Let $M_n(R)$ be the ring of the $n \times n$ matrices with elements in R and $A \in M_n(R)$. We will call *admissible operations* the following elementary operations:

(a) Exchange rows i and j.

(b) Exchange columns i and j.

(c) Add c times row i to row j, where c is the quotient of the Euclidean division of a_{ij} by a_{ii} and $a_{ii} \in R_1$.

(d) Add c times column i to column j, where c is the quotient of the Euclidean division of a_{ji} by a_{ii} and $a_{ii} \in R_1$.

(e) Multiply column i or row j by $u(a_{ii})$.

Proposition 4. *There exist algorithms to obtain the SF of a matrix over an Euclidean ring that have the following form:*

(i) $e_0 = 1, A^{(0)} = A$.

(ii) *For* $k > 0$,

$$e_k = f_0(e_{k-1}, A^{(k-1)}) \tag{5}$$
$$A^{(k)} = g(e_{k-1}, A^{(k-1)}) \tag{6}$$

(iii) *The algorithm stops when* $e_k = 0$ *and* $D = A^{(k)}$,

where $f_0 : \mathbb{N} \times M_n(R) \mapsto \mathbb{N}$ *and* $g : \mathbb{N} \times M_n(R) \mapsto M_n(R)$ *satisfy*

(a) *If* $\forall i, j$, *either* $A_{i,j} = B_{ij} = 0$ *or* $A_{i,j}, B_{ij} \neq 0$, *then* $f_0(e, A) = f_0(e, B)$.

(b) $g(e, A)$ *is obtained from* A *by performing an admissible operation that depends on* e.

Proof. See Appendix A. □

Condition (a) means that f_0 depends only on (1) its first argument and (2) which elements of the second argument are zero.

We can identify the algorithm with the functions f_0 and g. The interest of all this is not to prove existence, which is done in Hungerford (1980). The reason to go into this detail is that we need to use later a modification of this scheme. Let us see now with an example, why a modification is necessary.

Assume that the ring R is endowed with a topology (that makes the sum and product continuous) so that we can speak of random elements in R. Then, for a certain matrix A, we may have an estimate obtained with a sample of size T, say $\hat{A}_T$. Furthermore, suppose $\hat{A}_T$ is consistent in probability. We are interested in the SF D of A, so we could in principle, proceed by applying the exact algorithm to $\hat{A}_T$ obtaining a $\hat{D}_T$ with the hope that when $\hat{A}_T \xrightarrow{p} A$, $\hat{D}_T \xrightarrow{p} D$. Unfortunately, it is easy to prove with an example that this does not work. Let us consider the case of a 2×2 matrix with elements in the ring of the polynomials over $\mathbb{R}$. In order to achieve uniqueness, we set the highest order coefficient equal to one. Assume that $\hat{\varepsilon}_T \xrightarrow{p} 0$, $P[\hat{\varepsilon}_T \neq 0] = 1$ and the estimate is such that,

$$\hat{A}_T = \begin{pmatrix} x + \hat{\varepsilon}_T & 0 \\ 0 & x \end{pmatrix} \xrightarrow{p} A = D = \begin{pmatrix} x & 0 \\ 0 & x \end{pmatrix}$$

but if $\hat{D}_T$ is the SF of $\hat{A}_T$, then

$$\hat{D}_T = \begin{pmatrix} 1 & 0 \\ 0 & x^2 + \hat{\varepsilon}_T x \end{pmatrix} \xrightarrow{p} \begin{pmatrix} 1 & 0 \\ 0 & x^2 \end{pmatrix} \neq D$$

Consequently, we have to take a less direct approach to estimate the SF.

We assume now that R is endowed with a modulus function $a \in R \mapsto |a| \in \mathbb{R}$ such that for any $a, b \in R$,

(i) $|a+b| \leq |a| + |b|$.

(ii) $|ab| \leq |a| \cdot |b|$.

(iii) $a = 0$ if and only if $|a| = 0$.

To simplify matters, we may assume also that $|a| = |-a|$. In that case, $|a-b|$ is a metric.

Assumption 4 (Continuity of the Euclidean division)**.** *For any $(p,q) \in R \times R_1$, $\mu \in \mathbb{N}$ and $\varepsilon > 0$, there is some $\delta > 0$ such that for all $(p',q') \in R \times R_1$ such that $\varphi(p) \leq \mu$, $|p-p'| < \delta$ and $|q-q'| < \delta$ imply $|r-r'| < \varepsilon$, when r and r' are the remainders of the divisions of p by q and p' by q' respectively.*

Let $M_n(R)$ be the ring (and R-module) of the $n \times n$ matrices with entries in R. In analogy with the usual notation in linear algebra, we will write for $A \in M_n(R)$,

$$\|A\| = \sup_{\mathbf{a} \in R^n, \mathbf{a} \neq 0} \frac{\|A\mathbf{a}\|}{\|\mathbf{a}\|}. \tag{7}$$

where for $\mathbf{a} = (a_1, \ldots, a_n)$, $\|\mathbf{a}\| = \left(\sum_j |a_j|^2\right)^{1/2}$. The finiteness of the supremum in (7) can be proved in a similar fashion as in the vector space case. From now onwards, convergence of elements of R and matrices will be referred to the topologies generated by $|\cdot|$ and $\|\cdot\|$ respectively.

Now, we can introduce the 'approximate algorithm'.

Definition 1. *For any f_0 and g as in Proposition 4, and for a certain ε, the $\varepsilon-$approximate algorithm is the one obtained when we replace f_0 by f_ε that satisfies that $f_\varepsilon(e,A) = f_0(e,B)$, where $B_{ij} = 0$ when $|A_{ij}| < \varepsilon$ and otherwise, $B_{ij} = A_{ij}$.*

In other words, when the exact algorithm depends on whether $A_{ij} = 0$, the approximate algorithm depends on whether $|A_{ij}| < \varepsilon$.

Let us introduce some notation. For a matrix A, $\mathscr{S}(A,\varepsilon)$ is the output of the $\varepsilon-$approximate algorithm. When there is ambiguity, we will specify the ring in which we are operating as $\mathscr{S}(A,\varepsilon;R)$. Consequently, $\mathscr{S}(A,0)$ is the output of the exact algorithm, that is the SF of A. In particular, if we use the algorithm described in the appendix, we get the unique version in which the elements in the main diagonal of A belong to R_1 (in the polynomial case, they are monic). Let now $\hat{A}_T$ be an estimate of A.

Theorem 1. *If assumption 4 holds, the mapping u is continuous, $\hat{A}_T(z) = A + O_p(\xi_T)$ and*

$$\varepsilon_T \longrightarrow 0 \tag{8}$$

$$\varepsilon_T/\xi_T \longrightarrow \infty, \tag{9}$$

then $\|\mathscr{S}(\hat{A}_T, \varepsilon_T) - \mathscr{S}(A,0)\| = O_p(\xi_T)$.

Proof. See Appendix A. □

3.2 Automatic procedure

The central tool of the procedure is the approximate algorithm of Section 3.1, but how to use it is not completely straightworfard. We have to make some considerations before presenting it in full.

(1) To use Proposition 2.11 from Hungerford (1980) we need to specify in which ring we are considering the entries of the matrices. For the theoretical results of section 2, the ring considered is $\mathbb{R}_1(z) := \{f/g : f, g \in \mathbb{R}[z], \forall z, |z| \leq 1, g(z) \neq 0\}$, that is, the rational fractions without poles at any ω_k. That this is an Euclidean Ring and the continuity of the Euclidean division are both proved in the mathematical appendix (lemmas 1 and 3).

(2) To estimate D^{Φ} when $\Phi(z)$ is a polynomial matrix, we can take advantage of the fact that the SF of $\Phi(z)$ in the ring of polynomials $\mathbb{R}[z]$ is also the SF in $\mathbb{R}_1(z)$. That is because $\mathbb{R}[z] \subset \mathbb{R}_1(z)$, the invertible elements of $\mathbb{R}[z]$ are also invertible in $\mathbb{R}_1(z)$ and the divisibility relationship in $\mathbb{R}[z]$ is preserved in $\mathbb{R}_1(z)$.

(3) When we compute the SF, we can factorize its elements and remove all the factors without unit roots (which are invertible in R), since only unit roots are relevant. This way, we find the precise form of SF that appears in Proposition 1.

(4) When we estimate the SF via the approximate algorithm of the previous subsection, we get a diagonal matrix $\hat{D}^{\Phi}(z)$ whose entries' roots are neither exactly unitary, nor exactly the same as those of the determinant of the estimated VAR matrix $\hat{\Phi}(z)$. Since we are assuming that all unit roots belong to $\{\omega_k\}_k$, we can orthogonally project the unit roots of $\det\hat{\Phi}(z)$ onto $\{\omega_k\}_k$ and put them in the right row and column according to $\hat{D}^{\Phi}(z)$. More specifically, for each root $\hat{\omega}$ of $\det\Phi(z)$, we get its projection $\omega \in \{\omega_k\}_k$ and assign it to the place where it is the closest root of the diagonal entries of $\hat{D}^{\Phi}(z)$. Once all the roots are assigned, we get a new diagonal matrix $\tilde{D}^{\Phi}(z)$ similar to $\hat{D}^{\Phi}(z)$, but where all the roots are *exactly* unitary.

(5) The algorithms used to calculate the SF guarantee at each step that the elements of the main diagonal of $\hat{D}^{\Phi}(z)$ divide each other. This entails much more matrix operations and thus a greater estimation error. Hence, we found that the algorithm is more precise if we proceed in the following way: (i) apply the approximate algorithm wihout forcing the divisibility; (ii) apply the procedure described just above to make the roots exactly unitary and (iii) apply the full algorithm to get the SF with the diagonal elements dividing one each other. Step

(iii) is equivalent to make operations with integers, so is absolutely precise, whereas step (i) is simpler this way and thus, more accurate.

Considering all this, the algorithm of our package GVEC consists of the following steps:

(a) Fit a VAR(p), $\hat{\Phi}(z)$ to the data $y_1,\ldots,y_T$. Order p is provided by the user or determined by BIC, AIC or HQ.

(b) Obtain the preliminary SF estimate $\hat{D}^{(\Phi,0)}$ of $\hat{\Phi}(z)$ using the ε_T-approximate algorithm without divisibility.

