

Article

# Cointegration and Error Correction Mechanisms for Singular Stochastic Vectors

Matteo Barigozzi <sup>1</sup>, Marco Lippi <sup>2,\*</sup> and Matteo Luciani <sup>3</sup><sup>1</sup> Università di Bologna, Department of Economics, 40126 Bologna, Italy; matteo.barigozzi@unibo.it<sup>2</sup> Einaudi Institute for Economics and Finance, 00187 Roma, Italy<sup>3</sup> Federal Reserve Board of Governors, Washington, DC 20551, USA; matteo.luciani@frb.gov

\* Correspondence: mlippieief@gmail.com

Received: 28 March 2018; Accepted: 7 January 2020; Published: 4 February 2020



**Abstract:** Large-dimensional dynamic factor models and dynamic stochastic general equilibrium models, both widely used in empirical macroeconomics, deal with singular stochastic vectors, i.e., vectors of dimension  $r$  which are driven by a  $q$ -dimensional white noise, with  $q < r$ . The present paper studies cointegration and error correction representations for an  $I(1)$  singular stochastic vector  $\mathbf{y}_t$ . It is easily seen that  $\mathbf{y}_t$  is necessarily cointegrated with cointegrating rank  $c \geq r - q$ . Our contributions are: (i) we generalize Johansen's proof of the Granger representation theorem to  $I(1)$  singular vectors under the assumption that  $\mathbf{y}_t$  has rational spectral density; (ii) using recent results on singular vectors by Anderson and Deistler, we prove that for generic values of the parameters the autoregressive representation of  $\mathbf{y}_t$  has a finite-degree polynomial. The relationship between the cointegration of the factors and the cointegration of the observable variables in a large-dimensional factor model is also discussed.

**Keywords:** singular stochastic vectors; cointegration for singular vectors; Granger representation theorem; large-dimensional dynamic factor models)

**JEL Classification:** C0; C01; E0

## 1. Introduction

An  $r$ -dimensional stochastic vector  $\mathbf{y}_t$  such that  $\mathbf{y}_t = \mathbf{A}_0\mathbf{u}_t + \mathbf{A}_1\mathbf{u}_{t-1} + \dots$ , where the matrices  $\mathbf{A}_j$  are  $r \times q$  and  $\mathbf{u}_t$  is a  $q$ -dimensional white noise, with  $q < r$ , is said to be singular. Singular stochastic vectors have been systematically analyzed in a number of papers starting with (Anderson and Deistler 2008a, 2008b). A motivation for studying the consequences of singularity, as argued by these authors, is that the factors' vector in large-dimensional dynamic factor models (DFM), such as those introduced in Forni et al. (2000); Forni and Lippi (2001), (Stock and Watson 2002a, 2002b), is typically singular. Singularity is also an important feature of dynamic stochastic general equilibrium models (DSGE), see e.g., Sargent (1989), Canova (2007), pp. 230–2. Singularity as it arises in DFMs is presented in some detail below.

DFMs are based on the idea that all the observed variables in an economic system are driven by a few common (macroeconomic) shocks and by idiosyncratic components which may result from measurement errors and sectoral or regional shocks. Formally, each variable in the  $n$ -dimensional dataset  $x_{it}$ ,  $i = 1, 2, \dots, n$ ,  $t = 1, 2, \dots, T$ , is decomposed into the sum of a common component  $\chi_{it}$ , and an idiosyncratic component  $\epsilon_{it}$ :  $x_{it} = \chi_{it} + \epsilon_{it}$ , where  $\chi_{it}$  and  $\epsilon_{js}$  are orthogonal for all  $i, j, t, s$ . In the standard version of the DFM the common components are linear combinations of an  $r$ -dimensional vector of common factors  $\mathbf{F}_t = (F_{1t} F_{2t} \dots F_{rt})'$ ,

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

Now suppose that the observable variables  $x_{it}$  and the common factors  $\mathbf{F}_t$  are  $I(1)$  and that

$$(1 - L)\mathbf{F}_t = \mathbf{C}(L)\mathbf{u}_t, \quad (2)$$

where  $\mathbf{u}_t$  is a nonsingular  $q$ -dimensional white-noise vector<sup>1</sup>, the common shocks. A number of papers analyzing macroeconomic databases find strong empirical support for the assumption that the vector  $\mathbf{F}_t$  is singular, i.e., that  $q < r$ . See, for US datasets, [Giannone et al. \(2005\)](#); [Amengual and Watson \(2007\)](#); [Forni and Gambetti \(2010\)](#), [Luciani \(2015\)](#). For a Euro-area dataset, see [Barigozzi et al. \(2014\)](#).

Such results can be easily understood observing that usually the static Equation (1) is just a convenient representation derived from a “primitive” set of dynamic equations linking the common components  $\chi_{it}$  to the common shocks  $\mathbf{u}_t$ . As a simple example, suppose that the variables  $x_{it}$  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 we can set  $F_{1t} = f_t$ ,  $F_{2t} = f_{t-1} = F_{1,t-1}$ ,  $\lambda_{i1} = a_{i0}$ ,  $\lambda_{i2} = a_{i1}$ , so that Equations (1) and (2) take the form

$$x_{it} = \lambda_{i1}F_{1t} + \lambda_{i2}F_{2t} + \epsilon_{it} \quad \text{and} \quad \begin{pmatrix} F_{1t} \\ F_{2t} \end{pmatrix} = \begin{pmatrix} (1 - \alpha L)^{-1} \\ L(1 - \alpha L)^{-1} \end{pmatrix} u_t,$$

respectively. Here  $r = 2$  and  $q = 1$  so that  $\mathbf{F}_t$  is singular. For a general analysis of the relationship between representation (1) and “deeper” dynamic representations like (3), see e.g., [Forni et al. \(2009\)](#); [Stock and Watson \(2016\)](#).

Now suppose that the factors  $\mathbf{F}_t$  have been estimated. Obtaining  $\mathbf{u}_t$  and the impulse-response functions of the variables  $x_{it}$  with respect to  $\mathbf{u}_t$  (or structural shocks obtained by a linear transformation of  $\mathbf{u}_t$ ) requires the estimation of a VAR for the singular  $I(1)$  vector  $\mathbf{F}_t$ . On the other hand, the latter is necessarily cointegrated with cointegration rank  $c$  at least equal to  $r - q$  (the rank of the spectral density of  $(1 - L)\mathbf{F}_t$  does not exceed  $q$  at all frequencies and, therefore, at frequency zero).

Singular vectors of factors in an  $I(1)$  DFM and  $I(1)$  singular vectors in DSGE models provide strong motivation for studying singular  $I(1)$  vectors in a general time-series context. The main contributions of the paper are:

- (I) A generalization of Johansen’s proof of the Granger Representation Theorem (from MA to AR), this is Proposition 2. Consider an  $I(1)$  singular vector  $\mathbf{y}_t$ , with dimension  $r$ , rank  $q < r$ , and cointegrating rank  $c \geq r - q$ . Assuming that  $(1 - L)\mathbf{y}_t$  has an ARMA structure,  $\mathbf{S}(L)(1 - L)\mathbf{y}_t = \mathbf{B}(L)\mathbf{u}_t$  and that some simple additional conditions hold,  $\mathbf{y}_t$  has a representation as a vector error correction mechanism (VECM) with  $c$  error correction terms:

$$\mathbf{A}(L)\mathbf{y}_t = \mathbf{A}^*(L)(1 - L)\mathbf{y}_t + \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{y}_{t-1} - \mathbf{w}) = \mathbf{B}(0)\mathbf{u}_t, \quad (4)$$

where  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  are both  $r \times c$  and full rank,  $\boldsymbol{\beta}'\mathbf{y}_t - \mathbf{w}$  is  $I(0)$ ,  $\mathbf{A}(L)$  and  $\mathbf{A}^*(L)$  are  $r \times r$  rational matrices in  $L$ . Under the additional assumption that unity is the only zero of  $\mathbf{B}(L)$ , i.e., if  $z \neq 1$  then  $\mathbf{B}(z)$  is full rank,  $\mathbf{A}(L)$  and  $\mathbf{A}^*(L)$  are finite-degree matrix polynomials.

- (II) Assuming that the parameters of  $\mathbf{S}(L)$  and  $\mathbf{B}(L)$  may vary in an open subset of  $\mathbb{R}^\lambda$ , see Section 3.2 for the definition of  $\lambda$ , in Proposition 3 we show that all the assumptions used to obtain (4), and also the assumption that unity is the only possible zero of  $\mathbf{B}(L)$ , hold for generic values of

<sup>1</sup> Usually orthonormality is assumed. This is convenient but not necessary in the present paper.

the parameters. This implies that the matrices  $\mathbf{A}(L)$  and  $\mathbf{A}^*(L)$  are generically of finite degree, which is obviously not the case for nonsingular vectors.<sup>2</sup>

The paper is organized as follows. Section 2 is preliminary. We firstly recall recent results for stationary singular stochastic vectors with rational spectral density, see (Anderson and Deistler 2008a, 2008b). Secondly, we discuss cointegration and the cointegrating rank for  $I(1)$  singular stochastic vectors.

In Section 3 we prove our main results. We also obtain the permanent-transitory shock representation in the singular case:  $\mathbf{y}_t$  is driven by  $r - c$  permanent shocks, i.e.,  $r$  minus the cointegrating rank, the usual result. However, the number of transitory shocks is  $c - (r - q)$ , not  $c$  as in the nonsingular case.

Section 3 also contains an exercise carried out 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 nonsingular vectors, that 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  $x_{it}$  in a DFM. Our results on cointegration of the singular vector  $\mathbf{F}_t$  have the implication that  $p$ -dimensional subvectors of the  $n$ -dimensional common-component vector  $\chi_t$ , with  $p > r - c$ , are cointegrated. As a consequence, stationarity of the idiosyncratic components would imply that all  $p$ -dimensional subvectors of the  $n$ -dimensional dataset  $\mathbf{x}_t$  are cointegrated if  $p > r - c$ . 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 suggests that an estimation strategy robust to the assumption that the idiosyncratic components can be  $I(1)$  has to be preferred (for this aspect we refer to Barigozzi et al. 2019). Section 5 concludes. Some proofs, a discussion of some non-uniqueness problems arising with singularity and details on the simulations are collected in the Appendix.

## 2. Stationary and $I(1)$ Singular Vectors

### 2.1. Stationary Singular Vectors

As in this paper we only consider representation issues it is convenient to assume that all stochastic processes are defined for  $t \in \mathbb{Z}$ . Accordingly, the lag operator  $L$  is defined as  $Ly_t = y_{t-1}$  for  $t \in \mathbb{Z}$  (Bauer and Wagner (2012) also study  $I(1)$  and cointegrated processes for  $t \in \mathbb{Z}$ ).

We start by introducing results on singular vectors with an ARMA structure from (Anderson and Deistler 2008a, 2008b). Some preliminary definitions are needed.

#### Definition 1. (Zeros and Poles)

(A) When considering matrices  $\mathbf{V}(z)$  whose entries are rational functions of  $z \in \mathbb{C}$  we always assume that numerator and denominator of each entry have no common roots. If  $\mathbf{V}(z)$  is an  $r \times q$  matrix of rational functions, we say that  $z^*$  is a pole of  $\mathbf{V}(z)$  if it is a pole of some entry of  $\mathbf{V}(z)$ .

(B) Suppose that  $\mathbf{V}(z)$  is an  $r \times q$  matrix whose entries are polynomial functions of  $z \in \mathbb{C}$ , with  $q \leq r$ . We say that  $z^* \in \mathbb{C}$  is a zero of  $\mathbf{V}(z)$  if  $\text{rank}(\mathbf{V}(z^*)) < q$ , and that  $\mathbf{V}(z)$  is zeroless if it has no zeros, i.e.,  $\text{rank}(\mathbf{V}(z)) = q$  for all  $z \in \mathbb{C}$ .

<sup>2</sup> To our knowledge, the present paper is the first to study cointegration and error correction representations for  $I(1)$  singular vectors, the factors of  $I(1)$  dynamic factor models in particular. An error correction model in the DFM framework is studied in (Banerjee et al. 2014, 2017). 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 differences their results are not directly comparable to those in the present paper.

With a minor abuse of language, we may speak of zeros and poles of the corresponding matrix  $\mathbf{V}(L)$ . When a  $r \times r$  polynomial matrix  $\mathbf{S}(L)$  has all its zeros outside the unit circle we say that  $\mathbf{S}(L)$  is stable.

All the stationary vector processes considered have an ARMA structure. Precisely, the  $r$ -dimensional process  $\mathbf{y}_t$  has an ARMA structure with rank  $q$ ,  $q \leq r$ , if there exist

- (i) a non-singular  $q$ -dimensional white-noise process  $\mathbf{u}_t$ ,
- (ii) an  $r \times r$  stable polynomial matrix  $\mathbf{S}(z)$ , with  $\mathbf{S}(0) = \mathbf{I}_r$ ,
- (iii) an  $r \times q$  matrix  $\mathbf{B}(z)$  whose rank is  $q$  for all  $z$  with the exception of a finite subset of  $\mathbb{C}$ , such that

$$\mathbf{y}_t = \mathbf{V}(L)\mathbf{u}_t, \tag{5}$$

where  $\mathbf{V}(L) = \mathbf{S}(L)^{-1}\mathbf{B}(L)$ .

