

**Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.**

**Dynamic Factor Models, Cointegration, and Error Correction
Mechanisms**

Matteo Barigozzi, Marco Lippi, and Matteo Luciani

2016-018

Please cite this paper as:

Barigozzi, Matteo, Marco Lippi, and Matteo Luciani (2016). “Dynamic Factor Models, Cointegration, and Error Correction Mechanisms,” Finance and Economics Discussion Series 2016-018. Washington: Board of Governors of the Federal Reserve System, <http://dx.doi.org/10.17016/FEDS.2016.018>.

NOTE: Staff working papers in the Finance and Economics Discussion Series (FEDS) are preliminary materials circulated to stimulate discussion and critical comment. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors. References in publications to the Finance and Economics Discussion Series (other than acknowledgement) should be cleared with the author(s) to protect the tentative character of these papers.

Dynamic Factor Models, Cointegration, and Error Correction Mechanisms

Matteo BARIGOZZI¹

Marco LIPPI²

Matteo LUCIANI³

February 16, 2016

Abstract

The paper studies Non-Stationary Dynamic Factor Models such that: (1) the factors \mathbf{F}_t are $I(1)$ and singular, i.e. \mathbf{F}_t has dimension r and is driven by a q -dimensional white noise, the common shocks, with $q < r$, and (2) the idiosyncratic components are $I(1)$. We show that \mathbf{F}_t is driven by $r - c$ permanent shocks, where c is the cointegration rank of \mathbf{F}_t , and $q - (r - c) < c$ transitory shocks, thus the same result as in the non-singular case for the permanent shocks but not for the transitory shocks. Our main result is obtained by combining the classic Granger Representation Theorem with recent results by Anderson and Deistler on singular stochastic vectors: if $(1 - L)\mathbf{F}_t$ is singular and has *rational* spectral density then, for generic values of the parameters, \mathbf{F}_t has an autoregressive representation with a *finite-degree* matrix polynomial fulfilling the restrictions of a Vector Error Correction Mechanism with c error terms. This result is the basis for consistent estimation of Non-Stationary Dynamic Factor Models. The relationship between cointegration of the factors and cointegration of the observable variables is also discussed.

JEL subject classification: C0, C01, E0.

Key words and phrases: Dynamic Factor Models for $I(1)$ variables, Cointegration for singular vectors, Granger Representation Theorem for singular vectors.

¹m.barigozzi@lse.ac.uk – London School of Economics and Political Science, UK.

²ml@lippi.ws – Einaudi Institute for Economics and Finance, Roma, Italy.

³matteo.luciani@frb.gov – Federal Reserve Board of Governors, Washington DC, USA.

Massimo Franchi, and Rocco Mosconi read previous versions of the paper and gave suggestions for improvements. We also thank the participants to the Workshop on Estimation and Inference Theory for Cointegrated Processes in the State Space Representation, Technische Universität Dortmund, January 2016. Of course we are responsible for any remaining errors. The paper was written while Matteo Luciani was *chargé de recherches* F.R.S.-F.N.R.S., whose financial support he gratefully acknowledges. The views expressed here are those of the authors and do not necessarily reflect those of the Board of Governors or the Federal Reserve System.

1 Introduction

In the last fifteen years Large-Dimensional Dynamic Factor Models (DFM) have become increasingly popular in economic and econometric literature and they are nowadays commonly used by policy institutions, such as Central Banks and Ministries. These models are based on the idea that all the variables in an economic system are driven by a few common (macroeconomic) shocks, their residual dynamics being explained by idiosyncratic components which may result from measurement errors and sectoral or regional shocks.

Formally, each variable in the n -dimensional dataset x_{it} , $i = 1, 2, \dots, n$ can be decomposed into the sum of a common component χ_{it} , and an idiosyncratic component ϵ_{it} : $x_{it} = \chi_{it} + \epsilon_{it}$ (Forni et al., 2000; Forni and Lippi, 2001; Stock and Watson, 2002a,b). In the standard version of the DFM, which is adopted here, the common components are linear combinations of an r -dimensional vector of common factors $\mathbf{F}_t = (F_{1t} \ F_{2t} \ \dots \ F_{rt})'$,

$$\chi_{it} = \lambda_{i1}F_{1t} + \lambda_{i2}F_{2t} + \dots + \lambda_{ir}F_{rt} = \boldsymbol{\lambda}_i\mathbf{F}_t. \quad (1)$$

The vector \mathbf{F}_t is dynamically driven by the q -dimensional non-singular white-noise vector¹ $\mathbf{u}_t = (u_{1t} \ u_{2t} \ \dots \ u_{qt})'$, the common shocks:

$$\mathbf{F}_t = \mathbf{U}(L)\mathbf{u}_t, \quad (2)$$

where $\mathbf{U}(L)$ is an $r \times q$ matrix, (Stock and Watson, 2005; Bai and Ng, 2007; Forni et al., 2009). The dimension n of the dataset is assumed to be large as compared to r and q , which are independent of n , with $q \leq r$. More precisely, all assumptions and results are formulated assuming that both T , the number of observations for each x_{it} , and n , the number of variables, tend to infinity.

The assumption that the vector \mathbf{F}_t is singular, i.e. $r > q$, has received sound empirical support in a number of papers, see Giannone et al. (2005), Amengual and Watson (2007), Forni and Gambetti (2010), and Luciani (2015) for US macroeconomic databases and Barigozzi et al. (2014) for the Euro area. Such results can be easily understood observing that the static equation (1) is just a convenient representation derived from a “primitive” set of dynamic equations linking the common components χ_{it} to the common shocks \mathbf{u}_t . As a simple example, suppose that the variables x_{it} belong to a macroeconomic dataset and are driven by a common one-dimensional cyclical process f_t , such that $(1 - \alpha L)f_t = u_t$, where u_t is scalar white noise, and that the variables x_{it} load f_t dynamically:

$$x_{it} = a_{i0}f_t + a_{i1}f_{t-1} + \epsilon_{it}. \quad (3)$$

In this case representation (1) is obtained by setting $F_{1t} = f_t$, $F_{2t} = f_{t-1}$, $\lambda_{i1} = a_{i0}$, $\lambda_{i2} = a_{i1}$, while equation (2) takes the form

$$\begin{pmatrix} F_{1t} \\ F_{2t} \end{pmatrix} = \begin{pmatrix} (1 - \alpha L)^{-1} \\ (1 - \alpha L)^{-1}L \end{pmatrix} u_t,$$

so that $r = 2$, $q = 1$ and the dynamic equation (3) is replaced by the static representation $x_{it} = \lambda_{i1}F_{1t} + \lambda_{i2}F_{2t} + \epsilon_{it}$. For a general analysis of the relationship between representation (1) and “deeper” dynamic representations like (3), see e.g. Stock and Watson (2005), Forni et al.

¹Usually orthonormality is assumed. This is convenient but not necessary in the present paper.

(2009), see also Section 2.3 below. Singularity of \mathbf{F}_t , i.e. $r > q$, will be assumed throughout the present paper.

If the factors \mathbf{F}_t and the idiosyncratic terms are stationary, and hence the data x_{it} are stationary as well, the factors \mathbf{F}_t and the loadings $\boldsymbol{\lambda}_i$ can be consistently estimated using the first r principal components (Stock and Watson, 2002a,b). The common shocks \mathbf{u}_t and the function $\mathbf{U}(L)$ can then be estimated by a singular VAR for \mathbf{F}_t . Lastly, identification restrictions can be applied to the shocks \mathbf{u}_t and the function $\mathbf{U}(L)$ to obtain structural common shocks and impulse response functions, see Stock and Watson (2005), Forni et al. (2009).

In the present paper and in Barigozzi et al. (2016) we study the Large-Dimensional Non-Stationary Dynamic Factor Model under the assumption that the factors and the idiosyncratic components are $I(1)$, so that the observable variables x_{it} are $I(1)$ as well. Equation (1) remains unchanged while (2) is replaced by

$$(1 - L)\mathbf{F}_t = \mathbf{U}(L)\mathbf{u}_t. \quad (4)$$

In this case, the principal components of the stationary series $(1 - L)x_{it}$ provide an estimate of the differenced factors $(1 - L)\mathbf{F}_t$ and the loadings $\boldsymbol{\lambda}_i$. The factors \mathbf{F}_t can then be recovered by integration, see e.g. Bai and Ng (2004). Exceptions to this common practice are Bai (2004), Peña and Poncela (2004), in which the factors \mathbf{F}_t are directly estimated using the $I(1)$ variables x_{it} .

The main difference with respect to the stationary case arises with the estimation of \mathbf{u}_t and $\mathbf{U}(L)$, or structural common shocks and impulse-response functions. This requires the estimation of a VAR for the $I(1)$ factors \mathbf{F}_t . On the other hand, because \mathbf{F}_t is singular, then \mathbf{F}_t is quite obviously cointegrated (the spectral density of $(1 - L)\mathbf{F}_t$ is singular at all frequencies and therefore at frequency zero). Here we study the autoregressive representations of the singular cointegrated vector \mathbf{F}_t , while estimation is studied in Barigozzi et al. (2016).²

The paper is organized as follows. In Section 2 we recall recent results for singular stochastic vectors with rational spectral density, see Anderson and Deistler (2008a,b), and we discuss cointegration and the cointegration rank for $I(1)$ singular stochastic vectors: c , the cointegration rank, is equal to $r - q$, the minimum due to singularity, plus d , with $0 \leq d < q$.

In Section 3 we obtain the permanent-transitory shock representation in the singular case: \mathbf{F}_t is driven by $r - c = q - d$ permanent and $d = c - (r - q)$ transitory shocks, the same result as in the non-singular case for the permanent shocks but not for the transitory. Then we prove our main results. Assuming rational spectral density for the vector $(1 - L)\mathbf{F}_t$ and therefore that the entries of $\mathbf{U}(L)$ in (4) are rational functions of L , then *for generic values of the parameters of the matrix $\mathbf{U}(L)$* , \mathbf{F}_t has an autoregressive representation fulfilling the restrictions of a Vector Error Correction Mechanism (VECM) with c error terms:

$$\mathbf{A}(L)\mathbf{F}_t = \mathbf{A}^*(L)(1 - L)\mathbf{F}_t + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{F}_{t-1} = \mathbf{h} + \mathbf{R}\mathbf{u}_t, \quad (5)$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are both $r \times c$ and full rank, \mathbf{R} is $r \times q$, $\mathbf{A}(L)$ and $\mathbf{A}^*(L)$ are *finite-degree* matrix polynomials. These results are obtained by combining the Granger Representation Theorem (Engle and Granger, 1987) with Anderson and Deistler's results.

²To our knowledge, the present paper is the first to study cointegration and Error Correction representations for the singular factors of $I(1)$ Dynamic Factor Models. An Error Correction model in the DFM framework is studied in Banerjee et al. (2014a,b). However, their focus is on the relationship between the observable variables and the factors. Their Error Correction term is a linear combination of the variables x_{it} and the factors \mathbf{F}_t , which is stationary if the idiosyncratic components are stationary (so that the x 's and the factors are cointegrated). Because of this and other important differences their results are not directly comparable to those in the present paper.

Section 3 also contains an exercise carried on with simulated singular $I(1)$ vectors. We compare the results obtained by estimating an unrestricted VAR in the levels and a VECM. Though limited to a simple example, the results confirm what has been found for non-singular vectors, that is under cointegration the long-run features of impulse-response functions are better estimated using a VECM rather than an unrestricted VAR in the levels (Phillips, 1998).

In Section 4 we analyse cointegration of the observable variables x_{it} . Our results on cointegration of the factors \mathbf{F}_t have the obvious implication that p -dimensional subvectors of the n -dimensional common-component vector $\boldsymbol{\chi}_t$, with $p > q - d$, are cointegrated. Stationarity of the idiosyncratic components would imply that all p -dimensional subvectors of the n -dimensional dataset \mathbf{x}_t are cointegrated. For example, if $q = 3$ and $d = 1$, then all 3-dimensional subvectors in the dataset are cointegrated, a kind of regularity that we do not observe in actual large macroeconomic datasets. This motivates our assumption that the idiosyncratic components are $I(1)$ (some of the variables ϵ_{it} are $I(1)$). Section 5 concludes. Some long proofs, a discussion of some non-uniqueness problems arising with singularity and details on the simulations are collected in the Appendix.

2 Stationary and non-stationary Dynamic Factor Models

2.1 The Factors and Idiosyncratic Components are Stationary

Consider the Dynamic Factor Model

$$\mathbf{x}_t = \boldsymbol{\chi}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\chi}_t = \boldsymbol{\Lambda}\mathbf{F}_t, \quad (6)$$

where: (1) the observables \mathbf{x}_t , the common components $\boldsymbol{\chi}_t$, and the idiosyncratic components $\boldsymbol{\epsilon}_t$ are n -dimensional vectors, (2) \mathbf{F}_t is an r -dimensional vector of common factors, with r independent of n , (3) $\boldsymbol{\Lambda}$ is an $n \times r$ matrix, (4) \mathbf{F}_t is driven by a q -dimensional zero-mean white noise vector process \mathbf{u}_t , the common shocks, with $q < r$, (5) ϵ_{it} and u_{js} are orthogonal for all $i = 1, 2, \dots, n$, $j = 1, 2, \dots, q$, $t, s \in \mathbb{Z}$. Other assumptions concerning the asymptotic properties of the model, for $n \rightarrow \infty$, will not be used here and are therefore not reported (see the literature mentioned in the Introduction). The results in the present and the next section, though stated for the vector of the factors \mathbf{F}_t and the common shocks \mathbf{u}_t , hold for any singular stochastic vector under the assumptions specified below.

As observed in the Introduction, with some exceptions, the theory of model (6) has been developed under the assumption that \mathbf{x}_t , $\boldsymbol{\chi}_t$, $\boldsymbol{\epsilon}_t$ and \mathbf{F}_t are stationary. In addition to (6), it is often assumed that \mathbf{F}_t has a reduced-rank VAR representation:

$$\mathbf{A}(L)\mathbf{F}_t = \mathbf{R}\mathbf{u}_t, \quad (7)$$

where $\mathbf{A}(L)$ is a finite-degree $r \times r$ matrix polynomial and \mathbf{R} is $r \times q$. Moreover, it is well known that $\boldsymbol{\Lambda}$ and \mathbf{F}_t can be estimated by principal components, while an estimate of $\mathbf{A}(L)$, \mathbf{R} and \mathbf{u}_t can be obtained by standard techniques. Inversion of $\mathbf{A}(L)$ provides an estimate of the impulse-response functions of the observables to the common shocks:

$$\mathbf{x}_t = \boldsymbol{\Lambda}\mathbf{A}(L)^{-1}\mathbf{R}\mathbf{u}_t + \boldsymbol{\epsilon}_t.$$

Structural shocks and structural impulse-response functions can then be obtained, respectively, as $\mathbf{w}_t = \mathbf{Q}\mathbf{u}_t$ and $\boldsymbol{\Lambda}\mathbf{A}(L)^{-1}\mathbf{R}\mathbf{Q}^{-1}$, where the $q \times q$ matrix \mathbf{Q} is determined in the same way as in Structural VARs (Stock and Watson, 2005; Forni et al., 2009).

