

Article

A Parameterization of Models for Unit Root Processes: Structure Theory and Hypothesis Testing

Dietmar Bauer ^{1,*} , Lukas Matuschek ^{1,2}, Patrick de Matos Ribeiro ^{1,2} and Martin Wagner ^{3,4,5}

¹ Faculty of Business Administration and Economics, Bielefeld University, Universitätsstraße 25, 33615 Bielefeld, Germany; matuschek@statistik.tu-dortmund.de (L.M.); ribeiro@statistik.tu-dortmund.de (P.d.M.R.)

² Faculty of Statistics, TU Dortmund, Vogelpothsweg 87, 44227 Dortmund, Germany

³ Department of Economics, University of Klagenfurt, Universitätsstraße 65-67, 9020 Klagenfurt, Austria; Martin.Wagner@aau.at

⁴ Bank of Slovenia, Slovenska 35, 1505 Ljubljana, Slovenia

⁵ Institute for Advanced Studies, Josefstädter Straße 39, 1080 Vienna, Austria

* Correspondence: Dietmar.Bauer@uni-bielefeld.de; Tel.: +49-521-106-4873

Received: 18 April 2018; Accepted: 4 November 2020; Published: 10 November 2020



Abstract: We develop and discuss a parameterization of vector autoregressive moving average processes with arbitrary unit roots and (co)integration orders. The detailed analysis of the topological properties of the parameterization—based on the state space canonical form of Bauer and Wagner (2012)—is an essential input for establishing statistical and numerical properties of pseudo maximum likelihood estimators as well as, e.g., pseudo likelihood ratio tests based on them. The general results are exemplified in detail for the empirically most relevant cases, the (multiple frequency or seasonal) I(1) and the I(2) case. For these two cases we also discuss the modeling of deterministic components in detail.

Keywords: canonical form; cointegration; hypothesis testing; parameterization; state space representation; unit roots

1. Introduction

Since the seminal contribution of Clive W.J. Granger (1981) that introduced the concept of cointegration, the modeling of multivariate (economic) time series with models and methods that allow for unit roots and cointegration has become standard econometric practice with applications ranging from macroeconomics to finance to climate science.

The most prominent (parametric) model class for cointegration analysis are vector autoregressive (VAR) models, popularized by the important contributions of Søren Johansen and Katarina Juselius and their co-authors, see, e.g., the monographs Johansen (1995) and Juselius (2006). The popularity of VAR cointegration analysis stems not only from the (relative) simplicity of the model class, but also from the fact that the VAR cointegration literature is very well-developed and provides a large battery of tools for diagnostic testing, impulse response analysis, forecast error variance decompositions and the like. All this makes VAR cointegration analysis to a certain extent the benchmark in the literature.¹

The imposition of specific cointegration properties on an estimated VAR model becomes increasingly complicated as one moves away from the I(1) case. As discussed in Section 2, e.g., in the

¹ Please note that the original contribution to the estimation of cointegrating relationship has been least squares estimation in a non- or semi-parametric regression setting, see, e.g., Engle and Granger (1987). A recent survey of regression-based cointegration analysis is provided by Wagner (2018).

I(2) case a triple of indices needs to be chosen (fixed or determined via testing) to describe the cointegration properties. The imposition of cointegration properties in the estimation algorithm then leads to “switching” type algorithms that come together with non-trivial parameterization restrictions involving non-linear inter-relations, compare [Paruolo \(1996\)](#) or [Paruolo \(2000\)](#).² Mathematically, these complications arise from the fact that the unit root and cointegration properties are in the VAR setting related to rank restrictions on the autoregressive polynomial matrix and its derivatives.

Restricting cointegration analysis to VAR processes may be too restrictive. First, it is well-known since [Zellner and Palm \(1974\)](#) that VAR processes are not invariant with respect to marginalization, i.e., subsets of the variables of a VAR process are in general vector autoregressive moving average (VARMA) processes. Second, similar to the first argument, aggregation of VAR processes also leads to VARMA processes, an issue relevant, e.g., in the context of temporal aggregation and in mixed-frequency settings. Third, the linearized solutions to dynamic stochastic general equilibrium (DSGE) models are typically VARMA rather than VAR processes, see, e.g., [Campbell \(1994\)](#). Fourth, a VARMA model may be a more parsimonious description of the data generating process (DGP) than a VAR model, with parsimony becoming more important with increasing dimension of the process.³

If one accepts the above arguments as a motivation for considering VARMA processes in cointegration analysis, it is convenient to move to the—essentially equivalent (see [Hannan and Deistler 1988](#), chps. 1 and 2)—state space framework. A key challenge when moving from VAR to VARMA models—or state space models—is that *identification* becomes an important issue for the latter model class, whereas unrestricted VAR models are (reduced-form) identified. In other words, there are so-called equivalence classes of VARMA models that lead to the same dynamic behavior of the observed process. As is well-known, to achieve identification, restrictions have to be placed on the coefficient matrices in the VARMA case, e.g., zero or exclusion restrictions. A mapping attaching to every transfer function, i.e. the function relating the error sequence to the observed process, a unique VARMA (or state space) system from the corresponding class of observationally equivalent systems is called *canonical form*. Since not all entries of the coefficient matrices in canonical form are free parameters, for statistical analysis a so-called *parameterization* is required that maps the free parameters from coefficient matrices in canonical form into a parameter vector. These issues, including the importance of the properties such as continuity and differentiability of parameterizations, are discussed in detail in [Hannan and Deistler \(1988, chp. 2\)](#) and, of course, are also relevant for our setting in this paper.

The convenience of the state space framework for unit root and cointegration analysis stems from the fact that (static and dynamic) cointegration can be characterized by orthogonality constraints, see [Bauer and Wagner \(2012\)](#), once an appropriate basis for the state vector, which is a (potentially singular) VAR process of order one, is chosen. The integration properties are governed by the eigenvalue structure of unit modulus eigenvalues of the system matrix in the state equation. Eigenvalues of unit modulus and orthogonality constraints arguably are easier restrictions to deal with or to implement than the interrelated rank restrictions considered in the VAR or VARMA setting. The canonical form of [Bauer and Wagner \(2012\)](#) is designed for cointegration analysis by using a basis of the state vector that puts the unit root and cointegration properties to the center and forefront. Consequently, these results are key input for the present paper and are thus briefly reviewed in Section 3.

² The complexity of these inter-relations is probably well illustrated by the fact that only [Jensen \(2013\)](#) notes that “even though the I(2) models are formulated as submodels of I(1) models, some I(1) models are in fact submodels of I(2) models”.

³ The literature often uses VAR models as approximations, based on the fact that VARMA processes often can be approximated by VAR models with the order tending to infinity with the sample size at certain rates. This line of work goes back to [Lewis and Reinsel \(1985\)](#) for stationary processes and was extended to (co)integrated processes by [Saikkonen \(1992\)](#), [Saikkonen and Luukkonen \(1997\)](#) and [Bauer and Wagner \(2005\)](#). In addition to the issue of the existence and properties of a sequence of VAR approximations, the question whether a VAR approximation is parsimonious remains.

An important problem with respect to appropriately defining the “free parameters” in VARMA models is the fact that no continuous parameterization of all VARMA or state space models of a certain order n exists in the multivariate case (see [Hazewinkel and Kalman 1976](#)). This implies that the model set, M_n say, has to be partitioned into subsets on which continuous parameterizations exist, i.e., $M_n = \bigcup_{\Gamma \in G} M_\Gamma$ for some multi-index Γ varying in an index set G . Based on the canonical form of [Bauer and Wagner \(2012\)](#), the partitioning is according to systems—in addition to other restrictions such as fixed order n —with fixed unit root properties, to be precise over systems with given state space unit root structure. This has the advantage that, e.g., pseudo maximum likelihood (PML) estimation can straightforwardly be performed over systems with fixed unit root properties without any further ado, i.e., without having to consider (or ignore) rank restrictions on polynomial matrices. The definition and detailed discussion of the properties of this parameterization is the first main result of the paper.

The second main set of results, provided in Section 4, is a detailed discussion of the relationships between the different subsets of models M_Γ for different indices Γ and the parameterization of the respective model sets. Knowledge concerning these relations is important to understand the asymptotic behavior of PML estimators and pseudo likelihood ratio tests based on them. In particular, the structure of the closures of M , \overline{M} say, of the considered model set M has to be understood, since the difference $\overline{M} \setminus M$ cannot be avoided when maximizing the pseudo likelihood function⁴. Additionally, the inclusion properties between different sets M_Γ need to be understood, as this knowledge is important for developing hypothesis tests, in particular for developing hypothesis tests for the dimensions of cointegrating spaces. Hypotheses testing, with a focus on the MFI(1) and I(2) cases, is discussed in Section 5, which shows how the parameterization results of the paper can be used to formulate a large number of hypotheses on (static and polynomial) cointegrating relationships as considered in the VAR cointegration literature. This discussion also includes commonly used deterministic components such as intercept, seasonal dummies, and linear trend, as well as restrictions on these components.

The paper is organized as follows: Section 2 briefly reviews VAR and VARMA models with unit roots and cointegration and discusses some of the complications arising in the VARMA case in addition to the complications arising due to the presence of unit roots and cointegration already in the VAR case. Section 3 presents the canonical form and the parameterization based on it, with the discussion starting with the multiple frequency I(1)—MFI(1)—and I(2) cases prior to a discussion of the general case. This section also provides several important definitions like, e.g., of the state space unit root structure. Section 4 contains a detailed discussion concerning the topological structure of the model sets and Section 5 discusses testing of a large number of hypotheses on the cointegrating spaces commonly tested in the cointegration literature. The discussion in Section 5 focuses on the empirically most relevant MFI(1) and I(2) cases and includes the usual deterministic components considered in the literature. Section 6 briefly summarizes and concludes the paper. All proofs are relegated to the Appendices A and B.

Throughout we use the following notation: L denotes the lag operator, i.e., $L(\{x_t\}_{t \in \mathbb{Z}}) := \{x_{t-1}\}_{t \in \mathbb{Z}}$, for brevity written as $Lx_t = x_{t-1}$. For a matrix $\gamma \in \mathbb{C}^{s \times r}$, $\gamma' \in \mathbb{C}^{r \times s}$ denotes its conjugate transpose. For $\gamma \in \mathbb{C}^{s \times r}$ with full column rank $r < s$, we define $\gamma_\perp \in \mathbb{C}^{s \times (s-r)}$ of full column rank such that $\gamma' \gamma_\perp = 0$. I_p denotes the p -dimensional identity matrix, $0_{m \times n}$ the m times n zero matrix. For two matrices $A \in \mathbb{C}^{m \times n}$, $B \in \mathbb{C}^{k \times l}$, $A \otimes B \in \mathbb{C}^{mk \times nl}$ denotes the Kronecker product of A and B . For a complex valued quantity x , $\mathcal{R}(x)$ denotes its real part, $\mathcal{I}(x)$ its imaginary part and \bar{x} its complex conjugate. For a set V , \overline{V} denotes its closure.⁵ For two sets V and W , $V \setminus W$ denotes the difference of V and W , i.e., $\{v \in V : v \notin W\}$. For a square matrix A we denote the spectral radius (i.e., the maximum of the moduli of its eigenvalues) by $\lambda_{|\max|}(A)$ and by $\det(A)$ its determinant.

⁴ Below we often use the term “likelihood” as short form of “likelihood function”.

⁵ We are confident that this dual usage of notation does not lead to confusion.

2. Vector Autoregressive, Vector Autoregressive Moving Average Processes and Parameterizations

In this paper, we define VAR processes $\{y_t\}_{t \in \mathbb{Z}}$, $y_t \in \mathbb{R}^s$, as solution of

$$a(L)y_t = y_t + \sum_{j=1}^p a_j y_{t-j} = \varepsilon_t + \Phi d_t, \quad (1)$$

with $a(L) := I_s + \sum_{j=1}^p a_j L^j$, where $a_j \in \mathbb{R}^{s \times s}$ for $j = 1, \dots, p$, $\Phi \in \mathbb{R}^{s \times m}$, $a_p \neq 0$, a white noise process $\{\varepsilon_t\}_{t \in \mathbb{Z}}$, $\varepsilon_t \in \mathbb{R}^s$, with $\Sigma := \mathbb{E}(\varepsilon_t \varepsilon_t') > 0$ and a vector sequence $\{d_t\}_{t \in \mathbb{Z}}$, $d_t \in \mathbb{R}^m$, comprising deterministic components like, e.g., the intercept, seasonal dummies or a linear trend. Furthermore, we impose the *non-explosiveness* condition $\det a(z) \neq 0$ for all $|z| < 1$, with $a(z) := I_s + \sum_{j=1}^p a_j z^j$ and z denoting a complex variable.⁶

Thus, for given autoregressive order p , with—as defining characteristic of the order— $a_p \neq 0$, the considered class of VAR models with *specified* deterministic components $\{d_t\}_{t \in \mathbb{Z}}$ is given by the set of all polynomial matrices $a(z)$ such that (i) the non-explosiveness condition holds, (ii) $a(0) = I_s$ and (iii) $a_p \neq 0$; together with the set of all matrices $\Phi \in \mathbb{R}^{s \times m}$.

Equivalently, the model class can be characterized by a set of rational matrix functions $k(z) := a(z)^{-1}$, referred to as *transfer functions*, and the input-output description for the deterministic variables, i.e.,

$$\begin{aligned} V_{p,\Phi} &:= V_p \times \mathbb{R}^{s \times m}, \\ V_p &:= \left\{ k(z) = \sum_{j=0}^{\infty} k_j z^j = a(z)^{-1} : a(z) = I_s + \sum_{j=1}^p a_j z^j, \det a(z) \neq 0 \text{ for } |z| < 1, a_p \neq 0 \right\}. \end{aligned}$$

The associated parameter space is $\Theta_{p,\Phi} := \Theta_p \times \mathbb{R}^{sm} \subset \mathbb{R}^{s^2 p + sm}$, where the parameters

$$\theta := [\theta'_a, \theta'_\Phi]' = [\text{vec}(a_1)', \dots, \text{vec}(a_p)', \text{vec}(\Phi)']' \quad (2)$$

are obtained from stacking the entries of the matrices a_j and Φ , respectively.

Remark 1. In the above discussion the parameters, θ_Σ say, describing the variance covariance matrix Σ of ε_t are not considered. These can be easily included, similarly to Φ by, e.g., parameterizing positive definite symmetric $s \times s$ matrices via their lower triangular Cholesky factor. This leads to a parameter space $\Theta_{p,\Phi,\Sigma} \subset \mathbb{R}^{s^2 p + sm + \frac{s(s+1)}{2}}$. We omit θ_Σ for brevity, since typically no cross-parameter restrictions involving parameters corresponding to Σ are considered, whereas as discussed in Section 5 parameter restrictions involving—in this paper in the state space rather than the VAR setting—both elements of Θ_p and Φ , to, e.g., impose the absence of a linear trend in the cointegrating space, are commonly considered in the cointegration literature.⁷ The estimator of the variance covariance matrix Σ often equals the sample variance of suitable residuals $\hat{\varepsilon}_t(\theta)$ from (1), if there are no cross-restrictions between θ and θ_Σ . This holds, e.g., for the Gaussian pseudo maximum likelihood estimator. Thus, explicitly including θ_Σ and Θ_Σ in the discussion would only overload notation without adding any additional insights, given the simple nature of the parameterization of Σ .

⁶ Our definition of VAR processes differs to a certain extent from some widely used definitions in the literature. Given our focus on unit root and cointegration analysis we, unlike Hannan and Deistler (1988), allow for determinantal roots at the unit circle that, as is well known, lead to integrated processes. We also include deterministic components in our definition, i.e., we allow for a special case of exogenous variables, compare also Remark 2 below. There is, however, also a large part of the literature that refers to this setting simply as (cointegrated) vector autoregressive models, see, e.g., Johansen (1995) and Juselius (2006).

⁷ Of course, the statistical properties of the parameter estimators depend in many ways on the deterministic components.

Remark 2. Our consideration of deterministic components is a special case of including exogenous variables. We include exogenous deterministic variables with a static input-output behavior governed solely by the matrix Φ . More general exogenous variables that are dynamically related to the output $\{y_t\}_{t \in \mathbb{Z}}$ could be considered, thereby considering so-called VARX models rather than VAR models, which would necessitate considering in addition to the transfer function $k(z)$ also a transfer function $l(z)$, say, linking the exogenous variables dynamically to the output.

For the VAR case, the fact that the mapping assigning a given transfer function $k(z) \in V_p$, to a parameter vector $\theta_a \in \Theta_p$ —the parameterization—is continuous with continuously differentiable inverse is immediate.⁸ Homeomorphicity of a parameterization is important for the properties of parameter estimators, e.g., the ordinary least squares (OLS) or Gaussian PML estimator, compare the discussion in Hannan and Deistler (1988, Theorem 2.5.3 and Remark 1, p. 65).

For OLS estimation one typically considers the larger set V_p^{OLS} without the non-explosiveness condition and without the assumption $a_p \neq 0$:

$$V_p^{OLS} := \left\{ k(z) = \sum_{j=0}^{\infty} k_j z^j = a(z)^{-1} : a(z) = I_s + \sum_{j=1}^p a_j z^j \right\}.$$

Considering V_p^{OLS} allows for unconstrained optimization. It is well-known that for $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ as given above, the OLS estimator is consistent over the larger set V_p^{OLS} , i.e., without imposing non-explosiveness and also when specifying p too high. Alternatively, and closely related to OLS in the VAR case, the pseudo likelihood can be maximized over $\Theta_{p,\Phi}$. With this approach, maxima respectively suprema can occur at the boundary of the parameter space, i.e., maximization effectively has to consider $\bar{\Theta}_{p,\Phi}$. It is well-known that the PML estimator is consistent for the stable case (cf. Hannan and Deistler 1988, Theorem 4.2.1), but the maximization problem is complicated by the restrictions on the parameter space stemming from the non-explosiveness condition. Avoiding these complications and asymptotic equivalence of OLS and PML in the stable VAR case explains why VAR models are usually estimated by OLS.⁹

To be more explicit, ignore deterministic components for a moment and consider the case where the DGP is a stationary VAR process, i.e., a solution of (1) with $a(z)$ satisfying the stability condition $\det a(z) \neq 0$ for $|z| \leq 1$. Define the corresponding set of stable transfer functions by $V_{p,\bullet}$:

$$V_{p,\bullet} := \left\{ a(z)^{-1} \in V_p : \det a(z) \neq 0 \text{ for } |z| \leq 1, a_p \neq 0 \right\}.$$

Clearly, $V_{p,\bullet}$ is an open subset of V_p . If the DGP is a stationary VAR process, the above-mentioned consistency result of the OLS estimator over V_p^{OLS} implies that the probability that the estimated transfer function, $\hat{k}(z) = \hat{a}(z)^{-1}$ say, is contained in $V_{p,\bullet}$ converges to one as the sample size tends to infinity. Moreover, the asymptotic distribution of the estimated parameters is normal, under appropriate assumptions on $\{\varepsilon_t\}_{t \in \mathbb{Z}}$.

The situation is a bit more involved if the transfer function of the DGP corresponds to a point in the set $\bar{V}_{p,\bullet} \setminus V_{p,\bullet}$, which contains systems with unit roots, i.e., determinantal roots of $a(z)$ on the unit circle, as well as lower order autoregressive systems—with these two cases non-disjoint. The stable lower order case is relatively unproblematic from a statistical perspective. If, e.g., OLS estimation is performed over V_p^{OLS} , while the true model corresponds to an element in $V_{p^*,\bullet}$, with $p^* < p$, the OLS estimator is

⁸ The set V_p is endowed with the pointwise topology T_{pt} , defined in Section 3. For now, in the context of VAR models, it suffices to know that convergence in pointwise topology is equivalent to convergence of the VAR coefficient matrices a_1, \dots, a_p in the Frobenius norm.

⁹ Please note that in case of restricted estimation, i.e., zero restrictions or cross-equation restrictions, OLS is not asymptotically equivalent to PML in general.

still consistent, since $V_{p^*, \bullet} \subset V_p^{OLS}$. Furthermore, standard chi-squared pseudo likelihood ratio test based inference still applies. The integrated case, for a precise definition see the discussion below Definition 1, is a bit more difficult to deal with, as in this case not all parameters are asymptotically normally distributed and nuisance parameters may be present. Consequently, parameterizations that do not take the specific nature of unit root processes into account are not very useful for inference in the unit root case, see, e.g., Sims et al. (1990, Theorem 1). Studying the unit root and cointegration properties is facilitated by resorting to suitable parameterizations that “zoom in on the relevant characteristics”.

In case that the only determinantal root of $a(z)$ on the unit circle is at $z = 1$, the system corresponds to a so-called $I(d)$ process, with the integration order $d > 0$ made precise in Definition 1 below. Consider first the $I(1)$ case: As is well-known, the rank of the matrix $a(1)$ equals the dimension of the cointegrating space given in Definition 3 below—also referred to as the cointegrating rank. Therefore, determination of the rank of this matrix is of key importance. With the parameterization used so far, imposing a certain (maximal) rank on $a(1)$ implies complicated restrictions on the matrices $a_j, j = 1, \dots, p$. This in turn renders the correspondingly restricted optimization unnecessarily complicated and not conducive to develop tests for the cointegrating rank. It is more convenient to consider the so-called *vector error correction model* (VECM) representation of autoregressive processes, discussed in full detail in the monograph Johansen (1995). To this end let us first introduce the differencing operator at frequency $0 \leq \omega \leq \pi$

$$\Delta_\omega := \begin{cases} I_s - 2 \cos(\omega)L + L^2 & \text{for } 0 < \omega < \pi \\ I_s - \cos(\omega)L & \text{for } \omega \in \{0, \pi\} \end{cases} \quad (3)$$

For notational brevity, we omit the dependence on L in $\Delta_\omega(L)$, henceforth denoted as Δ_ω . Using this notation, the $I(1)$ error correction representation is given by

$$\begin{aligned} \Delta_0 y_t &= \Pi y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta_0 y_{t-j} + \varepsilon_t + \Phi d_t \\ &= \alpha \beta' y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta_0 y_{t-j} + \varepsilon_t + \Phi d_t, \end{aligned} \quad (4)$$

with the matrix $\Pi := -a(1) = -(I_s + \sum_{j=1}^p a_j)$ of rank $0 \leq r \leq s$ factorized into the product of two full rank matrices $\alpha, \beta \in \mathbb{R}^{s \times r}$ and $\Gamma_j := \sum_{m=j+1}^p a_m, j = 1, \dots, p - 1$.

This constitutes a reparameterization, where $k(z) \in V_p$ is now represented by the matrices $(\alpha, \beta, \Gamma_1, \dots, \Gamma_{p-1})$ and a corresponding parameter vector $\theta_a^{VECM} \in \Theta_{p,r}^{VECM}$. Please note that stacking the entries of the matrices does not lead to a homeomorphic mapping from V_p to $\Theta_{p,s}^{VECM}$, since for $0 < r \leq s$ the matrices α and β are not identifiable from the product $\alpha \beta'$, since $\alpha \beta' = \alpha M M^{-1} \beta' = \tilde{\alpha} \tilde{\beta}'$ for all regular matrices $M \in \mathbb{R}^{r \times r}$. One way to obtain identifiability is to introduce the restriction $\beta = [I_r, \beta^{*'}]'$, with $\beta^* \in \mathbb{R}^{(s-r) \times r}$ and $\alpha \in \mathbb{R}^{s \times r}$. With this additional restriction the parameter vector θ_a^{VECM} is given by stacking the vectorized matrices $\alpha, \beta^*, \Gamma_1, \dots, \Gamma_{p-1}$, similarly to (2). Then $\Theta_{p,r,\Phi}^{VECM} = \Theta_{p,r}^{VECM} \times \mathbb{R}^{sm} \subset \mathbb{R}^{ps^2 - (s-r)^2 + sm}$. Note for completeness that the normalization of $\beta = [I_r, \beta^{*'}]'$ may necessitate a re-ordering of the variables in $\{y_t\}_{t \in \mathbb{Z}}$ since—without potential reordering—this parameterization implies a restriction of generality as, e.g., processes, where the first variable is integrated, but does not cointegrate with the other variables, cannot be represented.

Define the following sets of transfer functions:

$$\begin{aligned} V_{p,r} &:= \left\{ a(z)^{-1} \in V_p : \det a(z) \neq 0 \text{ for } \{z : |z| = 1, z \neq 1\}, \text{rank}(a(1)) \leq r \right\}, \\ V_{p,r}^{RRR} &:= \left\{ a(z)^{-1} \in V_p^{OLS} : \text{rank}(a(1)) \leq r \right\}. \end{aligned}$$

The dimension of the parameter vector θ_a^{VECM} depends on the dimension of the cointegrating space, thus the parameterization of $k(z) \in V_{p,r}$ depends on r . The so-called reduced rank regression (RRR) estimator, given by the maximizer of the pseudo likelihood over $V_{p,r}^{RRR}$ is consistent, see, e.g., Johansen (1995, chp. 6). The RRR estimator uses an “implicit” normalization of β and thereby implicitly addresses the mentioned identification problem. However, for testing hypotheses involving the free parameters in α or β , typically the identifying assumption given above is used, as discussed in Johansen (1995, chp. 7).

Furthermore, since $V_{p,r} \subset V_{p,r^*}$ for $r < r^* \leq s$, with $\Theta_{p,r}^{VECM}$ a lower dimensional subset of Θ_{p,r^*}^{VECM} , pseudo likelihood ratio testing can be used to sequentially test for the rank r , starting with the hypothesis of a rank $r = 0$ against the alternative of a rank $0 < r \leq s$, and increasing the assumed rank consecutively until the null hypothesis is not rejected.

Ensuring that $\{y_t\}_{t \in \mathbb{Z}}$ generated from (4) is indeed an I(1) process, requires on the one hand that Π is of reduced rank, i.e., $r < s$ and on the other that the matrix

$$\alpha'_\perp \Gamma \beta_\perp := \alpha'_\perp \left(I_s - \sum_{j=1}^{p-1} \Gamma_j \right) \beta_\perp \tag{5}$$

has full rank. It is well-known that condition (5) is fulfilled on the complement of a “thin” algebraic subset of $V_{p,r}^{RRR}$, and is therefore, ignored in estimation, as it is “generically” fulfilled.¹⁰

The I(2) case is similar in structure to the I(1) case, but with two rank restrictions and one full rank condition to exclude even higher integration orders. The corresponding VECM is given by

$$\Delta_0^2 y_t = \alpha \beta' y_{t-1} - \Gamma \Delta_0 y_{t-1} + \sum_{j=1}^{p-2} \Psi_j \Delta_0^2 y_{t-j} + \varepsilon_t, \tag{6}$$

with α, β as defined in (4), Γ as defined in (5) and $\Psi_j := -\sum_{k=j+1}^{p-1} \Gamma_k$, $j = 1, \dots, p - 2$. From (5) we already know that reduced rank of

$$\alpha'_\perp \Gamma \beta_\perp =: \xi \eta', \tag{7}$$

with $\xi, \eta \in \mathbb{R}^{(s-r) \times m}$, $m < s - r$ is required for higher integration orders. The condition for the corresponding solution process $\{y_t\}_{t \in \mathbb{Z}}$ to be an I(2) process is given by full rank of

$$\tilde{\zeta}'_\perp \alpha'_\perp \left(\Gamma \beta (\beta' \beta)^{-1} (\alpha' \alpha)^{-1} \alpha' \Gamma + I_s - \sum_{j=1}^{p-2} \Psi_j \right) \beta_\perp \eta_\perp,$$

which again is typically ignored in estimation, just like condition (5) in the I(1) case. Thus, I(2) processes correspond to a “thin subset” of $V_{p,r}^{RRR}$, which in turn constitutes a “thin subset” of V_p^{OLS} . The fact that integrated processes correspond to “thin sets” in V_p^{OLS} implies that obtaining estimated systems with specific integration and cointegration properties requires restricted estimation based on parameterizations tailor made to highlight these properties.

Already for the I(2) case, formulating parameterizations that allow conveniently studying the integration and cointegration properties is a quite challenging task. Johansen (1997) contains several different (re-)parameterizations for the I(2) case and Paruolo (1996) defines “integration indices”, r_0, r_1, r_2 say, as the number of columns of the matrices $\beta \in \mathbb{R}^{s \times r_0}$, $\beta_1 := \beta_\perp \eta \in \mathbb{R}^{s \times r_1}$ and $\beta_2 := \beta_\perp \eta_\perp \in \mathbb{R}^{s \times r_2}$. Clearly, the indices r_0, r_1, r_2 are linked to the ranks of the above matrices Π and $\alpha'_\perp \Gamma \beta_\perp$, as $r_0 = r$ and $r_1 = m$ and the columns of $[\beta, \beta_1, \beta_2]$ form a basis of \mathbb{R}^s , such that $s = r_0 + r_1 + r_2$.

¹⁰ A similar property holds for $V_{p,r}^{RRR}$ being a “thin” subset of V_p^{OLS} . This implies that the probability that the OLS estimator calculated over V_p^{OLS} corresponds to an element $V_{p,r}^{RRR} \subset V_p^{OLS}$ is equal to zero in general.

It holds that $\{\beta'_2 y_t\}_{t \in \mathbb{Z}}$ is an I(2) process without cointegration and $\{\beta'_1 y_t\}_{t \in \mathbb{Z}}$ is an I(1) process without cointegration. The process $\{\beta'_1 y_t\}_{t \in \mathbb{Z}}$ is typically I(1) and in this case cointegrates with $\{\beta'_2 \Delta_0 y_t\}_{t \in \mathbb{Z}}$ to stationarity. Thus, there is a direct correspondence of these indices to the dimensions of the different cointegrating spaces—both static and dynamic (with precise definitions given below in Definition 3).¹¹ Please note that again, as already before in the I(1) case, different values of the integration indices r_0, r_1, r_2 , lead to parameter spaces of different dimensions. Furthermore, in these parameterizations matrices describing different cointegrating spaces are (i) not identified and (ii) linked by restrictions, compare the discussion in Paruolo (2000, sct. 2.2) and (7). These facts render the analysis of the cointegration properties in I(2) VAR systems complicated. Also, in the I(2) VAR case usually some forms of RRR estimators are considered over suitable subsets $V_{p,r,m}^{RRR}$ of $V_{p,r}^{RRR}$, again based on implicit normalizations. Inference, however, again requires one to consider parameterizations explicitly.

Estimation and inference issues are fundamentally more complex in the VARMA case than in the VAR case. This stems from the fact that unrestricted estimation—unlike in the VAR case—is not possible due to a lack of identification, as discussed below. This means that in the VARMA case identification and parameterization issues need to be tackled as the first step, compare the discussion in Hannan and Deistler (1988, chp. 2).

In this paper, we consider VARMA processes as solutions of the vector difference equation

$$y_t + \sum_{j=1}^p a_j y_{t-j} = \varepsilon_t + \sum_{j=1}^q b_j \varepsilon_{t-j} + \Phi d_t,$$

with $a(L) := I_s + \sum_{j=1}^p a_j L^j$, where $a_j \in \mathbb{R}^{s \times s}$ for $j = 1, \dots, p$, $a_p \neq 0$ and the non-explosiveness condition $\det(a(z)) \neq 0$ for $|z| < 1$. Similarly, $b(L) := I_s + \sum_{j=1}^q b_j L^j$, where $b_j \in \mathbb{R}^{s \times s}$ for $j = 1, \dots, q$, $b_q \neq 0$ and $\Phi \in \mathbb{R}^{s \times m}$. The transfer function corresponding to a VARMA process is $k(z) := a(z)^{-1} b(z)$.

It is well-known that without further restrictions the VARMA realization $(a(z), b(z))$ of the transfer function $k(z) = a(z)^{-1} b(z)$ is not identified, i.e., different pairs of polynomial matrices $(a(z), b(z))$ can realize the same transfer function $k(z)$. It is clear that $k(z) = a(z)^{-1} m(z)^{-1} m(z) b(z) = a(z)^{-1} b(z)$ for all non-singular polynomial matrices $m(z)$. Thus, the mapping π attaching the transfer function $k(z) = a(z)^{-1} b(z)$ to the pair of polynomial matrices $(a(z), b(z))$ is not injective.¹²

Consequently, we refer for given rational transfer function $k(z)$ to the class $\{(a(z), b(z)) : k(z) = a(z)^{-1} b(z)\}$ as a class of *observationally equivalent* VARMA realizations of $k(z)$. To achieve identification requires to define a canonical form, selecting one member of each class of observationally equivalent VARMA realizations for a set of considered transfer functions. A first step towards a canonical form is to only consider *left coprime* pairs $(a(z), b(z))$.¹³ However, left coprimeness is not sufficient for identification and thus further restrictions are required, leading to parameter vectors of smaller dimension than $\mathbb{R}^{s^2(p+q)}$. A widely used canonical form is the (reverse) echelon canonical form, see Hannan and Deistler (1988, Theorem 2.5.1, p. 59), based on (monic) normalizations of the diagonal elements of $a(z)$ and degree relationships between diagonal and off-diagonal elements as well as the entries in $b(z)$, which lead to zero restrictions. The (reverse) echelon canonical form in conjunction with a transformation to an error correction model was used in VARMA cointegration analysis in the I(1) case, e.g., in Poskitt (2006, Theorem 4.1), but, as for the VAR case, understanding the interdependencies of rank conditions already becomes complicated once one moves to the I(2) case.

¹¹ Below Example 3 we clarify how these indices are related to the state space unit root structure defined in Bauer and Wagner (2012, Definition 2) and link these to the dimensions of the cointegrating spaces in Section 5.2.

¹² Uniqueness of realizations in the VAR case stems from the normalization $m(z)b(z) = I_s$, which reduces the class of observationally equivalent VAR realizations of the same transfer function $k(z) = a(z)^{-1} b(z)$, with $b(z) = I_s$, to a singleton.