(c) Separate the unit roots of the elements in the main diagonal of $\hat{D}^{(\Phi,0)}$. For this, we use the criterion $|u|<1+\varepsilon_T$. We obtain for each $j=1,\ldots,n$, the approximately unit roots $u_{j\ell}$.

(d) Get the orthogonal projections v_i of the roots of $\det\hat{\Phi}(B)$ onto $\{\exp(2\pi ik/s)\}_{k=1}^{s}$.

(e) Each v_i is assigned to the position j^* of the diagonal where $j^*=\arg\min_j|v_i-u_{j\ell}|$. Each time one root is assigned, both v_i and $u_{j^*\ell}$ are removed. Once all roots are assigned, the elements of the diagonal matrix $\hat{D}^{(\Phi,1)}$ are calculated as the product of the degree-one factors corresponding to the unit roots assigned.

(f) Obtain the SF estimate $\hat{D}^{(\Phi,2)}$ of $\hat{\Phi}(z)$ using the exact algorithm on $\hat{D}^{(\Phi,1)}$.

(g) Calculate $\delta_1,\ldots,\delta_h$ as the distinct polynomials among the diagonal elements of $\hat{D}^{(\Phi,2)}$.

(h) Build the GVEC corresponding to $\delta_1,\ldots,\delta_h$. The order of the GVEC and the presence of deterministic trend and intercept is determined by BIC or AIC.

The parameter ε_T can be chosen by the user of the package. By default, we set $\varepsilon_T=\log\log T\cdot T^{-1/2}$. By the assumptions of theorem 1, ε_T should decrease more slowly than $T^{-1/2}$, so $\log\log T\cdot T^{-1/2}$ is very near the boundary.

Since the asymptotic efficiency of AIC is proved for the nonstationary case in Ing et al. (2007) one may think that AIC would be the right choice, but there are reasons to support the use of a more parsimonious criterion, such as BIC. One is that in fact, it is not necessary that the estimated VAR is consistent, as long as it captures consistently the unstable part of the model. This argument is developed in Proposition 5. This is strengthened by the fact that the unstable part of the model is superconsistent. Hence, in the example of section 5 we use BIC.

Even when the model $\Phi(B)y_t = \varepsilon_t$ is an infinite VAR, it can be approximated by a VAR(p). We would like to prove that it is possible to estimate consistently the SF of the infinite model using finite approximations. Unfortunately, while for the case of stable infinite VAR we know that under some assumptions the finite VAR approximation is consistent, to our knowledge there is no such result for the unstable case.

However, we can show that for the rational case, i. e., when $\Phi = M^{-1}A$, the approximation by a finite VAR works asymptotically well for our purposes even if the order of the VAR does not diverge to infinity.

Let $\tilde{S}(\hat{\Phi}, \varepsilon_T; \mathbb{R}[z])$ be the result of factorizing $S(\hat{\Phi}, \varepsilon_T; \mathbb{R}[z])$ and removing the factors with non-unitary roots. Then, $\tilde{S}(\hat{\Phi}, \varepsilon_T; \mathbb{R}[z])$ may be a consistent estimate of $S(\Phi, 0; \mathbb{R}_1(z))$ even if $\hat{\Phi}$ is not a consistent estimate of Φ. This happens when the true model is an unstable VARMA.

Proposition 5. *If $A(B)y_t = M(B)\varepsilon_t$, where A may have unit roots and $\hat{\Phi}(B)$ is the least squares estimator of $\Phi(B)y_t = \varepsilon_t$ with order greater than that of A, then, $\tilde{S}(\hat{\Phi}, \varepsilon_T; \mathbb{R}[z]) \xrightarrow{p} S(A, 0; \mathbb{R}_1(z))$.*

Proof. See Appendix A. □

On the other hand, Φ may be approximated by a VARMA model. We lack a precise analysis of the properties of the estimators of unstable VARMA models, but probably the properties of the univariate case (Ling and Li, 1998) still hold. This is a possible line of investigation for the future.

4 Monte Carlo

Since the automatic identification method described in Section 3.2 is not an absolutely straightforward application of the algorithm of section 3.1, but has instead some small heuristic modifications, it is convenient to have an empiric confirmation that the method actually works. We want to compare the performance of the method to more traditional tools. Thus, our exercise will have two parts. First, we will compare our method to one based in the Johansen cointegration test in cases in which the latter can be applied. In second place, we will turn to a variety of cases and show the performance of our method alone, since there is no one to compare in such a general framework.

For this purpose, we will consider $n \times 1$ (with $n = 2,3$) processes y_t with the form:

$$d(B)y_t = Q \cdot D^{\Psi}(B)\varepsilon_t, \tag{10}$$

where Q is a random matrix with $[0,1]-$uniform entries, $D^{\Psi}(B)$ is a certain Smith matrix whose choice is described below and ε_t is $n \times 1$ Gaussian white noise. For each model (10), we will simulate $M = 500$ time series of length $T = 25, 50, \ldots, 500$. For the simulations, we have translated the identificacion programs into MATLAB code that runs faster than our R package.

4.1 Comparison with Johansen test in the $I(1)$ case

Our first exercise with simulated series is to compare our method to the Johansen test. The way in which we will use the test is as follows:

(i) For each $r = 0, \ldots, n-1$ we perform the Johansen test for the null that the cointegration rank is less or equal that r, with signifcance level $\alpha = 0.05$.

(ii) We estimate the cointegration rank as $r_* + 1$, where r_* the greatest value of r for which the test rejects.

This procedure is only valid for $I(1)$ processes. Therefore, we limit the range of the polynomial $d(z)$ and matrices D^{Ψ} in (10) to $d(z) = 1 - z$ and

$$D^{\Psi}(z) = \begin{pmatrix} I_r & 0 \\ 0 & (1-z)I_{n-r} \end{pmatrix}. \tag{11}$$

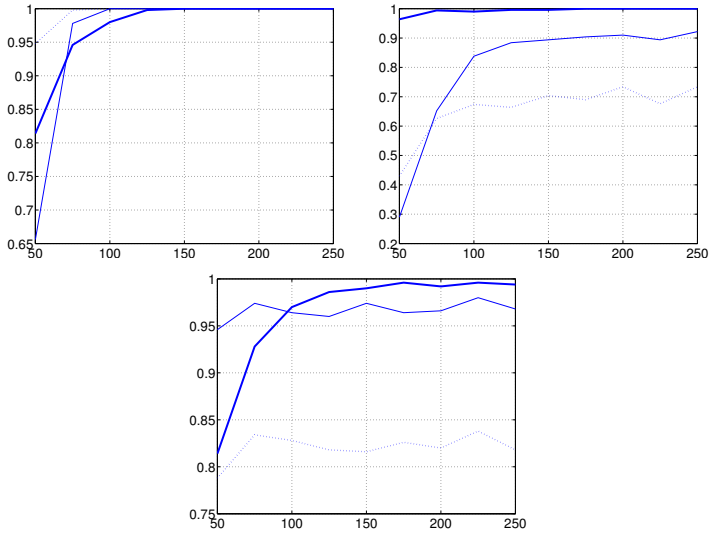

Figure 1: Probability of correct identification as a function of series length. Thick line: our method; continuous thin line: Johansen test with $\alpha = 0.01$; dotted line: Johansen test with $\alpha = 0.05$. Dimension $n = 2$ and series simulated according to (11) with (from top to bottom and from left to right) $r = 2$ (the series are actually stationary), $r = 1$ (cointegrated) and $r = 0$ (no cointegration).

Within this boundary, determining the SF is equivalent to determining the cointegration rank, so we can actually compare the probability of correctly identifying the cointegration rank with the Johansen test and with our procedure. We estimate both probabilities counting for each series length T how many times both methods get it right. In Figure 1 and Figure 2, we represent the length-dependent curves for dimensions 2 and 3.

Both in dimension 2 and 3, our method seems to do very well in the nontrivial cases when $0 < r < n$, clearly outperforming the Johansen test, particularly in $n = 3$. In the case $r = n$ there are no very large differences, since all methods detect quite easily that the series are not actually integrated. The only case in which the Johansen test works better is when $r = 0$, for series of short to moderate lengths, although for $d = 2$ this only happens for $\alpha = 0.01$. Also, even in the cases when

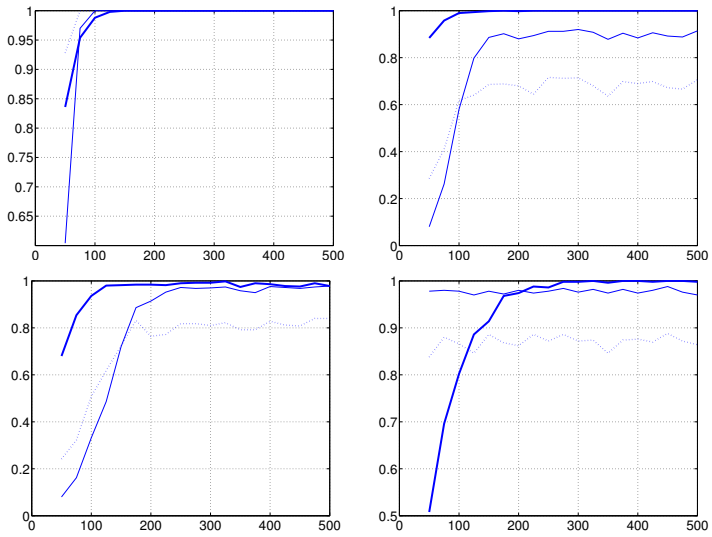

Figure 2: Probability of correct identification as a function of series length. Thick line: our method; continuous thin line: Johansen test with $\alpha = 0.01$; dotted line: Johansen test with $\alpha = 0.05$. Dimension $n = 3$ and series simulated according to (11) with (from top to bottom and from left to right) $r = 3$ (the series are actually stationary), $r = 2$ (cointegrated with rank 2), $r = 1$ (cointegrated with rank 1) and $r = 0$ (no cointegration).

the probability of correct identification with our method grows at a slower pace, it always starts at pretty decent levels for very short series, unlike the Johansen test, which yields very small probabilities in some cases.