The VAR representation (7) has a standard motivation as an approximation to an infinite autoregression with exponentially declining coefficients. However, as stated above, \mathbf{F}_t has reduced rank. Under reduced rank and rational spectral density for \mathbf{F}_t , Anderson and Deistler (2008a,b) prove that *generically* \mathbf{F}_t has a finite-degree autoregressive representation, so that no approximation argument is needed to motivate (7). A formal statement of this result requires the following definitions.

Definition 1. (Rational reduced-rank family) Assume that $r > q > 0$ and let \mathcal{G} be a set of ordered couples $(\mathbf{S}(L), \mathbf{C}(L))$, where:

- (i) $\mathbf{C}(L)$ is an $r \times q$ polynomial matrix of degree $s_1 \geq 0$.
- (ii) $\mathbf{S}(L)$ is an $r \times r$ polynomial matrix of degree $s_2 \geq 0$. $\mathbf{S}(0) = \mathbf{I}_r$.
- (iii) Denoting by \mathbf{p} the vector containing the $\lambda = rq(s_1 + 1) + r^2s_2$ coefficients of the entries of $\mathbf{C}(L)$ and $\mathbf{S}(L)$, we assume that $\mathbf{p} \in \Pi$, where Π is an open subset of \mathbb{R}^λ , and that for $\mathbf{p} \in \Pi$, if $\det(\mathbf{S}(z)) = 0$, then $|z| > 1$.

We say that the family of weakly stationary stochastic processes

$$\mathbf{F}_t = \mathbf{S}(L)^{-1}\mathbf{C}(L)\mathbf{u}_t, \quad (8)$$

where \mathbf{u}_t is a q -dimensional white noise with non-singular variance-covariance matrix and $(\mathbf{S}(L), \mathbf{C}(L))$ belongs to \mathcal{G} , is a rational reduced-rank family.

The notation \mathbf{F}_t^p , $\mathbf{C}^p(L)$, etc., though more rigorous, would be heavy and not really necessary. We use it only once in the proof of Proposition 2.

Note that (8) is the unique stationary solution of the ARMA equation

$$\mathbf{S}(L)\mathbf{F}_t = \mathbf{C}(L)\mathbf{u}_t. \quad (9)$$

Definition 2. (Genericity) Suppose that a statement $Q(\mathbf{p})$ depends on $\mathbf{p} \in \mathcal{A}$, where \mathcal{A} is an open subset of \mathbb{R}^λ . Then $Q(\mathbf{p})$ holds generically in \mathcal{A} if the subset \mathcal{N} of \mathcal{A} where it does not hold is nowhere dense in \mathcal{A} , i.e. the closure of \mathcal{N} has no internal points.

Proposition 1. (Anderson and Deistler) (I) Suppose that $\mathbf{V}(L)$ is an $r \times q$ matrix whose entries are rational functions of L , with $r > q$. If $\mathbf{V}(L)$ is zeroless, i.e. has rank q for all complex numbers z , then $\mathbf{V}(L)$ has a finite-degree stable left inverse, i.e. there exists a finite-degree polynomial $r \times r$ matrix $\mathbf{W}(L)$, such that (a) $\det(\mathbf{W}(z)) = 0$ implies $|z| > 1$, (b) $\mathbf{W}(L)\mathbf{V}(L) = \mathbf{V}(0)$. (II) Let $\mathbf{F}_t = \mathbf{S}(L)^{-1}\mathbf{C}(L)\mathbf{u}_t$ be a rational reduced-rank family with parameter set Π . For generic values of the parameters $\mathbf{p} \in \Pi$, $\mathbf{S}(z)^{-1}\mathbf{C}(z)$ is zeroless. In particular, generically $\mathbf{S}(1)^{-1}\mathbf{C}(1)$ and $\mathbf{S}(0)^{-1}\mathbf{C}(0) = \mathbf{C}(0)$ have full rank q .

For statement (I) see Deistler et al. (2010), Theorem 3. Statement (II) is a modified version of their Theorem 2. They obtain genericity with respect to the parameters of the state-space representation of (9), whereas in statement (II) we refer to the original parameters of the matrix polynomials in (9) (see Forni et al. (2015) for a proof). Another version of Anderson and Deistler's Theorem 2 is Proposition 2 in the present paper. Both statement (I) and our Proposition 2 are crucial for the proof of our main results in Propositions 3 and 4.

2.2 Non-Stationary factors

Suppose now that the vector \mathbf{F}_t and some of the idiosyncratic components ϵ_{it} are non-stationary while $(1 - L)\mathbf{F}_t$ and all the variables $(1 - L)\epsilon_{it}$ are stationary. The common

practice in this case consists in reducing the data x_{it} to stationarity by taking first differences and estimating the differenced factors $(1 - L)\mathbf{F}_t$ by means of the principal components of the variables $(1 - L)x_{it}$. Then usually impulse-response functions are obtained by estimating a VAR for $(1 - L)\mathbf{F}_t$. Of course this implies possible misspecification if \mathbf{F}_t is cointegrated, which is always the case when \mathbf{F}_t is singular.

To analyse cointegration and the autoregressive representations of the singular non-stationary vector \mathbf{F}_t let us firstly recall the definitions of $I(0)$, $I(1)$ and cointegrated vectors.

In the present paper we only consider stochastic vectors that are either weakly stationary with a rational spectral density matrix or such that their first difference is weakly stationary with rational spectral density. Assume that the n -dimensional vector \mathbf{y}_t is weakly stationary with rational spectral density, denoted by $\boldsymbol{\Sigma}_y(\theta)$. The matrix $\boldsymbol{\Sigma}_y(\theta)$ has constant rank ρ , with $\rho \leq n$, i.e. has the same rank ρ for θ almost everywhere in $[-\pi, \pi]$. We say that ρ is the rank of \mathbf{y}_t . Moreover, \mathbf{y}_t has moving average representations

$$\mathbf{y}_t = \mathbf{V}(L)\mathbf{v}_t, \quad (10)$$

where \mathbf{v}_t is a non singular ρ -dimensional white noise, $\mathbf{V}(L)$ is an $n \times \rho$ matrix whose entries are rational functions of L with no poles of modulus less or equal to unity. If $\text{rank}(\mathbf{V}(z)) = \rho$ for $|z| < 1$, then \mathbf{v}_t belongs to the space spanned by \mathbf{y}_τ , with $\tau \leq t$, and representation (10), as well as \mathbf{v}_t , is called *fundamental* (see Rozanov (1967), pp. 43–7)³.

Let \mathbf{z}_t be an r -dimensional weakly stationary with rational spectral density and assume that $\mathbf{z}_t \in L_2(\Omega, \mathcal{F}, P)$. Consider the difference equation

$$(1 - L)\boldsymbol{\zeta}_t = \mathbf{z}_t, \quad (11)$$

in the unknown process $\boldsymbol{\zeta}_t$. A solution of (11) is

$$\tilde{\mathbf{y}}_t = \begin{cases} \mathbf{z}_1 + \mathbf{z}_2 + \cdots + \mathbf{z}_t, & \text{for } t > 0 \\ \mathbf{0}, & \text{for } t = 0 \\ -(\mathbf{z}_0 + \mathbf{z}_{-1} \cdots + \mathbf{z}_{t+1}), & \text{for } t < 0. \end{cases} \quad (12)$$

All the solutions of (11) are $\mathbf{y}_t = \tilde{\mathbf{y}}_t + \mathbf{W}$, where \mathbf{W} is any r -dimensional stochastic vector belonging to $L_2(\Omega, \mathcal{F}, P)$, that is the particular solution $\tilde{\mathbf{y}}_t$ plus any solution of $(1 - L)\boldsymbol{\zeta}_t = \mathbf{0}$.

Definition 3. ($I(0)$, $I(1)$ and cointegrated vectors)

$I(0)$. The n -dimensional vector stochastic process \mathbf{y}_t is $I(0)$ if it is weakly stationary with rational spectral density $\boldsymbol{\Sigma}_y(\theta)$ and $\boldsymbol{\Sigma}_y(0) \neq \mathbf{0}$.

Integrated process of order 1, $I(1)$. The n -dimensional vector stochastic process \mathbf{y}_t is $I(1)$ if there exists an n -dimensional process \mathbf{z}_t , weakly stationary with rational spectral density, such that \mathbf{y}_t is a solution of the equation $(1 - L)\boldsymbol{\zeta}_t = \mathbf{z}_t$. The rank of \mathbf{y}_t is defined as the rank of \mathbf{z}_t .

Cointegration. Assume that the n -dimensional stochastic vector \mathbf{y}_t is $I(1)$ and denote by $\boldsymbol{\Sigma}_{\Delta y}(\theta)$ the spectral density of $(1 - L)\mathbf{y}_t$. The vector \mathbf{y}_t is cointegrated with cointegration rank c , with $0 < c < n$, if $\text{rank}(\boldsymbol{\Sigma}_{\Delta y}(0)) = n - c$. If ρ is the rank of \mathbf{y}_t , then $c \geq n - \rho$.

³When $n = p$, the condition that $\text{rank}(\mathbf{V}(z)) = p$ for $|z| < 1$ becomes $\det(\mathbf{V}(z)) \neq 0$ for $|z| < 1$.

Some comments are in order.

- C1. If \mathbf{y}_t has representation (10), because $\boldsymbol{\Sigma}_y(0) = (2\pi)^{-1}\mathbf{V}(1)\boldsymbol{\Gamma}_v\mathbf{V}(1)'$, where $\boldsymbol{\Gamma}_v$ is the covariance matrix of \mathbf{v}_t , \mathbf{y}_t is $I(0)$ if and only if $\mathbf{V}(1) \neq \mathbf{0}$.
- C2. Under the parameterization in Definition 1, for generic values of the parameters in Π , we have $\text{rank}(\mathbf{S}(1)^{-1}\mathbf{C}(1)) = q$ (see proposition 1(II)), so that

$$\text{rank}(\boldsymbol{\Sigma}_F(0)) = \text{rank}((2\pi)^{-1}\mathbf{S}(1)^{-1}\mathbf{C}(1)\boldsymbol{\Gamma}_u\mathbf{C}(1)'[\mathbf{S}(1)']^{-1}) = q,$$

where $\boldsymbol{\Gamma}_u$ is the covariance matrix of \mathbf{u}_t . Thus, generically, \mathbf{F}_t has rank q and is $I(0)$.

- C3. Assume that \mathbf{y}_t is such that $(1-L)\mathbf{y}_t$ is weakly stationary with rational spectral density and that

$$(1-L)\mathbf{y}_t = \mathbf{V}(L)\mathbf{v}_t \tag{13}$$

is one of its moving average representations. The process \mathbf{y}_t is $I(1)$ if and only if $\mathbf{V}(1) \neq \mathbf{0}$. Thus our definitions of $I(0)$ and $I(1)$ processes are equivalent to Definitions 3.2, and 3.3 in Johansen (1995), p. 35, with two minor differences: our assumption of rational spectral density and the time span of stochastic processes, $t \in \mathbb{Z}$ in the present paper, $t = 0, 1, \dots$ in Johansen's book.

- C4. Note that \mathbf{y}_t can be $I(1)$ even though some of its coordinate processes are $I(0)$. In particular, assuming that the idiosyncratic vector $\boldsymbol{\epsilon}_t$ is $I(1)$ does not prevent that some of the coordinates ϵ_{it} is $I(0)$.
- C5. If \mathbf{y}_t is $I(1)$ and cointegrated with cointegration rank c , there exist c linearly independent $n \times 1$ vectors \mathbf{c}_j , $j = 1, \dots, c$, such that the spectral density of $\mathbf{c}_j'(1-L)\mathbf{y}_t$ vanishes at frequency zero. The vectors \mathbf{c}_j are called cointegration vectors. Of course a set of cointegration vectors \mathbf{c}_j , $j = 1, \dots, c$, can be replaced by the set \mathbf{d}_j , $j = 1, \dots, c$, where the vectors \mathbf{d}_j are c independent linear combinations of the vectors \mathbf{c}_j .
- C6. In the literature on integrated and cointegrated vectors, the expression “ $I(1)$ process” is often ambiguous, sometimes it refers to a specific process \mathbf{y}_t which solves $(1-L)\boldsymbol{\zeta}_t = \mathbf{z}_t$, sometimes to the whole class $\mathbf{y}_t = \tilde{\mathbf{y}}_t + \mathbf{W}$. This minor abuse of language is very convenient and usually does not cause misunderstandings, see comment C7 below, Section 3.3, Proposition 3 in particular.
- C7. If \mathbf{y}_t is $I(1)$, cointegrated and has representation (13), the cointegration rank of \mathbf{y}_t is c if and only if the rank of $\mathbf{V}(1)$ is $n - c$. Moreover \mathbf{c} is a cointegration vector for \mathbf{y}_t if and only if $\mathbf{c}'\mathbf{V}(1) = \mathbf{0}$. In Appendix A.1 we show that if \mathbf{c} is a cointegration vector for \mathbf{y}_t , then \mathbf{y}_t can be determined (that is, a member of the class containing \mathbf{y}_t can be determined) such that $\mathbf{c}'\mathbf{y}_t$ is weakly stationary with rational spectral density. Thus our definition of cointegration is equivalent to that in Johansen (1995), p. 37.
- C8. Let \mathbf{y}_t be n -dimensional, $I(1)$ with rank $\rho < n$. The spectral density $\boldsymbol{\Sigma}_{\Delta y}(\theta)$ has rank ρ almost everywhere in $[-\pi, \pi]$, so that there exist vectors $\mathbf{d}_j(e^{-i\theta})$, $j = 1, 2, \dots, n - \rho$, such that $\mathbf{d}_j(e^{-i\theta})\boldsymbol{\Sigma}_{\Delta y}(\theta) = \mathbf{0}$ for all θ . In the time domain, $\mathbf{d}_j(L)(1-L)\mathbf{y}_t = \mathbf{0}$. Such exact linear dynamic relationships between the coordinates of \mathbf{y}_t (not possible when $\rho = n$) are closely linked to the non-uniqueness of autoregressive representations for singular vectors (see below in this section and Appendix B).

C9. Definition 3, Cointegration, does not rule out that an eigenvector of $\Sigma_{\Delta y}(\theta)$ constant, i.e. does not depend on θ . If \mathbf{d} is such an eigenvector, $\mathbf{d}(1-L)\mathbf{y}_t = 0$, which is a degenerate case of cointegration (again, not possible when $\rho = n$). More on this in comment C3 on Definition 4.

