Pseudo Maximum Likelihood Estimation of Cointegrated Multiple Frequency I(1) VARMA Processes Using the State Space Framework

Dissertation

zur Erlangung des Grades
eines Doktors der Naturwissenschaften

eingereicht an der Fakultät Statistik
der Technischen Universität Dortmund
vorgelegt von
Patrick de Matos Ribeiro

Dortmund, 29. September 2020

Gutachter:
Professor Dr. Martin Wagner, Alpen-Adria-Universität Klagenfurt
Professor Dr. Walter Krämer, Technische Universität Dortmund
Professor Dr. Dietmar Bauer, Universität Bielefeld

Inhaltsverzeichnis

Erklärung zu den Beiträgen der einzelnen Autoren zu den Artikeln **iii**

Introduction **1**

Acknowledgments **3**

A Parameterization of Models for Unit Root Processes: Structure Theory and Hypothesis Testing **1**
1.1 Introduction . . . 1
1.2 Vector Autoregressive, Vector Autoregressive Moving Average Processes and Parameterizations . . . 3
1.3 The Canonical Form and the Parameterization . . . 9
1.3.1 The Parameterization in the MFI(1) Case . . . 19
1.3.2 The Parameterization in the I(2) Case . . . 25
1.3.3 The Parameterization in the General Case . . . 28
1.4 The Topological Structure . . . 29
1.4.1 The Closure of Δ_Γ in Δ_n . . . 31
1.4.2 The Closure of M_Γ . . . 34
1.5 Testing Commonly Used Hypotheses in the MFI(1) and I(2) Cases . . . 35
1.5.1 The $MFI(1)$ Case . . . 36
1.5.2 The $I(2)$ Case . . . 42
1.6 Summary and Conclusions . . . 44

Pseudo Maximum Likelihood Parameter Estimation for Multiple Frequency I(1) Processes: A State Space Approach **47**
2.1 Introduction . . . 47
2.2 Setup and Assumptions . . . 49
2.3 Pseudo Maximum Likelihood Estimation . . . 54
2.4 Simulation Results . . . 59
2.5 Summary and Conclusion . . . 63

Stochastic Trends and Economic Fluctuations: Reconsidered From the State Space **64**
3.1 Introduction . . . 65
3.2 Methods . . . 66
3.2.1 The State Space Framework . . . 66
3.2.2 The Johansen Vector Error Correction Model . . . 69
3.2.3 The Approach of Stock and Watson . . . 69
3.3 Data Set . . . 70
3.4 Analysis of the Data Set of King et al. (1991) . . . 72
3.4.1 Choice of System Parameters . . . 72
3.4.2 Deterministic Components . . . 72
3.4.3 Determination of the Cointegrating Rank . . . 73

3.4.4 Estimation of the Cointegrating Space . . . 76
3.4.5 Testing Hypotheses on the Cointegrating Space . . . 77
3.5 Full Sample Analysis . . . 78
3.5.1 Choice of System Parameters . . . 78
3.5.2 Deterministic Components . . . 78
3.5.3 Determination of the Cointegrating Rank . . . 79
3.5.4 Estimation of the Cointegrating Space . . . 81
3.5.5 Testing Hypotheses on the Cointegrating Space . . . 82
3.6 Analysis of the Data Set From 1989 to 2018 . . . 83
3.6.1 Choice of System Parameters . . . 83
3.6.2 Deterministic Components . . . 83
3.6.3 Determination of the Cointegrating Rank . . . 84
3.6.4 Estimation of the Cointegrating Space . . . 85
3.6.5 Testing Hypotheses on the Cointegrating Space . . . 86
3.7 Summary and Conclusion . . . 87

Appendix of the First Chapter **88**
A.1 Proofs of the Results of Section 1.3 . . . 89
A.1.1 Proof of Lemma 1.1 . . . 89
A.1.2 Proof of Lemma 1.2 . . . 90
A.1.3 Proof of Theorem 1.2 . . . 90
A.2 Proofs of the Results of Section 1.4 . . . 91
A.2.1 Proof of Theorem 1.3 . . . 91
A.2.2 Proof of Theorem 1.4 . . . 92
A.2.3 Proof of Theorem 1.5 . . . 92

Appendix of the Second Chapter **95**
B.1 Preliminaries . . . 95
B.2 Proofs of the Theorems . . . 99
B.2.1 Proof of Theorem 2.1 . . . 101
B.2.2 Proof of Theorem 2.2 (Asymptotic Distribution) and Corollary 2.2 . . . 112

Appedix of the Third Chapter **127**
C.1 Supplementary Material . . . 127

Eidesstattliche Erklärung **129**

Erklärung zu den Beiträgen der einzelnen Autoren zu den Artikeln

Die vorliegende Dissertation besteht aus drei Artikeln, „*A Parameterization of Models for Unit Root Processes: Structure Theory and Hypothesis Testing*“, „*Pseudo Maximum Likelihood Parameter Estimation for Multiple Frequency I(1) Processes: A State Space Approach*“ und „*Stochastic Trends and Economic Fluctuations: Reconsidered From the State Space*“, die jeweils von Professor Dr. Dietmar Bauer, Professor Dr. Martin Wagner, Lukas Matuschek und mir verfasst wurden.
Die Resultate des ersten Artikels „*A Parameterization of Models for Unit Root Processes: Structure Theory and Hypothesis Testing*“ stammen von Lukas Matuschek und mir. Ausgehend davon verfasste Professor Dr. Dietmar Bauer einen Entwurf, bei dem die Gliederung des Artikels entstand. Bei der anschließenden Überarbeitung im Rahmen zweier Revisionen schrieb Professor Dr. Martin Wagner die Einleitung. Die Revisionen der übrigen Abschnitte wurden in enger Abstimmung und Diskussion mit Professor Dr. Martin Wagner durchgeführt, Abschnitte zwei und drei primär von mir, Abschnitte vier und fünf primär von Lukas Matuschek.
geschätzter Anteil der Autoren am Artikel: Patrick de Matos Ribeiro: 30%, Lukas Matuschek: 30%, Professor Dr. Martin Wagner: 20%, Professor Dr. Dietmar Bauer: 20%.
Den ersten Entwurf des zweiten Artikels „*Pseudo Maximum Likelihood Parameter Estimation for Multiple Frequency I(1) Processes: A State Space Approach*“ verfassten Professor Dr. Dietmar Bauer und Professor Dr. Martin Wagner. Nachdem die Autoren gemeinsam beschlossen hatten einen deterministischen linearen Trend, eine Konstante und Seasonal Dummies im Daten generierenden Prozess zuzulassen, führten Lukas Matuschek und ich die benötigten Änderungen und erforderlichen Ausweitungen der Beweise durch. Anschließend formulierte Professor Bauer den Beweis von Theorem 1 und die Einleitung neu. Des Weiteren schrieb ich Abschnitt vier, der die Simulationsstudien enthält. Der dafür verwendete Code wurde von Lukas Matuschek und mir erstellt und verwendet auch Code von Professor Dr. Dietmar Bauer zur Schätzung des Startwertes in der Optimierung.
geschätzter Anteil der Autoren am Artikel: Patrick de Matos Ribeiro: 35%, Lukas Matuschek: 25%, Professor Dr. Martin Wagner: 20%, Professor Dr. Dietmar Bauer: 20%.
Den dritten Artikel „*Stochastic Trends and Economic Fluctuations: Reconsidered From the State Space*“ habe ich mit Hilfe von Anregungen von Professor Dr. Martin Wagner eigenständig verfasst. Der verwendete Code für die Startschätzer stammt von Professor Dr. Dietmar Bauer, der Rest des Codes wurde von Lukas Matuschek und mir gemeinsam verfasst.
geschätzter Anteil der Autoren am Artikel: Patrick de Matos Ribeiro: 75%, Lukas Matuschek: 10%, Professor Dr. Martin Wagner: 10%, Professor Dr. Dietmar Bauer: 5%.

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 through finance to climate science. With the early exceptions of Aoki (1990) and Aoki and Havenner (1991) most authors focus on the VAR framework, most notably Johansen (1995) who developed vector error correction models for the empirically most relevant cases, the I(1) and the I(2) case. Johansen and Schaumburg (1999) extended the vector error correction from the I(1) to the multiple frequency I(1) case, which covers the case of seasonal integration.
Limiting 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, as shown by Amemiya and Wu (1972), 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.
Poskitt (2006) uses an error correction model for VARMA cointegration analysis in the I(1) case. However, the extension of this model to the case of higher integration orders or multiple unit roots is not straightforward. One possibility to overcome the difficulties for the cointegration analysis of VARMA processes is the usage of the state space framework, see, e. g., Hannan and Deistler (1988). This dissertation provides important tools for cointegration analysis in the state space framework, namely a continuous parameterization and a pseudo maximum likelihood estimator for the multiple frequency I(1) case.

Chapter 1 discusses the parameterization of state space processes of arbitrary integration orders. Since the state space representation of a stochastic process is not unique, a canonical form is necessary, which selects one unique state space representation. Since this canonical form places restrictions on the system matrices, not all entries of the matrices are free parameters. Some entries are restricted to be zero or depend on other entries. The parametrization is based on the canonical form of Bauer and Wagner (2012), which is particularly well suited for cointegration analysis. As shown by Hazewinkel and Kalman (1976), there is no continuous parameterization for all state space systems of a given system order. Therefore, we partition the set of all systems into subsets on which a continuous parameterization is possible. For this we use a multi-index, which is chosen in such a way that properties like the unit roots, integration orders and dimensions of the cointegrating spaces remain constant in each subset. In addition to deriving a continuous parametrization, which is almost everywhere continuously invertible, we find a generic subset which is open and dense in the set of all integrated processes with a state space representation of a given system order. Additionally, we discuss the topological structure of the subsets, defining a partial ordering of the multi-indices. This is important, since we show that if the multi-index is chosen to large the true transfer function is in the closure of the transfer functions in the subset, which potentially

allows for consistent estimation.
Finally, we discuss the implementation of hypotheses on the cointegrating ranks and spaces in the parametrization for the empirically most relevant cases, the multiple frequency I(1) and the I(2) case. We show that all hypotheses commonly tested for VAR processes in these cases can be implemented in the state space framework. This potentially allows for the derivation of pseudo likelihood ratio tests for these hypotheses.

Chapter 2 examines pseudo maximum likelihood estimation for multiple frequency I(1) processes. We derive the likelihood function for multiple frequency I(1) processes and show that the pseudo maximum likelihood estimator is consistent under relatively mild conditions. Additionally, we show that setting the starting values of the state process to zero does not affect the asymptotic properties of the pseudo maximum likelihood estimator. Consistency does not only hold for a correctly chosen multi-index, but also when the generic multi-index found in Chapter 1 is used. The cointegrating space is estimated super-consistently. The proof of consistency is independent of the parameterization, similar to the proof of Hannan and Deistler (1988) in the stationary case. For the case of a correctly chosen multi-index we additionally derive the asymptotic distribution of the pseudo maximum likelihood estimator, providing the ground work for pseudo likelihood ratio tests.
In a simulation study we compare the pseudo maximum likelihood estimator to reduced rank regression on vector autoregressive approximations by Johansen and Schaumburg (1999) and to the CCA subspace algorithm of Bauer and Buschmeier (2016). In the case of moving average roots close to the unit circle the pseudo maximum likelihood algorithm estimates the cointegrating space more precisely and makes better predictions than the other algorithms for our simulated systems.

Finally, Chapter 3 consists of an useful tutorial for the analysis of economic time series using the state space framework. Using the analysis of King, Plosser, Stock and Watson (1991) as an illustrative example, we demonstrate that all economically relevant questions examined by King et al. (1991) can also be analyzed using the state space framework. The analysis of King et al. (1991) is based on quarterly US economic data from 1949 to 1988. We compare the methods developed for the state space framework, namely the pseudo maximum likelihood estimator from Chapter 2 and the tests based upon it from Matuschek, de Matos Ribeiro, Bauer and Wagner (2020) to the methods used by King et al. (1991), i. e., the DOLS estimator of Stock and Watson (1993) and the tests for the cointegrating rank of Stock and Watson (1988) and to the vector error correction model for I(1) processes by Johansen (1995). In doing so, we point out which `matlab` procedures we use for the analysis. The results obtained with the three different approaches differ, which indicates that the results of empirical applications to time series of dimension six or more of sample sizes below two or three hundred should be interpreted with care.
Additionally, we test the robustness of the vector error correction model and the state space framework by repeating the analysis on an extended data set with quarterly US economic data from 1949 to 2018 and on the subset with data from 1989 to 2018. The results of both approaches differ for the three data sets. This may be a hint that there are structural breaks in the economic time series.

All simulations and computations for empirical applications have been performed in `matlab`. The code containing the respective procedures can be obtained from the author upon request.

Acknowledgments

I want to thank my supervisor and co-author Professor Dr. Martin Wagner for introducing me to this interesting topic and for many fruitful discussions. Although I did not always want to hear what he told me, his advice and constructive criticism improved this thesis a lot.
Financial support from the DFG within the projects BA 5404/1-1 and WA3427/1-1 is gratefully acknowledged.
I also want to thank Professor Dr. Bauer from Bielefeld University, our collaboration partner within those projects, who became a sort of second supervisor when Professor Dr. Wagner changed university and included me in his research group. Thanks for reviewing my dissertation and everything else in the past five years.
I want to thank Professor Walter Krämer for taking the time to review my dissertation.
Special thanks to my colleagues both in Dortmund and Bielefeld for helpful advice and pleasant collaboration. In particular I want to mention Lukas Matuschek who was my colleague both in Dortmund and Bielefeld, is a co-author of all papers in this dissertation and became a good friend during the past five years and Rainer Buschmeier and Dr. Yuanyuan Li who were part of the team on the DFG projects.
Lastly, I want to thank my parents for their support.

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

1.1 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 that allows by and large for least squares based estimation, 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 1.2, e. g., in the 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 complicated parameterization restrictions with complex 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

[1] 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).

[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".

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, Chapters 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 like continuity and differentiability of parameterizations, are discussed in detail in Hannan and Deistler (1988, Chapter 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 1.3.

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 like 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 1.4, 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 upon 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 function[4]. Additionally, the inclusion properties between different sets M_Γ need to be understood,

[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 has been extended to (co)integrated processes by Saikkonen (1992), Saikkonen and Luukkonen (1997) and Bauer and Wagner (2007). 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.

[4]Below we often use the term "likelihood" as short form of "likelihood function".

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 1.5, 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 like intercept, seasonal dummies and linear trend as well as restrictions on these components.

The paper is organized as follows: Section 1.2 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 1.3 presents the canonical form and the parameterization based upon 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 1.4 contains a detailed discussion concerning the topological structure of the model sets and Section 1.5 discusses testing of a large number of hypotheses on the cointegrating spaces commonly tested in the cointegration literature. The discussion in Section 1.5 focuses on the empirically most relevant MFI(1) and I(2) cases and includes the usual deterministic components considered in the literature. Section 1.6 briefly summarizes and concludes. All proofs are relegated to the appendix.

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 $r \leq s$, 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 $\overline{x}$ 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.

1.2 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

$$a(L)y_t = y_t + \sum_{j=1}^{p} a_j y_{t-j} = \varepsilon_t + \Phi d_t, \tag{1.1}$$

with $a(L) := I_s + \sum_{j=1}^{p} a_j L^j$, where $a_j \in \mathbb{R}^{s\times s}$ for $j = 1, \ldots, p$, $\Phi \in \mathbb{R}^{s\times m}$, $a_p \neq 0$, 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 $\Phi \in \mathbb{R}^{s\times m}$.

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

[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 1.2 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).

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

$$\boldsymbol{\theta} := [\boldsymbol{\theta}_a', \boldsymbol{\theta}_\Phi']' = [\mathrm{vec}(a_1)', \ldots, \mathrm{vec}(a_p)', \mathrm{vec}(\Phi)']' \tag{1.2}$$

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

Remark 1.1 *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 $\Theta_{p,\Phi,\Sigma} \subset \mathbb{R}^{s^2p+sm+\frac{s(s+1)}{2}}$. We omit θ_Σ for brevity, since typically no cross-parameter restrictions involving parameters corresponding to Σ are considered, whereas, as discussed in Section 1.5, 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] *In the absence of cross-parameter restrictions involving θ_Σ, the variance covariance matrix Σ is typically either estimated from least squares or reduced rank regression residuals (in a VAR setting) or concentrated out in pseudo maximum likelihood estimation. 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.*

Remark 1.2 *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 $\boldsymbol{\theta}_a \in \Theta_p$ – 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

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

[8] The set V_p is endowed with the *pointwise topology*, defined in Section 1.3. 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 $a_1, \ldots, a_p$ in the Frobenius norm.

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 $\overline{\Theta}_{p,\Phi}$. 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 (1.1) 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 $V_{p,\bullet}$:

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

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 $\overline{V}_{p,\bullet} \setminus V_{p,\bullet}$, 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 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 1.1, 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, Stock and Watson (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 1.1 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 1.3 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 $a_j,\ j = 1, \ldots, p$. 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 $0 \leq \omega \leq \pi$

$$\Delta_\omega \quad := \quad \begin{cases} I_s - 2\cos(\omega)L + L^2 & \text{for } 0 < \omega < \pi \\ I_s - \cos(\omega)L & \text{for } \omega \in \{0, \pi\} \end{cases}. \tag{1.3}$$

For notational brevity, we omit the dependence on L in $\Delta_\omega(L)$, henceforth denoted as Δ_ω. Using

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

this notation, the I(1) error correction representation is given by

$$\begin{aligned}\Delta_0 y_t &= \Pi y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta_0 y_{t-j} + \varepsilon_t + \Phi d_t \\ &= \alpha\beta' y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta_0 y_{t-j} + \varepsilon_t + \Phi d_t,\end{aligned} \tag{1.4}$$

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 $\Gamma_j := \sum_{m=j+1}^{p} a_m$, $j = 1, \ldots, p-1$.