Suppose that  $\mathbf{y}_t$  has also the representation  $\mathbf{y}_t = \tilde{\mathbf{S}}(L)^{-1}\tilde{\mathbf{B}}(L)\tilde{\mathbf{u}}_t$ , where  $\tilde{\mathbf{u}}_t$  is a  $\tilde{q}$ -dimensional nonsingular white noise. Denoting by  $\Sigma_{\mathbf{y}}(\theta)$  the spectral density of  $\mathbf{y}_t$ ,

$$\Sigma_{\mathbf{y}}(\theta) = (2\pi)^{-1}\mathbf{V}(e^{-i\theta})\Sigma_{\mathbf{u}}\mathbf{V}'(e^{i\theta}),$$

so that the rank of  $\Sigma_{\mathbf{y}}(\theta)$  is  $q$  for all  $\theta$ , with the exception of a finite subset of  $[-\pi, \pi]$ . As the spectral density is independent of the ARMA representation,  $q = \tilde{q}$  and  $\tilde{\mathbf{B}}(z)$  has rank  $q$  except for a finite subset of  $\mathbb{C}$ .

**Remark 1.** Let us recall that the equation

$$\mathbf{S}(L)\zeta_t = \mathbf{B}(L)\mathbf{u}_t,$$

in the unknown vector process  $\zeta_t$ , where  $\mathbf{S}(L)$  is stable, has only one stationary solution, and this is  $\mathbf{y}_t = \mathbf{S}(L)^{-1}\mathbf{B}(L)\mathbf{u}_t$ . Thus the ARMA process  $\mathbf{y}_t$  can also be defined as the stationary solution of  $\mathbf{S}(L)\zeta_t = \mathbf{B}(L)\mathbf{u}_t$ .

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

For example, assuming that  $\mathbf{p} \in \mathcal{A} = \mathbb{R}$ , the statement “The roots of the polynomial  $x^2 + \mathbf{p}x + 1$  are distinct” holds generically in  $\mathcal{A}$ .

**Definition 3. (Rational reduced-rank family of filters)** Assume that  $r > q$  and let  $\mathcal{G}$  be a set of ordered couples  $(\mathbf{S}(L), \mathbf{B}(L))$ , where:

- (i)  $\mathbf{B}(L)$  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  $\mathbf{B}(L)$  and  $\mathbf{S}(L)$ , we assume that  $\mathbf{p} \in \Pi$ , where  $\Pi$  is an open subset of  $\mathbb{R}^\lambda$  such that for  $\mathbf{p} \in \Pi$ , (1)  $\mathbf{S}(z)$  is stable, (2)  $\text{rank}(\mathbf{B}(z)) = q$  with the exception of a finite subset of  $\mathbb{C}$ .

We say that  $\mathcal{G}$  is a rational reduced-rank family of filters with parameter set  $\Pi$ .

The notation  $\mathbf{S}^{\mathbf{p}}(L)$ ,  $\mathbf{B}^{\mathbf{p}}(L)$ , though more rigorous, would be heavy and not really necessary. We use it only in Appendix A.1.

**Proposition 1.** Assume that  $r > q$ .

- (I) Suppose that  $\mathbf{V}(L)$  is an  $r \times q$  matrix polynomial in  $L$ . If  $\mathbf{V}(z)$  is zeroless then  $\mathbf{V}(L)$  has an  $r \times r$  finite-degree stable left inverse, i.e., there exists a finite-degree polynomial  $r \times r$  matrix  $\mathbf{W}(L)$  such that:

- (a)  $\mathbf{W}(0) = \mathbf{I}_r$ , (b)  $\det(\mathbf{W}(z)) = 0$  implies  $|z| > 1$ , (c)  $\mathbf{W}(L)\mathbf{V}(L) = \mathbf{V}(0)$ . Let  $\mathbf{y}_t$  be the stationary solution of  $\mathbf{S}(L)\zeta_t = \mathbf{B}(L)\mathbf{u}_t$  and suppose that  $\mathbf{B}(L)$  is zeroless. Then  $\mathbf{y}_t$  has a finite vector autoregressive representation (VAR)  $\mathbf{A}(L)\mathbf{y}_t = \mathbf{B}(0)\mathbf{u}_t$ , where  $\mathbf{A}(L) = \mathbf{N}(L)\mathbf{S}(L)$  and  $\mathbf{N}(L)$  is a finite-degree left inverse of  $\mathbf{B}(L)$ .
- (II) Assume that  $\mathbf{y}_t$  is the stationary solution of  $\mathbf{S}(L)\zeta_t = \mathbf{B}(L)\mathbf{u}_t$ , where  $(\mathbf{S}(L), \mathbf{B}(L))$  belongs to a rational reduced-rank family of filters with parameter set  $\Pi$ . For generic values of the parameters in  $\Pi$ ,  $\mathbf{B}(L)$  is zeroless so that  $\mathbf{y}_t$  has a finite VAR representation.

For statement (I) see Anderson and Deistler (2008a), Theorem 3. Statement (II) is a modified version of their Theorem 2, see for a proof Forni et al. (2009), p. 1327.

### 2.2. Fundamentalness

Assume that the  $r$ -dimensional vector  $\mathbf{y}_t$  has an ARMA structure, rank  $q$  and the moving average representation (5). If  $\text{rank}(\mathbf{B}(z)) = q$  for  $|z| < 1$ , then  $\mathbf{u}_t$  belongs to the space spanned by  $\mathbf{y}_{t-k}$ , with  $k \geq 0$ , and representation (5), as well as  $\mathbf{u}_t$ , is called fundamental (for these definitions and results see e.g., Rozanov (1967), pp. 43–7). Note that if (5) is fundamental  $\text{rank}(\mathbf{B}(0)) = q$ . Note also that when  $q = r$ , the condition that  $\text{rank}(\mathbf{B}(z)) = q$  for  $|z| < 1$  becomes  $\det(\mathbf{B}(z)) \neq 0$  for  $|z| < 1$ .

**Remark 2.** Note that in Proposition 1, part (II), we do not assume that  $\mathbf{u}_t$  is fundamental for  $\mathbf{y}_t$ . However, Proposition 1, (II), states that for generic values of  $\mathbf{p} \in \Pi$  the matrix  $\mathbf{B}(L)$  is zeroless and therefore  $\mathbf{u}_t$  is fundamental for  $\mathbf{y}_t$ .

### 2.3. I(1) Singular Vectors

To analyze cointegration and the autoregressive representations of singular non-stationary vectors let us first recall the definitions of  $I(0)$ ,  $I(1)$  and cointegrated vectors. This requires some preliminary definitions and results.

We denote by  $L_2(\Omega, \mathcal{F}, P)$  the space of the square-integrable functions on the probability space  $(\Omega, \mathcal{F}, P)$ . Let  $\mathbf{z}_t = (z_{1t} \ z_{2t} \ \dots \ z_{rt})'$ ,  $z_{ht} \in L_2(\Omega, \mathcal{F}, P)$ , be an  $r$ -dimensional stochastic process and consider the difference equation

$$(1 - L)\zeta_t = \mathbf{z}_t, \tag{6}$$

in the unknown  $r$ -dimensional process  $\zeta_t$ . A solution of (6) is

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

see e.g., Gregoir (1999), p. 439, Franchi and Paruolo (2019). All the solutions of (6) are  $\psi_t = \tilde{\psi}_t + \phi_t$ , where  $\phi_t = (\phi_{1t} \ \phi_{2t} \ \dots \ \phi_{rt})'$ ,  $\phi_{ht} \in L_2(\Omega, \mathcal{F}, P)$ , is a solution of the homogeneous equation  $(1 - L)\zeta_t = \mathbf{0}$ , so that  $\phi_t = \mathbf{K}$ , for some  $r$ -dimensional stochastic vector  $\mathbf{K}$ , for all  $t \in \mathbb{Z}$ . We say that the process  $\phi_t = \mathbf{K}$  is a constant stochastic process. Obviously a constant stochastic process  $\phi_t = \mathbf{K}$  is weakly stationary. Its spectral measure has the jump  $\Sigma_K$  at frequency zero. Thus  $\phi_t$  has a spectral density (has an absolutely continuous spectral measure) if and only if  $\Sigma_K = \mathbf{0}$ , i.e., if and only if  $\phi_t(\omega) = \mathbf{k}$ , where  $\mathbf{k} \in \mathbb{R}^r$ , for  $\omega$  almost everywhere in  $\Omega$ .

#### Definition 4. (I(0), I(1) and Cointegrated vectors)

**I(0).** An  $r$ -dimensional ARMA  $\mathbf{y}_t$  with spectral density  $\Sigma_y(\theta)$  is  $I(0)$  if  $\Sigma_y(0) \neq \mathbf{0}$ .

**I(1).** The  $r$ -dimensional vector stochastic process  $\mathbf{y}_t$  is  $I(1)$  if it is a solution  $(1 - L)\zeta_t = \mathbf{z}_t$  where  $\mathbf{z}_t$  is an  $r$ -dimensional  $I(0)$  process. The rank of  $\mathbf{y}_t$  is defined as the rank of  $\mathbf{z}_t$ .

**Cointegration.** Assume that the  $r$ -dimensional stochastic vector  $\mathbf{y}_t$  is  $I(1)$  and denote by  $\Sigma_{\Delta y}(\theta)$  the spectral density of  $(1 - L)\mathbf{y}_t$ . The vector  $\mathbf{y}_t$  is cointegrated with cointegrating rank  $c$ , with  $0 < c < r$ , if  $\text{rank}(\Sigma_{\Delta y}(0)) = r - c$ .

If  $q$  is the rank of  $\mathbf{y}_t$  and  $r \geq q$ , then  $c = r - q + d$ , where  $q > d > 0$ . Thus in the singular case,  $r > q$ ,  $\mathbf{y}_t$  is necessarily cointegrated with cointegrating rank at least equal to  $r - q$ .

If  $\mathbf{y}_t$  is  $I(1)$  and cointegrated with cointegrating rank  $c$ , there exist  $c$  linearly independent  $r \times 1$  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 cointegrating vectors and the set  $\mathbf{c}_j, j = 1, \dots, c$ , a complete set of cointegrating vectors. Of course a complete set of cointegrating 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  $\mathbf{c}_j$ .

**Lemma 1.** (I) Assume that  $\mathbf{y}_t$  has an ARMA structure and has the rational representation (5):  $\mathbf{y}_t = \mathbf{V}(L)\mathbf{u}_t$ . Then  $\mathbf{y}_t$  is  $I(0)$  if and only if  $\mathbf{V}(1) \neq \mathbf{0}$ .

(II) Assume  $(1 - L)\mathbf{y}_t$  has an ARMA structure and has the rational representation

$$(1 - L)\mathbf{y}_t = \mathbf{V}(L)\mathbf{u}_t. \tag{7}$$

The process  $\mathbf{y}_t$  is  $I(1)$  if and only if  $\mathbf{V}(1) \neq \mathbf{0}$ .

(III) If  $\mathbf{y}_t$  is  $I(1)$ , cointegrated and has representation (7), the cointegrating rank of  $\mathbf{y}_t$  is  $c$  if and only if the rank of  $\mathbf{V}(1)$  is  $r - c$ . Moreover  $\mathbf{c}$  is a cointegrating vector for  $\mathbf{y}_t$  if and only if  $\mathbf{c}'\mathbf{V}(1) = \mathbf{0}$ .

(IV) Assume that  $\mathbf{y}_t$  is  $I(1)$ .  $\mathbf{c}$  is a cointegrating vector for  $\mathbf{y}_t$  if and only if a scalar stochastic variable  $w \in L_2(\Omega, \mathcal{F}, P)$  can be determined such that  $\mathbf{c}'\mathbf{y}_t - w$  is stationary with an ARMA structure.