¹³ The pair $(a(z), b(z))$ is left coprime if all its left divisors are unimodular matrices. Unimodular matrices are polynomial matrices with constant non-zero determinant. Thus, pre-multiplication of, e.g., $a(z)$ with a unimodular matrix $u(z)$ does not affect the determinantal roots that shape the dynamic behavior of the solutions of VAR models.

In the VARMA case matters are further complicated by another well-known problem that makes statistical analysis considerably more involved compared to the VAR case. Although there exists a generalization of the autoregressive order to the VARMA case, such that any transfer function corresponding to a VARMA system has an order $n \in \mathbb{N}$ (with the precise definition given in the next section) it is known since [Hazewinkel and Kalman \(1976\)](#) that no continuous parameterization of all rational transfer functions of order n exists if $s > 1$. Therefore, if one wants to keep the above-discussed advantages that continuity of a parameterization provides, the set of transfer functions of order n , henceforth referred to as M_n , has to be partitioned into sets on which continuous parameterizations exist, i.e., $M_n = \bigcup_{\Gamma \in G} M_\Gamma$, for some index set G , as already mentioned in the introduction.¹⁴ For any given partitioning of the set M_n it is important to understand the relationships between the different subsets M_Γ , as well as the closures of the pieces M_Γ , since in case of misspecification of M_Γ points in $\overline{M_\Gamma} \setminus M_\Gamma$ cannot be avoided even asymptotically in, e.g., pseudo maximum likelihood estimation. These are more complicated issues in the VARMA case than in the VAR case, see the discussion in [Hannan and Deistler \(1988, Remark 1 after Theorem 2.5.3\)](#).

Based on these considerations, the following section provides and discusses a parameterization that focuses on unit root and cointegration properties, resorting to the state space framework that—as mentioned in the introduction—provides advantages for cointegration analysis. In particular, we derive an almost everywhere homeomorphic parameterization, based on partitioning the set of all considered transfer functions according to a multi-index Γ that contains, among other elements, the state space unit root structure. This implies that certain cointegration properties are invariant for all systems corresponding to a subset M_Γ , i.e., the parameterization allows to directly impose cointegration properties such as the “cointegration indices” of [Paruolo \(1996\)](#) mentioned before.

3. The Canonical Form and the Parameterization

As a first step we define the class of VARMA processes considered in this paper, using the differencing operator defined in (3):

Definition 1. *The s -dimensional real VARMA process $\{y_t\}_{t \in \mathbb{Z}}$ has unit root structure $\Omega := ((\omega_1, h_1), \dots, (\omega_l, h_l))$ with $0 \leq \omega_1 < \omega_2 < \dots < \omega_l \leq \pi, h_k \in \mathbb{N}, k = 1, \dots, l, l \geq 1$, if it is a solution of the difference equation*

$$\Delta_\Omega(y_t - \Phi d_t) := \prod_{k=1}^l \Delta_{\omega_k}^{h_k}(y_t - \Phi d_t) = v_t, \quad (8)$$

where $\{d_t\}_{t \in \mathbb{Z}}$ is an m -dimensional deterministic sequence, $\Phi \in \mathbb{R}^{s \times m}$ and $\{v_t\}_{t \in \mathbb{Z}}$ is a linearly regular stationary VARMA process, i.e., there exists a pair of left coprime matrix polynomials $(a(z), b(z))$, $\det a(z) \neq 0$, $|z| \leq 1$ such that $v_t = a(L)^{-1}b(L)(\varepsilon_t) =: c(L)(\varepsilon_t)$ for a white noise process $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ with $\mathbb{E}(\varepsilon_t \varepsilon_t') = \Sigma > 0$, with furthermore $c(z) \neq 0$ for $z = e^{i\omega_k}, k = 1, \dots, l$.

- The process $\{y_t\}_{t \in \mathbb{Z}}$ is called unit root process with unit roots $z_k := e^{i\omega_k}$ for $k = 1, \dots, l$, the set $F(\Omega) := \{\omega_1, \dots, \omega_l\}$ is the set of unit root frequencies and the integers $h_k, k = 1, \dots, l$ are the integration orders.
- A unit root process with unit root structure $((0, d))$, $d \in \mathbb{N}$, is an I(d) process.
- A unit root process with unit root structure $((\omega_1, 1), \dots, (\omega_l, 1))$ is an MFI(1), process.

A linearly regular stationary VARMA process has empty unit root structure $\Omega_0 := \{\}$.

¹⁴ When using the echelon canonical form, the partitioning is according to the so-called *Kronecker indices* related to a basis selection for the row-space of the *Hankel* matrix corresponding to the transfer function $k(z)$, see, e.g., [Hannan and Deistler \(1988, chp. 2.4\)](#) for a precise definition.

As discussed in [Bauer and Wagner \(2012\)](#) the state space framework is convenient for the analysis of VARMA unit root processes. Detailed treatments of the state space framework are given in [Hannan and Deistler \(1988\)](#) and—in the context of unit root processes—[Bauer and Wagner \(2012\)](#).

A state space representation of a unit root VARMA process is¹⁵

$$\begin{aligned} y_t &= Cx_t + \Phi d_t + \varepsilon_t, \\ x_{t+1} &= Ax_t + B\varepsilon_t, \end{aligned} \tag{9}$$

for a white noise process $\{\varepsilon_t\}_{t \in \mathbb{Z}}$, $\varepsilon_t \in \mathbb{R}^s$, a deterministic process $\{d_t\}_{t \in \mathbb{Z}}$, $d_t \in \mathbb{R}^m$ and the unobserved state process $\{x_t\}_{t \in \mathbb{Z}}$, $x_t \in \mathbb{C}^n$, $A \in \mathbb{C}^{n \times n}$, $B \in \mathbb{C}^{n \times s}$, $C \in \mathbb{C}^{s \times n}$ and $\Phi \in \mathbb{R}^{s \times m}$.

Remark 3. [Bauer and Wagner \(2012, Theorem 2\)](#) show that every real valued unit root VARMA process $\{y_t\}_{t \in \mathbb{Z}}$ as given in (8) has a real valued state space representation with $\{x_t\}_{t \in \mathbb{Z}}$ real valued and real valued system matrices (A, B, C) . Considering complex valued state space representations in (9) is merely for algebraic convenience, as in general some eigenvalues of A are complex valued. Note for completeness that [Bauer and Wagner \(2012\)](#) contains a detailed discussion why considering the A -matrix in the canonical form in (up to reordering) the Jordan normal form is useful for cointegration analysis. For the sake of brevity we abstain from including this discussion again in the present paper. The key aspect of this construction is its usefulness for cointegration analysis, which becomes visible in [Remark 4](#), where the “simple” unit root properties of blocks of the state vector are discussed.

The transfer function $k(z)$ with real valued power series coefficients corresponding to a real valued unit root process $\{y_t\}_{t \in \mathbb{Z}}$ as given in [Definition 1](#) is given by the rational matrix function $k(z) = \Delta_\Omega(z)^{-1}a(z)^{-1}b(z)$. The (possibly complex valued) matrix triple (A, B, C) realizes the transfer function $k(z)$ if and only if $\pi(A, B, C) := I_s + zC(I_n - zA)^{-1}B = k(z)$. Please note that as for VARMA realizations, for a transfer function $k(z)$ there exist multiple state space realizations (A, B, C) , with possibly different state dimensions n . A state space system (A, B, C) is *minimal* if there exists no state space system of lower state dimension realizing the same transfer function $k(z)$. The *order* of the transfer function $k(z)$ is the state dimension of a minimal system (A, B, C) realizing $k(z)$.

All minimal state space realizations of a transfer function $k(z)$ only differ in the basis of the state (cf. [Hannan and Deistler 1988, Theorem 2.3.4](#)), i.e., $\pi(A, B, C) = \pi(\tilde{A}, \tilde{B}, \tilde{C})$ for two minimal state space systems (A, B, C) and $(\tilde{A}, \tilde{B}, \tilde{C})$ is equivalent to the existence of a regular matrix $T \in \mathbb{C}^n$ such that $A = T\tilde{A}T^{-1}, B = T\tilde{B}, C = \tilde{C}T^{-1}$. Thus, the matrices A and \tilde{A} are similar for all minimal realizations of a transfer function $k(z)$.

By imposing restrictions on the matrices of a minimal state space system (A, B, C) realizing $k(z)$, [Bauer and Wagner \(2012, Theorem 2\)](#) provide a canonical form, i.e., a mapping of the set M_n of transfer functions with real valued power series coefficients defined below onto unique state space realizations (A, B, C) . The set M_n is defined as

$$M_n := \left\{ k(z) = \pi(A, B, C) \mid \begin{array}{l} \lambda_{|\max|}(A) \leq 1, \\ A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times s}, C \in \mathbb{R}^{s \times n}, (A, B, C) \text{ minimal} \end{array} \right\}.$$

To describe the necessary restrictions of the canonical form the following definition is useful:

¹⁵ Here and below we will only consider state space systems in so-called innovation representation, with the same error in both the output equation and the state equation. Since every state space system has an innovation representation this is no restriction, compare [Aoki \(1990, chp. 7.1\)](#).

Definition 2. A matrix $B = [b_{i,j}]_{i=1,\dots,c,j=1,\dots,s} \in \mathbb{C}^{c \times s}$ is positive upper triangular (p.u.t.) if there exist integers $1 \leq j_1 \leq j_2 \leq \dots \leq j_c \leq s$, such that for $j_i \leq s$ we have $b_{i,j} = 0, j < j_i, j_i < j_{i+1}, b_{i,j_i} \in \mathbb{R}^+$; i.e., B is of the form

$$B = \begin{bmatrix} 0 & \dots & 0 & b_{1,j_1} & * & \dots & * \\ 0 & & \dots & & 0 & b_{2,j_2} & * \\ \vdots & & & & & & \\ 0 & & & \dots & & 0 & b_{c,j_c} & * \end{bmatrix},$$

where the symbol $*$ indicates unrestricted complex-valued entries.

A unique state space realization of $k(z) \in M_n$ is given as follows (cf. Bauer and Wagner 2012, Theorem 2):

Theorem 1. For every transfer function $k(z) \in M_n$ there exists a unique minimal (complex) state space realization (A, B, C) such that

$$\begin{aligned} y_t &= Cx_{t,\mathbb{C}} + \Phi d_t + \varepsilon_t, \\ x_{t+1,\mathbb{C}} &= Ax_{t,\mathbb{C}} + B\varepsilon_t \end{aligned}$$

with:

(i) $A := \text{diag}(A_u, A_\bullet) := \text{diag}(A_{1,\mathbb{C}}, \dots, A_{l,\mathbb{C}}, A_\bullet)$, $A_u \in \mathbb{C}^{n_u \times n_u}, A_\bullet \in \mathbb{R}^{n_\bullet \times n_\bullet}$, where it holds for $k = 1, \dots, l$ that

– for $0 < \omega_k < \pi$:

$$A_{k,\mathbb{C}} := \begin{bmatrix} J_k & 0 \\ 0 & \bar{J}_k \end{bmatrix} \in \mathbb{C}^{2d^k \times 2d^k},$$

– for $\omega_k \in \{0, \pi\}$:

$$A_{k,\mathbb{C}} := J_k \in \mathbb{R}^{d^k \times d^k},$$

with

$$J_k := \begin{bmatrix} \bar{z}_k I_{d_1^k} & [I_{d_1^k}, 0_{d_1^k \times (d_2^k - d_1^k)}] & 0 & \dots & 0 \\ 0_{d_2^k \times d_1^k} & \bar{z}_k I_{d_2^k} & [I_{d_2^k}, 0_{d_2^k \times (d_3^k - d_2^k)}] & 0 & \vdots \\ 0 & 0 & \bar{z}_k I_{d_3^k} & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & [I_{d_{h_k-1}^k}, 0_{d_{h_k-1}^k \times (d_{h_k}^k - d_{h_k-1}^k)}] \\ 0 & 0 & \dots & 0 & \bar{z}_k I_{d_{h_k}^k} \end{bmatrix}, \tag{10}$$

where $0 < d_1^k \leq d_2^k \leq \dots \leq d_{h_k}^k$.

(ii) $B := [B'_u, B'_\bullet]' := [B'_{1,\mathbb{C}}, \dots, B'_{l,\mathbb{C}}, B'_\bullet]'$ and $C := [C_u, C_\bullet] := [C_{1,\mathbb{C}}, \dots, C_{l,\mathbb{C}}, C_\bullet]$ are partitioned accordingly. It holds for $k = 1, \dots, l$ that

– for $0 < \omega_k < \pi$:

$$B_{k,\mathbb{C}} := \begin{bmatrix} B_k \\ \bar{B}_k \end{bmatrix} \in \mathbb{C}^{2d^k \times s} \text{ and } C_{k,\mathbb{C}} := [C_k, \bar{C}_k] \in \mathbb{C}^{s \times 2d^k}.$$

– for $\omega_k \in \{0, \pi\}$:

$$\mathcal{B}_{k,\mathbb{C}} := \mathcal{B}_k \in \mathbb{R}^{d^k \times s} \text{ and } \mathcal{C}_{k,\mathbb{C}} := \mathcal{C}_k \in \mathbb{R}^{s \times d^k}.$$

- (iii) Partitioning \mathcal{B}_{k,h_k} in $\mathcal{B}_k = [\mathcal{B}'_{k,1}, \dots, \mathcal{B}'_{k,h_k}]'$ as $\mathcal{B}_{k,h_k} = [\mathcal{B}'_{k,h_k,1}, \dots, \mathcal{B}'_{k,h_k,h_k}]'$, with $\mathcal{B}_{k,h_k,j} \in \mathbb{C}^{(d_j^k - d_{j-1}^k) \times s}$ it holds that $\mathcal{B}_{k,h_k,j}$ is p.u.t. for $d_j^k > d_{j-1}^k$ for $j = 1, \dots, h_k$ and $k = 1, \dots, l$.
- (iv) For $k = 1, \dots, l$ define $\mathcal{C}_k = [\mathcal{C}_{k,1}, \mathcal{C}_{k,2}, \dots, \mathcal{C}_{k,h_k}]$, $\mathcal{C}_{k,j} = [\mathcal{C}_{k,j}^G, \mathcal{C}_{k,j}^E]$, with $\mathcal{C}_{k,j}^E \in \mathbb{C}^{s \times (d_j^k - d_{j-1}^k)}$ and $\mathcal{C}_{k,j}^G \in \mathbb{C}^{s \times d_{j-1}^k}$ for $j = 1, \dots, h_k$, with $d_0^k := 0$. Furthermore, define $\mathcal{C}_k^E := [\mathcal{C}_{k,1}^E, \dots, \mathcal{C}_{k,h_k}^E] \in \mathbb{C}^{s \times d_{h_k}^k}$. It holds that $(\mathcal{C}_k^E)' \mathcal{C}_k^E = I_{d_{h_k}^k}$ and $(\mathcal{C}_{k,j}^G)' \mathcal{C}_{k,i}^E = 0$ for $1 \leq i \leq j$ for $j = 2, \dots, h_k$ and $k = 1, \dots, l$.
- (v) $\lambda_{|\max|}(\mathcal{A}_\bullet) < 1$ and the stable subsystem $(\mathcal{A}_\bullet, \mathcal{B}_\bullet, \mathcal{C}_\bullet)$ of state dimension $n_\bullet = n - n_u$ is in echelon canonical form (cf. Hannan and Deistler 1988, Theorem 2.5.2).

Remark 4. As indicated in Remark 3 and discussed in detail in Bauer and Wagner (2012) considering complex valued quantities is merely for algebraic convenience. For econometric analysis, interest is, of course, on real valued quantities. These can be straightforwardly obtained from the representation given in Theorem 1 as follows. First define a transformation matrix (and its inverse):

$$T_{\mathbb{R},d} := \left[I_d \otimes \begin{bmatrix} 1 \\ i \end{bmatrix}, I_d \otimes \begin{bmatrix} 1 \\ -i \end{bmatrix} \right] \in \mathbb{C}^{2d \times 2d}, \quad T_{\mathbb{R},d}^{-1} := \frac{1}{2} \begin{bmatrix} I_d \otimes [1, -i] \\ I_d \otimes [1, i] \end{bmatrix}.$$

Starting from the complex valued canonical representation $(\mathcal{A}, \mathcal{B}, \mathcal{C})$, a real valued canonical representation

$$\begin{aligned} y_t &= C_{\mathbb{R}} x_{t,\mathbb{R}} + \Phi d_t + \varepsilon_t, \\ x_{t+1,\mathbb{R}} &= A_{\mathbb{R}} x_{t,\mathbb{R}} + B_{\mathbb{R}} \varepsilon_t, \end{aligned}$$

with real valued matrices $(A_{\mathbb{R}}, B_{\mathbb{R}}, C_{\mathbb{R}})$ follows from using the just defined transformation matrix. In particular it holds that:

$$\begin{aligned} A_{\mathbb{R}} &:= \text{diag}(A_{u,\mathbb{R}}, A_\bullet) := \text{diag}(A_{1,\mathbb{R}}, \dots, A_{l,\mathbb{R}}, A_\bullet), \\ B_{\mathbb{R}} &:= [\mathcal{B}'_{u,\mathbb{R}}, \mathcal{B}'_\bullet]' := [\mathcal{B}'_{1,\mathbb{R}}, \dots, \mathcal{B}'_{l,\mathbb{R}}, \mathcal{B}'_\bullet]', \\ C_{\mathbb{R}} &:= [\mathcal{C}_{u,\mathbb{R}}, \mathcal{C}_\bullet] := [\mathcal{C}_{1,\mathbb{R}}, \dots, \mathcal{C}_{l,\mathbb{R}}, \mathcal{C}_\bullet], \end{aligned}$$

with

$$(A_{k,\mathbb{R}}, B_{k,\mathbb{R}}, C_{k,\mathbb{R}}) := \begin{cases} (T_{\mathbb{R},d^k} \mathcal{A}_k T_{\mathbb{R},d^k}^{-1}, T_{\mathbb{R},d^k} \mathcal{B}_k, \mathcal{C}_k T_{\mathbb{R},d^k}^{-1}) & \text{if } 0 < \omega_k < \pi, \\ (\mathcal{A}_k, \mathcal{B}_k, \mathcal{C}_k) & \text{if } \omega_k \in \{0, \pi\}. \end{cases}$$

Before we turn to the real valued state process corresponding to the real valued canonical representation, we first consider the complex valued state process $\{x_{t,\mathbb{C}}\}_{t \in \mathbb{Z}}$ in more detail. This process is partitioned according to the partitioning of the matrices $\mathcal{C}_{k,\mathbb{C}}$ into $x_{t,\mathbb{C}} := [x'_{t,u}, x'_{t,\bullet}]' := [x'_{t,1,\mathbb{C}}, \dots, x'_{t,l,\mathbb{C}}, x'_{t,\bullet}]'$, where

$$x_{t,k,\mathbb{C}} := \begin{cases} [x'_{t,k}, \bar{x}'_{t,k}]' & \text{if } 0 < \omega_k < \pi, \\ x_{t,k} & \text{if } \omega_k \in \{0, \pi\}, \end{cases}$$

with

$$x_{t+1,k} = J_k x_{t,k} + B_k \varepsilon_t, \quad \text{for } k = 1, \dots, l.$$

For $k = 1, \dots, l$ the sub-vectors $x_{t,k}$ are further decomposed into $x_{t,k} := [(x_{t,k}^1)', \dots, (x_{t,k}^{h_k})']'$, with $x_{t,k}^j \in \mathbb{C}^{d_j^k}$ for $j = 1, \dots, h_k$ according to the partitioning $C_k = [C_{k,1}, \dots, C_{k,h_k}]$.

The partitioning of the complex valued process $\{x_{t,C}\}_{t \in \mathbb{Z}}$ leads to an analogous partitioning of the real valued state process $\{x_{t,\mathbb{R}}\}_{t \in \mathbb{Z}}$, $x_{t,\mathbb{R}} := [x'_{t,u,\mathbb{R}}, x'_{t,\bullet}]' := [x'_{t,1,\mathbb{R}}, \dots, x'_{t,l,\mathbb{R}}, x'_{t,\bullet}]'$, obtained from

$$x_{t,k,\mathbb{R}} := \begin{cases} T_{\mathbb{R},d^k} x_{t,k,C} & \text{if } 0 < \omega_k < \pi, \\ x_{t,k} & \text{if } \omega_k \in \{0, \pi\}, \end{cases}$$

with the corresponding block of the state equation given by

$$x_{t+1,k,\mathbb{R}} = A_{k,\mathbb{R}} x_{t,k,\mathbb{R}} + B_{k,\mathbb{R}} \varepsilon_t.$$

For $k = 1, \dots, l$ the sub-vectors $x_{t,k,\mathbb{R}}$ are further decomposed into $x_{t,k,\mathbb{R}} := [(x_{t,k,\mathbb{R}}^1)', \dots, (x_{t,k,\mathbb{R}}^{h_k})']'$, with $x_{t,k,\mathbb{R}}^j \in \mathbb{R}^{2d_j^k}$ if $0 < \omega_k < \pi$ and $x_{t,k,\mathbb{R}}^j \in \mathbb{R}^{d_j^k}$ if $\omega_k \in \{0, \pi\}$ for $j = 1, \dots, h_k$ and $C_{k,\mathbb{R}} := [C_{k,1,\mathbb{R}}, \dots, C_{k,h_k,\mathbb{R}}]$ decomposed accordingly.

Bauer and Wagner (2012, Theorem 3, p. 1328) show that the processes $\{x_{t,k,\mathbb{R}}^j\}_{t \in \mathbb{Z}}$ have unit root structure $((\omega_k, h_k - j + 1))$ for $j = 1, \dots, h_k$ and $k = 1, \dots, l$. Furthermore, for $j = 1, \dots, h_k$ and $k = 1, \dots, l$ the processes $\{x_{t,k,\mathbb{R}}^j\}_{t \in \mathbb{Z}}$ are not cointegrated, as defined in Definition 3 below. For $\omega_k = 0$, the process $\{x_{t,k,\mathbb{R}}^j\}_{t \in \mathbb{Z}}$ is the d_j^k -dimensional process of stochastic trends of order $h_1 - j + 1$, while the $2d_j^k$ components of $\{x_{t,k,\mathbb{R}}^j\}_{t \in \mathbb{Z}}$, for $0 < \omega_k < \pi$, and the d_j^k components of $\{x_{t,l,\mathbb{R}}^j\}_{t \in \mathbb{Z}}$, for $\omega_k = \pi$, are referred to as stochastic cycles of order $h_k - j + 1$ at their corresponding frequencies ω_k .

Remark 5. Parameterizing the stable part of the transfer function using the echelon canonical form is merely one possible choice. Any other canonical form of the stable subsystem and suitable parameterization based on it can be used instead for the stable subsystem.

Remark 6. Starting from a state space system (9) with matrices (A, B, C) in canonical form, a solution for $y_t, t > 0$ (with the solution for $t < 0$ obtained completely analogously)—for some $x_1 = [x'_{1,u}, x'_{1,\bullet}]'$ —is given by

$$y_t = \sum_{j=1}^{t-1} C_u A_u^{j-1} B_u \varepsilon_{t-j} + C_u A_u^{t-1} x_{1,u} + \sum_{j=1}^{t-1} C_{\bullet} A_{\bullet}^{j-1} B_{\bullet} \varepsilon_{t-j} + C_{\bullet} A_{\bullet}^{t-1} x_{1,\bullet} + \Phi d_t + \varepsilon_t.$$

Clearly, the term $C_u A_u^{t-1} x_{1,u}$ is stochastically singular and is effectively like a deterministic component, which may lead to an identification problem with Φd_t . If, the deterministic component Φd_t is rich enough to “absorb” $C_u A_u^{t-1} x_{1,u}$, then one solution of the identification problem is to set $x_{1,u} = 0$. Rich enough here means, e.g., in the $I(1)$ case with $A_u = I$ that d_t contains an intercept. Analogously, in the $MFI(1)$ case d_t has to contain seasonal dummy variables corresponding to all unit root frequencies. The term $C_{\bullet} A_{\bullet}^{t-1} x_{1,\bullet}$ decays exponentially and, therefore, does not impact the asymptotic properties of any statistical procedure. It is, therefore, inconsequential for statistical analysis but convenient (with respect to our definition of unit root processes) to set $x_{1,\bullet} = \sum_{j=1}^{\infty} A_{\bullet}^{j-1} B_{\bullet} \varepsilon_{1-j}$. This corresponds to the steady state or stationary solution of the stable block of the state equation, and renders $\{x_{t,\bullet}\}_{t \in \mathbb{N}}$ or, when the solution on \mathbb{Z} is considered, $\{x_{t,\bullet}\}_{t \in \mathbb{Z}}$ stationary. Please note that these issues with respect to starting values, potential identification problems and their impact or non-impact on statistical procedures also occur in the VAR setting.

Bauer and Wagner (2012, Theorem 2) show that minimality of the canonical state space realization (A, B, C) implies full row rank of the p.u.t. blocks $B_{k,h_k,j}$ of B_{k,h_k} . In addition to proposing the canonical form, Bauer and Wagner (2012) also provide details how to transform any minimal state space realization into canonical form: Given a minimal state space system (A, B, C) realizing the transfer function $k(z) \in M_n$, the first step is to find a similarity transformation T such that $\tilde{A} = TAT^{-1}$

is of the form given in (10) by using an eigenvalue decomposition, compare Chatelin (1993). In the second step the corresponding subsystem $(\tilde{A}_\bullet, \tilde{B}_\bullet, \tilde{C}_\bullet)$ is transformed to echelon canonical form as described in Hannan and Deistler (1988, chp. 2). These two transformations do not lead to a unique realization, because the restrictions on \mathcal{A} do not uniquely determine the *unstable subsystem* $(\mathcal{A}_u, \mathcal{B}_u, \mathcal{C}_u)$.

For example, in the case $\Omega = ((\omega_1, h_1)) = ((0, 1))$, $n_\bullet = 0$, $d_1^1 < s$, such that $(I_{d_1^1}, \mathcal{B}_1, \mathcal{C}_1)$ is a corresponding state space system, the same transfer function $k(z) = I_s + z\mathcal{C}_1(1 - z)^{-1}\mathcal{B}_1 = I_s + \mathcal{C}_1\mathcal{B}_1z(1 - z)^{-1}$ is realized also by all systems $(I_{d_1^1}, T\mathcal{B}_1, \mathcal{C}_1T^{-1})$, with some regular matrix $T \in \mathbb{C}^{d_1^1 \times d_1^1}$. To find a unique realization the product $\mathcal{C}_1\mathcal{B}_1$ needs to be uniquely decomposed into factors \mathcal{C}_1 and \mathcal{B}_1 . This is achieved by performing a QR decomposition of $\mathcal{C}_1\mathcal{B}_1$ (without pivoting) that leads to $\mathcal{C}_1'\mathcal{C}_1 = I$. The additional restriction of \mathcal{B}_1 being a p.u.t. matrix of full row rank then leads to a unique factorization of $\mathcal{C}_1\mathcal{B}_1$ into \mathcal{C}_1 and \mathcal{B}_1 . In the general case with an arbitrary unit root structure Ω , similar arguments lead to p.u.t. restrictions on sub-blocks $\mathcal{B}_{k,h_k,j}$ in \mathcal{B}_u and orthogonality restrictions on sub-blocks of \mathcal{C}_u .

The canonical form introduced in Theorem 1 was designed to be useful for cointegration analysis. To see this, first requires a definition of static and polynomial cointegration (cf. Bauer and Wagner 2012, Definitions 3 and 4).

Definition 3.

(i) Let $\tilde{\Omega} = ((\tilde{\omega}_1, \tilde{h}_1), \dots, (\tilde{\omega}_l, \tilde{h}_l))$ and $\Omega = ((\omega_1, h_1), \dots, (\omega_l, h_l))$ be two unit root structures. Then $\tilde{\Omega} \preceq \Omega$ if

- $F(\tilde{\Omega}) \subseteq F(\Omega)$.
- For all $\omega \in F(\tilde{\Omega})$ for \tilde{k} and k such that $\tilde{\omega}_{\tilde{k}} = \omega_k = \omega$ it holds that $\tilde{h}_{\tilde{k}} \leq h_k$.

Furthermore, $\tilde{\Omega} \prec \Omega$ if $\tilde{\Omega} \preceq \Omega$ and $\tilde{\Omega} \neq \Omega$. For two unit root structures $\tilde{\Omega} \preceq \Omega$ define the decrease $\delta_k(\Omega, \tilde{\Omega})$ of the integration order at frequency ω_k , for $k = 1, \dots, l$, as

$$\delta_k(\Omega, \tilde{\Omega}) := \begin{cases} h_k - \tilde{h}_{\tilde{k}} & \exists \tilde{k} : \tilde{\omega}_{\tilde{k}} = \omega_k \in F(\tilde{\Omega}), \\ h_k & \omega_k \notin F(\tilde{\Omega}) \end{cases}.$$

(ii) An s -dimensional unit root process $\{y_t\}_{t \in \mathbb{Z}}$ with unit root structure Ω is cointegrated of order $(\Omega, \tilde{\Omega})$, where $\tilde{\Omega} \prec \Omega$, if there exists a vector $\beta \in \mathbb{R}^s, \beta \neq 0$, such that $\{\beta'y_t\}_{t \in \mathbb{Z}}$ has unit root structure $\tilde{\Omega}$. In this case the vector β is a cointegrating vector (CIV) of order $(\Omega, \tilde{\Omega})$.

(iii) All CIVs of order $(\Omega, \tilde{\Omega})$ span the (static) cointegrating space of order $(\Omega, \tilde{\Omega})$.¹⁶

(iv) An s -dimensional unit root process $\{y_t\}_{t \in \mathbb{Z}}$ with unit root structure Ω is polynomially cointegrated of order $(\Omega, \tilde{\Omega})$, where $\tilde{\Omega} \prec \Omega$, if there exists a vector polynomial $\beta(z) = \sum_{m=0}^q \beta_m z^m, \beta_m \in \mathbb{R}^s, m = 0, \dots, q, \beta_q \neq 0$, for some integer $1 \leq q < \infty$ such that

- $\beta(L)'(\{y_t\}_{t \in \mathbb{Z}})$ has unit root structure $\tilde{\Omega}$,
- $\max_{k=1, \dots, l} \|\beta(e^{i\omega_k})\| \delta_k(\Omega, \tilde{\Omega}) \neq 0$.

In this case the vector polynomial $\beta(z)$ is a polynomial cointegrating vector (PCIV) of order $(\Omega, \tilde{\Omega})$.

(v) All PCIVs of order $(\Omega, \tilde{\Omega})$ span the polynomial cointegrating space of order $(\Omega, \tilde{\Omega})$.

¹⁶ The definition of cointegrating spaces as linear subspaces allows to characterize them by a basis and implies a well-defined dimension. These advantages, however, have the implication that the zero vector is an element of all cointegrating spaces, despite not being a cointegrating vector in our definition, where the zero vector is excluded. This issue is well-known of course in the cointegration literature.

Remark 7.

- (i) It is merely a matter of taste whether cointegrating spaces are defined in terms of their order $(\Omega, \tilde{\Omega})$ or their decrease $\delta(\Omega, \tilde{\Omega}) := (\delta_1(\Omega, \tilde{\Omega}), \dots, \delta_l(\Omega, \tilde{\Omega}))$, with $\delta_k(\Omega, \tilde{\Omega})$ as defined above. Specifying Ω and $\delta(\Omega, \tilde{\Omega})$ contains the same information as providing the order of (polynomial) cointegration.
- (ii) Notwithstanding the fact that CIVs and PCIVs in general may lead to changes of the integration orders at different unit root frequencies it may be of interest to “zoom in” on only one unit root frequency ω_k , thereby leaving the potential reductions of the integration orders at other unit root frequencies unspecified. This allows to—entirely similarly as in Definition 3—define cointegrating and polynomial cointegrating spaces of different orders at a single unit root frequency ω_k . Analogously one can also define cointegrating and polynomial cointegrating spaces of different orders for subsets of the frequencies in $F(\Omega)$.
- (iii) In principle the polynomial cointegrating spaces defined so far are infinite-dimensional as the polynomial degree is not bounded. However, since every polynomial vector $\beta(z)$ can be written as $\beta_0(z) + \beta_\Omega(z)\Delta_\Omega(z)$, where by definition $\{\Delta_\Omega y_t\}_{t \in \mathbb{Z}}$ has empty unit root structure, it suffices to consider PCIVs of polynomial degree smaller than the polynomial degree of $\Delta_\Omega(z)$. This shows that it is sufficient to consider finite dimensional polynomial cointegrating spaces. When considering, as in item (ii), (polynomial) cointegration only for one unit root it similarly suffices to consider polynomials of maximal degree equal to $h_k - 1$ for real unit roots and $2h_k - 1$ for complex unit roots. Thus, in the I(2) case it suffices to consider polynomials of degree one.
- (iv) The argument about maximal relevant polynomial degrees given in item (iii) can be made more precise and combined with the decrease in Ω achieved. Every polynomial vector $\beta(z)$ can be written as $\beta_0(z) + \beta_{\omega_k, \delta_k}(z)\Delta_{\omega_k}^{\delta_k}(z)$ for $\delta_k = 1, \dots, h_k$. By definition it holds that $\{\Delta_{\omega_k}^{\delta_k} y_t\}_{t \in \mathbb{Z}}$ has integration order $h_k - \delta_k$ at frequency ω_k . Thus, it suffices to consider PCIVs of polynomial degree smaller than δ_k for $\omega_k \in \{0, \pi\}$ or $2\delta_k$ for $0 < \omega_k < \pi$ when considering the polynomial cointegrating space at ω_k with decrease δ_k . In the MFI(1) case therefore, when considering only one unit root frequency, again only polynomials of degree one need to be considered. This space is often referred to in the literature as dynamic cointegration space.