We would like to point out that at least in this limited framework our method performs at least as well as the Johansen test even though ours is not restricted to the $I(1)$ case, unlike the Johansen test which limits itself to any of the $n+1$ possible values for the cointegration rank. In the next subsection, we show what happens when the true model is not among those that can be identify using the Johansen test.

4.2 Results in the $I(2)$ case

Now, we consider cases with

$$d(z) = (1-z)^2; \quad D^{\Psi}(z) = \begin{pmatrix} I_{r_0} & 0 & 0 \\ 0 & (1-z)I_{r_1} & 0 \\ 0 & 0 & (1-z)^2 I_{n-r_0-r_1} \end{pmatrix}. \tag{12}$$

Additionally, we also try a case with negative unit roots that can be interpreted as biannual seasonality, that is,

$$d(z) = 1-z^2; \quad D^{\Psi}(z) = \begin{pmatrix} I_r & 0 \\ 0 & (1-z^2)I_{n-r} \end{pmatrix}. \tag{13}$$

In Figure 3, we see that the convergence is somewhat slower in some cases, in particular when $r_0 = r_1 = 1$, but we got fairly good probabilities for long series, above 150 observations length. In Figure 4 for $n = 3$ we have some cases that are really tough. Since the convergence is slower, we represent the curves with the length parameter going up to 500 observations. In particular, the convergence of the case $r_0 = 2, r_1 = 0$, seems to be very slow. It is not surprising that for $n = 3$ is more difficult to identify correctly the case, since there are much more possible cases to distinguish. In fact, considering only roots equal to unity, the number of combinations for $n-$dimensional $I(d)$ series grows as $(d+1)^{n-1}$. However, we see that the seasonal case works pretty well even for relatively short series.

5 Real data example

To illustrate the use of the method and the package GVEC, we will identify a model for a data set similar to the one used in Hylleberg et al. (1990). In particular, y_t^1 will be the logarithm of the net disposable income of the UK and y_t^2 will be the logarithm of the expenditure in final consumption of the households, with t ranging from 1955:Q1 to 2012:Q1. Both series are depicted in Figure 5.

The AIC-selected order for the var is 5. The estimated VAR is

$$\hat{\Phi} = \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{pmatrix}$$

with

$$\begin{aligned}
\hat{\Phi}_{11} &= 1 - 0.400B - 0.105B^2 - 0.107B^3 - 0.462B^4 + 0.129B^5 \\
\hat{\Phi}_{12} &= -0.588B + 0.073B^2 + 0.112B^3 - 0.254B^4 + 0.603B^5 \\
\hat{\Phi}_{21} &= -0.021B + 0.014B^2 - 0.108B^3 + 0.091B^4 - 0.056B^5 \\
\hat{\Phi}_{22} &= 1 - 0.908B - 0.000B^2 + 0.0646B^3 - 0.973B^4 + 0.899B^5.
\end{aligned}$$

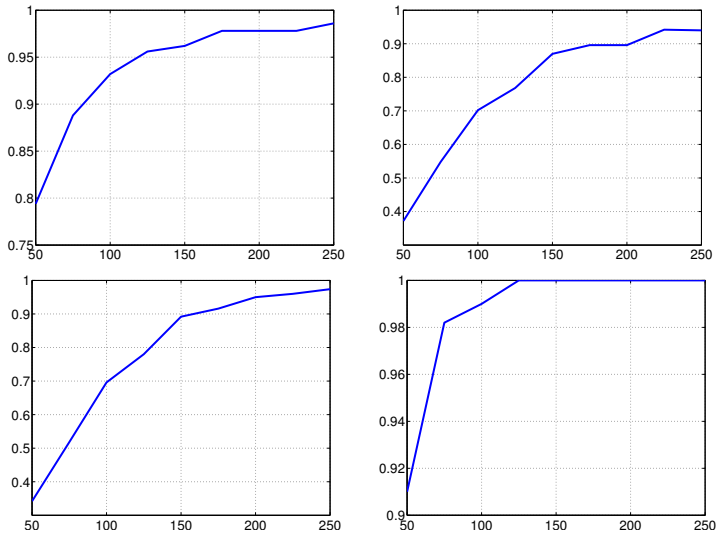

Figure 3: Probability of correct identification as a function of series length. Dimension $n = 2$ and series simulated according to (from top to bottom and from left to right) model (12): $r_0 = 1, r_1 = 0$; $r_0, r_1 = 1$; $r_0 = 2, r_1 = 0$ and model (13), $r = 1$.

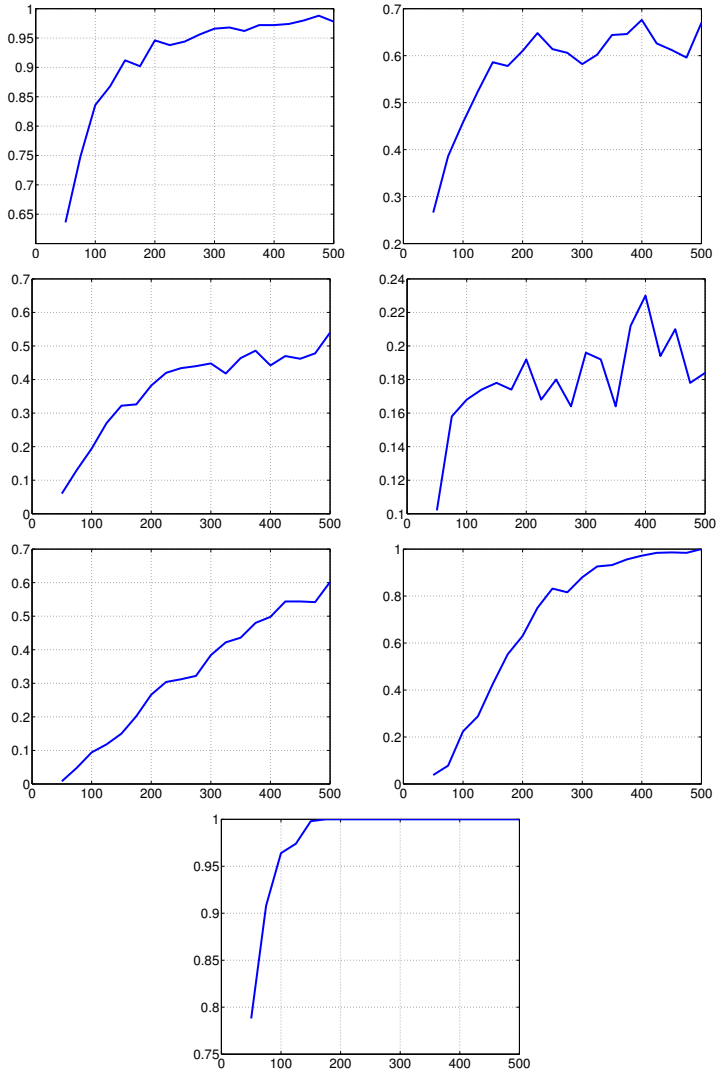

Figure 4: Probability of correct identification as a function of series length. Dimension $n = 3$ and series simulated according to (from top to bottom and from left to right) model (12): $r_0 = 1, r_1 = 0$; $r_0, r_1 = 1$; $r_0 = 1, r_1 = 2$; $r_0 = 2, r_1 = 0$; $r_0 = 2, r_1 = 1$; $r_0 = 3, r_1 = 0$; and model (13), $r = 2$.

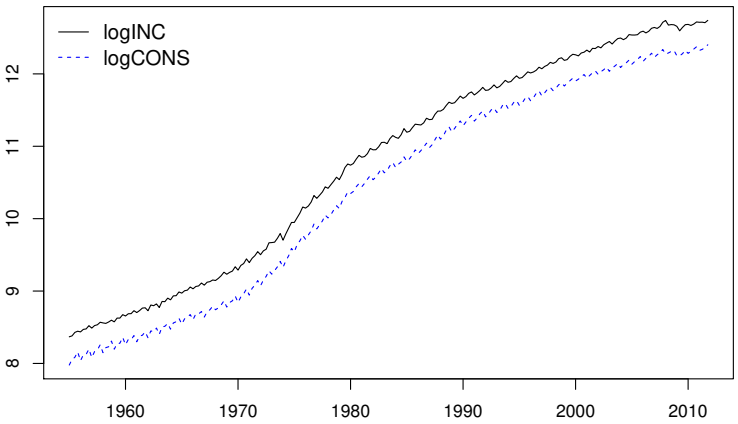

Figure 5: Logarithm of Income and Consumption of the UK.

We apply the automatic method with $\varepsilon_T = T^{-1/3}$ and get the matrix $\text{diag}(d_1, d_2)$ with $d_1 = 1 - B$ and $d_2 = (1-B)(1-B^4)$, or

$$\begin{pmatrix} \nabla & 0 \\ 0 & \nabla \cdot \nabla_s \end{pmatrix},$$

where $\nabla = 1 - B$ and $\nabla_s = 1 - B^s$.

Then, the GVEC is

$$\Gamma(B)\nabla_s \cdot \nabla y_t = \Pi_{0,1}^{(1)} \nabla_s y_{t-1} + \left(\Pi_{0,1}^{(1,-)} + \Pi_{0,1}^{(1,+)} B\right)(1-B^2)\nabla y_{t-1} + \Pi_{2,1}^{(1)}(1 - B - B^2 + B^3)\nabla y_{t-1} + \varepsilon_t.$$

where $\Gamma(B)$ has order $p = 4$ and the intercept μ equals to $(-0.00369, -0.00343)'$ and

$$\Pi_{0,1}^{(1)} = \begin{pmatrix} -0.185 & 0.139 \\ -0.717 & 0.667 \end{pmatrix} \quad \Pi_{1,1}^{(1,-)} = \begin{pmatrix} -0.046 & 0.023 \\ -0.090 & 0.104 \end{pmatrix}$$

$$\Pi_{1,1}^{(1,+)} = \begin{pmatrix} -0.020 & 0.038 \\ 0.190 & -0.069 \end{pmatrix} \quad \Pi_{2,1}^{(1)} = \begin{pmatrix} 0.032 & 0.001 \\ 0.070 & -0.019 \end{pmatrix}.$$

this means that according to our method both series are $I(2)$ with respect to frequency zero and $I(1)$ with respect to the seasonal frequencies. In this respect, this is consistent with the univariate results obtained with TRAMO/SEATS. The model identified is equivalent to a seasonal VEC model as the one analyzed in HEGY for the first differences. Therefore, the interpretation of the coefficients in the HEGY model applies here. For example, by extracting the eigenvector corresponding to the greatest eigenvalue of $\Pi_{0,1}^{(1)}$, we get that there is cointegration in the with respect to frequency zero and the cointegrating vector is quite close to $(1,-1)$.