**Proof.** (I) is an immediate consequence of  $\Sigma_y(0) = (2\pi)^{-1}\mathbf{V}(1)\Gamma_u\mathbf{V}(1)'$ , where  $\Gamma_u$  is the nonsingular covariance matrix of  $\mathbf{u}_t$ . (II) and (III) are obtained in the same way from  $\Sigma_{\Delta y}(0) = (2\pi)^{-1}\mathbf{V}(1)\Gamma_u\mathbf{V}(1)'$ . (IV) The process  $\mathbf{y}_t$  solves (6) with  $\mathbf{z}_t = \mathbf{V}(L)\mathbf{u}_t$ , so that, defining

$$\mu_t = \begin{cases} \mathbf{u}_1 + \mathbf{u}_2 + \dots + \mathbf{u}_t, & \text{for } t > 0 \\ \mathbf{0}, & \text{for } t = 0 \\ -(\mathbf{u}_0 + \mathbf{u}_{-1} \dots + \mathbf{u}_{t+1}), & \text{for } t < 0, \end{cases} \tag{8}$$

we have

$$\mathbf{y}_t = \mathbf{V}(L)\mu_t + \mathbf{K} = \left[ \mathbf{V}(1) + (1 - L) \frac{\mathbf{V}(L) - \mathbf{V}(1)}{1 - L} \right] \mu_t + \mathbf{K} = \mathbf{V}(1)\mu_t + \mathbf{V}^*(L)\mathbf{u}_t + \mathbf{K},$$

where (i) 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 unity, (ii)  $\mathbf{K}$  is a constant  $r$ -dimensional stochastic process. We have:

$$\mathbf{c}'\mathbf{y}_t = \mathbf{c}'\mathbf{V}(1)\mu_t + \mathbf{c}'\mathbf{V}^*(L)\mathbf{u}_t + \mathbf{c}'\mathbf{K}. \tag{9}$$

If  $\mathbf{c}$  is a cointegrating vector of  $\mathbf{y}_t$  we have  $\mathbf{c}'\mathbf{V}(1) = \mathbf{0}$ , so that

$$\mathbf{c}'\mathbf{y}_t = \mathbf{c}'\mathbf{V}^*(L)\mathbf{u}_t + \mathbf{c}'\mathbf{K}.$$

Setting  $w = \mathbf{c}'\mathbf{K}$ , the process  $\mathbf{c}'\mathbf{y}_t - w = \mathbf{c}'\mathbf{V}^*(L)\mathbf{u}_t$  has the desired properties. Note that  $w$  has the equivalent definition  $w = \mathbf{c}'\mathbf{y}_0 - \mathbf{c}'\mathbf{V}^*(L)\mathbf{u}_0$ . Conversely, suppose that  $w$  is such that  $\mathbf{c}'\mathbf{y}_t - w$  has an ARMA structure. By (9),

$$\mathbf{c}'\mathbf{y}_t - w = \mathbf{c}'\mathbf{V}(1)\mu_t + \mathbf{c}'\mathbf{V}^*(L)\mathbf{u}_t + \mathbf{c}'\mathbf{K} - w,$$



so that

$$\sqrt{E(\mathbf{c}'\mathbf{y}_t - w)^2} + \sqrt{E(\mathbf{c}'\mathbf{V}^*(L)\mathbf{u}_t)^2} + \sqrt{E(\mathbf{c}'\mathbf{K} - w)^2} \geq \sqrt{\mathbf{c}'\mathbf{V}(1)\boldsymbol{\Sigma}_{\mu_t}\mathbf{V}'(1)\mathbf{c}}.$$

The three terms on the left-hand side are finite and independent of  $t$ . As  $\boldsymbol{\Sigma}_{\mu_t} = |t|\boldsymbol{\Sigma}_u$  and  $\boldsymbol{\Sigma}_u$  is positive definite, the right-hand side diverges for  $|t| \rightarrow \infty$  unless  $\mathbf{c}'\mathbf{V}(1) = \mathbf{0}$ .  $\square$

Lemma 1 shows that 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: (i) our assumption of rational spectral density, (ii) the time span of the stochastic processes is  $t = 0, 1, \dots$  in Johansen’s book,  $t \in \mathbb{Z}$  in the present paper. Also, under the assumption that  $(1 - L)\mathbf{y}_t$  has an ARMA structure, our definition of cointegration is equivalent to that in Johansen (1995), p. 37.

### 3. Representation Theory for Singular I(1) Vectors

In Section 3.1 we prove our generalization to singular vectors of the Granger representation theorem (from MA to AR). We closely follow the proof in Johansen (1995), Theorem 4.5, p. 55–57. In Section 3.2 we show that, under a suitable parameterization, the matrix of the autoregressive representation is generically of finite degree.

#### 3.1. The Granger Representation Theorem (MA to AR)

Suppose that  $r \geq q, c > 0$  and  $r > c \geq r - q$ . Let  $\mathbf{B}(L)$  be an  $r \times q$  polynomial matrix of degree  $s_1 \geq 0$  and  $\mathbf{S}(L)$  an  $r \times r$  polynomial matrix of degree  $s_2 \geq 0$  with  $\mathbf{S}(0) = \mathbf{I}_r$ .

**Assumption 1.**  $\mathbf{S}(L)$  is stable.

**Assumption 2.** If  $z^*$  is a zero of  $\mathbf{B}(z)$  (i.e.  $\text{rank}(\mathbf{B}(z^*)) < q$ ) then either  $z^* = 1$  or  $|z^*| > 1$ .

Assumption 2 implies that the rank of  $\mathbf{B}(0)$  is  $q$ . The next is a stronger version of Assumption 2:

**Assumption 3.** If  $z^*$  is a zero of  $\mathbf{B}(z)$  then  $z^* = 1$ .

**Assumption 4.**  $\text{rank}(\mathbf{B}(1)) = r - c$ .

Under Assumption 1, let  $\mathbf{y}_t$  be a solution of the equation

$$(1 - L)\boldsymbol{\zeta}_t = \mathbf{S}(L)^{-1}\mathbf{B}(L)\mathbf{u}_t. \tag{10}$$

We have

$$\mathbf{y}_t = \mathbf{S}(L)^{-1}\mathbf{B}(L)\boldsymbol{\mu}_t + \mathbf{K}, \tag{11}$$

where  $\boldsymbol{\mu}_t$  is defined in (8) and  $\mathbf{K}$  is a constant stochastic process. By Assumption 4,  $\mathbf{S}(1)^{-1}\mathbf{B}(1) \neq \mathbf{0}$ , so that  $\mathbf{y}_t$  is  $I(1)$  with cointegrating rank  $c$ , see Lemma 1, (II) and (III).

Consider the finite Taylor expansion of  $\mathbf{B}(z)$  around  $z = 1$ :

$$\mathbf{B}(z) = \mathbf{B}(1) - (1 - z)\mathbf{B}'(1) + (1 - z)^2\mathbf{B}''(1) + \dots .$$

Assumption 4 implies that

$$\mathbf{B}(1) = \boldsymbol{\zeta}\boldsymbol{\eta}',$$

where  $\boldsymbol{\zeta}$  is  $r \times (r - c)$  of rank  $r - c$ ,  $\boldsymbol{\eta}$  is  $q \times (r - c)$  of rank  $r - c$ , see Lancaster and Tismenetsky (1985, p. 97, Proposition 3). The Taylor expansion above can be rewritten as

$$\mathbf{B}(z) = \boldsymbol{\zeta}\boldsymbol{\eta}' + (1 - z)\mathbf{B}^* + (1 - z)^2\mathbf{E}(z), \tag{12}$$

where  $\mathbf{B}^* = -\mathbf{B}'(1)$  and  $\mathbf{E}(z)$  is a polynomial matrix.

Let  $\boldsymbol{\zeta}_\perp$  be an  $r \times c$  matrix whose columns are orthogonal to all columns of  $\boldsymbol{\zeta}$ : (i) the columns of  $\boldsymbol{\zeta}_\perp$  are a complete set of cointegrating vectors for  $\mathbf{B}(L)\mathbf{u}_t$ , (ii) the columns of the matrix  $\mathbf{S}'(1)\boldsymbol{\zeta}_\perp$  are a complete set of cointegrating vectors for  $\mathbf{y}_t$ . Regarding (i), using (11) and (12), we have

$$\boldsymbol{\zeta}'_\perp \mathbf{S}(L)\mathbf{y}_t = \boldsymbol{\zeta}'_\perp \mathbf{B}(L)\boldsymbol{\mu}_t + \boldsymbol{\zeta}'_\perp \mathbf{S}(1)\mathbf{K} = (\boldsymbol{\zeta}'_\perp \mathbf{B}^* + (1-L)\boldsymbol{\zeta}'_\perp \mathbf{E}(L))\mathbf{u}_t + \boldsymbol{\zeta}'_\perp \mathbf{S}(1)\mathbf{K}, \tag{13}$$

so that  $\boldsymbol{\zeta}'_\perp \mathbf{S}(L)\mathbf{y}_t - \boldsymbol{\zeta}'_\perp \mathbf{S}(1)\mathbf{K}$  has an ARMA structure. Regarding (ii), see the proof of Proposition 2.

**Assumption 5.**  $\text{rank} \begin{bmatrix} \boldsymbol{\zeta}'_\perp \mathbf{B}^* \\ \boldsymbol{\eta}' \end{bmatrix} = \text{rank} \begin{bmatrix} \boldsymbol{\zeta}'_\perp \mathbf{B}^* \\ \boldsymbol{\zeta}' \boldsymbol{\zeta}' \boldsymbol{\eta}' \end{bmatrix} = q.$

Define  $\mathbf{S}^*(L) = \frac{\mathbf{S}(L) - \mathbf{S}(1)}{1-L}.$

**Assumption 6.**  $\boldsymbol{\zeta}'_\perp (\mathbf{B}^* - \mathbf{S}^*(1)\mathbf{S}(1)^{-1}\boldsymbol{\zeta}'\boldsymbol{\eta}') \neq \mathbf{0}.$

**Remark 3.** Let  $\mathbf{y}_t$  be a solution of (10) so that  $(1-L)\mathbf{y}_t$  is stationary and  $\mathbf{S}(L)[(1-L)\mathbf{y}_t] = \mathbf{B}(L)\mathbf{u}_t$ . Assumption 2, and therefore 3, implies that  $\mathbf{u}_t$  is fundamental for  $(1-L)\mathbf{y}_t$ , see Section 2.2.

We are now ready for our main representation result.

**Proposition 2.** (I) Weak form. Suppose that Assumptions 1, 2, 4, 5 and 6 hold and let  $\mathbf{y}_t$  be a solution of the difference Equation (10), so that  $\mathbf{y}_t = \mathbf{S}(L)^{-1}\mathbf{B}(L)\boldsymbol{\mu}_t + \mathbf{K}$ , with  $\boldsymbol{\mu}_t$  defined in (8) and  $\mathbf{K}$  a constant stochastic process. Set  $\boldsymbol{\beta} = \mathbf{S}(1)'\boldsymbol{\zeta}_\perp$ . Then a  $c$ -dimensional stochastic vector  $\mathbf{w}$  can be determined such that (i)  $\boldsymbol{\beta}'\mathbf{y}_t - \mathbf{w}$  is  $I(0)$ , (ii)  $\mathbf{y}_t$  has the error correction representation

$$\mathbf{A}(L)\mathbf{y}_t = \mathbf{A}^*(L)(1-L)\mathbf{y}_t + \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{y}_{t-1} - \mathbf{w}) = \mathbf{B}(0)\mathbf{u}_t, \tag{14}$$

where  $\mathbf{A}(L)$  is a rational  $r \times r$  matrix with no poles in or on the unit circle,  $\mathbf{A}(1) = \mathbf{I}_r$ ,  $\mathbf{A}^*(L) = (\mathbf{A}(L) - \mathbf{A}(1)L)(1-L)^{-1}$ ,  $\boldsymbol{\alpha}$  is  $r \times c$  and full rank,  $\boldsymbol{\alpha}\boldsymbol{\beta}' = \mathbf{A}(1)$ .

(II) Strong form. Under Assumptions 1, 3, 4, 5 and 6, statement (I) holds with an  $r \times r$  stable, finite-degree matrix polynomial  $\mathbf{A}(L)$ .

**Proof.** Multiply both sides of  $(1-L)\mathbf{S}(L)\mathbf{y}_t = \mathbf{B}(L)\mathbf{u}_t$  by the  $r \times r$  invertible matrix  $\boldsymbol{\Xi} = \begin{pmatrix} \boldsymbol{\zeta}'_\perp \\ \boldsymbol{\zeta}' \end{pmatrix}.$

We obtain

$$\begin{aligned} (1-L)\boldsymbol{\Xi}\mathbf{S}(L)\mathbf{y}_t &= \boldsymbol{\Xi}\mathbf{B}(L)\mathbf{u}_t \\ &= \left\{ \begin{pmatrix} \mathbf{0}_{c \times q} \\ \boldsymbol{\zeta}' \boldsymbol{\zeta}' \boldsymbol{\eta}' \end{pmatrix} + (1-L) \begin{pmatrix} \boldsymbol{\zeta}'_\perp \mathbf{B}^* \\ \boldsymbol{\zeta}' \mathbf{B}^* \end{pmatrix} + (1-L)^2 \begin{pmatrix} \boldsymbol{\zeta}'_\perp \mathbf{E}(L) \\ \boldsymbol{\zeta}' \mathbf{E}(L) \end{pmatrix} \right\} \mathbf{u}_t \\ &= \begin{pmatrix} (1-L)\mathbf{I}_c & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{r-c} \end{pmatrix} \left\{ \begin{pmatrix} \boldsymbol{\zeta}'_\perp \mathbf{B}^* \\ \boldsymbol{\zeta}' \boldsymbol{\zeta}' \boldsymbol{\eta}' \end{pmatrix} + (1-L) \begin{pmatrix} \boldsymbol{\zeta}'_\perp \mathbf{E}(L) \\ \boldsymbol{\zeta}' \mathbf{B}^* \end{pmatrix} + (1-L)^2 \begin{pmatrix} \mathbf{0}_{c \times q} \\ \boldsymbol{\zeta}' \mathbf{E}(L) \end{pmatrix} \right\} \mathbf{u}_t. \end{aligned} \tag{15}$$

Taking the first  $c$  rows in (15),

$$(1-L)\boldsymbol{\zeta}'_\perp \mathbf{S}(L)\mathbf{y}_t = (1-L) (\boldsymbol{\zeta}'_\perp \mathbf{B}^* + (1-L)\boldsymbol{\zeta}'_\perp \mathbf{E}(L)) \mathbf{u}_t.$$

This implies that

$$\boldsymbol{\zeta}'_\perp \mathbf{S}(L)\mathbf{y}_t = (\boldsymbol{\zeta}'_\perp \mathbf{B}^* + (1-L)\boldsymbol{\zeta}'_\perp \mathbf{E}(L)) \mathbf{u}_t + \mathbf{w}, \tag{16}$$



where  $\mathbf{w}$  is a  $c$ -dimensional constant stochastic vector. Comparing with (13),  $\mathbf{w} = \boldsymbol{\zeta}'_{\perp} \mathbf{S}(1)\mathbf{K}$ . On the other hand,

$$\begin{aligned} \boldsymbol{\zeta}'_{\perp} \mathbf{S}(1)\mathbf{y}_t - \mathbf{w} &= (\boldsymbol{\zeta}'_{\perp} \mathbf{S}(L)\mathbf{y}_t - \mathbf{w}) - (\boldsymbol{\zeta}'_{\perp} \mathbf{S}(L)\mathbf{y}_t - \boldsymbol{\zeta}'_{\perp} \mathbf{S}(1)\mathbf{y}_t) \\ &= (\boldsymbol{\zeta}'_{\perp} \mathbf{S}(L)\mathbf{y}_t - \mathbf{w}) - \boldsymbol{\zeta}'_{\perp} \mathbf{S}^*(L)(1-L)\mathbf{y}_t \\ &= (\boldsymbol{\zeta}'_{\perp} \mathbf{S}(L)\mathbf{y}_t - \mathbf{w}) - \boldsymbol{\zeta}'_{\perp} \mathbf{S}^*(L)\mathbf{S}(L)^{-1}\mathbf{B}(L)\mathbf{u}_t \\ &= \left\{ \boldsymbol{\zeta}'_{\perp} (\mathbf{B}^* - \mathbf{S}^*(1)\mathbf{S}(1)^{-1}\boldsymbol{\zeta}\boldsymbol{\eta}') + (1-L)\mathcal{H}(L) \right\} \mathbf{u}_t, \end{aligned} \tag{17}$$

where the last equality has been obtained using (16) and  $\mathcal{H}(L)$  is a suitable polynomial matrix. Thus  $\boldsymbol{\beta}'\mathbf{y}_t - \mathbf{w} = \boldsymbol{\zeta}'_{\perp} \mathbf{S}(1)\mathbf{y}_t - \mathbf{w}$  has an ARMA structure. Moreover, by Assumption 6,  $\boldsymbol{\beta}'\mathbf{y}_t - \mathbf{w}$  is  $I(0)$ .