To illustrate the advantages of the canonical form for cointegration analysis consider

$$y_t = \sum_{k=1}^l \sum_{j=1}^{h_k} C_{k,j,\mathbb{R}} x_{t,k,\mathbb{R}}^j + C_\bullet x_{t,\bullet} + \Phi d_t + \varepsilon_t.$$

By Remark 4, the process $\{x_{t,k,\mathbb{R}}^j\}_{t \in \mathbb{Z}}$ is not cointegrated. This implies that $\beta \in \mathbb{R}^s, \beta \neq 0$, reduces the integration order at unit root z_k to $h_k - j$ if and only if $\beta' [C_{k,1,\mathbb{R}}, \dots, C_{k,j,\mathbb{R}}] = 0$ and $\beta' C_{k,j+1,\mathbb{R}} \neq 0$ or equivalently $\beta' [C_{k,1}, \dots, C_{k,j}] = 0$ and $\beta' C_{k,j+1} \neq 0$ (using the transformation to the complex matrices of the canonical form, as discussed in Remark 4, and that $\beta' [C_k, \bar{C}_k] = 0$ if and only if $\beta' C_k = 0$). Thus, the CIVs are characterized by orthogonality to sub-blocks of C_u .

The real valued representation given in Remark 4 used in its partitioned form just above immediately leads to necessary orthogonality constraint for polynomial cointegration of degree one:

$$\begin{aligned} \beta(L)'(y_t) &= \beta(L)'(C_{u,\mathbb{R}} x_{t,u,\mathbb{R}} + C_\bullet x_{t,\bullet} + \Phi d_t + \varepsilon_t) \\ &= \beta'_0 C_{u,\mathbb{R}} x_{t,u,\mathbb{R}} + \beta'_1 C_{u,\mathbb{R}} x_{t-1,u,\mathbb{R}} + \beta(L)'(C_\bullet x_{t,\bullet} + \Phi d_t + \varepsilon_t) \\ &= \beta'_0 C_{u,\mathbb{R}} (A_{u,\mathbb{R}} x_{t-1,u,\mathbb{R}} + B_{u,\mathbb{R}} \varepsilon_{t-1}) + \beta'_1 C_{u,\mathbb{R}} x_{t-1,u,\mathbb{R}} + \beta(L)'(C_\bullet x_{t,\bullet} + \Phi d_t + \varepsilon_t) \\ &= (\beta'_0 C_{u,\mathbb{R}} A_{u,\mathbb{R}} + \beta'_1 C_{u,\mathbb{R}}) x_{t-1,u,\mathbb{R}} + \beta'_0 C_{u,\mathbb{R}} B_{u,\mathbb{R}} \varepsilon_{t-1} + \beta(L)'(C_\bullet x_{t,\bullet} + \Phi d_t + \varepsilon_t) \\ &= (\beta'_0 C_u A_u + \beta'_1 C_u) x_{t-1,u} + \beta'_0 C_u B_u \varepsilon_{t-1} + \beta(L)'(C_\bullet x_{t,\bullet} + \Phi d_t + \varepsilon_t) \end{aligned}$$

follows. Since all terms except the first are stationary or deterministic, a necessary condition for a reduction of the unit root structure is the orthogonality of $[\beta'_0 \quad \beta'_1]'$ to sub-blocks of $\begin{bmatrix} C_{u,\mathbb{R}} A_{u,\mathbb{R}} \\ C_{u,\mathbb{R}} \end{bmatrix}$ or

sub-blocks of the complex matrix $\begin{bmatrix} C_u A_u \\ C_u \end{bmatrix}$. Please note, however, that this orthogonality condition is not sufficient for $[\beta'_0, \beta'_1]'$ to be a PCIV, because it does not imply $\max_{k=1, \dots, J} \|\beta(e^{i\omega_k})\| \delta_k(\Omega, \tilde{\Omega}) \neq 0$. For a detailed discussion of polynomial cointegration, when considering also higher polynomial degrees, see Bauer and Wagner (2012, sct. 5).

The following examples illustrate cointegration analysis in the state space framework for the empirically most relevant, i.e., the I(1), MFI(1) and I(2) cases.

Example 1 (Cointegration in the I(1) case). *In the I(1) case, neglecting the stable subsystem and the deterministic components for simplicity, it holds that*

$$\begin{aligned} y_t &= C_1 x_{t,1} + \varepsilon_t, & y_t, \varepsilon_t &\in \mathbb{R}^s, x_{t,1} \in \mathbb{R}^{d_1}, C_1 \in \mathbb{R}^{s \times d_1}, \\ x_{t+1,1} &= x_{t,1} + B_1 \varepsilon_t, & B_1 &\in \mathbb{R}^{d_1 \times s}. \end{aligned}$$

The vector $\beta \in \mathbb{R}^s, \beta \neq 0$, is a CIV of order $((0, 1), \{\})$ if and only if $\beta' C_1 = 0$.

Example 2 (Cointegration in the MFI(1) case with complex unit root z_k). *In the MFI(1) case with unit root structure $\Omega = ((\omega_k, 1))$ and complex unit root z_k , neglecting the stable subsystem and the deterministic components for simplicity, it holds that*

$$\begin{aligned} y_t &= C_{k,\mathbb{R}} x_{t,k,\mathbb{R}} + \varepsilon_t \\ &= \begin{bmatrix} C_k & \bar{C}_k \end{bmatrix} \begin{bmatrix} x_{t,k} \\ \bar{x}_{t,k} \end{bmatrix} + \varepsilon_t, \\ y_t, \varepsilon_t &\in \mathbb{R}^s, x_{t,k,\mathbb{R}} \in \mathbb{R}^{2d_1^k}, x_{t,k} \in \mathbb{C}^{d_1^k}, C_{k,\mathbb{R}} \in \mathbb{R}^{s \times 2d_1^k}, C_k \in \mathbb{C}^{s \times d_1^k}, \\ \begin{bmatrix} x_{t+1,k} \\ \bar{x}_{t+1,k} \end{bmatrix} &= \begin{bmatrix} \bar{z}_k I_{d_1^k} & 0 \\ 0 & z_k I_{d_1^k} \end{bmatrix} \begin{bmatrix} x_{t,k} \\ \bar{x}_{t,k} \end{bmatrix} + \begin{bmatrix} B_k \\ \bar{B}_k \end{bmatrix} \varepsilon_t, & B_k \in \mathbb{C}^{d_1^k \times s}. \end{aligned}$$

The vector $\beta \in \mathbb{R}^s, \beta \neq 0$, is a CIV of order $(\Omega, \{\})$ if and only if

$$\beta' C_k = 0 \text{ (and thus } \beta' \bar{C}_k = 0 \text{)}.$$

The vector polynomial $\beta(z) = \beta_0 + \beta_1 z$, with $\beta_0, \beta_1 \in \mathbb{R}^s, [\beta'_0, \beta'_1]' \neq 0$, is a PCIV of order $(\Omega, \{\})$ if and only if

$$[\beta'_0, \beta'_1] \begin{bmatrix} \bar{z}_k C_k & z_k \bar{C}_k \\ C_k & \bar{C}_k \end{bmatrix} = 0, \tag{11}$$

which is equivalent to

$$(\bar{z}_k \beta'_0 + \beta'_1) C_k = 0.$$

The fact that the matrix in (11) has a block structure with two blocks of conjugate complex columns implies some additional structure also on the space of PCIVs, here with polynomial degree one. More specifically it holds that if $\beta_0 + \beta_1 z$ is a PCIV of order $(\Omega, \{\})$, also $-\beta_1 + (\beta_0 + 2 \cos(\omega_k) \beta_1) z$ is a PCIV of order $(\Omega, \{\})$. This follows from

$$\begin{aligned} (\bar{z}_k (-\beta_1)' + (\beta_0 + 2 \cos(\omega_k) \beta_1)') C_k &= (\beta'_0 + (2\mathcal{R}(z_k) - \bar{z}_k) \beta'_1) C_k \\ &= (\beta'_0 + z_k \beta'_1) C_k \\ &= z_k (\bar{z}_k \beta'_0 + \beta'_1) C_k = 0. \end{aligned}$$

Thus, the space of PCIVs of degree (up to) one inherits some additional structure emanating from the occurrence of complex eigenvalues in complex conjugate pairs.

Example 3 (Cointegration in the I(2) case). In the I(2) case, neglecting the stable subsystem and the deterministic components for simplicity, it holds that

$$\begin{aligned} y_t &= C_{1,1}^E x_{t,1}^E + C_{1,2}^G x_{t,2}^G + C_{1,2}^E x_{t,2}^E + \varepsilon_t, \\ y_t, \varepsilon_t &\in \mathbb{R}^s, x_{t,1}^E, x_{t,2}^G \in \mathbb{R}^{d_1^1}, x_{t,2}^E \in \mathbb{R}^{d_2^1 - d_1^1}, C_{1,1}^E, C_{1,2}^G \in \mathbb{R}^{s \times d_1^1}, C_{1,2}^E \in \mathbb{R}^{s \times (d_2^1 - d_1^1)}, \\ x_{t+1,1}^E &= x_{t,1}^E + x_{t,2}^G + \mathcal{B}_{1,1} \varepsilon_t, \\ x_{t+1,2}^G &= x_{t,2}^G + \mathcal{B}_{1,2,1} \varepsilon_t, \\ x_{t+1,2}^E &= x_{t,2}^E + \mathcal{B}_{1,2,2} \varepsilon_t, \quad \mathcal{B}_{1,1} \in \mathbb{R}^{d_1^1 \times s}, \mathcal{B}_{1,2,1} \in \mathbb{R}^{d_1^1 \times s}, \mathcal{B}_{1,2,2} \in \mathbb{R}^{(d_2^1 - d_1^1) \times s}. \end{aligned}$$

The vector $\beta \in \mathbb{R}^s, \beta \neq 0$ is a CIV of order $((0, 2), (0, 1))$ if and only if

$$\beta' C_{1,1}^E = 0 \quad \text{and} \quad \beta' [C_{1,2}^G, C_{1,2}^E] \neq 0.$$

The vector $\beta \in \mathbb{R}^s, \beta \neq 0$, is a CIV of order $((0, 2), \{\})$ if and only if

$$\beta' [C_{1,1}^E, C_{1,2}^G, C_{1,2}^E] = 0.$$

The vector polynomial $\beta(z) = \beta_0 + \beta_1 z$, with $\beta_0, \beta_1 \in \mathbb{R}^s$ is a PCIV of order $((0, 2), \{\})$ if and only if

$$[\beta_0', \beta_1'] \begin{bmatrix} C_{1,1}^E & C_{1,1}^E + C_{1,2}^G & C_{1,2}^E \\ C_{1,1}^E & C_{1,2}^G & C_{1,2}^E \end{bmatrix} = 0 \quad \text{and} \quad \beta(1) = \beta_0 + \beta_1 \neq 0.$$

The above orthogonality constraint indicates that the two cases $C_{1,2}^G = 0$ and $C_{1,2}^G \neq 0$ have to be considered separately for polynomial cointegration analysis. Consider first the case $C_{1,2}^G = 0$. In this case the orthogonality constraints imply $\beta_0' C_{1,1}^E = 0, \beta_1' C_{1,1}^E = 0$ and $(\beta_0 + \beta_1)' C_{1,2}^E = 0$. Thus, the vector $\beta_0 + \beta_1$ is a CIV of order $((0, 2), \{\})$ and therefore $\beta(z) = \beta_0 + \beta_1 z$ is of “non-minimum” degree, one in this case rather than zero ($\beta_0 + \beta_1$). For a formal definition of minimum degree PCIVs see [Bauer and Wagner \(2003, Definition 4\)](#). In case $C_{1,2}^G \neq 0$ there are PCIVs of degree one that are not simple transformations of static CIVs. Consider $\beta(z) = \beta_0 + \beta_1 z = \gamma_1(1 - z) + \gamma_2$ such that $\{\gamma_1'(y_t - y_{t-1}) + \gamma_2' y_t\}_{t \in \mathbb{Z}}$ is stationary. The integrated contribution to $\{\gamma_1'(y_t - y_{t-1})\}_{t \in \mathbb{Z}}$ is given by $\gamma_1'(1 - L)(\{C_{1,1}^E x_{t,1}^E\}_{t \in \mathbb{Z}}) = \{\gamma_1' C_{1,1}^E x_{t-1,2}^G + \gamma_1' C_{1,1}^E \mathcal{B}_{1,1} \varepsilon_{t-1}\}_{t \in \mathbb{Z}}$, with $\gamma_1' C_{1,1}^E \neq 0$. This term is eliminated by $\{\gamma_2' C_{1,2}^G x_{t,2}^G\}_{t \in \mathbb{Z}}$ in $\{\gamma_2' y_t\}_{t \in \mathbb{Z}}$, if $\gamma_1' C_{1,1}^E + \gamma_2' C_{1,2}^G = 0$, which is only possible if $C_{1,2}^G \neq 0$. Additionally, $\gamma_2' [C_{1,1}^E, C_{1,2}^E] = 0$ needs to hold, such that there is no further integrated contribution to $\{\gamma_2' y_t\}_{t \in \mathbb{Z}}$. Neither γ_1 nor γ_2 are CIVs since both violate the necessary conditions given in the definition of CIVs, which implies that $\beta(z)$ is indeed a “minimum degree” PCIV.

As was shown above, the unit root and cointegration properties of $\{y_t\}_{t \in \mathbb{Z}}$ depend on the sub-blocks of C_u and the eigenvalue structure of A_u . We, therefore, define the more encompassing state space unit root structure containing information on the geometrical and algebraic multiplicities of the eigenvalues of A_u (cf. [Bauer and Wagner 2012, Definition 2](#)).

Definition 4. A unit root process $\{y_t\}_{t \in \mathbb{Z}}$ with a canonical state space representation as given in [Theorem 1](#) has state space unit root structure

$$\Omega_S := \left((\omega_1, d_{h_1}^1, \dots, d_{h_1}^1), \dots, (\omega_l, d_{h_l}^1, \dots, d_{h_l}^1) \right)$$

where $0 \leq d_1^k \leq d_2^k \leq \dots \leq d_{h_k}^k \leq s$ for $k = 1, \dots, l$. For $\{y_t\}_{t \in \mathbb{Z}}$ with empty unit root structure $\Omega_S := \{\}$.

Remark 8. The state space unit root structure Ω_S contains information concerning the integration properties of the process $\{y_t\}_{t \in \mathbb{Z}}$, since the integers d_j^k , $k = 1, \dots, l$, $j = 1, \dots, h_k$ describe (multiplied by two for k such that $0 < \omega_k < \pi$) the numbers of non-cointegrated stochastic trends or cycles of corresponding integration orders, compare again Remark 4. As such, Ω_S describes properties of the stochastic process $\{y_t\}_{t \in \mathbb{Z}}$ —and, therefore, the state space unit root structure Ω_S partitions unit root processes according to these (co-)integration properties. These (co-)integration properties, however, are invariant to a chosen canonical representation, or more generally invariant to whether a VARMA or state space representation is considered. For all minimal state representations of a unit root process $\{y_t\}_{t \in \mathbb{Z}}$ these indices—being related to the Jordan normal form—are invariant.

As mentioned in Section 2, Paruolo (1996, Definition 3) introduces integration indices at frequency zero as a triple of integers (r_0, r_1, r_2) . These correspond to the numbers of columns of the matrices β, β_1, β_2 in the error correction representation of I(2) VAR processes, see, e.g., Johansen (1997, sct. 3). Here, r_2 is the number of stochastic trends of order two, i.e., $r_2 = d_1^1$. Furthermore, r_1 is the number of stochastic trends of order one that do not cointegrate with $\beta_2' \Delta_0 \{y_t\}_{t \in \mathbb{Z}}$ and hence $r_1 = d_2^1 - d_1^1$. Therefore, the integration indices at frequency zero are in one-one correspondence with the state space unit root structure $\Omega_S = ((0, d_1^1, d_2^1))$ for I(2) processes and the dimension $s = r_0 + r_1 + r_2$ of the process.

The canonical form given in Theorem 1 imposes p.u.t. structures on sub-blocks of the matrix \mathcal{B}_u . The occurrence of these blocks—related to $d_j^k > d_{j-1}^k$ —is determined by the state space unit root structure Ω_S . The number of free entries in these p.u.t.-blocks, however, is not determined by Ω_S . Consequently, we need structure indices $p \in \mathbb{N}_0^{n_u}$ indicating for each row the position of a potentially restricted positive element, as formalized below:

Definition 5 (Structure indices). For the block $\mathcal{B}_u \in \mathbb{C}^{n_u \times s}$ of the matrix \mathcal{B} of a state space realization $(\mathcal{A}, \mathcal{B}, \mathcal{C})$ in canonical form, define the corresponding structure indices $p \in \mathbb{N}_0^{n_u}$ as

$$p_i := \begin{cases} 0 & \text{if the } i\text{-th row of } \mathcal{B}_u \text{ is not part of a p.u.t. block,} \\ j & \text{if the } i\text{-th row of } \mathcal{B}_u \text{ is part of a p.u.t. block and its } j\text{-th entry is restricted to be positive.} \end{cases}$$

Remark 9. Since sub-blocks of \mathcal{B}_u corresponding to complex unit roots are of the form $\mathcal{B}_{k,\mathbb{C}} = [\mathcal{B}'_k, \overline{\mathcal{B}}'_k]'$, the entries restricted to be positive are located in the same columns and rows of both \mathcal{B}_k and $\overline{\mathcal{B}}_k$. Thus, the structure indices p_i of the corresponding rows are identical for \mathcal{B}_k and $\overline{\mathcal{B}}_k$. Therefore, it would be possible to omit the parts of p corresponding to the blocks $\overline{\mathcal{B}}_k$. It is, however, as will be seen in Definition 9, advantageous for the comparison of unit root structures and structure indices that p is a vector with n_u entries.

Example 4. Consider the following state space system:

$$\begin{aligned} y_t &= [C_{1,1}^E \quad C_{1,2}^G \quad C_{1,2}^E] x_t + \varepsilon_t \quad y_t, \varepsilon_t \in \mathbb{R}^2, x_t \in \mathbb{R}^3, C_{1,1}^E, C_{1,2}^G, C_{1,2}^E \in \mathbb{R}^{2 \times 1} \quad (12) \\ x_{t+1} &= \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} x_t + \begin{bmatrix} \mathcal{B}_{1,1} \\ \mathcal{B}_{1,2,1} \\ \mathcal{B}_{1,2,2} \end{bmatrix} \varepsilon_t, \quad x_0 = 0, \quad \mathcal{B}_{1,1}, \mathcal{B}_{1,2,1}, \mathcal{B}_{1,2,2} \in \mathbb{R}^{1 \times 2}. \end{aligned}$$

In canonical form $\mathcal{B}_{1,2,1}$ and $\mathcal{B}_{1,2,2}$ are p.u.t. matrices and $\mathcal{B}_{1,1}$ is unrestricted. If, e.g., the second entry $b_{1,2,1,2}$ of $\mathcal{B}_{1,2,1}$ and the first entry $b_{1,2,2,1}$ of $\mathcal{B}_{1,2,2}$ are restricted to be positive, then

$$\mathcal{B} = \begin{bmatrix} * & * \\ 0 & b_{1,2,1,2} \\ b_{1,2,2,1} & * \end{bmatrix},$$

where the symbol $*$ denotes unrestricted entries. In this case $p = [0, 2, 1]'$.

For given state space unit root structure Ω_S the matrix \mathcal{A}_u is fully determined. The parameterization of the set of feasible matrices \mathcal{B}_u for given structure indices p and of the set of stable subsystems $(\mathcal{A}_\bullet, \mathcal{B}_\bullet, \mathcal{C}_\bullet)$ for given Kronecker indices α_\bullet (cf. Hannan and Deistler 1988, chp. 2.) is straightforward, since the entries in these matrices are either unrestricted, restricted to zero or restricted to be positive. Matters are a bit more complicated for \mathcal{C}_u . One possibility to parameterize the set of possible matrices \mathcal{C}_u for a given state space unit root structure Ω_S is to use real and complex valued Givens rotations (cf. Golub and van Loan 1996, chp. 5.1).

Definition 6 (Real Givens rotation). *The real Givens rotation $R_{q,i,j}(\theta) \in \mathbb{R}^{q \times q}$, $\theta \in [0, 2\pi)$ is defined as*

$$R_{q,i,j}(\theta) := \begin{bmatrix} I_{i-1} & & & & 0 \\ & \cos(\theta) & 0 & \sin(\theta) & \\ & 0 & I_{j-1-i} & 0 & \\ & -\sin(\theta) & 0 & \cos(\theta) & \\ 0 & & & & I_{q-j} \end{bmatrix}.$$

Remark 10. *Givens rotations allow transforming any vector $v = [v_1, v_2, \dots, v_q]' \in \mathbb{R}^q$ into a vector of the form $[\tilde{v}_1, 0, \dots, 0]'$ with $\tilde{v}_1 \geq 0$. This is achieved by the following algorithm:*

1. Set $j = 1$, $v_1^{(1)} = v_1$ and $v^{(1)} = v$.
2. Represent $[v_1^{(j)}, v_{q-j+1}]'$ using polar coordinates as $[v_1^{(j)}, v_{q-j+1}]' = [r_j \cos(\theta_{q-j}), r_j \sin(\theta_{q-j})]'$, with $r_j \geq 0$ and $\theta_{q-j} \in [0, 2\pi)$. If $r_j = 0$, set $\theta_{q-j} = 0$ (cf. Otto 2011, chp. 1.5.3, p. 39). Then $R_{2,1,2}(\theta_{q-j})[v_1^{(j)}, v_{q-j+1}]' = [v_1^{(j+1)}, 0]'$ such that $v^{(j+1)} = R_{q,1,q-j+1}(\theta_{q-j})v^{(j)} = [v_1^{(j+1)}, v_2, \dots, v_{q-j}, 0, \dots, 0]'$, with $v_1^{(j+1)} \geq 0$.
3. If $j = q - 1$, stop. Else increment j by one ($j \rightarrow j + 1$) and continue at step 2.

This algorithm determines a unique vector $\theta = [\theta_1, \dots, \theta_{q-1}]'$ for every vector $v \in \mathbb{R}^q$.

Remark 11. *The determinant of real Givens rotations is equal to one, i.e., $\det(R_{s,i,j}(\theta)) = 1$ for all $s, i, j \in \mathbb{N}$ and all $\theta \in [0, 2\pi)$. Thus, it is not possible to factorize an orthonormal matrix Q with $\det(Q) = -1$ into a product of Givens rotations. This obvious fact has implications for the parameterization of \mathcal{C} -matrices as is detailed below.*

Definition 7 (Complex Givens rotation). *The complex Givens rotation $Q_{q,i,j}(\varphi) \in \mathbb{C}^{q \times q}$, $\varphi := [\varphi_1, \varphi_2]' \in \Theta_{\mathbb{C}} := [0, \pi/2) \times [0, 2\pi)$, is defined as*

$$Q_{q,i,j}(\varphi) := \begin{bmatrix} I_{i-1} & & & & 0 \\ & \cos(\varphi_1) & 0 & \sin(\varphi_1)e^{i\varphi_2} & \\ & 0 & I_{j-1-i} & 0 & \\ & -\sin(\varphi_1)e^{-i\varphi_2} & 0 & \cos(\varphi_1) & \\ 0 & & & & I_{q-j} \end{bmatrix}.$$

Remark 12. *Complex Givens rotations allow transforming any vector $v = [v_1, v_2, \dots, v_q]' \in \mathbb{C}^q$ into a vector of the form $[\tilde{v}_1, 0, \dots, 0]'$ with $\tilde{v}_1 \in \mathbb{C}$. This is achieved by the following algorithm:*

1. Set $j = 1$, $v_1^{(1)} = v_1$ and $v^{(1)} = v$.
2. Represent $[v_1^{(j)}, v_{q-j+1}]'$ using polar coordinates as $[v_1^{(j)}, v_{q-j+1}]' = [a_j e^{i\varphi_{a,j}}, b_j e^{i\varphi_{b,j}}]'$, with $a_j, b_j \geq 0$ and $\varphi_{a,j}, \varphi_{b,j} \in [0, 2\pi)$. If $v_1^{(j)} = 0$, set $\varphi_{a,j} = 0$ and if $v_{q-j+1} = 0$, set $\varphi_{b,j} = 0$ (cf. Otto 2011, chp. 8.1.3, p. 222).

3. Set

$$\varphi_{q-j,1} = \begin{cases} \tan^{-1} \left(\frac{b_j}{a_j} \right) & \text{if } a_j > 0, \\ \pi/2 & \text{if } a_j = 0, b_j > 0, \\ 0 & \text{if } a_j = 0, b_j = 0, \end{cases}$$

$$\varphi_{q-j,2} = \varphi_{a,j} - \varphi_{b,j} \text{ mod } 2\pi.$$

Then $Q_{2,1,2}(\varphi_{q-j})[v_1^{(j)}, v_{q-j+1}]' = [v_1^{(j+1)}, 0]'$ such that $v^{(j+1)} = Q_{q,1,q-j+1}(\theta_{q-1})v^{(j)} = [v_1^{(j+1)}, v_2, \dots, v_{q-j}, 0]'$, with $v_1^{(j+1)} \in \mathbb{C}$.

4. If $j = q - 1$, stop. Else increment j by one ($j \rightarrow j + 1$) and continue at step 2.

This algorithm determines a unique vector $\varphi = [\varphi_{1,1}, \varphi_{1,2}, \dots, \varphi_{q-1,2}]'$ for every vector $v \in \mathbb{C}^q$.

To set the stage for the general case, we start the discussion of the parameterization of the set of matrices (A, B, C) in canonical form with the MFI(1) and I(2) cases. These two cases display all ingredients required later for the general case. The MFI(1) case illustrates the usage of either real or complex Givens rotations, depending on whether the considered C -block corresponds to a real or complex unit root. The I(2) case highlights recursive orthogonality constraints on the parameters of the C -block, which are related to the polynomial cointegration properties (cf. Example 3).

3.1. The Parameterization in the MFI(1) Case

The state space unit root structure of an MFI(1) process is given by $\Omega_S = ((\omega_1, d_1^1), \dots, (\omega_l, d_1^l))$. For the corresponding state space system (A, B, C) in canonical form, the sub-blocks of A_u are equal to $J_k = \bar{z}_k I_{d_1^k}$, the sub-blocks B_k of B_u are p.u.t. and $C_k' C_k = I_{d_1^k}$, for $k = 1, \dots, l$.

Starting with the sub-blocks of C_u , it is convenient to separate the discussion of the parameterization of C_u -blocks into the real case, where $\omega_k \in \{0, \pi\}$ and $C_k \in \mathbb{R}^{s \times d_1^k}$, and the complex case with $0 < \omega_k < \pi$ and $C_k \in \mathbb{C}^{s \times d_1^k}$. For the case of real unit roots the two cases $d_1^k < s$ and $d_1^k = s$ have to be distinguished. For brevity of notation refer to the considered real block simply as $C \in \mathbb{R}^{s \times d}$. Using this notation, the set of matrices to be parameterized is

$$O_{s,d} := \{C \in \mathbb{R}^{s \times d} | C' C = I_d\}.$$

The parameterization of $O_{s,d}$ is based on the combination of real Givens rotations, as given in Definition 6, that allow transforming every matrix in $O_{s,d}$ to the form $[I_d, 0'_{(s-d) \times d}]'$ for $d < s$. For $d = s$, Givens rotations allow transforming every matrix $C \in O_{s,s}$ either to I_s or $I_s^- := \text{diag}(I_{s-1}, -1)$, since, compare Remark 11, for the transformed matrix $\tilde{C}^{(s)}$ it holds that $\det(C) = \det(\tilde{C}^{(s)}) \in \{-1, 1\}$. This is achieved with the following algorithm:

1. Set $j = 1$ and $C^{(1)} = C$.
2. Transform the entries $[c_{j,j}, \dots, c_{j,d}]$ in the j -th row of $C^{(j)}$, to $[\tilde{c}_{j,j}, 0, \dots, 0]$, $\tilde{c}_{j,j} \geq 0$. Since this is a row vector, this is achieved by right-multiplication of $C^{(j)}$ with transposed Givens rotations and the required parameters are obtained via the algorithm described in Remark 10. The first $j - 1$ entries of the j -th row remain unchanged. Denote the transformed matrix by $C^{(j+1)}$.
3. If $j = d - 1$ stop. Else increment j by one ($j \rightarrow j + 1$) and continue at step 2.
4. Collect all parameters used for the Givens rotations in steps 1 to 3 in a parameter vector θ_R . Steps 1–3 correspond to a QR decomposition of $C' = Q\tilde{C}'$, with an orthonormal matrix Q given by the product of the Givens rotations. Please note that the first $j - 1$ entries of the j -th column of $\tilde{C} = C^{(d)}$ are equal to zero by construction.
5. Set $j = 0$ and $C^{(0)} = \tilde{C}$.

6. Collect the entries in column $d - j$ of $\tilde{C}^{(j)}$ which have not been transformed to zero by previous transformations into the vector $[c_{d-j,d-j}, c_{d+1,d-j}, \dots, c_{s,d-j}]'$. Using the algorithm described in Remark 10 transform this vector to $[\tilde{c}_{d-j,d-j}, 0, \dots, 0]'$ by left-multiplication of $\tilde{C}^{(j)}$ with Givens rotations. Since Givens rotations are orthonormal, the transformed matrix $\tilde{C}^{(j+1)}$ is still orthonormal implying for its entries $\tilde{c}_{d-j,d-j} = 1$ and $\tilde{c}_{i,d-j} = 0$ for all $i < d - j$. An exception occurs if $d = s$. In this case $c_{d-j,d-j} \in \{-1, 1\}$ and no Givens rotations are defined.
7. If $j = d - 1$ stop. Else increment j by one ($j \rightarrow j + 1$) and continue at step 6.
8. Collect all parameters used for the Givens rotations in steps 5 to 7 in a parameter vector θ_L .

The parameter vector $\theta = [\theta'_L, \theta'_R]'$, contains the angles of the employed Givens rotations and provides one way of parameterizing $O_{s,d}$. The following Lemma 1 demonstrates the usefulness of this parameterization.

Lemma 1 (Properties of the parameterization of $O_{s,d}$). Define for $d \leq s$ a mapping $\theta \rightarrow C_O(\theta)$ from $\Theta_{\mathbb{O}}^{\mathbb{R}} := [0, 2\pi)^{d(s-d)} \times [0, 2\pi)^{d(d-1)/2} \rightarrow O_{s,d}$ by

$$\begin{aligned}
 C_O(\theta) &:= \left[\prod_{i=1}^d \prod_{j=1}^{s-d} R_{s,i,d+j}(\theta_{L,(s-d)(i-1)+j}) \right]' \begin{bmatrix} I_d \\ 0_{(s-d) \times d} \end{bmatrix} \left[\prod_{i=1}^{d-1} \prod_{j=1}^i R_{d,d-i,d-i+j}(\theta_{R,i(i-1)/2+j}) \right] \\
 &:= R_L(\theta_L)' \begin{bmatrix} I_d \\ 0_{(s-d) \times d} \end{bmatrix} R_R(\theta_R),
 \end{aligned}$$

with $\theta := [\theta'_L, \theta'_R]'$, where $\theta_L := [\theta_{L,1}, \dots, \theta_{L,d(s-d)}]'$ and $\theta_R := [\theta_{R,1}, \dots, \theta_{R,d(d-1)/2}]'$. The following properties hold:

- (i) $O_{s,d}$ is closed and bounded.
- (ii) The mapping $C_O(\cdot)$ is infinitely often differentiable.

For $d < s$, it holds that

- (iii) For every $C \in O_{s,d}$ there exists a vector $\theta \in \Theta_{\mathbb{O}}^{\mathbb{R}}$ such that

$$C = C_O(\theta) = R_L(\theta_L)' \begin{bmatrix} I_d \\ 0_{(s-d) \times d} \end{bmatrix} R_R(\theta_R).$$

The algorithm discussed above defines the inverse mapping $C_O^{-1} : O_{s,d} \rightarrow \Theta_{\mathbb{O}}^{\mathbb{R}}$.

- (iv) The inverse mapping $C_O^{-1}(\cdot)$ —the parameterization of $O_{s,d}$ —is infinitely often differentiable on the pre-image of the interior of $\Theta_{\mathbb{O}}^{\mathbb{R}}$. This is an open and dense subset of $O_{s,d}$.

For $d = s$, it holds that

- (v) $O_{s,s}$ is a disconnected space in $\mathbb{R}^{s \times s}$ with two disjoint non-empty closed subsets $O_{s,s}^+ := \{C \in \mathbb{R}^{s \times s} | C'C = I_s, \det(C) = 1\}$ and $O_{s,s}^- := \{C \in \mathbb{R}^{s \times s} | C'C = I_s, \det(C) = -1\}$.
- (vi) For every $C \in O_{s,s}^+$ there exists a vector $\theta \in \Theta_{\mathbb{O}}^{\mathbb{R}}$ such that

$$C = C_O(\theta) = R_L(\theta_L)' \begin{bmatrix} I_d \end{bmatrix} R_R(\theta_R) = R_R(\theta_R).$$

In this case, steps 1-4 of the algorithm discussed above define the inverse mapping $C_O^{-1} : O_{s,s}^+ \rightarrow \Theta_{\mathbb{O}}^{\mathbb{R}}$.

- (vii) Define $v := [\pi, \dots, \pi]' \in \mathbb{R}^{s(s-1)/2}$. Then a parameterization of $O_{s,s}$ is given by

$$C_{\mathbb{O}}^{\pm}(C) = \begin{cases} v + C_O^{-1}(C) & \text{if } C \in O_{s,s}^+ \\ -(v + C_O^{-1}(CI_s^-)) & \text{if } C \in O_{s,s}^- \end{cases}.$$

The parameterization is infinitely often differentiable with infinitely often differentiable inverse on an open and dense subset of $O_{s,s}$.