6 Conclusions

We have tried to unify the treatment of cointegration for practitioners, in doing so we have contributed in providing unified theoretic framework for a broad class of situations. To this end we have provided a theoretical framework that covers any process that can be represented by an autoregressive model with unit roots. Our approach covers I(1), I(2), multicointegration, polynomial and seasonal cointegration. From the practioner perspective we put forward an automatic method to identify GVEC models based on an estimate of the Smith Forms of the autoregressive model. Our method is competitive with Johansen test, in the restricted cases in which the latter applies, with the advantage that the new method can cope with other practical situations.

Further research can be conducted to extend the theoretical framework to fractional integration or other forms of long memory and to relax some assumptions such as the rationality of the transfer function. Also, in order to apply these ideas to large dimension, or even to panel data, the algorithms need to be adapted, since in their actual form they are only suited to moderate dimensions.

Acknowledgements: Mariano Matilla-García is grateful to the following research grants: ECO2015-65637-P from the Ministerio de Economía y Competitividad de Espana and

FEDER funds, and to the program of Groups of Excellence of the Región de Murcia, the Fundación Séneca , Science and Technology Agency of the Región de Murcia, project 19884/GERM/15.

A Appendix

Proof of Proposition 1. First, the ranks $r_{k,j}$ are uniquely determined because

$$r_{k,j} = \dim_{\mathbb{C}^n} \langle a(\omega_k) : a(B)' y_t \sim I(d_k - j) \rangle.$$

By the uniqueness of the ranks, it suffices to prove that for every process y_t, Ψ can be represented as in (a) and then, there are exactly $n - \sum_{\ell \geq j} s_{k,\ell}$ cointegration vectors whose values at ω_k are linearly independent.

The first step is a consequence of Proposition 2.11 from Hungerford (1980), applied to the ring $R = \{f/g : \forall k, g(\omega_k) \neq 0\}$. We have then a representation $\Psi(z) = U(z) D^{\Psi}(z) V(z)$, where the elements of the diagonal of $D^{\Psi}(z)$ divide one each other $d_{i,i}(z) | d_{i+1,i+1}(z)$. This allows to represent y_t as

$$d(B) y_t = U(B) D^{\Psi}(B) V(B) \varepsilon_t. \tag{14}$$

It is clear that we can factorize $D^{\Psi}(z)$ as $D^{\Psi}(z) = D_0(z) \cdot \ldots \cdot D_g(z)$. If we focus on a certain k, all the other unit roots can be moved to the left or the right so we can write $\Psi(z) = U_k(z) D_k(z) V_k(z)$. where $\det U(\omega_k), \det V(\omega_k) \neq 0$. If we choose $G(z)$ as the last $r_{k,j}$ rows of $U_k(z)^{-1}$, then we get

$$G(B) y_t = [0_{1 \times n - r_{k,j}} : \ldots] d(B)^{-1} D_k(B) V_k(B)) \varepsilon_t.$$

Since the last $r_{k,j}$ elements of the diagonal of $D^{\Psi}(z)$ are divided by $(1 - \omega_k^{-1} z)^j$, we get that $G(B) y_t \sim I_k(d_k - j)$. On the other hand, $G(\omega_k)$ is full rank for otherwise $U_k(z)$ would have a root at ω_k, which is contradictory with the conditions of the SF.

Thus, there are at least $r'_{k,j} := n - \sum_{\ell \geq j} s_{k,\ell}$ cointegrating vectors $a(z) \in \mathscr{A}$ such that $\{a(\omega_k)\}_{a \in \mathscr{A}}$ are linearly independent. Hence $r_{k,j} \geq r'_{k,j}$. To see that this is actually an identity, let us assume that $r_{k,j} > r'_{k,j}$. Then, we arrange the cointegration relationships in a $r_{k,j} \times n$ matrix $A_1(z)$. Then $A_1(B) y_t = \tilde{U}_1(B) \tilde{D}_1(B) \tilde{V}_1(B) \tilde{\varepsilon}_{1t}$. Now,

let A_2 be a $(n-r_{k,j})\times n$ constant matrix A_2 such that $[A_1(\omega_k)' : A_2']$ is invertible and let $A_2 y_t = \tilde{U}_2(B)\tilde{D}_2(B)\tilde{V}_2(B)\tilde{\varepsilon}_{2t}$. Now, we can represent y_t as

$$y_t = \tilde{U}(B)\begin{pmatrix} \tilde{D}_1(B) & 0 \\ 0 & \tilde{D}_2(B) \end{pmatrix}\tilde{V}(B)\begin{pmatrix} \tilde{\varepsilon}_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

where

$$\tilde{U}(B) = \begin{pmatrix} A_1(B) \\ A_2 \end{pmatrix}^{-1}\begin{pmatrix} \tilde{U}_1(B) & 0 \\ 0 & \tilde{U}_2(B) \end{pmatrix} \qquad \tilde{V}(B) = \begin{pmatrix} \tilde{V}_1(B) & 0 \\ 0 & \tilde{V}_2(B) \end{pmatrix}$$

To see (b), if $D^{\Psi}(z)|d(z)$ we just need to multiply (14) by $\mathrm{adj}V(B)D^{\Psi}(B)^{-1}\mathrm{adj}U(B)$, so we get.

$$\mathrm{adj}V(B)d(B)D^{\Psi}(B)^{-1}\mathrm{adj}U(B)y_t = \det U(B)\det V(B)\varepsilon_t,$$

where $\det U(B)\det V(B)$ has no unit roots. If $D^{\Psi}\nmid d$, then some roots of $D^{\Psi}(z)$ have to remain in the RHS of the representation, so m has unit roots. □

Lemma 1. *The ring of polynomials $\mathbb{C}[z]$ and $R = \{f/g : f,g \in \mathbb{C}[z], \forall k, g(\omega_k)\neq 0\}$ are PIDs.*

Proof. Every Euclidean ring is a PID (see Hungerford, 1980, theorem 3.9). The fact that $\mathbb{C}[z]$ is an Euclidean ring is elementary. We will prove that for R. We can write any $f/g \in R$ as $f = hk/g$, where h has roots only among $\{\omega_\ell\}_{\ell=0}^{s-1}$ and k has none there. Then, if we denote by ∂p the degree of polynomial p, we can define $\varphi(f) = \partial h$. To divide $f_1 = h_1k_1/g_1$ by $f_2 = h_2k_2/g_2$, we first divide h_1 by h_2, so $h_1 = qh_2 + r$. Then $f_1 = (qh_2+r)k_1 = qh_2k_1/g_1 + rk_1/g_1 = \tilde{q}f_2 + rk_1/g_1$, where $\tilde{q} = (g_2k_1g_1^{-1}k_2^{-1})q$ and $\varphi(rk_1/g_1) = \partial r < \partial h_2 = \varphi(f_2)$. □

Lemma 2. *Let $f(z)$ be a holomorphic function in $\Omega \subset \mathbb{C}$ such that $\bar{f}(z) = f(\bar{z})$ and $p(z)$ a polynomial with real coefficients and nonzero roots. Let us assume that the roots of p that have nonnegative imaginary part are $\{\theta_k\}_k$ with multiplicities m_k. Then, $f(z) = h(z)p(z) + r(z)$ where $h(z)$ is a holomorphic function in Ω and*

$$r(z) = \sum_{\mathrm{Im}\theta_k=0}\sum_{\ell=1}^{m_k} c_{k,\ell}\frac{p(z)}{(z-\theta_k)^\ell} + \sum_{\mathrm{Im}\theta_k>0}\sum_{\ell=1}^{m_k}\left(c_{k,\ell,0}+c_{k,\ell,1}z\right)\frac{p(z)}{(z^2-2\mathrm{Re}\theta_k z+|\theta_k|^2)^\ell},$$

with $c_{k,\ell}, c_{k,\ell,0}, c_{k,\ell,1} \in \mathbb{R}$.

Proof. Let $q_{k,\ell}(z) = p(z)/(z-\theta_k)^\ell$. Is is easy to see that there are $c_{k,\ell} \in \mathbb{C}$ such that at each θ_k, $\sum_k \sum_{\ell=1}^{m_k} c_{k,\ell} q_{k,\ell}(z)$ and its derivatives up to $m_k - 1$ coincide with $f(z)$. It suffices to write down the identities and see that they form a triangular linear system.

Let now $q_{k,\ell,0}(z) = p(z)/(z^2 - 2\mathrm{Re}\theta_k + |\theta_k|^2)^\ell$ and $q_{k,\ell,1}(z) = q_{k,\ell,0}(z)z$. We will see that $\mathscr{B} = \{q_{k,\ell} : \mathrm{Im}\theta_k = 0\} \cup \{q_{k,\ell,0}, q_{k,\ell,1} : \mathrm{Im}\theta_k > 0\}$ are linearly independent in V, the space of the polynomials of degree up to $\partial p - 1$. Let us assume that there is a linear combination of the elements of $\mathscr{B}$ that equals zero. Then,

$$\sum_{\mathrm{Im}\theta_k=0} \sum_{\ell=1}^{m_k} c_{k,\ell} \frac{1}{z(z-\theta_k)^\ell} + \sum_{\mathrm{Im}\theta_k>0} \sum_{\ell=1}^{m_k} c_{k,\ell,0} \frac{1}{z(z^2 - 2\mathrm{Re}\theta_k z + |\theta_k|^2)^\ell} + \sum_{\mathrm{Im}\theta_k>0} \sum_{\ell=1}^{m_k} c_{k,\ell,1} \frac{1}{(z^2 - 2\mathrm{Re}\theta_k z + |\theta_k|^2)^\ell} = 0. \quad (15)$$

Now, we can integrate the last identity along a closed path encircling only θ_k. By the Residue Theorem (Rudin, 1987), for real θ_k we obtain that $2\pi i c_{k,\ell}/\theta_k = 0$, whereas for a pair of conjugate roots, we get $2\pi c_{k,\ell,0}/(\theta_k 2\mathrm{Im}\theta_k) + 2\pi c_{k,\ell,1}/(2\mathrm{Im}\theta_k) = 0$ and $-2\pi c_{k,\ell,0}/(\bar{\theta}_k 2\mathrm{Im}\theta_k) - 2\pi c_{k,\ell,1}/(2\mathrm{Im}\theta_k) = 0$. Consequently, $c_{k,\ell,0} = c_{k,\ell,1} = 0$. Since the number of elements in $\mathscr{B}$ equals the dimension of V, they form a basis.