Joining (16) with the last  $r - c$  rows of (15),

$$\begin{pmatrix} \mathbf{I}_c & \mathbf{0} \\ \mathbf{0} & (1-L)\mathbf{I}_{r-c} \end{pmatrix} \boldsymbol{\Xi} \mathbf{S}(L)\mathbf{y}_t - \begin{pmatrix} \mathbf{I}_c \\ \mathbf{0}_{(r-c) \times c} \end{pmatrix} \mathbf{w} = \mathbf{M}(L)\mathbf{u}_t, \tag{18}$$

where

$$\mathbf{M}(L) = \begin{pmatrix} \boldsymbol{\zeta}'_{\perp} \mathbf{B}^* \\ \boldsymbol{\zeta}' \boldsymbol{\zeta} \boldsymbol{\eta}' \end{pmatrix} + (1-L) \begin{pmatrix} \boldsymbol{\zeta}'_{\perp} \boldsymbol{\Xi} \mathbf{E}(L) \\ \boldsymbol{\zeta}' \mathbf{B}^* \end{pmatrix} + (1-L)^2 \begin{pmatrix} \mathbf{0}_{c \times q} \\ \boldsymbol{\zeta}' \boldsymbol{\Xi} \mathbf{E}(L) \end{pmatrix}. \tag{19}$$

By (15) and (19),

$$\mathbf{B}(L) = \boldsymbol{\Xi}^{-1} \begin{pmatrix} (1-L)\mathbf{I}_c & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{r-c} \end{pmatrix} \mathbf{M}(L).^3 \tag{20}$$

By Assumption 5,  $\mathbf{M}(z)$  has no zero at  $z = 1$ , see (19). On the other hand, (i) if  $z^*$  is a zero of  $\mathbf{M}(z)$  then  $z^*$  is a zero of  $\mathbf{B}(z)$ , (ii) if  $z^*$  is a zero of  $\mathbf{B}(z)$ ,  $z^* \neq 1$ , then  $z^*$  is a zero of  $\mathbf{M}(z)$ . Therefore, Assumption 3 implies that  $\mathbf{M}(z)$  is zeroless and viceversa. Under Assumption 2, the zeros of  $\mathbf{M}(z)$  lie outside the unit circle. In order to conclude the proof we need inverting  $\mathbf{M}(L)$  in (18).

(I) Under Assumption 3, Proposition 1, part (I), states that there exists an  $r \times r$  stable, finite-degree polynomial matrix  $\mathbf{N}(L) = \mathbf{I}_r + \mathbf{N}_1 L + \dots + \mathbf{N}_p L^p$ , for some  $p$ , such that: (i)  $\mathbf{N}(0) = \mathbf{I}_r$ , (ii)  $\mathbf{N}(L)\mathbf{M}(L) = \mathbf{M}(0)$ .

(II) Under Assumption 2, by a standard procedure we remove all the zeros of  $\mathbf{M}(z)$  which lie outside the unit circle<sup>4</sup>, then use Proposition 1, part (I), to left-invert the residual zeroless polynomial, thus obtaining an  $r \times r$  rational matrix  $\mathbf{N}(L)$  such that (i)  $\mathbf{N}(L)$  has no poles in or on the unit circle (possible poles of  $\mathbf{N}(L)$  are the zeros of  $\mathbf{M}(L)$ , which lie outside the unit circle), (ii)  $\mathbf{N}(0) = \mathbf{I}_r$ , (iii)  $\mathbf{N}(L)\mathbf{M}(L) = \mathbf{M}(0)$ . See also Deistler et al. (2010).

Defining

$$\mathbf{A}(L) = \boldsymbol{\Xi}^{-1} \mathbf{N}(L) \begin{pmatrix} \mathbf{I}_c & \mathbf{0} \\ \mathbf{0} & (1-L)\mathbf{I}_{r-c} \end{pmatrix} \boldsymbol{\Xi} \mathbf{S}(L) = \boldsymbol{\Xi}^{-1} \mathbf{N}(L) \begin{pmatrix} \boldsymbol{\zeta}'_{\perp} \\ (1-L)\boldsymbol{\zeta}' \end{pmatrix} \mathbf{S}(L)$$

and using  $\mathbf{M}(0) = \boldsymbol{\Xi} \mathbf{B}(0)$ , we have

$$\mathbf{A}(L)\mathbf{y}_t - \boldsymbol{\Xi}^{-1} \mathbf{N}(1) \begin{pmatrix} \mathbf{I}_c \\ \mathbf{0}_{(r-c) \times c} \end{pmatrix} \mathbf{w} = \mathbf{B}(0)\mathbf{u}_t,$$

<sup>3</sup> In the square case,  $r = q$ , Assumption 3 holds if and only if  $\mathbf{M}(z)$  is unimodular.

<sup>4</sup> If  $z^*$  is a zero of  $\mathbf{M}(z)$ , multiply  $\mathbf{M}(z)$  by an invertible  $r \times r$  matrix  $\mathbf{Q}_{z^*}$  such that  $z^*$  is a zero of, say, the first row of  $\mathbf{Q}_{z^*} \mathbf{M}(z)$ . Then multiply by the  $r \times r$  diagonal matrix with  $(z - z^*)^{-1}$  in position (1, 1) and unity elsewhere on the main diagonal. Iterating, all the zeros of  $\mathbf{M}(z)$  are removed.

with  $\mathbf{A}(0) = \mathbf{I}_r$ . Defining  $\mathbf{A}^*(L) = (\mathbf{A}(L) - \mathbf{A}(1)L)(1 - L)^{-1}$ ,

$$\mathbf{A}^*(L)(1 - L)\mathbf{y}_t + \mathbf{A}(1)\mathbf{y}_{t-1} - \Xi^{-1}\mathbf{N}(1) \begin{pmatrix} \mathbf{I}_c \\ \mathbf{0}_{(r-c) \times c} \end{pmatrix} \mathbf{w} = \mathbf{B}(0)\mathbf{u}_t.$$

Defining

$$\boldsymbol{\alpha} = \Xi^{-1}\mathbf{N}(1) \begin{pmatrix} \mathbf{I}_c \\ \mathbf{0}_{(r-c) \times c} \end{pmatrix},$$

we see that  $\mathbf{A}(1) = \boldsymbol{\alpha}\boldsymbol{\beta}'$  and

$$\mathbf{A}^*(L)(1 - L)\mathbf{y}_t + \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{y}_{t-1} - \mathbf{w}) = \mathbf{B}(0)\mathbf{u}_t.$$

□

Some remarks are in order.

**Remark 4.** (I) Under our assumption of an ARMA structure, Assumption 1 corresponds to Definition 3.1 in Johansen’s book, see p. 34. Assumption 2 is Johansen’s Assumption 1 (see p. 14), adapted for singularity. Assumption 3 has no counterpart in Johansen’s nonsingular framework. In Section 3.2 we show that under the parameterization adopted in Definition 5, Assumption 3 holds generically.

(II) Simplifying the model by taking  $\mathbf{S}(L) = \mathbf{I}_r$ , Assumption 5 generalizes to the singular case Johansen’s assumption that  $\boldsymbol{\zeta}'_{\perp}\mathbf{C}^*\boldsymbol{\eta}_{\perp}$  is full rank (see Theorem 4.5, p. 55;  $\mathbf{C}^*$  corresponds to our  $\mathbf{B}^*$ ). For, assuming that  $r = q$ , multiplying the matrix in Assumption 5 by the nonsingular matrix  $\begin{pmatrix} \boldsymbol{\eta}_{\perp} & \boldsymbol{\eta} \end{pmatrix}$ , we obtain that Assumption 5 holds if and only if  $\boldsymbol{\zeta}'_{\perp}\mathbf{B}^*\boldsymbol{\eta}_{\perp}$  is full rank. Assumption 5 is used in the proof of Proposition 2 to invert the matrix  $\mathbf{M}(L)$ , which remains on the right-hand side after the removal of the unit roots, see Equation (18), which is the same rôle played by Johansen’s assumption in his proof.

(III) Under  $\mathbf{S}(L) = \mathbf{I}_r$ , assumption 6 simplifies to  $\boldsymbol{\zeta}'_{\perp}\mathbf{B}^* \neq \mathbf{0}$ . If  $d > 0$  Assumption 6 is a consequence of Assumption 5. For, if  $d > 0$  then  $r - c = q - d < q$ . On the other hand,  $r - c$  is the number of rows of  $\boldsymbol{\eta}'$ , so that Assumption 5 holds only if Assumption 6 holds. In particular, if  $r = q$  and  $c = d > 0$ , Assumption 6 is redundant. However if  $r > q$  and  $d = 0$ , so that the rank of  $\boldsymbol{\eta}'$  is  $q$ , then Assumption 5 holds even if  $\boldsymbol{\zeta}'_{\perp}\mathbf{B}^* = \mathbf{0}$ . Assumption 6 is necessary in Proposition 2 to prove that the error correction term is  $I(0)$ , not only stationary.

**Remark 5.** Uniqueness issues arise with autoregressive representations of singular vectors. For example, suppose that  $c = r - q > 0$ , so that  $d = 0$ . Representation (14) has an  $(r - q)$ -dimensional error correction term  $\boldsymbol{\beta}'\mathbf{y}_t - \mathbf{w}$ . On the other hand, in this case  $\mathbf{B}(1)$  has full rank  $q$ , so that Proposition 1 (I) applies and, in spite of cointegration,  $\mathbf{y}_t$  has an autoregressive representation in differences

$$\mathbf{D}(L)\mathbf{S}(L)(1 - L)\mathbf{y}_t = \mathbf{B}(0)\mathbf{u}_t.$$

In Appendix B.1 we sketch a proof of the statement that in general,  $\mathbf{y}_t$  has VECM representations with a number of error correction terms ranging from  $d$  to  $c$ . However, as we show in Appendix B.2, different autoregressive representations of  $\mathbf{y}_t$  produce the same impulse-response functions. Both in this and the companion paper Barigozzi et al. (2019) the number of error correction terms in the error correction representation for reduced-rank  $I(1)$  vectors is always the maximum  $c$ . It is worth reporting that, in our experiments with simulated data, the best results in estimation of singular VECMs are obtained using  $c$  as the number of error correction terms.

**Remark 6.** Assume for simplicity that  $\mathbf{S}(L) = \mathbf{I}_r$ . From equation (17):

$$\mathbf{e}_t = \boldsymbol{\beta}'\mathbf{y}_t - \mathbf{w} = \boldsymbol{\zeta}'_{\perp}\mathbf{y}_t - \mathbf{w} = \{\boldsymbol{\zeta}'_{\perp}\mathbf{B}^* + (1 - L)\mathcal{H}(L)\} \mathbf{u}_t.$$

If  $r = q$ , Assumption 5 implies that  $\zeta'_{\perp} \mathbf{B}^*$  has rank  $c$ , so that no  $c$ -dimensional vector  $\mathbf{d} \neq \mathbf{0}$  can be determined such that some of the coordinates of  $\mathbf{d}\mathbf{e}_t$  is stationary but not  $I(0)$ . Thus, according to the definition introduced in Franchi and Paruolo (2019), p. 1181, the error term  $\mathbf{e}_t$  is a “non-cointegrated  $I(0)$  process.” When  $r > q$  and  $c \leq q$ , i.e.,  $r \leq 2q - d$ , elementary examples can be produced in which  $\mathbf{e}_t$  is an  $I(0)$  but not a non-cointegrated  $I(0)$  process (one is given in Appendix A.2). Thus Assumption 6 only implies that  $\mathbf{e}_t$  is  $I(0)$ . Of course, under  $c \leq q$ , the assumption that  $\zeta'_{\perp} (\mathbf{B}^* - \mathbf{S}^*(1)\mathbf{S}(1)^{-1}\zeta\eta')$  has rank  $c$ , an enhancement of Assumption 6, implies that  $\mathbf{e}_t$  is a non-cointegrated  $I(0)$  process. On the other hand, if  $c > q$ , i.e.,  $r > 2q - d$ ,  $\mathbf{e}_t$  cannot be a non-cointegrated  $I(0)$  process.