Remark 13. The following arguments illustrate why C_O^{-1} is not continuous on the pre-image of the boundary of $\Theta_O^{\mathbb{R}}$: Consider the unit sphere $O_{3,1} = \{C \in \mathbb{R}^3 | C'C = \|C\|_2 = 1\}$. One way to parameterize the unit sphere is to use degrees of longitude and latitude. Two types of discontinuities occur: After fixing the location of the zero degree of longitude, i.e., the prime meridian, its anti-meridian is described by both 180° W and 180° E. Using the half-open interval $[0, 2\pi)$ in our parametrization causes a similar discontinuity. Second, the degree of longitude is irrelevant at the north pole. As seen in Remark 10, with our parameterization a similar issue occurs when the first two entries of C to be compared are both equal to zero. In this case the parameter of the Givens rotation is set to zero, although every θ will produce the same result. Both discontinuities clearly occur on a thin subset of $O_{s,d}$.

As in the parametrization of the VAR I(1)-case in the VECM framework, where the restriction $\beta = [I_{s-d}, \beta^*]'$ can only be imposed when the upper $(s - d) \times (s - d)$ block of the true β_0 of the DGP is of full rank (cf. Johansen 1995, chp. 5.2), the set where the discontinuities occur can effectively be changed by a permutation of the components of the observed time series. This corresponds to redefining the locations of the prime meridian and the poles.

Remark 14. Please note that the parameterization partitions the parameter vector θ into two parts $\theta_L \in [0, 2\pi)^{d(s-d)}$ and $\theta_R \in [0, 2\pi)^{(d-1)d/2}$. Since changing the parameter values in θ_R does not change the column space of $C_O(\theta)$, which, as seen above, determines the cointegrating vectors, θ_L fully characterizes the (static) cointegrating space. Please note that the dimension of θ_L is $d(s - d)$ and thus coincides with the number of free parameters in β in the VECM framework (cf. Johansen 1995, chp. 5.2).

Example 5. Consider the matrix

$$C = \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{2} \\ \frac{1}{\sqrt{2}} & \frac{1}{2} \end{bmatrix}$$

with $d = 2$ and $s = 3$. As discussed, the static cointegrating space is characterized by the left kernel of this matrix. The left kernel of a matrix in $\mathbb{R}^{3 \times 2}$ with full rank two is given by a one-dimensional space, with the corresponding basis vector parameterized, when normalized to length one, by two free parameters. Thus, for the characterization of the static cointegrating space two parameters are required, which exactly coincides with the dimension of θ_L given in Remark 14. The parameters in θ_R correspond to the choice of a basis of the image of C . Having fixed the two-dimensional subspace through θ_L , only one free parameter for the choice of an orthonormal basis remains, which again coincides with the dimension given in Remark 14. To obtain the parameter vector, the starting point is a QR decomposition of $C' = R_R(\theta_R)\tilde{C}'$. In this example $R_R(\theta_R) = R_{2,1,2}(\theta_{R,1})$, with $\theta_{R,1}$ to be determined. To find $\theta_{R,1}$, solve $[0 \quad \frac{1}{\sqrt{2}}]R_{2,1,2}(\theta_{R,1})' = [r \quad 0]$ for $r \geq 0$ and $\theta_{R,1} \in [0, 2\pi)$. In other words, find $r \geq 0$ and $\theta_{R,1} \in [0, 2\pi)$ such that $[0 \quad \frac{1}{\sqrt{2}}] = r[\cos(\theta_{R,1}) \quad \sin(\theta_{R,1})]$, which leads to $r = \frac{1}{\sqrt{2}}$, $\theta_{R,1} = \frac{\pi}{2}$. Thus, the orthonormal matrix $R_R(\theta_R)$ is equal to $R_{2,1,2}(\frac{\pi}{2})$ and the transpose of the upper triangular matrix \tilde{C}' is equal to:

$$\tilde{C} = \tilde{C}^{(0)} = C \cdot R_{2,1,2} \left(\frac{\pi}{2} \right)' = \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{2} \\ \frac{1}{\sqrt{2}} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{2} & \frac{1}{\sqrt{2}} \\ \frac{1}{2} & -\frac{1}{\sqrt{2}} \end{bmatrix}.$$

Second, transform the entries in the lower 1×2 -sub-block of $\tilde{C}^{(0)}$ to zero, starting with the last column. For this find $\theta_{L,2} \in [0, 2\pi)$ such that $R_{3,2,3}(\theta_{L,2}) [0 \quad \frac{1}{\sqrt{2}} \quad -\frac{1}{\sqrt{2}}]' = [0 \quad 1 \quad 0]'$, i.e., $[\frac{1}{\sqrt{2}} \quad -\frac{1}{\sqrt{2}}]' = r [\cos(\theta_{L,2}) \quad \sin(\theta_{L,2})]$. This yields $r = 1, \theta_{L,2} = \frac{7\pi}{4}$. Next compute $\tilde{C}^{(1)} = R_{3,2,3}(\frac{7\pi}{4})\tilde{C}^{(0)}$:

$$\tilde{C}^{(1)} = R_{3,2,3} \left(\frac{7\pi}{4} \right) \cdot C \cdot R_{2,1,2} \left(\frac{\pi}{2} \right)' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{2} \\ \frac{1}{\sqrt{2}} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \\ \frac{1}{\sqrt{2}} & 0 \end{bmatrix}.$$

In the final step find $\theta_{L,1} \in [0, 2\pi)$ such that $R_{3,1,3}(\theta_{L,1}) [\frac{1}{\sqrt{2}} \quad 0 \quad \frac{1}{\sqrt{2}}]' = [1 \quad 0 \quad 0]'$, i.e., $[\frac{1}{\sqrt{2}} \quad \frac{1}{\sqrt{2}}]' = r [\cos(\theta_{L,1}) \quad \sin(\theta_{L,1})]$. The solution is $r = 1, \theta_{L,1} = \frac{\pi}{4}$. Combining the transformations leads to

$$R_{3,1,3} \left(\frac{\pi}{4} \right) \cdot R_{3,2,3} \left(\frac{7\pi}{4} \right) \cdot C \cdot R_{2,1,2} \left(\frac{\pi}{2} \right)' = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{2} \\ \frac{1}{\sqrt{2}} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

The parameter vector for this matrix is therefore $\theta = [\theta'_L, \theta'_R]' = [[\frac{\pi}{4}, \frac{7\pi}{4}], [\frac{\pi}{2}]]'$ with $\theta = C_O^{-1}(C)$.

In case of complex unit roots, referring for brevity again to the considered block C_k simply as $C \in \mathbb{C}^{s \times d}$, the set of matrices to be parameterized is

$$U_{s,d} := \{ C \in \mathbb{C}^{s \times d} \mid C'C = I_d \}.$$

The parameterization of this set is based on the combination of complex Givens rotations, as given in Definition 7, which can be used to transform every matrix in $U_{s,d}$ to the form $[D_d, 0'_{(s-d) \times d}]'$ with a diagonal matrix D_d whose diagonal elements are of unit modulus. This transformation is achieved with the following algorithm:

1. Set $j = 1$ and $C^{(1)} = C$.
2. Transform the entries $[c_{j,j}, \dots, c_{j,d}]$ in the j -th row of $C^{(j)}$, to $[\tilde{c}_{j,j}, 0, \dots, 0]$. Since this is a row vector, this is achieved by right-multiplication of C with transposed Givens rotations and the required parameters are obtained via the algorithm described in Remark 12. The first $j - 1$ entries of the j -th row remain unchanged. Denote the transformed matrix by $C^{(j+1)}$.
3. If $j = d - 1$ stop. Else increment j by one ($j \rightarrow j + 1$) and continue at step 2.
4. Collect all parameters used for the Givens rotations in steps 1 to 3 in a parameter vector φ_R . Step 1–3 corresponds to a QR decomposition of $C' = QC'$, with a unitary matrix Q given by the product of the Givens rotations. Please note that the first $j - 1$ entries of the j -th column of $\tilde{C} = C^{(d)}$ are equal to zero by construction.
5. Set $j = 0$ and $\tilde{C}^{(0)} = \tilde{C}$.
6. Collect the entries in column $d - j$ of $\tilde{C}^{(j)}$ which have not been transformed to zero by previous transformations into the vector $[c_{d-j,d-j}, c_{d+1,d-j}, \dots, c_{s,d-j}]'$. Using the algorithm described in Remark 12 transform this vector to $[\tilde{c}_{d-j,d-j}, 0, \dots, 0]'$ by left-multiplication of $\tilde{C}^{(j)}$ with Givens rotations. Since Givens rotations are unitary, the transformed matrix $\tilde{C}^{(j+1)}$ is still unitary implying for its entries $|\tilde{c}_{d-j,d-j}| = 1$ and $\tilde{c}_{i,d-j} = 0$ for all $i < d - j$. An exception occurs if $d = s$. In this case $|c_{d-j,d-j}| = 1$ and no Givens rotations are defined.
7. If $j = d - 1$ stop. Else increment j by one ($j \rightarrow j + 1$) and continue at step 6.
8. Collect all parameters used for the Givens rotations in steps 5 to 7 in a parameter vector φ_L .

9. Transform the diagonal entries of the transformed matrix $\tilde{C}^{(d)} = [D_d, 0'_{(s-d) \times d}]'$ into polar coordinates and collect the angles in a parameter vector φ_D .

The following lemma demonstrates the usefulness of this parameterization.

Lemma 2 (Properties of the parametrization of $U_{s,d}$). Define for $d \leq s$ a mapping $\varphi \rightarrow C_U(\varphi)$ from $\Theta_U^{\mathbb{C}} := \Theta_{\mathbb{C}}^{d(s-d)} \times \Theta_{\mathbb{C}}^{(d-1)d/2} \times [0, 2\pi)^d \rightarrow U_{s,d}$ by

$$\begin{aligned} C_U(\varphi) &:= \left[\prod_{i=1}^d \prod_{j=1}^{s-d} Q_{s,i,d+j}(\varphi_{L,(s-d)(i-1)+j}) \right]' \begin{bmatrix} D_d(\varphi_D) \\ 0_{(s-d) \times d} \end{bmatrix} \left[\prod_{i=1}^{d-1} \prod_{j=1}^i Q_{d,d-i,d-i+j}(\varphi_{R,i(i-1)/2+j}) \right] \\ &:= Q_L(\varphi_L)' \begin{bmatrix} D_d(\varphi_D) \\ 0_{(s-d) \times d} \end{bmatrix} Q_R(\varphi_R), \end{aligned}$$

with $\varphi := [\varphi'_L, \varphi'_R, \varphi'_D]'$, where $\varphi_L = [\varphi_{L,1}, \dots, \varphi_{L,d(s-d)}]'$, $\varphi_R := [\varphi_{R,1}, \dots, \varphi_{R,d(d-1)/2}]'$ and $\varphi_D := [\varphi_{D,1}, \dots, \varphi_{D,d}]'$ and where $D_d(\varphi_D) = \text{diag}(e^{i\varphi_{D,1}}, \dots, e^{i\varphi_{D,d}})$. The following properties hold:

- (i) $U_{s,d}$ is closed and bounded.
- (ii) The mapping $C_U(\varphi)$ is infinitely often differentiable.
- (iii) For every $C \in U_{s,d}$ a vector $\varphi \in \Theta_U^{\mathbb{C}}$ exists such that

$$C = C_U(\varphi) = Q_L(\varphi_L)' \begin{bmatrix} D_d(\varphi_D) \\ 0_{(s-d) \times d} \end{bmatrix} Q_R(\varphi_R).$$

The algorithm discussed above defines the inverse mapping $C_U^{-1} : U_{s,d} \rightarrow \Theta_U^{\mathbb{R}}$.

- (iv) The inverse mapping $C_U^{-1}(\cdot)$ —the parameterization of $U_{s,d}$ —is infinitely often differentiable on an open and dense subset of $U_{s,d}$.

Remark 15. Note the partitioning of the parameter vector φ into the parts φ_L, φ_D and φ_R . The component φ_L fully characterizes the column space of $C_U(\varphi)$, i.e., φ_L determines the cointegrating spaces.

Example 6. Consider the matrix

$$C = \begin{bmatrix} \frac{1-i}{2} & \frac{1-i}{2} \\ \frac{1+i}{2} & -\frac{1-i}{2} \\ 0 & 0 \end{bmatrix}.$$

The starting point is again a QR decomposition of $C' = Q_R(\varphi_R)\tilde{C}' = Q_{2,1,2}(\varphi_{R,1})\tilde{C}'$. To find a complex Givens rotation such that $[\frac{1-i}{2} \quad \frac{1-i}{2}]Q_{2,1,2}(\varphi_{R,1})' = [r e^{i\varphi_a} \quad 0]$ with $r > 0$, transform the entries of $[\frac{1-i}{2} \quad \frac{1-i}{2}]'$ into polar coordinates. The equation $[\frac{1-i}{2} \quad \frac{1-i}{2}]' = [a e^{i\varphi_a} \quad b e^{i\varphi_b}]'$ has the solutions $a = b = \frac{1}{\sqrt{2}}$ and $\varphi_a = \varphi_b = \frac{7\pi}{4}$. Using the results of Remark 12, the parameters of the Givens rotation are $\varphi_{R,1,1} = \tan^{-1}(\frac{b}{a}) = \frac{\pi}{4}$ and $\varphi_{R,1,2} = \varphi_a - \varphi_b = 0$. Right-multiplication of C with $Q_{2,1,2}([\frac{\pi}{4}, 0])'$ leads to

$$\tilde{C} = C Q_{2,1,2}([\frac{\pi}{4}, 0])' = C \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}' = \begin{bmatrix} \frac{1-i}{\sqrt{2}} & 0 \\ 0 & -\frac{1-i}{\sqrt{2}} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} D_2(\varphi_D) \\ 0_{1 \times 2} \end{bmatrix}.$$

Since the entries in the lower 1×2 -sub-block of \tilde{C} are already equal to zero, the remaining complex Givens rotations are $Q_{3,2,3}([0, 0]) = Q_{3,1,3}([0, 0]) = I_3$. Finally, the parameter values corresponding to the diagonal matrix $D_2(\varphi_D) = \text{diag}(e^{i\varphi_{D,1}}, e^{i\varphi_{D,2}}) = \text{diag}(\frac{1-i}{\sqrt{2}}, -\frac{1-i}{\sqrt{2}})$ are $\varphi_{D,1} = \frac{3\pi}{4}$ and $\varphi_{D,2} = \frac{5\pi}{4}$.

The parameter vector for this matrix is therefore $\varphi = [\varphi'_L, \varphi'_R, \varphi'_D] = [[0, 0, 0, 0], [\frac{\pi}{4}, 0], [\frac{3\pi}{4}, \frac{5\pi}{4}]]'$, with $\varphi = C_U^{-1}(C)$.

Components of the Parameter Vector

Based on the results of the preceding sections we can now describe the parameter vectors for the general case. The dimensions of the parameter vectors of the respective blocks of the system matrices $(\mathcal{A}, \mathcal{B}, \mathcal{C})$ depend on the multi-index Γ , consisting of the state space unit root structure Ω_S , the structure indices p and the Kronecker indices α_\bullet for the stable subsystem. A parameterization of the set of all systems in canonical form with given multi-index Γ for the MFI(1) case, therefore, combines the following components:

- $\theta_{B,f} := [\theta'_{B,f,1}, \dots, \theta'_{B,f,l}]' \in \Theta_{B,f} = \mathbb{R}^{d_{B,f}}$, with:

$$\theta_{B,f,k} := \begin{cases} [b_{1,p_1^k+1}^k, b_{1,p_1^k+2}^k, \dots, b_{1,s}^k, b_{2,p_2^k+1}^k, \dots, b_{d_1^k,s}^k]' & \text{for } \omega_k \in \{0, \pi\}, \\ [\mathcal{R}(b_{1,p_1^k+1}^k), \mathcal{I}(b_{1,p_1^k+1}^k), \mathcal{R}(b_{1,p_1^k+2}^k), \dots, \mathcal{I}(b_{1,s}^k), \mathcal{R}(b_{2,p_2^k+1}^k), \dots, \mathcal{I}(b_{d_1^k,s}^k)]' & \text{for } 0 < \omega_k < \pi, \end{cases}$$

for $k = 1, \dots, l$, with p_j^k denoting the j -th entry of the structure indices p corresponding to \mathcal{B}_k . The vectors $\theta_{B,f,k}$ contain the real and imaginary parts of free entries in \mathcal{B}_k not restricted by the p.u.t. structures.

- $\theta_{B,p} := [\theta'_{B,p,1}, \dots, \theta'_{B,p,l}]' \in \Theta_{B,p} = \mathbb{R}_+^{d_{B,p}}$: The vectors $\theta_{B,p,k} := [b_{1,p_1^k}^k, \dots, b_{d_1^k,p_1^k}^k]'$ contain the entries in \mathcal{B}_k restricted by the p.u.t. structures to be positive reals.
- $\theta_{C,E} := [\theta'_{C,E,1}, \dots, \theta'_{C,E,l}]' \in \Theta_{C,E} \subset \mathbb{R}^{d_{C,E}}$: The parameters for the matrices \mathcal{C}_k as discussed in Lemma 1 and Lemma 2.
- $\theta_\bullet \in \Theta_{\bullet,\alpha} \subset \mathbb{R}^{d_\bullet}$: The parameters for the stable subsystem in echelon canonical form for Kronecker indices α_\bullet .

Example 7. Consider an MFI(1) process with $\Omega_S = ((0,2), (\frac{\pi}{2}, 2))$, $p = [1, 3, 1, 2, 1, 2]'$, $n_\bullet = 0$, and system matrices

$$\mathcal{A} = \text{diag}(1, 1, i, i, -i, -i),$$

$$\mathcal{B} = \begin{bmatrix} 1 & -1 & 2 \\ 0 & 0 & 2 \\ \hline 1 & 1+i & 1-i \\ 0 & 2 & i \\ \hline 1 & 1-i & 1+i \\ 0 & 2 & -i \end{bmatrix}, \quad \mathcal{C} = \left[\begin{array}{cc|cc|cc} 0 & \frac{1}{\sqrt{2}} & \frac{1-i}{2} & \frac{1-i}{2} & \frac{1+i}{2} & \frac{1+i}{2} \\ -\frac{1}{\sqrt{2}} & \frac{1}{2} & \frac{1+i}{2} & \frac{-1-i}{2} & \frac{1-i}{2} & \frac{-1+i}{2} \\ \hline \frac{1}{\sqrt{2}} & \frac{1}{2} & 0 & 0 & 0 & 0 \end{array} \right],$$

in canonical form. For this example it holds that $\theta_{B,f} = [[-1, 2], [1, 1, 1, -1, 0, 1]]'$, $\theta_{B,p} = [[1, 2], [1, 2]]$ and

$$\theta_{C,E} = \left[\left[\left[\frac{\pi}{4}, \frac{7\pi}{4} \right], \left[\frac{\pi}{2} \right] \right], \left[[0, 0, 0, 0], \left[\frac{\pi}{4}, 0 \right], \left[\frac{3\pi}{4}, \frac{5\pi}{4} \right] \right] \right]'$$

with parameter values corresponding to the C-blocks collected in $\theta_{C,E}$ considered in Examples 5 and 6.

3.2. The Parameterization in the I(2) Case

The canonical form provided above for the general case has the following form for I(2) processes with unit root structure $\Omega_s = ((0, d_1^1, d_2^1))$:

$$A = \begin{bmatrix} I_{d_1^1} & I_{d_1^1} & 0 & 0 \\ 0 & I_{d_1^1} & 0 & 0 \\ 0 & 0 & I_{d_2^1-d_1^1} & 0 \\ 0 & 0 & 0 & A_\bullet \end{bmatrix}, \quad B = \begin{bmatrix} B_{1,1} \\ B_{1,2,1} \\ B_{1,2,2} \\ B_\bullet \end{bmatrix}, \quad C = [C_{1,1}^E \quad C_{1,2}^G \quad C_{1,2}^E \quad C_\bullet],$$

where $0 < d_1^1 \leq d_2^1 \leq s$, $B_{1,2,1}$ and $B_{1,2,2}$ are p.u.t., $C_{1,1}^E \in O_{s,d_1^1}$, $C_{1,2}^E \in O_{s,d_2^1-d_1^1}$, $(C_{1,1}^E)'C_{1,2}^E = 0_{d_1^1 \times d_2^1}$, $(C_{1,1}^E)'C_{1,2}^G = 0_{d_1^1 \times d_1^1}$, $(C_{1,2}^E)'C_{1,2}^G = 0_{(d_2^1-d_1^1) \times d_1^1}$ and $(A_\bullet, B_\bullet, C_\bullet)$ is in echelon canonical form with Kronecker indices α_\bullet . All matrices are real valued.

The parameterizations of the p.u.t. matrices $B_{1,2,1}$ and $B_{1,2,2}$ are as discussed above. The entries of $B_{1,1}$ are unrestricted and thus included in the parameter vector $\theta_{B,f}$ containing also the free entries in $B_{1,2,1}$ and $B_{1,2,2}$. The subsystem $(A_\bullet, B_\bullet, C_\bullet)$ is parameterized using the echelon canonical form.

The parameterization of $C_{1,1}^E \in O_{s,d_1^1}$ proceeds as in the MFI(1) case, using $C_O^{-1}(C_{1,1}^E)$. The parameterization of $C_{1,2}^E$ has to take the restriction of orthogonality of $C_{1,2}^E$ to $C_{1,1}^E$ into account, thus the set to be parameterized is given by

$$O_{s,d_2^1-d_1^1}(C_{1,1}^E) := \{C_{1,2}^E \in \mathbb{R}^{s \times (d_2^1-d_1^1)} \mid (C_{1,1}^E)'C_{1,2}^E = 0_{d_1^1 \times (d_2^1-d_1^1)}, (C_{1,2}^E)'C_{1,2}^E = I_{d_2^1-d_1^1}\}. \quad (13)$$

The parameterization of this set again uses real Givens rotations. For $C \in O_{s,d_2^1-d_1^1}(C_{1,1}^E)$ it follows that $R_L(\theta_L)C = [0'_{d_1^1 \times (d_2^1-d_1^1)}, \tilde{C}']'$ for a matrix \tilde{C} such that $\tilde{C}'\tilde{C} = I_{d_2^1-d_1^1}$ with $R_L(\theta_L)$ corresponding to $C_{1,1}^E$. The matrix \tilde{C} is parameterized as discussed in Lemma 1.

Corollary 1 (Properties of the parameterization of $O_{s,d_2^1-d_1^1}(C_{1,1}^E)$). Define for $d_1^1 < d_2^1 \leq s$ a mapping $\tilde{\theta} \rightarrow C_{O,d_2^1-d_1^1}(\tilde{\theta}; C_{1,1}^E)$ from $\Theta_{O,d_2^1}^{\mathbb{R}} := [0, 2\pi)^{(d_2^1-d_1^1)(s-d_2^1)} \times [0, 2\pi)^{(d_2^1-d_1^1)(d_2^1-d_1^1-1)/2} \rightarrow O_{s,d_2^1-d_1^1}(C_{1,1}^E)$ by

$$C_{O,d_2^1-d_1^1}(\tilde{\theta}; C_{1,1}^E) := R_L(\theta_L)' \begin{bmatrix} 0_{d_1^1 \times (d_2^1-d_1^1)} \\ C_O(\tilde{\theta}) \end{bmatrix},$$

where θ_L denotes the parameter values corresponding to $[\theta_L', \theta_R']' = C_O^{-1}(C_{1,1}^E)$ as defined in Lemma 1. The following properties hold:

- (i) $O_{s,d_2^1-d_1^1}(C_{1,1}^E)$ is closed and bounded.
- (ii) The mapping $C_{O,d_2^1-d_1^1}(\tilde{\theta}; C_{1,1}^E)$ is infinitely often differentiable.

For $d_2^1 < s$, it holds

- (iii) For every $C_{1,2}^E \in O_{s,d_2^1-d_1^1}(C_{1,1}^E)$ there exists a vector $\tilde{\theta} = [\tilde{\theta}_L', \tilde{\theta}_R']' \in \Theta_{O,d_2^1-d_1^1}^{\mathbb{R}}$ such that

$$C_{1,2}^E = C_{O,d_2^1-d_1^1}(\tilde{\theta}; C_{1,1}^E) = R_L(\theta_L)' \begin{bmatrix} 0_{d_1^1 \times (d_2^1-d_1^1)} \\ R_L(\tilde{\theta}_L)' \begin{bmatrix} I_{d_2^1-d_1^1} \\ 0_{(s-d_2^1) \times (d_2^1-d_1^1)} \end{bmatrix} R_R(\tilde{\theta}_R) \end{bmatrix}.$$

The algorithm discussed above Lemma 1 defines the inverse mapping $C_{O,d_2^1-d_1^1}^{-1}$.

- (iv) The inverse mapping $C_{O,d_2^1-d_1^1}^{-1}(\cdot; C_{1,1}^E)$ —the parameterization of $O_{s,d_2^1-d_1^1}(C_{1,1}^E)$ —is infinitely often differentiable on the pre-image of the interior of $\Theta_{O,d_2^1-d_1^1}^{\mathbb{R}}$. This is an open and dense subset of $O_{s,d_2^1-d_1^1}(C_{1,1}^E)$.

For $d_2^1 = s$, it holds that

(v) $O_{s,s-d_1^1}(\mathcal{C}_{1,1}^E)$ is a disconnected space with two disjoint non-empty closed subsets:

$$\begin{aligned}
 O_{s,s-d_1^1}^+(\mathcal{C}_{1,1}^E) &:= \\
 &\{ \mathcal{C}_{1,2}^E \in \mathbb{R}^{s \times (s-d_1^1)} \mid (\mathcal{C}_{1,1}^E)' \mathcal{C}_{1,2}^E = 0_{d_1^1 \times (s-d_1^1)}, (\mathcal{C}_{1,2}^E)' \mathcal{C}_{1,2}^E = I_{s-d_1^1}, \det([\mathcal{C}_{1,1}^E, \mathcal{C}_{1,2}^E]) = 1 \}, \\
 O_{s,s-d_1^1}^-(\mathcal{C}_{1,1}^E) &:= \\
 &\{ \mathcal{C}_{1,2}^E \in \mathbb{R}^{s \times (s-d_1^1)} \mid (\mathcal{C}_{1,1}^E)' \mathcal{C}_{1,2}^E = 0_{d_1^1 \times (s-d_1^1)}, (\mathcal{C}_{1,2}^E)' \mathcal{C}_{1,2}^E = I_{s-d_1^1}, \det([\mathcal{C}_{1,1}^E, \mathcal{C}_{1,2}^E]) = -1 \}.
 \end{aligned}$$

(vi) For every $O_{s,s-d_1^1}^+(\mathcal{C}_{1,1}^E)$ there exists a vector $\tilde{\theta} \in \Theta_{O,d_2^1-d_1^1}^{\mathbb{R}}$ such that

$$\mathcal{C}_{1,2}^E = C_{O,s-d_1^1}(\tilde{\theta}; \mathcal{C}_{1,1}^E) = R_R(\tilde{\theta}_R).$$

Steps 1–4 of the algorithm discussed above Lemma 1 define the inverse mapping $C_{O,s-d_1^1}^{-1}(\cdot; \mathcal{C}_{1,1}^E) :$

$$O_{s,s-d_1^1}^+(\mathcal{C}_{1,1}^E) \rightarrow \Theta_{O,s-d_1^1}^{\mathbb{R}}.$$

(vii) Define $v := [\pi, \dots, \pi]' \in \mathbb{R}^{(s-d_1^1)(s-d_1^1-1)/2}$. Then a parameterization of $O_{s,s-d_1^1}(\mathcal{C}_{1,1}^E)$ is given by

$$C_{O,s-d_1^1}^{\pm}(\mathcal{C}_{1,2}^E; \mathcal{C}_{1,1}^E) = \begin{cases} v + C_{O,s-d_1^1}^{-1}(\mathcal{C}_{1,2}^E; \mathcal{C}_{1,1}^E) & \text{if } C \in O_{s,s-d_1^1}^+(\mathcal{C}_{1,1}^E) \\ -(v + C_{O,s-d_1^1}^{-1}(\mathcal{C}_{1,2}^E I_{s-d_1^1}^-; \mathcal{C}_{1,1}^E)) & \text{if } C \in O_{s,s-d_1^1}^-(\mathcal{C}_{1,1}^E). \end{cases}$$

The parameterization is infinitely often differentiable with infinitely often differentiable inverse on an open and dense subset of $O_{s,s}$.

The proof of Corollary 1 uses the same arguments as the proof of Lemma 1 and is, therefore, omitted. It remains to provide a parameterization for $\mathcal{C}_{1,2}^G$ restricted to be orthogonal to both $\mathcal{C}_{1,1}^E$ and $\mathcal{C}_{1,2}^E$. Thus, the set to be parametrized is given by

$$O_{s,G}(\mathcal{C}_{1,1}^E, \mathcal{C}_{1,2}^E) := \{ \mathcal{C}_{1,2}^G \in \mathbb{R}^{s \times d_1^1} \mid (\mathcal{C}_{1,1}^E)' \mathcal{C}_{1,2}^G = 0_{d_1^1 \times d_1^1}, (\mathcal{C}_{1,2}^E)' \mathcal{C}_{1,2}^G = 0_{(d_2^1-d_1^1) \times d_1^1} \}.$$

The parameterization of $O_{s,G}(\mathcal{C}_{1,1}^E, \mathcal{C}_{1,2}^E)$ is straightforward: Left multiplication of $\mathcal{C}_{1,2}^G$ with $R_L(\theta_L)$ as defined in Lemma 1 and of the lower $(s-d_1^1) \times d_1^1$ -block with $R_L(\tilde{\theta}_L)$ as defined in Corollary 1 transforms the upper $d_2^1 \times d_1^1$ -block to zero and collects the free parameters in the lower $(s-d_1^1) \times d_1^1$ -block. Clearly, this is a bijective and infinitely often differentiable mapping on $O_{s,G}(\mathcal{C}_{1,1}^E, \mathcal{C}_{1,2}^E)$ and thus a useful parameterization, since the matrix $\mathcal{C}_{1,2}^G$ is only multiplied with two constant invertible matrices. The entries of the matrix product are then collected in a parameter vector as shown in Corollary 2.

Corollary 2 (Properties of the parameterization of $O_{s,G}(\mathcal{C}_{1,1}^E, \mathcal{C}_{1,2}^E)$). Define for given matrices $\mathcal{C}_{1,1}^E \in O_{s,d_1^1}$ and $\mathcal{C}_{1,2}^E \in O_{s,d_2^1-d_1^1}(\mathcal{C}_{1,1}^E)$ a mapping $\lambda \rightarrow C_{O,G}(\lambda; \mathcal{C}_{1,1}^E, \mathcal{C}_{1,2}^E)$ from $\mathbb{R}^{d_1^1(s-d_2^1)} \rightarrow O_{s,G}(\mathcal{C}_{1,1}^E, \mathcal{C}_{1,2}^E)$ by

$$C_{O,G}(\lambda; \mathcal{C}_{1,1}^E, \mathcal{C}_{1,2}^E) := R_L(\theta_L)' \left[R_L(\tilde{\theta}_L)' \begin{bmatrix} 0_{d_1^1 \times d_1^1} \\ 0_{(d_2^1-d_1^1) \times 1} \cdots 0_{(d_2^1-d_1^1) \times 1} \\ \lambda_1 \cdots \lambda_{d_1^1} \\ \lambda_{d_1^1+1} \cdots \lambda_{2d_1^1} \\ \vdots \vdots \\ \lambda_{d_1^1(s-d_2^1-1)+1} \cdots \lambda_{d_1^1(s-d_2^1)} \end{bmatrix} \right],$$

where θ_L denotes the parameter values corresponding to $[\theta'_L, \theta'_R]' = C_O^{-1}(C_{1,1}^E)$ as defined in Lemma 1 and $\tilde{\theta}_L$ denotes the parameter values corresponding to $[\tilde{\theta}'_L, \tilde{\theta}'_R]' = C_{O,d_2^1-d_1^1}^{-1}(C_{1,2}^E; C_{1,1}^E)$ as defined in Corollary 1. The set $O_{s,G}(C_{1,1}^E, C_{1,2}^E)$ is closed and both $C_{O,G}$ as well as $C_{O,G}^{-1}(\cdot)$ —the parameterization of $O_{s,G}(C_{1,1}^E, C_{1,2}^E)$ —are infinitely often differentiable.

Components of the Parameter Vector

In the I(2) case, the multi-index Γ contains the state space unit root structure $\Omega_S = ((0, d_1^1, d_2^1))$, the structure indices $p \in \mathbb{N}_0^{d_1^1+d_2^1}$, encoding the p.u.t. structures of $\mathcal{B}_{1,2,1}$ and $\mathcal{B}_{1,2,2}$, and the Kronecker indices α_\bullet for the stable subsystem. The parameterization of the set of all systems in canonical form with given multi-index Γ for the I(2) case uses the following components:

- $\theta_{B,f} := \theta_{B,f,1} \in \Theta_{B,f} = \mathbb{R}^{d_{B,f}}$: The vector $\theta_{B,f,1}$ contains the free entries in \mathcal{B}_1 not restricted by the p.u.t. structure, collected in the same order as for the matrices \mathcal{B}_k in the MFI(1) case.
- $\theta_{B,p} := \theta_{B,p,1} \in \Theta_{B,p} = \mathbb{R}_+^{d_{B,p}}$: The vector $\theta_{B,p,1} := \left[b_{d_1^1+1, p_{d_1^1+1}}^1, \dots, b_{d_1^1+d_2^1, p_{d_1^1+d_2^1}}^1 \right]'$ contains the entries in \mathcal{B}_1 restricted by the p.u.t. structures to be positive reals.
- $\theta_{C,E} := [\theta'_{C,E,1,1}, \theta'_{C,E,1,2}]' \in \Theta_{C,E} \subset \mathbb{R}^{d_{C,E}}$: The parameters for the matrices $C_{1,1}^E$ as in the MFI(1) case and $C_{1,2}^E$ as discussed in Corollary 1.
- $\theta_{C,G} \in \Theta_{C,G} = \mathbb{R}^{d_{C,G}}$: The parameters for the matrix $C_{1,2}^G$ as discussed in Corollary 2.
- $\theta_\bullet \in \Theta_{\bullet, \alpha_\bullet} \subset \mathbb{R}^{d_\bullet}$: The parameters for the stable subsystem in echelon canonical form for Kronecker indices α_\bullet .