Now, suppose that \mathbf{F}_t is $I(1)$, that

$$(1-L)\mathbf{F}_t = \mathbf{S}(L)^{-1}\mathbf{C}(L)\mathbf{u}_t = \mathbf{U}(L)\mathbf{u}_t$$

and that the rank of the $r \times q$ matrix $\mathbf{U}(z)$ is q for almost all $z \in \mathbb{C}$, i.e. that the rank of \mathbf{F}_t is q . Then:

- (i) Obviously, as already observed in the definition of cointegration, \mathbf{F}_t has at least $r - q$ cointegration vectors: $c \geq r - q$.
- (ii) If we assume that the couple $(\mathbf{S}(L), \mathbf{C}(L))$ is parameterized as in Definition 1, then, by Proposition 1(II), we can argue that generically $\mathbf{S}(z)^{-1}\mathbf{C}(z)$ has full rank q for all z and therefore the cointegration rank of \mathbf{F}_t is generically $r - q$. However, the rank of $\mathbf{S}(z)^{-1}\mathbf{C}(z)$ at $z = 1$ has a special interpretation as the number of long-run equilibrium relationships between the processes F_{ft} . Such number usually has a theoretical or behavioral motivation, so that it cannot be modified by any genericity argument. As a consequence we adopt a different parameterization for families of $I(1)$ vectors in which the cointegration rank c is fixed, with $r > c \geq r - q$, see Definition 4.
- (iii) If the family has $c = r - q$, then generically $\text{rank}(\mathbf{S}(1)^{-1}\mathbf{C}(1)) = q$ and Proposition 1(I) can be applied. In spite of cointegration, generically $(1-L)\mathbf{F}_t$ has a finite-degree autoregressive representation

$$\mathbf{A}(L)(1-L)\mathbf{F}_t = \mathbf{C}(0)\mathbf{u}_t. \quad (14)$$

- (iv) If the family has $c > r - q$, then no autoregressive representation exists for $(1-L)\mathbf{F}_t$, finite or infinite. However, we prove that, if c is equal or greater than $r - q$, generically \mathbf{F}_t has a representation as a VECM.

$$\mathbf{A}(L)\mathbf{F}_t = \mathbf{A}^*(L)(1-L)\mathbf{F}_t + \mathbf{A}(1)\mathbf{F}_{t-1} = \mathbf{h} + \mathbf{C}(0)\mathbf{u}_t, \quad (15)$$

where the rank of $\mathbf{A}(1)$ is c , and $\mathbf{A}(L)$ and $\mathbf{A}^*(L)$ are *finite-degree* matrix polynomials.

- (v) In the singular case the autoregressive representation of \mathbf{F}_t is in general not unique (as it is the case with full-rank vectors). For example, when $c = r - q$, (14) and (15) are two different autoregressive representations for \mathbf{F}_t , the first with no error terms, the second with $r - q$ error terms. However, as we show in Appendix B, different autoregressive representations of \mathbf{F}_t produce the same impulse-response functions. Proposition 3 in the next section proves the existence of representation (15), which has the maximum number of error terms.

The existence of representation (15) for reduced-rank $I(1)$ vectors, where $\mathbf{A}^*(L)$ is of finite degree, is our main result. Its proof combines the Granger Representation Theorem with the results summarized in Proposition 1.

2.3 “Trivial” and “primitive” cointegration vectors of \mathbf{F}_t

Denote by d the number of cointegrating vectors exceeding the minimum $r - q$, so that $c = r - q + d$. Of course $q > d \geq 0$. There is an interesting class of vectors \mathbf{F}_t for which it is possible to distinguish between $r - q$ cointegrating vectors that merely arise in the construction of \mathbf{F}_t , and d additional cointegrating vectors with a possible structural interpretation.

Let \mathbf{f}_t be a non-singular q -dimensional vector of primitive factors \mathbf{f}_t with representation

$$(1 - L)\mathbf{f}_t = \mathbf{U}_f(L)\mathbf{u}_t, \quad (16)$$

and assume that the variables x_{it} load \mathbf{f}_t and its lags up to the p -th:

$$x_{it} = \mathbf{a}_{i0}\mathbf{f}_t + \mathbf{a}_{i1}\mathbf{f}_{t-1} + \cdots + \mathbf{a}_{ip}\mathbf{f}_{t-p} + \epsilon_{it}. \quad (17)$$

This is a generalization of the example used in the Introduction to motivate the singularity of the vector \mathbf{F}_t , see equation (3). Model (16)–(17) is transformed into the standard form (1) by introducing the r dimensional vector

$$\mathbf{F}_t = (\mathbf{f}'_t \quad \mathbf{f}'_{t-1} \quad \cdots \quad \mathbf{f}'_{t-p})',$$

where $r = q(p + 1)$. We have

$$\mathbf{x}_t = \mathbf{\Lambda}\mathbf{F}_t + \boldsymbol{\epsilon}_t, \quad \mathbf{\Lambda} = \begin{pmatrix} \mathbf{a}_{10} & \mathbf{a}_{11} & \cdots & \mathbf{a}_{1p} \\ \mathbf{a}_{20} & \mathbf{a}_{21} & \cdots & \mathbf{a}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_{n0} & \mathbf{a}_{n1} & \cdots & \mathbf{a}_{np} \end{pmatrix}, \quad (1-L)\mathbf{F}_t = \begin{pmatrix} \mathbf{U}_f(L) \\ L\mathbf{U}_f(L) \\ \vdots \\ L^p\mathbf{U}_f(L) \end{pmatrix} \mathbf{u}_t = \mathbf{U}(L)\mathbf{u}_t. \quad (18)$$

It is immediately seen that \mathbf{F}_t has $r - q = qp$ cointegration vectors

$$\mathbf{t}^{h,k} = \begin{pmatrix} t_1^{h,k} & t_2^{h,k} & \cdots & t_r^{h,k} \end{pmatrix},$$

$h = 1, \dots, q, k = 1, \dots, p$, where

$$t_j^{h,k} = \begin{cases} 1 & \text{if } j = h \\ -1 & \text{if } j = kq + h \\ 0 & \text{otherwise.} \end{cases}$$

These can be called trivial cointegration vectors, as they merely result from the construction of \mathbf{F}_t by stacking the vectors \mathbf{f}_{t-k} , $k = 0, \dots, p$.

On the other hand, if \mathbf{f}_t is cointegrated, with cointegration rank c_f , $q > c_f > 0$, and cointegrating vectors \mathbf{s}^m , $m = 1, \dots, c_f$, then the cointegration rank of \mathbf{F}_t is $c = r - q + c_f$ with the additional c_f cointegrating vectors obtained by augmenting each \mathbf{s}^m with qp zeros. Thus cointegration of the primitive factors \mathbf{f}_t naturally translates into cointegration of \mathbf{F}_t .

Two observations are in order. Firstly, the above distinction between primitive and trivial cointegration has heuristic interest but is limited to representation (18), the latter being only one among infinitely many equivalent standard representations of (16)–(17). If \mathbf{H} is an $r \times r$ invertible matrix,

$$\mathbf{x}_t = \mathbf{\Lambda}^*\mathbf{F}_t^* + \boldsymbol{\epsilon}_t, \quad (1 - L)\mathbf{F}_t^* = \mathbf{U}^*(L)\mathbf{u}_t,$$

where $\mathbf{F}_t^* = \mathbf{H}^{-1}\mathbf{F}_t$, $\mathbf{\Lambda}^* = \mathbf{\Lambda}\mathbf{H}$, $\mathbf{U}^*(L) = \mathbf{H}^{-1}\mathbf{U}(L)$, is another factor representation for the variables x_{it} . The cointegrating vectors of \mathbf{F}_t^* are linear combinations of the vectors $\mathbf{t}^{h,k}$ and \mathbf{s}^m , so that primitive and trivial cointegration get mixed. In particular, if the model is estimated by principal components of the variables x_{it} , in general the estimated factors $\hat{\mathbf{F}}_t$ approximate the space spanned by \mathbf{F}_t , not \mathbf{F}_t^* itself, so that no distinction between trivial and primitive cointegrating vectors of $\hat{\mathbf{F}}_t$ is possible.

Secondly, as the elementary example below shows, not all vectors \mathbf{F}_t can be put in the form (16)–(17). Let $r = 2$, $q = 1$ and

$$\mathbf{U}^*(L) = \begin{pmatrix} 1 + L \\ L^2 \end{pmatrix}.$$

Suppose that there exists an invertible matrix

$$\mathbf{H} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix},$$

such that

$$\mathbf{U}(L) = \mathbf{H}\mathbf{U}^*(L) = \begin{pmatrix} \alpha + \alpha L + \beta L^2 \\ \gamma + \gamma L + \delta L^2 \end{pmatrix}$$

has the form in (18), third equation. The second row of $\mathbf{U}(L)$ would equal the first multiplied by L :

$$\gamma + \gamma L + \delta L^2 = (\alpha + \alpha L + \beta L^2)L = \alpha L + \alpha L^2 + \beta L^3.$$

This implies that $\gamma = \beta = \alpha = \delta = 0$. Thus no representation (16)–(17) exists for $\mathbf{U}^*(L)$.

3 Representation theory for reduced rank $I(1)$ vectors

3.1 Families of cointegrated vectors

Consider the equation

$$(1 - L)\boldsymbol{\zeta}_t = \mathbf{S}(L)^{-1}\mathbf{C}(L)\mathbf{u}_t = \mathbf{U}(L)\mathbf{u}_t, \quad (19)$$

where $\det \mathbf{S}(L)$ has no roots of modulus less or equal to unity and $\mathbf{C}(1) \neq \mathbf{0}$, so that $\mathbf{U}(1) \neq \mathbf{0}$. Suppose that $u_{jt} \in L_2(\Omega, \mathcal{F}, P)$, for $j = 1, \dots, q$, where (Ω, \mathcal{F}, P) is a probability space. It is easily seen that all the solutions of (19) are the processes

$$\mathbf{F}_t = \tilde{\mathbf{F}}_t + \mathbf{W}, \quad t \in \mathbb{Z}, \quad (20)$$

where \mathbf{W} is an r -dimensional stochastic vector with $W_k \in L_2(\Omega, \mathcal{F}, P)$, $f = 1, \dots, r$, and

$$\tilde{\mathbf{F}}_t = \mathbf{U}(L)\boldsymbol{\nu}_t = \mathbf{S}(L)^{-1}\mathbf{C}(L)\boldsymbol{\nu}_t, \quad \text{where} \quad \boldsymbol{\nu}_t = \begin{cases} \mathbf{u}_1 + \mathbf{u}_2 + \dots + \mathbf{u}_t, & \text{for } t > 0 \\ \mathbf{0}, & \text{for } t = 0 \\ -(\mathbf{u}_0 + \mathbf{u}_{-1} + \dots + \mathbf{u}_{t+1}), & \text{for } t < 0. \end{cases} \quad (21)$$

Because \mathbf{F}_t is a solution of (19), $(1 - L)\mathbf{F}_t = \mathbf{U}(L)\mathbf{u}_t$ is stationary with rational spectral density. Moreover, as we assume $\mathbf{C}(1) \neq \mathbf{0}$, so that $\mathbf{U}(1) = \mathbf{S}(1)^{-1}\mathbf{C}(1) \neq \mathbf{0}$, then \mathbf{F}_t is $I(1)$.

Now, because $\mathbf{S}(1)^{-1}$ is a non-singular $r \times r$ matrix, the cointegration rank of \mathbf{F}_t only depends on the rank of $\mathbf{C}(1)$. Precisely, if c is the cointegration rank of \mathbf{F}_t , then $c = r -$

rank($\mathbf{C}(1)$), so that $r > c \geq r - q$. Moreover, there exist an $r \times (r - c)$ matrix $\boldsymbol{\xi}$ and a $q \times (r - c)$ matrix $\boldsymbol{\eta}$, both of full rank $r - c \leq q$, such that

$$\mathbf{C}(1) = \boldsymbol{\xi}\boldsymbol{\eta}', \quad (22)$$

see Lancaster and Tismenetsky (1985, p. 97, Proposition 3). The matrix $\mathbf{C}(L)$ has the (finite) Taylor expansion

$$\mathbf{C}(L) = \mathbf{C}(1) - (1 - L)\mathbf{C}'(1) + \frac{1}{2}(1 - L)^2\mathbf{C}''(1) - \dots$$

Gathering all terms after the second and using (22),

$$\mathbf{C}(L) = \boldsymbol{\xi}\boldsymbol{\eta}' - (1 - L)\mathbf{C}'(1) + (1 - L)^2\mathbf{C}_1(L), \quad (23)$$

where $\mathbf{C}_1(L)$ is a polynomial matrix.

Representation (23) can be used for a very convenient parameterization of $\mathbf{C}(L)$.

Definition 4. (*Rational reduced-rank I(1) family with cointegration rank c*) Assume that $r > q > 0$, $r > c \geq r - q$ and let \mathcal{G} be a set of couples $(\mathbf{S}(L), \mathbf{C}(L))$, where:

(i) The matrix $\mathbf{C}(L)$ has the parameterization

$$\mathbf{C}(L) = \boldsymbol{\xi}\boldsymbol{\eta}' + (1 - L)\mathbf{D} + (1 - L)^2\mathbf{E}(L), \quad (24)$$

where $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are $r \times (r - c)$ and $q \times (r - c)$ respectively, \mathbf{D} is an $r \times q$ matrix and $\mathbf{E}(L)$ is an $r \times q$ matrix polynomial of degree $s_1 \geq 0$.

(ii) $\mathbf{S}(L)$ is an $r \times r$ polynomial matrix of degree $s_2 \geq 0$. $\mathbf{S}(0) = \mathbf{I}_r$.

(iii) Denoting by \mathbf{p} the vector containing the $\lambda = (r - c)(r + q) + rq(1 + s_1) + r^2s_2$ coefficients of the matrices in (24) and in $\mathbf{S}(L)$, we assume that $\mathbf{p} \in \Pi$, where Π is an open subset of \mathbb{R}^λ , and that for $\mathbf{p} \in \Pi$, if $\det(\mathbf{S}(z)) = 0$ then $|z| > 1$.

We say that the family of processes \mathbf{F}_t , such that $(1 - L)\mathbf{F}_t = \mathbf{S}(L)^{-1}\mathbf{C}(L)\mathbf{u}_t$, where \mathbf{u}_t is a q -dimensional non-singular white noise and $(\mathbf{S}(L), \mathbf{C}(L))$ belongs to \mathcal{G} , is a rational reduced-rank I(1) family with cointegration rank c .

Three comments are in order.

- C1. Generically the matrices $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ have full rank $r - c$, so that \mathbf{F}_t is I(1) with cointegration rank c .
- C2. In Remark 2 in Appendix A.2, we show that generically \mathbf{F}_t has rank q .
- C3. Suppose that $r = 4$, $q = 1$, $s_1 = s_2 = 0$. For all $\mathbf{p} \in \Pi$, there exists a vector \mathbf{d} orthogonal to the 4-dimensional columns $\boldsymbol{\xi}\boldsymbol{\eta}'$, \mathbf{D} , \mathbf{E}_0 . Thus $\mathbf{d}'\mathbf{C}(L) = \mathbf{0}$ and therefore $\mathbf{d}(1 - L)\mathbf{F}_t = 0$, the degenerate case of cointegration mentioned in comment C9 on Definition 3. However, if $s_2 > 0$ or s_1 is big enough as compared to r , degenerate cointegration can be ruled out generically.