### 3.2. Generically, $\mathbf{A}(L)$ Is a Finite-Degree Polynomial

Suppose that the couple  $(\mathbf{S}(L), \mathbf{B}(L))$  is parameterized as in Definition 3. It is easy to see that  $\mathbf{B}(1)$  has generically rank  $q$ , so that generically the cointegrating rank of  $\mathbf{y}_t$  is  $r - q$ . In particular, if  $r = q$  cointegration is non-generic.

It is quite easy to see that this paradoxical result only depends on the choice of a parameter set that is unfit to study cointegration. Our starting point here is that a specific value of  $c$  between  $r - q$  and  $r - 1$  has a motivation in economic theory or in statistical inference, and must be therefore built in the parameter set. Thus in Definition 5 below the family of filters is redefined so that generically the cointegrating rank is equal to a given  $c$  between  $r - q$  and  $r - 1$ .

**Definition 5. (Rational reduced-rank family of filters with cointegrating rank  $c$ )** Assume that  $r > q$ ,  $c > 0$  and  $r > c \geq r - q$ . Let  $\mathcal{G}$  be a set of couples  $(\mathbf{S}(L), \mathbf{B}(L))$ , where:

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

$$\mathbf{B}(L) = \zeta\eta' + (1 - L)\mathbf{B}^* + (1 - L)^2\mathbf{E}(L),$$

where  $\zeta$  and  $\eta$  are  $r \times (r - c)$  and  $q \times (r - c)$  respectively,  $\mathbf{B}^*$  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  $\lambda = (r - c)(r + q) + rq(s_1 + 2) + r^2s_2$  coefficients of the matrices  $\mathbf{S}(L)$ ,  $\zeta$ ,  $\eta$ ,  $\mathbf{B}^*$  and  $\mathbf{E}(L)$ , we assume that  $\mathbf{p} \in \Pi$ , where  $\Pi$  is an open subset of  $\mathbb{R}^{\lambda}$  such that for  $\mathbf{p} \in \Pi$ : (1)  $\mathbf{S}(z)$  is stable, (2)  $\text{rank}(\mathbf{B}(z)) = q$  with the exception of a finite subset of  $\mathbb{C}$ , (3)  $\text{rank}(\mathbf{B}(1)) = \text{rank}(\zeta\eta') = r - c$ .

We say that  $\mathcal{G}$  is a rational reduced-rank family of filters with cointegrating rank  $c$ .

**Proposition 3.** Assume that  $r > q$ . Let  $\mathbf{y}_t$  be a  $I(1)$  solution of Equation (10), where  $(\mathbf{S}(L), \mathbf{B}(L))$  belongs to a rational reduced-rank family of filters with cointegrating rank  $c$ . For generic values of the parameters in  $\Pi$ , Assumptions 1, 3, 4, 5 and 6 hold. Thus the Strong Form of Proposition 2 holds and  $\mathbf{y}_t$  has an error correction representation

$$\mathbf{A}(L)\mathbf{y}_t = \mathbf{A}^*(L)(1 - L)\mathbf{y}_t + \alpha(\beta'\mathbf{y}_{t-1} - \mathbf{w}) = \mathbf{B}(0)\mathbf{u}_t,$$

where  $\mathbf{A}(L)$  is a finite-degree polynomial matrix.

**Proof.** Part (iii) of Definition 5 implies that Assumptions 1 and 4 hold for all  $\mathbf{p} \in \Pi$ . The sets where Assumptions 5 and 6 do not hold are the intersections of the open set  $\Pi$  with the algebraic varieties

$$(a) \text{rank} \left[ \begin{pmatrix} \zeta'_{\perp} \mathbf{B}^* \\ \eta' \end{pmatrix} \right] < q, \quad (b) \zeta'_{\perp} (\mathbf{B}^* - \mathbf{S}^*(1)\mathbf{S}(1)^{-1}\zeta\eta') = \mathbf{0}$$

(the variety described by (a) is obtained by equating to zero the determinant of all the  $q \times q$  submatrices of the  $r \times q$  matrix between brackets). It is easy to see that the varieties (a) and (b) are not trivial,

i.e., that their dimension is lower than  $\lambda$ . Thus Assumptions 5 and 6 hold generically. The same result holds for Assumption 3. The points of  $\Pi$  where it is not fulfilled belong to a lower-dimensional algebraic variety. This is proved in A.1, see in particular Lemma A4.  $\square$

**Remark 7.** It is easy to see that, assuming that  $c \leq q$ ,  $\text{rank}(\xi'_{\perp}(\mathbf{B}^* - \mathbf{S}^*(1)\mathbf{S}(1)^{-1}\xi\eta')) = c$  holds generically in  $\Pi$ . Thus, in that case, the error term  $\beta\mathbf{y}_t - \mathbf{w}$  is generically a non-cointegrated  $I(0)$  process, see Remark 6.

**Remark 8.** A general comment on genericity results is in order. Theorems like Proposition 3 or Proposition 1, part (II), show that the subset where some statement does not hold belong to some algebraic variety of lower dimension (see the proof of Proposition 3 in particular), and is therefore negligible from a topological point of view. This suggests the working hypothesis that such subset is negligible from an economic or statistical point of view as well. If, for example, economic theory produces a singular vector  $\mathbf{y}_t$  with cointegration rank  $c$ , we may find reasonable to conclude that  $\mathbf{y}_t$  has representation (14) with a finite autoregressive polynomial. However, a greater degree of certainty is obtained by checking that the parameters of  $(\mathbf{S}(L), \mathbf{B}(L))$ , that are implicit in the theory, do not necessarily lie in one of the three algebraic varieties described in the proof of Proposition 3.

Definition 5 does not assume that  $\mathbf{B}(L)$  has no zeros inside the unit circle. Thus we have not assumed that  $\mathbf{u}_t$  is fundamental for  $(1 - L)\mathbf{F}_t$ , see Section 2.2. However, Proposition 3 shows that for generic values of the parameters in  $\Pi$ , the assumptions of Proposition 2, strong form, hold, Assumption 3 in particular, so that  $\mathbf{B}(L)$  has no zeros of non-unit modulus and therefore inside the unit circle. Thus:

**Proposition 4.** Assume that  $r > q$ . Let  $\mathbf{y}_t$  be a solution of Equation (10), where  $(\mathbf{S}(L), \mathbf{B}(L))$  belongs to a rational reduced-rank family of filters with cointegrating rank  $c$ . For generic values of the parameters in  $\Pi$ ,  $\mathbf{u}_t$  is fundamental for  $(1 - L)\mathbf{y}_t$ .

**Remark 9.** Note that Propositions 3 and 4 do not hold in the nonsingular case, where no genericity argument can be used to rule out non-unit zeros of  $\mathbf{B}(L)$ , either inside or outside the unit circle. In particular, fundamentalness of  $\mathbf{u}_t$  for  $(1 - L)\mathbf{y}_t$  is not generic if  $r = q$ .

### 3.3. Permanent and Transitory Shocks

Let  $\eta_{\perp}$  be a  $q \times d$  matrix whose columns are independent and orthogonal to the columns of  $\eta$ , and let

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

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

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

We have

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

where  $\mathbf{G}_1(L) = (\mathbf{B}^* + (1 - L)\mathbf{E}(L))\bar{\eta}_{\perp}$ , and  $\mathbf{G}_2(L) = (\mathbf{B}^* + (1 - L)\mathbf{E}(L))\bar{\eta}$ . All the solutions of the difference equation  $(1 - L)\mathbf{y}_t = \mathbf{S}(L)^{-1}\mathbf{C}(L)\mathbf{u}_t$  are

$$\mathbf{y}_t = \mathbf{S}(L)^{-1} [\mathbf{G}_1(L)\mathbf{v}_{1t} + \mathbf{G}_2(L)\mathbf{v}_{2t} + \mathbf{T}_t] + \mathbf{K}, \tag{21}$$

where  $\mathbf{K}$  is a constant stochastic process, and

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

As  $\boldsymbol{\zeta}$  is full rank, we see that  $\mathbf{y}_t$  is driven by the  $q - d = r - c$  permanent shocks  $\mathbf{v}_{2t}$ , and by the  $d$  temporary shocks  $\mathbf{v}_{1t}$ . In representation (21), 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 cointegrating rank, as usual. However, the number of transitory shocks is only  $d = c - (r - q)$ , as though  $r - q$  transitory shocks had a zero coefficient.

### 3.4. VECMs and Unrestricted VARs in The Levels

Several papers have addressed the issue of whether and when an error correction model or an unrestricted VAR in the levels should be used for estimation in the case of nonsingular cointegrated vectors: [Sims et al. \(1990\)](#) have shown that the parameters of a cointegrated VAR 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 evidence in favour of the VECM specification in the singular case.

- (I) We generate  $\mathbf{y}_t$  using a specification of (14) with  $r = 4, q = 3, d = 2$ , so that  $c = r - q + d = 3$ . The  $4 \times 4$  matrix  $\mathbf{A}(L)$  is of degree 2. The impulse-response functions are identified by assuming that the upper  $3 \times 3$  submatrix of  $\mathbf{B}(0)$  is lower triangular (see [Appendix C](#) for details). We replicate the generation of  $\mathbf{y}_t$  1000 times for  $T = 100, 500, 1000, 5000$ .
- (II) For each replication, we estimate a (misspecified) VAR in differences (DVAR), a VAR in the levels (LVAR) and a VECM, as in [Johansen \(1988 1991\)](#), assuming known  $c$ , the degree of  $\mathbf{A}(L)$  and that of  $\mathbf{A}^*(L)$ . For the VAR in differences the impulse-response functions for  $(1 - L)\mathbf{y}_t$  are cumulated to obtain impulse-response function for  $\mathbf{y}_t$ . The root mean square error between estimated and actual impulse-response functions is computed for each replication using all 12 impulse-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. The performance of the misspecified DVAR is uniformly poor with the exception of lag zero.

**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  $y_t$  to the shocks  $u_t$ . Estimation is carried out using three different autoregressive representations: a VAR for  $(1 - L)y_t$  (DVAR), a VAR for  $y_t$  (LVAR), and a VECM with  $c = r - q + d$  error terms (VECM). The results are based on 1000 replications. For the data generating process see Appendix C. The RMSEs are obtained averaging over all replications and all  $4 \times 3$  responses.

#### 4. Cointegration of the Observable Variables in a DFM

Consider again the factor model  $x_{it} = \chi_{it} + \epsilon_{it}$ , rewritten here as

$$x_t = \chi_t + \epsilon_t, \quad \chi_t = \Lambda F_t, \tag{22}$$

where  $\Lambda$  is  $n \times r$ , with  $n > r$ . The relationship between cointegration of the factors  $F_t$  and cointegration of the variables  $x_{it}$  is now considered.

Let us recall that the the common factors  $F_{jt}$  are assumed to be orthogonal to the idiosyncratic components  $\epsilon_{ks}$  for all  $i, j, t, s$ , i.e.,  $E\chi_t \epsilon_s' = \mathbf{0}_{n \times n}$ . for all  $t, s$ , see the Introduction. The other assumptions on model (22) are asymptotic, see e.g., Forni et al. (2000); Forni and Lippi (2001); (Stock and Watson 2002a, 2002b), and put no restriction on the matrix  $\Lambda$  and the vector  $\epsilon_t$  for a given finite  $n$ . In particular, the first  $r$  eigenvalues of the matrix  $\Lambda \Lambda'$  must diverge as  $n \rightarrow \infty$ , but this has no implications on the rank of the matrix  $\Lambda$  corresponding to, say,  $n = 10$ . However, as we see in Proposition 5 (iii), if the idiosyncratic components are  $I(0)$ , then, independently of  $\Lambda$ , all  $p$ -dimensional subvectors of  $x_t$  are cointegrated for  $p > q - d$ , which is at odds with what is observed in the macroeconomic datasets analyzed in the empirical Dynamic Factor Model literature. This motivates assuming that  $\epsilon_t$  is  $I(1)$ . In that case, see Proposition 5 (i), cointegration of  $x_t$  requires that both the common and the idiosyncratic components are cointegrated. Some results are collected in the statement below.

**Proposition 5.** Let  $x_t^{(p)} = \chi_t^{(p)} + \epsilon_t^{(p)} = \Lambda^{(p)} F_t + \epsilon_t^{(p)}$  be a  $p$ -dimensional subvector of  $x_t$ ,  $p \leq n$ . Denote by  $c_\chi^p$  and  $c_\epsilon^p$  the cointegrating rank of  $\chi_t^{(p)}$  and  $\epsilon_t^{(p)}$  respectively. Both range from  $p$ , stationarity, to 0, no cointegration.