To see that the coefficients are real, notice that the left hand side of (15) equals $r(z)/(zp(z)) = f(z)/(zp(z)) - h(z)/z$. Then, we can apply again the Residue Theorem with the same paths. Since $\overline{f(z)/p(z)} = f(\bar{z})/p(\bar{z})$ and $h(z)$ is holomorphic, then $\mathrm{Res}(f(z)/(zp(z)) - h(z)/z, \theta_k) \in i\mathbb{R}$. This entails that by solving the equations, we can see that $c_{k,\ell}, c_{k,\ell,0}, c_{k,\ell,1} \in \mathbb{R}$. □

Proof of Proposition 2. From the identity $\Phi(z) = U^\Phi(z) D^\Phi(z) V^\Phi(z)$, we get $\Phi(z) = \sum_{i=1}^n u_i(z) v_i(z)' d_i(z)$, where the vectors $u_i(z)$ and $v_i(z)'$ are respectively the columns of $U^\Phi(z)$ and the rows of $V^\Phi(z)$, and $d_i(z)$ is the ith element in the main diagonal of $D(z)$. We can group terms with common d_i, so

$$\Phi(z) = \sum_{j=1}^r A_j^{(1)}(z) \delta_j(z). \quad (16)$$

On the other hand, let us write $\delta_{j+1} = c_j \delta_j$. We can divide c_1 between z. We have then, $c_1 = z g_1 + k$, with $k \in \mathbb{R}$ and $k \neq 0$. Hence, $\delta_1 = k^{-1}\delta_2 - z g_1 \delta_1$. If we replace δ_1 in (16), we get

$$\Phi(z) = A_r^{(1)}(z)\delta_h(z) + \ldots + \left(A_2^{(1)}(z) + k^{-1}A_1^{(1)}(z)\right)\delta_2(z) - A_1^{(1)}(z)g_1(z)z\delta_1(z) = \sum_{j>1}^{h} A_j^{(2)}(z)\delta_j(z) + A_j^{(2)}(z)\delta_j(z)z$$

We can repeat this device to obtain the form

$$A_r^{(h-1)}(z)\delta_h(z) + \sum_{j=1}^{h-1} A_j^{(r-1)}(z)\delta_j(z)z = B_h^{(1)}(z)\delta_h(z) + \sum_{j=1}^{h-1} B_j^{(1)}(z)\delta_j(z)z. \quad (17)$$

The last step consists of rewriting the terms of the last sum in (17). We divide the elements of $B_1^{(1)}(z)$ by $c_1(z)$ using lemma 2, so we get $B_1^{(1)}(z) = c_1(z)Q_1(z) + R_1(z)$, where $R_1(z)$ is a polynomial matrix whose elements have the form given by lemma 2. Then, if c_1 has roots θ_k with multiplicities m_k, then

$$\Phi(z) = \sum_{\mathrm{Im}\theta_k = 0} \sum_{\ell=1}^{m_k} \Pi_{k\ell}^{(1)} \Delta_{1k\ell}(z) + \sum_{\mathrm{Im}\theta_k > 0} \sum_{\ell=1}^{m_k} \left(\Pi_{k\ell}^{(1,+)} + \Pi_{k\ell}^{(1,-)} z\right)\Delta_{1k\ell}(z) + B_1^{(2)}\delta_1(z)z + \sum_{j=2}^{h-1} B_j^{(2)}(z)\delta_j(z)z + B_r^{(2)}(z)\delta_h(z),$$

where $B_1^{(2)} = R_1$. We proceed dividing in turn each $B_j^{(j)}$ by f_j. Finally $\Pi_j = A_j^{(r-1)}$. □

Proof of Proposition 3. The VAR representation of the process, $\Phi(B)y_t = \varepsilon_t$, can be written as $U(B)D(B)V(B)y_t = \varepsilon_t$, where for ease of notation we omit the superscript $\cdot^{\Phi}$. We can write the elements of the main diagonal of $D(z)$ as $d_i(z) = \tilde{d}_i(z)g_i(z)$, where $\tilde{d}_i(z)$ has all the unit roots of $d_i(z)$. Then, we can move the non-unit roots to the right and get the representation $U(B)\tilde{D}(B)\tilde{V}(B)y_t = \varepsilon_t$. We will prove first the case when all unit roots are equal to 1. Then, exists r, s such that $\delta_1 = (1-z)^s$ and $\delta_h = (1-z)^r$. We will write $\Delta_i(z) = (1-z)^i$. Let us define $\Delta_i(B)y_t = y_t^{(i)}$.

We will write down the form of the least squares estimator. Then by defining $Y = \left[y_T^{(r)}, y_{T-1}^{(r)}, \ldots\right]$ and $\varepsilon = \left[\varepsilon_T, \varepsilon_{T-1}, \ldots\right]$, the parameter matrix $\beta = [\Gamma_1, \ldots, \Gamma_p : \Pi]$, $\Pi = [\Pi^{(1)}, \ldots, \Pi^{(r)}]$ and $X = [X^{(1)'} : X^{(2)'}]'$, where

$$X^{(1)} = \begin{pmatrix} y_{T-1}^{(r)} & y_{T-2}^{(r)} & \cdots & \\ y_{T-2}^{(r)} & y_{T-3}^{(r)} & \cdots & \\ \cdots & \cdots & \cdots & \cdots \\ y_{T-p}^{(r)} & y_{T-p-1}^{(r)} & \cdots & \end{pmatrix} \quad X^{(2)} = \begin{pmatrix} y_T^{(r-1)} & y_{T-1}^{(r-1)} & \cdots & \\ y_T^{(r-2)} & y_{T-1}^{(r-2)} & \cdots & \\ \cdots & \cdots & \cdots & \cdots \\ y_T^{(s)} & y_{T-1}^{(s)} & \cdots & \end{pmatrix}.$$

Now, we can write the model as $Y = \beta X + \varepsilon$ and the least squares estimator is $\hat{\beta} = YX'(XX')^{-1}$ so that $\hat{\beta} = \beta + \varepsilon X'(XX')^{-1}$. We need to analyze the asymptotic behavior of $\varepsilon X'$ and XX'.

The process $Y_t = (y_t^{(s)'}, \ldots, y_t^{(r-1)'})'$ satisfies $Y_t = \mathbf{D}Y_{t-1} + A_t$ with

$$\mathbf{D} = \begin{pmatrix} I_n & I_n & 0 & \ldots & 0 \\ 0 & I_n & I_n & \ldots & 0 \\ 0 & 0 & I_n & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & I_n \end{pmatrix}$$

and $A_t = (0, \ldots, 0, C(B)\varepsilon_t)'$, $C = \tilde{V}^{-1}\delta_r D^{-1} U^* = \tilde{V}(B)^{-1}\mathrm{diag}(\delta_r/\delta_s, \ldots, 1)U^*$. By means of the Beveridge-Nelson decomposition we can represent Y_t in a similar way as in section 2 of Tsay and Tiao (1990) (hereinafter, cited as TT). In particular, we write $Y_t = Z_t + M_t$, where $Z_t = \mathbf{D}Z_{t-1} + B_t$ and $B_t = (0, \ldots, 0, C(1)\varepsilon_t)'$ and $M_{\ell,t} = O_p(t^{r-\ell-1})$. We can also write the components of Z_t as $Z_{r-1,t} = C(1)\sum_{j=1}^{t}\varepsilon_j$ and $Z_{\ell,t} = Z_{\ell,t-1} + Z_{\ell+1,t}$.

Let us simplify the notation by setting $\xi_t = C(1)\varepsilon_t$, $Z_{r-1,t} = \sum_{\tau=1}^{t}\xi_\tau$ and $Z_{\ell,t} = \sum_{\tau=1}^{t} Z_{\ell+1,\tau}$. Now, for $u \in [0,1]$, we define $W_{\ell,t}(u) = t^{\ell-r+1/2} Z_{\ell,[tu]}$ for $\ell < r$. Then

$$\begin{aligned} W_{r-1,t}(u) &= t^{-1/2}\sum_{\tau=1}^{[tu]}\xi_\tau \\ W_{\ell,t}(u) &= t^{\ell-r+1/2}\sum_{\tau=1}^{[tu]} Z_{\ell+1,\tau} = t^{-1}\sum_{\tau=1}^{[tu]} W_{\ell+1,\tau}(u) \end{aligned}$$

Using theorem 3.1 from TT, we get $W_{\ell,t}(u) \xrightarrow{w} \Gamma_\ell(u)$, where $\Gamma_{r-1}(u) = B_\xi(u)$, $\Gamma_\ell(u) = \int_0^u \Gamma_{\ell+1}(v)dv$ and $B(u)$ is a Brownian motion with covariance matrix $C(1)C(1)'$. It also holds

$$T^{i+j-2r}\sum_t^T Z_{i,t}Z'_{j,t} = T^{-1}\sum_t^T W_{i,T}\left(\frac{t}{T}\right)W_{j,T}\left(\frac{t}{T}\right) \to \int_0^1 \Gamma_i(u)\Gamma_j(u)'du$$

and

$$T^{\ell-r}\sum_t^T Z_{i,t}\xi'_t = T^{-1/2}\sum_t^T W_{i,T}\left(\frac{t}{T}\right)\xi_t \to \int_0^1 \Gamma_i(u)dB_\xi(u)'du$$

$$T^{\ell-r}\sum_t^T Z_{i,t}\varepsilon'_t = T^{-1/2}\sum_t^T W_{i,T}\left(\frac{t}{T}\right)\varepsilon_t \to \int_0^1 \Gamma_i(u)dB_\varepsilon(u)'du.$$