Denoting by $\boldsymbol{\xi}_\perp$ an $r \times c$ matrix whose columns are linearly independent and orthogonal to all columns of $\boldsymbol{\xi}$, the columns of $\boldsymbol{\xi}_\perp$ and $\check{\boldsymbol{\xi}}_\perp = \mathbf{S}'(1)\boldsymbol{\xi}_\perp$ are full sets of independent cointegrating vectors for $\mathbf{S}(L)\mathbf{F}_t$ and \mathbf{F}_t respectively.

3.2 Permanent and transitory shocks

Let \mathbf{F}_t be a rational reduced-rank $I(1)$ family with cointegration rank $c = r - q + d$, $q > d \geq 0$. Let $\boldsymbol{\eta}_\perp$ be a $q \times d$ matrix whose columns are independent and orthogonal to the columns of $\boldsymbol{\eta}$, and let

$$\bar{\boldsymbol{\eta}} = \boldsymbol{\eta}(\boldsymbol{\eta}'\boldsymbol{\eta})^{-1}, \quad \bar{\boldsymbol{\eta}}_\perp = \boldsymbol{\eta}_\perp(\boldsymbol{\eta}'_\perp\boldsymbol{\eta}_\perp)^{-1}.$$

Defining $\mathbf{v}_{1t} = \boldsymbol{\eta}'_\perp \mathbf{u}_t$, and $\mathbf{v}_{2t} = \boldsymbol{\eta}' \mathbf{u}_t$, we have

$$\mathbf{u}_t = \bar{\boldsymbol{\eta}}_\perp \mathbf{v}_{1t} + \bar{\boldsymbol{\eta}} \mathbf{v}_{2t} = \begin{pmatrix} \bar{\boldsymbol{\eta}}_\perp & \bar{\boldsymbol{\eta}} \end{pmatrix} \begin{pmatrix} \mathbf{v}_{1t} \\ \mathbf{v}_{2t} \end{pmatrix}$$

We have

$$\mathbf{C}(L)\mathbf{u}_t = [\mathbf{C}(L) \begin{pmatrix} \bar{\boldsymbol{\eta}}_\perp & \bar{\boldsymbol{\eta}} \end{pmatrix}] \begin{pmatrix} \mathbf{v}_{1t} \\ \mathbf{v}_{2t} \end{pmatrix} = (1-L)\mathbf{G}_1(L)\mathbf{v}_{1t} + (\boldsymbol{\xi} + (1-L)\mathbf{G}_2(L))\mathbf{v}_{2t}. \quad (25)$$

where $\mathbf{G}_1(L) = (\mathbf{D} + (1-L)\mathbf{E}(L))\bar{\boldsymbol{\eta}}_\perp$, and $\mathbf{G}_2(L) = (\mathbf{D} + (1-L)\mathbf{E}(L))\bar{\boldsymbol{\eta}}$. Using (25), we can write all the solutions of the difference equation $(1-L)\mathbf{F}_t = \mathbf{S}(L)^{-1}\mathbf{C}(L)\mathbf{u}_t$ as

$$\mathbf{F}_t = \mathbf{S}(L)^{-1} [\mathbf{G}_1(L)\mathbf{v}_{1t} + \mathbf{G}_2(L)\mathbf{v}_{2t} + \mathbf{T}_t] + \mathbf{W}, \quad (26)$$

where $\mathbf{W} \in L_2(\Omega, \mathcal{F}, P)$, and

$$\mathbf{T}_t = \begin{cases} \boldsymbol{\xi}(\mathbf{v}_{21} + \mathbf{v}_{22} + \cdots + \mathbf{v}_{2t}), & \text{for } t > 0 \\ \mathbf{0}, & \text{for } t = 0 \\ -\boldsymbol{\xi}(\mathbf{v}_{20} + \mathbf{v}_{2,-1} + \cdots + \mathbf{v}_{2,t+1}), & \text{for } t < 0. \end{cases}$$

As $\boldsymbol{\xi}$ is full rank, we see that \mathbf{F}_t is driven by the $q - d = r - c$ permanent shocks \mathbf{v}_{2t} , and by the d temporary shocks \mathbf{v}_{1t} . In representation (26), the component \mathbf{T}_t is the common-trend of Stock and Watson (1988). Note that the number of permanent shocks is obtained as r minus the cointegration rank, as usual. However, the number of transitory shocks is obtained as the complement of the number of permanent shocks to q , not to r , as though $r - q$ transitory shocks had a zero coefficient.

3.3 Error Correction representations

We now prove the Granger Representation Theorem for singular stochastic vectors \mathbf{F}_t belonging to a rational reduced-rank $I(1)$ family with cointegration rank c and parameters in the open set $\Pi \in \mathbb{R}^\lambda$. Our line of reasoning combines arguments used in the proof of Granger's Theorem in the non-singular case (see e.g. Johansen (1995), Theorem 4.5, p. 55-57) and the results in Proposition 1 on singular stochastic vectors.

From Definition 4, \mathbf{F}_t is a solution of the equation

$$\mathbf{S}(L)(1-L)\boldsymbol{\zeta}_t = \mathbf{C}(L)\mathbf{u}_t = (\boldsymbol{\xi}\boldsymbol{\eta}' + (1-L)\mathbf{D} + (1-L)^2\mathbf{E}(L))\mathbf{u}_t,$$

and has the representation $\mathbf{F}_t = \tilde{\mathbf{F}}_t + \mathbf{W}$, where $\tilde{\mathbf{F}}_t$ is defined in (21) and \mathbf{W} is an r -dimensional stochastic vector.

Generically, the matrix $\zeta = \begin{pmatrix} \xi'_\perp \\ \xi' \end{pmatrix}$ is $r \times r$ and invertible (see comment C1 on Definition 4). We have

$$\begin{aligned} (1-L)\zeta\mathbf{S}(L)\mathbf{F}_t &= \zeta\mathbf{C}(L)\mathbf{u}_t = \left\{ \begin{pmatrix} \mathbf{0}_{c \times q} \\ \xi'_\perp \xi \eta' \end{pmatrix} + (1-L) \begin{pmatrix} \xi'_\perp \mathbf{D} \\ \xi' \mathbf{D} \end{pmatrix} + (1-L)^2 \begin{pmatrix} \xi'_\perp \mathbf{E}(L) \\ \xi' \mathbf{E}(L) \end{pmatrix} \right\} \mathbf{u}_t \\ &= \begin{pmatrix} (1-L)\mathbf{I}_c & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{r-c} \end{pmatrix} \left\{ \begin{pmatrix} \xi'_\perp \mathbf{D} \\ \xi'_\perp \xi \eta' \end{pmatrix} + (1-L) \begin{pmatrix} \xi'_\perp \mathbf{E}(L) \\ \xi' \mathbf{D} \end{pmatrix} + (1-L)^2 \begin{pmatrix} \mathbf{0}_{c \times q} \\ \xi' \mathbf{E}(L) \end{pmatrix} \right\} \mathbf{u}_t. \end{aligned} \quad (27)$$

Taking the first c rows,

$$(1-L)\xi'_\perp \mathbf{S}(L)\mathbf{F}_t = (1-L) \left(\xi'_\perp \mathbf{D} + (1-L)\xi'_\perp \mathbf{E}(L) \right) \mathbf{u}_t. \quad (28)$$

In Appendix A.1 we prove that if \mathbf{F}_t is such that (28) holds and $\xi'_\perp \mathbf{S}(L)\mathbf{F}_t$ is weakly stationary with rational spectral density, then, in $\mathbf{F}_t = \tilde{\mathbf{F}}_t + \mathbf{W}$, \mathbf{W} must be chosen such that

$$\xi'_\perp \mathbf{S}(L)\mathbf{F}_t = \mathbf{k} + \left(\xi'_\perp \mathbf{D} + (1-L)\xi'_\perp \mathbf{E}(L) \right) \mathbf{u}_t, \quad (29)$$

where \mathbf{k} is a c -dimensional constant vector and $\xi'_\perp \mathbf{S}(1)\mathbf{W} = \mathbf{k}$. Now,

$$\xi'_\perp \mathbf{S}(1)\mathbf{F}_t = \xi'_\perp \mathbf{S}(L)\mathbf{F}_t - \xi'_\perp \mathbf{S}^*(L)(1-L)\mathbf{F}_t, \quad (30)$$

where $\mathbf{S}^*(L)$ is the polynomial $(\mathbf{S}(L) - \mathbf{S}(1))/(1-L)$. Obviously $\xi'_\perp \mathbf{S}^*(L)(1-L)\mathbf{F}_t$ is weakly stationary with rational spectral density. As a consequence, $\xi'_\perp \mathbf{S}(L)\mathbf{F}_t$ is weakly stationary with rational spectral density if and only if $\xi'_\perp \mathbf{S}(1)\mathbf{F}_t$ is weakly stationary with rational spectral density. Moreover, equation (30), replacing $\xi'_\perp \mathbf{S}(L)\mathbf{F}_t$ with the left-hand side of (29) and $(1-L)\mathbf{F}_t$ with $\mathbf{S}(L)^{-1}(\xi \eta' + (1-L)\mathbf{D} + (1-L)^2\mathbf{E}(L))\mathbf{u}_t$, becomes

$$\xi'_\perp \mathbf{S}(1)\mathbf{F}_t = \mathbf{k} + \left\{ \left(\xi'_\perp \mathbf{D} - \xi'_\perp \mathbf{S}^*(1)\mathbf{S}(1)^{-1}\xi \eta' \right) + (1-L)\mathcal{M}(L) \right\} \mathbf{u}_t.$$

As $\xi'_\perp \mathbf{D} - \xi'_\perp \mathbf{S}^*(1)\mathbf{S}(1)^{-1}\xi \eta' \neq \mathbf{0}$ generically, $\xi'_\perp \mathbf{S}(1)\mathbf{F}_t$ is generically $I(0)$.

In conclusion, assuming that \mathbf{F}_t is such that $\xi'_\perp \mathbf{S}(L)\mathbf{F}_t$ is weakly stationary with rational spectral density and mean \mathbf{k} :

$$\begin{aligned} \begin{pmatrix} \mathbf{I}_c & \mathbf{0} \\ \mathbf{0} & (1-L)\mathbf{I}_{r-c} \end{pmatrix} \zeta\mathbf{S}(L)\mathbf{F}_t &= \\ &= \begin{pmatrix} \mathbf{k} \\ \mathbf{0}_{(r-c) \times 1} \end{pmatrix} + \left\{ \begin{pmatrix} \xi'_\perp \mathbf{D} \\ \xi'_\perp \xi \eta' \end{pmatrix} + (1-L) \begin{pmatrix} \xi'_\perp \mathbf{E}(L) \\ \xi' \mathbf{D} \end{pmatrix} + (1-L)^2 \begin{pmatrix} \mathbf{0}_{c \times q} \\ \xi' \mathbf{E}(L) \end{pmatrix} \right\} \mathbf{u}_t. \end{aligned}$$

Denote by $\mathbf{M}(L)$ the matrix between curly brackets. The following statement is proved in Appendix A.2.

Proposition 2. *Assume that the family of $I(1)$ processes \mathbf{F}_t , such that $(1-L)\mathbf{F}_t = \mathbf{S}(L)^{-1}\mathbf{C}(L)\mathbf{u}_t$, is a rational reduced-rank $I(1)$ family with cointegration rank c and parameter set Π . Then, for generic values of the parameters in Π , the $r \times q$ matrix $\mathbf{M}(z)$ is zeroless. In particular, generically,*

$$\text{rank}(\mathbf{M}(1)) = \text{rank} \left(\begin{pmatrix} \xi'_\perp \mathbf{D} \\ \xi'_\perp \xi \eta' \end{pmatrix} \right) = q. \quad (31)$$

For $r = q$, (31) is equivalent to the condition that $\xi'_\perp \mathbf{C}^* \eta_\perp$ has full rank in Johansen (1995), Theorem 5.4, p. 55 (Johansen's matrix \mathbf{C}^* is equal to our \mathbf{D}). To see this, observe that if $r = q$ then $(\eta_\perp \quad \eta)$ is $r \times r$ and invertible, and that

$$\mathbf{M}(1) = (\eta_\perp \quad \eta)^{-1} \begin{pmatrix} \xi'_\perp \mathbf{D} \eta_\perp & \xi'_\perp \mathbf{D} \eta \\ \mathbf{0}_{r-c \times c} & \xi' \xi \eta' \eta \end{pmatrix}.$$

As $\xi' \xi \eta' \eta$ is non singular, the determinant of $\mathbf{M}(1)$ vanishes if and only if the determinant of $\xi'_\perp \mathbf{D} \eta_\perp$ vanishes.

A consequence of Proposition 2 and Proposition 1(I) is that generically there exists a finite-degree $r \times r$ polynomial matrix

$$\mathbf{N}(L) = \mathbf{I}_r + \mathbf{N}_1 L + \cdots + \mathbf{N}_p L^p,$$

for some p , such that: (i) $\mathbf{N}(L)\mathbf{M}(L) = \mathbf{M}(0)$, i.e. $\mathbf{N}(L)$ is a left inverse of $\mathbf{M}(L)$; (ii) all the roots of $\det(\mathbf{N}(L))$ lie outside the unit circle, so that $\mathbf{N}(1)$ has full rank.

In conclusion, for generic values of the parameters in Π ,

$$\mathbf{A}(L)\mathbf{F}_t = \mathbf{h} + \mathbf{C}(0)\mathbf{u}_t,$$

where

$$\begin{aligned} \mathbf{A}(L) &= \mathbf{I}_r + \mathbf{A}_1 L + \cdots + \mathbf{A}_P L^P = \zeta^{-1} \mathbf{N}(L) \begin{pmatrix} \mathbf{I}_c & \mathbf{0} \\ \mathbf{0} & (1-L)\mathbf{I}_{r-c} \end{pmatrix} \zeta \mathbf{S}(L) \\ &= \zeta^{-1} \mathbf{N}(L) \begin{pmatrix} \xi'_\perp \\ (1-L)\xi' \end{pmatrix} \mathbf{S}(L), \end{aligned} \quad (32)$$

with $P = p + 1 + s_2$, and

$$\mathbf{h} = \mathbf{A}(1) \begin{pmatrix} \mathbf{k} \\ \mathbf{0}_{(r-c) \times 1} \end{pmatrix}. \quad (33)$$

Defining

$$\boldsymbol{\alpha} = \zeta^{-1} \mathbf{N}(1) \begin{pmatrix} \mathbf{I}_c \\ \mathbf{0}_{(r-c) \times c} \end{pmatrix}, \quad \boldsymbol{\beta} = \mathbf{S}(1)' \xi'_\perp, \quad (34)$$

both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ have generically rank c (regarding $\boldsymbol{\alpha}$, remember that $\mathbf{N}(1)$ has full rank) and $\mathbf{A}(1) = \boldsymbol{\alpha} \boldsymbol{\beta}'$. Lastly, define

$$\mathbf{A}^*(L) = (1-L)^{-1} (\mathbf{A}(L) - \mathbf{A}(1)L). \quad (35)$$

We have proved the following statement.