This constitutes a reparameterization, where $k(z) \in V_p$ is now represented by the matrices $(\alpha, \beta, \Gamma_1, \ldots, \Gamma_{p-1})$ and a corresponding parameter vector $\boldsymbol{\theta}_a^{\text{VECM}} \in \Theta_{p,r}^{\text{VECM}}$. Note that stacking the entries of the matrices does not lead to a homeomorphic mapping from V_p to $\Theta_{p,s}^{\text{VECM}}$, 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 $\beta = [I_r, \beta^{*\prime}]'$, with $\beta^* \in \mathbb{R}^{(s-r)\times r}$ and $\alpha \in \mathbb{R}^{s\times r}$. With this additional restriction the parameter vector $\boldsymbol{\theta}_a^{\text{VECM}}$ is given by stacking the vectorized matrices $\alpha, \beta^*, \Gamma_1, \ldots, \Gamma_{p-1}$, similarly to (1.2). Then, $\Theta_{p,r,\Phi}^{\text{VECM}} = \Theta_{p,r}^{\text{VECM}} \times \mathbb{R}^{sm} \subset \mathbb{R}^{ps^2-(s-r)^2+sm}$. Note for completeness that the normalization of $\beta = [I_r, \beta^{*\prime}]'$ 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:

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

The dimension of the parameter vector $\boldsymbol{\theta}_a^{\text{VECM}}$ 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, Chapter 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, Chapter 7).

Furthermore, since $V_{p,r} \subset V_{p,r^*}$ for $r < r^* \leq s$, with $\Theta_{p,r}^{\text{VECM}}$ a lower dimensional subset of $\Theta_{p,r^*}^{\text{VECM}}$, 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 (1.4) 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

$$\alpha_\perp'\Gamma\beta_\perp := \alpha_\perp'\left(I_s - \sum_{j=1}^{p-1}\Gamma_j\right)\beta_\perp \tag{1.5}$$

has full rank. It is well-known that condition (1.5) 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

$$\Delta_0^2 y_t = \alpha\beta' y_{t-1} - \Gamma\Delta_0 y_{t-1} + \sum_{j=1}^{p-2}\Psi_j \Delta_0^2 y_{t-j} + \varepsilon_t, \tag{1.6}$$

[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.

with α, β as defined in (1.4), Γ as defined in (1.5) and $\Psi_j := -\sum_{k=j+1}^{p-1} \Gamma_k$, $j = 1, \ldots, p-2$. From (1.5) we already know that reduced rank of

$$\alpha'_\perp \Gamma \beta_\perp \quad =: \quad \xi \eta', \tag{1.7}$$

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

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

which again is typically ignored in estimation, just like condition (1.5) 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 to conveniently study 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 $s = r_0 + r_1 + r_2$. 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 $\{\beta' y_t\}_{t\in\mathbb{Z}}$ 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 1.3).[11] Note that again, as already before in the I(1) case, different values of p and ranks r and m, respectively 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, Section 2.2) and (1.7). 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, Chapter 2).

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

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

with $a(L) := I_s + \sum_{j=1}^{p} a_j L^j$, where $a_j \in \mathbb{R}^{s\times s}$ for $j = 1, \ldots, p$, $a_p \neq 0$ 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 $j = 1, \ldots, q$, $b_q \neq 0$ 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

[11] Below Example 1.3 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 1.5.2.

$(a(z), b(z))$ can realize the same transfer function $k(z)$. It is clear that for all non-singular polynomial matrices $m(z)$ the equality $k(z) = a(z)^{-1}m(z)^{-1}m(z)b(z) = a(z)^{-1}b(z)$ holds. 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 has been 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.

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 like the "cointegration indices" of Paruolo (1996) mentioned before.

[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.

[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, Chapter 2.4) for a precise definition.

1.3 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 (1.3):

Definition 1.1 *The s-dimensional real VARMA process* $\{y_t\}_{t\in\mathbb{Z}}$ *has* unit root structure $\Omega := ((\omega_1, h_1), \ldots, (\omega_l, h_l))$ *with* $0 \leq \omega_1 < \omega_2 < \cdots < \omega_l \leq \pi, h_k \in \mathbb{N}$, $k = 1, \ldots, l$, $l \geq 1$, *if it is a solution of the difference equation*

$$\Delta_\Omega(y_t - \Phi d_t) := \prod_{k=1}^{l} \Delta_{\omega_k}^{h_k}(y_t - \Phi d_t) = v_t, \tag{1.8}$$

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* $z = e^{i\omega_k}, k = 1, \ldots, l$.

- *The process* $\{y_t\}_{t\in\mathbb{Z}}$ *is called* unit root process *with* unit roots $z_k := e^{i\omega_k}$ *for* $k = 1, \ldots, l$, *the set* $F(\Omega) := \{\omega_1, \ldots, \omega_l\}$ *is the set of* unit root frequencies *and the integers* h_k, $k = 1, \ldots, l$ *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* $((\omega_1, 1), \ldots, (\omega_l, 1))$ *is an* MFI(1), *process.*

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

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 is[15]

$$\begin{aligned} y_t &= Cx_t + \Phi d_t + \varepsilon_t, \\ x_{t+1} &= Ax_t + B\varepsilon_t, \end{aligned} \tag{1.9}$$

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 1.3 *Bauer and Wagner (2012, Theorem 2) show that every real valued unit root VARMA process* $\{y_t\}_{t\in\mathbb{Z}}$ *as given in* (1.8) *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* (1.9) *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 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 1.4, 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 1.1 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 $\pi(A, B, C) := I_s + zC(I_n - zA)^{-1}B = k(z)$. Note that, as

[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, Chapter 7.1).

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 $(\mathcal{A}, \mathcal{B}, \mathcal{C})$. The set M_n is defined as

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

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

Definition 1.2 *A matrix $B = [b_{i,j}]_{i=1,\ldots,c, j=1,\ldots,s} \in \mathbb{C}^{c \times s}$ is positive upper triangular (p.u.t.) if there exist integers $1 \leq j_1 \leq j_2 \leq \cdots \leq j_c \leq s + 1$, such that for $j_i \leq s$ we have $b_{i,j} = 0, j < j_i, j_i < j_{i+1}$, $b_{i,j_i} \in \mathbb{R}^+$. For $j_i = s + 1$ it holds that $b_{i,j} = 0, 1 \leq j \leq s$, i. e., B is of the form*

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

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 1.1 *For every transfer function $k(z) \in M_n$ there exists a unique minimal (complex) state space realization $(\mathcal{A}, \mathcal{B}, \mathcal{C})$ such that*

$$\begin{aligned} y_t &= \mathcal{C}x_{t,\mathbb{C}} + \varepsilon_t, \\ x_{t+1,\mathbb{C}} &= \mathcal{A}x_{t,\mathbb{C}} + \mathcal{B}\varepsilon_t \end{aligned}$$

with:

(i) $\mathcal{A} := diag(\mathcal{A}_u, \mathcal{A}_\bullet) := diag(\mathcal{A}_{1,\mathbb{C}}, \ldots, \mathcal{A}_{l,\mathbb{C}}, \mathcal{A}_\bullet)$, $\mathcal{A}_u \in \mathbb{C}^{n_u \times n_u}, \mathcal{A}_\bullet \in \mathbb{R}^{n_\bullet \times n_\bullet}$, *where it holds for $k = 1, \ldots, l$ that*

- *for $0 < \omega_k < \pi$:*

$$\mathcal{A}_{k,\mathbb{C}} := \begin{bmatrix} J_k & 0 \\ 0 & \overline{J}_k \end{bmatrix} \in \mathbb{C}^{2d^k \times 2d^k},$$

- *for $\omega_k \in \{0, \pi\}$:*

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

with

$$J_k := \begin{bmatrix} \overline{z_k} I_{d_1^k} & [I_{d_1^k}, 0_{d_1^k \times (d_2^k - d_1^k)}] & 0 & \cdots & 0 \\ 0_{d_2^k \times d_1^k} & \overline{z_k} I_{d_2^k} & [I_{d_2^k}, 0_{d_2^k \times (d_3^k - d_2^k)}] & 0 & \vdots \\ 0 & 0 & \overline{z_k} I_{d_3^k} & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & [I_{d_{h_k-1}^k}, 0_{d_{h_k-1}^k \times (d_{h_k}^k - d_{h_k-1}^k)}] \\ 0 & 0 & \cdots & 0 & \overline{z_k} I_{d_{h_k}^k} \end{bmatrix}, \tag{1.10}$$

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

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

- *for* $0 < \omega_k < \pi$*:*

$$\mathcal{B}_{k,\mathbb{C}} := \begin{bmatrix} \mathcal{B}_k \\ \overline{\mathcal{B}}_k \end{bmatrix} \in \mathbb{C}^{2d^k \times s} \text{ and } \mathcal{C}_{k,\mathbb{C}} := \begin{bmatrix} \mathcal{C}_k, & \overline{\mathcal{C}}_k \end{bmatrix} \in \mathbb{C}^{s \times 2d^k}.$$

- *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* $\mathcal{B}_k = [\mathcal{B}_{k,1}', \ldots, \mathcal{B}_{k,h_k}']'$ *as* $\mathcal{B}_{k,h_k} = [\mathcal{B}_{k,h_k,1}', \ldots, \mathcal{B}_{k,h_k,h_k}']'$*, 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* $j = 1, \ldots, h_k$ *and* $k = 1, \ldots, l$.

(iv) For $k = 1, \ldots, l$ *define* $\mathcal{C}_k = [\mathcal{C}_{k,1}, \mathcal{C}_{k,2}, \ldots, \mathcal{C}_{k,h_k}], \mathcal{C}_{k,j} = [\mathcal{C}_{k,j}^G, \mathcal{C}_{k,j}^E]$*, 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* $j = 1, \ldots, h_k$*, with* $d_0^k := 0$*. Further, define* $\mathcal{C}_k^E := [\mathcal{C}_{k,1}^E, \ldots, \mathcal{C}_{k,h_k}^E] \in \mathbb{C}^{s \times d_{h_k}^k}$*. 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* $j = 2, \ldots, h_k$ *and* $k = 1, \ldots, l$.

(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 1.4 *As indicated in Remark 1.3 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 1.1 as follows. First, define a transformation matrix (and its inverse):*

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

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

$$\begin{aligned} y_t &= \mathcal{C}_\mathbb{R} x_{t,\mathbb{R}} + \varepsilon_t, \\ x_{t+1,\mathbb{R}} &= \mathcal{A}_\mathbb{R} x_{t,\mathbb{R}} + \mathcal{B}_\mathbb{R} \varepsilon_t, \end{aligned}$$

with real valued matrices $(\mathcal{A}_\mathbb{R}, \mathcal{B}_\mathbb{R}, \mathcal{C}_\mathbb{R})$ *follows from using the just defined transformation matrix. In particular it holds that:*

$$\begin{aligned} \mathcal{A}_\mathbb{R} &:= diag(\mathcal{A}_{u,\mathbb{R}}, \mathcal{A}_\bullet) &&:= diag(\mathcal{A}_{1,\mathbb{R}}, \ldots, \mathcal{A}_{l,\mathbb{R}}, \mathcal{A}_\bullet), \\ \mathcal{B}_\mathbb{R} &:= [\mathcal{B}_{u,\mathbb{R}}', \mathcal{B}_\bullet']' &&:= [\mathcal{B}_{1,\mathbb{R}}', \ldots, \mathcal{B}_{l,\mathbb{R}}', \mathcal{B}_\bullet']', \\ \mathcal{C}_\mathbb{R} &:= [\mathcal{C}_{u,\mathbb{R}}, \mathcal{C}_\bullet] &&:= [\mathcal{C}_{1,\mathbb{R}}, \ldots, \mathcal{C}_{l,\mathbb{R}}, \mathcal{C}_\bullet], \end{aligned}$$

with

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

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* $x_{t,\mathbb{C}} := [x'_{t,u},x'_{t,\bullet}]' := [x'_{t,1,\mathbb{C}},\ldots,x'_{t,l,\mathbb{C}},x'_{t,\bullet}]'$*, where*

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

with

$$x_{t+1,k} = J_kx_{t,k}+\mathcal{B}_k\varepsilon_t, \qquad \text{for } k=1,\ldots,l.$$

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

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

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

with the corresponding block of the state equation given by

$$x_{t+1,k,\mathbb{R}} = \mathcal{A}_{k,\mathbb{R}}x_{t,k,\mathbb{R}}+\mathcal{B}_{k,\mathbb{R}}\varepsilon_t.$$

For $k=1,\ldots,l$ *the sub-vectors* $x_{t,k,\mathbb{R}}$ *are further decomposed into* $x_{t,k,\mathbb{R}} := [(x_{t,k,\mathbb{R}}^1)',\ldots,(x_{t,k,\mathbb{R}}^{h_k})']'$*, 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* $j=1,\ldots,h_k$ *and* $\mathcal{C}_{k,\mathbb{R}} := [\mathcal{C}_{k,1,\mathbb{R}},\ldots,\mathcal{C}_{k,h_k,\mathbb{R}}]$ *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* $j=1,\ldots,h_k$ *and* $k=1,\ldots,l$*. Furthermore, for* $j=1,\ldots,h_k$ *and* $k=1,\ldots,l$ *the processes* $\{x_{t,k,\mathbb{R}}^j\}_{t\in\mathbb{Z}}$ *are not cointegrated, as defined in Definition 1.3 below. For* $\omega_k=0$*, the process* $\{x_{t,k,\mathbb{R}}^j\}_{t\in\mathbb{Z}}$ *is the* d_k^j*-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 1.5 *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 upon it can be used instead for the stable subsystem.*

Remark 1.6 *Starting from a state space system* (1.9) *with matrices* $(\mathcal{A},\mathcal{B},\mathcal{C})$ *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*

$$y_t = \sum_{j=1}^{t-1}\mathcal{C}_u\mathcal{A}_u^{j-1}\mathcal{B}_u\varepsilon_{t-j}+\mathcal{C}_u\mathcal{A}_u^{t-1}x_{1,u}+\sum_{j=1}^{t-1}\mathcal{C}_\bullet\mathcal{A}_\bullet^{j-1}\mathcal{B}_\bullet\varepsilon_{t-j}+ \mathcal{C}_\bullet\mathcal{A}_\bullet^{t-1}x_{1,\bullet}+\Phi d_t+\varepsilon_t.$$

Clearly, the term $\mathcal{C}_u\mathcal{A}_u^{t-1}x_{1,u}$ 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" $\mathcal{C}_u\mathcal{A}_u^{t-1}x_{1,u}$, 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 $\mathcal{A}_u = I$ 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 $\mathcal{C}_\bullet\mathcal{A}_\bullet^{t-1}x_{1,\bullet}$ 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 $x_{1,\bullet} = \sum_{j=1}^{\infty}\mathcal{A}_\bullet^{j-1}\mathcal{B}_\bullet\varepsilon_{1-j}$. 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. 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 $(\mathcal{A},\ \mathcal{B},\ \mathcal{C})$ implies full row rank of the p.u.t. blocks $\mathcal{B}_{k,h_k,j}$ of $\mathcal{B}_{k,h_k}$. 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 (1.10) 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, Chapter 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 $k(z) = I_s + z\mathcal{C}_1(1-z)^{-1}\mathcal{B}_1 = I_s + \mathcal{C}_1\mathcal{B}_1 z(1-z)^{-1}$ 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 1.1 has been 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 1.3 *(i) Let $\tilde{\Omega} = ((\tilde{\omega}_1, \tilde{h}_1), \ldots, (\tilde{\omega}_{\tilde{l}}, \tilde{h}_{\tilde{l}}))$ and $\Omega = ((\omega_1, h_1), \ldots, (\omega_l, h_l))$ 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 $\tilde{h}_{\tilde{k}} \le h_k$.*

Further, $\tilde{\Omega} \prec \Omega$ 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 $k = 1, \ldots, l$, as

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

(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* $\beta(z) = \sum_{m=0}^{q} \beta_m z^m$, $\beta_m \in \mathbb{R}^s$, $m = 0, \ldots, q$, $\beta_q \neq 0$, *for some integer* $1 \leq q < \infty$ *such that*

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

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 $(\Omega, \tilde{\Omega})$.

Remark 1.7 *(i) It is merely a matter of taste whether cointegrating spaces are defined in terms of their order* $(\Omega, \tilde{\Omega})$ *or their decrease* $\delta(\Omega, \tilde{\Omega}) := (\delta_1(\Omega, \tilde{\Omega}), \ldots, \delta_l(\Omega, \tilde{\Omega}))$, *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 1.3 – 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* $\delta_k = 1, \ldots, h_k$. *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

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

By Remark 1.4, 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 $\beta'[\mathcal{C}_{k,1,\mathbb{R}}, \ldots, \mathcal{C}_{k,j,\mathbb{R}}] = 0$ and

[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.

$\beta'\mathcal{C}_{k,j+1,\mathbb{R}} \neq 0$ or equivalently $\beta'[\mathcal{C}_{k,1},\ldots,\mathcal{C}_{k,j}] = 0$ and $\beta'\mathcal{C}_{k,j+1} \neq 0$ (using the transformation to the complex matrices of the canonical form, as discussed in Remark 1.4, and that $\beta'[\mathcal{C}_k, \overline{\mathcal{C}}_k] = 0$ if and only if $\beta'\mathcal{C}_k = 0$). Thus, the CIVs are characterized by orthogonality to sub-blocks of $\mathcal{C}_u$.