Example 8. Consider an I(2) process with $\Omega_S = ((0, 1, 2))$, $p = [0, 1, 1]'$, $n_\bullet = 0$ and system matrices

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 2 & -2 \\ 1 & -1 & 3 \\ 2 & 0 & 1 \end{bmatrix}, \quad C = \left[\begin{array}{c|c|c} 0 & -1 & \frac{1}{\sqrt{2}} \\ \hline -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{1}{2} \\ \hline \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{1}{2} \end{array} \right]'$$

In this case, $\theta_{B,f,1} = [-1, 2, -2, -1, 3, 0, 1]'$, $\theta_{B,p,1} = [1, 2]'$. It follows from

$$\begin{aligned} R_{3,1,2} \left(\frac{7\pi}{4} \right) R_{3,1,3} \left(\frac{\pi}{2} \right) C_{1,1}^E &= [1 \ 0 \ 0]', \\ R_{3,1,2} \left(\frac{7\pi}{4} \right) R_{3,1,3} \left(\frac{\pi}{2} \right) C_{1,2}^E &= \left[0 \ \frac{1}{\sqrt{2}} \ \frac{-1}{\sqrt{2}} \right]' \quad \text{and} \quad R_{2,1,2} \left(\frac{7\pi}{4} \right) \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \\ R_{3,1,2} \left(\frac{7\pi}{4} \right) R_{3,1,3} \left(\frac{\pi}{2} \right) C_{1,2}^G &= [0 \ 1 \ 1]' \quad \text{and} \quad R_{2,1,2} \left(\frac{7\pi}{4} \right) \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \sqrt{2} \end{bmatrix}, \end{aligned}$$

that $\theta_{C,E} = [\theta'_{C,E,1,1}, \theta_{C,E,1,2}]' = \left[\left[\frac{\pi}{2}, \frac{7\pi}{4} \right], \left[\frac{7\pi}{4} \right] \right]'$ and $\theta_{C,G} = [\sqrt{2}]$.

3.3. The Parameterization in the General Case

Inspecting the canonical form shows that all relevant building blocks are already present in the MFI(1) and the I(2) cases and can be combined to deal with the general case: The entries in \mathcal{B}_u are either unrestricted or follow restrictions according to given structure indices p , and the parameter space is chosen accordingly, as discussed for the MFI(1) and I(2) cases. The restrictions on the matrices C_u and its blocks C_k require more sophisticated parameterizations of parts of unitary or orthonormal matrices as well as of orthogonal complements. These are dealt with in Lemmas 1 and 2 and Corollaries 1 and 2 above. The extension of Corollaries 1 and 2 to complex matrices and to matrices which are orthogonal to a larger number of blocks of C_k is straightforward.

The following theorem characterizes the properties of parameterizations for sets M_Γ of transfer functions with (general) multi-index Γ and describes the relations between sets of transfer functions and the corresponding sets Δ_Γ of triples $(\mathcal{A}, \mathcal{B}, \mathcal{C})$ of system matrices in canonical form, defined below. Discussing the continuity and differentiability of mappings on sets of transfer functions and on sets of matrix triples also requires the definition of a topology on both sets.

Definition 8.

- (i) The set of transfer functions of order n , M_n , is endowed with the pointwise topology T_{pt} : First, identify transfer functions with their impulse response sequences. Then, a sequence of transfer functions $k_i(z) = I_s + \sum_{j=1}^{\infty} K_{j,i}z^j$ converges in T_{pt} to $k_0(z) = I_s + \sum_{j=1}^{\infty} K_{j,0}z^j$ if and only if for every $j \in \mathbb{N}$ it holds that $K_{j,i} \xrightarrow{i \rightarrow \infty} K_{j,0}$.
- (ii) The set of all triples $(\mathcal{A}, \mathcal{B}, \mathcal{C})$ in canonical form corresponding to transfer functions with multi-index Γ is called Δ_Γ . The set Δ_Γ is endowed with the topology corresponding to the distance $d((A_1, B_1, C_1), (A_2, B_2, C_2)) := \|A_1 - A_2\|_{Fr} + \|B_1 - B_2\|_{Fr} + \|C_1 - C_2\|_{Fr}$.

Please note that in the definition of the pointwise topology convergence does not need to be uniform in j and moreover, the power series coefficients do not need to converge to zero for $j \rightarrow \infty$ and hence the concept can also be used for unstable systems.

Theorem 2. The set M_n can be partitioned into pieces M_Γ , where $\Gamma := \{\Omega_S, p, \alpha_\bullet\}$, i.e.,

$$M_n = \bigcup_{\Gamma = \{\Omega_S, p, \alpha_\bullet\} | n_u(\Omega_S) + n_\bullet(\alpha_\bullet) = n} M_\Gamma,$$

where $n_u(\Omega_S) := \sum_{k=1}^l \sum_{j=1}^{h_k} d_j^k \delta_k$, with $\delta_k = 1$ for $\omega_k \in \{0, \pi\}$ and $\delta_k = 2$ for $0 < \omega_k < \pi$ is the state dimension of the unstable subsystem $(\mathcal{A}_u, \mathcal{B}_u, \mathcal{C}_u)$ with state space unit root structure Ω_S and $n_\bullet(\alpha_\bullet) := \sum_{i=1}^s \alpha_{\bullet,i}$ is the state dimension of the stable subsystem with Kronecker indices $\alpha_\bullet = (\alpha_{\bullet,1}, \dots, \alpha_{\bullet,s})$, $\alpha_{\bullet,i} \in \mathbb{N}_0$. For every multi-index Γ there exists a parameter space $\Theta_\Gamma \subset \mathbb{R}^{d(\Gamma)}$ for some integer $d(\Gamma)$, endowed with the Euclidean norm, and a function $\phi_\Gamma : \Delta_\Gamma \rightarrow \Theta_\Gamma$, such that for every $(\mathcal{A}, \mathcal{B}, \mathcal{C}) \in \Delta_\Gamma$ the parameter vector $\theta := \phi_\Gamma(\mathcal{A}, \mathcal{B}, \mathcal{C}) \in \Theta_\Gamma$ is composed of:

- The parameter vector $\theta_{B,f} = [\theta'_{B,f,1}, \dots, \theta'_{B,f,l}]' \in \Theta_{B,f} = \mathbb{R}^{d_{B,f}}$, collecting the (real and imaginary parts of) non-restricted entries in $\mathcal{B}_k, k = 1, \dots, l$ as described in the MFI(1) case.
- The parameter vector $\theta_{B,p} = [\theta'_{B,p,1}, \dots, \theta'_{B,p,l}]' \in \Theta_{B,p} = \mathbb{R}_+^{d_{B,p}}$, collecting the entries in $\mathcal{B}_k, k = 1, \dots, l$, restricted by the p.u.t. forms to be positive reals in a similar fashion as described for \mathcal{B}_1 in the I(2) case.
- The parameter vector $\theta_{C,E} = [\theta'_{C,E,1}, \dots, \theta'_{C,E,l}]' \in \Theta_{C,E} \subset \mathbb{R}^{d_{C,E}}$, $\theta_{C,E,k} = [\theta'_{C,E,k,1}, \dots, \theta'_{C,E,k,h_k}]'$ collecting the parameters $\theta_{C,E,k,j}$ for all blocks $\mathcal{C}_{k,j}^E, k = 1, \dots, l$ and $j = 1, \dots, h_k$, obtained using Givens rotations (see Lemmas 1 and 2 and Corollary 1 and its extension to complex matrices).
- The parameter vector $\theta_{C,G} = [\theta'_{C,G,1}, \dots, \theta'_{C,G,l}]' \in \Theta_{C,G} = \mathbb{R}^{d_{C,G}}$, $\theta_{C,G,k} = [\theta'_{C,G,k,2}, \dots, \theta'_{C,G,k,h_k}]'$ collecting the parameters $\theta_{C,G,k,j}$ (real and imaginary parts for complex roots) for $\mathcal{C}_{k,j}^G, k = 1, \dots, l$ and $j = 2, \dots, h_k$, subject to the orthogonality restrictions (see Corollary 2 and its extension to complex matrices).
- The parameter vector $\theta_\bullet \in \Theta_\bullet \subset \mathbb{R}^{d_\bullet}$ collecting the free entries in echelon canonical form with Kronecker indices α_\bullet .

- (i) The mapping $\psi_\Gamma : M_\Gamma \rightarrow \Delta_\Gamma$ that attaches a triple $(\mathcal{A}, \mathcal{B}, \mathcal{C})$ in canonical form to a transfer function in M_Γ is continuous. It is the inverse (restricted to M_Γ) of the T_{pt} -continuous function $\pi : (A, B, C) \mapsto k(z) = I_s + zC(I_n - zA)^{-1}B$.
- (ii) Every parameter vector $\theta = [\theta'_{B,f}, \theta'_{B,p}, \theta'_{C,E}, \theta'_{C,G}, \theta'_\bullet]' \in \Theta_\Gamma \subset \Theta_{B,f} \times \Theta_{B,p} \times \Theta_{C,E} \times \Theta_{C,G} \times \Theta_\bullet$ corresponds to a triple $(\mathcal{A}(\theta), \mathcal{B}(\theta), \mathcal{C}(\theta)) \in \Delta_\Gamma$ and a transfer function $k(z) = \pi(\mathcal{A}(\theta), \mathcal{B}(\theta), \mathcal{C}(\theta)) \in M_\Gamma$. The mapping $\phi_\Gamma^{-1} : \theta \rightarrow (\mathcal{A}(\theta), \mathcal{B}(\theta), \mathcal{C}(\theta))$ is continuous on Θ_Γ .

(iii) For every multi-index Γ the set of points in Δ_Γ , where the mapping ϕ_Γ is continuous, is open and dense in Δ_Γ .

As mentioned in Section 2, the parameterization of Φ is straightforward. The $s \times m$ entries of Φ are collected in a parameter vector d . Thus, there is a one-to-one correspondence between state space realizations $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \Phi) \in \Delta_\Gamma \times \mathbb{R}^{s \times m}$ and parameter vectors $\tau = [\theta', d']' \in \Theta_\Gamma \times \mathbb{R}^{sm}$. The same holds true for parameters used for the symmetric, positive definite innovation matrix $\Sigma \in \mathbb{R}^{s \times s}$ obtained, e.g., from a lower triangular Cholesky factor of Σ .

4. The Topological Structure

The parameterization of M_n in Theorem 2 partitions M_n into subsets M_Γ for a selection of multi-indices Γ . To every multi-index Γ there exists a corresponding associated parameter set Θ_Γ . Thus, in practical applications, maximizing the pseudo likelihood requires choosing the multi-index Γ . Maximizing the pseudo likelihood over the set M_Γ effectively amounts to including also all elements in the closure of M_Γ , because of continuity of the parameterization. It is thus necessary to characterize the closures of the sets M_Γ .

Moreover, maximizing the pseudo likelihood function over all possible multi-indices is time-consuming and not desirable. Fortunately, the results discussed below show that there exists a generic multi-index Γ_g such that $M_n \subset \overline{M_{\Gamma_g}}$. This generic choice corresponds to the set of all stable systems of order n corresponding to the generic neighborhood of the echelon canonical form. This multi-index, therefore, is a natural starting point for estimation.

However, in particular for hypotheses testing, it will be necessary to maximize the pseudo likelihood over sets of transfer functions of order n with specific state space unit root structure Ω_S , denoted as $M(\Omega_S, n_\bullet)$ below, where n_\bullet denotes the dimension of the stable part of the state. We show below that also in this case there exists a generic multi-index $\Gamma_g(\Omega_S, n_\bullet)$ such that $M(\Omega_S, n_\bullet) \subset \overline{M_{\Gamma_g(\Omega_S, n_\bullet)}}$.

The main tool to obtain these results is investigating the properties of the mappings ψ_Γ , that map transfer functions in M_Γ to triples $(A, B, C) \in \Delta_\Gamma$, as well as analyzing the closures of the sets Δ_Γ . The relation between parameter vectors $\theta \in \Theta_\Gamma$ and triples of system matrices $(A, B, C) \in \Delta_\Gamma$ is easier to understand than the relation between Δ_Γ and M_Γ , due to the results of Theorem 2. Consequently, this section focuses on the relations between Δ_Γ and M_Γ —and their closures—for different multi-indices Γ .

To define the closures we embed the sets Δ_Γ of matrices in canonical form with multi-indices Γ corresponding to transfer functions of order n into the space Δ_n of all conformable complex matrix triples (A, B, C) with $A \in \mathbb{C}^{n \times n}$, where additionally $\lambda_{|max|}(A) \leq 1$. Since the elements of Δ_n are matrix triples, this set is isomorphic to a subset of the finite dimensional space \mathbb{C}^{n^2+2ns} , equipped with the Euclidean topology. Please note that Δ_n also contains non-minimal state space realizations, corresponding to transfer functions of lower order.

Remark 16. In principle the set Δ_n also contains state space realizations of transfer functions $k(z) = I_s + \sum_{j=1}^\infty K_j z^j$ with complex valued coefficients K_j . Since the subset of Δ_n of state space systems realizing transfer functions with real valued K_j is closed in Δ_n , realizations corresponding to transfer functions with coefficients with non-zero imaginary part are irrelevant for the analysis of the closures of the sets Δ_Γ .

After investigating the closure of Δ_Γ in Δ_n , denoted by $\overline{\Delta_\Gamma}$, we consider the set of corresponding transfer functions $\pi(\overline{\Delta_\Gamma})$. Since we effectively maximize the pseudo likelihood over $\overline{\Delta_\Gamma}$, we have to understand for which multi-indices $\tilde{\Gamma}$ the set $\pi(\Delta_{\tilde{\Gamma}})$ is a subset of $\pi(\overline{\Delta_\Gamma})$. Moreover, we find a covering of $\pi(\overline{\Delta_\Gamma}) \subset \bigcup_{i \in \mathcal{I}} M_{\Gamma_i}$. This restricts the set of multi-indices Γ that may occur as possible multi-indices of the limit of a sequence in $\pi(\Delta_\Gamma)$ and thus the set of transfer functions that can be obtained by maximization of the pseudo likelihood.

The sets M_Γ , are embedded into the vector space M of all causal transfer functions $k(z) = I_s + \sum_{j=1}^\infty K_j z^j$. The vector space M is isomorphic to the infinite dimensional space $\Pi_{j \in \mathbb{N}} \mathbb{R}_j^{s \times s}$ equipped with the pointwise topology. Since, as mentioned above, maximization of the pseudo likelihood function over M_Γ effectively includes $\overline{M_\Gamma}$, it is important to determine for any given multi-index Γ , the multi-indices $\tilde{\Gamma}$ for which the set $M_{\tilde{\Gamma}}$ is a subset of $\overline{M_\Gamma}$. Please note that $\overline{M_\Gamma}$ is not necessarily equal to $\pi(\overline{\Delta_\Gamma})$. The continuity of π , as shown in Theorem 2 (i), implies the following inclusions:

$$M_\Gamma = \pi(\Delta_\Gamma) \subset \pi(\overline{\Delta_\Gamma}) \subset \overline{M_\Gamma}.$$

In general all these inclusions are strict. For a discussion in case of stable transfer functions see Hannan and Deistler (1988, Theorem 2.5.3).

We first define a partial ordering on the set of multi-indices Γ . Subsequently we examine the closure $\overline{\Delta_\Gamma}$ in Δ_n and finally we examine the closures $\overline{M_\Gamma}$ in M .

Definition 9.

- (i) For two state space unit root structures Ω_S and $\tilde{\Omega}_S$ with corresponding matrices $\mathcal{A}_u \in \mathbb{C}^{n_u \times n_u}$ and $\tilde{\mathcal{A}}_u \in \mathbb{C}^{\tilde{n}_u \times \tilde{n}_u}$ in canonical form, it holds that $\tilde{\Omega}_S \leq \Omega_S$ if and only if there exists a permutation matrix S such that

$$S \mathcal{A}_u S' = \begin{bmatrix} \tilde{\mathcal{A}}_u & \tilde{J}_{12} \\ 0 & \tilde{J}_2 \end{bmatrix}.$$

Moreover, $\tilde{\Omega}_S < \Omega_S$ holds if additionally $\tilde{\Omega}_S \neq \Omega_S$.

- (ii) For two state space unit root structures Ω_S and $\tilde{\Omega}_S$ and dimensions of the stable subsystems $n_\bullet, \tilde{n}_\bullet \in \mathbb{N}_0$ we define

$$(\tilde{\Omega}_S, \tilde{n}_\bullet) \leq (\Omega_S, n_\bullet) \text{ if and only if } \tilde{\Omega}_S \leq \Omega_S, \tilde{n}_\bullet \leq n_\bullet.$$

Strict inequality holds, if at least one of the two inequalities above holds strictly.

- (iii) For two pairs (Ω_S, p) and $(\tilde{\Omega}_S, \tilde{p})$ with corresponding matrices $\mathcal{A}_u \in \mathbb{C}^{n_u \times n_u}$ and $\tilde{\mathcal{A}}_u \in \mathbb{C}^{\tilde{n}_u \times \tilde{n}_u}$ in canonical form, it holds that $(\tilde{\Omega}_S, \tilde{p}) \leq (\Omega_S, p)$ if and only if there exists a permutation matrix S such that

$$S \mathcal{A}_u S' = \begin{bmatrix} \tilde{\mathcal{A}}_u & \tilde{J}_{12} \\ 0 & \tilde{J}_2 \end{bmatrix}, \quad S p = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix},$$

where $p_1 \in \mathbb{N}_0^{\tilde{n}_u}$ and \tilde{p} restricts at least as many entries as p_1 , i.e., $\tilde{p}_i \geq (p_1)_i$ holds for all $i = 1, \dots, \tilde{n}_u$. Moreover, $(\tilde{\Omega}_S, \tilde{p}) < (\Omega_S, p)$ holds if additionally $(\tilde{\Omega}_S, \tilde{p}) \neq (\Omega_S, p)$.

- (iv) Let $\alpha_\bullet = (\alpha_{\bullet,1}, \dots, \alpha_{\bullet,s}), \alpha_{\bullet,i} \in \mathbb{N}_0$ and $\tilde{\alpha}_\bullet = (\tilde{\alpha}_{\bullet,1}, \dots, \tilde{\alpha}_{\bullet,s}), \tilde{\alpha}_{\bullet,i} \in \mathbb{N}_0$. Then $\tilde{\alpha}_\bullet \leq \alpha_\bullet$ if and only if $\tilde{\alpha}_{\bullet,i} \leq \alpha_{\bullet,i}, i = 1, \dots, s$. Moreover, $\tilde{\alpha}_\bullet < \alpha_\bullet$ holds, if at least one inequality is strict (compare Hannan and Deistler 1988, sct. 2.5).

Finally, define

$$\tilde{\Gamma} = (\tilde{\Omega}_S, \tilde{p}, \tilde{\alpha}_\bullet) \leq \Gamma = (\Omega_S, p, \alpha_\bullet) \text{ if and only if } (\tilde{\Omega}_S, \tilde{p}) \leq (\Omega_S, p) \text{ and } \tilde{\alpha}_\bullet \leq \alpha_\bullet.$$

Strict inequality holds, if at least one of the inequalities above holds strictly.

Please note that (i) implies that $\tilde{\Omega}_S$ only contains unit roots that are also contained in Ω_S , with the integration orders \tilde{h}_k of the unit roots in $\tilde{\Omega}_S$ smaller or equal to the integration orders of the respective unit roots in Ω_S . Thus, denoting the unit root structures corresponding to $\tilde{\Omega}_S$ and Ω_S by $\tilde{\Omega}$ and Ω , it follows that $\tilde{\Omega}_S \leq \Omega_S$ implies $\tilde{\Omega} \preceq \Omega$. The reverse does not hold as, e.g., for $\Omega_S = ((0, 1, 1))$ (where

hence $\Omega = ((0, 2))$ and $\tilde{\Omega}_S = ((0, 2))$ (with $\tilde{\Omega} = ((0, 1))$) it holds that $\tilde{\Omega} \prec \Omega$, but neither $\tilde{\Omega}_S \leq \Omega_S$ nor $\Omega_S \leq \tilde{\Omega}_S$ holds as here

$$\mathcal{A}_u = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \tilde{\mathcal{A}}_u = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

This partial ordering is convenient for the characterization of the closure of Δ_Γ .

4.1. The Closure of Δ_Γ in Δ_n

Please note that the block-structure of \mathcal{A} implies that every system in Δ_Γ can be separated in two subsystems $(\mathcal{A}_u, \mathcal{B}_u, \mathcal{C}_u)$ and $(\mathcal{A}_\bullet, \mathcal{B}_\bullet, \mathcal{C}_\bullet)$. Define $\Delta_{\Omega_S, p} := \Delta_{(\Omega_S, p, \{\})}$ as the set of all state space realizations in canonical form corresponding to state space unit root structure Ω_S , structure indices p and $n_\bullet = 0$. Analogously define $\Delta_{\alpha_\bullet} := \Delta_{(\{\}, \{\}, \alpha_\bullet)$ as the set of all state space realizations in canonical form with $\Omega_S = \{\}$ and Kronecker indices α_\bullet . Examining $\overline{\Delta_{\Omega_S, p}}$ and $\overline{\Delta_{\alpha_\bullet}}$ separately simplifies the analysis.

4.1.1. The Closure of $\Delta_{\Omega_S, p}$

The canonical form imposes a lot of structure, i.e., restrictions on the matrices \mathcal{A} , \mathcal{B} and \mathcal{C} . By definition $\Delta_{\Omega_S, p} = \Delta_{\Omega_S, p}^A \times \Delta_{\Omega_S, p}^B \times \Delta_{\Omega_S, p}^C$ and the closures of the three matrices can be analyzed separately. $\Delta_{\Omega_S, p}^A$ and $\Delta_{\Omega_S, p}^C$ are very easy to investigate. The structure of \mathcal{A} is fully determined by Ω_S and consequently $\Delta_{\Omega_S, p}^A$ consists of a single matrix \mathcal{A} which immediately implies that $\overline{\Delta_{\Omega_S, p}^A} = \Delta_{\Omega_S, p}^A$. The matrix \mathcal{C} , compare Theorem 1 is composed of blocks \mathcal{C}_k^E that are sub-blocks of unitary (or orthonormal) matrices and blocks \mathcal{C}_k^G that have to fulfill (recursive) orthogonality constraints. The corresponding sets were shown to be closed in Lemmas 1 and 2 and Corollaries 1 and 2. Thus, $\overline{\Delta_{\Omega_S, p}^C} = \Delta_{\Omega_S, p}^C$.

It remains to discuss $\overline{\Delta_{\Omega_S, p}^B}$. The structure indices p defining the p.u.t. structures of the matrices \mathcal{B}_k restrict some entries to be positive. Combining all the parameters—unrestricted with complex values parameterized by real and imaginary part and the positive entries—into a parameter vector leads to an open sub-set of \mathbb{R}^m for some m . For convergent sequences of systems with fixed Ω_S and p , limits of entries restricted to be positive may be zero. When this happens, two cases have to be distinguished. First, all p.u.t. sub-matrices still have full row rank. In this case the limiting system, $(\mathcal{A}_0, \mathcal{B}_0, \mathcal{C}_0)$ say, is still minimal and can be transformed to a system in canonical form $(\tilde{\mathcal{A}}_0, \tilde{\mathcal{B}}_0, \tilde{\mathcal{C}}_0)$ with fewer unrestricted entries in $\tilde{\mathcal{B}}_0$.

Second, if at least one of the row ranks of the p.u.t. blocks decreases in the limit, the limiting system is no longer minimal. Consequently, $(\tilde{\Omega}_S, \tilde{p}) < (\Omega_S, p)$ in the limit.

To illustrate this point consider again Example 4 with Equation (12) rewritten as

$$x_{t+1,1} = x_{t,1} + x_{t,2} + \mathcal{B}_{1,1}\varepsilon_t, \quad x_{t+1,2} = x_{t,2} + \mathcal{B}_{1,2,1}\varepsilon_t, \quad x_{t+1,3} = x_{t,3} + \mathcal{B}_{1,2,2}\varepsilon_t.$$

If $\mathcal{B}_{1,2,1} = [0, b_{1,2,1,2}] \neq 0$ and $\mathcal{B}_{1,2,2} = [b_{1,2,2,1}, b_{1,2,2,2}] \neq 0$, $b_{1,2,2,1} > 0$, it holds that $\{y_t\}_{t \in \mathbb{Z}}$ is an I(2) process with state space unit root structure $\Omega_S = ((0, 1, 2))$.

Now consider a sequence of systems with all parameters except for $b_{1,2,1,2}$ constant and $b_{1,2,1,2} \rightarrow 0$. The limiting system is then given by

$$y_t = \mathcal{C}_{1,1}^E x_{t,1} + \mathcal{C}_{1,2}^G x_{t,2} + \mathcal{C}_{1,2}^E x_{t,3} + \varepsilon_t,$$

$$\begin{bmatrix} x_{t+1,1} \\ x_{t+1,2} \\ x_{t+1,3} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t,1} \\ x_{t,2} \\ x_{t,3} \end{bmatrix} + \begin{bmatrix} b_{1,1,1} & b_{1,1,2} \\ 0 & 0 \\ b_{1,2,2,1} & b_{1,2,2,2} \end{bmatrix} \varepsilon_t, \quad x_{1,1} = x_{1,2} = x_{1,3} = 0.$$

In the limiting system $x_{t,2} = 0$ is redundant and $\{y_t\}_{t \in \mathbb{Z}}$ is an I(1) process rather than an I(2) process. Dropping $x_{t,2}$ leads to a state space realisation of the limiting system $\{y_t\}_{t \in \mathbb{Z}}$ given by

$$y_t = C_{1,1}^E x_{t,1} + C_{1,2}^E x_{t,3} + \varepsilon_t = \tilde{C} \tilde{x}_t + \varepsilon_t, \quad \tilde{x}_t \in \mathbb{R}^2,$$

$$\tilde{x}_{t+1} = \begin{bmatrix} x_{t+1,1} \\ x_{t+1,3} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t,1} \\ x_{t,3} \end{bmatrix} + \begin{bmatrix} b_{1,1,1} & b_{1,1,2} \\ b_{1,2,2,1} & b_{1,2,2,2} \end{bmatrix} \varepsilon_t = \tilde{x}_t + \tilde{B} \varepsilon_t, \quad x_{1,1} = x_{1,3} = 0.$$

In case \tilde{B} has full rank, the above system is minimal. Since $b_{1,2,2,1} > 0$, the matrix \tilde{B} needs to be transformed into p.u.t. format. By definition all systems in the sequence, with $b_{1,2,1,2} \neq 0$, have structure indices $p = [0, 2, 1]'$ as discussed in Example 12. The limiting system—in case of full rank of \tilde{B} —has indices $\tilde{p} = [1, 2]'$. To relate to Definition 9 choose the permutation matrix

$$S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \text{ to arrive at}$$

$$S \mathcal{A}_u S' = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} I_2 & \tilde{J}_{12} \\ 0 & \tilde{J}_2 \end{bmatrix}, \quad S p = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} (p_1)_1 \\ (p_1)_2 \\ p_2 \end{bmatrix}.$$

This shows that $(\tilde{p})_i > (p_1)_i, i = 1, 2$ and thus the limiting system has a smaller multi-index Γ than the systems of the sequence. In case \tilde{B} has reduced rank equal to one a further reduction in the system order to $n = 1$ along similar lines as discussed is possible, again leading to a limiting system with smaller multi-index Γ .

The discussion shows that the closure of $\Delta_{\Omega_s, p}^B$ is related to lower order systems in the sense of Definition 9. The precise statement is given in Theorem 3 after a discussion of the closure of the stable subsystems.

4.1.2. The Closure of Δ_{α_\bullet}

Consider a convergent sequence of systems $\{(\mathcal{A}_j, \mathcal{B}_j, \mathcal{C}_j)\}_{j \in \mathbb{N}}$ in Δ_{α_\bullet} and denote the limiting system by (A_0, B_0, C_0) . Clearly, $\lambda_{|\max|}(A_0) \leq 1$ holds true for the limit A_0 of the sequence $\{\mathcal{A}_j\}_{j \in \mathbb{N}}$ with $\lambda_{|\max|}(\mathcal{A}_j) < 1$ for all j . Therefore, two cases have to be discussed for the limit:

- If $\lambda_{|\max|}(A_0) < 1$, the potentially non-minimal limiting system (A_0, B_0, C_0) corresponds to a minimal state space realization with Kronecker indices smaller or equal to α_\bullet (cf. Hannan and Deistler 1988, Theorem 2.5.3).
- If $\lambda_{|\max|}(A_0) = 1$, the limiting matrix A_0 is similar to a block matrix $\tilde{A} = \text{diag}(\tilde{J}_2, \tilde{A}_\bullet)$, where all eigenvalues of \tilde{J}_2 have unit modulus and $\lambda_{|\max|}(\tilde{A}_\bullet) < 1$.

The first case is well understood, compare Hannan and Deistler (1988, chp. 2), since the limit in this case corresponds to a stable transfer function. In the second case the limiting system can be separated into two subsystems $(\tilde{J}_2, \tilde{B}_u, \tilde{C}_u)$ and $(\tilde{A}_\bullet, \tilde{B}_\bullet, \tilde{C}_\bullet)$, according to the block diagonal structure of \tilde{A} . The state space unit root structure of the limiting system (A_0, B_0, C_0) depends on the multiplicities of the eigenvalues of the matrix \tilde{J}_2 and is greater (in the sense of Definition 9) than the empty state space unit root structure. At the same time the Kronecker indices of the subsystem $(\tilde{A}_\bullet, \tilde{B}_\bullet, \tilde{C}_\bullet)$ are smaller than α_\bullet , compare again Hannan and Deistler (1988, chp. 2). Since the Kronecker indices impose restrictions on some entries of the matrices \mathcal{A}_j and thus also on A_0 , the block \tilde{J}_2 and consequently also the limiting state space unit root structure might be subject to further restrictions.

4.1.3. The Conformable Index Set and the Closure of Δ_Γ

The previous subsection shows that the closure of Δ_Γ does not only contain systems corresponding to transfer functions with multi-index smaller or equal to Γ , but also systems that are related in a different way that is formalized below.

Definition 10 (Conformable index set). Given a multi-index $\Gamma = (\Omega_S, p, \alpha_\bullet)$, the set of conformable multi-indices $\mathcal{K}(\Gamma)$ contains all multi-indices $\tilde{\Gamma} = (\tilde{\Omega}_S, \tilde{p}, \tilde{\alpha}_\bullet)$, where:

- The pair $(\tilde{\Omega}_S, \tilde{p})$ with corresponding matrix $\tilde{\mathcal{A}}_u$ in canonical form extends (Ω_S, p) with corresponding matrix \mathcal{A}_u in canonical form, i.e., there exists a permutation matrix S such that

$$S \tilde{\mathcal{A}}_u S' = \begin{bmatrix} \mathcal{A}_u & 0 \\ 0 & \tilde{J}_2 \end{bmatrix} \quad \text{and} \quad S \tilde{p} = \begin{bmatrix} p \\ \tilde{p}_2 \end{bmatrix},$$

- $\tilde{\alpha}_\bullet \leq \alpha_\bullet$.
- $\tilde{n}_u + \tilde{n}_\bullet = n_u + n_\bullet$.

Please note that the definition implies $\Gamma \in \mathcal{K}(\Gamma)$. The importance of the set $\mathcal{K}(\Gamma)$ is clarified in the following theorem:

Theorem 3. Transfer functions corresponding to state space realizations with multi-index $\tilde{\Gamma} \leq \Gamma$ are contained in the set $\pi(\overline{\Delta_\Gamma})$. The set $\pi(\overline{\Delta_\Gamma})$ is contained in the union of all sets $M_{\tilde{\Gamma}}$ for $\tilde{\Gamma} \leq \Gamma$ with $\tilde{\Gamma}$ conformable to Γ , i.e.,

$$\bigcup_{\tilde{\Gamma} \leq \Gamma} M_{\tilde{\Gamma}} \subset \pi(\overline{\Delta_\Gamma}) \subset \bigcup_{\tilde{\Gamma} \in \mathcal{K}(\Gamma)} \bigcup_{\tilde{\Gamma} \leq \tilde{\Gamma}} M_{\tilde{\Gamma}}.$$

Theorem 3 provides a characterization of the transfer functions corresponding to systems in the closure of Δ_Γ . The conformable set $\mathcal{K}(\Gamma)$ plays a key role here, since it characterizes the set of all minimal systems that can be obtained as limits of convergent sequences from within the set Δ_Γ . Conformable indices extend the matrix \mathcal{A}_u corresponding to the unit root structure by the block \tilde{J}_2 .

The second inclusion in Theorem 3 is potentially strict, depending on the Kronecker indices α_\bullet in Γ . Equality holds, e.g., in the following case:

Corollary 3. For every multi-index Γ with $n_\bullet = 0$ the set of conformable indices consists only of Γ , which implies $\pi(\overline{\Delta_\Gamma}) = \bigcup_{\tilde{\Gamma} \leq \Gamma} M_{\tilde{\Gamma}}$.

4.2. The Closure of M_Γ

It remains to investigate the closure of M_Γ in M . Hannan and Deistler (1988, Theorem 2.6.5 (ii) and Remark 3, p. 73) show that for any order n , there exist Kronecker indices $\alpha_{\bullet,g} = \alpha_{\bullet,g}(n)$ corresponding to the generic neighborhood $M_{\alpha_{\bullet,g}}$ for transfer functions of order n such that

$$M_{\bullet,n} := \bigcup_{\alpha_\bullet | n_\bullet(\alpha_\bullet) = n} M_{\alpha_\bullet} \subset \overline{M_{\alpha_{\bullet,g}}},$$

where $M_{\alpha_\bullet} := \pi(\Delta_{\alpha_\bullet})$. Here $M_{\bullet,n}$ denotes the set of all transfer functions of order n with state space realizations (A, B, C) satisfying $\lambda_{|\max|}(A) < 1$. Every transfer function in $M_{\bullet,n}$ can be approximated by a sequence of transfer functions in $M_{\alpha_{\bullet,g}}$.