Proposition 3. (Granger Representation Theorem for reduced-rank $I(1)$ vectors)
Let \mathbf{F}_t be a rational reduced-rank $I(1)$ family with cointegration rank c and parameter set Π , so that

$$(1-L)\mathbf{F}_t = \mathbf{S}(L)^{-1} \mathbf{C}(L)\mathbf{u}_t.$$

For generic values of the parameters in Π , (i) \mathbf{F}_t can be determined such that $\boldsymbol{\beta}' \mathbf{F}_t = \xi'_\perp \mathbf{S}(1) \mathbf{F}_t$ is weakly stationary with rational spectral density and mean \mathbf{k} , (ii) \mathbf{F}_t has the Error Correction representation

$$\mathbf{A}(L)\mathbf{F}_t = \mathbf{A}^*(L)(1-L)\mathbf{F}_t + \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{F}_{t-1} = \mathbf{h} + \mathbf{C}(0)\mathbf{u}_t. \quad (36)$$

where the $r \times r$ finite-degree polynomial matrices $\mathbf{A}(L)$ and $\mathbf{A}^*(L)$, the full-rank $r \times c$ matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, the r -dimensional constant vector \mathbf{h} , have been defined above in (32), (35), (34), (33), respectively. Generically, $\boldsymbol{\beta}' \mathbf{F}_t$ is $I(0)$.

In Definition 4 we have not assumed that \mathbf{u}_t is fundamental for $(1-L)\mathbf{F}_t$. However,

$$\mathbf{C}(L) = \zeta^{-1} \begin{pmatrix} (1-L)\mathbf{I}_c & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{r-c} \end{pmatrix} \mathbf{M}(L). \quad (37)$$

Therefore, by Proposition 2, generically the matrix $\mathbf{C}(L)$ has full rank q for $|z| < 1$. Thus, see Rozanov (1967), pp. 43–7:

Proposition 4. *Assume that the family of $I(1)$ processes \mathbf{F}_t , such that $(1-L)\mathbf{F}_t = \mathbf{S}(L)^{-1}\mathbf{C}(L)\mathbf{u}_t$, is a rational reduced-rank $I(1)$ family with cointegration rank c and parameter set Π . For generic values of the parameters the vector \mathbf{u}_t is fundamental for the vector $(1-L)\mathbf{F}_t$.*

As recalled in Section 2.2, if $(1-L)\mathbf{F}_t = \mathbf{S}(L)^{-1}\mathbf{C}(L)\mathbf{u}_t$, then \mathbf{u}_t is fundamental for $(1-L)\mathbf{F}_t$ if it belongs to the space spanned by $(1-L)\mathbf{F}_\tau$, $\tau \leq t$. This is not inconsistent with the fact that when $c > r - q$, so that $\mathbf{C}(z)$ is not full rank for $z = 1$, there are not autoregressive representations for $(1-L)\mathbf{F}_t$, either finite or infinite, and therefore no representations of the form

$$\mathbf{u}_t = (\mathcal{A}_0 + \mathcal{A}_1L + \mathcal{A}_2L^2 + \dots)(1-L)\mathbf{F}_t = \lim_{n \rightarrow \infty} \sum_{k=0}^n \mathcal{A}_k(1-L)\mathbf{F}_{t-k},$$

where the matrices \mathcal{A}_j are $q \times r$. Indeed, fundamentalness of \mathbf{u}_t only implies that \mathbf{u}_t can be obtained as the limit of linear combinations of the vectors $(1-L)\mathbf{F}_\tau$, $\tau \leq 0$, i.e.

$$\mathbf{u}_t = \lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} \mathcal{B}_k^{(n)}(1-L)\mathbf{F}_{t-k},$$

where the coefficient matrices depend both on k and n .⁴

Lastly, let us recall that neither \mathbf{F}_t nor \mathbf{u}_t are identified in the factor model

$$\mathbf{x}_t = \mathbf{\Lambda}\mathbf{F}_t + \boldsymbol{\epsilon}_t, \quad (1-L)\mathbf{F}_t = \mathbf{S}(L)^{-1}\mathbf{C}(L)\mathbf{u}_t.$$

In particular, if $\mathbf{F}_t = \mathbf{H}\mathbf{F}_t^*$, where \mathbf{H} is $r \times r$ and invertible,

$$\mathbf{x}_t = \mathbf{\Lambda}^*\mathbf{F}_t^* + \boldsymbol{\epsilon}_t, \quad (1-L)\mathbf{F}_t^* = \mathbf{S}^*(L)^{-1}\mathbf{C}^*(L)\mathbf{u}_t,$$

where $\mathbf{\Lambda}^* = \mathbf{\Lambda}\mathbf{H}$, $\mathbf{S}^*(L) = \mathbf{H}^{-1}\mathbf{S}(L)\mathbf{H}$, $\mathbf{C}^*(L) = \mathbf{H}^{-1}\mathbf{C}(L)$ (see also Section 2.3). In particular, if $\mathbf{H}^{-1} = \zeta\mathbf{S}(1)$, the first c coordinates of \mathbf{F}_t^* are $I(0)$ and the remaining $r - c = q - d$ are $I(1)$.⁵ Moreover, the c coordinates of the error vector $\boldsymbol{\beta}^{*'}\mathbf{F}_{t-1}^*$ in representation (36) for \mathbf{F}_t^* are linear combinations of the $I(0)$ factors alone.

3.4 VECMs and unrestricted VARs in the levels

Several papers have addressed the issue whether and when an Error Correction model or an unrestricted VAR in the levels should be used for estimation in the case of non-singular cointegrated vectors: Sims et al. (1990) have shown that the parameters of a cointegrated VAR

⁴For example, if $x_t = (1-L)u_t$, where u_t is a univariate white noise, then u_t is fundamental for x_t although no autoregressive representation $x_t + a_1x_{t-1} + a_2x_{t-2} + \dots = u_t$ exists. See Brockwell and Davis (1991), p. 111, Problem 3.8.

⁵More precisely, the first c equations in $(1-L)\mathbf{F}_t^* = \mathbf{S}^*(L)^{-1}(L)\mathbf{C}^*(L)\mathbf{u}_t$ have an $I(0)$ solution (the argument goes as in the discussion of equation (27)). This is consistent with Definition 3(ii), some of the coordinates of a $I(1)$ vector can be $I(0)$.

are consistently estimated using an unrestricted VAR in the levels; on the other hand, Phillips (1998) shows that if the variables are cointegrated, the long-run features of the impulse-response functions are consistently estimated only if the unit roots are explicitly taken into account, that is within a VECM specification. The simulation exercise described below provides some evidence in favour of the VECM specification in the singular case.

We generate \mathbf{F}_t using a specification of (36) with $r = 4$, $q = 3$, $d = 2$, so that $c = r - q + d = 3$. The 4×4 matrix $\mathbf{A}(L)$ is of degree 2. Moreover, the upper 3×3 submatrix of $\mathbf{C}(0)$ is lower triangular (see Appendix C for details). We estimate a VECM as in Johansen (1988, 1991) and assuming c , the degree of $\mathbf{A}(L)$ and the identification restrictions known. We replicate the generation 1000 times for $T = 100, 500, 1000, 5000$. For each replication, we estimate a (misspecified) VAR in differences, a VAR in the levels and a VECM, assuming known c and the degree of $\mathbf{A}(L)$ and $\mathbf{A}^*(L)$. The Root Mean Square Error between estimated and actual impulse-response functions is computed for each replication using all 12 responses and averaged over all replications. The results are shown in Table 1. We see that the RMSE of both the VECM and the LVAR decreases as T increases. However, for all values of T , the RMSE of the VECM stabilizes as the lag increases, whereas it deteriorates for the LVAR, in line with the claim that the long-run response of the variables are better estimated with the VECM.

For the sake of simplicity, in the simulation exercise the “structural shocks” are determined by imposing restrictions on the response of \mathbf{F}_t to the shocks. Precisely, the upper 3×3 submatrix of $\mathbf{C}(0)$ is lower triangular. However, as we recalled in the Introduction and Section 2.3, neither the factors \mathbf{F}_t nor their response to the shocks, have a direct economic interpretation. In empirical work with actual data, identification of the structural shocks \mathbf{v}_t , which result as a linear transformation of \mathbf{u}_t , is usually obtained by imposing restrictions on the impulse-response functions of the variables x_{it} with respect to \mathbf{v}_t .

Table 1: MONTE CARLO SIMULATIONS. VECM: $r = 4, q = 3, c = 3$.

	lags	DVAR	LVAR	VECM		lags	DVAR	LVAR	VECM
$T = 100$	0	0.06	0.05	0.05	$T = 500$	0	0.02	0.02	0.02
	4	0.26	0.18	0.17		4	0.23	0.07	0.07
	20	0.30	0.37	0.22		20	0.25	0.14	0.09
	40	0.30	0.45	0.22		40	0.25	0.21	0.09
	80	0.30	0.57	0.22		80	0.25	0.32	0.09
$T = 1000$	0	0.02	0.02	0.02	$T = 5000$	0	0.01	0.01	0.01
	4	0.23	0.05	0.05		4	0.22	0.02	0.02
	20	0.25	0.09	0.07		20	0.25	0.03	0.03
	40	0.25	0.13	0.07		40	0.25	0.04	0.03
	80	0.25	0.22	0.07		80	0.25	0.06	0.03

Root Mean Squared Errors at different lags, when estimating the impulse response functions of the simulated variables \mathbf{F}_t to the common shocks \mathbf{u}_t . Estimation is carried out using three different autoregressive representations: a VAR for $(1-L)\mathbf{F}_t$ (DVAR), a VAR for \mathbf{F}_t (LVAR), and a VECM with $c = r - q + d$ error correction terms (VECM). Results are based on 1000 replications. For the data generating process see Appendix C. The RMSEs are obtained averaging over all replications and all 4×3 impulse responses.

4 Cointegration of the variables x_{it}

The relationship between cointegration of the factors \mathbf{F}_t and cointegration of the variables x_{it} is now considered. Let us firstly observe that, regarding model (6), neither the assumptions (1) through (5) listed in Section 2, nor the asymptotic conditions (see e.g. Forni et al. (2009)) say much on the matrix $\mathbf{\Lambda}$ and the vector $\boldsymbol{\epsilon}_t$ for a given finite n . In particular, the first r eigenvalues of the matrix $\mathbf{\Lambda}\mathbf{\Lambda}'$ must diverge as $n \rightarrow \infty$, but this has no implications on the rank of the matrix $\mathbf{\Lambda}$ corresponding to, say, $n = 10$. Moreover, as we see in Proposition 5(iii), if the idiosyncratic components are $I(0)$, then all p -dimensional subvectors of \mathbf{x}_t are cointegrated for $p > q - d$, which is at odds with what is observed in the macroeconomic datasets analysed in the empirical Dynamic Factor Model literature. This motivates assuming that $\boldsymbol{\epsilon}_t$ is $I(1)$. In that case, see Proposition 5(i), cointegration of \mathbf{x}_t requires that both the common and the idiosyncratic components are cointegrated. Some results are collected in the statement below.

Proposition 5. *Let $\mathbf{x}_t^{(p)} = \boldsymbol{\chi}_t^{(p)} + \boldsymbol{\epsilon}_t^{(p)} = \mathbf{\Lambda}^{(p)}\mathbf{F}_t + \boldsymbol{\epsilon}_t^{(p)}$ be a p -dimensional subvector of \mathbf{x}_t , $p \leq n$. Denote by c_χ^p and c_ϵ^p the cointegration rank of $\boldsymbol{\chi}_t^{(p)}$ and $\boldsymbol{\epsilon}_t^{(p)}$ respectively. Both range from p , stationarity, to 0, no cointegration.*

(i) $\mathbf{x}_t^{(p)}$ is cointegrated only if $\boldsymbol{\chi}_t^{(p)}$ and $\boldsymbol{\epsilon}_t^{(p)}$ are both cointegrated.

(ii) If $p > q - d$ then $\boldsymbol{\chi}_t^{(p)}$ is cointegrated. If $p \leq q - d$ and $\mathbf{\Lambda}^{(p)}$ is full rank then $\boldsymbol{\chi}_t^{(p)}$ is not cointegrated. If $p \leq q - d$ and $\text{rank}(\mathbf{\Lambda}^{(p)}) < p$ then $\boldsymbol{\chi}_t^{(p)}$ is cointegrated.

(iii) Let $V^\chi \subseteq \mathbb{R}^p$ and $V^\epsilon \subseteq \mathbb{R}^p$ be the cointegration spaces of $\boldsymbol{\chi}_t^{(p)}$ and $\boldsymbol{\epsilon}_t^{(p)}$ respectively. The vector $\mathbf{x}_t^{(p)}$ is cointegrated if and only if the intersection of V^χ and V^ϵ contains non-zero vectors. In particular, if $p > q - d$ and $c^\epsilon > q - d$ then $\mathbf{x}_t^{(p)}$ is cointegrated.

Proof. Because χ_{it} and ϵ_{js} are orthogonal for all i, j, t, s , see Assumption (5) for model (6), the spectral densities of $(1 - L)\mathbf{x}_t^{(p)}$, $(1 - L)\boldsymbol{\chi}_t^{(p)}$, $(1 - L)\boldsymbol{\epsilon}_t^{(p)}$ fulfill:

$$\boldsymbol{\Sigma}_{\Delta x}^{(p)}(\theta) = \boldsymbol{\Sigma}_{\Delta \chi}^{(p)}(\theta) + \boldsymbol{\Sigma}_{\Delta \epsilon}^{(p)}(\theta) \quad \theta \in [-\pi, \pi]. \quad (38)$$

Now, (38) implies that

$$\lambda_p \left(\boldsymbol{\Sigma}_{\Delta x}^{(p)}(0) \right) \geq \lambda_p \left(\boldsymbol{\Sigma}_{\Delta \chi}^{(p)}(0) \right) + \lambda^{(p)} \left(\boldsymbol{\Sigma}_{\Delta \epsilon}^{(p)}(0) \right), \quad (39)$$

where $\lambda_p(A)$ denotes the smallest eigenvalue of the hermitian matrix A ; this is one of the Weyl's inequalities, see Franklin (2000), p. 157, Theorem 1. Because spectral density matrices are non-negative definite, the right hand side in (39) vanishes if and only if both terms on the right hand side vanish, i.e. the spectral density of $\Delta \mathbf{x}_t^{(p)}$ is singular at zero if and only if the spectral densities of $\Delta \boldsymbol{\chi}_t^{(p)}$ and $\Delta \boldsymbol{\epsilon}_t^{(p)}$ are singular at zero. By definition 3, (i) is proved.