The real valued representation given in Remark 1.4 used in its partitioned form just above immediately leads to necessary orthogonality constraint for polynomial cointegration of degree one:

$$
\begin{aligned}
&\beta(L)'(y_t) = \\
&\beta(L)'(\mathcal{C}_{u,\mathbb{R}}x_{t,u,\mathbb{R}} + \mathcal{C}_\bullet x_{t,\bullet} + \Phi d_t + \varepsilon_t) = \\
&\beta_0'\mathcal{C}_{u,\mathbb{R}}x_{t,u,\mathbb{R}} + \beta_1'\mathcal{C}_{u,\mathbb{R}}x_{t-1,u,\mathbb{R}} + \beta(L)'(\mathcal{C}_\bullet x_{t,\bullet} + \Phi d_t + \varepsilon_t) = \\
&\beta_0'\mathcal{C}_{u,\mathbb{R}}(\mathcal{A}_{u,\mathbb{R}}x_{t-1,u,\mathbb{R}} + \mathcal{B}_{u,\mathbb{R}}\varepsilon_{t-1}) + \beta_1'\mathcal{C}_{u,\mathbb{R}}x_{t-1,u,\mathbb{R}} + \beta(L)'(\mathcal{C}_\bullet x_{t,\bullet} + \Phi d_t + \varepsilon_t) = \\
&(\beta_0'\mathcal{C}_{u,\mathbb{R}}\mathcal{A}_{u,\mathbb{R}} + \beta_1'\mathcal{C}_{u,\mathbb{R}})x_{t-1,u,\mathbb{R}} + \beta_0'\mathcal{C}_{u,\mathbb{R}}\mathcal{B}_{u,\mathbb{R}}\varepsilon_{t-1} + \beta(L)'(\mathcal{C}_\bullet x_{t,\bullet} + \Phi d_t + \varepsilon_t) = \\
&(\beta_0'\mathcal{C}_u\mathcal{A}_u + \beta_1'\mathcal{C}_u)x_{t-1,u} + \beta_0'\mathcal{C}_u\mathcal{B}_u\varepsilon_{t-1} + \beta(L)'(\mathcal{C}_\bullet x_{t,\bullet} + \Phi d_t + \varepsilon_t)
\end{aligned}
$$

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 $\begin{bmatrix} \beta_0' & \beta_1' \end{bmatrix}'$ to sub-blocks of $\begin{bmatrix} \mathcal{C}_{u,\mathbb{R}}\mathcal{A}_{u,\mathbb{R}} \\ \mathcal{C}_{u,\mathbb{R}} \end{bmatrix}$ or sub-blocks of the complex matrix $\begin{bmatrix} \mathcal{C}_u\mathcal{A}_u \\ \mathcal{C}_u \end{bmatrix}$. Note, however, that this orthogonality condition is not sufficient for $[\beta_0', \beta_1']'$ to be a PCIV, because it does not imply $\max_{k=1,\ldots,l}\|\beta(e^{i\omega_k})\|\delta_k(\Omega, \tilde{\Omega}) \neq 0$. For a detailed discussion of polynomial cointegration, when considering also higher polynomial degrees, see Bauer and Wagner (2012, Section 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 1.1 (Cointegration in the I(1) case) *In the I(1) case, neglecting the stable subsystem and the deterministic components for simplicity, it holds that*

$$
\begin{aligned}
y_t &= \mathcal{C}_1 x_{t,1} + \varepsilon_t, \quad y_t, \varepsilon_t \in \mathbb{R}^s, x_{t,1} \in \mathbb{R}^{d_1^1}, \mathcal{C}_1 \in \mathbb{R}^{s \times d_1^1}, \\
x_{t+1,1} &= x_{t,1} + \mathcal{B}_1\varepsilon_t, \quad \mathcal{B}_1 \in \mathbb{R}^{d_1^1 \times s}.
\end{aligned}
$$

The vector $\beta \in \mathbb{R}^s, \beta \neq 0$, is a CIV of order $((0,1),\{\})$ if and only if $\beta'\mathcal{C}_1 = 0$.

Example 1.2 (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*

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

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

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

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

$$[\beta_0', \beta_1'] \begin{bmatrix} \overline{z}_k\mathcal{C}_k & z_k\overline{\mathcal{C}}_k \\ \mathcal{C}_k & \overline{\mathcal{C}}_k \end{bmatrix} = 0, \tag{1.11}$$

which is equivalent to

$$(\overline{z}_k\beta_0' + \beta_1')\mathcal{C}_k = 0.$$

The fact that the matrix in (1.11) *has a block structure with two blocks of conjugate complex columns implies some additional structure also on 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*

$$\begin{aligned}(\overline{z}_k(-\beta_1)' + (\beta_0 + 2\cos(\omega_k)\beta_1)')\mathcal{C}_k &= (\beta_0' + (2\mathcal{R}(z_k) - \overline{z}_k)\beta_1')\mathcal{C}_k \\ &= (\beta_0' + z_k\beta_1')\mathcal{C}_k \\ &= z_k(\overline{z}_k\beta_0' + \beta_1')\mathcal{C}_k = 0.\end{aligned}$$

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

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

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

$$\beta'\mathcal{C}_{1,1}^E = 0 \quad \textit{and} \quad \beta'[\mathcal{C}_{1,2}^G, \mathcal{C}_{1,2}^E] \neq 0.$$

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

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

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*

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

The above orthogonality constraint indicates that the two cases $\mathcal{C}_{1,2}^G = 0$ *and* $\mathcal{C}_{1,2}^G \neq 0$ *have to be considered separately for polynomial cointegration analysis. Consider first the case* $\mathcal{C}_{1,2}^G = 0$*. In this case the orthogonality constraints imply* $\beta_0'\mathcal{C}_{1,1}^E = 0$*,* $\beta_1'\mathcal{C}_{1,1}^E = 0$ *and* $(\beta_0+\beta_1)'\mathcal{C}_{1,2}^E = 0$*. 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* $\mathcal{C}_{1,2}^G \neq 0$ *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* $\gamma_1'(1-L)(\{\mathcal{C}_{1,1}^E x_{t,1}^E\}_{t\in\mathbb{Z}}) = \{\gamma_1'\mathcal{C}_{1,1}^E x_{t-1,2}^G + \gamma_1'\mathcal{C}_{1,1}^E\mathcal{B}_{1,1}\varepsilon_{t-1}\}_{t\in\mathbb{Z}}$*, with* $\gamma_1'\mathcal{C}_{1,1}^E \neq 0$*. This term is eliminated by* $\{\gamma_2'\mathcal{C}_{1,2}^G x_{t,2}^G\}_{t\in\mathbb{Z}}$ *in* $\{\gamma_2' y_t\}_{t\in\mathbb{Z}}$*, if* $\gamma_1'\mathcal{C}_{1,1}^E + \gamma_2'\mathcal{C}_{1,2}^G = 0$*, which is only possible if* $\mathcal{C}_{1,2}^G \neq 0$*. Additionally,* $\gamma_2'[\mathcal{C}_{1,1}^E, \mathcal{C}_{1,2}^E] = 0$ *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 has been shown above, the unit root and cointegration properties of $\{y_t\}_{t\in\mathbb{Z}}$ depend on the sub-blocks of $\mathcal{C}_u$ and the eigenvalue structure of $\mathcal{A}_u$. We therefore define the more encompassing state space unit root structure containing information on the geometrical and algebraic multiplicities of the eigenvalues of $\mathcal{A}_u$, cf. Bauer and Wagner (2012, Definition 2).

Definition 1.4 *A unit root process $\{y_t\}_{t\in\mathbb{Z}}$ with a canonical state space representation as given in Theorem 1.1 has* state space unit root structure

$$\Omega_S \quad := \quad \big((\omega_1, d_1^1, \ldots, d_{h_1}^1), \ldots, (\omega_l, d_1^l, \ldots, d_{h_l}^l)\big)$$

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

Remark 1.8 *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, $k = 1, \ldots, l$, $j = 1, \ldots, h_k$ 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 1.4. 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 1.2, 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, Section 3). Here, r_2 is the number of stochastic trends of order two, i. e., $r_2 = d_1^1$. Further, 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 1.1 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 1.5 (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*

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

Remark 1.9 *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 1.9, advantageous for the comparison of unit root structures and structure indices that p is a vector with n_u entries.*

Example 1.4 *Consider the following state space system:*

$$\begin{aligned} y_t &= \begin{bmatrix} \mathcal{C}_{1,1}^E & \mathcal{C}_{1,2}^G & \mathcal{C}_{1,2}^E \end{bmatrix} x_t + \varepsilon_t \qquad y_t,\ \varepsilon_t \in \mathbb{R}^2,\ x_t \in \mathbb{R}^3,\ \mathcal{C}_{1,1}^E, \mathcal{C}_{1,2}^G, \mathcal{C}_{1,2}^E \in \mathbb{R}^{2\times 1} \quad (1.12) \\ x_{t+1} &= \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} x_t + \begin{bmatrix} \mathcal{B}_{1,1} \\ \mathcal{B}_{1,2,1} \\ \mathcal{B}_{1,2,2} \end{bmatrix} \varepsilon_t, \quad x_0 = 0, \quad \mathcal{B}_{1,1}, \mathcal{B}_{1,2,1}, \mathcal{B}_{1,2,2} \in \mathbb{R}^{1\times 2}. \end{aligned}$$

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 $\alpha_\bullet$, cf. Hannan and Deistler (1988, Chapter 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, Chapter 5.1).

Definition 1.6 (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 1.10 *Givens rotations allow to transform any vector $v = [v_1, v_2, ..., v_q]' \in \mathbb{R}^q$ into a vector of the form $[\tilde{v}_1, 0, ..., 0]'$ 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, Chapter 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 $v^{(j+1)} = R_{q,1,q-j+1}(\theta_{q-j})v^{(j)} = [v_1^{(j+1)}, v_2, \ldots, v_{q-j}, 0, \ldots, 0]'$, with $v_1^{(j+1)} \geq 0$.*
3. *If $j = q - 1$, stop. Else increment j by one ($j \to j + 1$) and continue at step 2.*

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

Remark 1.11 *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 a 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 1.7 (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 1.12 *Complex Givens rotations allow to transform any vector $v = [v_1, v_2, ..., v_q]' \in \mathbb{C}^q$ into a vector of the form $[\tilde{v}_1, 0, ..., 0]'$ 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, Chapter 8.1.3, p. 222).*

3. *Set*

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

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

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

This algorithm determines a unique vector $\varphi = [\varphi_{1,1}, \varphi_{1,2}, ..., \varphi_{q-1,2}]'$ *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 $(\mathcal{A}, \mathcal{B}, \mathcal{C})$ 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 $\mathcal{C}$-block corresponds to a real or complex unit root. The I(2) case highlights recursive orthogonality constraints on the parameters of the $\mathcal{C}$-block, which are related to the polynomial cointegration properties (cf. Example 1.3).

1.3.1 The Parameterization in the MFI(1) Case

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

Starting with the sub-blocks of $\mathcal{C}_u$, it is convenient to separate the discussion of the parameterization of $\mathcal{C}_u$-blocks into the real case, where $\omega_k \in \{0, \pi\}$ and $\mathcal{C}_k \in \mathbb{R}^{s \times d_1^k}$, and the complex case with $0 < \omega_k < \pi$ and $\mathcal{C}_k \in \mathbb{C}^{s \times d_1^k}$. 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 $\mathcal{C} \in \mathbb{R}^{s \times d}$. 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 1.6, that allow to transform 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 to transform every matrix $\mathcal{C} \in O_{s,s}$ either to I_s or $I_s^- := \text{diag}(I_{s-1}, -1)$, since, compare Remark 1.11, for the transformed matrix $\tilde{\mathcal{C}}^{(s)}$ it holds that $\det(\mathcal{C}) = \det(\tilde{\mathcal{C}}^{(s)}) \in \{-1, 1\}$. This is achieved with the following algorithm:

1. Set $j = 1$ and $\mathcal{C}^{(1)} = \mathcal{C}$.

2. Transform the entries $[c_{j,j}, \ldots, c_{j,d}]$ in the j-th row of $\mathcal{C}^{(j)}$, to $[\tilde{c}_{j,j}, 0, \ldots, 0]$, $\tilde{c}_{j,j} \geq 0$. Since this is a row vector, this is achieved by right-multiplication of $\mathcal{C}^{(j)}$ with transposed Givens rotations and the required parameters are obtained via the algorithm described in Remark 1.10. The first $j-1$ entries of the j-th row remain unchanged. Denote the transformed matrix by $\mathcal{C}^{(j+1)}$.

3. If $j = d-1$ stop. Else increment j by one $(j \to j+1)$ and continue at step 2.

4. Collect all parameters used for the Givens rotations in steps 1 to 3 in a parameter vector $\boldsymbol{\theta}_R$. Steps 1-3 correspond to a QR decomposition of $\mathcal{C}' = Q\tilde{\mathcal{C}}'$, with an orthonormal matrix Q given by the product of the Givens rotation. Note that the first $j-1$ entries of the j-th column of $\tilde{\mathcal{C}} = \mathcal{C}^{(d)}$ are equal to zero by construction.

5. Set $j = 0$ and $\tilde{\mathcal{C}}^{(0)} = \tilde{\mathcal{C}}$.

6. Collect the entries in column $d-j$ of $\tilde{\mathcal{C}}^{(j)}$ which have not been transformed to zero by previous transformations into the vector $[c_{d-j,d-j}, c_{d+1,d-j}, \ldots, c_{s,d-j}]'$. Using the algorithm described in Remark 1.10 transform this vector to $[\tilde{c}_{d-j,d-j}, 0, \ldots, 0]'$ by left-multiplication of $\tilde{\mathcal{C}}^{(j)}$ with Givens rotations. Since Givens rotations are orthonormal, the transformed matrix $\tilde{\mathcal{C}}^{(j+1)}$ 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 ($j \to j+1$) and continue at step 6.

8. Collect all parameters used for the Givens rotations in steps 5 to 7 in a parameter vector $\boldsymbol{\theta}_L$.

The parameter vector $\boldsymbol{\theta} = [\boldsymbol{\theta}_L', \boldsymbol{\theta}_R']'$, contains the angles of the employed Givens rotations and provides one way of parameterizing $O_{s,d}$. The following Lemma 1.1 demonstrates the usefulness of this parameterization.

Lemma 1.1 (Properties of the parameterization of $O_{s,d}$) *Define for $d \leq s$ a mapping $\boldsymbol{\theta} \to C_O(\boldsymbol{\theta})$ from $\Theta_O^{\mathbb{R}} := [0, 2\pi)^{d(s-d)} \times [0, 2\pi)^{d(d-1)/2} \to O_{s,d}$ by*

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

with $\boldsymbol{\theta} := [\boldsymbol{\theta}_L', \boldsymbol{\theta}_R']'$, where $\boldsymbol{\theta}_L := [\theta_{L,1}, \ldots, \theta_{L,d(s-d)}]'$ and $\boldsymbol{\theta}_R := [\theta_{R,1}, \ldots, \theta_{R,d(d-1)/2}]'$. 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 $\mathcal{C} \in O_{s,d}$ there exists a vector $\boldsymbol{\theta} \in \Theta_O^{\mathbb{R}}$ such that

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

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

(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 $\Theta_O^{\mathbb{R}}$. 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 $\mathcal{C} \in O^+_{s,s}$ there exists a vector $\boldsymbol{\theta} \in \Theta^{\mathbb{R}}_O$ such that*

$$\mathcal{C} = C_O(\boldsymbol{\theta}) = R_L(\boldsymbol{\theta}_L)' \begin{bmatrix} I_d \end{bmatrix} R_R(\boldsymbol{\theta}_R) = R_R(\boldsymbol{\theta}_R).$$

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

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

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

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

Remark 1.13 *The following arguments illustrate why C_O^{-1} is not continuous on the pre-image of the boundary of $\Theta^{\mathbb{R}}_O$: 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 1.10, 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, Chapter 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 1.14 *Note that the parameterization partitions the parameter vector $\boldsymbol{\theta}$ into two parts $\boldsymbol{\theta}_L \in [0, 2\pi)^{d(s-d)}$ and $\boldsymbol{\theta}_R \in [0, 2\pi)^{(d-1)d/2}$. Since changing the parameter values in $\boldsymbol{\theta}_R$ does not change the column space of $C_O(\boldsymbol{\theta})$, which, as seen above, determines the cointegrating vectors, $\boldsymbol{\theta}_L$ fully characterizes the (static) cointegrating space. Note that the dimension of $\boldsymbol{\theta}_L$ is $d(s-d)$ and thus coincides with the number of free parameters in β in the VECM framework ,cf. Johansen (1995, Chapter 5.2).*

Example 1.5 *Consider the matrix*

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

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 $\boldsymbol{\theta}_L$ given in Remark 1.14. The parameters in $\boldsymbol{\theta}_R$ correspond to the choice of a basis of the image of C. Having fixed the two-dimensional subspace through $\boldsymbol{\theta}_L$, only one free parameter for the choice of an orthonormal basis remains, which again coincides with the dimension given in Remark 1.14. To obtain the parameter vector, the starting point is a QR decomposition of $C' = R_R(\boldsymbol{\theta}_R)\tilde{C}'$. In this example $R_R(\boldsymbol{\theta}_R) = R_{2,1,2}(\theta_{R,1})$, with $\theta_{R,1}$ to be determined. To find $\theta_{R,1}$, solve $\begin{bmatrix} 0 & \frac{1}{\sqrt{2}} \end{bmatrix} R_{2,1,2}(\theta_{R,1})' = \begin{bmatrix} r & 0 \end{bmatrix}$ 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 $\begin{bmatrix} 0 & \frac{1}{\sqrt{2}} \end{bmatrix} = r \begin{bmatrix} \cos(\theta_{R,1}) & \sin(\theta_{R,1}) \end{bmatrix}$,

which leads to $r = \frac{1}{\sqrt{2}}$, $\theta_{R,1} = \frac{\pi}{2}$. *Thus, the orthonormal matrix* $R_R(\boldsymbol{\theta}_R)$ *is equal to* $R_{2,1,2}\left(\frac{\pi}{2}\right)$ *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 $\boldsymbol{\theta} = [\boldsymbol{\theta}_L', \boldsymbol{\theta}_R']' = \left[\left[\frac{\pi}{4}, \frac{7\pi}{4}\right], \left[\frac{\pi}{2}\right]\right]'$ *with* $\boldsymbol{\theta} = C_O^{-1}(C)$.

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

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

The parameterization of this set is based on the combination of complex Givens rotations, as given in Definition 1.7, 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 $\mathcal{C}^{(1)} = \mathcal{C}$.

2. Transform the entries $[c_{j,j}, \ldots, c_{j,d}]$ in the j-th row of $\mathcal{C}^{(j)}$, to $[\tilde{c}_{j,j}, 0, \ldots, 0]$. Since this is a row vector, this is achieved by right-multiplication of $\mathcal{C}$ with transposed Givens rotations and the required parameters are obtained via the algorithm described in Remark 1.12. The first $j-1$ entries of the j-th row remain unchanged. Denote the transformed matrix by $\mathcal{C}^{(j+1)}$.