It can be easily seen that a generic neighborhood also exists for systems with state space unit root structure Ω_S and without stable subsystem: Set the structure indices p to have a minimal number of elements restricted in p.u.t. sub-blocks of \mathcal{B}_u , i.e., for any block $\mathcal{B}_{k,h_k,j} \in \mathbb{C}^{n_{k,h_k,j} \times s}$, or $\mathcal{B}_{k,h_k,j} \in \mathbb{R}^{n_{k,h_k,j} \times s}$ in case of a real unit root, set the corresponding structure indices to $p = [1, \dots, n_{k,h_k,j}]$. Any p.u.t. matrix can be approximated by a matrix in this generic neighborhood with some positive entries restricted by the p.u.t. structure tending to zero. Combining these results with Theorem 3 implies the existence of a generic neighborhood for the canonical form considered in this paper:

Theorem 4. Let $M(\Omega_S, n_\bullet)$ be the set of all transfer functions $k(z) \in M_{n_u(\Omega_S)+n_\bullet}$ with state space unit root structure Ω_S . For every Ω_S and n_\bullet , there exists a multi-index $\Gamma_g := \Gamma_g(\Omega_S, n_\bullet)$ such that

$$M(\Omega_S, n_\bullet) \subset \overline{M_{\Gamma_g}}. \tag{14}$$

Moreover, it holds that $M(\Omega_S, n_\bullet) \subset \overline{M_{\alpha_{\bullet,g}(n)}}$ for every Ω_S and n_\bullet satisfying $n_u(\Omega_S) + n_\bullet \leq n$.

Theorem 4 is the basis for choosing a generic multi-index Γ for maximizing the pseudo likelihood function. For every Ω_S and n_\bullet there exists a generic piece that—in its closure—contains all transfer functions of order $n_u(\Omega_S) + n_\bullet$ and state space unit root structure Ω_S : The set of transfer functions corresponding to the multi-index with the largest possible structure indices p in the sense of Definition 9 (iii) and generic Kronecker indices for the stable subsystem. Choosing these sets and their corresponding parameter spaces as model sets is, therefore, the most convenient choice for numerical maximization, if only Ω_S and n_\bullet are known.

If, e.g., only an upper bound for the system order n is known and the goal is only to obtain consistent estimators, using $\alpha_{\bullet,g}(n)$ is a feasible choice, since all transfer functions in the closure of the set $M_{\alpha_{\bullet,g}(n)}$ can be approximated arbitrarily well, regardless of their potential state space unit root structure Ω_S , $n_u(\Omega_S) \leq n$. For testing hypotheses, however, it is important to understand the topological relations between sets corresponding to different multi-indices Γ . In the following we focus on the multi-indices $\Gamma_g(\Omega_S, n_\bullet)$ for arbitrary Ω_S and n_\bullet .

The closure of $M(\Omega_S, n_\bullet)$ contains also transfer functions that have a different state space unit root structure than Ω_S . Considering convergent sequences of state space realizations $(A_j, B_j, C_j)_{j \in \mathbb{N}}$ of transfer functions in $M(\Omega_S, n_\bullet)$, the state space unit root structure of $(A_0, B_0, C_0) := \lim_{j \rightarrow \infty} (A_j, B_j, C_j)$ may differ in three ways:

- For sequences $(A_j, B_j, C_j)_{j \in \mathbb{N}}$ in canonical form rows of $B_{u,j}$ can tend to zero, which reduces the state space unit root structure as discussed in Section 4.1.1.
- Stable eigenvalues of A_j may converge to the unit circle, thereby extending the unit root structure.
- Off-diagonal entries of the sub-block $\mathcal{A}_{u,j}$ of $A_j = T_j A_j T_j^{-1}$ may be converging to zeros in the sub-block $\mathcal{A}_{u,0}$ of the limit $A_0 = T_0 A_0 T_0^{-1}$ in canonical form, resulting in a different attainable state space unit root structure. Here $T_j \in \mathbb{C}^{n \times n}$ for all $j \in \mathbb{N}$ are regular matrices transforming A_j to canonical form and $T_0 \in \mathbb{C}^{n \times n}$ transforms A_0 accordingly.

The first change of Ω_S described above results in a transfer function with smaller state space unit root structure according to Definition 9 (ii). The implications of the other two cases are summarized in the following definition:

Definition 11 (Attainable unit root structures). For given n_\bullet and Ω_S the set $\mathcal{A}(\Omega_S, n_\bullet)$ of attainable unit root structures contains all pairs $(\tilde{\Omega}_S, \tilde{n}_\bullet)$, where $\tilde{\Omega}_S$ with corresponding matrix $\tilde{\mathcal{A}}_u$ in canonical form extends Ω_S with corresponding matrix \mathcal{A}_u in canonical form, i.e., there exists a permutation matrix S such that

$$S \mathcal{A}_u S' = \begin{bmatrix} \tilde{\mathcal{A}}_u & J_{12} \\ 0 & J_2 \end{bmatrix},$$

where $\tilde{\mathcal{A}}_u$ can be obtained by replacing off-diagonal entries in \mathcal{A}_u by zeros and where $\tilde{n}_\bullet := n_\bullet - d_j$ with d_j the dimension of $J_2 \in \mathbb{C}^{d_j \times d_j}$.

Remark 17. It is a direct consequence of the definition of $\mathcal{A}(\Omega_S, n_\bullet)$ that $(\tilde{\Omega}_S, \tilde{n}_\bullet) \in \mathcal{A}(\Omega_S, n_\bullet)$ implies $\mathcal{A}(\tilde{\Omega}_S, \tilde{n}_\bullet) \subset \mathcal{A}(\Omega_S, n_\bullet)$.

Theorem 5.

- (i) M_Γ is T_{pt} -open in $\overline{M_\Gamma}$ (see Definition 8 for a definition of T_{pt}).

(ii) For every generic multi-index Γ_g corresponding to Ω_S and n_\bullet it holds that

$$\begin{aligned} \pi(\overline{\Delta_{\Gamma_g}}) &\subset \bigcup_{\check{\Gamma} \in \mathcal{K}(\Gamma_g)} \bigcup_{\check{\Gamma} \leq \check{\Gamma}} M_{\check{\Gamma}} \\ &\subset \bigcup_{(\check{\Omega}_S, \check{n}_\bullet) \in \mathcal{A}(\Omega_S, n_\bullet)} \bigcup_{(\check{\Omega}_S, \check{n}_\bullet) \leq (\check{\Omega}_S, \check{n}_\bullet)} M(\check{\Omega}_S, \check{n}_\bullet) = \overline{M_{\Gamma_g}}. \end{aligned}$$

Theorem 5 has important consequences for statistical analysis, e.g., PML estimation, since—as stated several times already—maximizing the pseudo likelihood function over Θ_Γ effectively amounts to calculating the supremum over the larger set $\overline{M_\Gamma}$. Depending on the choice of Γ the following asymptotic behavior may occur:

- If Γ is chosen correctly and the estimator of the transfer function is consistent, openness of M_Γ in its closure implies that the probability of the estimator being an interior point of M_Γ tends to one asymptotically. Since the mapping attaching the parameters to the transfer function is continuous on an open and dense set, consistency in terms of transfer functions, therefore, implies generic consistency of the parameter estimators.
- If the multi-index is incorrectly chosen to equal Γ , estimator consistency is still possible if the true multi-index $\Gamma_0 < \Gamma$, as in this case $M_{\Gamma_0} \subset \overline{M_\Gamma}$. This is in some sense not too surprising and something that is also well-known in the simpler VAR framework where consistency of OLS can be established when the true autoregressive order is smaller than the order chosen for estimation. Analogous to the lag number in the VAR case, thus, a necessary condition for consistency is to choose the system order larger or equal to the true system order.

Finally, note that Theorem 5 also implies the following result relevant for the determination of the unit root structure, further discussed in Sections 5.1.1 and 5.2.1:

Corollary 4. For every pair $(\check{\Omega}_S, \check{n}_\bullet) \in \mathcal{A}(\Omega_S, n_\bullet)$ it holds that

$$\overline{M(\check{\Omega}_S, \check{n}_\bullet)} \subset \overline{M(\Omega_S, n_\bullet)}.$$

5. Testing Commonly Used Hypotheses in the MFI(1) and I(2) Cases

This section discusses a large number of hypotheses, respectively restrictions, on cointegrating spaces, adjustment coefficients and deterministic components often tested in the empirical literature. As with the VECM framework, as discussed for the I(2) case in Section 2, testing hypotheses on the cointegrating spaces or adjustment coefficients may necessitate different reparameterizations.

5.1. The MFI(1) Case

The two by far most widely used cases of MFI(1) processes are $I(1)$ processes and seasonally (co-)integrated processes for quarterly data with state space unit root structure $((0, d_1^1), (\pi/2, d_1^2), (\pi, d_1^3))$. In general, assuming for notational simplicity $\omega_1 = 0$ and $\omega_l = \pi$, it holds that for $t > 0$ and $x_{1,\mu} = 0$ we have

$$\begin{aligned}
 y_t &= \sum_{k=1}^l C_{k,\mathbb{R}} x_{t,k,\mathbb{R}} + C_{\bullet} x_{t,\bullet} + \Phi d_t + \varepsilon_t \\
 &= C_1 x_{t,1} + \sum_{k=2}^{l-1} (C_k x_{t,k} + \bar{C}_k \bar{x}_{t,k}) + C_l x_{t,l} + C_{\bullet} x_{t,\bullet} + \Phi d_t + \varepsilon_t \\
 &= C_1 \mathcal{B}_1 \sum_{j=1}^{t-1} \varepsilon_{t-j} + 2 \sum_{k=2}^{l-1} \mathcal{R} \left(C_k \mathcal{B}_k \sum_{j=1}^{t-1} (\bar{z}_k)^{j-1} \varepsilon_{t-j} \right) + C_l \mathcal{B}_l \sum_{j=1}^{t-1} (-1)^{j-1} \varepsilon_{t-j} \\
 &\quad + C_{\bullet} \sum_{j=1}^{t-1} \mathcal{A}_{\bullet}^{j-1} \mathcal{B}_{\bullet} \varepsilon_{t-j} + C_{\bullet} \mathcal{A}_{\bullet}^{t-1} x_{1,\bullet} + \Phi d_t + \varepsilon_t \\
 &= C_1 \mathcal{B}_1 \sum_{j=1}^{t-1} \varepsilon_{t-j} + 2 \sum_{k=2}^{l-1} \sum_{j=1}^{t-1} \left(\mathcal{R}(C_k \mathcal{B}_k) \cos(\omega_k(j-1)) + \mathcal{I}(C_k \mathcal{B}_k) \sin(\omega_k(j-1)) \right) \varepsilon_{t-j} \\
 &\quad + C_l \mathcal{B}_l \sum_{j=1}^{t-1} (-1)^{j-1} \varepsilon_{t-j} + C_{\bullet} \sum_{j=1}^{t-1} \mathcal{A}_{\bullet}^{j-1} \mathcal{B}_{\bullet} \varepsilon_{t-j} + C_{\bullet} \mathcal{A}_{\bullet}^{t-1} x_{1,\bullet} + \Phi d_t + \varepsilon_t.
 \end{aligned}$$

The above equation provides an additive decomposition of $\{y_t\}_{t \in \mathbb{Z}}$ into stochastic trends and cycles, the deterministic and stationary components. The stochastic cycles at frequency $0 < \omega_k < \pi$ are, of course, given by the combination of sine and cosine terms. For the MFI(1) case this can also be seen directly from considering the real valued canonical form discussed in Remark 4, with the matrices

$$\mathcal{A}_{k,\mathbb{R}} \text{ for } k = 2, \dots, l-1, \text{ given by } \mathcal{A}_{k,\mathbb{R}} = I_{d_1^k} \otimes \begin{pmatrix} \cos(\omega_k) & -\sin(\omega_k) \\ \sin(\omega_k) & \cos(\omega_k) \end{pmatrix} \text{ in this case.}$$

The ranks of $C_k \mathcal{B}_k$ are equal to the integers d_1^k in $\Omega_S = ((\omega_1, d_1^1), \dots, (\omega_l, d_1^l))$. The number of stochastic trends is equal to d_1^1 , the number of stochastic cycles at frequency ω_k is equal to $2d_1^k$ for $k = 2, \dots, l-1$ and equal to d_1^l if $k = l$, as discussed in Section 3.

Moreover, in the MFI(1) case, d_1^k is linked to the *complex cointegrating rank* r_k at frequency ω_k , defined in Johansen (1991) and Johansen and Schaumburg (1999) in the VECM case as the rank of the matrix $\Pi_k := -a(z_k)$. For VARMA processes with arbitrary integration orders the complex cointegrating rank r_k at frequency ω_k is $r_k := \text{rank}(-k^{-1}(z_k))$, where $k(z)$ is the transfer function, with $r_k = s - d_1^k$ in the MFI(1) case. Thus, in the MFI(1) case, determination of the state space unit root structure corresponds to determination of the complex cointegrating ranks in the VECM case.

In the VECM setting, the matrix Π_k is usually factorized into $\Pi_k = \alpha_k \beta_k'$, as presented for the I(1) case in Section 2. For $\omega_k = \{0, \pi\}$ the column space of β_k gives the cointegrating space of the process at frequency ω_k . For $0 < \omega_k < \pi$ the relation between the column space of β_k and the space of CIVs and PCIVs at the corresponding frequency is more involved. The columns of β_k are orthogonal to the columns of C_k , the sub-block of \mathcal{C} from a state space realization $(\mathcal{A}, \mathcal{B}, \mathcal{C})$ in canonical form corresponding to the VAR process. Analogously, the column space of the matrix α_k , containing the so-called *adjustment coefficients*, is orthogonal to the row space of the sub-block \mathcal{B}_k of \mathcal{B} .

Both integers d_1^k and r_k are related to the dimensions of the static and dynamic cointegrating spaces in the MFI(1) case: For $\omega_k \in \{0, \pi\}$, the cointegrating rank $r_k = s - d_1^k$ coincides with the dimension of the static cointegrating space at frequency ω_k . Furthermore, the dimension of the static cointegrating space at frequency $0 < \omega_k < \pi$ is bounded from above by $r_k = s - d_1^k$, since it is spanned by at most $s - d_1^k$ vectors $\beta \in \mathbb{R}^s$ orthogonal to the complex valued matrix C_k . The dimension of the dynamic cointegrating space at $0 < \omega_k < \pi$ is equal to $2r_k = 2(s - d_1^k)$. Identifying again $\beta(z) = \beta_0 + \beta_1 z$ with the vector $[\beta_0', \beta_1']'$, a basis of the dynamic cointegrating space at $0 < \omega_k < \pi$ is then given by the column space of the product

$$\begin{bmatrix} \gamma_0 & \tilde{\gamma}_0 \\ \gamma_1 & \tilde{\gamma}_1 \end{bmatrix} := \begin{bmatrix} I_s & 0_{s \times s} \\ -\cos(\omega_k)I_s & \sin(\omega_k)I_s \end{bmatrix} \begin{bmatrix} \mathcal{R}(\beta_k) & \mathcal{I}(\beta_k) \\ -\mathcal{I}(\beta_k) & \mathcal{R}(\beta_k) \end{bmatrix},$$

with the columns of $\beta_k \in \mathbb{C}^{s \times (s-d_1^k)}$ spanning the orthogonal complement of the column space of C_k , i.e., β_k is of full rank and $\beta_k' C_k = (\mathcal{R}(\beta_k)' - i\mathcal{I}(\beta_k)') C_k = 0$. This holds true, since both factors are of full rank and $[\gamma_0', \gamma_1']'$ satisfies $(\bar{z}_k \gamma_0' + \gamma_1') C_k = 0$, which corresponds to the necessary condition given in Example 2 for the columns of $[\gamma_0', \gamma_1']'$ to be PCIVs. The latter implies $(\bar{z}_k \tilde{\gamma}_0' + \tilde{\gamma}_1') C_k = 0$ also for $[\tilde{\gamma}_0', \tilde{\gamma}_1']'$, highlighting again the additional structure of the cointegrating space emanating from the complex conjugate pairs or eigenvalues (and matrices) as discussed in Example 2.

Please note that the relations between r_k and d_1^k discussed above only hold in the MFI(1) and I(1) special cases. For higher orders of integration no such simple relations exist.

In the MFI(1) setting the deterministic component typically includes a constant, seasonal dummies and a linear trend. As discussed in Remark 6, a sufficiently rich set of deterministic components allows to absorb non-zero initial values $x_{1,\mu}$.

5.1.1. Testing Hypotheses on the State Space Unit Root Structure

Using the generic sets of transfer functions M_{Γ_g} presented in Theorem 4, we can construct pseudo likelihood ratio tests for different hypotheses $H_0 : (\Omega_S, n_\bullet) = (\Omega_{S,0}, n_{\bullet,0})$ against chosen alternatives. Note, however, that by the results of Theorem 5 the null hypothesis includes all pairs $(\Omega_S, n_\bullet) \in \mathcal{A}(\Omega_{S,0}, n_{\bullet,0})$ as well as all pairs (Ω_S, n_\bullet) that are smaller than a pair $(\tilde{\Omega}_S, \tilde{n}_\bullet) \in \mathcal{A}(\Omega_{S,0}, n_{\bullet,0})$.

As common in the VECM setting, first consider hypotheses at a single frequency ω_k . For an MFI(1) process, the hypothesis of a state space unit root structure equal to $\Omega_{S,0} = ((\omega_k, d_{1,0}^k))$ corresponds to the hypothesis of the (complex) cointegrating rank r_k at frequency ω_k being equal to $r_0 = s - d_{1,0}^k$. Maximization of the pseudo likelihood function over the set $\overline{M((\omega_k, d_{1,0}^k)), n - \delta_k d_{1,0}^k}$ —with a suitably chosen order n —leads to estimates that may be arbitrary close to transfer functions with different state space unit root structures Ω_S . These include Ω_S with additional unit root frequencies $\omega_{\tilde{k}}$, with the integers $d_{1,\tilde{k}}^k$ restricted only by the order n . Therefore, focusing on a single frequency ω_k does not rule out a more complicated true state space unit root structure. Assume $n \geq \delta_k s$ with $\delta_k = 1$ for $\omega_k \in \{0, \pi\}$ and $\delta_k = 2$ else. Corollary 4 shows that

$$\overline{M(\{\}, n)} \supset \overline{M((\omega_k, 1), n - \delta_k)} \supset \dots \supset \overline{M((\omega_k, s), n - s\delta_k)}$$

since, e.g., $((\omega_k, 1), n - \delta_k) \in \mathcal{A}(\{\}, n)$.

Analogously to the procedure of testing for the complex cointegrating rank r_k in the VECM setting, these inclusions can be employed to test for d_1^k : Start with the hypothesis of $d_1^k = s$ against the alternative of $0 \leq d_1^k < s$ and decrease the assumed d_1^k consecutively until the test does not reject the null hypothesis.

Furthermore, one can formulate hypotheses on d_1^k jointly at different frequencies ω_k . Again, there exist inclusions based on the definition of the set of attainable state space unit root structures and Corollary 4, which can be used to consecutively test hypotheses on Ω_S .

5.1.2. Testing Hypotheses on CIVs and PCIVs

Johansen (1995) considers in the I(1) case three types of hypotheses on the cointegrating space spanned by the columns of β that are each motivated by examples from economic research: The different cases correspond to different types of hypotheses related to restrictions implied by economic theory.

- (i) $H_0 : \beta = H\varphi, \beta \in \mathbb{R}^{s \times r}, H \in \mathbb{R}^{s \times t}, \varphi \in \mathbb{R}^{t \times r}, r \leq t < s$: The cointegrating space is known to be a subspace of the column space of H (which is of full column rank).

- (ii) $H'_0 : \beta = [b, \varphi], \beta \in \mathbb{R}^{s \times r}, b \in \mathbb{R}^{s \times t}, \varphi \in \mathbb{R}^{s \times r-t}, 0 < t \leq r$: Some cointegrating relations are known.
- (iii) $H''_0 : \beta = [H_1 \varphi_1, \dots, H_c \varphi_c], \beta \in \mathbb{R}^{s \times r}, H_j \in \mathbb{R}^{s \times t_j}, \varphi_j \in \mathbb{R}^{t_j \times r_j}, r_j \leq t_j \leq s$, for $j = 1, \dots, c$ such that $\sum_{j=1}^c r_j = r$. Cointegrating relations are known to be in the column spaces of matrices H_k (which are of full column rank).

As discussed in Example 1, cointegration at $\omega_k = 0$ occurs if and only if a vector β_j satisfies $\beta'_j C_1 = 0$. In other words, the column space of C_1 is the orthocomplement of the cointegrating space spanned by the columns of β and hypotheses on β restrict entries of C_1 .

The first type of hypothesis, H_0 , implies that the column space of C_1 is equal to the orthocomplement of the column space of $H\varphi$. Assume w.l.o.g. $H \in O_{s,t}, \varphi_{\perp} \in O_{t,t-r}$ and $H_{\perp} \in O_{s,s-t}$, such that the columns of $[H\varphi_{\perp}, H_{\perp}]$ form an orthonormal basis for the orthocomplement of the cointegrating space. Consider now the mapping:

$$C_1^r(\check{\theta}_L, \theta_R) := \left[H \cdot \check{R}_L(\check{\theta}_L)' \begin{bmatrix} I_{t-r} \\ 0_{r \times (t-r)} \end{bmatrix}, H_{\perp} \right] \cdot R_R(\theta_R), \tag{15}$$

where $\check{R}_L(\check{\theta}_L) := \prod_{i=1}^{t-r} \prod_{j=1}^r R_{t,i,t-r+j}(\theta_{L,r(i-1)+j}) \in \mathbb{R}^{t \times t}$ and $R_R(\theta_R) \in \mathbb{R}^{(s-r) \times (s-r)}$ as in Lemma 1. From this one can derive a parameterization of the set of matrices C_1^r corresponding to H_0 , analogously to Lemma 1. The difference of the number of free parameters under the null hypothesis and under the alternative is the difference between the number of free parameters in $\theta_L \in [0, 2\pi)^{r(s-r)}$ and $\check{\theta}_L \in [0, 2\pi)^{r(t-r)}$, implying a reduction of the number of free parameters of $r(s-t)$ under the null hypothesis. This necessarily coincides with the number of degrees of freedom of the corresponding test statistic in the VECM setting (cf. Johansen 1995, Theorem 7.2).

The second type of hypothesis, H'_0 , is also straightforwardly parameterized: In this case a subspace of the cointegrating space is known and given by the column space of $b \in \mathbb{R}^{s \times t}$. Assume w.l.o.g. $b \in O_{s,t}$. The orthocomplement of $\beta = [b, \varphi]$ is given by the set of matrices C_1 satisfying the restriction $b' C_1 = 0$, i.e., the set $O_{s,d_1}(b)$ defined in (13). The parameterization of this set has already been discussed. The reduction of the number of free parameters under the null hypothesis is $t(s-r)$ which again coincides with the number of degrees of freedom of the corresponding test statistic in the VECM setting (cf. Johansen 1995, Theorem 7.3).

Finally, the third type of hypothesis, H''_0 , is the most difficult to parameterize in our setting. As an illustrative example consider the case $H''_0 : \beta = [H_1 \varphi_1, H_2 \varphi_2], \beta \in \mathbb{R}^{s \times r}, H_1 \in \mathbb{R}^{s \times t_1}, H_2 \in \mathbb{R}^{s \times t_2}, \varphi_1 \in \mathbb{R}^{t_1 \times r_1}, \varphi_2 \in \mathbb{R}^{t_2 \times r_2}, r_j \leq t_j \leq s$ and $r_1 + r_2 = r$. W.l.o.g. choose $H_b \in O_{s,t_b}$ such that its columns span the t_b -dimensional intersection of the column spaces of H_1 and H_2 and choose $\tilde{H}_j \in O_{s,\tilde{t}_j}(H_b), j = 1, 2$ such that the columns of \tilde{H}_j and H_b span the column space of H_j . Define $\tilde{H} := [\tilde{H}_1, \tilde{H}_2, H_b] \in O_{s,\tilde{t}}$, with $\tilde{t} = \tilde{t}_1 + \tilde{t}_2 + t_b$. Let w.l.o.g. $\tilde{H}_{\perp} \in O_{s,s-\tilde{t}}(\tilde{H})$ and define $p_j := \min(r_j, \tilde{t}_j), q_j := \max(r_j, \tilde{t}_j)$ for $j = 1, 2$ and $p_b = q_1 - \tilde{t}_1 + q_2 - \tilde{t}_2$. A parameterization of $\beta^r \in O_{s,r}$ satisfying the restrictions under the null hypothesis can be derived from the following mapping:

$$\beta^r(\theta_H, \theta_{R,\beta}) := \tilde{H} \cdot R_H(\theta_H)' \begin{bmatrix} I_{p_1} & 0_{p_1 \times p_2} & 0_{p_1 \times p_b} \\ \mathbf{0}_{(q_1-r_1) \times p_1} & \mathbf{0}_{(q_1-r_1) \times p_2} & \mathbf{0}_{(q_1-r_1) \times p_b} \\ 0_{p_2 \times p_1} & I_{p_2} & 0_{p_2 \times p_b} \\ 0_{(q_2-r_2) \times p_1} & \mathbf{0}_{(q_2-r_2) \times p_2} & \mathbf{0}_{(q_2-r_2) \times p_b} \\ 0_{p_b \times p_1} & 0_{p_b \times p_2} & I_{p_b} \\ \mathbf{0}_{(\tilde{t}-q_1-q_2) \times p_1} & \mathbf{0}_{(\tilde{t}-q_1-q_2) \times p_2} & \mathbf{0}_{(\tilde{t}-q_1-q_2) \times p_b} \end{bmatrix} \cdot R_R(\theta_{R,\beta}),$$

where $R_R(\theta_{R,\beta}) \in \mathbb{R}^{r \times r}$ as in Lemma 1 and $R_H(\theta_H) := R_H((\theta_{H_1}, \theta_{H_2}, \theta_{H_b})) := R_{H_1}(\theta_{H_1}) R_{H_2}(\theta_{H_2}) R_{H_b}(\theta_{H_b}) \in \mathbb{R}^{\tilde{t} \times \tilde{t}}$ is a product of Givens rotations corresponding to the entries in the blocks highlighted by bold font. The three matrices are defined as follows:

$$\begin{aligned}
 R_{H_1}(\boldsymbol{\theta}_{H_1}) &:= \prod_{i=1}^{p_1} \prod_{j=1}^{\tilde{t}-q_2-r_1} R_{t_i, \delta_{H_1}(j)+j}(\boldsymbol{\theta}_{H_1, (\tilde{t}-q_2-r_1)(i-1)+j}), \quad \delta_{H_1}(j) := \begin{cases} p_1 & \text{if } j \leq q_1 - r_1 \\ \tilde{t}_1 + \tilde{t}_2 + p_b & \text{else,} \end{cases} \\
 R_{H_2}(\boldsymbol{\theta}_{H_2}) &:= \prod_{i=1}^{p_2} \prod_{j=1}^{\tilde{t}-q_1-r_2} R_{t_i, p_1+i, \delta_{H_2}(j)+j}(\boldsymbol{\theta}_{H_2, (\tilde{t}-q_1-r_2)(i-1)+j}), \quad \delta_{H_2}(j) := \begin{cases} \tilde{t}_1 + p_2 & \text{if } j \leq q_2 - r_2 \\ \tilde{t}_1 + \tilde{t}_2 + p_b & \text{else,} \end{cases} \\
 R_{H_b}(\boldsymbol{\theta}_{H_b}) &:= \prod_{i=1}^{p_b} \prod_{j=1}^{\tilde{t}-q_1-q_2} R_{t_i, p_1+p_2+i, \tilde{t}_1+\tilde{t}_2+p_b+j}(\boldsymbol{\theta}_{H_b, (\tilde{t}-q_1-q_2)(i-1)+j}).
 \end{aligned}$$

Consequently, a parameterization of the orthocomplement of the cointegrating space is based on the mapping:

$$C_1^r(\boldsymbol{\theta}_H, \boldsymbol{\theta}_{R,C}) := \left[\tilde{H} \cdot R_H(\boldsymbol{\theta}_H)' \begin{bmatrix} 0_{p_1 \times (q_1-r_1)} & 0_{p_1 \times (q_2-r_2)} & 0_{p_1 \times (\tilde{t}-q_1-q_2)} \\ I_{q_1-r_1} & 0_{(q_1-r_1) \times (q_2-r_2)} & 0_{(q_1-r_1) \times (\tilde{t}-q_1-q_2)} \\ 0_{p_2 \times (q_1-r_1)} & 0_{p_2 \times (q_2-r_2)} & 0_{p_2 \times (\tilde{t}-q_1-q_2)} \\ 0_{(q_2-r_2) \times (q_1-r_1)} & I_{q_2-r_2} & 0_{(q_2-r_2) \times (\tilde{t}-q_1-q_2)} \\ 0_{p_b \times (q_1-r_1)} & 0_{p_b \times (q_2-r_2)} & 0_{p_b \times (\tilde{t}-q_1-q_2)} \\ 0_{(\tilde{t}-q_1-q_2) \times (q_1-r_1)} & 0_{(\tilde{t}-q_1-q_2) \times (q_2-r_2)} & I_{\tilde{t}-q_1-q_2} \end{bmatrix}, \tilde{H}_\perp \right] \cdot R_R(\boldsymbol{\theta}_{R,C}),$$

where $R_H(\boldsymbol{\theta}_H) \in \mathbb{R}^{\tilde{t} \times \tilde{t}}$ as above and $R_R(\boldsymbol{\theta}_{R,C}) \in \mathbb{R}^{(s-r) \times (s-r)}$ as in Lemma 1. Please note that for all $\boldsymbol{\theta}_H, \boldsymbol{\theta}_{R,\beta}$ and $\boldsymbol{\theta}_{R,C}$ it holds that $\beta^r(\boldsymbol{\theta}_H, \boldsymbol{\theta}_{R,\beta})' C_1^r(\boldsymbol{\theta}_H, \boldsymbol{\theta}_{R,C}) = 0_{r \times (s-r)}$. The number of parameters restricted under H_0'' is equal to $r_1(q_1 - r_1) + r_2(q_2 - r_2) + (r_1 + r_2)(\tilde{t} - q_1 - q_2) + (s - r)(s - r + 1)/2$, and thus, through q_1 and q_2 , depends on the dimension t_b of the intersection of the columns spaces of H_1 and H_2 . The reduction of the number of free parameters matches the degrees of freedom of the test statistics in Johansen (1995, Theorem 7.5), if β is identified, which is the case if $r_1 \leq \tilde{t}_1$ and $r_2 \leq \tilde{t}_2$.

Using the mapping $\beta^r(\cdot)$ as a basis for a parameterization allows to introduce another type of hypotheses of the form:

(iv) $H_0''' : \beta_\perp = \mathcal{C}_1 = [H_1\varphi_1, \dots, H_c\varphi_c], \beta_\perp \in \mathbb{R}^{s \times (s-r)}, H_j \in O_{s,t_j}, \varphi_j \in O_{t_j,r_j}, r_j \leq t_j \leq s$, for $j = 1, \dots, c$ such that $\sum_{j=1}^c r_j = s - r$. The ortho-complement of the cointegrating space is contained in the column spaces of the (full rank) matrices H_k .

This type of hypothesis allows, e.g., to test for the presence of cross-unit cointegrating relations (cf. Wagner and Hlouskova 2009, Definition 1) in, e.g., multi-country data sets.

Hypotheses on the cointegrating space at frequency $\omega_k = \pi$ can be treated analogously to hypotheses on the cointegrating space at frequency $\omega_k = 0$.

Testing hypotheses on cointegrating spaces at frequencies $0 < \omega_k < \pi$ has to be discussed in more detail, as one also has to consider the space spanned by PCIVs, compare Example 2. There are $2(s - d_1^k)$ linearly independent PCIVs of the form $\beta(z) = \beta_0 + \beta_1 z$. Every PCIV corresponds to a vector $z_k \beta_0 + \beta_1 \in \mathbb{C}^s$ orthogonal to \mathcal{C}_k and consequently hypotheses on the space spanned by PCIVs can be transformed to hypotheses on the complex column space of $\mathcal{C}_k \in \mathbb{C}^{s \times d_1^k}$.