Without loss of generality we can assume that $\mathbf{S}(L) = \mathbf{I}_r$. By substituting (26) in (6), we obtain

$$\mathbf{x}_t = \mathbf{\Lambda} [(\mathbf{G}_1(L)\mathbf{v}_{1t} + \mathbf{G}_2(L)\mathbf{v}_{2t} + \mathbf{T}_t) + \mathbf{W}] + \boldsymbol{\epsilon}_t, \quad (40)$$

where on the right hand side the only non-stationary term is \mathbf{T}_t and (possibly) $\boldsymbol{\epsilon}_t$. By recalling that $\mathbf{T}_t = \boldsymbol{\xi} \sum_{s=1}^t \mathbf{v}_{2s}$ where $\boldsymbol{\xi}$ is of dimension $r \times (q - d)$ and rank $q - d$, and by defining $\mathcal{G}_t = \mathbf{\Lambda}[\mathbf{G}_1(L)\mathbf{v}_{1t} + \mathbf{G}_2(L)\mathbf{v}_{2t} + \mathbf{Z}]$ and $\mathcal{T}_t = \sum_{s=1}^t \mathbf{v}_{2s}$, we can rewrite (40) as:

$$\mathbf{x}_t = \mathbf{\Lambda}\boldsymbol{\xi}\mathcal{T}_t + \mathcal{G}_t + \boldsymbol{\epsilon}_t.$$

For $\mathbf{x}_t^{(p)}$:

$$\mathbf{x}_t^{(p)} = \boldsymbol{\chi}_t^{(p)} + \boldsymbol{\epsilon}_t^{(p)} = \boldsymbol{\Lambda}^{(p)} \boldsymbol{\xi} \mathcal{T}_t + \boldsymbol{\mathcal{G}}_t^{(p)} + \boldsymbol{\epsilon}_t^{(p)},$$

where $\boldsymbol{\Lambda}^{(p)}$ and $\boldsymbol{\mathcal{G}}_t^{(p)}$ have an obvious definition. Of course cointegration of the common components $\boldsymbol{\chi}_t^{(p)}$ is equivalent to cointegration of $\boldsymbol{\Lambda}^{(p)} \boldsymbol{\xi} \mathcal{T}_t$, which in turn is equivalent to $\text{rank}(\boldsymbol{\Lambda}^{(p)} \boldsymbol{\xi}) < p$. Statement (ii) follows from

$$\text{rank}(\boldsymbol{\Lambda}^{(p)} \boldsymbol{\xi}) \leq \min(\text{rank}(\boldsymbol{\Lambda}^{(p)}), \text{rank}(\boldsymbol{\xi})).$$

The first part of (iii) is obvious. Assume now that $p > q - d$. If $c_\chi^p + c_\epsilon^p = \dim(V^\chi) + \dim(V^\epsilon) = p - (q - d) + c_\epsilon^p > p$, i.e. if $c_\epsilon^p > q - d$, then the intersection between V^χ and V^ϵ is non-trivial, so that $\mathbf{x}_t^{(p)}$ is cointegrated. \square

5 Summary and conclusions

The paper studies representation theory for Dynamic Factor Models when the factors are $I(1)$ and singular, and the idiosyncratic components are $I(1)$. Singular $I(1)$ vectors are cointegrated, with cointegration rank c equal to $r - q$, the dimension of \mathbf{F}_t minus its rank, plus d , with $0 \leq d < q$. We prove that if $(1 - L)\mathbf{F}_t$ has rational spectral density, then generically F_t has an Error Correction representation with c error terms and a finite autoregressive matrix polynomial. Moreover, \mathbf{F}_t is driven by $r - c$ permanent shocks and d transitory shocks, with $r - c + d = q$, not r as in the non-singular case. These results are obtained by combining the standard results on cointegration with recent results on singular stochastic vectors.

Using simulated data generated by a simple singular VECM, confirms previous results, obtained for non-singular vectors, showing that under cointegration the long-run features of impulse-response functions are better estimated using a VECM rather than a VAR in the levels.

In Section 4 we argue that stationarity of the idiosyncratic components would produce an amount of cointegration for the observable variables x_{it} that is not observed in the datasets that are standard in Dynamic Factor Model literature, see e.g. Stock and Watson (2002a,b, 2005), Forni et al. (2009). Thus the idiosyncratic vector in those datasets is likely to be $I(1)$.

The results in this paper are the basis for estimation of $I(1)$ Dynamic Factor Models with cointegrated factors, see the companion paper (Barigozzi et al., 2016).

References

- Amengual, D. and M. W. Watson (2007). Consistent estimation of the number of dynamic factors in a large N and T panel. *Journal of Business and Economic Statistics* 25, 91–96.
- Anderson, B. D. and M. Deistler (2008a). Generalized linear dynamic factor models—a structure theory. *IEEE Conference on Decision and Control*.
- Anderson, B. D. and M. Deistler (2008b). Properties of zero-free transfer function matrices. *SICE Journal of Control, Measurement and System Integration* 1, 284–292.
- Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data. *Journal of Econometrics* 122, 137–183.

- Bai, J. and S. Ng (2004). A PANIC attack on unit roots and cointegration. *Econometrica* 72, 1127–1177.
- Bai, J. and S. Ng (2007). Determining the number of primitive shocks in factor models. *Journal of Business and Economic Statistics* 25, 52–60.
- Banerjee, A., M. Marcellino, and I. Masten (2014a). Forecasting with factor-augmented error correction models. *International Journal of Forecasting* 30, 589–612.
- Banerjee, A., M. Marcellino, and I. Masten (2014b). Structural FECM: Cointegration in large-scale structural FAVAR models. *Working Paper*, <http://www.econlab.si/igor.masten>.
- Barigozzi, M., A. M. Conti, and M. Luciani (2014). Do euro area countries respond asymmetrically to the common monetary policy? *Oxford Bulletin of Economics and Statistics* 76, 693–714.
- Barigozzi, M., M. Lippi, and M. Luciani (2016). Non-stationary dynamic factor models for large datasets. <http://arxiv.org/abs/1602.02398>.
- Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods* (Second ed.). New York: Springer.
- Deistler, M., B. D. Anderson, A. Filler, C. Zinner, and W. Chen (2010). Generalized linear dynamic factor models: An approach *via* singular autoregressions. *European Journal of Control*, 211–224.
- Engle, R. F. and C. W. J. Granger (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica* 55, 251–76.
- Forni, M. and L. Gambetti (2010). The dynamic effects of monetary policy: A structural factor model approach. *Journal of Monetary Economics* 57, 203–216.
- Forni, M., D. Giannone, M. Lippi, and L. Reichlin (2009). Opening the Black Box: Structural Factor Models versus Structural VARs. *Econometric Theory* 25, 1319–1347.
- Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The Generalized Dynamic Factor Model: Identification and Estimation. *The Review of Economics and Statistics* 82, 540–554.
- Forni, M., M. Hallin, M. Lippi, and P. Zaffaroni (2015). Dynamic factor models with infinite-dimensional factor spaces: one-sided representations. *Journal of Econometrics* 185, 359–371.
- Forni, M. and M. Lippi (2001). The Generalized Dynamic Factor Model: Representation Theory. *Econometric Theory* 17, 1113–1141.
- Forni, M. and M. Lippi (2010). The Unrestricted Dynamic Factor Model: One-sided Representation Results. *Journal of Econometrics* 163, 23–28.
- Franklin, J. N. (2000). *Matrix Theory* (Second ed.). New York: Dover Publications.
- Giannone, D., L. Reichlin, and L. Sala (2005). Monetary policy in real time. In M. Gertler and K. Rogoff (Eds.), *NBER Macroeconomics Annual 2004*. MIT Press.

- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12, 231–254.
- Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica* 59, 1551–80.
- Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models* (First ed.). Oxford: Oxford University Press.
- Lancaster, P. and M. Tismenetsky (1985). *The theory of matrices* (Second ed.). New York: Academic Press.
- Luciani, M. (2015). Monetary policy and the housing market: A structural factor analysis. *Journal of Applied Econometrics* 30, 199–218.
- Peña, D. and P. Poncela (2004). Nonstationary dynamic factor analysis. *Journal of Statistical Planning and Inference* 136, 1237–1257.
- Phillips, P. C. (1998). Impulse response and forecast error variance asymptotics in nonstationary VARs. *Journal of Econometrics* 83, 21–56.
- Rozanov, Y. A. (1967). *Stationary Random Processes*. San Francisco: Holden-Day.
- Sims, C., J. H. Stock, and M. W. Watson (1990). Inference in linear time series models with some unit roots. *Econometrica* 58, 113–144.
- Stock, J. H. and M. W. Watson (1988). Testing for common trends. *Journal of the American Statistical Association* 83, 1097–1107.
- Stock, J. H. and M. W. Watson (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97, 1167–1179.
- Stock, J. H. and M. W. Watson (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics* 20, 147–162.
- Stock, J. H. and M. W. Watson (2005). Implications of dynamic factor models for VAR analysis. Working Paper 11467, NBER.
- van der Waerden, B. L. (1953). *Modern Algebra* (Second ed.), Volume I. New York: Frederick Ungar.
- Watson, M. W. (1994). Vector autoregressions and cointegration. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics, Volume IV*. Elsevier Science.

Appendix A Proofs

Appendix A.1 Stationary solutions of $(1 - L)\mathbf{y}_t = (1 - L)\boldsymbol{\zeta}_t$

Consider firstly the difference equation in the unknown g -dimensional vector process $\boldsymbol{\zeta}_t$:

$$(1 - L)\boldsymbol{\zeta}_t = \mathbf{w}_t, \quad t \in \mathbb{Z}, \quad (\text{A1})$$

where \mathbf{w}_t is a g -dimensional stochastic process with $w_{jt} \in L_2(\Omega, \mathcal{F}, P)$. Define a solution of (A1) as a process \mathbf{y}_t , $t \in \mathbb{Z}$, $Y_{jt} \in L_2(\Omega, \mathcal{F}, P)$, such that $(1 - L)\mathbf{Y}_t = \mathbf{w}_t$. If $\tilde{\mathbf{y}}_t$ is a solution, then all the solutions of (A1) are $\tilde{\mathbf{y}}_t + \mathbf{W}$, where \mathbf{W} is a g -dimensional stochastic variable with $W_{jt} \in L_2(\Omega, \mathcal{F}, P)$, i.e. a particular solution plus a constant stochastic process (a solution of the homogeneous equation $(1 - L)\mathbf{y}_t = \mathbf{0}$).

Assume that \mathbf{z}_t is a g -dimensional, weakly stationary process with a moving-average representation $\mathbf{z}_t = \mathbf{Z}(L)\mathbf{v}_t$, where:

- (i) \mathbf{v}_t is an s -dimensional non-singular white noise and $s \leq g$ belonging to $L_2(\Omega, \mathcal{F}, P)$,
- (ii) $\mathbf{Z}(L)$ is a $g \times s$ square-summable matrix such that z_{kt} and v_{jt} , for $t \in \mathbb{Z}$, $k = 1, 2, \dots, g$, $j = 1, 2, \dots, s$, span the same subspace of $L_2(\Omega, \mathcal{F}, P)$.⁶

Now consider the equation

$$(1 - L)\boldsymbol{\zeta}_t = (1 - L)\mathbf{z}_t. \quad (\text{A2})$$

Because \mathbf{z}_t trivially fulfills (A2), all the solutions of (A2) are

$$\mathbf{y}_t = \mathbf{z}_t + \mathbf{K}$$

where \mathbf{K} is a g -dimensional stochastic vector belonging to $L_2(\Omega, \mathcal{F}, P)$. We want to determine the conditions such that the solution \mathbf{y}_t is weakly stationary with a spectral density.

Let

$$\mathbf{K} = \sum_{k=-\infty}^{\infty} \mathbf{P}_k \mathbf{v}_k + \mathbf{H}.$$

be the orthogonal projection of \mathbf{K} on the space spanned by \mathbf{v}_k , $k \in \mathbb{Z}$. Setting $\mathbf{V} = \sum_{k=-\infty}^{\infty} \mathbf{P}_k \mathbf{v}_k$, because \mathbf{H} is orthogonal to \mathbf{V} and \mathbf{z}_τ , $\tau \in \mathbb{Z}$,

$$\mathbf{E}(\mathbf{y}_t \mathbf{y}'_{t-k}) = \mathbf{E}(\mathbf{z}_t \mathbf{z}'_{t-k}) + \mathbf{E}(\mathbf{V} \mathbf{V}') + \mathbf{E}(\mathbf{H} \mathbf{H}') + \mathbf{E}(\mathbf{z}_t \mathbf{V}') + \mathbf{E}(\mathbf{V} \mathbf{z}'_{t-k}).$$

Given k , the last two terms tend to zero when t tends to infinity (by the same argument used to prove that the autocovariances of a moving average tend to zero as the lag tends to infinity). Weak stationarity of \mathbf{y}_t implies that

$$\mathbf{E}(\mathbf{z}_t \mathbf{V}') + \mathbf{E}(\mathbf{V} \mathbf{z}'_{t-k}) = 0,$$

for all $t \in \mathbb{Z}$. On the other hand, given t , $\mathbf{E}(\mathbf{V} \mathbf{z}'_{t-k})$ tends to zero as k tends to infinity (again the argument on autocovariances), so that $\mathbf{E}(\mathbf{z}_t \mathbf{V}') = 0$ for all t . Orthogonality of \mathbf{V} to all \mathbf{z}_t implies orthogonality to all \mathbf{v}_t , see assumption (ii) above. As \mathbf{V} is an average of \mathbf{v}_k , $k \in \mathbb{Z}$, then $\mathbf{V} = \mathbf{0}$. In conclusion, all the stationary solutions of (A2) are $\mathbf{y}_t = \mathbf{z}_t + \mathbf{K}$ with \mathbf{K} orthogonal to \mathbf{z}_t for all $t \in \mathbb{Z}$.

Lastly, the spectral measure of \mathbf{y}_t has a jump at frequency zero unless the variance-covariance matrix of \mathbf{K} is zero. Thus \mathbf{y}_t has a spectral density if and only if \mathbf{K} is a constant

⁶A very weak condition. However, if $\mathbf{Z}(L)$ is a band-pass filter, (ii) does not hold.

vector (i.e. $\mathbf{K}(\omega) = \mathbf{k}$ almost surely in Ω). In that case the spectral densities of \mathbf{y}_t and \mathbf{z}_t coincide.