3. If $j = d - 1$ stop. Else increment j by one ($j \to j + 1$) and continue at step 2.

4. Collect all parameters used for the Givens rotations in steps 1 to 3 in a parameter vector $\boldsymbol{\varphi}_R$. Step 1-3 corresponds to a QR decomposition of $\mathcal{C}' = Q\tilde{\mathcal{C}}'$, with a unitary matrix Q given by the product of the Givens rotations. Note that the first $j-1$ entries of the j-th column of $\tilde{\mathcal{C}} = \mathcal{C}^{(d)}$ are equal to zero by construction.

5. Set $j = 0$ and $\tilde{\mathcal{C}}^{(0)} = \tilde{\mathcal{C}}$.

6. Collect the entries in column $d-j$ of $\tilde{\mathcal{C}}^{(j)}$ which have not been transformed to zero by previous transformations into the vector $[c_{d-j,d-j}, c_{d+1,d-j}, \ldots, c_{s,d-j}]'$. Using the algorithm described in Remark 1.12 transform this vector to $[\tilde{c}_{d-j,d-j}, 0, \ldots, 0]'$ by left-multiplication of $\tilde{\mathcal{C}}^{(j)}$ with Givens rotations. Since Givens rotations are unitary, the transformed matrix $\tilde{\mathcal{C}}^{(j+1)}$ 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 ($j \to j+1$) and continue at step 6.

8. Collect all parameters used for the Givens rotations in steps 5 to 7 in a parameter vector $\boldsymbol{\varphi}_L$.

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

The following lemma demonstrates the usefulness of this parameterization.

Lemma 1.2 (Properties of the parametrization of $U_{s,d}$) *Define for $d \leq s$ a mapping $\boldsymbol{\varphi} \to C_U(\boldsymbol{\varphi})$ from $\Theta_U^{\mathbb{C}} := \Theta_{\mathbb{C}}^{d(s-d)} \times \Theta_{\mathbb{C}}^{(d-1)d/2} \times [0, 2\pi)^d \to U_{s,d}$ by*

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

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

(i) $U_{s,d}$ is closed and bounded.

(ii) The mapping $C_U(\boldsymbol{\varphi})$ is infinitely often differentiable.

(iii) For every $\mathcal{C} \in U_{s,d}$ a vector $\boldsymbol{\varphi} \in \Theta_U^{\mathbb{C}}$ exists such that

$$\mathcal{C} = C_U(\boldsymbol{\varphi}) = Q_L(\boldsymbol{\varphi}_L)' \begin{bmatrix} D_d(\boldsymbol{\varphi}_D) \\ 0_{(s-d)\times d} \end{bmatrix} Q_R(\boldsymbol{\varphi}_R).$$

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

(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 1.15 *Note the partitioning of the parameter vector $\boldsymbol{\varphi}$ into the parts $\boldsymbol{\varphi}_L$,$\boldsymbol{\varphi}_D$ and $\boldsymbol{\varphi}_R$. The component $\boldsymbol{\varphi}_L$ fully characterizes the column space of $C_U(\boldsymbol{\varphi})$, i. e., $\boldsymbol{\varphi}_L$ determines the cointegrating spaces.*

Example 1.6 *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 $C' = Q_R(\boldsymbol{\varphi}_R)\tilde{C}' = Q_{2,1,2}(\varphi_{R,1})\tilde{C}'$. To find a complex Givens rotation such that $\begin{bmatrix} \frac{1-i}{2} & \frac{1-i}{2} \end{bmatrix} Q_{2,1,2}(\varphi_{R,1})' = \begin{bmatrix} re^{i\varphi_a} & 0 \end{bmatrix}$ with $r > 0$, transform the entries of $\begin{bmatrix} \frac{1-i}{2} & \frac{1-i}{2} \end{bmatrix}'$ into polar coordinates. The equation $\begin{bmatrix} \frac{1-i}{2} & \frac{1-i}{2} \end{bmatrix}' = \begin{bmatrix} ae^{i\varphi_a} & be^{i\varphi_b} \end{bmatrix}'$

has the solutions $a = b = \frac{1}{\sqrt{2}}$ *and* $\varphi_a = \varphi_b = \frac{7\pi}{4}$*. Using the results of Remark 1.12, 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* $Q_{2,1,2}\left(\left[\frac{\pi}{4}, 0\right]\right)'$ *leads to*

$$\tilde{C} = CQ_{2,1,2}\left(\left[\frac{\pi}{4}, 0\right]\right)' = C\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}' = \begin{bmatrix} \frac{1-i}{\sqrt{2}} & 0 \\ 0 & \frac{-1-i}{\sqrt{2}} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} D_2(\boldsymbol{\varphi}_D) \\ 0_{1\times 2} \end{bmatrix}.$$

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* $D_2(\boldsymbol{\varphi}_D) = diag(e^{i\varphi_{D,1}}, e^{i\varphi_{D,2}}) = diag(\frac{1-i}{\sqrt{2}}, \frac{-1-i}{\sqrt{2}})$ *are* $\varphi_{D,1} = \frac{3\pi}{4}$ *and* $\varphi_{D,2} = \frac{5\pi}{4}$*.*
The parameter vector for this matrix is therefore $\boldsymbol{\varphi} = [\boldsymbol{\varphi}'_L, \boldsymbol{\varphi}'_R, \boldsymbol{\varphi}'_D]' = \left[[0,0,0,0], \left[\frac{\pi}{4}, 0\right], \left[\frac{3\pi}{4}, \frac{5\pi}{4}\right]\right]'$*, with* $\boldsymbol{\varphi} = C_U^{-1}(C)$*.*

Components of the Parameter Vector

Based on the results of the preceding sections we can now describe the parameter vectors for the MFI(1) 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 $\alpha_\bullet$ 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:

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

$$\boldsymbol{\theta}_{B,f,k} := \begin{cases} [b^k_{1,p^k_1+1}, b^k_{1,p^k_1+2}, \ldots, b^k_{1,s}, b^k_{2,p^k_2+1}, \ldots, b^k_{d^k_1,s}]' & \text{for } \omega_k \in \{0, \pi\}, \\ [\mathcal{R}(b^k_{1,p^k_1+1}), \mathcal{I}(b^k_{1,p^k_1+1}), \mathcal{R}(b^k_{1,p^k_1+2}), \ldots, \mathcal{I}(b^k_{1,s}), \mathcal{R}(b^k_{2,p^k_2+1}), \ldots, \mathcal{I}(b^k_{d^k_1,s})]' & \\ & \text{for } 0 < \omega_k < \pi, \end{cases}$$

  for $k = 1, \ldots, l$, with p^k_j 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.

- $\boldsymbol{\theta}_{B,p} := [\boldsymbol{\theta}'_{B,p,1}, ..., \boldsymbol{\theta}'_{B,p,l}]' \in \Theta_{B,p} = \mathbb{R}_+^{d_{B,p}}$: The vectors $\boldsymbol{\theta}_{B,p,k} := \left[b^k_{1,p^k_1}, \ldots, b^k_{d^k_1,p^k_{d^k_1}}\right]'$ contain the entries in $\mathcal{B}_k$ restricted by the p.u.t. structures to be positive reals.

- $\boldsymbol{\theta}_{C,E} := [\boldsymbol{\theta}'_{C,E,1}, ..., \boldsymbol{\theta}'_{C,E,l}]' \in \Theta_{C,E} \subset \mathbb{R}^{d_{C,E}}$: The parameters for the matrices $\mathcal{C}_k$ as discussed in Lemma 1.1 and Lemma 1.2.

- $\boldsymbol{\theta}_\bullet \in \Theta_{\bullet,\alpha} \subset \mathbb{R}^{d_\bullet}$: The parameters for the stable subsystem in echelon canonical form for Kronecker indices $\alpha_\bullet$.

Example 1.7 *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*

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

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

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

with parameter values corresponding to the C-blocks collected in $\boldsymbol{\theta}_{C,E}$ *considered in Examples 1.5 and 1.6.*

1.3.2 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 $\Omega_s = ((0, d_1^1, d_2^1))$:

$$\mathcal{A} = \begin{bmatrix} I_{d_1^1} & I_{d_1^1} & 0 & 0 \\ 0 & I_{d_1^1} & 0 & 0 \\ 0 & 0 & I_{d_2^1 - d_1^1} & 0 \\ 0 & 0 & 0 & \mathcal{A}_\bullet \end{bmatrix}, \mathcal{B} = \begin{bmatrix} \mathcal{B}_{1,1} \\ \mathcal{B}_{1,2,1} \\ \mathcal{B}_{1,2,2} \\ \mathcal{B}_\bullet \end{bmatrix},$$
$$\mathcal{C} = \begin{bmatrix} \mathcal{C}_{1,1}^E & \mathcal{C}_{1,2}^G & \mathcal{C}_{1,2}^E & \mathcal{C}_\bullet \end{bmatrix},$$

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

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

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

$$\begin{aligned} O_{s,d_2^1 - d_1^1}(\mathcal{C}_{1,1}^E) := \{&\mathcal{C}_{1,2}^E \in \mathbb{R}^{s \times (d_2^1 - d_1^1)} | (\mathcal{C}_{1,1}^E)'\mathcal{C}_{1,2}^E = 0_{d_1^1 \times (d_2^1 - d_1^1)}, \\ &(\mathcal{C}_{1,2}^E)'\mathcal{C}_{1,2}^E = I_{d_2^1 - d_1^1}\}. \end{aligned} \tag{1.13}$$

The parameterization of this set again uses real Givens rotations. For $\mathcal{C} \in O_{s,d_2^1 - d_1^1}(\mathcal{C}_{1,1}^E)$ it follows that $R_L(\boldsymbol{\theta}_L)\mathcal{C} = [0'_{d_1^1 \times (d_2^1 - d_1^1)}, \tilde{\mathcal{C}}']'$ for a matrix $\tilde{\mathcal{C}}$ such that $\tilde{\mathcal{C}}'\tilde{\mathcal{C}} = I_{d_2^1 - d_1^1}$ with $R_L(\boldsymbol{\theta}_L)$ corresponding to $\mathcal{C}_{1,1}^E$. The matrix $\tilde{\mathcal{C}}$ is parameterized as discussed in Lemma 1.1.

Corollary 1.1 (Properties of the parameterization of $O_{s,d_2^1 - d_1^1}(\mathcal{C}_{1,1}^E)$) *Define for* $d_1^1 < d_2^1 \leq s$ *a mapping* $\tilde{\boldsymbol{\theta}} \to C_{O,d_2^1 - d_1^1}(\tilde{\boldsymbol{\theta}}; C_{1,1}^E)$ *from* $\Theta_{O,d_2^1}^{\mathbb{R}} := [0, 2\pi)^{(d_2^1 - d_1^1)(s - d_2^1)} \times [0, 2\pi)^{(d_2^1 - d_1^1)(d_2^1 - d_1^1 - 1)/2}$ *to* $O_{s,d_2^1 - d_1^1}(\mathcal{C}_{1,1}^E)$ *by*

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

where $\boldsymbol{\theta}_L$ *denotes the parameter values corresponding to* $[\boldsymbol{\theta}_L', \boldsymbol{\theta}_R']' = C_O^{-1}(\mathcal{C}_{1,1}^E)$ *as defined in Lemma 1.1. The following properties hold:*

(i) $O_{s,d_2^1 - d_1^1}(\mathcal{C}_{1,1}^E)$ *is closed and bounded.*

(ii) The mapping $C_{O,d_2^1 - d_1^1}(\tilde{\boldsymbol{\theta}}; \mathcal{C}_{1,1}^E)$ *is infinitely often differentiable.*

For $d_2^1 < s$, *it holds*

(iii) *For every* $\mathcal{C}^E_{1,2} \in O_{s,d^1_2-d^1_1}(\mathcal{C}^E_{1,1})$ *there exists a vector* $\tilde{\boldsymbol{\theta}} = [\tilde{\boldsymbol{\theta}}'_L, \tilde{\boldsymbol{\theta}}'_R]' \in \Theta^{\mathbb{R}}_{O,d^1_2-d^1_1}$ *such that*

$$\mathcal{C}^E_{1,2} = C_{O,d^1_2-d^1_1}(\tilde{\boldsymbol{\theta}}; \mathcal{C}^E_{1,1}) = R_L(\boldsymbol{\theta}_L)' \begin{bmatrix} & 0_{d^1_1\times(d^1_2-d^1_1)} & \\ R_L(\tilde{\boldsymbol{\theta}}_L)' & I_{d^1_2-d^1_1} & R_R(\tilde{\boldsymbol{\theta}}_R) \\ & 0_{(s-d^1_2)\times(d^1_2-d^1_1)} & \end{bmatrix}.$$

The algorithm discussed above Lemma 1.1 defines the inverse mapping $C^{-1}_{O,d^1_2-d^1_1}$.

(iv) *The inverse mapping* $C^{-1}_{O,d^1_2-d^1_1}(\cdot; \mathcal{C}^E_{1,1})$ *- the parameterization of* $O_{s,d^1_2-d^1_1}(\mathcal{C}^E_{1,1})$ *- is infinitely often differentiable on the pre-image of the interior of* $\Theta^{\mathbb{R}}_{O,d^1_2-d^1_1}$. *This is an open and dense subset of* $O_{s,d^1_2-d^1_1}(\mathcal{C}^E_{1,1})$.

For $d^1_2 = s$, *it holds that*

(v) $O_{s,s-d^1_1}(\mathcal{C}^E_{1,1})$ *is a disconnected space with two disjoint non-empty closed subsets:*

$$\begin{aligned} &O^+_{s,s-d^1_1}(\mathcal{C}^E_{1,1}) := \\ &\quad \{\mathcal{C}^E_{1,2} \in \mathbb{R}^{s\times(s-d^1_1)} | (\mathcal{C}^E_{1,1})'\mathcal{C}^E_{1,2} = 0_{d^1_1\times(s-d^1_1)}, (\mathcal{C}^E_{1,2})'\mathcal{C}^E_{1,2} = I_{s-d^1_1}, \\ &\quad \det([\mathcal{C}^E_{1,1}, \mathcal{C}^E_{1,2}]) = 1\}, \\ &O^-_{s,s-d^1_1}(\mathcal{C}^E_{1,1}) := \\ &\quad \{\mathcal{C}^E_{1,2} \in \mathbb{R}^{s\times(s-d^1_1)} | (\mathcal{C}^E_{1,1})'\mathcal{C}^E_{1,2} = 0_{d^1_1\times(s-d^1_1)}, (\mathcal{C}^E_{1,2})'\mathcal{C}^E_{1,2} = I_{s-d^1_1}, \\ &\quad \det([\mathcal{C}^E_{1,1}, \mathcal{C}^E_{1,2}]) = -1\}. \end{aligned}$$

(vi) *For every* $O^+_{s,s-d^1_1}(\mathcal{C}^E_{1,1})$ *there exists a vector* $\tilde{\boldsymbol{\theta}} \in \Theta^{\mathbb{R}}_{O,d^1_2-d^1_1}$ *such that*

$$\mathcal{C}^E_{1,2} = C_{O,s-d^1_1}(\tilde{\boldsymbol{\theta}}; \mathcal{C}^E_{1,1}) = R_R(\tilde{\boldsymbol{\theta}}_R).$$

The first four steps of the algorithm discussed above Lemma 1.1 define the inverse mapping $C^{-1}_{O,s-d^1_1}(\cdot; \mathcal{C}^E_{1,1}) : O^+_{s,s-d^1_1}(\mathcal{C}^E_{1,1}) \to \Theta^{\mathbb{R}}_{O,s-d^1_1}$.

(vii) *Define* $v := [\pi, \ldots, \pi]' \in \mathbb{R}^{(s-d^1_1)(s-d^1_1-1)/2}$. *Then, a parameterization of* $O_{s,s-d^1_1}(\mathcal{C}^E_{1,1})$ *is given by*

$$C^{\pm}_{O,s-d^1_1}(\mathcal{C}^E_{1,2}; \mathcal{C}^E_{1,1}) \quad = \quad \begin{cases} v + C^{-1}_{O,s-d^1_1}(\mathcal{C}^E_{1,2}; \mathcal{C}^E_{1,1}) \\ \textit{if } C \in O^+_{s,s-d^1_1}(\mathcal{C}^E_{1,1}) \\ -(v + C^{-1}_{O,s-d^1_1}(\mathcal{C}^E_{1,2} I^-_{s-d^1_1}; \mathcal{C}^E_{1,1})) \\ \textit{if } C \in O^-_{s,s-d^1_1}(\mathcal{C}^E_{1,1}) \end{cases}$$

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 1.1 uses the same arguments as the proof of Lemma 1.1 and is therefore omitted. It remains to provide a parameterization for $\mathcal{C}^G_{1,2}$ restricted to be orthogonal to both $\mathcal{C}^E_{1,1}$ and $\mathcal{C}^E_{1,2}$. Thus, the set to be parametrized is given by

$$\begin{aligned} &O_{s,G}(\mathcal{C}^E_{1,1}, \mathcal{C}^E_{1,2}) := \\ &\quad \{\mathcal{C}^G_{1,2} \in \mathbb{R}^{s\times d^1_1} | (\mathcal{C}^E_{1,1})'\mathcal{C}^G_{1,2} = 0_{d^1_1\times d^1_1}, (\mathcal{C}^E_{1,2})'\mathcal{C}^G_{1,2} = 0_{(d^1_2-d^1_1)\times d^1_1}\}. \end{aligned}$$

The parameterization of $O_{s,G}(\mathcal{C}^E_{1,1}, \mathcal{C}^E_{1,2})$ is straightforward: Left multiplication of $\mathcal{C}^G_{1,2}$ with $R_L(\boldsymbol{\theta}_L)$ as defined in Lemma 1.1 and of the lower $(s-d^1_1)\times d^1_1$- block with $R_L(\tilde{\boldsymbol{\theta}}_L)$ as defined in Corollary 1.1

transforms the upper $d_2^1 \times d_1^1$-block to zero and collects the free parameters in the lower $(s-d_2^1)\times d_1^1$-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 1.2.