Since, $M_{\ell,t} = O_p(t^{r-\ell-1})$, we get for $i, j < r$,

$$T^{i+j-2r}\sum_t^T y_t^{(i)}y_t^{(j)\prime} \to \int_0^1 \Gamma_i(u)\Gamma_j(u)'du$$

$$T^{i-r}\sum_t^T y_t^{(i)}\xi'_t \to \int_0^1 \Gamma_i(u)dB_\xi(u)'du$$

$$T^{i-r}\sum_t^T y_t^{(i)}\varepsilon'_t \to \int_0^1 \Gamma_i(u)dB_\varepsilon(u)'du$$

Now, we can put $\varepsilon X'(XX')^{-1} = UA^{-1}$ as

$$\left[\sum_t^T \varepsilon_t y_{t-1}^{(r)}, \ldots, \sum_t^T \varepsilon_t y_{t-p}^{(r)} : \sum_t^T \varepsilon_t y_{t-1}^{(r-1)}, \ldots, \sum_t^T \varepsilon_t y_{t-1}^{(s)}\right] \times$$

$$\times \begin{pmatrix} (\sum_t^T y_{t-i}^{(r)} y_{t-i}^{(r)\prime})_{i,j} & (\sum_t^T y_{t-i}^{(r)} y_{t-1}^{(r-j)\prime})_{i,j} \\ (\sum_t^T y_{t-1}^{(r-i)} y_{t-j}^{(r)\prime})_{i,j} & (\sum_t^T y_{t-1}^{(r-i)} y_{t-1}^{(r-j)\prime})_{i,j} \end{pmatrix}^{-1}$$

Now, let $L_T^* = \mathrm{diag}(T^{-1/2}I_{pn}, L_T)$. Then, $(\hat{\beta} - \beta)L_T^* = U^*A^{*-1}$ equals

$$\left[T^{-1/2}\sum_t^T \varepsilon_t y_{t-1}^{(r)}, \ldots, T^{-1/2}\sum_t^T \varepsilon_t y_{t-p}^{(r)} : T^{-1}\sum_t^T \varepsilon_t y_{t-1}^{(r-1)}, \ldots, T^{-r+s}\sum_t^T \varepsilon_t y_{t-1}^{(s)}\right] \times$$

$$\times \begin{pmatrix} (T^{-1}\sum_t^T y_{t-i}^{(r)} y_{t-i}^{(r)\prime})_{i,j} & (T^{-1/2+j-r}\sum_t^T y_{t-i}^{(r)} y_{t-1}^{(r-j)\prime})_{i,j} \\ (T^{-1/2+i-r}\sum_t^T y_{t-1}^{(r-i)} y_{t-j}^{(r)\prime})_{i,j} & (T^{2r-i-j}\sum_t^T y_{t-1}^{(r-i)} y_{t-1}^{(r-j)\prime})_{i,j} \end{pmatrix}^{-1}$$

Now, for $U^* = (U_1^*, U_2^*)$, $U_1 \xrightarrow{d} N(\Xi, \Omega)$,

$$U_1^* \xrightarrow{d} \left(\Big[\int_0^1 \Gamma_{r-1}(u) dB_\varepsilon(u)' \Big]', \ldots, \Big[\int_0^1 \Gamma_s(u) dB_\varepsilon(u)' \Big]' \right)$$

and

$$A^* \xrightarrow{d} \begin{pmatrix} \Gamma_p^{(r)} & \mathbf{0} \\ \mathbf{0} & \Omega \end{pmatrix}$$

where $\Gamma_p^{(r)}$ is the covariance matrix of $(y_{t-1}^{(r)}, \ldots, y_{t-p}^{(r)})'$ and

$$\Omega = \Big(\int_0^1 \Gamma_i(u)\Gamma_j(u)' du \Big)_{i,j}.$$

We will sketch now the proof for the general case. In order to make the notation less cumbersome, we denote the multi-index (j,k,ℓ) as α. We consider its values ordered with the lexicographical order. Then, $y_t^{(\alpha)} = \Delta_\alpha(B) y_t$. Now, $(\hat{\beta} - \beta) = UA^{-1}$ equals

$$\Big[\sum_t^T \varepsilon_t y_{t-1}^{(h)\prime}, \ldots, \sum_t^T \varepsilon_t y_{t-p}^{(h)\prime} : \Big(\sum_t^T \varepsilon_t y_{t-1}^{(\alpha)} \Big)_\alpha' \Big] \times$$
$$\begin{pmatrix} (\sum_t^T y_{t-\mu}^{(h)} y_{t-\nu}^{(h)\prime})_{\mu,\nu} & (\sum_t^T y_{t-\mu}^{(h)} y_{t-1}^{(\alpha)\prime})_{\mu,\alpha} \\ (\sum_t^T y_{t-1}^{(\alpha)} y_{t-\nu}^{(h)\prime})_{\alpha,\nu} & (\sum_t^T y_{t-1}^{(\alpha)} y_{t-1}^{(\beta)\prime})_{\alpha,\beta} \end{pmatrix}^{-1}.$$

Now, $\mathscr{B}_1 = \{\Delta^{(h)}\} \cup \{\Delta^{(\alpha)}\}_\alpha$ is a basis of the space of polynomials of degree up to $q = \partial\Delta^{(h)}$. On the other hand, let us define $L_{\nu,\tau}(z) = \Delta^{(h)}(z)(1-\theta_\tau^{-1}z)^{-\nu}$, where $\tau = 0, \ldots, s-1$ and $\nu = 1, \ldots, m_\tau$, where m_τ is the multiplicity of θ_τ in $\Delta^{(h)}$. It is easy to prove that $\mathscr{B}_2 = \{\Delta^{(h)}\} \cup \{L_{\nu,\tau}\}_{\nu,\tau}$ is also a basis by the Residue Theorem as in lemma 2. Then, there exists an invertible matrix Q such that $Q \cdot \big(\Delta^{(\alpha)}\big)_{(\alpha)} = \big(L(0)', \ldots, L(s-1)'\big)'$ and $L(\tau) = (L_{1,\tau}(z), \ldots, L_{m_\tau,\tau}(z))'$ contains the elements of $\mathscr{B}_2$ associated to the τth root. Then, $UA^{-1}Q'^{-1} = UQ\big(QAQ'\big)^{-1} = VB^{-1}$.

Let us denote $z_{\nu,\tau,t} = L_{\nu,\tau}(B)y_t = (1-\theta_\tau^{-1}z)^{-\nu}C(B)\varepsilon_t$. When $0 < \tau < b$, $z_{\nu,\tau,t}$ and $z_{\nu,s-\tau,t}$ are conjugate. We can also make a transformation similar to that in section 3 of TT to transform the conjugate pairs into pairs of real and imaginary parts $w_(u_{\nu,\tau,t} = \nu,\tau,t,v_{\nu,\tau,t})'$. Then, there is an invertible matrix P such that $VB^{-1} = WC^{-1}P$, $W = VP$ and $C = PBP'$. Consequently, $UA^{-1}Q'^{-1}P'^{-1}$ has the form

$$\left[W_{-1} : W_0, \ldots, W_b\right] \times \begin{pmatrix} C_{-1,-1} & C_{-1,0} & \ldots & C_{-1,b} \\ C_{0,-1} & C_{0,0} & \ldots & C_{0,b} \\ \ldots & \ldots & \ldots & \ldots \\ C_{b,-1} & C_{b,0} & \ldots & C_{b,b} \end{pmatrix}^{-1},$$

where $W_{-1} = [\sum_t \varepsilon_t y_t^{(h)\prime}, \ldots, \sum_t \varepsilon_t y_{t-p}^{(h)\prime}]$, $W_\tau = \left(\sum_t \varepsilon_t z'_{1,\tau,t-1}, \ldots, \sum_t \varepsilon_t z'_{m_\tau,\tau,t-1}\right)'$ for $\tau = 0, \ldots, b$, and

$$C_{-1,-1} = \left(\sum_t^T y_{t-u}^{(h)} y_{t-v}^{(h)\prime}\right)_{u,v} \qquad C_{-1,\tau} = \left(\sum_t^T y_{t-u}^{(h)} z_{\nu,\tau,t-1}{}'\right)_{u,\nu}$$
$$C_{\sigma,-1} = C'_{-1,\sigma} \quad C_{\sigma,\tau} = \left(\sum_t^T z_{\mu,\sigma,t-1} z_{\nu,\tau,t-1}{}'\right)_{\mu,\nu}.$$

We can deal now with the W's and C's in a similar fashion as the unity case, along the lines of section 4 of TT. □

Proof of Proposition 4. It suffices to see that the algorithm described in annex B (i) can be described as (5)-(6) with $\varepsilon = 0$, (ii) it stops and (iii) when it stops, the state variable D is the Smith form of A.

The first assertion can be proved as follows: let the state variable e_k comprise all the flags, an additional flag indicating whether $r = 0$ and the current line number. Then, all the actions in the algorithm are either (a) control flow sentences and changes of r, that are (5) or (b) *admissible operations*, that are (6). Thus, we can write a meta-algorithm that runs over the algorithm of the annex as follows:

1: $k, nline \leftarrow 1$
2: **loop**

3: **if** $sentence(nline)$ is type (a) **then**
4: perform $e_k = f_0(e_{k-1}, A^{(k-1)})$
5: **else**
6: perform $A^{(k)} = g(e_{k-1}, A^{(k-1)})$ and $nline \leftarrow nline + 1$
7: **end if**
8: $k \leftarrow k+1$
9: **end loop**

In order to see that the algorithm always stops and that when it stops, we get the Smith form, we can easily adapt the proof in Hungerford (1980), page 340, replacing the arguments based on the finiteness of the divisors of an element in the ring, by the fact that the degree function takes values in $\mathbb{N}$ and thus it can decrease only a finite number of times. □

Proof of Theorem 1. To simplify the proof, we assume that the algorithm does not involve the condition $r = 0$, but just conditions of the form $a_{ij} = 0$. The proof can be easily adapted then by considering an augmented matrix $\tilde{A} = [A : R]$, where $R = (r_{ij})_{ij}$ and r_{ij} is the remainder of the Euclidean division of a_{ii} and a_{ij}.