Using the results above we can prove the statement in comment C7 on Definition 3, Cointegration. If \mathbf{y}_t is such that $(1 - L)\mathbf{y}_t = \mathbf{V}(L)\mathbf{v}_t$, then $\mathbf{y}_t = \tilde{\mathbf{y}}_t + \mathbf{W}$, where

$$\tilde{\mathbf{y}}_t = \mathbf{V}(L)\boldsymbol{\mu}_t \quad \text{where} \quad \boldsymbol{\mu}_t = \begin{cases} \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_t, & \text{for } t > 0 \\ \mathbf{0}, & \text{for } t = 0 \\ -(\mathbf{v}_0 + \mathbf{v}_{-1} + \cdots + \mathbf{v}_{t+1}), & \text{for } t < 0 \end{cases}$$

and \mathbf{W} is an r -dimensional stochastic vector. If \mathbf{c} is a cointegration vector, then

$$(1 - L)\mathbf{c}'\mathbf{y}_t = \mathbf{c}'\mathbf{V}(L)\mathbf{v}_t = \mathbf{c}' \left[\mathbf{V}(1) + (1 - L) \frac{\mathbf{V}(L) - \mathbf{V}(1)}{1 - L} \right] \mathbf{v}_t = (1 - L)\mathbf{c}'\mathbf{V}^*(L)\mathbf{v}_t,$$

where trivially the entries of $\mathbf{V}^*(L) = (\mathbf{V}(L) - \mathbf{V}(1))/(1 - L)$ are rational functions of L with no poles of modulus less or equal to 1. From the last equation we obtain

$$\mathbf{c}'\mathbf{y}_t = \mathbf{c}'\tilde{\mathbf{y}}_t + \mathbf{c}'\mathbf{W} = \mathbf{c}'\mathbf{V}^*(L)\mathbf{v}_t + w,$$

where w is a stochastic variable. The process $\mathbf{c}'\mathbf{y}_t$ is weakly stationary with rational spectral density if and only if w is a constant with probability one. On the other hand,

$$\mathbf{c}'\tilde{\mathbf{y}}_t = \mathbf{c}'\mathbf{V}(L)\boldsymbol{\mu}_t = (1 - L)\mathbf{c}'\mathbf{V}^*(L)\boldsymbol{\mu}_t = \mathbf{c}'\mathbf{V}^*(L)\mathbf{v}_t,$$

so that $\mathbf{c}'\mathbf{W} = w$. In conclusion, $\mathbf{c}'\mathbf{y}_t$ is weakly stationary with rational spectral density if and only if, in the solution $\mathbf{y}_t = \tilde{\mathbf{y}}_t + \mathbf{W}$, the stochastic vector \mathbf{W} is chosen such that $\mathbf{c}'\mathbf{W}$ is constant with probability one.

The same reasoning applies to equation (28), to prove that $\boldsymbol{\xi}'_{\perp} \mathbf{S}(L)\mathbf{F}_t$ is weakly stationary with rational spectral density if and only if (29) holds and $\boldsymbol{\xi}'_{\perp} \mathbf{S}(1)\mathbf{W} = \mathbf{k}$.

Appendix A.2 Proof of Proposition 2

With one exception at the end of the proof, we keep using the notation $\mathbf{C}(z)$, $\mathbf{M}(z)$, etc., avoiding explicit dependence on $\mathbf{p} \in \Pi$ (see Definition 1).

Remark 1. Suppose that the statement $S(\mathbf{p})$, depending on a vector $\mathbf{p} \in \Pi$, is equivalent to a set of polynomial equations for the parameters, for example the statement that all the $q \times q$ minors of $\mathbf{M}(1)$ vanish, i.e. that $\text{rank}(\mathbf{M}(1)) < q$. Statement $S(\mathbf{p})$ is true either for a nowhere dense subset of Π or for the whole Π . Thus, if the statement is false for one point in Π , it is generically false in Π . Moreover, $S(\mathbf{p})$ can be obviously extended to any $\mathbf{p} \in \mathbb{R}^{\lambda}$ and, as Π is an open subset of \mathbb{R}^{λ} , if the statement S is false for one point in \mathbb{R}^{λ} , then it is generically false in \mathbb{R}^{λ} and therefore in Π .

Remark 2. Remark 1 can be used to show that generically the stochastic vectors of a reduced rank $I(1)$ family, see Definition 4, are of rank q . Let \mathbf{p}^* be a point in \mathbb{R}^{λ} such that $\boldsymbol{\xi} = \mathbf{0}$, $\boldsymbol{\eta} = \mathbf{0}$, $\mathbf{S}(L) = \mathbf{I}_r$, $\mathbf{E}(L) = \mathbf{0}$ and let \mathbf{D} be of rank q (note that \mathbf{p}^* does not necessarily belong to Π). Given $\theta^* \in [-\pi, \pi]$, $\theta^* \neq 0$, the matrix $\mathbf{S}(e^{-i\theta^*})^{-1}\mathbf{C}(e^{-i\theta^*})$ is equal to $(1 - e^{-i\theta^*})\mathbf{D}$ at \mathbf{p}^* and has therefore rank q , so that, by Remark 1, has rank q generically in Π . Thus, generically in Π , the spectral density of $(1 - L)\mathbf{F}_t$, which is a rational function of $e^{-i\theta}$, has rank q except for a finite subset of $[-\pi, \pi]$ (depending on \mathbf{p}).

Remark 3. Consider the polynomials

$$A(z) = a_0 z^n + a_1 z^{n-1} + \cdots + a_n, \quad B(z) = b_0 z^m + b_1 z^{m-1} + \cdots + a_m$$

and let α_i , $i = 1, \dots, n$ and β_j , $j = 1, \dots, m$, be the roots of A and B respectively. Suppose that $a_0 \neq 0$ and $b_0 \neq 0$. Then, see van der Waerden (1953, pp. 83-8),

$$a_0^m b_0^n \prod_{i,j} (\alpha_i - \beta_j) = R(a_0, a_1, \dots, a_n; b_0, b_1, \dots, b_m),$$

where R is a polynomial function. The function R is called the resultant of A and B . The resultant vanishes if and only if A and B have a common root. Now suppose that the coefficients a_i and b_j are polynomial functions of $\mathbf{p} \in \Pi$. Then, by Remark 1, if there exists a point $\tilde{\mathbf{p}} \in \Pi$ (or $\tilde{\mathbf{p}} \in \mathbb{R}^\lambda$) such that $a_0(\tilde{\mathbf{p}}) \neq 0$, $b_0(\tilde{\mathbf{p}}) \neq 0$, and $R(\tilde{\mathbf{p}}) \neq 0$, then generically A and B have no common roots.

Remark 4. Recall that a zero of $\mathbf{M}(z)$ is a complex number z^* such that $\text{rank}(\mathbf{M}(z^*)) < q$ (see Proposition 1). If $\mathbf{M}(z)$ has two $q \times q$ submatrices whose determinants have no common roots, then $\mathbf{M}(z)$ is zeroless.

Starting with

$$\mathbf{C}(z) = \boldsymbol{\xi} \boldsymbol{\eta}' + (1-z) \mathbf{D} + (1-z)^2 \mathbf{E}(z),$$

we obtain, see Section 3.3,

$$\begin{aligned} \zeta \mathbf{C}(z) &= \begin{pmatrix} (1-z) \mathbf{I}_c & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{r-c} \end{pmatrix} \left\{ \begin{pmatrix} \boldsymbol{\xi}'_{\perp} \mathbf{D} \\ \boldsymbol{\xi}' \boldsymbol{\xi} \boldsymbol{\eta}' \end{pmatrix} + (1-z) \begin{pmatrix} \boldsymbol{\xi}'_{\perp} \mathbf{E}(z) \\ \boldsymbol{\xi}' \mathbf{D} \end{pmatrix} + (1-z)^2 \begin{pmatrix} \mathbf{0}_{c \times q} \\ \boldsymbol{\xi}' \mathbf{E}(z) \end{pmatrix} \right\} \\ &= \begin{pmatrix} (1-z) \mathbf{I}_c & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{r-c} \end{pmatrix} \mathbf{M}(z). \end{aligned}$$

With no loss of generality we can assume that $r = q + 1$, see Remark 4. We denote by $\mathbf{M}_1(z)$ and $\mathbf{M}_2(z)$ the $q \times q$ matrices obtained by dropping the first and the last row of $\mathbf{M}(z)$ respectively. The degrees of the polynomials $\det(\mathbf{M}_1(z))$ and $\det(\mathbf{M}_2(z))$ are $d_1 = (q-d)(s_1+2) + d(s_1+1)$ and $d_2 = (q-d-1)(s_1+2) + (d+1)(s_1+1)$.

Let us now define a subfamily of $\mathbf{M}(z)$, denoted by $\underline{\mathbf{M}}(z)$, obtained by specifying $\boldsymbol{\eta}$, $\boldsymbol{\xi}$, $\boldsymbol{\xi}'_{\perp}$, \mathbf{D} and $\mathbf{E}(z)$ in the following way:

$$\underline{\boldsymbol{\eta}}' = (\mathbf{0}_{(q-d) \times d} \mathbf{I}_{q-d}), \quad \underline{\boldsymbol{\xi}} = \begin{pmatrix} \mathbf{I}_{q-d} \\ \mathbf{0}_{c \times (q-d)} \end{pmatrix}, \quad \underline{\boldsymbol{\xi}}'_{\perp} = \begin{pmatrix} \mathbf{K} \\ \mathbf{H} \end{pmatrix}, \quad \underline{\mathbf{D}} = (\mathbf{H}' \quad \mathbf{0}_{(q+1) \times (q-d)}), \quad \underline{\mathbf{E}}(z) = \begin{pmatrix} \mathbf{E}_1(z) \\ \mathbf{E}_2(z) \\ \mathbf{E}_3(z) \end{pmatrix},$$

where

$$\mathbf{K} = (\mathbf{0}_{1 \times (q-d)} \quad 1 \quad \mathbf{0}_{1 \times d}), \quad \mathbf{H} = (\mathbf{0}_{d \times (q+1-d)} \quad \mathbf{I}_d),$$

$$\mathbf{E}_1(z) = \begin{pmatrix} & k_1(z) & h_1(z) & \cdots & 0 \\ \mathbf{0}_{(q-d) \times d} & & \ddots & \ddots & \\ & 0 & \cdots & \ddots & h_{q-d-1}(z) \\ & & & & k_{q-d}(z) \end{pmatrix}$$

$$\mathbf{E}_2(z) = (e(z) \quad \mathbf{0}_{1 \times (q-1)})$$

$$\mathbf{E}_3(z) = \begin{pmatrix} f_1(z) & g_1(z) & \cdots & 0 \\ & \ddots & \ddots & \\ 0 & \cdots & f_d(z) & g_d(z) \end{pmatrix}, \quad \mathbf{0}_{d \times (q-d-1)}$$

the polynomial entries e , k_i , h_i , f_i and g_i being of degree s_1 . We have:

$$\underline{\mathbf{M}}(z) = \begin{pmatrix} \mathbf{0}_{1 \times d} & \mathbf{0}_{1 \times (q-d)} \\ \mathbf{I}_d & \mathbf{0}_{d \times (q-d)} \\ \mathbf{0}_{(q-d) \times d} & \mathbf{I}_{q-d} \end{pmatrix} + (1-z) \begin{pmatrix} \mathbf{E}_2(z) \\ \mathbf{E}_3(z) \\ \mathbf{0}_{(q-d) \times q} \end{pmatrix} + (1-z)^2 \begin{pmatrix} \mathbf{0}_{1 \times q} \\ \mathbf{0}_{d \times q} \\ \mathbf{E}_1(z) \end{pmatrix}.$$

Notice that $\underline{\mathbf{M}}(z)$ has zero entries except for the diagonal joining the positions $(1, 1)$ and (q, q) , and the diagonal joining $(2, 1)$ and $(q+1, q)$. The matrices $\underline{\mathbf{M}}_1(z)$ and $\underline{\mathbf{M}}_2(z)$, obtained by dropping the first and the last row of $\underline{\mathbf{M}}(z)$, respectively, are upper- and lower-triangular, respectively. Moreover,

$$\begin{aligned} \det(\underline{\mathbf{M}}_1(z)) &= [1 + (1-z)f_1(z)] \cdots [(1 + (1-z)f_d(z)] \\ &\quad \times [1 + (1-z)^2 k_1(z)] \cdots [1 + (1-z)^2 k_{q-d}(z)] \\ \det(\underline{\mathbf{M}}_2(z)) &= (1-z)^{2q-d-1} e(z) [g_1(z) \cdots g_d(z)] [h_1(z) \cdots h_{q-d-1}(z)] \end{aligned}$$

Now:

- (i) The leading coefficient of $\det(\underline{\mathbf{M}}_1(z))$, call it Q_1 , corresponding to z^{d_1} , is the product of the leading coefficients of the polynomials $k_j(z)$, $j = 1, \dots, (q-d)$ and $f_i(z)$, $i = 1, \dots, d$. Trivially, there exist values for the parameters of the polynomials k_j and f_i , such that $Q_1 \neq 0$. Let ω_1 be the vector of such parameters and $\underline{\mathbf{M}}_1^{\omega_1}(z)$ the matrix $\underline{\mathbf{M}}_1(z)$ corresponding to the parameters in ω_1 .
- (ii) Now observe, firstly, that the polynomials $\det(\underline{\mathbf{M}}_1(z))$ and $\det(\underline{\mathbf{M}}_2(z))$ have no parameters in common, and, secondly, that the parameters of $\det(\underline{\mathbf{M}}_2(z))$ vary in an open set (each one is a subvector of the parameters vector of $\mathbf{M}(z)$, which varies in the open set Π). As a consequence, there exist parameters for the polynomials e , g_j and h_i , call ω_2 the vector of such parameters, such that (1) the leading coefficient of $\det(\underline{\mathbf{M}}_2^{\omega_2}(z))$ does not vanish, (2) $\det(\underline{\mathbf{M}}_1^{\omega_1}(z))$ and $\det(\underline{\mathbf{M}}_2^{\omega_2}(z))$ have no roots in common. This implies that, as the leading coefficient of $\det(\underline{\mathbf{M}}_1^{\omega_1}(z))$ does not vanish as well, the resultant of $\det(\underline{\mathbf{M}}_1^{\omega_1}(z))$ and $\det(\underline{\mathbf{M}}_2^{\omega_2}(z))$ does not vanish, see Remark 2.
- (iii) Combining the parameters in ω_1 and ω_2 , we determine a point $\mathbf{p} \in \Pi$ such that, at \mathbf{p} , the leading coefficient of $\det(\underline{\mathbf{M}}_1(z))$ and $\det(\underline{\mathbf{M}}_2(z))$ and their resultant do not vanish. As the leading coefficients and the resultant of $\det(\underline{\mathbf{M}}_1(z))$ and $\det(\underline{\mathbf{M}}_2(z))$ are polynomial functions of the parameters in \mathbf{p} , then by Remarks 2 and 3, $\mathbf{M}(z)$ is generically zeroless.

Appendix B Non uniqueness

In Proposition 3 we prove that a singular $I(1)$ vector has a finite Error Correction representation with c error correction terms. However, as anticipated in Section 2, this representation is not unique since: (i) different Error Correction representations can be obtained in which the number of error terms varies between d and $r - q + d$, (ii) the left inverse of the matrix $\mathbf{M}(L)$ may be not unique. We discuss these two causes of non-uniqueness for representation (36) below. In Appendix B.3 we show that all such representation produce the same impulse-response functions.