Corollary 1.2 (Properties of the parameterization of $O_{s,G}(\mathcal{C}_{1,1}^E, \mathcal{C}_{1,2}^E)$) *Define a mapping $\boldsymbol{\lambda} \to C_{O,G}(\boldsymbol{\lambda}; \mathcal{C}_{1,1}^E, \mathcal{C}_{1,2}^E)$ from $\mathbb{R}^{d_1^1(s-d_2^1)} \to O_{s,G}(\mathcal{C}_{1,1}^E, \mathcal{C}_{1,2}^E)$ for given matrices $\mathcal{C}_{1,1}^E \in O_{s,d_1^1}$ and $\mathcal{C}_{1,2}^E \in O_{s,d_2^1-d_1^1}(\mathcal{C}_{1,1}^E)$ by*

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

where $\boldsymbol{\theta}_L$ denotes the parameter values corresponding to $[\boldsymbol{\theta}_L', \boldsymbol{\theta}_R']' = C_O^{-1}(\mathcal{C}_{1,1}^E)$ as defined in Lemma 1.1 and $\tilde{\boldsymbol{\theta}}_L$ denotes the parameter values corresponding to the parameter vector $[\tilde{\boldsymbol{\theta}}_L', \tilde{\boldsymbol{\theta}}_R']' = C_{O,d_2^1-d_1^1}^{-1}(\mathcal{C}_{1,2}^E; \mathcal{C}_{1,1}^E)$ as defined in Corollary 1.1. The set $O_{s,G}(\mathcal{C}_{1,1}^E, \mathcal{C}_{1,2}^E)$ is closed and both $C_{O,G}$ as well as $C_{O,G}^{-1}(\cdot)$ - the parameterization of $O_{s,G}(\mathcal{C}_{1,1}^E, \mathcal{C}_{1,2}^E)$ - 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 $\alpha_\bullet$ 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:

- $\boldsymbol{\theta}_{B,f} := \boldsymbol{\theta}_{B,f,1} \in \Theta_{B,f} = \mathbb{R}^{d_{B,f}}$: The vector $\boldsymbol{\theta}_{B,f,1}$ 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.

- $\boldsymbol{\theta}_{B,p} := \boldsymbol{\theta}_{B,p,1} \in \Theta_{B,p} = \mathbb{R}_+^{d_{B,p}}$: The vector $\boldsymbol{\theta}_{B,p,1} := \left[b^1_{d^1-d^1_{h_1}+1,p^1_{d^1-d^1_{h_1}+1}}, \ldots, b^1_{d_1^1,p_{d_1^1}^1} \right]'$ contains the entries in $\mathcal{B}_1$ restricted by the p.u.t. structures to be positive reals.

- $\boldsymbol{\theta}_{C,E} := [\boldsymbol{\theta}_{C,E,1,1}', \boldsymbol{\theta}_{C,E,1,2}']' \in \Theta_{C,E} \subset \mathbb{R}^{d_{C,E}}$: The parameters for the matrices $\mathcal{C}_{1,1}^E$ as in the MFI(1) case and $\mathcal{C}_{1,2}^E$ as discussed in Corollary 1.1.

- $\boldsymbol{\theta}_{C,G} \in \Theta_{C,G} = \mathbb{R}^{d_{C,G}}$: The parameters for the matrix $\mathcal{C}_{1,2}^G$ as discussed in Corollary 1.2.

- $\boldsymbol{\theta}_\bullet \in \Theta_{\bullet,\alpha} \subset \mathbb{R}^{d_\bullet}$: The parameters for the stable subsystem in echelon canonical form for Kronecker indices $\alpha_\bullet$.

Example 1.8 *Consider an I(2) process with $\Omega_S = ((0,1,2))$, $p = [0,1,1]'$, $n_\bullet = 0$ and system matrices*

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

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

$$R_{3,1,2}\left(\frac{7\pi}{4}\right) R_{3,1,3}\left(\frac{\pi}{2}\right) \mathcal{C}^E_{1,1} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}',$$

$$R_{3,1,2}\left(\frac{7\pi}{4}\right) R_{3,1,3}\left(\frac{\pi}{2}\right) \mathcal{C}^E_{1,2} = \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}' \quad \textit{and} \quad R_{2,1,2}\left(\frac{7\pi}{4}\right) \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

$$R_{3,1,2}\left(\frac{7\pi}{4}\right) R_{3,1,3}\left(\frac{\pi}{2}\right) \mathcal{C}^G_{1,2} = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}' \quad \textit{and} \quad R_{2,1,2}\left(\frac{7\pi}{4}\right) \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \sqrt{2} \end{bmatrix},$$

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

1.3.3 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 $\mathcal{C}_u$ and its blocks $\mathcal{C}_k$ require more sophisticated parameterizations of parts of unitary or orthonormal matrices as well as of orthogonal complements. These are dealt with in Lemmas 1.1 and 1.2 and Corollaries 1.1 and 1.2 above. The extension of Corollaries 1.1 and 1.2 to complex matrices and to matrices, which are orthogonal to a larger number of blocks of $\mathcal{C}_k$, 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 1.8 *(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* $K_{j,i} \overset{i\to\infty}{\to} K_{j,0}$.

(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* $d((A_1, B_1, C_1), (A_2, B_2, C_2)) := \|A_1 - A_2\|_{Fr} + \|B_1 - B_2\|_{Fr} + \|C_1 - C_2\|_{Fr}$.

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 $j \to \infty$ and hence, the concept can also be used for unstable systems.

Theorem 1.2 *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* $\alpha_\bullet = (\alpha_{\bullet,1}, \ldots, \alpha_{\bullet,s}), \alpha_{\bullet,i} \in \mathbb{N}_0$.
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* $\phi_\Gamma : \Delta_\Gamma \to \Theta_\Gamma$, *such that for every* $(\mathcal{A}, \mathcal{B}, \mathcal{C}) \in \Delta_\Gamma$ *the parameter vector* $\boldsymbol{\theta} := \phi_\Gamma(\mathcal{A}, \mathcal{B}, \mathcal{C}) \in \Theta_\Gamma$ *is composed of:*

- *The parameter vector* $\boldsymbol{\theta}_{B,f} = [\boldsymbol{\theta}'_{B,f,1}, ..., \boldsymbol{\theta}'_{B,f,l}]' \in \Theta_{B,f} = \mathbb{R}^{d_{B,f}}$*, collecting the (real and imaginary parts of) non-restricted entries in* $\mathcal{B}_k, k = 1, \ldots, l$ *as described in the MFI(1) case.*

- *The parameter vector* $\boldsymbol{\theta}_{B,p} = [\boldsymbol{\theta}'_{B,p,1}, ..., \boldsymbol{\theta}'_{B,p,l}]' \in \Theta_{B,p} = \mathbb{R}_+^{d_{B,p}}$*, collecting the entries in* $\mathcal{B}_k, k = 1, \ldots, l$*, 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* $\boldsymbol{\theta}_{C,E} = [\boldsymbol{\theta}'_{C,E,1}, ..., \boldsymbol{\theta}'_{C,E,l}]' \in \Theta_{C,E} \subset \mathbb{R}^{d_{C,E}}$*, consisting of the vectors* $\boldsymbol{\theta}_{C,E,k} = [\boldsymbol{\theta}'_{C,E,k,1}, \ldots, \boldsymbol{\theta}'_{C,E,k,h_k}]'$ *collecting the parameters* $\boldsymbol{\theta}_{C,E,k,j}$ *for all blocks* $\mathcal{C}^E_{k,j}$*,* $k = 1, \ldots, l$ *and* $j = 1, \ldots, h_k$*, obtained using Givens rotations (see Lemmas 1.1 and 1.2 and Corollary 1.1 and its extension to complex matrices).*

- *The parameter vector* $\boldsymbol{\theta}_{C,G} = [\boldsymbol{\theta}'_{C,G,1}, ..., \boldsymbol{\theta}'_{C,G,l}]' \in \Theta_{C,G} = \mathbb{R}^{d_{C,G}}$*, consisting of the vectors* $\boldsymbol{\theta}_{C,G,k} = [\boldsymbol{\theta}'_{C,G,k,2}, \ldots, \boldsymbol{\theta}'_{C,G,k,h_k}]'$ *collecting the parameters* $\boldsymbol{\theta}_{C,G,k,j}$ *(real and imaginary parts for complex roots) for* $\mathcal{C}^G_{k,j}$*,* $k = 1, \ldots, l$ *and* $j = 2, \ldots, h_k$*, subject to the orthogonality restrictions (see Corollary 1.2 and its extension to complex matrices).*

- *The parameter vector* $\boldsymbol{\theta}_\bullet \in \Theta_\bullet \subset \mathbb{R}^{d_\bullet}$ *collecting the free entries in echelon canonical form with Kronecker indices* $\alpha_\bullet$*.*

(i) The mapping $\psi_\Gamma : M_\Gamma \to \Delta_\Gamma$ *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 $\boldsymbol{\theta} = [\boldsymbol{\theta}'_{B,f}, \boldsymbol{\theta}'_{B,p}, \boldsymbol{\theta}'_{C,E}, \boldsymbol{\theta}'_{C,G}, \boldsymbol{\theta}'_\bullet]' \in \Theta_\Gamma \subset \Theta_{B,f} \times \Theta_{B,p} \times \Theta_{C,E} \times \Theta_{C,G} \times \Theta_\bullet$ *corresponds to a triple* $(\mathcal{A}(\boldsymbol{\theta}), \mathcal{B}(\boldsymbol{\theta}), \mathcal{C}(\boldsymbol{\theta})) \in \Delta_\Gamma$ *and a transfer function* $k(z) = \pi(\mathcal{A}(\boldsymbol{\theta}), \mathcal{B}(\boldsymbol{\theta}), \mathcal{C}(\boldsymbol{\theta})) \in M_\Gamma$*. The mapping* $\phi_\Gamma^{-1} : \boldsymbol{\theta} \to (\mathcal{A}(\boldsymbol{\theta}), \mathcal{B}(\boldsymbol{\theta}), \mathcal{C}(\boldsymbol{\theta}))$ *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 1.2, the parameterization of Φ is straightforward. The $s \times m$ entries of Φ are collected in a parameter vector $\boldsymbol{d}$. 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 $\boldsymbol{\tau} = [\boldsymbol{\theta}', \boldsymbol{d}']' \in \Theta_\Gamma \times \mathbb{R}^{sm}$. 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.

1.4 The Topological Structure

The parameterization of M_n in Theorem 1.2 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 the 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 1.2. 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. Note that Δ_n also contains non-minimal state space realizations, corresponding to transfer functions of lower order.

Remark 1.16 *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 $\overline{M_\Gamma}$. Note that $\overline{M_\Gamma}$ is not necessarily equal to $\pi(\overline{\Delta_\Gamma})$. The continuity of π, as shown in Theorem 1.2 (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 closures $\overline{\Delta}_\Gamma$ in Δ_n and finally we examine the closures $\overline{M}_\Gamma$ in M.

Definition 1.9 *(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*

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

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

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

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

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

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 $i = 1, \ldots, \tilde{n}_u$. Moreover, $(\tilde{\Omega}_S, \tilde{p}) < (\Omega_S, p)$ holds if additionally $(\tilde{\Omega}_S, \tilde{p}) \neq (\Omega_S, p)$.

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

Finally, define

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

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 $\tilde{\Omega} \preceq \Omega$, 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.

1.4.1 The Closure of Δ_Γ in Δ_n

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 $\alpha_\bullet$. Examining $\overline{\Delta_{\Omega_S,p}}$ and $\overline{\Delta_{\alpha_\bullet}}$ separately simplifies the analysis.

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 $\Delta_{\Omega_S,p} = \Delta^{\mathcal{A}}_{\Omega_S,p} \times \Delta^{\mathcal{B}}_{\Omega_S,p} \times \Delta^{\mathcal{C}}_{\Omega_S,p}$ and the closures of the three matrices can be analyzed separately. $\Delta^{\mathcal{A}}_{\Omega_S,p}$ and $\Delta^{\mathcal{C}}_{\Omega_S,p}$ are very easy to investigate. The structure of $\mathcal{A}$ is fully determined by Ω_S and consequently $\Delta^{\mathcal{A}}_{\Omega_S,p}$ consists of a single matrix $\mathcal{A}$, which immediately implies that $\overline{\Delta^{\mathcal{A}}_{\Omega_S,p}} = \Delta^{\mathcal{A}}_{\Omega_S,p}$. The matrix $\mathcal{C}$, compare Theorem 1.1 is composed of blocks $\mathcal{C}^E_k$ that are sub-blocks of unitary (or orthonormal) matrices and blocks $\mathcal{C}^G_k$ that have to fulfill (recursive) orthogonality constraints. The corresponding sets have been shown to be closed in Lemmas 1.1 and 1.2 and Corollaries 1.1 and 1.2. Thus, $\overline{\Delta^{\mathcal{C}}_{\Omega_S,p}} = \Delta^{\mathcal{C}}_{\Omega_S,p}$.

It remains to discuss $\overline{\Delta^{\mathcal{B}}_{\Omega_S,p}}$. 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 subset 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 1.4 with equation (1.12) 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 $b_{1,2,1,2} \to 0$. The limiting system is then given by

$$\begin{aligned} y_t &= \mathcal{C}^E_{1,1}x_{t,1} + \mathcal{C}^G_{1,2}x_{t,2} + \mathcal{C}^E_{1,2}x_{t,3} + \varepsilon_t, \\ \begin{bmatrix} x_{t+1,1} \\ x_{t+1,2} \\ x_{t+1,3} \end{bmatrix} &= \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t,1} \\ x_{t,2} \\ x_{t,3} \end{bmatrix} + \begin{bmatrix} b_{1,1,1} & b_{1,1,2} \\ 0 & 0 \\ b_{1,2,2,1} & b_{1,2,2,2} \end{bmatrix} \varepsilon_t, \qquad x_{1,1} = x_{1,2} = x_{1,3} = 0. \end{aligned}$$

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 realization of the limiting system $\{y_t\}_{t\in\mathbb{Z}}$ given by

$$\begin{aligned} y_t &= \mathcal{C}^E_{1,1}x_{t,1} + \mathcal{C}^E_{1,2}x_{t,3} + \varepsilon_t = \tilde{C}\tilde{x}_t + \varepsilon_t, \qquad \tilde{x}_t \in \mathbb{R}^2, \\ \tilde{x}_{t+1} = \begin{bmatrix} x_{t+1,1} \\ x_{t+1,3} \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t,1} \\ x_{t,3} \end{bmatrix} + \begin{bmatrix} b_{1,1,1} & b_{1,1,2} \\ b_{1,2,2,1} & b_{1,2,2,2} \end{bmatrix} \varepsilon_t = \tilde{x}_t + \tilde{B}\varepsilon_t, \qquad x_{1,1} = x_{1,3} = 0. \end{aligned}$$

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 1.12. The limiting system - in case of full rank of $\tilde{B}$ - has indices $\tilde{p} = [1, 2]'$. To relate to Definition 1.9 choose the permutation matrix $S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ to arrive at

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

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 $\Delta^{\mathcal{B}}_{\Omega_S,p}$ is related to lower order systems in the sense of Definition 1.9. The precise statement is given in Theorem 1.3 after a discussion of the closure of the stable subsystems.

The Closure of $\Delta_{\alpha_\bullet}$

Consider a convergent sequence of systems $\{(\mathcal{A}_j, \mathcal{B}_j, \mathcal{C}_j)\}_{j\in\mathbb{N}}$ in $\Delta_{\alpha_\bullet}$ 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 $\alpha_\bullet$, 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 $\tilde{A} = \operatorname{diag}(\tilde{J}_2, \tilde{A}_\bullet)$, 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, Chapter 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 1.9) 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 $\alpha_\bullet$, compare again Hannan and Deistler (1988, Chapter 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.

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 1.10 (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*

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

- $\tilde{\alpha}_\bullet \leq \alpha_\bullet$.

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

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 1.3 *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 $M_{\check{\Gamma}}$ for $\check{\Gamma} \leq \tilde{\Gamma}$ with $\tilde{\Gamma}$ conformable to Γ, i. e.,*

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

Theorem 1.3 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 1.3 is potentially strict, depending on the Kronecker indices $\alpha_\bullet$ in Γ. Equality holds, e. g., in the following case:

Corollary 1.3 *For every multi-index Γ with $n_\bullet = 0$ the set of conformable indices consists only of Γ, which implies $\pi(\overline{\Delta_\Gamma}) = \bigcup_{\tilde{\Gamma} \leq \Gamma} M_{\tilde{\Gamma}}$.*

1.4.2 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

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

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 $p = [1, \ldots, n_{k,h_k,j}]$. 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 1.3 implies the existence of a generic neighborhood for the canonical form considered in this paper:

Theorem 1.4 *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*

$$M(\Omega_S, n_\bullet) \quad \subset \quad \overline{M_{\Gamma_g}}. \tag{1.14}$$

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 1.4 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 1.9 (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 their limits $(A_0, B_0, C_0) := \lim_{j\to\infty}(A_j, B_j, C_j)$ may differ in three ways:

- For sequences $(\mathcal{A}_j, \mathcal{B}_j, \mathcal{C}_j)_{j\in\mathbb{N}}$ in canonical form rows of $\mathcal{B}_{u,j}$ can tend to zero, which reduces the state space unit root structure as discussed in Section 1.4.1.

- 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 $\mathcal{A}_j = T_j A_j T_j^{-1}$ may be converging to zeros in the sub-block $\mathcal{A}_{u,0}$ of the limit $\mathcal{A}_0 = T_0 A_0 T_0^{-1}$ 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 1.9 (ii). The implications of the other two cases are summarized in the following definition:

Definition 1.11 (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*

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

where $\check{\mathcal{A}}_u$ 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 1.17 *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 1.5 *(i) M_Γ is T_{pt}-open in $\overline{M_\Gamma}$.*
(ii) For every generic multi-index Γ_g corresponding to Ω_S and $n_\bullet$ it holds that

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

Theorem 1.5 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 1.5 also implies the following result relevant for the determination of the unit root structure, further discussed in Sections 1.5.1 and 1.5.2:

Corollary 1.4 *For every pair $(\tilde{\Omega}_S, \tilde{n}_\bullet) \in \mathcal{A}(\Omega_S, n_\bullet)$ it holds that*

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

1.5 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. Similarly to the VECM framework, as discussed for the I(2) case in Section 1.2, testing

hypotheses on the cointegrating spaces or adjustment coefficients may necessitate different reparameterizations.