- (i)  $x_t^{(p)}$  is cointegrated only if  $\chi_t^{(p)}$  and  $\epsilon_t^{(p)}$  are both cointegrated.
- (ii) If  $p > q - d$  then  $\chi_t^{(p)}$  is cointegrated. If  $p \leq q - d$  and  $\text{rank}(\Lambda^{(p)}) < p$  then  $\chi_t^{(p)}$  is cointegrated.
- (iii) Let  $V^\chi \subseteq \mathbb{R}^p$  and  $V^\epsilon \subseteq \mathbb{R}^p$  be the cointegrating spaces of  $\chi_t^{(p)}$  and  $\epsilon_t^{(p)}$  respectively. The vector  $x_t^{(p)}$  is cointegrated if and only if the intersection of  $V^\chi$  and  $V^\epsilon$  contains non-zero vectors. In particular, (a) if  $p > q - d$  and  $c^\epsilon > q - d$  then  $x^{(p)}$  is cointegrated, (b) if  $p > q - d$  and  $\epsilon_t^{(p)}$  is stationary then  $x^{(p)}$  is cointegrated.

**Proof.** Because  $\chi_{it}$  and  $\epsilon_{js}$  are orthogonal for all  $i, j, t, s$ , the spectral densities of  $(1 - L)x_t^{(p)}, (1 - L)\chi_t^{(p)}, (1 - L)\epsilon_t^{(p)}$  fulfill:

$$\Sigma_{\Delta x}^{(p)}(\theta) = \Sigma_{\Delta \chi}^{(p)}(\theta) + \Sigma_{\Delta \epsilon}^{(p)}(\theta) \quad \theta \in [-\pi, \pi]. \tag{23}$$



Now, (23) implies that

$$\lambda_p \left( \Sigma_{\Delta x}^{(p)}(0) \right) \geq \lambda_p \left( \Sigma_{\Delta \chi}^{(p)}(0) \right) + \lambda^{(p)} \left( \Sigma_{\Delta \epsilon}^{(p)}(0) \right), \tag{24}$$

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 the spectral density matrices are non-negative definite, the right hand side in (24) vanishes if and only if both terms on the right hand side vanish, i.e., the spectral density of  $\Delta x_t^{(p)}$  is singular at zero if and only if the spectral densities of  $\Delta \chi_t^{(p)}$  and  $\Delta \epsilon_t^{(p)}$  are singular at zero. By definition 4, (i) is proved.

Without loss of generality we can assume that  $S(L) = I_r$ . By substituting (21) in (22), we obtain

$$x_t = \Lambda \left[ (\mathbf{G}_1(L)\mathbf{v}_{1t} + \mathbf{G}_2(L)\mathbf{v}_{2t} + \mathbf{T}_t) + \mathbf{K} \right] + \epsilon_t, \tag{25}$$

where on the right hand side the only non-stationary terms are  $\mathbf{T}_t$  and possibly  $\epsilon_t$ . By recalling that  $\mathbf{T}_t = \xi \sum_{s=1}^t \mathbf{v}_{2s}$  where  $\xi$  is of dimension  $r \times (q - d)$  and rank  $q - d$ , and by defining  $\mathcal{G}_t = \Lambda[\mathbf{G}_1(L)\mathbf{v}_{1t} + \mathbf{G}_2(L)\mathbf{v}_{2t} + \mathbf{K}]$  and  $\mathcal{T}_t = \sum_{s=1}^t \mathbf{v}_{2s}$ , we can rewrite (25) as

$$x_t = \Lambda \xi \mathcal{T}_t + \mathcal{G}_t + \epsilon_t.$$

For  $x_t^{(p)}$ :

$$x_t^{(p)} = \chi_t^{(p)} + \epsilon_t^{(p)} = \Lambda^{(p)} \xi \mathcal{T}_t + \mathcal{G}_t^{(p)} + \epsilon_t^{(p)},$$

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

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

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  $V^\chi$  and  $V^\epsilon$  is non-trivial, so that  $x_t^{(p)}$  is cointegrated.  $\square$

### 5. Summary and Conclusions

The paper studies representation theory for singular  $I(1)$  stochastic vectors, the factors of an  $I(1)$  Dynamic Factor Model in particular. Singular  $I(1)$  vectors are cointegrated, with a cointegrating rank  $c$  equal to  $r - q$ , the dimension of  $y_t$  minus its rank, plus  $d$ , with  $0 \leq d < q$ .

If  $(1 - L)y_t$  has rational spectral density, under assumptions that generalize to the singular case those in Johansen (1995), we show that  $y_t$  has an error correction representation with  $c$  error terms, thus generalizing the Granger representation theorem (from MA to AR) to the singular case. Important consequences of singularity are that generically: (i) the autoregressive matrix polynomial of the error correction representation is of finite degree, (ii) the white noise vector driving  $(1 - L)y_t$  is fundamental.

We find that  $y_t$  is driven by  $r - c$  permanent shocks and  $d = c - (r - q)$  transitory shocks, not  $c$  as in the nonsingular case.

Using simulated data generated by a simple singular VECM, confirms previous results, obtained for nonsingular 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 in a DFM produce an amount of cointegration for the observable variables  $x_{it}$  that is not observed in the datasets that are standard in applied Dynamic Factor Model literature. Thus the idiosyncratic vector in those datasets is likely to be  $I(1)$ , so that an estimation strategy robust to the assumption that some of the idiosyncratic variables  $\epsilon_{it}$  are  $I(1)$  should be preferred.

The results in this paper are the basis for estimation of  $I(1)$  Dynamic Factor Models with cointegrated factors, which is developed in the companion paper (Barigozzi et al. 2019).

**Author Contributions:** All authors contributed equally to the paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** Dietmar Bauer, Manfred Deistler, Massimo Franchi, Martin Wagner, three anonymous referees and the Editors of this Special Issue gave important 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. Part of this paper was written while Matteo Luciani was chargé de recherches F.R.S.-F.N.R.S., and he gratefully acknowledges their financial support. Of course we are responsible for any remaining errors.

**Disclaimer:** The views expressed in this paper are those of the authors and do not necessarily reflect those of the Board of Governors or the Federal Reserve System.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A. Proofs

### Appendix A.1. Assumption 3 Holds Generically

Proving that Assumption 3 holds generically is equivalent to proving that  $\mathbf{M}(z)$  is generically zeroless, see the argument below Equation (20).

We need some preliminary results. Lemma A1, though quite easy, is not completely standard and is therefore carefully stated and proved below. Regarding notation, to avoid possible misunderstandings, let us recall that vectors and matrices are always denoted by boldface symbols, while light symbols denote scalars, see Lemmas A1 and A2 in particular.

**Lemma A1.** Let  $A_j, j = 1, \dots, s$ , be scalar polynomials defined on  $\mathbb{R}^\lambda$ , let  $\mathbf{p} \in \mathbb{R}^\lambda$  and  $Q(\mathbf{p})$  be the statement

$$A_j(\mathbf{p}) = 0, \quad \text{for } j = 1, \dots, s,$$

for example the statement that all the  $q \times q$  minors of  $\mathbf{M}(1)$  vanish, i.e., that  $\text{rank}(\mathbf{M}(1)) < q$ . Let  $\Pi$  be an open subset of  $\mathbb{R}^\lambda$ . If  $Q$  is false for one point  $\mathbf{p}^* \in \mathbb{R}^\lambda$ , then  $Q$  is generically false in  $\Pi$ .

**Proof.** Let  $\mathcal{N}$  be the closure in  $\Pi$  (in the topology of  $\Pi$ ) of the subset of  $\Pi$  where  $Q$  is true. Suppose that  $Q$  is not generically false in  $\Pi$ . Then the interior of  $\mathcal{N}$  in  $\Pi$ , call it  $\mathcal{N}^\circ$ , is not empty. As  $\Pi$  is open,  $\mathcal{N}^\circ$  is open both in the topology of  $\Pi$  and of  $\mathbb{R}^\lambda$ . On the other hand a polynomial function defined on  $\mathbb{R}^\lambda$  vanishes on an open set if and only if it vanishes on the whole  $\mathbb{R}^\lambda$ , which contradicts the existence of a point in  $\mathbb{R}^\lambda$  where  $Q$  is false.  $\square$

**Lemma A2.** Consider the scalar polynomials

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

with  $a_0 \neq 0$  and  $b_0 \neq 0$ , and let  $\alpha_i, i = 1, \dots, n$  and  $\beta_j, j = 1, \dots, m$ , be the roots of  $A$  and  $B$ , respectively. Then: (i)

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

where  $R$  is a polynomial function which is called the resultant of  $A$  and  $B$ . (ii) The resultant vanishes if and only if  $A$  and  $B$  have a common root. (iii) Suppose that the coefficients  $a_i$  and  $b_j$  are polynomial functions of  $\mathbf{p} \in \Pi$ , where  $\Pi$  is an open subset of  $\mathbb{R}^\lambda$ . If there exists a point  $\mathbf{p}^* \in \mathbb{R}^\lambda$  such that  $a_0(\mathbf{p}^*) \neq 0$ ,  $b_0(\mathbf{p}^*) \neq 0$ , and  $R(\mathbf{p}^*) \neq 0$ , then generically in  $\Pi$  the polynomials  $A$  and  $B$  have no common roots.

**Proof.** For (i) and (ii) see van der Waerden (1953, pp. 83-8). Statement (iii) is an obvious consequence of (ii) and Lemma A1. □

**Lemma A3.** Recall that a zero of  $\mathbf{M}(z)$  is a complex number  $z^*$  such that  $\text{rank}(\mathbf{M}(z^*)) < q$ . If  $\mathbf{M}(z)$  has two  $q \times q$  submatrices whose determinants have no common roots, then  $\mathbf{M}(z)$  is zeroless.

**Proof.** If  $z^*$  is a zero of  $\mathbf{M}(z)$ , then  $z^*$  is a zero of all the  $q \times q$  submatrices of  $\mathbf{M}(z)$ . □

For the statement and proof of our last result it is convenient to make explicit the dependence of the matrix  $\mathbf{M}(z)$  and its submatrices on the vector  $\mathbf{p}$ . Thus we use  $\mathbf{M}^{\mathbf{p}}(z)$ , etc. The parameters of the matrix  $\mathbf{S}(L)$  play no role here. Hence, with no loss of generality, we assume  $s_2 = 0$ , so that  $\lambda = (r - c)(r + q) + rq(s_1 + 2)$ . Lemmas A2–A4 below imply that Assumption 3 holds generically in  $\Pi$ .

**Lemma A4.** Let  $\mathbf{M}_1^{\mathbf{p}}(z), \mathbf{M}_2^{\mathbf{p}}(z), \dots$  be all the  $q \times q$  submatrices of  $\mathbf{M}^{\mathbf{p}}(z)$  and let  $\mathcal{L}_i^{\mathbf{p}}$  be the leading coefficient of  $\det \mathbf{M}_i^{\mathbf{p}}(z)$  and  $R_{ij}^{\mathbf{p}}$  is the resultant of  $\det \mathbf{M}_i^{\mathbf{p}}(z)$  and  $\det \mathbf{M}_j^{\mathbf{p}}(z)$ . There exist  $i, j, \mathbf{p}^* \in \mathbb{R}^\lambda$  such that

$$\mathcal{L}_i^{\mathbf{p}^*} \mathcal{L}_j^{\mathbf{p}^*} \neq 0$$

and

$$R_{ij}^{\mathbf{p}^*} \neq 0.$$

**Proof.** Assume that  $r = q + 1$ . To each  $\mathbf{p} \in \Pi$  there corresponds the matrix

$$\mathbf{M}^{\mathbf{p}}(z) = \begin{pmatrix} \boldsymbol{\zeta}'_{\perp} \mathbf{B}^* \\ \boldsymbol{\zeta}' \boldsymbol{\zeta} \boldsymbol{\eta}' \end{pmatrix} + (1 - z) \begin{pmatrix} \boldsymbol{\zeta}'_{\perp} \mathbf{E}(z) \\ \boldsymbol{\zeta}' \mathbf{B}^* \end{pmatrix} + (1 - z)^2 \begin{pmatrix} \mathbf{0}_{c \times q} \\ \boldsymbol{\zeta}' \mathbf{E}(z) \end{pmatrix}.$$

□

Of course, the definition of  $\mathbf{M}^{\mathbf{p}}(z)$  makes sense for all  $\mathbf{p} \in \mathbb{R}^\lambda$ , see Equation (19). Let  $\mathbf{M}_1^{\mathbf{p}}(z)$  and  $\mathbf{M}_2^{\mathbf{p}}(z)$  be the matrices obtained from  $\mathbf{M}^{\mathbf{p}}(z)$  by removing the first and the last row respectively. We have:

$$\begin{aligned} \text{degree}[\det(\mathbf{M}_1^{\mathbf{p}}(z))] &\leq (q - d)(s_1 + 2) + d(s_1 + 1) = d_1, \\ \text{degree}[\det(\mathbf{M}_2^{\mathbf{p}}(z))] &\leq (q - d - 1)(s_1 + 2) + (d + 1)(s_1 + 1) = d_2. \end{aligned}$$

We will construct a point  $\mathbf{p}^* \in \mathbb{R}^\lambda$  such that: (A) the coefficient of  $z^{d_1}$  in  $\det(\mathbf{M}_1^{\mathbf{p}^*}(z))$  and the coefficient of  $z^{d_2}$  in  $\det(\mathbf{M}_2^{\mathbf{p}^*}(z))$  (the leading coefficients) do not vanish, (B) the resultant of  $\det(\mathbf{M}_1^{\mathbf{p}^*}(z))$  and  $\det(\mathbf{M}_2^{\mathbf{p}^*}(z))$  does not vanish.