We will denote by $(e_k, A^{(k)})$ the pair we obtain as the result of iterating (5)-(6) starting from A, whereas $(\hat{e}_k, \hat{A}^{(k)})$ is got by iterating the $\varepsilon-$approximate version of (5)-(6) starting from $\hat{A}$. We will see that $\forall k$,

$$P[\hat{e}_\ell = e_\ell, \forall \ell \leq k] \to 1 \tag{18}$$

$$\hat{A}^{(k)} - A^{(k)} = w_T^k, \tag{19}$$

where $\forall \delta > 0, \exists M > 0, T_0$ such that $\forall T \geq T_0$, $P[\xi_T^{-1} \|w_T^k\| > M | \hat{e}_\ell = e_\ell, \forall \ell \leq k] < \delta$. For any random variable that satisfies this property of w_T^k, we write $O_{cp}(\xi_T)$, that is, *conditional order ξ_T in probability*. It is easy to see that this property behaves in a similar fashion to the usual order in probability, in particular, the product of two $O_{cp}(\xi_T)$ and $O_{cp}(\eta_T)$ sequences is $O_{cp}(\xi_T \eta_T)$.

We will prove (18) and (19) by induction in k. For $k = 0$ it is trivial. Now, let us assume that it holds for $k-1$ and we will prove it for k.

First, we will see that $P[\hat{e}_\ell = e_\ell, \forall \ell \leq k] \to 1$. Let i and j be indexes such that given $\hat{e}_{k-1} = e_{k-1}$, f_{ε_T} takes a certain value α if $|\hat{a}_{ij}^{(k-1)}| < \varepsilon$ and a value β if $|\hat{a}_{ij}^{(k-1)}| \geq \varepsilon$. We will use that

$$P\left[\hat{e}_\ell = e_\ell, \forall \ell \leq k\right] = P\left[\hat{e}_k = e_k | \hat{e}_\ell = e_\ell, \forall \ell \leq k-1\right] P\left[\hat{e}_\ell = e_\ell, \forall \ell \leq k-1\right]. \quad (20)$$

There are two cases, either $a_{ij}^{(k-1)} = 0$ or $a_{ij}^{(k-1)} \neq 0$. If $a_{ij}^{(k-1)} = 0$, then

$$P\left[\hat{e}_\ell = e_\ell, \forall \ell \leq k\right] = P\left[|\hat{a}_{ij}^{(k-1)}| \leq \varepsilon_T | \hat{e}_\ell = e_\ell, \forall \ell \leq k-1\right] P\left[\hat{e}_\ell = e_v, \forall \ell \leq k-1\right]$$

From (18) and $\xi_T^{-1}\varepsilon_T \longrightarrow \infty$, it follows,

$$P\left[|\hat{a}_{ij}^{(k-1)}| \leq \varepsilon_T | \hat{e}_\ell = e_\ell, \forall \ell \leq k-1\right] = \quad (21)$$

$$P\left[\xi_T^{-1}|\hat{a}_{ij}^{(k-1)} - a_{ij}^{(k-1)}| \leq \xi_T^{-1}\varepsilon_T | \hat{e}_\ell = e_\ell, \forall \ell \leq k-1\right] \to 1.$$

Let us now consider the case that $a_{ij}^{(k-1)} \neq 0$. Then,

$$P\left[|\hat{a}_{ij}^{(k-1)}| \leq \varepsilon_T | \hat{e}_\ell = e_\ell, \forall \ell \leq k-1\right] = \quad (22)$$

$$P\left[|\hat{a}_{ij}^{(k-1)} - a_{ij}^{(k-1)}| \leq \varepsilon_T | \hat{e}_\ell = e_\ell, \forall \ell \leq k-1\right] \to 1,$$

because $|\hat{a}_{ij}^{(k-1)} - a_{ij}^{(k-1)}| \to |a_{ij}^{(k-1)}| > 0$ in probability and $\varepsilon_T \to 0$.

Consequently, (18) is proved. Let us see (19). We denote by $\hat{S}^{(k-1)}$ and $S^{(k-1)}$ the matrices of the admissible operation (d) performed on $\hat{A}^{(k-1)}$ and $A^{(k-1)}$ respectively. Then,

$$\hat{A}^{(k)} - A^{(k)} = \hat{A}^{(k-1)}\hat{S}^{(k-1)} - A^{(k-1)}S^{(k-1)} = \quad (23)$$

$$= \left[\hat{A}^{(k-1)} - A^{(k-1)}\right]\hat{S}^{(k-1)} + A^{(k-1)}\left[\hat{S}^{(k-1)} - S^{(k-1)}\right]. \quad (24)$$

For row operations, we find a similar identity. Thus, we just need to prove that $\hat{S}^{(k-1)} - S^{(k-1)} = O_{cp}(\xi_T)$ and use that in turn, this implies $\hat{S}^{(k-1)} = O_{cp}(1)$.

The matrix $\hat{S}^{(k-1)} - S^{(k-1)}$ only has a nonzero element, that is the difference between the quotient $\hat{c}$ of the Euclidean division of $\hat{a}_{ij}^{(k-1)}$ between $\hat{a}_{ii}^{(k-1)}$ in $\hat{S}^{(k-1)}$

and its counterpart c from $A^{(k-1)}$. The degrees of the $\hat{a}_{ij}^{(k)}$ are bounded conditionally to $\hat{e}_\ell = e_\ell, \forall \ell \leq k$, so we can use assumption 4 and

$$\hat{a}_{ii}^{(k-1)} - a_{ii}^{(k-1)} = O_{cp}(\xi_T) \tag{25}$$

$$\hat{a}_{ij}^{(k-1)} - a_{ij}^{(k-1)} = O_{cp}(\xi_T), \tag{26}$$

and get that $\hat{c} - c = O_{cp}(\xi_T)$ and thus $\hat{S}^{(k-1)} - S^{(k-1)} = O_{cp}(\xi_T)$.

The case of the admissible operations (c) and (e) are similar, while (a) and (b) are isometries and thus they satisfy trivially the condition. □

Lemma 3. *The Euclidean division and the mapping u are continuous in* $R = \mathbb{R}[z]$.

Proof. If $\delta < 1$, then $|q - q'| < \delta$ entails $\varphi(q') = \varphi(q)$. Thus, the degrees of the polynomials, and consequently the number of operations involved in the division are bounded. Since the algorithm only requires addition, multiplication and division by the lead coefficient of the divisor, the continuity is granted.

The mapping u in this case boils down to calculate the inverse of the leading coefficient of the argument and by definition, the leading coefficient is always nonzero. □

Before proving Proposition 5, we need some preliminary results. In TT, it is proved that for purely nonstationary processes, that is, processes that satisfy a $A(B)y_t = M(B)\varepsilon_t$ such that $\det\Phi(z)$ has only unit roots, the autoregressive least squares estimates are consistent. For processes that are nonstationary, but not purely nonstationary, that is, when $\det A(z)$ has roots on and outside the unit circle, the purely nonstationary part of the estimates (in some sense that is specified in the proof of proposition 5) is consistent.

We will prove that $\tilde{S}(A, 0; \mathbb{R}[z])$ does not depend on the stationary part and that $\hat{\Phi} \to A^*$ such that the $\tilde{S}(A^*, 0; \mathbb{R}[z]) = \tilde{S}(A, 0; \mathbb{R}[z])$ and thus $\tilde{S}(\hat{\Phi}, \varepsilon_t; \mathbb{R}[z]) \xrightarrow{p} \tilde{S}(A, 0; \mathbb{R}[z])$.

Now, we see that if we can decompose the autoregressive polynomial into stable and purely unstable components, only the purely unstable component determines $\tilde{S}(A, 0; \mathbb{R}_1(z))$.

Lemma 4. *Let* $\Phi_s(z)$ *be stable (i.e., without unit roots). Then, for any* $\Phi_n(z)$*, possibly with unit roots,* $\mathscr{S}((\Phi_n^{-1} + \Phi_s^{-1})^{-1}, 0; \mathbb{R}_1(z)) = \mathscr{S}(\Phi_n, 0; \mathbb{R}_1(z))$.

Proof. For ease of notation, we drop the superscript $\cdot^{\Phi}$. Let $\Phi_n = UDV$ and $(\Phi_n^{-1} + \Phi_s^{-1})^{-1} = \Theta^{-1}\Phi$, where Φ and Θ are left-coprime. Then $\Theta^{-1}\Phi = \left[DV\,\mathrm{adj}\Phi_s \det U + \mathrm{adj}U \det\Phi_s\right]^{-1} \cdot \left[DV \det\Phi_s \det U\right]$. By theorem 2.1.1 in Hannan and Deistler (1988), we know that there exists some unimodular matrix C such that $\Phi = CDV \det\Phi_s \det U$. Therefore $(\Phi_n^{-1} + \Phi_s^{-1})^{-1} = \Theta^{-1}CDV \det\Phi_s \det U$ and thus, its SF is D. □

Proof of Proposition 5. We will use the representation $Y_t = FY_{t-1} + a_t$, where $Y_t = (y_t', \ldots, y_{t-p+1}')'$, $a_t = L\Theta(B)\varepsilon_t$, F is the companion matrix of Φ and L comprises the first n columns of I_{np}. Let $J = PFP^{-1}$ be the Jordan form of F. We can decompose it in the stable and non-stable parts as $J = \mathrm{diag}(J_s, J_n)$.

If we call $U_t = PY_t$, then $U_t = (U_{st}, U_{nt})'$, where U_{st} and U_{nt} are the stable and unstable components. In TT it is proved that $\hat{J}_s \to J_s^*$ and $\hat{J}_n \to J_n$, where $J_s^* \neq J_s$ when $\Theta \neq 0$, but we can see that J_s^* is stable, that is, it has all its eigenvalues inside the unit disk. Since U_{st} is stable, $U_{st} = \sum_{k\geq 0} \Psi_k \varepsilon_{t-k}$, with $\sum_k \|\Psi_k\|^2 < +\infty$. Then, $J_s^* = \Gamma(0)^{-1}\Gamma(1)$, where $\Gamma(0) = \sum_{k\geq 0} \Psi_k \Psi_k'$ and $\Gamma(1) = \sum_{k\geq 0} \Psi_{k+1} \Psi_k'$.