1.5.1 The $MFI(1)$ Case

The two by far most widely used cases of MFI(1) processes are $I(1)$ processes and seasonally 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 $x_{1,u} = 0$

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

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 directly seen from considering the real valued canonical form discussed in Remark 1.4, with the matrices $\mathcal{A}_{k,\mathbb{R}}$ for $k = 2, \ldots, l-1$, given by $\mathcal{A}_{k,\mathbb{R}} = I_{d_1^k} \otimes \begin{pmatrix} \cos(\omega_k) & -\sin(\omega_k) \\ \sin(\omega_k) & \cos(\omega_k) \end{pmatrix}$ in this case.

The ranks of $\mathcal{C}_k \mathcal{B}_k$ are equal to the integers d_1^k in $\Omega_S = ((\omega_1, d_1^1), \ldots, (\omega_l, d_1^l))$. 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 $k = 2, \ldots, l-1$ and equal to d_1^l if $k = l$, as discussed in Section 1.3.

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 $r_k := \operatorname{rank}(-k^{-1}(z_k))$, 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 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 1.2. 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 $\mathcal{C}_k$, 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 $\mathcal{C}_k$. 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 $\mathcal{C}_k$, i. e., β_k is of full rank and $\beta_k' \mathcal{C}_k = (\mathcal{R}(\beta_k)' - i\mathcal{I}(\beta_k)')\mathcal{C}_k = 0$. This holds true, since both factors are of full rank and $[\gamma_0', \gamma_1']'$ satisfies $(\overline{z}_k \gamma_0' + \gamma_1')\mathcal{C}_k = 0$, which corresponds to the necessary condition given in Example 1.2 for the columns of $[\gamma_0', \gamma_1']'$ to be PCIVs. The latter implies $(\overline{z}_k \tilde{\gamma}_0' + \tilde{\gamma}_1')\mathcal{C}_k = 0$ 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 1.2.

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 1.6, a sufficiently rich set of deterministic components allows to absorb non-zero initial values $x_{1,u}$.

Testing Hypotheses on the State Space Unit Root Structure

Using the generic sets of transfer functions M_{Γ_g} presented in Theorem 1.4, 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 1.5 the null hypothesis includes all pairs $(\Omega_S, n_\bullet) \in \mathcal{A}(\Omega_{S,0}, n_{\bullet,0})$ as well as all pairs $(\Omega_S, n_\bullet)$ 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 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 $d_1^{\tilde{k}}$ 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 1.4 shows that

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

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

Analogously to the procedure of testing for the 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 1.4, which can be used to consecutively test hypotheses on Ω_S.

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 dif-

ferent 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) $H_0'' : \beta = [H_1\varphi_1, \ldots, H_c\varphi_c], \beta \in \mathbb{R}^{s\times r}, H_j \in \mathbb{R}^{s\times t_j}, \varphi_j \in \mathbb{R}^{t_j\times r_j}, r_j \leq t_j \leq s$, for $j = 1, \ldots, c$ 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 1.1, cointegration at $\omega_k = 0$ occurs if and only if a vector β_j satisfies $\beta_j'\mathcal{C}_1 = 0$. In other words, the column space of $\mathcal{C}_1$ is the orthocomplement of the cointegrating space spanned by the columns of β and hypotheses on β restrict entries of $\mathcal{C}_1$.

The first type of hypothesis, H_0, implies that the column space of $\mathcal{C}_1$ 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:

$$\mathcal{C}_1^r(\check{\boldsymbol{\theta}}_L, \boldsymbol{\theta}_R) \quad := \quad \left[H \cdot \check{R}_L(\check{\boldsymbol{\theta}}_L)' \begin{bmatrix} I_{t-r} \\ 0_{r\times(t-r)} \end{bmatrix}, \quad H_\perp\right] \cdot R_R(\boldsymbol{\theta}_R), \tag{1.15}$$

where $\check{R}_L(\check{\boldsymbol{\theta}}_L) := \prod_{i=1}^{t-r}\prod_{j=1}^{r} R_{t,i,t-r+j}(\theta_{L,r(i-1)+j}) \in \mathbb{R}^{t\times t}$ and $R_R(\boldsymbol{\theta}_R) \in \mathbb{R}^{(s-r)\times(s-r)}$ as in Lemma 1.1. From this one can derive a parameterization of the set of matrices $\mathcal{C}_1^r$ corresponding to H_0, analogously to Lemma 1.1. 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 $\boldsymbol{\theta}_L \in [0, 2\pi)^{r(s-r)}$ and $\check{\boldsymbol{\theta}}_L \in [0, 2\pi)^{r(t-r)}$, 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 $\mathcal{C}_1$ satisfying the restriction $b'\mathcal{C}_1 = 0$, i. e., the set $O_{s,d_1}(b)$ defined in (1.13). 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 $\tilde{t} = \tilde{t}_1 + t_b + \tilde{t}_2$. 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:

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

where $R_H(\boldsymbol{\theta}_H) := R_H\left((\boldsymbol{\theta}_{H_1}, \boldsymbol{\theta}_{H_2}, \boldsymbol{\theta}_{H_b})\right) := R_{H_1}(\boldsymbol{\theta}_{H_1}) R_{H_2}(\boldsymbol{\theta}_{H_2}) R_{H_b}(\boldsymbol{\theta}_{H_b}) \in \mathbb{R}^{\tilde{t}\times\tilde{t}}$ is a product of Givens rotations corresponding to the entries in the blocks highlighted by bold font and where $R_R(\boldsymbol{\theta}_{R,\beta}) \in \mathbb{R}^{r\times r}$ as in Lemma 1.1. The three matrices are defined as follows:

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

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

$$
\mathcal{C}_1^r(\boldsymbol{\theta}_H, \boldsymbol{\theta}_{R,\mathcal{C}}) :=
\begin{bmatrix}
\tilde{H} \cdot R_H(\boldsymbol{\theta}_H)' \begin{bmatrix}
0_{p_1\times(q_1-r_1)} & 0_{p_1\times(q_2-r_2)} & 0_{p_1\times(\tilde{t}-q_1-q_2)} \\
I_{q_1-r_1} & 0_{(q_1-r_1)\times(q_2-r_2)} & 0_{(q_1-r_1)\times(\tilde{t}-q_1-q_2)} \\
0_{p_2\times(q_1-r_1)} & 0_{p_2\times(q_2-r_2)} & 0_{p_2\times(\tilde{t}-q_1-q_2)} \\
0_{(q_2-r_2)\times(q_1-r_1)} & I_{q_2-r_2} & 0_{(q_2-r_2)\times(\tilde{t}-q_1-q_2)} \\
0_{p_b\times(q_1-r_1)} & 0_{p_b\times(q_2-r_2)} & 0_{p_b\times(\tilde{t}-q_1-q_2)} \\
0_{(\tilde{t}-q_1-q_2)\times(q_1-r_1)} & 0_{(\tilde{t}-q_1-q_2)\times(q_2-r_2)} & I_{\tilde{t}-q_1-q_2}
\end{bmatrix}, \quad \tilde{H}_\perp
\end{bmatrix} \cdot R_R(\boldsymbol{\theta}_{R,\mathcal{C}}),
$$

where $R_H(\boldsymbol{\theta}_H) \in \mathbb{R}^{\tilde{t}\times\tilde{t}}$ as above and $R_R(\boldsymbol{\theta}_{R,\mathcal{C}}) \in \mathbb{R}^{(s-r)\times(s-r)}$ as in Lemma 1.1. Note that for all $\boldsymbol{\theta}_H$, $\boldsymbol{\theta}_{R,\beta}$ and $\boldsymbol{\theta}_{R,\mathcal{C}}$ it holds that $\beta^r(\boldsymbol{\theta}_H, \boldsymbol{\theta}_{R,\beta})' \mathcal{C}_1^r(\boldsymbol{\theta}_H, \boldsymbol{\theta}_{R,\mathcal{C}}) = 0_{r\times(s-r)}$. 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) $H_0''' : \beta_\perp = \mathcal{C}_1 = [H_1\varphi_1, \ldots, H_c\varphi_c], \beta_\perp \in \mathbb{R}^{s\times(s-r)}, H_j \in O_{s,t_j}, \varphi_j \in O_{t_j,r_j}, r_j \leq t_j \leq s$, for $j = 1, \ldots, c$ such that $\sum_{j=1}^c r_j = s - r$. The orthocomplement 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 1.2. 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

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

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 $\mathcal{C}_k$ 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 $\mathcal{C}_k$ 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 (1.15), with $\check{R}_L(\check{\boldsymbol{\theta}}_L)$ and $R_R(\theta_R)$ replaced by $\check{Q}_L(\check{\boldsymbol{\varphi}}_L) := \prod_{i=1}^{t-r}\prod_{j=1}^{r} Q_{t,i,t-r+j}(\varphi_{L,r(i-1)+j}) \in \mathbb{R}^{t\times t}$ and $D_d(\boldsymbol{\varphi}_D)Q_R(\boldsymbol{\varphi}_R)$ as in Lemma 1.2.

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 $\mathcal{C}_k$. To parameterize the set of matrices restricted according to these null hypotheses, Lemma 1.2 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 $b_j \in \mathbb{R}^s$, $j = 1, \ldots, m$ 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 $\beta_2(z) = b_j z$, $j = 1, \ldots, m$. In light of the discussion above the necessary condition for these two polynomials to be PCIVs is equivalent to $b_j'\mathcal{C}_k = 0$, for $j = 1, \ldots, m$. 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 $\mathcal{C}_k$. A suitable parameterization is derived from the following mapping

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

where $\boldsymbol{\theta}_b \in [0, 2\pi)^{m(s-m)}$ and $C_U(\boldsymbol{\varphi}) := C_U(\boldsymbol{\varphi}_L, \boldsymbol{\varphi}_D, \boldsymbol{\varphi}_R) \in U_{s-m,d_1^k}$ as in Lemma 1.2. 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.

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 $\mathcal{B}_1 A\psi = 0$. 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 $\mathcal{C}_1$ and $\mathcal{B}_1$. We therefore consider the set of feasible matrices $\mathcal{B}_1'$ as a subset in $O_{s,s-r}$ and use the mapping $\mathcal{B}_1'(\check{\boldsymbol{\theta}}_L, \boldsymbol{\theta}_R) = [A\check{R}_L(\check{\boldsymbol{\theta}}_L)'[I_{t-r}, 0_{r\times(t-r)}]', A_\perp]R_R(\boldsymbol{\theta}_R)$ to derive a parameterization, while $\mathcal{C}_1'$ is restricted to be a p.u.t. matrix and the set of feasible matrices $\mathcal{C}_1'$ is parameterized accordingly.

As a second type of hypothesis Juselius (2006, Section 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 (1.13) for the set of feasible matrices $\mathcal{B}'_1$ and the parameterization of the set of p.u.t. matrices for the set of feasible matrices $\mathcal{C}'_1$, 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.

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 (1.4), 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, Section 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 $y_t = \mathcal{C}x_t + \Phi d_t + \varepsilon_t$, compare (1.9). 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:

$$
\begin{array}{lll}
1. & H(r) & : \ \Phi d_t = \mathcal{C}_1\tilde{\Phi}_2 t^2 + \Phi_1 t + \Phi_0, \\
2. & H^*(r) & : \ \Phi d_t = \Phi_1 t + \Phi_0, \\
3. & H_1(r) & : \ \Phi d_t = \mathcal{C}_1\tilde{\Phi}_1 t + \Phi_0, \\
4. & H_1^*(r) & : \ \Phi d_t = \Phi_0, \\
5. & H_2(r) & : \ \Phi d_t = \mathcal{C}_1\tilde{\Phi}_0,
\end{array}
$$

where $\Phi_0, \Phi_1 \in \mathbb{R}^s$ and $\tilde{\Phi}_0, \tilde{\Phi}_1, \tilde{\Phi}_2 \in \mathbb{R}^{d_1^1}$. The component $\mathcal{C}_1\tilde{\Phi}_0$ captures the influence of the initial value $\mathcal{C}_1 x_{1,1}$ 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 $\mathcal{C}_k\tilde{\Phi}_k z_k^t + \overline{\mathcal{C}_k\tilde{\Phi}_k(z_k)^t}$, $\tilde{\Phi}_k \in \mathbb{C}^{d_1^k}$, similarly allowing for a reinterpretation of these components as influence of the initial values $x_{1,k}$ on the output.

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.

1.5.2 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 $x_{1,u} = 0$ of the system in canonical form in this setting is given by

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

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 1.2. 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 1.3.

Again, both the integers d_1^1 and d_2^1 and the ranks r, 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 $\mathcal{C}_{1,1}^E$ 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 $\mathcal{C}_{1,2}^G$, since this space is the orthocomplement of the column space of $[\mathcal{C}_{1,1}^E, \mathcal{C}_{1,2}^G, \mathcal{C}_{1,2}^E]$. 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 1.3.

Consider the matrices β, β_1 and β_2 as defined in Section 1.2. 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 $\mathcal{C}_{1,1}^E$. The same relation holds true for β_1 and the sub-block $\mathcal{C}_{1,2}^E$. With respect to polynomial cointegration, Bauer and Wagner (2012) show that the rank of $\mathcal{C}_{1,2}^G$ determines the number of minimum degree polynomial cointegrating relations, as discussed in Example 1.3. If $\mathcal{C}_{1,2}^G = 0$, 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 $\{\beta' y_t\}_{t\in\mathbb{Z}}$ 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 $x_{1,u}$: The solution to the state space equations for $t > 0$ and $x_{1,u} \neq 0$ is given by:

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

Hence, if $\Phi d_t = \Phi_0 + \Phi_1 t$, the output equation contains the terms $\mathcal{C}_{1,1}^E x_{1,1}^E + \mathcal{C}_{1,2}^G x_{1,2}^G + \mathcal{C}_{1,2}^E x_{1,2}^E - \mathcal{C}_{1,1}^E x_{1,2}^G + \Phi_0$ and $(\mathcal{C}_{1,1}^E x_{1,2}^G + \Phi_1)t$. Again, this implies non-identifiability, which is resolved by assuming $x_{1,u} = 0$, compare Remark 1.6.

Testing Hypotheses on the State Space Unit Root Structure

To simplify notation we use

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

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

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

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.

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.

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 $[\mathcal{C}_{1,1}^E, \mathcal{C}_{1,2}^E]$ 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 $\varphi' K' [\mathcal{C}_{1,1}^E, \mathcal{C}_{1,2}^E] = 0$. 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 (1.15), use the mapping

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

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

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 $b' [\mathcal{C}_{1,1}^E, \mathcal{C}_{1,2}^E] = 0$. Assume w.l.o.g. $b \in O_{s,t}$ and parameterize the set of feasible matrices $\mathcal{C}_{1,1}^E$ using $O_{s,d_1^1}(b)$ as defined in (1.13) and the set of feasible matrices $\mathcal{C}_{1,2}^E$ using $O_{s,d_2^1 - d_1^1}([b, C_{1,1}^E])$. Alternatively, parameterize the set of feasible matrices jointly as elements $[\mathcal{C}_{1,1}^E, \mathcal{C}_{1,2}^E] \in O_{s,s-r}(b)$.

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., $\mathcal{C}^E_{1,1}$. 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 $[\mathcal{C}^E_{1,1}, \mathcal{C}^E_{1,2}]$ by $\mathcal{C}^E_{1,1}$.

The hypothesis of no minimum degree polynomial cointegrating relations implies the restriction $\mathcal{C}^G_{1,2} = 0$, compare Example 1.3. Therefore, we can test all hypotheses considered in Johansen (2006) also in our more general setting.

Testing Hypotheses on the Adjustment Coefficients

Hypotheses on α and ξ as defined in (1.6) and (1.7) correspond to hypotheses on the spaces spanned by the rows of $\mathcal{B}_{1,2,1}$ and $\mathcal{B}_{1,2,2}$. For VAR processes integrated of order two, the row space of $\mathcal{B}_{1,2,1}$ is equal to the orthogonal complement of the column space of $[\alpha, \alpha_\perp \xi]$, while the row space of $\mathcal{B}_{1,2} := [\mathcal{B}'_{1,2,1}, \mathcal{B}'_{1,2,2}]'$ is equal to the orthogonal complement of the column space of α. The restrictions corresponding to hypotheses on α and ξ can be implemented analogously to the restrictions corresponding to hypotheses on α_1 in Section 1.5.1, reversing the roles of the relevant sub-blocks in B_u and $\mathcal{C}_u$ accordingly.

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, Kongsted and Jorgensen (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 (1.9). Choose, e. g., $\Phi d_t = \Phi_0 + \Phi_1 t$, as in the I(1) case. In analogy to Section 1.5.1, 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

$$\Phi_0 \quad = \quad [\mathcal{C}^E_{1,1}, \mathcal{C}^E_{1,2}]\phi_0 + \tilde{\mathcal{C}}_{1,2}\tilde{\phi}_0 + \mathcal{C}_\perp \check{\phi}_0,$$

where the columns of $\tilde{\mathcal{C}}_{1,2}$ are a basis for the column space of $\mathcal{C}^G_{1,2}$, which does not necessarily have full column rank, and the columns of $\mathcal{C}_\perp$ span the orthocomplement of the column space of $[\mathcal{C}^E_{1,1}, \mathcal{C}^E_{1,2}, \tilde{\mathcal{C}}_{1,2}]$. 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.

1.6 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 to subsequently develop statistical inference procedures. Bauer and Wagner (2012) provide 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 upon 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 developed for or 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, $\overline{M}_\Gamma$ 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, utilizing 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 1.5 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 upon

them for determining cointegrating ranks that allow to perform 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.