Let us firstly define a family of matrices, denoted by  $\underline{\mathbf{M}}(z)$ , obtained by specifying  $\boldsymbol{\eta}, \boldsymbol{\zeta}, \boldsymbol{\zeta}'_{\perp}, \mathbf{B}^*$  and  $\mathbf{E}(z)$  in the following way:

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

where:

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

$$\mathbf{E}_1(z) = \begin{pmatrix} & k_1(z) & h_1(z) & \cdots & 0 \\ \mathbf{0}_{(q-d) \times d} & & \ddots & \ddots & \\ & & & \ddots & h_{q-d-1}(z) \\ & 0 & \cdots & & k_{q-d}(z) \end{pmatrix}, \quad \mathbf{E}_2(z) = \begin{pmatrix} e(z) & \mathbf{0}_{1 \times (q-1)} \end{pmatrix},$$

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

the entries  $e, k_i, h_i, f_i$  and  $g_i$  being scalar polynomials of degree  $s_1$ .

We denote by  $\mathbf{q}_1$  the vector including the coefficients of the polynomials  $f_i, i = 1, \dots, d$  and  $k_i, i = 1, \dots, (q-d)$ , a total of  $q(s_1 + 1)$  coefficients, by  $\mathbf{q}_2$  the vector including the coefficients of the polynomials  $e, g_i, i = 1, \dots, d$  and  $h_i, i = 1, \dots, (q-d-1)$ , a total of  $q(s_1 + 1)$  coefficients, by  $\mathbf{q}_0$  the vector including the zeros and the ones in the definition of  $\underline{\xi}, \underline{\eta}, \mathbf{B}^*, \mathbf{E}$ , and define  $\mathbf{q} = (\mathbf{q}_0 \mathbf{q}_1 \mathbf{q}_2)$ , which is a  $\lambda$ -dimensional parameter vector. We put no restriction on  $\mathbf{q}_1$  and  $\mathbf{q}_2$ , so that both can take any value in  $\mathbb{R}^\nu$ , with  $\nu = q(s_1 + 1)$ . Note that  $\mathbf{q}$  does not necessarily belong to  $\Pi$ . We have:

$$\underline{\mathbf{M}}^{\mathbf{q}}(z) = \begin{pmatrix} \mathbf{0}_{1 \times d} & \mathbf{0}_{1 \times (q-d)} \\ \mathbf{I}_d & \mathbf{0}_{d \times (q-d)} \\ \mathbf{0}_{(q-d) \times d} & \mathbf{I}_{q-d} \end{pmatrix} + (1-z) \begin{pmatrix} \mathbf{E}_2(z) \\ \mathbf{E}_3(z) \end{pmatrix} + (1-z)^2 \begin{pmatrix} \mathbf{0}_{1 \times q} \\ \mathbf{0}_{d \times q} \\ \mathbf{E}_1(z) \end{pmatrix}. \tag{A1}$$

The matrix  $\underline{\mathbf{M}}^{\mathbf{q}}(z)$  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  $\underline{\mathbf{M}}_1^{\mathbf{q}}(z)$  and  $\underline{\mathbf{M}}_2^{\mathbf{q}}(z)$  are upper- and lower-triangular, respectively, and

$$\det(\underline{\mathbf{M}}_1^{\mathbf{q}}(z)) = [1 + (1-z)f_1(z)] \cdots [(1 + (1-z)f_d(z)]$$

$$\times [1 + (1-z)^2k_1(z)] \cdots [1 + (1-z)^2k_{q-d}(z)] = \mathcal{L}_{1,d_1}^{\mathbf{q}} z^{d_1} + \cdots + \mathcal{L}_{1,0}^{\mathbf{q}}$$

$$\det(\underline{\mathbf{M}}_2^{\mathbf{q}}(z)) = (1-z)^{2q-d-1} e(z) [g_1(z) \cdots g_d(z)] [h_1(z) \cdots h_{q-d-1}(z)] = \mathcal{L}_{2,d_2}^{\mathbf{q}} z^{d_2} + \cdots + \mathcal{L}_{2,0}^{\mathbf{q}}.$$

Note that  $\det(\underline{\mathbf{M}}_1^{\mathbf{q}}(z))$  does not depend on  $\mathbf{q}_2$ , while  $\det(\underline{\mathbf{M}}_2^{\mathbf{q}}(z))$  does not depend on  $\mathbf{q}_1$ . Thus we use the notation  $\delta_1^{\mathbf{q}_1}(z) = \det(\underline{\mathbf{M}}_1^{\mathbf{q}}(z)), \delta_2^{\mathbf{q}_2}(z) = \det(\underline{\mathbf{M}}_2^{\mathbf{q}}(z)), \mathcal{M}_{1,d_1}^{\mathbf{q}_1} = \mathcal{L}_{1,d_1}^{\mathbf{q}}, \mathcal{M}_{2,d_2}^{\mathbf{q}_2} = \mathcal{L}_{2,d_2}^{\mathbf{q}}$ . Now:

- (i) Let  $\mathbf{q}_2^* \in \mathbb{R}^\nu$  be such that none of the leading coefficients of the polynomials  $e, g_i$  and  $h_i$  vanishes. Of course  $\mathcal{M}_{2,d_2}^{\mathbf{q}_2^*} = d_2 \neq 0$ .
- (ii) Let  $\check{z}$  be a root of  $\delta_2^{\mathbf{q}_2^*}(z)$ . If  $\check{z} = 1$  then  $\check{z}$  is not a root of  $\delta_1^{\mathbf{q}_1}(z)$  for all  $\mathbf{q}_1 \in \mathbb{R}^\nu$ . Suppose that  $\check{z}$  is a root of  $g_j(z)$ , for some  $j$ . As the parameters of the polynomials  $f_i$  and  $k_i$  are free to vary in  $\mathbb{R}^\nu$ , then, generically in  $\mathbb{R}^\nu, \delta_1^{\mathbf{q}_1}(\check{z}) \neq 0$ . Iterating for all roots of  $\delta_2^{\mathbf{q}_2^*}(z)$ , generically in  $\mathbb{R}^\nu, \delta_1^{\mathbf{q}_1}(z)$  and  $\delta_2^{\mathbf{q}_2^*}(z)$  have no roots in common. Moreover, generically in  $\mathbb{R}^\nu, \mathcal{M}_{1,d_1}^{\mathbf{q}_1} = d_1 \neq 0$ . Thus, there exists  $\mathbf{q}_1^*$  such that (a)  $\mathcal{M}_{1,d_1}^{\mathbf{q}_1^*} = d_1 \neq 0$ , (b)  $\delta_1^{\mathbf{q}_1^*}(z)$  and  $\delta_2^{\mathbf{q}_2^*}(z)$  have no roots in common.
- (iii) Now let  $\mathbf{p}^* = (\mathbf{q}_0 \mathbf{q}_1^* \mathbf{q}_2^*)$ , so that

$$\det(\underline{\mathbf{M}}_1^{\mathbf{p}^*}(z)) = \delta_1^{\mathbf{q}_1^*}(z), \quad \det(\underline{\mathbf{M}}_2^{\mathbf{p}^*}(z)) = \delta_2^{\mathbf{q}_2^*}(z).$$

Using (i) and (ii), (A) the leading coefficients of  $\det(\mathbf{M}_1^{\mathbf{P}^*}(z))$  and  $\det(\mathbf{M}_2^{\mathbf{P}^*}(z))$  do not vanish, (B)  $\det(\mathbf{M}_1^{\mathbf{P}^*}(z))$  and  $\det(\mathbf{M}_2^{\mathbf{P}^*}(z))$  have no root in common so that their resultant does not vanish. This proves the proposition for  $r = q + 1$ .

Generalizing this result to  $r > q + 1$  is easy. Let us define the family  $\underline{\mathbf{N}}(z)$  in the following way: (a) specify  $\boldsymbol{\eta}'$ ,  $\boldsymbol{\xi}$ ,  $\mathbf{E}_1(z)$  and  $\mathbf{E}_3(z)$  as in the definition of  $\underline{\mathbf{M}}(z)$ , (b) then let

$$\mathbf{K} = \begin{pmatrix} \mathbf{0}_{(r-q) \times (r-d-1)} & (\mathbf{0}_{1 \times (r-q-1)} \mathbf{1})' & \mathbf{0}_{(r-q) \times d} \end{pmatrix}, \quad \mathbf{H} = \begin{pmatrix} \mathbf{0}_{d \times (r-d)} & \mathbf{I}_d \end{pmatrix},$$

$$\underline{\boldsymbol{\xi}}'_{\perp} = \begin{pmatrix} \mathbf{K} \\ \mathbf{H} \end{pmatrix}, \quad \underline{\mathbf{D}} = \begin{pmatrix} \mathbf{H}' & \mathbf{I}_{r \times (q-d)} \end{pmatrix}, \quad \mathbf{E}_2(z) = \begin{pmatrix} \mathbf{0}_{(r-q) \times q} \\ (e(z) \quad \mathbf{0}_{1 \times (q-1)}) \end{pmatrix}.$$

We have:

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

It is easy to see that the  $(q + 1) \times q$  lower submatrix of  $\underline{\mathbf{N}}(z)$  is identical to the matrix  $\underline{\mathbf{M}}^q(z)$  in (A1).

*Appendix A.2. if  $R > Q$  and  $C \leq Q$ , Assumptions 5 and 6 Do Not Imply That  $\mathbf{e}_t$  Is a Non-Cointegrated  $I(0)$  Process.*

Let  $r = 3, q = 2, \mathbf{S}(L) = \mathbf{I}_3$ ,

$$\boldsymbol{\xi} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{\eta} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \boldsymbol{\xi}_{\perp} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{B}^* = \begin{pmatrix} a & b \\ 1 & 0 \\ 1 & 0 \end{pmatrix}.$$

In this case  $c = 2$  and  $d = 1$ , so that  $c = q$  (see Remark 6). We have

$$\begin{pmatrix} \boldsymbol{\xi}'_{\perp} \mathbf{B}^* \\ \boldsymbol{\eta}' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

We see that Assumptions 5 and 6 hold. However,  $\text{rank}(\boldsymbol{\xi}'_{\perp} \mathbf{B}^*) = 1$ , so that  $\mathbf{e}_t$ , though being  $I(0)$ , is not a non-cointegrated  $I(0)$  process. On the other hand, if the  $(3,2)$  entry of  $\mathbf{B}^*$  is 1 instead of 0,  $\mathbf{e}_t$  is non-cointegrated.

### Appendix B. Non Uniqueness

In Proposition 3 we prove that a singular  $I(1)$  vector with cointegrating rank  $c$  has a finite error correction representation with  $c$  error terms. On the other hand, as we have seen in Remark 5, when  $c = r - q$  the singular vector  $\mathbf{y}_t$  has also an autoregressive representation in the differences, i.e., a representation with zero error terms. In Appendix B.1 we give an example hinting that  $\mathbf{y}_t$  has error correction representations with any number of error terms between  $d$  and  $c$ . However, in Appendix B.2 we show that all such representations produce the same impulse-response functions.

#### Appendix B.1. Alternative Representations with Different Numbers of Error Terms

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

$$\begin{aligned} \zeta' &= \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \\ \eta' &= \begin{pmatrix} 1 & 2 \end{pmatrix} \\ \zeta'_{\perp} &= \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \end{aligned}$$

We have,

$$(1 - L) \begin{pmatrix} \zeta'_{\perp} \\ \zeta' \end{pmatrix} \mathbf{y}_t = \begin{pmatrix} 1 - L & 0 & 0 \\ 0 & 1 - L & 0 \\ 0 & 0 & 1 \end{pmatrix} \left\{ \begin{pmatrix} b_{11}^* - b_{21}^* & b_{12}^* - b_{22}^* \\ b_{21}^* - b_{31}^* & b_{22}^* - b_{32}^* \\ 3 & 6 \end{pmatrix} + (1 - L)\hat{\mathbf{E}}(L) \right\} \mathbf{u}_t,$$

where  $(1 - L)\hat{\mathbf{E}}(L)$  gathers the second and third terms in  $\mathbf{M}(L)$ . If the assumptions of Proposition 2 hold, we obtain an error correction representation with error terms

$$\zeta'_{\perp} \mathbf{y}_t = \begin{pmatrix} y_{1t} - y_{2t} \\ y_{2t} - y_{3t} \end{pmatrix}.$$

However, we also have

$$\begin{aligned} (1 - L) \begin{pmatrix} \zeta'_{\perp} \\ \zeta' \end{pmatrix} \mathbf{y}_t &= \begin{pmatrix} 1 - L & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &\times \left\{ \begin{pmatrix} b_{11}^* - b_{21}^* & b_{12}^* - b_{22}^* \\ (1 - L)(b_{21}^* - b_{31}^*) & (1 - L)(b_{22}^* - b_{32}^*) \\ 3 & 6 \end{pmatrix} + (1 - L)\check{\mathbf{E}}(L) \right\} \mathbf{u}_t = \begin{pmatrix} 1 - L & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \check{\mathbf{M}}(L) \mathbf{u}_t. \end{aligned}$$

Under suitable assumptions on the coefficients  $b_{ij}^*$  and  $\check{\mathbf{E}}(L)$ , assuming in particular that the matrix

$$\begin{pmatrix} b_{11}^* - b_{21}^* & b_{12}^* - b_{22}^* \\ 3 & 6 \end{pmatrix}$$

is nonsingular, the matrix  $\check{\mathbf{M}}(L)$  is zeroless and has therefore a finite-degree left inverse. Proceeding as in Proposition 2, we obtain an alternative error correction representation with just one error term, namely  $y_{1t} - y_{2t}$ .