Let us see that all eigenvalues of J_s^* have modulus less than one. Let u be an eigenvector and λ its eigenvalue, so $\Gamma(1)u = \lambda\Gamma(0)u$ and then $u'\Gamma(1)u = \lambda u'\Gamma(0)u$. Let us consider the infinite sequences $x = (x_0, x_1, \ldots)$, where $x_j \in \mathbb{R}^n$ endowed with the scalar product $\langle x, y\rangle = \sum_k x_k' y_k$. Then, by defining $a = (\Psi_0' u, \Psi_1' u, \ldots)$ and $b = (\Psi_1' u, \Psi_2' u \ldots)$, then $u'\Gamma(0)u = \langle a, a\rangle$ and $u'\Gamma(1)u = \langle a, b\rangle$. Since $\|b\| \leq \|a\|$, we conclude that $|\lambda| < 1$ unless a and b are linearly dependent, but his implies that $\Psi_k' u = \alpha^k w$. Then, $\alpha = \lambda$ and necessarily $|\alpha| < 1$ because Ψ_k is square-summable.

We can recover the infinite MA representation of y_t as $y_t = L'P^{-1}(1 - JB)^{-1}PL\varepsilon_t$. Since $J = \mathrm{diag}(J_s, J_n)$, then

$$\hat{\Phi}(z)^{-1} \to L'P^{-1} \begin{pmatrix} (1 - J_s^* B)^{-1} & 0 \\ 0 & 0 \end{pmatrix} PL + L'P^{-1} \begin{pmatrix} 0 & 0 \\ 0 & (1 - J_n B)^{-1} \end{pmatrix} PL$$

Since the first part is stable and the second unstable, we conclude by applying lemma 4. □

B Algorithm

For $a \in R$, T_{ij}^{a} is the unitary matrix that multiplied on the right side adds the ith column to the jth one, whereas matrix S_{ij} swaps columns i and j.

for $i = 0$ to $n-1$ **do**
  for $j = 0$ to n **do**
    $A_{ij} \leftarrow u(A_{ii})A_{ij}$
  end for
  $flag_{div} \leftarrow TRUE$
  while $flag_{div}$ **do**
    $flag_{z} \leftarrow TRUE$
    while $flag_{z}$ **do**
      $flag_{row} \leftarrow TRUE$
      while $flag_{row}$ **do**
        MAKE ROW ZEROS
      end while
      $flag_{col} \leftarrow TRUE$
      while $flag_{col}$ **do**
        MAKE COLUMN ZEROS
      end while
      $flag_{z} \leftarrow FALSE$
      for $j = i+1$ to n **do**
        if $A_{ij} \neq 0$ or $A_{ji} \neq 0$ **then**
          $flag_{z} \leftarrow TRUE$
        end if
      end for
    end while
    BOX DIVISIBILITY
  end while
end for

SUBALGORITHM MAKE ROW ZEROS.

$nzcount \leftarrow 0$

for $j = i+1, \ldots, n$ **do**
  if $A_{ij} \neq 0$ **then**
    $nzcount \leftarrow nzcount + 1$
    $r \leftarrow$ remainder from ($A_{ij} = A_{ii}q + r$)
    if $r = 0$ **then**
      $q \leftarrow$ quotient from ($A_{ij} = A_{ii}q + r$)
      $A \leftarrow AT_{ij}^{q}$
      $U \leftarrow (T_{ij}^{q})^{-1}U$
    else
      $q \leftarrow$ quotient from ($A_{ij} = A_{ii}q + r$)
      $A \leftarrow AT_{ij}^{q}S_{ij}$
      $U \leftarrow S_{ij}^{-1}(T_{ij}^{q})^{-1}U$
    end if
  end if
end for
if $nzcount = 0$ **then**
  $flag_{row} \leftarrow FALSE$
end if

SUBALGORITHM MAKE COLUMN ZEROS

$nzcount \leftarrow 0$
for $j = i+1$ to n **do**
  if $A_{ji} \neq 0$ **then**
    $nzcount \leftarrow nzcount + 1$
    $r \leftarrow$ remainder from ($A_{ij} = A_{ii}q + r$)
    if $\|r\| < \varepsilon$ **then**
      $q \leftarrow$ quotient from ($A_{ij} = A_{ii}q + r$)
      $D \leftarrow T_{ij}^{q}(q)A$
      $V \leftarrow V(T_{ij}^{q})^{-1}$
    else
      $q \leftarrow$ quotient from ($A_{ij} = A_{ii}q + r$)
      $A \leftarrow S_{ij}T_{ij}^{q}A$
      $U \leftarrow U(T_{ij}^{q})^{-1}(S_{ij})^{-1}$
    end if

```
    end if
end for
if nzcount = 0 then
    flag_col ← FALSE
end if
```

SUBALGORITHM BOX DIVISIBILITY

```
flag_div ← FALSE
for j = i + 1 to n do
    for k = i + 1 to n do
        r ← remainder from ( A_ik = A_ii q + r)
        if r ≠ 0 then
            flag_div ← TRUE
            D ← S_ij A
            V ← V S_ij^-1
            break for
        end if
    end for
    if flag_div = TRUE then
        break for
    end if
end for
```

References

Barrio Castro, T., and Osborn, D. (2011). HEGY Tests in the presence of moving averages. *Oxford Bulletin of Economics and Statistics*, 73: 691–704. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1468-0084.2011.00633.x/abstract.

Engle, R., and Granger, C. (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica*, 55: 251–76. URL http://www.jstor.org/stable/1913236.

Franchi, M. (2006). A general representation theorem for integrated vector autoregressive processes. Discussion paper 06-16, University of Copenhagen. URL http://www.economics.ku.dk/research/publications/wp/2006/0616.pdf/.

Franchi, M. (2007). The integration order of vector autoregressive processes. *Econometric Theory*, 23: 546–553. URL http://www.jstor.org/stable/4497802.

Granger, C. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 23: 121–130. URL http://www.sciencedirect.com/science/article/pii/0304407681900798.

Granger, C. W., and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of time series analysis*, 1: 15–29. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9892.1980.tb00297.x/abstract.

Granger, C. W. J., and Lee, T. (1989). Investigation of production, sales and inventory relations using multicointegration and non-symmetric error correction models. *Journal of Applied Econometrics*, 4 (S1): S145–S159. URL http://onlinelibrary.wiley.com/doi/10.1002/jae.3950040508/abstract.

Haldrup, N., and Salmon, M. (1998). Representations of I(2) Cointegrated Systems Using the Smith-McMillan Form. *Journal of Econometrics*, 84: 303–325. URL http://www.sciencedirect.com/science/article/pii/S0304407697000882.

Hannan, E. J., and Deistler, M. (1988). *The statistical theory of linear systems*. John Wiley and Sons, New-York.

Hungerford, T. W. (1980). *Algebra*. Springer, New-York.

Hylleberg, S., Engle, R. F., Granger, C. W. J., and Yoo, B. S. (1990). Seasonal integration and cointegration. *Journal of Econometrics*, 44: 215–238. URL http://www.sciencedirect.com/science/article/pii/030440769090080D.

Ing, C.-K., Sin, C., and Yu, S.-H. (2007). Efficient selection of the order of an AR(∞): a unified approach without knowing the order of integratedness. Discussion paper, Institute of Statistical Science, Academia Sinica.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12: 231–254. URL http://www.sciencedirect.com/science/article/pii/0165188988900413.

Johansen, S. (1992). A representation of vector autoregressive processes integrated of order 2. *Econometric Theory*, 8: 188–202. URL http://www.jstor.org/stable/3532439.

Johansen, S., and Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration with applications to the demand for money. *Oxford Bulletin of Economics and Statistics*, 52: 169–210. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1468-0084.1990.mp52002003.x/abstract.

Levin, A., and Lin, C. F. (1993). Unit root tests in panel data: Asymptotic and finite-sample properties. Discussion paper, University of California, San Diego.

Levin, A., Lin, C. F., and Chu, C. S. J. (2002). Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics*, 108: 1–24. URL http://www.sciencedirect.com/science/article/pii/S0304407601000987.

Lewis, R., and Reinsel, G. C. (1985). Prediction of multivariate time series by autoregressive model fitting. *Journal of Multivariate Analysis*, 16: 393–411. URL http://www.sciencedirect.com/science/article/pii/0047259X85900272.

Ling, S., and Li, W. K. (1998). Limiting distributions of maximum likelihood estimators for unstable autoregressive moving-average time series with general autoregressive heteroscedastic errors. *The Annals of Statistics*, 26: 84–125. URL http://www.jstor.org/stable/119980.

Maravall, A. (2008). Notes on programs TRAMO and SEATS. Discussion paper, Bank of Spain. URL http://www.bde.es/f/webbde/SES/servicio/Programas_estadisticos_y_econometricos/Notas_introductorias_TRAMO_SEATS/ficheros/Part_III_Seats.pdf.

Paruolo, P. (1996). On the determination of integration indices in I(2) systems. *Journal of Econometrics*, 72: 313–356. URL http://www.sciencedirect.com/science/article/pii/0304407695017259.

Rudin, W. (1987). *Real and complex analysis*. McGraw-Hill, New York.

Tsay, R. S., and Tiao, G. C. (1990). Asymptotic properties of multivariate nonstationary processes with applications to autoregressions. *The Annals of Statistics*, 18: 220–250. URL http://www.jstor.org/stable/2241542.

Worthington, A. C., and Higgs, H. (2010). Assessing financial integration in the European Union equity markets: Panel unit root and multivariate cointegration and causality evidence. *Journal of Economic Integration*, pages 457–479. URL http://www.jstor.org/stable/23000868.

Yap, S. F., and Reinsel, G. C. (1995). Estimation and testing for unit roots in a partially nonstationary vector autoregressive moving average model. *Journal of the American Statistical Association*, 90: 253–267. URL http://www.jstor.org/stable/2291150.

Yu, J., and Lee, L. (2010). Estimation of unit root spatial dynamic panel data models. *Econometric Theory*, 26: 1332–1362. URL http://www.jstor.org/stable/40800885.