Consider, e.g., an extension of the first type of hypothesis of the form

$$\begin{aligned}
 H_0^k : \begin{bmatrix} \gamma_0 & \tilde{\gamma}_0 \\ \gamma_1 & \tilde{\gamma}_1 \end{bmatrix} &= \begin{bmatrix} I_s & 0_{s \times s} \\ -\cos(\omega_k) I_s & \sin(\omega_k) I_s \end{bmatrix} \begin{bmatrix} (\tilde{H}_0 \tilde{\phi}_0 - \tilde{H}_1 \tilde{\phi}_1) & (\tilde{H}_0 \tilde{\phi}_1 + \tilde{H}_1 \tilde{\phi}_0) \\ -(\tilde{H}_0 \tilde{\phi}_1 + \tilde{H}_1 \tilde{\phi}_0) & (\tilde{H}_0 \tilde{\phi}_0 - \tilde{H}_1 \tilde{\phi}_1) \end{bmatrix} \\
 &= \begin{bmatrix} I_s & 0_{s \times s} \\ -\cos(\omega_k) I_s & \sin(\omega_k) I_s \end{bmatrix} \begin{bmatrix} \tilde{H}_0 & \tilde{H}_1 \\ -\tilde{H}_1 & \tilde{H}_0 \end{bmatrix} \begin{bmatrix} \tilde{\phi}_0 & \tilde{\phi}_1 \\ -\tilde{\phi}_1 & \tilde{\phi}_0 \end{bmatrix},
 \end{aligned}$$

with $\tilde{H}_0, \tilde{H}_1 \in \mathbb{R}^{s \times t}, \tilde{\phi}_0, \tilde{\phi}_1 \in \mathbb{R}^{t \times r}, r \leq t < s$, which implies that the column space of C_k is equal to the orthocomplement of the column space of $(\tilde{H}_0 + i\tilde{H}_1)(\tilde{\phi}_0 + i\tilde{\phi}_1)$. This general hypothesis encompasses, e.g., the hypothesis $[\gamma'_0, \gamma'_1]' = H\phi = [H'_0, H'_1]'\phi$, with $H \in \mathbb{R}^{2s \times t}, H_0, H_1 \in \mathbb{R}^{s \times t}, \phi \in \mathbb{R}^{t \times r}$, by setting $\tilde{\phi}_0 := \tilde{\phi}_1 := \tilde{\phi}, \tilde{H}_0 := H_0$ and $\tilde{H}_1 := -(\cos(\omega_k)H_0 + H_1) / \sin(\omega_k)$. The extension is tailored to include the pairwise structure of PCIVs and to simplify transformation into hypotheses on the complex matrix C_k used in the parameterization. The parameterization of the set of matrices corresponding to H_0^k is derived from a mapping of the form given in (15), with $\check{R}_L(\check{\theta}_L)$ and $R_R(\theta_R)$ replaced by $\check{Q}_L(\check{\varphi}_L) := \prod_{i=1}^{t-r} \prod_{j=1}^r Q_{t,i,t-r+j}(\varphi_{L,r(i-1)+j}) \in \mathbb{R}^{t \times t}$ and $D_d(\varphi_D)Q_R(\varphi_R)$ as in Lemma 2.

Similarly, the three other types of hypotheses on the cointegrating spaces considered above can be extended to hypotheses on the space of PCIVs in the MFI(1) case. They translate into hypotheses on complex valued matrices β_k orthogonal to C_k . To parameterize the set of matrices restricted according to these null hypotheses, Lemma 2 is used. Thus, the restrictions implied by the extensions of all four types of hypotheses to hypotheses on the dynamic cointegrating spaces at frequencies $0 < \omega_k < \pi$ for MFI(1) processes can be implemented using Givens rotations.

A different case of interest is the hypothesis of at least m linearly independent CIVs $b_j \in \mathbb{R}^s, j = 1, \dots, m$ with $0 < m \leq s - d_1^k$, i.e., an m -dimensional static cointegrating space at frequency $0 < \omega_k < \pi$, which we discuss as another illustrative example to the procedure for the case of cointegration at complex unit roots.

For the dynamic cointegrating space, this hypothesis implies the existence of $2m$ linearly independent PCIVs of the form $\beta_1(z) = b_j$ and $\beta_2(z) = b_j z, j = 1, \dots, m$. In light of the discussion above the necessary condition for these two polynomials to be PCIVs is equivalent to $b'_j C_k = 0$, for $j = 1, \dots, m$. This restriction is similar to H'_0 discussed above, except for the fact that the cointegrating vectors b_j are not fully specified. This hypothesis is equivalent to the existence of an m -dimensional real kernel of C_k . A suitable parameterization is derived from the following mapping

$$C(\theta_b, \varphi) := R_L(\theta_b) \begin{bmatrix} 0_{m \times d_1^k} \\ C_U(\varphi) \end{bmatrix},$$

where $\theta_b \in [0, 2\pi)^{m(s-m)}$ and $C_U(\varphi) := C_U(\varphi_L, \varphi_D, \varphi_R) \in U_{s-m, d_1^k}$ as in Lemma 2. The difference in the number of free parameters without restrictions and with restrictions is equal to $m(s - m)$.

The hypotheses can also be tested jointly for the cointegrating spaces of several unit roots.

5.1.3. Testing Hypotheses on the Adjustment Coefficients

As in the case of hypotheses on the cointegrating spaces β_k , hypotheses on the adjustment coefficients α_k are typically formulated as hypotheses on the column spaces of α_k . We only focus on hypotheses on the real valued α_1 corresponding to frequency zero. Analogous hypotheses may be considered for α_k at frequencies $\omega_k \neq 0$, using the same ideas.

The first type of hypothesis on α_1 is of the form $H_\alpha : \alpha_1 = A\psi, A \in \mathbb{R}^{s \times t}, \psi \in \mathbb{R}^{t \times r}$ and therefore, can be rewritten as $B_1 A \psi = 0$. W.l.o.g. let $A \in O_{s,t}$ and $A_\perp \in O_{s,s-t}$. We deal with this type of hypothesis as with $H_0 : \beta = H\varphi$ in the previous section by simply reversing the roles of C_1 and B_1 . We, therefore, consider the set of feasible matrices B'_1 as a subset in $O_{s,s-r}$ and use the mapping $B'_1(\check{\theta}_L, \theta_R) = [A\check{R}_L(\check{\theta}_L)'[I_{t-r}, 0_{r \times (t-r)}]', A_\perp]R_R(\theta_R)$ to derive a parameterization, while C'_1 is restricted to be a p.u.t. matrix and the set of feasible matrices C'_1 is parameterized accordingly.

As a second type of hypothesis Juselius (2006, sct. 11.9, p. 200) discusses $H'_\alpha : \alpha_{1,\perp} = H\psi, H \in \mathbb{R}^{s \times t}, \psi \in \mathbb{R}^{t \times (s-r)}$, linked to the absence of permanent effects of shocks $H_\perp \varepsilon_t$ on any of the variables of the system. Assume w.l.o.g. $H_\perp \in O_{s,s-t}$. Using the parameterization of $O_{s-r}(H_\perp)$ defined in (13) for the set of feasible matrices B'_1 and the parameterization of the set of p.u.t. matrices for the set of feasible matrices C'_1 , implements this restriction.

The restrictions on H_α reduce the number of free parameters by $r(s - t)$ and the restrictions implied by H'_α lead to a reduction by $t(s - r)$ free parameters, compared to the unrestricted case,

which matches in both cases the number of degrees of freedom of the corresponding test statistic in the VECM framework.

5.1.4. Restrictions on the Deterministic Components

Including an unrestricted constant in the VECM equation $\Delta_0 y_t = \varepsilon_t + \Phi_0$ leads to a linear trend in the solution process $y_t = \sum_{j=1}^t (\varepsilon_j + \Phi_0) + y_1 = \sum_{j=1}^t \varepsilon_j + y_1 + \Phi_0 t$, for $t > 1$. If one restricts the constant to $\Phi_0 = \alpha \tilde{\Phi}_0, \tilde{\Phi}_0 \in \mathbb{R}^r$ in a general VECM equation as given in (4), with $\Pi = \alpha \beta'$ of rank r , no summation to linear trends in the solution process occurs, while a constant non-zero mean is still present in the cointegrating relations, i.e., the process $\{\beta' y_t\}_{t \in \mathbb{Z}}$. Analogously an unrestricted linear trend $\Phi_1 t$ in the VECM equation leads to a quadratic trend of the form $\Phi_1 t(t-1)/2$ in the solution process, which is excluded by the restriction $\Phi_1 t = \alpha \tilde{\Phi}_1 t$.

In the VECM framework, compare Johansen (1995, sct. 5.7, p. 81), five restrictions related to the coefficients corresponding to the constant and the linear trend are commonly considered:

1. $H(r) : \Phi d_t = \Phi_1 t + \Phi_0$, i.e., unrestricted constant and linear trend,
2. $H^*(r) : \Phi d_t = \alpha \tilde{\Phi}_1 t + \Phi_0$, i.e., unrestricted constant, linear trend restricted to cointegrating relations,
3. $H_1(r) : \Phi d_t = \Phi_0$, i.e., unrestricted constant, no linear trend,
4. $H_1^*(r) : \Phi d_t = \alpha \tilde{\Phi}_0$, i.e., constant restricted to cointegrating relations, no linear trend,
5. $H_2(r) : \Phi d_t = 0$, i.e., no deterministic components present,

with $\Phi_0, \Phi_1 \in \mathbb{R}^s$ and $\tilde{\Phi}_0, \tilde{\Phi}_1 \in \mathbb{R}^r$ and the following consequences for the solution processes: Under $H(r)$ the solution process contains a quadratic trend in the direction of the common trends, i.e., in $\{\beta'_\perp y_t\}_{t \in \mathbb{Z}}$, and a linear trend in the direction of the cointegrating relations, i.e., in $\{\beta' y_t\}_{t \in \mathbb{Z}}$. Under $H^*(r)$ the quadratic trend is not present. $H_1(r)$ features a linear trend only in the directions of the common trends, $H_2(r)$ a constant only in these directions. Under $H_1^*(r)$ the constant is also present in the directions of the cointegrating relations.

In the state space framework the deterministic components can be added in the output equation $y_t = Cx_t + \Phi d_t + \varepsilon_t$, compare (9). Consequently, the above considered hypotheses can be imposed by formulating linear restrictions on Φ . These can be directly parameterized by including the following deterministic components in the five considered cases:

1. $H(r) : \Phi d_t = C_1 \tilde{\Phi}_2 t^2 + \Phi_1 t + \Phi_0$,
2. $H^*(r) : \Phi d_t = \Phi_1 t + \Phi_0$,
3. $H_1(r) : \Phi d_t = C_1 \tilde{\Phi}_1 t + \Phi_0$,
4. $H_1^*(r) : \Phi d_t = \Phi_0$,
5. $H_2(r) : \Phi d_t = C_1 \tilde{\Phi}_0$,

where $\Phi_0, \Phi_1 \in \mathbb{R}^s$ and $\tilde{\Phi}_0, \tilde{\Phi}_1, \tilde{\Phi}_2 \in \mathbb{R}^{d_1}$. The component $C_1 \tilde{\Phi}_0$ captures the influence of the initial value $C_1 x_{1,1}$ in the output equation.

In the VECM framework for the seasonal MFI(1) case, with $\Pi_k = \alpha_k \beta'_k$ of rank r_k for $0 < \omega_k < \pi$, the deterministic component usually includes restricted seasonal dummies of the form $\alpha_k \tilde{\Phi}_k z_k^t + \overline{\alpha_k \tilde{\Phi}_k (z_k)^t}$, $\tilde{\Phi}_k \in \mathbb{C}^{r_k}$ to avoid summation in the directions of the stochastic trends. The state space framework allows to straightforwardly include seasonal dummies in the output equation in the form of $\Phi_k z_k^t + \overline{\Phi_k (z_k)^t}$, $\Phi_k \in \mathbb{C}^s$. Again, it is of interest whether these components are unrestricted or whether they take the form of $C_k \tilde{\Phi}_k z_k^t + \overline{C_k \tilde{\Phi}_k (z_k)^t}$, $\tilde{\Phi}_k \in \mathbb{C}^{d_k^1}$, similarly allowing for a reinterpretation of these components as influence of the initial values $x_{1,k}$ on the output.

Please note that $\Phi_k z_k^t + \overline{\Phi_k (z_k)^t}$ is equivalently given by $\check{\Phi}_{k,1} \sin(\omega_k t) + \check{\Phi}_{k,2} \cos(\omega_k t)$ using real coefficients $\check{\Phi}_{k,1}, \check{\Phi}_{k,2} \in \mathbb{R}^s$ and the desired restrictions can be implemented accordingly.

5.2. The I(2) Case

The state space unit root structure of I(2) processes is of the form $\Omega_S = ((0, d_1^1, d_2^1))$, where the integer d_1^1 equals the dimension of $x_{t,1}^E$, and d_2^1 equals the dimension of $[(x_{t,2}^G)', (x_{t,2}^E)']'$. Recall that the solution for $t > 0$ and $x_{1,\mu} = 0$ of the system in canonical form in this setting is given by

$$\begin{aligned} y_t &= C_{1,1}^E x_{t,1}^E + C_{1,2}^G x_{t,2}^G + C_{1,2}^E x_{t,2}^E + C_{\bullet} x_{t,\bullet} + \Phi d_t + \varepsilon_t \\ &= C_{1,1}^E \mathcal{B}_{1,2,1} \sum_{k=1}^{t-1} \sum_{j=1}^k \varepsilon_{t-j} + (C_{1,1}^E \mathcal{B}_{1,1} + C_{1,2}^G \mathcal{B}_{1,2,1} + C_{1,2}^E \mathcal{B}_{1,2,2}) \sum_{j=1}^{t-1} \varepsilon_{t-j} \\ &\quad + C_{\bullet} \sum_{j=1}^{t-1} \mathcal{A}_{\bullet}^{j-1} \mathcal{B}_{\bullet} \varepsilon_{t-j} + C_{\bullet} \mathcal{A}_{\bullet}^{t-1} x_{1,\bullet} + \Phi d_t + \varepsilon_t. \end{aligned}$$

For VAR processes integrated of order two the integers d_1^1 and d_2^1 of the corresponding state space unit root structure are linked to the ranks of the matrices $\Pi = \alpha\beta'$ (denoted as $r = r_0$) and $\alpha'_{\perp} \Gamma \beta_{\perp} = \xi \eta'$ (denoted as $m = r_1$) in the VECM setting, as discussed in Section 2. It holds that $r = s - d_2^1$ and $m = d_2^1 - d_1^1$. The relation of the state space unit root structure to the cointegration indices r_0, r_1, r_2 was also discussed in Section 3.

Again, both the integers d_1^1 and d_2^1 and the ranks r and m , and consequently also the indices r_0, r_1 and r_2 , are closely related to the dimensions of the spaces spanned by CIVs and PCIVs. In the I(2) case the static cointegrating space of order $((0, 2), (0, 1))$ is the orthocomplement of the column space of $C_{1,1}^E$ and thus of dimension $s - d_1^1$. The dimension of the space spanned by CIVs of order $((0, 2), \{\})$ is equal to $s - d_2^1 - r_{c,G}$, where $r_{c,G}$ denotes the rank of $C_{1,2}^G$, since this space is the orthocomplement of the column space of $[C_{1,1}^E, C_{1,2}^G, C_{1,2}^E]$. The space spanned by the PCIVs $\beta_0 + \beta_1 z$ of order $((0, 2), \{\})$ is of dimension smaller or equal to $2s - d_1^1 - d_2^1$, due to the orthogonality constraint on $[\beta'_0, \beta'_1]'$ given in Example 3.

Consider the matrices β, β_1 and β_2 as defined in Section 2. From a state space realization $(\mathcal{A}, \mathcal{B}, \mathcal{C})$ in canonical form corresponding to a VAR process it immediately follows that the columns of β_2 span the same space as the columns of the sub-block $C_{1,1}^E$. The same relation holds true for β_1 and the sub-block $C_{1,2}^E$. With respect to polynomial cointegration, Bauer and Wagner (2012) show that the rank of $C_{1,2}^G$ determines the number of minimum degree polynomial cointegrating relations, as discussed in Example 3. If $C_{1,2}^G = 0$, then there exists no vector γ , such that $\{\gamma' y_t\}_{t \in \mathbb{Z}}$ is integrated and cointegrated with $\{\beta'_2 \Delta_0 y_t\}_{t \in \mathbb{Z}}$. In this case $\{\beta'_1 y_t\}_{t \in \mathbb{Z}}$ is a stationary process.

The deterministic components included in the I(2) setting are typically a constant and a linear trend. As in the MFI(1) case, identifiability problems occur, if we consider a non-zero initial state $x_{1,\mu}$: The solution to the state space equations for $t > 0$ and $x_{1,\mu} \neq 0$ is given by:

$$y_t = \sum_{j=1}^{t-1} C \mathcal{A}^{j-1} \mathcal{B} \varepsilon_{t-j} + C_{1,1}^E (x_{1,1}^E + x_{1,2}^G (t-1)) + C_{1,2}^G x_{1,2}^G + C_{1,2}^E x_{1,2}^E + C_{\bullet} \mathcal{A}_{\bullet}^{t-1} x_{1,\bullet} + \Phi d_t + \varepsilon_t.$$

Hence, if $\Phi d_t = \Phi_0 + \Phi_1 t$, the output equation contains the terms $C_{1,1}^E x_{1,1}^E + C_{1,2}^G x_{1,2}^G + C_{1,2}^E x_{1,2}^E - C_{1,1}^E x_{1,2}^G + \Phi_0$ and $(C_{1,1}^E x_{1,2}^G + \Phi_1) t$. Again, this implies non-identifiability, which is resolved by assuming $x_{1,\mu} = 0$, compare Remark 6.

5.2.1. Testing Hypotheses on the State Space Unit Root Structure

To simplify notation we use

$$\overline{M}(d_1^1, d_2^1) := \begin{cases} \overline{M}(((0, d_1^1, d_2^1)), n - d_1^1 - d_2^1) & \text{if } d_1^1 > 0, \\ \overline{M}(((0, d_2^1)), n - d_2^1) & \text{if } d_1^1 = 0, d_2^1 > 0, \\ \overline{M}_{\bullet, n} & \text{if } d_1^1 = d_2^1 = 0, \end{cases}$$

with $n \geq d_1^1 + d_2^1$. Here $\overline{M}(d_1^1, d_2^1)$ for $d_1^1 + d_2^1 > 0$ denotes the closure of the set of transfer functions of order n that possess a state space unit root structure of either $\Omega_S = ((0, d_1^1, d_2^1))$ or $\Omega_S = ((0, d_2^1))$ in case of $d_1^1 = 0$, while $\overline{M}(0, 0)$ denotes the closure of the set of all stable transfer functions of order n .

Considering the relations between the different sets of transfer functions given in Corollary 4 shows that the following relations hold (assuming $s \geq 4$; the columns are arranged to include transfer functions with the same dimension of \mathcal{A}_u):

$$\begin{array}{ccccccc} \overline{M}(0,0) & \supset & \overline{M}(0,1) & \supset & \overline{M}(1,0) & & \\ & & & & \cup & & \\ & & & & \overline{M}(0,2) & \supset & \overline{M}(1,1) & \supset & \overline{M}(2,0) \\ & & & & & & \cup & & \cup \\ & & & & & & \overline{M}(0,3) & \supset & \overline{M}(1,2) \\ & & & & & & & & \cup \\ & & & & & & & & \overline{M}(0,4) \end{array}$$

Please note that $\overline{M}(d_1^1, d_2^1)$ corresponds to $H_{s-d_2^1, d_2^1-d_1^1} = H_{r, r_1}$ in Johansen (1995). Therefore, the relationships between the subsets match the ones in Johansen (1995, Table 9.1) and the ones found by Jensen (2013). The latter type of inclusions appear for instance for $\overline{M}(0, 2)$, containing transfer functions corresponding to $I(1)$ processes, which is a subset of the set $\overline{M}(1, 0)$ of transfer functions corresponding to $I(2)$ processes.

The same remarks as in the MFI(1) case also apply in the I(2) case: When testing for $H_0 : \Omega_S = ((0, d_{1,0}^1, d_{2,0}^1))$, all attainable state space unit root structures $\mathcal{A}(((0, d_{1,0}^1, d_{2,0}^1)))$ have to be included in the null hypothesis.

5.2.2. Testing Hypotheses on CIVs and PCIVs

Johansen (2006) discusses several types of hypotheses on the cointegrating spaces of different orders. These deal with properties of β , joint properties of $[\beta, \beta_1]$ or the occurrence of non-trivial polynomial cointegrating relations. Boswijk and Paruolo (2017), moreover, discuss testing hypotheses on the loading matrices of common trends (corresponding in our setting to testing hypotheses on \mathcal{C}_1).

We commence with hypotheses of the form $H_0 : \beta = K\varphi$ and $H'_0 : \beta = [b, \varphi]$ just as in the MFI(1) case at unit root one, since hypotheses on β correspond to hypotheses on its orthocomplement spanned by $[C_{1,1}^E, C_{1,2}^E]$ in the VARMA framework:

Hypotheses of the form $H_0 : \beta = K\varphi, K \in \mathbb{R}^{s \times t}, \varphi \in \mathbb{R}^{t \times r}$ imply $\varphi' K' [C_{1,1}^E, C_{1,2}^E] = 0$. W.l.o.g. let $K \in O_{s,t}$ and $K_\perp \in O_{s,s-t}$. As in the parameterization under H_0 in the MFI(1) case at unit root one, compare (15), use the mapping

$$[C_{1,1}^{E,r}, C_{1,2}^{E,r}](\check{\theta}_L, \theta_R) := \left[K \cdot \check{R}_L(\check{\theta}_L)' \begin{bmatrix} I_{t-r} \\ 0_{r \times (t-r)} \end{bmatrix}, K_\perp \right] \cdot R_R(\theta_R),$$

to derive a parameterization of the set of feasible matrices $[C_{1,1}^E, C_{1,2}^E]$, i.e., a joint parameterization of both sets of matrices $C_{1,1}^E$ and $C_{1,2}^E$, where $[C_{1,1}^E, C_{1,2}^E] \in O_{s,s-r}$.

Hypotheses of the form $H'_0 : \beta = [b, \varphi], b \in \mathbb{R}^{s \times t}, \varphi \in \mathbb{R}^{s \times (r-t)}, 0 < t \leq r$ are equivalent to $b' [C_{1,1}^E, C_{1,2}^E] = 0$. Assume w.l.o.g. $b \in O_{s,t}$ and parameterize the set of feasible matrices $C_{1,1}^E$ using $O_{s,d_1^1}(b)$ as defined in (13) and the set of feasible matrices $C_{1,2}^E$ using $O_{s,d_2^1-d_1^1}([b, C_{1,1}^E])$. Alternatively, parameterize the set of feasible matrices jointly as elements $[C_{1,1}^E, C_{1,2}^E] \in O_{s,s-r}(b)$.

Applications using the VECM framework allow for testing hypotheses on $[\beta, \beta_1]$. In the VARMA framework, these correspond to hypotheses on the orthogonal complement of $[\beta, \beta_1]$, i.e., $C_{1,1}^E$. Implementation of different types of hypotheses on $[\beta, \beta_1]$ proceeds as for similar hypotheses on β in the MFI(1) case at unit root one, replacing $[C_{1,1}^E, C_{1,2}^E]$ by $C_{1,1}^E$.

The hypothesis of no minimum degree polynomial cointegrating relations implies the restriction $C_{1,2}^G = 0$, compare Example 3. Therefore, we can test all hypotheses considered in Johansen (2006) also in our more general setting.

5.2.3. Testing Hypotheses on the Adjustment Coefficients

Hypotheses on α and ζ as defined in (6) and (7) correspond to hypotheses on the spaces spanned by the rows of $B_{1,2,1}$ and $B_{1,2,2}$. For VAR processes integrated of order two, the row space of $B_{1,2,1}$ is equal to the orthogonal complement of the column space of $[\alpha, \alpha_\perp \zeta]$, while the row space of $B_{1,2} := [B'_{1,2,1}, B'_{1,2,2}]'$ is equal to the orthogonal complement of the column space of α . The restrictions corresponding to hypotheses on α and ζ can be implemented analogously to the restrictions corresponding to hypotheses on α_1 in Section 5.1.3, reversing the roles of the relevant sub-blocks in B_u and C_u accordingly.

5.2.4. Restrictions on the Deterministic Components

The I(2) case is, with respect to the modeling of deterministic components, less well studied than the MFI(1) case. In most theory papers they are simply left out, with the notable exception Rahbek et al. (1999), dealing with the inclusion of a constant term in the I(2)-VECM representation. The main reason for this appears to be the way deterministic components in the defining vector error correction representation translate into deterministic components in the corresponding solution process. An unrestricted constant in the VECM for I(2) processes leads to a linear trend in $\{\beta'_1 y_t\}_{t \in \mathbb{Z}}$ and a quadratic trend in $\{\beta'_2 y_t\}_{t \in \mathbb{Z}}$, while an unrestricted linear trend results in quadratic and cubic trends in the respective directions. Already in the I(1) case discussed above five different cases—with respect to integration and asymptotic behavior of estimators and tests—need to be considered separately. An all encompassing discussion of the restrictions on the coefficients of a constant and a linear trend in the I(2) case requires the specification of even more cases. As an alternative approach in the VECM framework, deterministic components could be dealt with by replacing y_t with $y_t - \Phi d_t$ in the VECM equation. This has recently been considered in Johansen and Nielsen (2018) and is analogous to our approach in the state space framework.

As before, in the MFI(1) or I(1) case, the analysis of (the impact of) deterministic components is straightforward in the state space framework, which effectively stems from their additive inclusion in the Granger-type representation, compare (9). Choose, e.g., $\Phi d_t = \Phi_0 + \Phi_1 t$, as in the I(1) case. In analogy to Section 5.1.4, linear restrictions of deterministic components in relation to the static and polynomial cointegrating spaces can be embedded in a parameterization. Focusing on Φ_0 , e.g., this is achieved by

$$\Phi_0 = [C_{1,1}^E, C_{1,2}^E] \phi_0 + \check{C}_{1,2} \check{\phi}_0 + C_\perp \check{\phi}_0,$$

where the columns of $\check{C}_{1,2}$ are a basis for the column space of $C_{1,2}^G$, which does not necessarily have full column rank, and the columns of C_\perp span the orthocomplement of the column space of $[C_{1,1}^E, C_{1,2}^E, \check{C}_{1,2}]$. The matrix Φ_1 can be decomposed analogously. The corresponding parametrization then allows to consider different restricted versions of deterministic components and to study the asymptotic behavior of estimators and tests for these cases.

6. Summary and Conclusions

Vector autoregressive moving average (VARMA) processes, which can be cast equivalently in the state space framework, may be useful for empirical analysis compared to the more restrictive class of vector autoregressive (VAR) processes for a variety of reasons. These include invariance with respect to marginalization and aggregation, parsimony as well as the fact that the log-linearized solutions to DSGE models are typically VARMA processes rather than VAR processes. To realize the potential of these advantages necessitates, in our view, to develop cointegration analysis for VARMA processes to a

similar extent as it is developed for VAR processes. The necessary first steps of this research agenda are to develop a set of structure theoretical results that allow subsequently developing statistical inference procedures. [Bauer and Wagner \(2012\)](#) provides the very first step of this agenda by providing a *canonical form* for unit root processes in the state space framework, which is shown in that paper to be very convenient for cointegration analysis.

Based on the earlier canonical form paper this paper derives a state space model *parameterization* for VARMA processes with unit roots using the state space framework. The canonical form and a fortiori the parameterization based on it are constructed to facilitate the investigation of the unit root and (static and polynomial) cointegration properties of the considered process. Furthermore, the paper shows that the framework allows to test a large variety of hypotheses on cointegrating ranks and spaces, clearly a key aspect for the usefulness of any method to analyze cointegration. In addition to providing general results, throughout the paper all results are discussed in detail for the multiple frequency I(1) and I(2) cases, which cover the vast majority of applications.

Given the fact that (as shown in [Hazewinkel and Kalman 1976](#)) VARMA unit root processes cannot be continuously parameterized, the set of all unit root processes (as defined in this paper) is partitioned according to a multi-index Γ that includes the state space unit root structure. The parameterization is shown to be a diffeomorphism on the interior of the considered sets. The topological relationships between the sets forming the partitioning of all transfer functions considered are studied in great detail for three reasons: First, pseudo maximum likelihood estimation effectively amounts to maximizing the pseudo likelihood function over the closures of sets of transfer functions, \overline{M}_Γ in our notation. Second, related to the first item, the relations between subsets of M_Γ have to be understood in detail as knowledge concerning these relations is required for developing (sequential) pseudo likelihood-ratio tests for the numbers of stochastic trends or cycles. Third, of particular importance for the implementation of, e.g., pseudo maximum likelihood estimators, we discuss the existence of *generic pieces*.

In this respect we derive two results: First, for correctly specified state space unit root structure and system order of the stable subsystem—and thus correctly specified system order—we explicitly describe generic indices $\Gamma_g(\Omega_S, n_\bullet)$ such that $M_{\Gamma_g(\Omega_S, n_\bullet)}$ is open and dense in the set of all transfer functions with state space unit root structure Ω_S and system order of the stable subsystem n_\bullet . This result forms the basis for establishing consistent estimators of the transfer functions—and via continuity of the parameterization—of the parameter estimators when the state space unit root structure and system order are known. Second, in case only an upper bound on the system order is known (or specified), we show the existence of a generic multi-index $\Gamma_{\alpha_\bullet, g(n)}$ for which the set of corresponding transfer functions $M_{\Gamma_{\alpha_\bullet, g(n)}}$ is open and dense in the set \overline{M}_n of all non-explosive transfer functions whose order (or McMillan degree) is bounded by n . This result is the basis for consistent estimation (on an open and dense subset) when only an upper bound of the system order is known. In turn this estimator is the starting point for determining Ω_S , using the subset relationships alluded to above in the second point. For the MFI(1) and I(2) cases we show in detail that similar subset relations (concerning cointegrating ranks) as in the cointegrated VAR MFI(1) and I(2) cases hold, which suggests constructing similar sequential test procedures for determining the cointegrating ranks as in the VAR cointegration literature.

Section 5 is devoted to a detailed discussion of testing hypotheses on the cointegrating spaces, again for both the MFI(1) and the I(2) case. In this section, particular emphasis is put on modeling deterministic components. The discussion details how all usually formulated and tested hypotheses concerning (static and polynomial) cointegrating vectors, potentially in combination with (un-)restricted deterministic components, in the VAR framework can also be investigated in the state space framework.

Altogether, the paper sets the stage to develop pseudo maximum likelihood estimators, investigate their asymptotic properties (consistency and limiting distributions) and tests based on them for determining cointegrating ranks that allow performing cointegration analysis for cointegrated

VARMA processes. The detailed discussion of the MFI(1) and I(2) cases benefits the development of statistical theory dealing with these cases undertaken in a series of companion papers.

Author Contributions: The authors of the paper have contributed equally, via joint efforts, regarding both ideas, research, and writing. Conceptualization, all authors; methodology, all authors; formal analysis, P.d.M.R. and L.M.; investigation, all authors; writing—original draft preparation, P.d.M.R. and L.M.; writing—review and editing, all authors.; project administration, D.B. and M.W.; funding acquisition, D.B. and M.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation-Projekt Nummer 276051388) which is gratefully acknowledged. We acknowledge support for the publication costs by the Deutsche Forschungsgemeinschaft and the Open Access Publication Fund of Bielefeld University.

Acknowledgments: We thank the editors, Rocco Mosconi and Paolo Paruolo, as well as anonymous referees for helpful suggestions. The views expressed in this paper are solely those of the authors and not necessarily those of the Bank of Slovenia or the European System of Central Banks. On top of this the usual disclaimer applies.

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

Appendix A. Proofs of the Results of Section 3

Appendix A.1. Proof of Lemma 1

(i) Let C_j be a sequence in $O_{s,d}$ converging to C_0 for $j \rightarrow \infty$. By continuity of matrix multiplication

$$C_0' C_0 = \left(\lim_{j \rightarrow \infty} C_j \right)' \lim_{j \rightarrow \infty} C_j = \lim_{j \rightarrow \infty} (C_j' C_j) = I_d.$$

Thus, $C_0 \in O_{s,d}$, which shows that $O_{s,d}$ is closed. By construction $[C'C]_{i,i} = \sum_{j=1}^s c_{j,i}^2$. Since $[C'C]_{i,i} = 1$ for all $C \in O_{s,d}$ and $i = 1, \dots, d$, the entries of C are bounded.

(ii) By definition $C_O(\theta)$ is a product of matrices whose elements are either constant or infinitely often differentiable functions of the elements of θ .

(iii) The algorithm discussed above Lemma 1 maps every $C \in O_{s,d}$ to $[I_d, O'_{s-d \times d}]'$. Since $R_{q,i,j}(\theta)^{-1} = R_{q,i,j}(\theta)'$ for all q, i, j and θ , C can be obtained by multiplying $[I_d, O'_{s-d \times d}]'$ with the transposed Givens rotations.

(iv) As discussed, $C_O^{-1}(\cdot)$ is obtained from a repeated application of the algorithm described in Remark 10. In each step two entries are transformed to polar coordinates. According to Amann and Escher (2008, chp. 8, p. 204) the transformation to polar coordinates is infinitely often differentiable with infinitely often differentiable inverse for $\theta > 0$ (and hence $r > 0$), i.e., on the interior of the interval $[0, \pi)$. Thus, C_O^{-1} is a concatenation of functions which are infinitely often differentiable on the interior of $\Theta_O^{\mathbb{R}}$ and is thus infinitely often differentiable, if $\theta_j > 0$ for all components of θ .

Clearly, the interior of $\Theta_O^{\mathbb{R}}$ is open and dense in $\Theta_O^{\mathbb{R}}$. By the definition of continuity the pre-image of the interior of $\Theta_O^{\mathbb{R}}$ is open in $O_{s,d}$. By (iii) there exists a θ_0 for arbitrary $C_0 \in O_{s,d}$ such that $C_O(\theta_0) = C_0$. Since the interior of $\Theta_O^{\mathbb{R}}$ is dense in $\Theta_O^{\mathbb{R}}$ there exists a sequence θ_j in the interior of $\Theta_O^{\mathbb{R}}$ such that $\theta_j \rightarrow \theta_0$. Then $C_O(\theta_j) \rightarrow C_0$ because of the continuity of C_O . Since $C_O(\theta_j)$ is a sequence in the pre-image of the interior of $\Theta_O^{\mathbb{R}}$, it follows that the pre-image of the interior of $\Theta_O^{\mathbb{R}}$ is dense in $O_{s,d}$.

(v) For any $C \in O_{s,s}$ it holds that $1 = \det(C'C) = \det(C)^2$ and $\det(C) \in \mathbb{R}$, which implies $\det(C) \in \{-1, 1\}$. Since the determinant is a continuous function on square matrices, both sets $O_{s,s}^+$ and $O_{s,s}^-$ are disjoint and closed.

(vi) The proof proceeds analogously to the proof of (iii).