This example should be sufficient to convey the idea that  $\mathbf{y}_t$  admits error correction representations with a minimum  $d$  and a maximum  $c = r - q + d$  of error terms.

The problem of error correction representations, with different numbers of error terms, has been recently addressed in Deistler and Wagner (2017). An implication of their main result (see Theorem 1, p. 41) is that if  $\mathbf{y}_t$  has the error correction representation

$$\tilde{\mathbf{A}}(L)\mathbf{y}_t = \tilde{\mathbf{A}}^*(L)(1 - L)\mathbf{y}_t + \tilde{\mathbf{A}}(1)\mathbf{y}_{t-1} = \tilde{\mathbf{B}}\tilde{\mathbf{u}}_t,$$

and  $\text{rank}(\tilde{\mathbf{A}}(1)) < c$  (the number of error terms is not the maximum), then  $\tilde{\mathbf{A}}(L)$  and  $\tilde{\mathbf{B}}$  are not left coprime.

The consequences of Deistler and Wagner’s paper have not yet been developed. In Propositions 2 and 3 we have only considered representations with  $c$  error terms. On non-uniqueness of autoregressive representations for singular vectors with rational spectral density see also Chen et al. (2011); Anderson et al. (2012); Forni et al. (2015).



Appendix B.2. Uniqueness of Impulse-Response Functions

Suppose that the assumptions of Proposition 2, weak form, hold. Let  $\mathbf{y}_t$  be a solution of Equation (10), so that

$$(1 - L)\mathbf{y}_t = \mathbf{S}(L)^{-1}\mathbf{B}(L)\mathbf{u}_t, \tag{A2}$$

and suppose that  $\mathbf{y}_t$  has the autoregressive representation

$$\tilde{\mathbf{A}}(L)\mathbf{y}_t = \tilde{\mathbf{B}}\tilde{\mathbf{u}}_t, \tag{A3}$$

where  $\tilde{\mathbf{A}}(L)$  is a rational matrix with poles outside the unit circle,  $\tilde{\mathbf{A}}(0) = \mathbf{I}_r$ ,  $\tilde{\mathbf{u}}_t$  is a nonsingular  $q$ -dimensional white noise,  $\tilde{\mathbf{B}}$  is a full rank  $r \times q$  matrix<sup>5</sup>. We have

$$\tilde{\mathbf{A}}(L)[(1 - L)\mathbf{y}_t] = (1 - L)\tilde{\mathbf{B}}\tilde{\mathbf{u}}_t. \tag{A4}$$

The assumption that  $\tilde{\mathbf{B}}$  is full rank and the argument used e.g., in Brockwell and Davis (1991), p. 111, Problem 3.8, imply that  $\tilde{\mathbf{u}}_t$  is fundamental for  $(1 - L)\mathbf{y}_t$ . Thus  $\tilde{\mathbf{u}}_t = \mathbf{Q}\mathbf{u}_t$ , where  $\mathbf{Q}$  is a nonsingular  $q \times q$  matrix (see Rozanov (1967), p. 57), and  $\tilde{\mathbf{B}}\tilde{\mathbf{u}}_t = [\tilde{\mathbf{B}}\mathbf{Q}]\mathbf{u}_t$ .

On the other hand, from (A2) and (A4):

$$\tilde{\mathbf{A}}(L)\mathbf{S}(L)^{-1}\mathbf{B}(L)\mathbf{u}_t = (1 - L)[\tilde{\mathbf{B}}\mathbf{Q}]\mathbf{u}_t. \tag{A5}$$

As  $\mathbf{u}_t$  is nonsingular,  $\tilde{\mathbf{A}}(L)\mathbf{S}(L)^{-1}\mathbf{B}(L) = (1 - L)[\tilde{\mathbf{B}}\mathbf{Q}]$ . Setting  $L = 0$  we have  $\tilde{\mathbf{B}}\mathbf{Q} = \mathbf{B}(0)$ , so that (A3) becomes

$$\tilde{\mathbf{A}}(L)\mathbf{y}_t = \mathbf{B}(0)\mathbf{u}_t \tag{A6}$$

while (A5) becomes

$$\tilde{\mathbf{A}}(L)\mathbf{S}(L)^{-1}\mathbf{B}(L)\mathbf{u}_t = (1 - L)\mathbf{B}(0)\mathbf{u}_t. \tag{A7}$$

The impulse-response function of  $\mathbf{y}_t$  to  $\mathbf{u}_t$  resulting from (A6) is  $\mathbf{H}(L)\mathbf{B}(0)$ , where  $\mathbf{H}(L)\tilde{\mathbf{A}}(L) = \mathbf{I}_r$ . Multiplying both sides of (A7) by  $\mathbf{H}(L)$  we obtain

$$\mathbf{S}(L)^{-1}\mathbf{B}(L) = (1 - L)\mathbf{H}(L)\mathbf{B}(0),$$

so that  $\mathbf{H}(L)\mathbf{B}(0)$  is obtained by cumulating  $\mathbf{S}(L)^{-1}\mathbf{B}(L)$  and is therefore independent of  $\tilde{\mathbf{A}}(L)$ .

Appendix C. Data Generating Process for the Simulations

The simulation results of Section 3.4 are obtained using the following specification of (14):

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

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

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

---

<sup>5</sup> Multiplying both sides of (A3) by  $(1 - L)$  and using (A2), we obtain  $\tilde{\mathbf{A}}(L)\mathbf{S}(L)^{-1}\mathbf{B}(L)\mathbf{u}_t = (1 - L)\tilde{\mathbf{B}}\tilde{\mathbf{u}}_t$ . Comparing the spectral densities of right- and left-hand terms, it is easy to prove that  $\tilde{\mathbf{u}}_t$  must be a  $q$ -dimensional, nonsingular white noise and the rank of  $\tilde{\mathbf{B}}$  must be  $q$ .

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

Regarding the generation of the data, the diagonal entries of the matrix  $\mathcal{U}_1$  are drawn from a uniform distribution between 0.5 and 0.8, while the extra-diagonal entries are drawn from a uniform distribution between 0 and 0.3.  $\mathcal{U}_1$  is then multiplied by a scalar so that its largest eigenvalue is 0.6. The matrix  $\mathbf{G}$  is generated as in Bai and Ng (2007): (1)  $\tilde{\mathbf{G}}$  is an  $r \times r$  diagonal matrix of rank  $q$  where  $\tilde{g}_{ii}$  is drawn from the uniform distribution between 0.8 and 1.2, (2)  $\check{\mathbf{G}}$  is obtained by orthogonalizing an  $r \times r$  uniform random matrix, (3)  $\mathbf{G}$  is equal to the first  $q$  columns of the matrix  $\check{\mathbf{G}}\check{\mathbf{G}}^{1/2}$ . Lastly, the orthogonal matrix  $\mathbf{H}$  is such that the upper  $3 \times 3$  submatrix of  $\mathbf{G}\mathbf{H}$  is lower triangular. The results are based on 1000 replications. The matrices  $\mathcal{U}_1$ ,  $\mathbf{G}$  and  $\mathbf{H}$  are generated only once (the numerical values are available on request) 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.

## References

- Amengual, Dante, and Mark 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, Brian DO, and Manfred Deistler. 2008a. Generalized linear dynamic factor models—A structure theory. Paper presented at IEEE Conference on Decision and Control, Cancun, Mexico, December 9–11.
- Anderson, Brian DO, and Manfred Deistler. 2008b. Properties of zero-free transfer function matrices. *SICE Journal of Control, Measurement and System Integration* 1: 284–92.
- Anderson, Brian DO, Manfred Deistler, Weitian Chen, and Alexander Filler. 2012. Autoregressive models of singular spectral matrices. *Automatica* 48: 2843–49.
- Bai, Jushan, and Serena Ng. 2007. Determining the number of primitive shocks in factor models. *Journal of Business and Economic Statistics* 25: 52–60.
- Banerjee, Anindya, Massimiliano Marcellino, and Igor Masten. 2014. Forecasting with factor-augmented error correction models. *International Journal of Forecasting* 30: 589–612.
- Banerjee, Anindya, Massimiliano Marcellino, and Igor Masten. 2017. Structural FECM: Cointegration in large-scale structural FAVAR models. *Journal of Applied Econometrics* 32: 1069–86.
- Barigozzi, Matteo, Antonio M. Conti, and Matteo Luciani. 2014. Do euro area countries respond asymmetrically to the common monetary policy? *Oxford Bulletin of Economics and Statistics* 76: 693–714.
- Barigozzi, Matteo, Marco Lippi, and Matteo Luciani. 2019. Large-dimensional dynamic factor models: Estimation of impulse-response functions with  $I(1)$  cointegrated factors. *arXiv* arXiv:1602:02398.
- Bauer, Dietmar, and Martin Wagner. 2012. A State Space Canonical Form For Unit Root Processes. *Econometric Theory* 28: 1313–49.
- Brockwell, Peter J., and Richard A. Davis. 1991. *Time Series: Theory and Methods*, 2nd ed. New York: Springer.
- Canova, Fabio. 2007. *Methods for Applied Macroeconomics*. Princeton: Princeton University Press.
- Chen, Weitian, Brian DO Anderson, Manfred Deistler, and Alexander Filler. 2011. Solutions of Yule-Walker equations for singular AR processes. *Journal of Time Series Analysis* 32: 531–38.
- Deistler, Manfred, Brian DO Anderson, A. Filler, Ch. Zinner, and W. Chen. 2010. Generalized linear dynamic factor models: An approach via singular autoregressions. *European Journal of Control* 16: 211–24.
- Deistler, Manfred, and Martin Wagner. 2017. Cointegration in singular ARMA models. *Economics Letters* 155: 39–42.
- Forni, Mario, and Luca Gambetti. 2010. The dynamic effects of monetary policy: A structural factor model approach. *Journal of Monetary Economics* 57: 203–16.
- Forni, Mario, Domenico Giannone, Marco Lippi, and Lucrezia Reichlin. 2009. Opening the Black Box: Structural Factor Models versus Structural VARs. *Econometric Theory* 25: 1319–47.
- Forni, Mario, Marc Hallin, Marco Lippi, and Lucrezia Reichlin. 2000. The Generalized Dynamic Factor Model: Identification and Estimation. *The Review of Economics and Statistics* 82: 540–54.
- Forni, Mario, Marc Hallin, Marco Lippi, and Paolo Zaffaroni. 2015. Dynamic factor models with infinite-dimensional factor spaces: One-sided representations. *Journal of Econometrics* 185: 359–71.
- Forni, Mario, and Marco Lippi. 2001. The Generalized Dynamic Factor Model: Representation Theory. *Econometric Theory* 17: 1113–41.

- Franchi, Massimo, and Paolo Paruolo. 2019. A general inversion theorem for cointegration. *Econometric Reviews* 38: 1176–201.
- Franklin, J. N. 2000. *Matrix Theory*, 2nd ed. New York: Dover Publications.
- Giannone, Domenico, Lucrezia Reichlin, and Luca Sala. 2005. Monetary policy in real time. In *NBER Macroeconomics Annual 2004*. Edited by Mark Gertler and Kenneth Rogoff. Cambridge: MIT Press, chp. 3, pp. 161–224.
- Gregoir, Stéphane. 1999. Multivariate Time Series With Various Hidden Unit Roots, Part I. *Econometric Theory* 15: 435–68.
- Johansen, Søren. 1988. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12: 231–54.
- Johansen, Søren. 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59: 1551–80.
- Johansen, Søren. 1995. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, 1st ed. Oxford: Oxford University Press.
- Lancaster, Peter, and Miron Tismenetsky. 1985. *The Theory of Matrices*, 2nd ed. New York: Academic Press.
- Luciani, Matteo. 2015. Monetary policy and the housing market: A structural factor analysis. *Journal of Applied Econometrics* 30: 199–218.
- Phillips, Peter C.B. 1998. Impulse response and forecast error variance asymptotics in nonstationary VARs. *Journal of Econometrics* 83: 21–56.
- Rozanov, Yu. A. 1967. *Stationary Random Processes*. San Francisco: Holden-Day.
- Sargent, Thomas J. 1989. Two Models of Measurements and the Investment Accelerator. *Journal of Political Economy* 97: 251–87.
- Sims, Christopher, James H. Stock, and Mark W. Watson. 1990. Inference in linear time series models with some unit roots. *Econometrica* 58: 113–44.
- Stock, James H., and Mark W. Watson. 1988. Testing for common trends. *Journal of the American Statistical Association* 83: 1097–107.
- Stock, James H., and Mark W. Watson. 2002a. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97: 1167–79.
- Stock, James H., and Mark W. Watson. 2002b. Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics* 20: 147–62.
- Stock, James H., and Mark W. Watson. 2016. Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of Macroeconomics*. Edited by John B. Taylor and Harald Uhlig. Amsterdam: North Holland, Elsevier, vol. 2A, chp. 8, pp. 415–525.
- Van der Waerden, Bartel Leendert. 1953. *Modern Algebra*, 2nd ed. New York: Frederick Ungar, vol. I.
- Watson, Mark W. 1994. Vector autoregressions and cointegration. In *Handbook of Econometrics*. Edited by Robert F. Engle and Daniel L. McFadden. Amsterdam: North Holland, Elsevier, vol. 4, chp. 47, pp. 2843–915.