Pseudo Maximum Likelihood Parameter Estimation for Multiple Frequency I(1) Processes: A State Space Approach

2.1 Introduction

Time series in many different scientific disciplines show trending behavior and additionally seasonality that is not adequately modeled using deterministic terms such as sinus and cosinus terms of different frequencies. Often such time series are analyzed jointly as they show co-trending of different sort at some or all frequencies.
The predominant approach to modeling such multivariate time series is the vector autoregressive (VAR) framework of Johansen and coworkers, see, e. g., Johansen and Schaumburg (1999) or Johansen (1995). The VAR framework uses the vector error correction (VECM) representation for estimation and inference which is well studied for (seasonally) integrated processes.
It is well known that the more flexible vector autoregressive moving average (VARMA) model is better suited than the VAR framework for modeling multivariate time series for a number of reasons: First, VARMA processes are closed under marginalization such that the vector of a subset of variables of an VARMA process also is a VARMA process, whereas subsets of VAR processes also are VARMA processes. Second, temporally or spatially aggregating VARMA processes again leads to a VARMA process. Third, the solutions to dynamic stochastic general equilibrium models, that are commonly used in economics, are VARMA rather than VAR processes. And finally, the VARMA framework which in a certain sense is equivalent to the state space framework, arguably is more parsimonious for high dimensional time series than the VAR framework.
Therefore, this paper deals with the estimation of VARMA models for seasonally integrated processes in the state space framework. Recently Bauer and Wagner (2012) provided a canonical form for such models in state space representation and Bauer, Matuschek, de Matos Ribeiro and Wagner (2020) discuss a specific parameterization for such models.
While the literature on estimation and inference for (seasonally) integrated processes for the VAR framework is well developed, the same is not true for the VARMA framework. Early exceptions are Aoki (1990) and Aoki and Havenner (1997) who focus on processes integrated at frequency 0 (that is, showing a unit root at $z = 1$) but not at other seasonal frequencies. The same is true for Yap and Reinsel (1995) who provide estimation algorithms and asymptotic distributions for the corresponding estimates for I(1) VARMA processes. Another contribution in this respect is given by Ribarits and Hanzon (2014) who make extensive use of a VECM representation for state space models. Lütkepohl and Claessen (1997) and Poskitt (2006) provide similar developments for the VARMA framework based on echelon forms.

For seasonally integrated processes theory almost exclusively focuses on VAR processes: Johansen and Schaumburg (1999) extend the VECM representation from the I(1) to the seasonally integrated case of order one (in the following called *multiple frequency I(1)*, see Definition 2.1 for an explicit definition). Cubadda (2001) and Cubadda and Omtzigt (2005) further develop the estimation algorithms as well as the theory. Further results for more general processes for the VAR case can be found in Lee (1992) and Gregoir (1999).
In the VARMA case Bauer and Buschmeier (2016) discuss estimation for MFI(1) processes using so called subspace algorithms. They show that these algorithms can be used under the assumption of known state space order to obtain consistent estimators for the system for stationary as well as seasonally integrated processes, but do not provide results on the limiting distribution. These estimators thus can serve as consistent initial guesses for subsequent pseudo maximum likelihood estimation which numerically can be performed using the parameterization provided in Bauer et al. (2020).
Based on these preliminary results this paper discusses the asymptotic properties of pseudo maximum likelihood (PML) estimators obtained by maximizing the Gaussian pseudo likelihood function. In particular the contributions of the paper are:

(i) We discuss in depth two different versions of the PML estimator depending on the treatment of initial values. Taking initial values equal to zero leads to so called *prediction error* estimation. Alternatively, initial values corresponding to the stationary distributions for the stationary part of the state can be used. We show that the two different versions are related but different.

(ii) We formulate sufficient assumptions for the consistency of the estimators. This is done in a coordinate-free fashion, drawing heavily from Hannan and Deistler (1988, Chapter 4). Hereby, the pseudo likelihood function is assumed to be maximized over a set of transfer functions M_Γ such that the data generating transfer function k_0 is inside the closure $\overline{M_\Gamma}$ of this set. Here M_Γ denotes the set of all transfer functions with multi-index Γ, defined in Bauer et al. (2020), see also the explanation above Assumption 2.1 below. We show that the corresponding estimator in this situation is consistent without requiring compactness of the set M_Γ or the corresponding parameter set. The restriction $k_0 \in \overline{M_\Gamma}$ requires that the state space order is not underestimated. But it does not impose any assumption of correctly specified integration or cointegration properties such as the cointegrating rank.

(iii) We derive the asymptotic distribution of the estimated parameter vector in the situation of correct specification of the multi-index Γ. As usual, it is shown that all parameters corresponding to cointegrating spaces are estimated super-consistently with mixed normal distributions, while the remaining parameters corresponding, e. g., to short term dynamics are asymptotically normally distributed with the standard rate square root sample size. Our results cover many different cases with respect to the in- or exclusion of deterministic terms.

(iv) We demonstrate in a simulation study that our approach compares favorably to the subspace estimators of Bauer and Buschmeier (2016) as well as VAR approximations in some situations. The superiority in our simulations increases – as expected – with zeros tending to the unit circle.

Consequently, the paper clarifies the asymptotic properties of PML estimators for general MFI(1) processes. Based on the results a number of specification tests can be performed with more results in this respect being derived in the companion paper Matuschek et al. (2020).

This paper is structured as follows. In Section 2.2 we describe the data generating processes considered in this paper and the parameter space. In Section 2.3 we show the consistency of the pseudo maximum likelihood estimator and derive the asymptotic distribution of the parameters. In Section 2.4 we compare the estimation quality of the cointegrating space and the prediction performance of the pseudo maximum likelihood, the Johansen Schaumburg, cf. Johansen and Schaumburg (1999), and the CCA subspace algorithm, cf. Bauer and Buschmeier (2016), in a small simulation study. Section 2.5 summarizes and concludes the paper. All proofs are relegated

to the appendix.
Notation in this paper is as follows: I_p denotes the p-dimensional identity matrix, $0^{m\times n}$ the m times n zero matrix. For a square matrix X we denote the spectral radius (i. e., the maximum of the modulus of its eigenvalues) by $\lambda_{|\max|}(X)$. We denote the smallest eigenvalue of a symmetric matrix X by $\lambda_{\min}(X)$. For a $m \times n$ matrix X with $n < m$ of full rank $X_\perp$ denotes a $m \times (m-n)$ matrix of full rank such that $X'X_\perp = 0^{n\times(m-n)}$ and $X'_\perp X_\perp = I_{m-n}$. L denotes the backshift operator such that $L(\{y_t\}_{t\in\mathbb{Z}}) = \{y_{t-1}\}_{t\in\mathbb{Z}}$ for a process $\{y_t\}_{t\in\mathbb{Z}}$. The Kronecker delta $\delta_{i,j}$ is one if $i = j$ and zero if $i \neq j$. For two matrices $A \in \mathbb{C}^{m\times n}$ and $B \in \mathbb{C}^{k\times l}$, $A \otimes B \in \mathbb{C}^{(mk)\times(nl)}$ denotes their Kronecker product. For $c \in \mathbb{C}$, $\mathcal{R}(c)$ denotes its real part and $\mathcal{I}(c)$ its imaginary part. For a matrix or vector X its complex conjugate is denoted by $\overline{X}$, its complex conjugate transpose by X'. For a set Θ, $\bar{\Theta}$ denotes the closure and Θ^c denotes the complement (within a suitable ambient space). Convergence in distribution is denoted by $\xrightarrow{d}$. For $x \in \mathbb{R}^n$ and $r \in \mathbb{R}$, $B(x,r)$ denotes the open ball with center x and radius r. For finite sequences $\{a_t\}_{t=1,\dots,T}, \{b_t\}_{t=1,\dots,T}$, we define $\langle a_t, b_t\rangle := T^{-1}\sum_{t=1}^{T} a_t b'_t$ analogously to Johansen and Nielsen (2018).

2.2 Setup and Assumptions

This paper deals with the same class of VARMA processes $\{y_t\}_{t\in\mathbb{Z}}$, $y_t \in \mathbb{R}^s$ as discussed in Bauer and Wagner (2012, p. 1316-1317): We refer to a stochastic process $\{y_t\}_{t\in\mathbb{Z}}$ as a VARMA process, if there exist integers $p, q \geq 0$ and matrices $A_j \in \mathbb{R}^{s\times s}$, $j = 1,\dots,p$ and $B_j \in \mathbb{R}^{s\times s}$, $j = 1,\dots,q$, where $A_p \neq 0, B_q \neq 0$, and a real-valued white noise process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$, with $\mathbb{E}(\varepsilon_t\varepsilon'_t) = \Sigma > 0$ (detailed assumptions are stated below), such that[17]

$$y_t + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} = \varepsilon_t + B_1\varepsilon_{t-1} + \cdots + B_q\varepsilon_{t-q}, \quad t \in \mathbb{Z}. \tag{2.1}$$

Defining the matrix polynomials $a(z) := I_s + A_1 z + \cdots + A_p z^p$ and $b(z) := I_s + B_1 z + \cdots + B_q z^q$ where $z \in \mathbb{C}$, the pair $(a(z), b(z))$ is called a VARMA system corresponding to the stochastic process $\{y_t\}_{t\in\mathbb{Z}}$. We define the difference operator at frequency $0 \leq \omega \leq \pi$ as

$$\Delta_\omega(L) := \begin{cases} 1 - L, & \omega = 0 \\ 1 - 2\cos(\omega)L + L^2, & \omega \in (0,\pi) \\ 1 + L, & \omega = \pi. \end{cases}$$

For a set of frequencies $\Omega := \{\omega_1,\dots,\omega_l\}$, we denote $z_k := e^{i\omega_k}$ for $k = 1,\dots,l$ and $\mathbf{\Delta}_\Omega := \prod_{k=1}^{l}\Delta_{\omega_k}(L)$. We are now ready to define MFI(1) processes:

Definition 2.1 *A stochastic process $\{y_t\}_{t\in\mathbb{Z}}$ is a multiple frequency I(1) process (MFI(1) process) with unit root frequencies $\Omega = \{\omega_1,\dots,\omega_l\}$, $0 \leq \omega_1 < \omega_2 < \cdots < \omega_l \leq \pi$, if there exists a deterministic process[18] $\{\mathbf{d}_t\}_{t\in\mathbb{Z}}$, $\mathbf{d}_t \in \mathbb{R}^s$ such that*

$$\mathbf{\Delta}_\Omega(L)(y_t - \mathbf{d}_t) = v_t, \quad t \in \mathbb{Z},$$

where $\{v_t\}_{t\in\mathbb{Z}}$ is the stationary solution of a VARMA system $(a(z), b(z))$ fulfilling $\det(a(z)) \neq 0$ for $|z| \leq 1$, $\det(b(z)) \neq 0$ for all $|z| < 1$ and $b(z_k) \neq 0$ for $k = 1,\dots,l$.

Remark 2.1 *This definition of MFI(1) processes differs slightly from the definition in Bauer and Wagner (2012), since we allow the presence of a general deterministic process $\{\mathbf{d}_t\}_{t\in\mathbb{Z}}$, where Bauer and Wagner (2012) only allow for processes such that $\mathbf{\Delta}_\Omega\mathbf{d}_t = 0$. This is done to account for a linear trend term that is often included in empirical modeling.*
Second, the symbol Ω in this paper is used for the set of frequencies and not for the unit root structure. In Bauer and Wagner (2012) for an MFI(1) process the unit root structure is denoted as $((\omega_1, 1), ..., (\omega_l, 1))$. Since in this paper we only deal with MFI(1) processes, the simpler notation of only listing the unit root frequencies suffices.

[17] Note that contrary to definitions in the literature we here do not assume any stability condition. Thus, not all VARMA processes are stationary in our setting.
[18] Here and below a process $\{\mathbf{d}_t\}_{t\in\mathbb{Z}}$ is called deterministic, if $\mathbf{d}_t = \mathbb{E}(\mathbf{d}_t)$.

Similar as in Bauer and Wagner (2012, Theorem 2), it can be shown that every MFI(1) process $\{y_t\}_{t\in\mathbb{Z}}$ has a unique state space representation of the form

$$\begin{aligned} y_t &= \underbrace{\begin{bmatrix} \mathcal{C}_u & \mathcal{C}_\bullet \end{bmatrix}}_{\mathcal{C}} \begin{bmatrix} x_{t,u} \\ x_{t,\bullet} \end{bmatrix} + \mathbf{d}_t + h_t + \varepsilon_t \\ \begin{bmatrix} x_{t+1,u} \\ x_{t+1,\bullet} \end{bmatrix} &= \underbrace{\begin{bmatrix} \mathcal{A}_u & 0 \\ 0 & \mathcal{A}_\bullet \end{bmatrix}}_{\mathcal{A}} \begin{bmatrix} x_{t,u} \\ x_{t,\bullet} \end{bmatrix} + \underbrace{\begin{bmatrix} \mathcal{B}_u \\ \mathcal{B}_\bullet \end{bmatrix}}_{\mathcal{B}} \varepsilon_t \\ \begin{bmatrix} x_{1,u} \\ x_{1,\bullet} \end{bmatrix} &= \begin{bmatrix} 0 \\ \sum_{j=0}^{\infty} \mathcal{A}_\bullet^j \mathcal{B}_\bullet \varepsilon_{-j} \end{bmatrix} \end{aligned} \tag{2.2}$$

where $\mathcal{A} \in \mathbb{R}^{n\times n}$, $\mathcal{B} \in \mathbb{R}^{n\times s}$, $\mathcal{C} \in \mathbb{R}^{s\times n}$, $\{\mathbf{d}_t\}_{t\in\mathbb{Z}}$ is deterministic and $\{h_t\}_{t\in\mathbb{Z}}$ fulfills $\mathbf{\Delta}_\Omega(L)h_t = 0$ for all $t \in \mathbb{Z}$. Additionally,

- all eigenvalues of $\mathcal{A}_u$ are simple (i. e., their algebraic and geometric multiplicity coincide) and have unit modulus;
- $\lambda_{|\max|}(\mathcal{A}_\bullet) < 1$ and $\lambda_{|\max|}(\underline{\mathcal{A}}) \leq 1$ for $\underline{\mathcal{A}} := \mathcal{A} - \mathcal{B}\mathcal{C}$;
- the state space system $(\mathcal{A}, \mathcal{B}, \mathcal{C})$ is minimal, i. e., there is no alternative state space representation of $\{y_t\}_{t\in\mathbb{Z}}$ with a smaller state dimension;
- the subsystem $(\mathcal{A}_u, \mathcal{B}_u, \mathcal{C}_u)$ is of the form

  $$\mathcal{A}_u = \operatorname{diag}(\mathcal{A}_1, \ldots, \mathcal{A}_l), \quad \mathcal{B}_u = \begin{bmatrix} \mathcal{B}_1' & \ldots & \mathcal{B}_l' \end{bmatrix}', \quad \mathcal{C}_u = \begin{bmatrix} \mathcal{C}_1 & \ldots & \mathcal{C}_l \end{bmatrix},$$

  where the blocks correspond to the unit root frequencies Ω of $\{y_t\}_{t\in\mathbb{Z}}$ with $0 \leq \omega_1 < \omega_2 < \cdots < \omega_l \leq \pi$. If $\omega_k \in \{0, \pi\}$ we have $\mathcal{A}_k := z_k I_{c_k}$, $B_k := \mathcal{B}_k$ where B_k is positive upper triangular (p.u.t.)[19] and $C_k := \mathcal{C}_k$ where $C_k' C_k = I_{c_k}$. If $\omega_k \notin \{0, \pi\}$ we have

  $$\mathcal{A}_k := \begin{bmatrix} \cos(\omega_k) I_{c_k} & -\sin(\omega_k) I_{c_k} \\ \sin(\omega_k) I_{c_k} & \cos(\omega_k) I_{c_k} \end{bmatrix}, \ \mathcal{B}_k := \begin{bmatrix} \mathcal{B}_k^r \\ \mathcal{B}_k^i \end{bmatrix}, \ \mathcal{C}_k := \begin{bmatrix} \mathcal{C}_k^r & \mathcal{C}_k^i \end{bmatrix},$$

  where $B_k := \frac{1}{2}(\mathcal{B}_k^r - i\mathcal{B}_k^i)$ is p.u.t. and $C_k := (\mathcal{C}_k^r + i\mathcal{C}_k^i)$ is such that $C_k' C_k = I_{c_k}$;
- the state space representation of the (stable) subsystem $(\mathcal{A}_\bullet, \mathcal{B}_\bullet, \mathcal{C}_\bullet)$ is in echelon canonical form.

For notational simplicity we assume throughout the paper $\omega_1 = 0$ and $\omega_l = \pi$. Further, we will use the notation Δ_n for the set of all triples of system matrices (A, B, C), $A \in \mathbb{R}^{n\times n}$, in the canonical form, that is fulfilling all restrictions listed above.
This canonical form is associated to a multi-index Γ, whose properties are examined in Bauer et al. (2020, Theorem 2). Γ determines

- the state space unit root structure $\Omega_S = ((\omega_1, c_1), ..., (\omega_l, c_l))$,
- the indices $p \in \mathbb{N}^c$ with $c = \sum_{k=1}^{l} c_k \delta_k$ (where $\delta_k = 1$ for $k \in \{1, l\}$ and $\delta_k = 2$ else) denoting the dimension of $\mathcal{A}_u$, where p_i denotes the column index of the entry of the i-th row of $\mathcal{B}_u$ that is restricted to be a positive real,
- the Kronecker indices (see, e. g., Hannan and Deistler (1988, Chapter 2.4) for a precise definition) $\alpha_\bullet$ of the stable subsystem.

[19] Here a matrix $M = [m_{i,j}] \in \mathbb{C}^{r\times s}, r \leq s$, of full row rank is called *positive upper triangular (p.u.t.)*, if there exists indices $1 \leq m_1 < ... < m_r \leq s$ such that $m_{i,j} = 0$ if $j < m_i, m_{i,m_i} \in \mathbb{R}, m_{i,m_i} > 0$ for $i = 1, ..., r$.