(vii) A function defined on two disjoint subsets is infinitely often differentiable if and only if the two functions restricted to the subsets are infinitely often differentiable. The same arguments as used in (iv) together with the results in (ii) imply that $C_O^{-1} : O_{s,s}^+ \rightarrow \Theta_O^{\mathbb{R}}$ and $C_O^{\pm}(\cdot)|_{O_{s,s}^+}$ are infinitely often differentiable with infinitely often differentiable inverse on an open subset of

$O_{s,s}^+$. Clearly, the multiplication with I_s^- is infinitely often differentiable with infinitely often differentiable inverse, which implies that $C_0^\pm(\cdot)|_{O_{s,s}^-}$ is infinitely often differentiable with infinitely often differentiable inverse on an open subset of $O_{s,s}^-$, from which the result follows.

Appendix A.2. Proof of Lemma 2

- (i) Let C_j be a sequence in $U_{s,d}$ converging to C_0 for $j \rightarrow \infty$. By continuity of matrix multiplication

$$C_0' C_0 = (\lim_{j \rightarrow \infty} C_j)' \lim_{j \rightarrow \infty} C_j = \lim_{j \rightarrow \infty} (C_j' C_j) = I_d.$$

Thus, $C_0 \in U_{s,d}$, which shows that $U_{s,d}$ is closed. By construction $[C' C]_{i,i} = \sum_{j=1}^s |c_{j,i}|^2$. Since $[C' C]_{i,i} = 1$ for all $C \in U_{s,d}$ and $i = 1, \dots, d$, the entries of C are bounded.

- (ii) By definition $C_U(\varphi)$ is a product of matrices whose elements are either constant or infinitely often differentiable functions of the elements of φ .
- (iii) The algorithm discussed above Lemma 2 maps every $C \in U_{s,d}$ to $[D_d(\varphi_D), 0'_{s-d \times d}]'$ with $D_d(\varphi_D) = \text{diag}(e^{i\varphi_{D,1}}, \dots, e^{i\varphi_{D,d}})$. Since $Q_{q,i,j}(\varphi)^{-1} = Q_{q,i,j}(\varphi)'$ for all q, i, j and φ , C can be obtained by multiplying $[D_d(\varphi_D), 0'_{s-d \times d}]'$ with the transposed Givens rotations.
- (iv) The algorithms in Remark 12 and above Lemma 2 describe C_U^{-1} in detail. The determination of an element of φ_L or φ_R uses the transformation of two complex numbers into polar coordinates in step 2 of Remark 12, which according to Amann and Escher (2008, chp. 8, p. 204) is infinitely often differentiable with infinitely often differentiable inverse except for non-negative reals, which are the complement of an open and dense subset of the complex plane. Step 3 of Remark 12 uses the formulas $\varphi_1 = \tan^{-1} \left(\frac{b}{a} \right)$, which is infinitely often differentiable for $a > 0$, and $\varphi_2 = \varphi_a - \varphi_b \text{ mod } 2\pi$, which is infinitely often differentiable for $\varphi_a \neq \varphi_b$, which occurs on an open and dense subset of $[0, 2\pi) \times [0, 2\pi)$. For the determination of an element of φ_D a complex number of modulus one is transformed in polar coordinates which is infinitely often differentiable on an open and dense subset of complex numbers of modulus one compare again Amann and Escher (2008, chp. 8, p. 204). Thus, C_U^{-1} is a concatenation of functions which are infinitely often differentiable on open and dense subsets of their domain of definition and is thus infinitely often differentiable on an open and dense subset of $U_{s,d}$.

Appendix A.3. Proof of Theorem 2

- (i) The multi-index Γ is unique for a transfer function $k \in M_n$, since it only contains information encoded in the canonical form. Therefore, M_Γ is well defined. Since conversely for every transfer function $k \in M_n$ a multi-index Γ can be found, M_Γ constitutes a partitioning of M_n . Furthermore, using the canonical form, it is straightforward to see that the mapping attaching the triple $(\mathcal{A}, \mathcal{B}, \mathcal{C}) \in \Delta_\Gamma$ in canonical form to a transfer function $k \in M_\Gamma$ is homeomorphic (bijective, continuous, with continuous inverse): Bijectivity is a consequence of the definition of the canonical form. T_{pt} continuity of the transfer function as a function of the matrix triples is obvious from the definition of T_{pt} . Continuity of the inverse can be shown by constructing the canonical form starting with an overlapping echelon form (which is continuous according to Hannan and Deistler 1988, chp. 2) and subsequently transforming the state basis to reach the canonical form. This involves the calculation of a Jordan normal form with fixed structure. This is an analytic mapping (cf. Chatelin 1993, Theorem 4.4.3). Finally, the restrictions on C and B are imposed. For given multi-index Γ these transformations are continuous (as discussed above they involve QR decompositions to obtain unitary block columns for the blocks of C , rotations to p.u.t form with fixed structure for the blocks of B and transformations to echelon canonical form for the stable part).
- (ii) The construction of the triple $(\mathcal{A}(\theta), \mathcal{B}(\theta), \mathcal{C}(\theta))$ for given θ and Γ is straightforward: \mathcal{A}_u is uniquely determined by Γ . Since $\theta_{B,p}$ contains the entries of \mathcal{B}_u restricted to be positive and $\theta_{B,f}$

contains the free parameters of \mathcal{B}_u , the mapping $\theta_{B,p} \times \theta_{B,f} \rightarrow \mathcal{B}_u$ is continuous. The mapping $\theta_{\bullet} \rightarrow (\mathcal{A}_{\bullet}, \mathcal{B}_{\bullet}, \mathcal{C}_{\bullet})$ is continuous (cf. Hannan and Deistler 1988, Theorem 2.5.3 (ii)). The mapping $\theta_{C,E} \times \theta_{C,G} \rightarrow \mathcal{C}_u$ consists of iterated applications of C_O , and C_U (compare Lemmas 1 and 2) which are differentiable and thus continuous and iterated applications of the extensions of the mappings C_{O,d_2-d_1} and $C_{O,G}$ (compare Corollaries 1 and 2) to general unit root structures and to complex matrices. The proof that these functions are differentiable is analogous to the proofs of Lemma 1 and Lemma 2.

- (iii) The definitions of $\theta_{B,f}$ and $\theta_{B,p}$ immediately imply that they depend continuously on \mathcal{B}_u . The parameter vector θ_{\bullet} depends continuously on $(\mathcal{A}_{\bullet}, \mathcal{B}_{\bullet}, \mathcal{C}_{\bullet})$ (cf. Hannan and Deistler 1988, Theorem 2.5.3 (ii)). The existence of an open and dense subset of matrices \mathcal{C}_u such that the mapping attaching parameters to the matrices is continuous follows from arguments contained in the proofs of Lemmas 1 and 2.

Appendix B. Proofs of the Results of Section 4

Appendix B.1. Proof of Theorem 3

For the first inclusion the proof can be divided into two parts, discussing the stable and the unstable subsystem separately. The result with regard to the stable subsystem is due to Hannan and Deistler (1988, Theorem 2.5.3 (iv)). For the unstable subsystem $(\tilde{\Omega}_S, \tilde{p}) \leq (\Omega_S, p)$ implies the existence of a matrix S as described in Definition 9. Partition $S = \begin{bmatrix} S_1 \\ S_2 \end{bmatrix}$ such that $S_1 p = p_1 \geq \tilde{p}$. Let \tilde{k} be an arbitrary transfer function in $M_{\tilde{\Gamma}} = \pi(\Delta_{\tilde{\Gamma}})$ with corresponding state space realization $(\tilde{\mathcal{A}}, \tilde{\mathcal{B}}, \tilde{\mathcal{C}}) \in \Delta_{\tilde{\Gamma}}$. Then, we find matrices B_1 and C_1 such that for the state space realization given by $\mathcal{A} = S \begin{bmatrix} \tilde{\mathcal{A}} & \tilde{J}_{12} \\ 0 & \tilde{J}_2 \end{bmatrix} S'$, $\mathcal{B} = S \begin{bmatrix} \tilde{\mathcal{B}} \\ B_1 \end{bmatrix}$ and $\mathcal{C} = [\tilde{\mathcal{C}} \ C_1] S'$ it holds that $(\mathcal{A}, \mathcal{B}, \mathcal{C}) \in \Delta_{\tilde{\Gamma}}$. Then, $(\mathcal{A}_j, \mathcal{B}_j, \mathcal{C}_j) = (\mathcal{A}, S \text{diag}(I_{n_1}, j^{-1}I_{n_2})S' \mathcal{B}, \mathcal{C}) \in \Delta_{\tilde{\Gamma}}$, where n_i is the number of rows of S_i for $i = 1, 2$ converges for $j \rightarrow \infty$ to $(\mathcal{A}, S \begin{bmatrix} \tilde{\mathcal{B}} \\ 0 \end{bmatrix}, \mathcal{C}) \in \overline{\Delta_{\tilde{\Gamma}}}$, which is observationally equivalent to $(\tilde{\mathcal{A}}, \tilde{\mathcal{B}}, \tilde{\mathcal{C}})$. Consequently, $\tilde{k} = \pi \left(\mathcal{A}, S \begin{bmatrix} \tilde{\mathcal{B}} \\ 0 \end{bmatrix}, \mathcal{C} \right) \in \pi(\overline{\Delta_{\tilde{\Gamma}}})$.

To show the second inclusion, consider a sequence of systems $(\mathcal{A}_j, \mathcal{B}_j, \mathcal{C}_j) \in \Delta_{\tilde{\Gamma}}$, $j \in \mathbb{N}$ converging to $(A_0, B_0, C_0) \in \overline{\Delta_{\tilde{\Gamma}}}$. We need to show $\tilde{\Gamma} \in \bigcup_{\tilde{\Gamma} \in \mathcal{K}(\Gamma)} \{\tilde{\Gamma} \leq \tilde{\Gamma}\}$, where $\tilde{\Gamma}$ is the multi-index corresponding to (A_0, B_0, C_0) .

For the stable system we can separate the subsystem $(\mathcal{A}_{j,s}, \mathcal{B}_{j,s}, \mathcal{C}_{j,s})$ remaining stable in the limit and the part with eigenvalues of A_j tending to the unit circle. As discussed in Section 4.1.2, $(\mathcal{A}_{j,s}, \mathcal{B}_{j,s}, \mathcal{C}_{j,s})$ converges to the stable subsystem $(A_{0,\bullet}, B_{0,\bullet}, C_{0,\bullet})$ whose Kronecker indices can only be smaller than or equal to α_{\bullet} (cf. Hannan and Deistler 1988, Theorem 2.5.3).

The remaining subsystem consists of the unstable subsystem of $(\mathcal{A}_j, \mathcal{B}_j, \mathcal{C}_j)$ which converges to $(A_{0,u}, B_{0,u}, C_{0,u})$ and the second part of the stable subsystem containing all stable eigenvalues of A_j converging to the unit circle. The limiting combined subsystem $(A_{0,c}, B_{0,c}, C_{0,c})$ is such that $A_{0,c}$ is block diagonal. If the limiting combined subsystem is minimal and $B_{0,u}$ has a structure corresponding to p , this shows that the pair $(\tilde{\Omega}_S, \tilde{p})$ extends (Ω_S, p) in accordance with the definition of $\mathcal{K}(\Gamma)$.

Since the limiting subsystem is not necessarily minimal and $B_{0,u}$ has not necessarily a structure corresponding to p , eliminating coordinates of the state and adapting the corresponding structure indices p may result in a pair $(\tilde{\Omega}_S, \tilde{p})$ that is smaller than the pair $(\tilde{\Omega}_S, \tilde{p})$ corresponding to an element of $\mathcal{K}(\Gamma)$.

Appendix B.2. Proof of Theorem 4

The multi-index Γ contains three components: $\Omega_S, p, \alpha_{\bullet}$. For given Ω_S the selection of the structures indices p_{\max} introducing the fewest restrictions, such that in its boundary all possible p.u.t. matrices occur, was discussed in Section 4.2. Choosing this maximal element p_{\max} then implies that

all systems of given state space unit root structure correspond to a multi-index that is smaller than or equal to $(\Omega_S, p_{\max}, \beta_\bullet)$, where β_\bullet is a Kronecker index corresponding to state space dimension n_\bullet . For the Kronecker indices of order n_\bullet it is known that there exists one index $\alpha_{\bullet,g}$ such that $M_{\alpha_{\bullet,g}}$ is open and dense in $\overline{M_{n_\bullet}}$. The set $M_{\Omega_S, p_{\max}, \beta_\bullet}$ is, therefore, contained in $\overline{M_{\Omega_S, p_{\max}, \alpha_{\bullet,g}}}$ which implies (14) with $\Gamma_g(\Omega_S, n_\bullet) := (\Omega_S, p_{\max}, \alpha_{\bullet,g})$.

For the second claim choose an arbitrary state space realization $(\mathcal{A}, \mathcal{B}, \mathcal{C})$ in canonical form such that $\pi(\mathcal{A}, \mathcal{B}, \mathcal{C}) \in M(\Omega_S, n_\bullet)$ for arbitrary Ω_S . Define the sequence $(A_j, B_j, C_j)_{j \in \mathbb{N}}$ by $A_j = (1 - j^{-1})\mathcal{A}$, $B_j = (1 - j^{-1})\mathcal{B}$, $C_j = \mathcal{C}$. Then $\lambda_{|\max|}(A_j) < 1$ holds for all j , which implies $\pi(A_j, B_j, C_j) \in \overline{M_{\Gamma_{\alpha_{\bullet,g}(n)}}$ for every $n \geq n_u(\Omega_S) + n_\bullet$ and every j . The continuity of π implies $\pi(\mathcal{A}, \mathcal{B}, \mathcal{C}) = \lim_{j \rightarrow \infty} \pi(A_j, B_j, C_j) \in \overline{M_{\Gamma_{\alpha_{\bullet,g}(n)}}$.

Appendix B.3. Proof of Theorem 5

(i) Assume that there exists a sequence $k_i \in \overline{M_\Gamma}$ converging to a transfer function $k_0 \in M_\Gamma$. For such a sequence the size of the Jordan blocks for every unit root are identical from some i_0 onwards since eigenvalues depend continuously on the matrices (cf. Chatelin 1993): Thus, the stable part of the transfer functions k_i must converge to the stable part of the transfer function k_0 , since the sum of the algebraic multiplicity of all eigenvalues inside the open unit disc cannot drop in the limit. Since V_α (the set of all stable transfer functions with Kronecker index α) is open in $\overline{V_\alpha}$ according to Hannan and Deistler (1988, Theorem 2.5.3) this implies that the stable part of k_i has Kronecker index α_\bullet from some i_0 onwards.

For the unstable part of the transfer function note that in M_Γ for every unit root z_j the rank of $(A - z_j I_n)^r$ is equal for every r . Thus, the maximum over $\overline{M_\Gamma}$ cannot be larger due to lower semi-continuity of the rank. It follows that for $k_i \rightarrow k_0$ the ranks of $(A - z_j I_n)^r$ for all $|z_j| = 1$ and for all $r \in \mathbb{N}_0$ are identical to the ranks corresponding to k_0 from some point onwards showing that k_i has the same state space unit root structure as k_0 from some i_0 onwards. Finally, the p.u.t. structure of sub-blocks of \mathcal{B}_k clearly introduces an open set being defined via strict inequalities. This shows that $k_i \in M_\Gamma$ from some i_0 onwards implying that M_Γ is open in $\overline{M_\Gamma}$.

(ii) The first inclusion was shown in Theorem 3. Comparing Definitions 10 and 11 we see $\bigcup_{\tilde{\Gamma} \in \mathcal{K}(\Gamma_g)} M_{\tilde{\Gamma}} \subset \bigcup_{(\check{\Omega}_S, \check{n}_\bullet) \in \mathcal{A}(\Omega_S, n_\bullet)} M(\check{\Omega}_S, \check{n}_\bullet)$. By the definition of the partial ordering (compare Definition 9) $\bigcup_{\tilde{\Gamma} \leq \Gamma_g} M_{\tilde{\Gamma}} \subset \bigcup_{(\check{\Omega}_S, \check{n}_\bullet) \leq (\Omega_S, n_\bullet)} M(\check{\Omega}_S, \check{n}_\bullet)$ holds. Together these two statements imply the second inclusion.

$\bigcup_{(\check{\Omega}_S, \check{n}_\bullet) \in \mathcal{A}(\Omega_S, n_\bullet)} \bigcup_{(\check{\Omega}_S, \check{n}_\bullet) \leq (\hat{\Omega}_S, \hat{n}_\bullet)} M(\check{\Omega}_S, \check{n}_\bullet) \subset \overline{M_{\Gamma_g(\Omega_S, n_\bullet)}}$ is a consequence of the following two statements:

- (a) If $M(\check{\Omega}_S, \check{n}_\bullet) \subset \overline{M(\Omega_S, n_\bullet)}$, then $\bigcup_{(\check{\Omega}_S, \check{n}_\bullet) \leq (\hat{\Omega}_S, \hat{n}_\bullet)} M(\check{\Omega}_S, \check{n}_\bullet) \subset \overline{M(\Omega_S, n_\bullet)}$.
- (b) If $(\check{\Omega}_S, \check{n}_\bullet) \in \mathcal{A}(\Omega_S, n_\bullet)$, then $M(\check{\Omega}_S, \check{n}_\bullet) \subset \overline{M(\Omega_S, n_\bullet)}$.

For (a) note that for an arbitrary transfer function $\check{k} \in M(\check{\Omega}_S, \check{n}_\bullet)$ with $(\check{\Omega}_S, \check{n}_\bullet) \leq (\hat{\Omega}_S, \hat{n}_\bullet)$ there is a multi-index $\check{\Gamma}$ such that $\check{k} \in M_{\check{\Gamma}}$. By the definition of the partial ordering (compare Definition 9) we find a multi-index $\tilde{\Gamma} \geq \check{\Gamma}$ such that $M_{\tilde{\Gamma}} \subset M(\check{\Omega}_S, \check{n}_\bullet)$. By Theorem 3 and the continuity of π we have $M_{\tilde{\Gamma}} \subset \pi(\overline{\Delta_{\tilde{\Gamma}}}) \subset \overline{M_{\tilde{\Gamma}}}$. Since $M(\check{\Omega}_S, \check{n}_\bullet) \subset \overline{M(\Omega_S, n_\bullet)}$ by assumption, $\check{k} \in \overline{M_{\tilde{\Gamma}}} \subset \overline{M(\check{\Omega}_S, \check{n}_\bullet)} \subset \overline{M(\Omega_S, n_\bullet)}$ which finishes the proof of (a).

With respect to (b) note that by Definition 11, $\mathcal{A}(\Omega_S, n_\bullet)$ contains transfer functions with two types of state space unit root structures. First, $\tilde{\mathcal{A}}_u$ corresponding to state space unit root $\hat{\Omega}_S$ may be of the form

$$S \tilde{\mathcal{A}}_u S' = \begin{bmatrix} \mathcal{A}_u & J_{12} \\ 0 & J_2 \end{bmatrix}. \tag{A1}$$

Second, $\check{\mathcal{A}}_u$ corresponding to state space unit root $\check{\Omega}_S$ may be of the form (A1) where off-diagonal elements of \mathcal{A}_u are replaced by zero. To prove (b) we need to show that for both cases the corresponding transfer function is contained in $\overline{M(\Omega_S, n_\bullet)}$.

We start by showing that in the second case the transfer function \check{k} is contained in $\overline{M(\check{\Omega}_S, \check{n}_\bullet)}$, where $\check{\Omega}_S$ is the state space unit root structure corresponding to $\check{\mathcal{A}}_u$ in (A1). For this, consider the sequence

$$A_j = \begin{bmatrix} 1 & j^{-1} \\ 0 & 1 \end{bmatrix}, \quad B_j = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad C_j = \begin{bmatrix} C_1 & C_2 \end{bmatrix}.$$

Clearly, every system (A_j, B_j, C_j) corresponds to an $I(2)$ process, while the limit for $j \rightarrow \infty$ corresponds to an $I(1)$ process. This shows that it is possible in the limit to trade one $I(2)$ component with two $I(1)$ components leading to more transfer functions in the T_{pt} closure of $M_{\Gamma_g(\Omega_S, n_\bullet)}$ than only the ones included in $\pi(\overline{\Delta_{\Gamma_g(\Omega_S, n_\bullet)}})$, where the off-diagonal entry in A_j is restricted to equal one and hence the corresponding sequence of systems in the canonical form diverges to infinity. In a sense these systems correspond to “points at infinity”: For the example given above we obtain the canonical form

$$\mathcal{A}_j = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \mathcal{B}_j = \begin{bmatrix} B_1 \\ B_2/j \end{bmatrix}, \quad \mathcal{C}_j = \begin{bmatrix} C_1 & jC_2 \end{bmatrix}.$$

Thus, the corresponding parameter vector for the entries in $\mathcal{B}_{j,2}$ converges to zero and the ones corresponding to $\mathcal{C}_{j,2}$ to infinity.

Generalizing this argument shows that every transfer function corresponding to a pair $(\check{\Omega}_S, \check{n}_\bullet)$ in $\mathcal{A}(\check{\Omega}_S, \check{n}_\bullet)$, where $\check{\mathcal{A}}_u$ can be obtained by replacing off-diagonal entries of \mathcal{A}_u with zero, can be reached from within $M(\check{\Omega}_S, \check{n}_\bullet)$.

To prove $\check{k} \in \overline{M(\Omega_S, n_\bullet)}$ in the first case, where the state space unit root structure is extended as visible in Equation (A1), consider the sequence:

$$\tilde{A}_j = \begin{bmatrix} 1 & 1 \\ 0 & 1 - j^{-1} \end{bmatrix}, \quad \tilde{B}_j = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad \tilde{C}_j = \begin{bmatrix} C_1 & C_2 \end{bmatrix},$$

corresponding to the following system in canonical form (except that the stable subsystem is not necessarily in echelon canonical form)

$$\tilde{\mathcal{A}}_j = \begin{bmatrix} 1 & 0 \\ 0 & 1 - j^{-1} \end{bmatrix}, \quad \tilde{\mathcal{B}}_j = \begin{bmatrix} B_1 + jB_2 \\ -jB_2 \end{bmatrix}, \quad \tilde{\mathcal{C}}_j = \begin{bmatrix} C_1 & C_1 - C_2/j \end{bmatrix}.$$

This sequence shows that there exists a sequence of transfer functions corresponding to $I(1)$ processes with one common trend that converge to a transfer function corresponding to an $I(2)$ system. Again, in the canonical form this cannot happen as there the $(1, 2)$ entry of \tilde{A}_j would be restricted to be equal to zero. At the same time note that the dimension of the stable system is reduced due to one component of the state changing from the stable to the unit root part.

Now for a unit root structure $\check{\Omega}_S$ such that $(\check{\Omega}_S, \check{n}_\bullet) \in \mathcal{A}(\Omega_S, n_\bullet)$, satisfying

$$S \tilde{\mathcal{A}}_u S' = \begin{bmatrix} \mathcal{A}_u & J_{12} \\ 0 & J_2 \end{bmatrix},$$

the Jordan blocks corresponding to Ω_S are sub-blocks of the ones corresponding to $\tilde{\Omega}_S$, potentially involving a reordering of coordinates using the permutation matrix S . Taking as the approximating sequence of transfer functions $\tilde{k}_j \in M_{\Gamma_g(\Omega_S, n_\bullet)} \rightarrow k_0 \in M_{\Gamma_g(\tilde{\Omega}_S, \tilde{n}_\bullet)}$ that have the same structure $\tilde{\Omega}_S$ but replacing J_2 by $\frac{j-1}{j}J_2$ leads to processes with state space unit root structure Ω_S .

For the stable part of \tilde{k}_j we can separate the part containing poles tending to the unit circle (contained in J_2) and the remaining transfer function $\tilde{k}_{j,s}$, which has Kronecker indices $\tilde{\alpha} \leq \alpha_\bullet$. However, the results of Hannan and Deistler (1988, Theorem 2.5.3) then imply that the limit remains in $\overline{M_{\alpha_\bullet}}$ and hence allows for an approximating sequence in M_{α_\bullet} .

Both results combined constitute the whole set of attainable state space unit root structures in Definition 11 and prove (b).

As follows from Corollary 4, $\overline{M(\Omega_S, n_\bullet)} = \overline{M_{\Gamma_g(\Omega_S, n_\bullet)}}$. Thus, (b) implies $\bigcup_{(\tilde{\Omega}_S, \tilde{n}_\bullet) \in \mathcal{A}(\Omega_S, n_\bullet)} M(\tilde{\Omega}_S, \tilde{n}_\bullet) \subset \overline{M_{\Gamma_g(\Omega_S, n_\bullet)}}$ and (a) adds the second union showing the subset inclusion.

It remains to show equality for the last set inclusion. Thus, we need to show that for $k_j \in M_{\Gamma_g(\Omega_S, n_\bullet)}, k_j \rightarrow k_0$, it holds that $k_0 \in M(\tilde{\Omega}_S, \tilde{n}_\bullet)$, where $(\tilde{\Omega}_S, \tilde{n}_\bullet) \leq (\check{\Omega}_S, \check{n}_\bullet) \in \mathcal{A}(\Omega_S, n_\bullet)$. To this end note that the rank of a matrix is a lower semi-continuous function such that for a sequence of matrices E_j with limit E_0 , we have

$$\text{rank}(\lim_{j \rightarrow \infty} E_j) = \text{rank}(E_0) \leq \liminf_{j \rightarrow \infty} \text{rank}(E_j).$$

Then, consider a sequence $k_j(z) \in M_{\Gamma_g(\Omega_S, n_\bullet)}, j \in \mathbb{N}$. We can find a converging sequence of systems (A_j, B_j, C_j) realizing $k_j(z)$. Therefore, choosing $E_j = (A_j - z_k I_n)^r$ we obtain that

$$\text{rank}((A_0 - z_k I_n)^t) \leq n - \sum_{r=1}^t d_{j, h_k - r + 1}^k,$$

since $k_j(z) \in M_{\Gamma_g(\Omega_S, n_\bullet)}$ implies that the number $d_{j, h_k - r + 1}^k$ of the generalized eigenvalues at the unit roots is governed by the entries of the state space unit root structure Ω_S . This implies that $\sum_{r=1}^t d_{j, h_k - r + 1}^k \leq \sum_{r=1}^t d_{0, h_k - r + 1}^k$ for $t = 1, 2, \dots, n$. Consequently, the limit has at least as many chains of generalized eigenvalues of each maximal length as dictated by the state space unit root structure Ω_S for each unit root of the limiting system.

Rearranging the rows and columns of the Jordan normal form using a permutation matrix S it is then obvious that either the limiting matrix A_0 has additional eigenvalues, where thus

$$S A_0 S' = \begin{bmatrix} \mathcal{A}_j & \tilde{J}_{12} \\ 0 & \tilde{J}_2 \end{bmatrix}$$

must hold. Or upper diagonal entries in \mathcal{A}_j must be changed from ones to zeros in order to convert some of the chains to lower order. One example in this respect was given above:

For $A_j = \begin{bmatrix} 1 & 1/j \\ 0 & 1 \end{bmatrix}$ the rank of $(A_j - I_2)^r$ is equal to 1 for $r = 1$ and 0 for $r = 2$. For the limit we obtain $A_0 = I_2$ and hence the rank is zero for $r = 1, 2$. The corresponding indices are $d_{j,1}^1 = 1, d_{j,2}^1 = 1$ for the approximating sequence and $d_{0,1}^1 = 0, d_{0,2}^1 = 2$ for the limit respectively. Summing these indices starting from the last one, one obtains $d_{j,2}^1 = 1 \leq d_{0,2}^1 = 2$ and $d_{j,1}^1 + d_{j,2}^1 = 2 \leq d_{0,1}^1 + d_{0,2}^1 = 2$.

Hence the state space unit root structure corresponding to (A_0, B_0, C_0) must be attainable according to Definition 11. The number of stable state components must decrease accordingly.

Finally, the limiting system (A_0, B_0, C_0) is potentially not minimal. In this case the pair $(\tilde{\Omega}_S, \tilde{n}_\bullet)$ is reduced to a smaller one, concluding the proof.

References

- Amann, Herbert, and Joachim Escher. 2008. *Analysis III*. Basel: Birkhäuser Basel.
- Aoki, Massanao. 1990. *State Space Modeling of Time Series*. New York: Springer.
- Bauer, Dietmar, and Martin Wagner. 2003. On Polynomial Cointegration in the State Space Framework. *Mimeo*. [CrossRef]
- Bauer, Dietmar, and Martin Wagner. 2005. Autoregressive Approximations of Multiple Frequency I(1) Processes. IHS Economics Series, Institut für Höhere Studien–Institute for Advanced Studies (IHS) Vienna, No. 174. Available online: <http://hdl.handle.net/10419/72306> (accessed on 3 November 2020).
- Bauer, Dietmar, and Martin Wagner. 2012. A State Space Canonical Form for Unit Root Processes. *Econometric Theory* 28: 1313–49. [CrossRef]
- Boswijk, H. Peter, and Paolo Paruolo. 2017. Likelihood Ratio Tests of Restrictions on Common Trends Loading Matrices in I(2) VAR Systems. *Econometrics* 5: 28. [CrossRef]
- Campbell, John Y. 1994. Inspecting the Mechanism: An Analytical Approach to the Stochastic Growth Model. *Journal of Monetary Economics* 33: 463–506. [CrossRef]
- Chatelin, Françoise. 1993. *Eigenvalues of Matrices*. New York: John Wiley & Sons.
- Engle, Robert F., and Clive W.J. Granger. 1987. Cointegration and Error Correction: Representation, Estimation and Testing. *Econometrica* 55: 251–76. [CrossRef]
- Golub, Gene H., and Charles F. van Loan. 1996. *Matrix Computations*, 3rd ed. Baltimore: The Johns Hopkins University Press.
- Granger, Clive W.J. 1981. Some Properties of Time Series Data and Their Use in Econometric Model Specification. *Journal of Econometrics* 16: 121–30. [CrossRef]
- Hannan, Edward J., and Manfred Deistler. 1988. *The Statistical Theory of Linear Systems*. New York: John Wiley & Sons.
- Hazewinkel, Michiel, and Rudolf E. Kalman. 1976. Invariants, Canonical Forms and Moduli for Linear, Constant, Finite Dimensional, Dynamical Systems. In *Mathematical Systems Theory*. Edited by Giovanni Marchesini and Sanjoy Kumar Mitter. Berlin: Springer, chp. 4, pp. 48–60.
- Jensen, Andreas N. 2013. The Nesting Structure of the Cointegrated Vector Autoregressive Models. Paper presented at the QED Conference 2013, Vienna, Austria, May 3–4.
- Johansen, Søren. 1991. Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica* 59: 1551–80. [CrossRef]
- Johansen, Søren. 1995. *Likelihood-Based Inference in Cointegrated Vector Auto-Regressive Models*. Oxford: Oxford University Press.
- Johansen, Søren. 1997. Likelihood Analysis of the I(2) Model. *Scandinavian Journal of Statistics* 24: 433–62. [CrossRef]
- Johansen, Søren. 2006. Statistical Analysis of Hypotheses on the Cointegrating Relations in the I(2) Model. *Journal of Econometrics* 132: 81–115. [CrossRef]
- Johansen, Søren, and Morton Ø. Nielsen. 2018. The Cointegrated Vector Autoregressive Model with General Deterministic Terms. *Journal of Econometrics* 202: 214–29. [CrossRef]
- Johansen, Søren, and Ernst Schaumburg. 1999. Likelihood Analysis of Seasonal Cointegration. *Journal of Econometrics* 88: 301–39. [CrossRef]
- Juselius, Katarina. 2006. *The Cointegrated VAR Model: Methodology and Applications*. Oxford: Oxford University Press.
- Lewis, Richard, and Gregory C. Reinsel. 1985. Prediction of Multivariate Time Series by Autoregressive Model Fitting. *Journal of Multivariate Analysis* 16: 393–411. [CrossRef]
- Otto, Markus. 2011. *Rechenmethoden für Studierende der Physik im ersten Jahr*. Heidelberg: Spektrum Akademischer Verlag.
- Paruolo, Paolo. 1996. On the Determination of Integration Indices in I(2) Systems. *Journal of Econometrics* 72: 313–56. [CrossRef]
- Paruolo, Paolo. 2000. Asymptotic Efficiency of the Two Stages Estimator in I(2) Systems. *Econometric Theory* 16: 524–50. [CrossRef]

- Poskitt, Donald S. 2006. On the Identification and Estimation of Nonstationary and Cointegrated ARMAX Systems. *Econometric Theory* 22: 1138–75. [[CrossRef](#)]
- Rahbek, Anders, Hans C. Kongsted, and Clara Jorgensen. 1999. Trend Stationarity in the I(2) Cointegration Model. *Journal of Econometrics* 90: 265–89. [[CrossRef](#)]
- Saikkonen, Pentti. 1992. Estimation and Testing of Cointegrated Systems by an Autoregressive Approximation. *Econometric Theory* 8: 1–27. [[CrossRef](#)]
- Saikkonen, Pentti, and Ritva Luukkonen. 1997. Testing Cointegration in Infinite Order Vector Autoregressive Processes. *Journal of Econometrics* 81: 93–126. [[CrossRef](#)]
- Sims, Christopher A., James H. Stock, and Mark W. Watson. 1990. Inference in Linear Time Series Models with Some Unit Roots. *Econometrica* 58: 113–44. [[CrossRef](#)]
- Wagner, Martin. 2018. Estimation and Inference for Cointegrating Regressions. In *Oxford Research Encyclopedia of Economics and Finance*. Oxford: Oxford University Press.
- Wagner, Martin, and Jaroslava Hlouskova. 2009. The Performance of Panel Cointegration Methods: Results from a Large Scale Simulation Study. *Econometric Reviews* 29: 182–223. [[CrossRef](#)]
- Zellner, Arnold, and Franz C. Palm. 1974. Time Series Analysis and Simultaneous Equation Econometric Models. *Journal of Econometrics* 2: 17–54. [[CrossRef](#)]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).