Appendix B.1 Alternative representations with different numbers of error correction terms

Let, for simplicity, $\mathbf{S}(L) = \mathbf{I}_r$ and consider the following example, with $r = 3$, $q = 2$, $c = 2$, so that $d = 1$:

$$\begin{aligned}\boldsymbol{\xi}' &= (1 \quad 1 \quad 1) \\ \boldsymbol{\eta}' &= (1 \quad 2) \\ \boldsymbol{\xi}'_{\perp} &= \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}\end{aligned}$$

We have,

$$(1-L) \begin{pmatrix} \boldsymbol{\xi}'_{\perp} \\ \boldsymbol{\xi}' \end{pmatrix} \mathbf{F}_t = \begin{pmatrix} 1-L & 0 & 0 \\ 0 & 1-L & 0 \\ 0 & 0 & 1 \end{pmatrix} \left\{ \begin{pmatrix} d_{11} - d_{21} & d_{12} - d_{22} \\ d_{21} - d_{31} & d_{22} - d_{32} \\ 3 & 6 \end{pmatrix} + (1-L)\mathbf{G}(L) \right\} \mathbf{u}_t,$$

where $(1-L)\mathbf{G}(L)$ gathers the second and third terms within curly brackets in the second line of (27). If the first matrix within the curly brackets has full rank, we can proceed as in Proposition 3 and obtain an Error Correction representation with error terms

$$\boldsymbol{\xi}'_{\perp} \mathbf{F}_t = \begin{pmatrix} F_{1t} - F_{2t} \\ F_{2t} - F_{3t} \end{pmatrix}.$$

However, we also have

$$\begin{aligned}(1-L) \begin{pmatrix} \boldsymbol{\xi}'_{\perp} \\ \boldsymbol{\xi}' \end{pmatrix} \mathbf{F}_t &= \begin{pmatrix} 1-L & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \left\{ \begin{pmatrix} d_{11} - d_{21} & d_{12} - d_{22} \\ (1-L)(d_{21} - d_{31}) & (1-L)(d_{22} - d_{32}) \\ 3 & 6 \end{pmatrix} \right. \\ &\quad \left. + (1-L)\tilde{\mathbf{G}}(L) \right\} \mathbf{u}_t = \begin{pmatrix} 1-L & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tilde{\mathbf{M}}(L) \mathbf{u}_t.\end{aligned}$$

Assuming that the matrix

$$\begin{pmatrix} d_{11} - d_{21} & d_{12} - d_{22} \\ 3 & 6 \end{pmatrix}$$

is non-singular, the matrix $\tilde{\mathbf{M}}(L)$ is zeroless and has therefore a finite-degree left inverse. Proceeding as in Proposition 3, we obtain an alternative Error Correction representation with just one error term, namely $F_{1t} - F_{2t}$.

This example can be generalized to show that generically \mathbf{F}_t admits Error Correction representations with a minimum d and a maximum $r - q + d$ of error terms. In particular, if $d = 0$, in addition to an Error Correction representation, \mathbf{F}_t generically has a finite-degree autoregressive representation with no error terms (i.e. a VAR), consistently with the fact that in this case $\mathbf{C}(L)$ is generically zeroless.

Experiments with simulated and actual data suggest that the best results in estimation of singular VECMs are obtained using c (the maximum number of) error correction terms.

Appendix B.2 The left inverse of $\mathbf{M}(L)$ is not necessarily unique

In the proof of Proposition 3 we have used the fact that generically the matrix $\mathbf{M}(L)$ has a finite-degree left inverse $\mathbf{N}(L)$. We now give some examples in which $\mathbf{N}(L)$ is not unique. This is a well known fact, see also Forni and Lippi (2010), Forni et al. (2015).

Consider

$$(1 - L)F_t = \begin{pmatrix} 1 + aL \\ 1 + bL \end{pmatrix} u_t, \quad (\text{B3})$$

with $r = 2$, $q = 1$, $d = 0$, $c = 1$, with $a \neq b$. In this case $\mathbf{A}(L)$ is zeroless. An autoregressive representation can be obtained by elementary manipulations. Rewrite (B3) as

$$\begin{aligned} (1 - L)F_{1t} &= u_t + au_{t-1} \\ (1 - L)F_{2t} &= u_t + bu_{t-1} \end{aligned} \quad (\text{B4})$$

Taking $(b - a)\mathbf{C}(L)\mathbf{u}_t$, we get

$$u_t = \frac{b(1 - L)F_{1t} - a(1 - L)F_{2t}}{b - a}.$$

This can be used to get rid of u_{t-1} in (B4) and obtain

$$\left[\mathbf{I}_2 - \begin{pmatrix} \frac{ab}{b-a} & \frac{a^2}{b-a} \\ \frac{b^2}{b-a} & -ab \end{pmatrix} L \right] (1 - L)\mathbf{F}_t = \begin{pmatrix} 1 \\ 1 \end{pmatrix} u_t, \quad (\text{B5})$$

which is an autoregressive representation in first differences.

Model (B4), slightly modified, can be used to illustrate non-uniqueness in the left inversion of $\mathbf{M}(L)$. Consider

$$\begin{aligned} (1 - L)F_{1t} &= u_t + au_{t-1} \\ (1 - L)F_{2t} &= u_t + bu_{t-1} \\ (1 - L)F_{3t} &= u_t + cu_{t-1}. \end{aligned} \quad (\text{B6})$$

Taking any vector $\mathbf{h} = (h_1 \ h_2 \ h_3)$, orthogonal to $(a \ b \ c)$, we get rid of u_{t-1} in (B6) and obtain an autoregressive representation in the differences. However, unlike in (B4), here the vectors \mathbf{h} span a 2-dimensional space, thus producing an infinite set of autoregressive representations.

In the example just above non-uniqueness can also be seen as the consequence of the fact that the three stochastic variables $F_{j,t-1}$, $j = 1, 2, 3$, are linearly dependent. Therefore, projecting each of the F_{jt} onto the space spanned by $F_{j,t-1}$, $j = 1, 2, 3$, one would find a non-invertible covariance matrix, thus a unique projection of course but many representations of it as linear combinations of $F_{j,t-1}$, $j = 1, 2, 3$.

We do not address this problem systematically in the present paper. However, in the empirical analysis of Barigozzi et al. (2016) we find no hint of singular covariance matrices.

Appendix B.3 Uniqueness of impulse-response functions

Start with representation (19)

$$(1-L)\mathbf{F}_t = \mathbf{S}(L)^{-1}\mathbf{C}(L)\mathbf{u}_t = \mathbf{U}(L)\mathbf{u}_t = \mathbf{U}_0\mathbf{u}_t + \mathbf{U}_1\mathbf{u}_{t-1} + \dots = \mathbf{C}(0)\mathbf{u}_t + \mathbf{U}_1\mathbf{u}_{t-1} + \dots \quad (\text{B7})$$

We assume that \mathbf{u}_t is fundamental for $(1-L)\mathbf{F}_t$, see Proposition 4. The impulse response function of \mathbf{F}_t to \mathbf{u}_t is

$$\mathbf{H}_j = \mathbf{U}_0 + \dots + \mathbf{U}_j, \quad i = 0, 1, \dots$$

Now suppose that \mathbf{F}_t fulfills the autoregressive equation

$$\mathbf{B}(L)\mathbf{F}_t = (\mathbf{I}_r + \mathbf{B}_1L + \dots + \mathbf{B}_mL^m)\mathbf{F}_t = \tilde{\mathbf{m}} + \tilde{\mathbf{R}}\tilde{\mathbf{u}}_t \quad (\text{B8})$$

where: (i) $\tilde{\mathbf{R}}$ is a full-rank $r \times q$ matrix, (ii) $\tilde{\mathbf{u}}_t$ is q -dimensional white noise, (iii) $\tilde{\mathbf{u}}_t$ is orthogonal to $(1-L)\mathbf{F}_\tau$, for $\tau \geq 0$. Applying $(1-L)$ to both sides of (B8) we obtain

$$\mathbf{B}(L)\mathbf{U}(L)\mathbf{u}_t = (1-L)\tilde{\mathbf{R}}\tilde{\mathbf{u}}_t. \quad (\text{B9})$$

Assumption (i) and the argument mentioned in footnote 4 imply that $\tilde{\mathbf{u}}_t$ belongs to the space spanned by \mathbf{u}_τ , for $\tau \geq 0$, call it $\mathcal{H}_{u,t}$. Now consider the projection

$$\tilde{\mathbf{u}}_t = \mathbf{G}_0\mathbf{u}_t + \mathbf{G}_1\mathbf{u}_{t-1} + \dots$$

Multiplying both sides by \mathbf{u}'_{t-k} and taking expected values:

$$\mathbb{E}\tilde{\mathbf{u}}_t\mathbf{u}'_{t-k} = \mathbf{G}_k\boldsymbol{\Gamma}_u.$$

By assumption (iii), $\tilde{\mathbf{u}}_t$ is orthogonal to $\mathcal{H}_{(1-L)\mathbf{F},t-k}$, for $k > 0$, which is equal to $\mathcal{H}_{u,t-k}$, for $k > 0$ (a consequence of the fundamentalness of \mathbf{u}_t in (B7)). Thus $\mathbf{G}_k = \mathbf{0}$, for $k > 0$ and

$$\tilde{\mathbf{u}}_t = \mathbf{G}_0\mathbf{u}_t,$$

where \mathbf{G}_0 is a non-singular $q \times q$ matrix. Therefore, with no loss of generality, equation (B8) can be rewritten with $\mathbf{R}\mathbf{u}_t$ instead of $\tilde{\mathbf{R}}\tilde{\mathbf{u}}_t$ and (B9) becomes

$$\mathbf{B}(L)\mathbf{U}(L)\mathbf{u}_t = (1-L)\mathbf{R}\mathbf{u}_t.$$

As \mathbf{u}_t is a non-singular q dimensional white noise, this implies

$$\mathbf{B}(L)\mathbf{U}(L) = (1-L)\mathbf{R},$$

so that:

$$\begin{aligned} \mathbf{U}_0 &= \mathbf{R}, \\ \mathbf{B}_1\mathbf{R} + \mathbf{U}_1 &= -\mathbf{R}, \quad \mathbf{U}_1 = -(\mathbf{I}_r + \mathbf{B}_1)\mathbf{R}, \\ \mathbf{B}_2\mathbf{R} + \mathbf{B}_1\mathbf{U}_1 + \mathbf{U}_2 &= \mathbf{0}, \quad \mathbf{U}_2 = (\mathbf{B}_1 + \mathbf{B}_1^2 - \mathbf{B}_2)\mathbf{R}, \\ &\vdots \end{aligned}$$

and therefore

$$\begin{aligned} \mathbf{H}_0 &= \mathbf{U}_0 = \mathbf{R}, \\ \mathbf{H}_1 &= \mathbf{U}_0 + \mathbf{U}_1 = -\mathbf{B}_1\mathbf{R}, \\ \mathbf{H}_2 &= \mathbf{U}_0 + \mathbf{U}_1 + \mathbf{U}_2 = (\mathbf{B}_1^2 - \mathbf{B}_2)\mathbf{R}, \\ &\vdots \end{aligned}$$

On the other hand, the impulse-response function implicit in (B8) is given by the coefficient matrices of $\mathbf{K}(L)\mathbf{R}$, where

$$\mathbf{K}(L)\mathbf{B}(L) = (\mathbf{I}_r + \mathbf{K}_1L + \dots)\mathbf{B}(L) = \mathbf{I}_r.$$

It is easily seen that $\mathbf{K}_j\mathbf{R} = \mathbf{H}_j$.

Note that we are not making assumptions on $\mathbf{B}(1)$ in equation (B8). When $d = 0$, equation (B8) can be the autoregressive model in differences that results from left-inverting $\mathbf{U}(L)$ (no error correction term):

$$\tilde{\mathbf{B}}(L)(1 - L)\mathbf{F}_t = \mathbf{B}(L)\mathbf{F}_t = \mathbf{C}(0)\mathbf{u}_t.$$

Replacing \mathbf{u}_t with any other white noise vector $\mathbf{w}_t = \mathbf{Q}\mathbf{u}_t$, as we do when the shocks are identified according to restrictions based on economic theory, produces different impulse-response functions that are however independent of the autoregressive representation of \mathbf{F}_t .

Appendix C Data Generating Process for the Simulations

The simulation results of Section 3.3 are obtained using the following specification of (36):

$$\mathbf{A}(L)\mathbf{F}_t = \mathbf{A}^*(L)(1 - L)\mathbf{F}_t + \alpha\beta'\mathbf{F}_{t-1} = \mathbf{C}(0)\mathbf{u}_t = \mathbf{G}\mathbf{H}\mathbf{u}_t, \quad (\text{C10})$$

where $r = 4$, $q = 3$, $c = 3$, the degree of $\mathbf{A}(L)$ is 2, so that the degree of $\mathbf{A}^*(L)$ is 1. $\mathbf{A}(L)$ is generated using the factorization

$$\mathbf{A}(L) = \mathbf{U}(L)\mathcal{M}(L)\mathcal{V}(L),$$

where $\mathbf{U}(L)$ and $\mathcal{V}(L)$ are $r \times r$ matrix polynomials with all their roots outside the unit circle, and

$$\mathcal{M}(L) = \begin{pmatrix} (1 - L)\mathbf{I}_{r-c} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_c \end{pmatrix}$$

(see Watson, 1994). To get a VAR(2) we set $\mathbf{U}(L) = \mathbf{I}_r - \mathbf{U}_1L$, and $\mathcal{V}(L) = \mathbf{I}_r$, and then, by rewriting $\mathcal{M}(L) = \mathbf{I}_r - \mathcal{M}_1L$, we get $\mathbf{A}_1 = \mathcal{M}_1 + \mathbf{U}_1$, and $\mathbf{A}_2 = -\mathcal{M}_1\mathbf{U}_1$.

The data are then generated as follows. The diagonal elements of the matrix \mathbf{U}_1 are drawn from a uniform distribution between 0.5 and 0.8, while the extra-diagonal elements from a uniform distribution between 0 and 0.3. \mathbf{U}_1 is then standardized to ensure that its largest eigenvalue is 0.6. The matrix \mathbf{G} is generated as in Bai and Ng (2007). Let $\tilde{\mathbf{G}}$ be a $r \times r$ diagonal matrix of rank q with non-zero entries \tilde{g}_{ii} drawn from the uniform distribution between 0.8 and 1.2, and let $\check{\mathbf{G}}$ be a random $r \times r$ orthogonal matrix. Then, \mathbf{G} is equal to the first q columns of the matrix $\check{\mathbf{G}}\tilde{\mathbf{G}}^{1/2}$. Lastly, the matrix \mathbf{H} is such that the upper 3×3 submatrix of $\mathbf{G}\mathbf{H}$ is lower triangular.

Results are based on 1000 replications. The matrices \mathbf{U}_1 , \mathbf{G} and \mathbf{H} are simulated only once so that the set of impulse responses to be estimated is the same for all replications, whereas the vector \mathbf{u}_t is redrawn from $\mathcal{N}(\mathbf{0}, \mathbf{I}_4)$ at each replication.