Consequently, Γ also determines the system order $n = c + n_\bullet$, where $\mathcal{A}_\bullet \in \mathbb{R}^{n_\bullet \times n_\bullet}$. Bauer et al. (2020) show that the multi-index Γ is a property of the system such that the set of all systems of order n where all eigenvalues of $\mathcal{A}$ are within the closed unit disc can be partitioned into sets Δ_Γ. That is, $\Delta_n = \cup_\Gamma \Delta_\Gamma$ where $\Delta_\Gamma \cap \Delta_{\tilde{\Gamma}} = \varnothing$ for $\Gamma \neq \tilde{\Gamma}$. Conversely, every process $\{y_t\}_{t\in\mathbb{Z}}$ generated by a state space system (2.2) such that $(\mathcal{A}, \mathcal{B}, \mathcal{C}) \in \Delta_\Gamma$ and $\lambda_{|\max|}(\underline{\mathcal{A}}) \leq 1$ is an MFI(1) process with unit roots determined by the eigenvalues of $\mathcal{A}$. In this paper we only consider processes of this form:

Assumption 2.1 *The process $\{y_t\}_{t\in\mathbb{Z}}, y_t \in \mathbb{R}^s$, is generated by a state space system $(\mathcal{A}_\circ, \mathcal{B}_\circ, \mathcal{C}_\circ) \in \Delta_\Gamma$ corresponding to $k_\circ(z) = I_s + z\mathcal{C}_\circ(I_n - z\mathcal{A}_\circ)^{-1}\mathcal{B}_\circ \in M_\Gamma$ using (2.2) where $x_{1,u} = 0, x_{1,\bullet} = \sum_{j=0}^{\infty} \mathcal{A}_\bullet^j \mathcal{B}_\bullet \varepsilon_{-j}$ where the errors $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are a strictly stationary martingale difference sequence satisfying:*

- $\mathbb{E}(\varepsilon_t|\mathcal{F}_{t-1}) = 0$, *where* $\mathcal{F}_{t-1} = \sigma(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots)$.
- $\mathbb{E}(\varepsilon_t\varepsilon_t') = \mathbb{E}(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}) = \Sigma_\circ > 0$.
- $\mathbb{E}(|\varepsilon_t|^4) < \infty$.

Furthermore, the deterministic process $\mathbf{d}_t$ is given by

$$\begin{aligned} \mathbf{d}_t \quad := \quad & d_1 + \sum_{k=2}^{l-1} \left[d_k^r \cos(\omega_k(t-1)) + d_k^i \sin(\omega_k(t-1))\right] + d_l(-1)^{t-1} \\ & + \sum_{k=l+1}^{m} \left[d_k^r \cos(\omega_k(t-1)) + d_k^i \sin(\omega_k(t-1))\right] + d_{m+1}t \end{aligned} \tag{2.3}$$

where $d_1, d_l, d_{m+1} \in \mathbb{R}^s$, and $d_k^r, d_k^i \in \mathbb{R}^s$, for $k = 2, \dots, l-1$ and $k = l+1, \dots, m$ and where $\omega_{l+1}, \dots, \omega_m$ are frequencies, which are not unit root frequencies in Ω.
Finally, we assume that $\lambda_{|max|}(\underline{\mathcal{A}}_\circ) < 1$ (strict minimum-phase condition).

The properties of processes generated according to Assumptions 2.1 are discussed in detail in (Bauer and Wagner, 2012). Of particular importance for this paper is the concept of cointegration:

Definition 2.2 *An s-dimensional MFI(1) process $\{y_t\}_{t\in\mathbb{Z}}$ with unit root frequencies Ω is called cointegrated of order $(\Omega, \tilde{\Omega})$, where $\tilde{\Omega} \subsetneq \Omega$, 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 frequencies $\tilde{\Omega}$. In this case the vector β is called cointegrating vector (CIV) of order $(\Omega, \tilde{\Omega})$.*
An s-dimensional MFI(1) process $\{y_t\}_{t\in\mathbb{Z}}$ with unit root frequencies Ω is called polynomially cointegrated of order $(\Omega, \tilde{\Omega})$, where $\tilde{\Omega} \subsetneq \Omega$, 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

(i) $\beta(L)'\{y_t\}_{t\in\mathbb{Z}}$ has unit root frequencies $\tilde{\Omega}$,

(ii) $\max_{k=1,\dots,l}(|\beta(z_k)|\delta_k) \neq 0$ where $\delta_k = 1$ if $\omega_k \notin \tilde{\Omega}$ and $\delta_k = 0$ else.

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

Remark 2.2 *Consider the analysis of PCIVs of order $(\Omega, \tilde{\Omega})$ with $\tilde{\Omega} \subseteq \Omega \setminus \{\omega_k\}$. As discussed in Bauer and Wagner (2012, Section 5), it is sufficient to consider CIVs if $\omega_k \in \{0, \pi\}$ and PCIVs of degree one else.*
The decomposition $y_t = \sum_{k=1}^{l} \mathcal{C}_k x_{t,k} + \mathcal{C}_\bullet x_{t,\bullet} + \mathbf{d}_t + h_t + \varepsilon_t$ with $x_{t+1,k} = \mathcal{A}_k x_{t,k} + \mathcal{B}_k \varepsilon_t$ following from the block diagonality of $\mathcal{A}$ implies that for $\omega_k \in \{0, \pi\}$ the vector $0 \neq \beta \in \mathbb{R}^s$ is a CIV if and only if $\beta'\mathcal{C}_k x_{t,k}$ is stationary, since the other terms do not contain processes integrated at frequency ω_k. This is the case if and only if $\beta' C_k = 0$, because as shown in Bauer and Wagner (2012, Theorem 2) C_k has full column rank as a consequence of minimality and the state component

process $\{x_{t,k}\}_{t\in\mathbb{Z}}$ *is not cointegrated.*
For $\omega_k \notin \{0,\pi\}$ *the polynomial* $\beta(z)=\beta_0+\beta_1 z$ *is a PCIV of order* $(\Omega,\tilde{\Omega})$ *if and only if condition (ii) of Definition 2.2 is fulfilled and* $(\beta_0'+\beta_1'L)\mathcal{C}_k x_{t,k}$ *is stationary, since the other terms do not contain processes integrated at frequency* ω_k*. Since*

$$(\beta_0'+\beta_1'L)\mathcal{C}_k x_{t,k} = \beta_0'\mathcal{C}_k\mathcal{B}_k\varepsilon_{t-1} + (\beta_0'\mathcal{C}_k\mathcal{A}_k+\beta_1'\mathcal{C}_k)\, x_{t-1,k},$$

$\beta(z)$ *is a PCIV if and only if*

$$\begin{bmatrix}\beta_0' & \beta_1'\end{bmatrix}\begin{bmatrix}\mathcal{C}_k\mathcal{A}_k\\ \mathcal{C}_k\end{bmatrix}=0$$

and $\beta(z_k)\neq 0$*.*

The following lemma shows that the space of these PCIVs is isomorphic to the left kernel of the complex matrix C_k.

Lemma 2.1 *Let* $\omega_k\notin\{0,\pi\}$ *and let* $\gamma\in\mathbb{C}^s$ *be a vector satisfying* $\gamma\neq 0$ *and* $\gamma' C_k=0$*. Then* $\beta(z)=\gamma(1-z_k z)+\overline{\gamma}(1-\overline{z}_k z)$ *is a (real valued) PCIV of order* $(\Omega,\tilde{\Omega})$ *with* $\tilde{\Omega}\subseteq\Omega\setminus\{\omega_k\}$*. Conversely, for any PCIV* $\beta(z)$ *of polynomial degree 1 of order* $(\Omega,\tilde{\Omega})$ *with* $\tilde{\Omega}\subseteq\Omega\setminus\{\omega_k\}$ *we find a vector* $\gamma\in\mathbb{C}^s$*,* $\gamma\neq 0$*, such that* $\beta(z)=\gamma(1-z_k z)+\overline{\gamma}(1-\overline{z}_k z)$ *and* $\gamma' C_k=0$*.*

Therefore, the space spanned by CIVs respectively PCIVs of polynomial degree one is linked to the cointegration and polynomial cointegration properties of the process $\{y_t\}_{t\in\mathbb{Z}}$. This motivates the following definition linking the integer c_k with cointegrating ranks r_k:

Definition 2.3 *The complex cointegrating space of the process* $\{y_t\}_{t\in\mathbb{Z}}$ *at frequency* ω_k *is defined as the orthogonal complement of the column space of* C_k*. The cointegrating rank* r_k *at frequency* ω_k *is the dimension* $s-c_k$ *of the complex cointegrating space at* ω_k*.*

Note that the state space system $(\mathcal{A},\mathcal{B},\mathcal{C})$ is related to the transfer function $k(z)=I_s+z\mathcal{C}(I_n-z\mathcal{A})^{-1}\mathcal{B}=I_s+\sum_{j=0}^{\infty}\mathcal{C}\mathcal{A}^j\mathcal{B}z^{j+1}\in M_n$. Here, M_n denotes the set of all rational transfer functions of order[20] n with poles inside the closed unit disc. Let M_Γ denote the set of all transfer functions within M_n that correspond to $(\mathcal{A},\mathcal{B},\mathcal{C})\in\Delta_\Gamma$. Bauer et al. (2020) show that $M_n=\bigcup_\Gamma M_\Gamma$.
The inverse transfer function is $k^{-1}(z)=I_s-z\mathcal{C}(I_n-z\underline{\mathcal{A}})^{-1}\mathcal{B}$, where $\underline{\mathcal{A}}=\mathcal{A}-\mathcal{B}\mathcal{C}$. As in the VAR framework the matrices $\Pi_j=k^{-1}(z_j)=\alpha_j\beta_j'\in\mathbb{C}^{s\times s},\alpha_j,\beta_j\in\mathbb{C}^{s\times r_j}$, play a big role in this paper. It follows from the inversion that $\beta_j'C_j=0$ for all $j=1,...,l$. Furthermore $\Pi_1,\Pi_l\in\mathbb{R}^{s\times s}$. Thus, the complex cointegrating spaces are determined by the values of the transfer function at the unit roots, as are the cointegrating ranks. If the matrix $\Pi_j=k^{-1}(z_j)$ is of full rank, then the corresponding frequency ω_j is not contained in Ω.
In this paper consistency is derived in a coordinate free fashion analogously to the results of Hannan and Deistler (1988, Section 2.4). For this we will consider the maximization of the pseudo likelihood function over the product space $\overline{M_\Gamma}\otimes\underline{\Sigma}\otimes\Theta_D$ for appropriate Γ to be discussed below. Here, $\underline{\Sigma}\subset\mathbb{R}^{s\times s}$ denotes the set of all $s\times s$ non-negative definite symmetric matrices. A parameterization can be obtained for example using $\sigma\in\mathbb{R}^{s(s+1)/2}$ of vectorizations of the lower diagonal part (including the diagonal) of the Cholesky factor. This set can be equipped with the Euclidean topology.
The set Θ_D contains the parameters for the deterministic terms. In this respect note that with matrices of the form $\mathcal{A}_j$ we have:

$$\mathcal{C}_j\mathcal{A}_j^t\mathcal{B}_j x=\mathcal{C}_j\begin{bmatrix}\mathcal{R}(z_j^tB_j)\\ \mathcal{I}(z_j^tB_j)\end{bmatrix}x=\mathcal{C}_j\begin{bmatrix}\cos(\omega_j t)\mathcal{B}_j^r-\sin(\omega_j t)\mathcal{B}_j^i\\ \cos(\omega_j t)\mathcal{B}_j^i+\sin(\omega_j t)\mathcal{B}_j^r\end{bmatrix}x=d_j^r\cos(\omega_j t)+d_j^i\sin(\omega_j t)$$

for appropriate vectors $d_j^r=\mathcal{C}_j e_j^r, d_j^i=\mathcal{C}_j e_j^i\in\mathbb{R}^s$. Consequently, parts of the deterministic terms could also originate from non-zero state components of $x_{1,u}$. The results below will show that the

[20] The order of a transfer function is defined as the state dimension of a minimal state space representation of the rational transfer function, see Hannan and Deistler (1988, Chapter 2).

coefficients corresponding to these directions cannot be estimated consistently. Therefore, in the estimation we will set these directions equal to zero and only estimate the orthogonal complement: For $j = 1, ..., l$ let $d_j^r + id_j^i = [C_j, C_{j,\perp}][e_j', e_{j,\perp}']$. Then, we will use $e_j = 0$ and $e_{j,\perp} \in \mathbb{C}^{s-c_j}$. For $j = 1, l$ the vector $e_{j,\perp}$ is real. Consequently, Θ_D collects for $j = 1, \ldots, l$ the vectors $e_{j,\perp}$ and for $j = l+1, \ldots, m$ the real vectors d_j^r, d_j^i. Also, this space is equipped with the Euclidean topology. Finally, consider $M_\Gamma \subset M_n$. We equip this set with the pointwise topology T_{pt}.[21]
For the multi-index Γ we will consider two different scenarios:

- $\Gamma_0(n) = (\{\}, (), \alpha_{g,n})$ where $\alpha_{g,n}$ refers to the generic echelon neighborhood for systems of order n. This set contains all stable transfer functions of order n corresponding to the generic Kronecker indices. Bauer et al. (2020, Theorem 4) show that $\overline{M_{\Gamma_0(n)}} = M_n$. Thus, even if all considered systems are stable, the closure contains all systems corresponding to unit root processes and in particular all MFI(1) systems of order n. Here it is only assumed that the true order n of the data generating system is known. In fact only $n \geq n_0$ is required below such that $k_\circ \in \overline{M_{\Gamma_0(n)}}$.

- Γ: In this scenario we assume that the whole multi-index Γ is known and used in the estimation.

For the results on the asymptotic distribution of the parameter values we will only use the second scenario. In this situation of correctly specified multi-index Γ we use the parameterization introduced in Bauer et al. (2020): Let $\Theta_\Gamma \subset \Theta_{C,E} \times \Theta_{B,f} \times \Theta_{B,p} \times \Theta_\bullet \subset \mathbb{R}^{d(\Gamma)}$, such that a parameter vector $\theta \in \Theta_\Gamma$ is composed of

- the parameter vector $\theta_{C,E} \in \Theta_{C,E} \subset \mathbb{R}^{n_u}$ collecting parameters for the block columns of the unitary matrices $C_k = C_k(\theta_u)$, moreover $\theta_{C,E} = [\ \theta_{C,L}' \quad \theta_{C,D}' \quad \theta_{C,R}'\ ]'$ where $\theta_{C,L} = [\ \theta_{C,L,1}' \quad \ldots \quad \theta_{C,L,l}'\ ]'$ contains the parameters determining the column space of C_k and thus, the complex cointegrating space for each frequency ω_k and $\theta_{C,D} = [\ \theta_{C,D,2}' \quad \ldots \quad \theta_{C,D,l-1}'\ ]'$ and $\theta_{C,R} = [\ \theta_{C,R,1}' \quad \ldots \quad \theta_{C,R,l}'\ ]'$ describe the choice of the basis of the column space, see Bauer et al. (2020, Lemmas 1 and 2) for details,[22]

- the parameter vector $\theta_{B,f} \in \Theta_{B,f} = \mathbb{C}^{n_{B,f}}$ collecting the non-restricted entries in all B_k,

- the parameter vector $\theta_{B,p} \in \Theta_{B,p} = \mathbb{R}_+^{n_{B,p}}$ collecting the positive real entries in all B_k restricted due to the p.u.t. form,

- the parameter vector $\theta_\bullet \in \Theta_\bullet \subset \mathbb{R}^{n_{sta}}$ collecting the free entries in the echelon canonical form of the stable subsystem $(\mathcal{A}_\bullet, \mathcal{B}_\bullet, \mathcal{C}_\bullet)$ with Kronecker indices $\alpha_\bullet$.

Then, the true parameter vector $\theta_\circ$ will be assumed to be an inner point of Θ_Γ. From Bauer et al. (2020) it follows that $k_\circ$ always is an interior point of $\overline{M_\Gamma}$. Hence, $\theta_\circ$ only is no interior point of the parameter set, if it refers to a boundary in $\Theta_{C,E}$ as all other components of the parameter space are open sets (or points on the boundary of the strict minimum-phase condition such that $\lambda_{|max|}(\underline{\mathcal{A}}) = 1$). We will always assume that the true system is strictly minimum-phase such that $\lambda_{|max|}(\underline{\mathcal{A}_\circ}) < 1$). Clearly, the complement of this boundary constitutes an open and dense set such that points in the interior occur generically. For all other points suitable reparameterizations can be easily found. We will not deal with these issues here.

[21] A sequence of transfer functions $k_n = \sum_{j=0}^{\infty} K_{j,n} z^j$ converges in T_{pt} if and only if $K_{j,n} \to K_{j,0}$ for some $K_{j,0}$ for $j \in \mathbb{N}_0$.

[22] In fact the parameterization is changed slightly: $C_k(\theta_k) = R_L(\theta_{L,k})'[I_{c_k}, 0]' R_R(\theta_{R,k})$ in the real case where R_L and R_R are products of Givens rotations. While in Bauer et al. (2020) the matrix R_R is built rowwise, in this paper we change the ordering of the parameters and build the matrix columnwise. Clearly, this is a simple design choice not changing any of the results in Bauer et al. (2020).

2.3 Pseudo Maximum Likelihood Estimation

In this paper we will use the Gaussian likelihood function for estimation. This requires calculations of the expectation and the variance for the stacked vector of observations. To this end, note that for $t \geq 1$ the equations in (2.2) can be solved for

$$\begin{aligned} y_t &= \sum_{j=1}^{t-1} \mathcal{C}\mathcal{A}^{j-1}\mathcal{B}\varepsilon_{t-j} + \mathcal{C}_\bullet \mathcal{A}_\bullet^{t-1} x_{1,\bullet} + \mathbf{d}_t + \varepsilon_t \\ &= k(L)\varepsilon_t + \mathcal{C}_\bullet \mathcal{A}_\bullet^{t-1} x_{1,\bullet} + \mathbf{d}_t \end{aligned}$$

where – with slight abuse of notation – we set all processes to zero for $t \leq 0$ for notational convenience. Thus, e. g., $k(L)\varepsilon_t = \sum_{i=0}^{t-1} K_i \varepsilon_{t-i}$.
For the calculation of the pseudo likelihood function we will consider two different approaches that are commonly used in the literature: The pseudo maximum likelihood approach (PML) uses the stationary distribution of the stable part of the state. Accordingly we set $x_{1,u} = 0$ and $x_{1,\bullet}$ such that $\mathbb{E}(x_{1,\bullet}) = 0$ and $\mathrm{Var}(x_{1,\bullet}) = P_\bullet(\theta, \sigma)$, such that $P_\bullet(\theta, \sigma)$ solves the Ljapunov equation

$$P_\bullet = \mathcal{A}_\bullet(\theta) P_\bullet \mathcal{A}_\bullet(\theta)' + \mathcal{B}_\bullet(\theta)\Sigma(\sigma)\mathcal{B}_\bullet(\theta)'.$$

This can only be done when maximizing over M_Γ and hence the dimension $n_\bullet$ of the stable part is known. Note that consequently, the covariances $\mathrm{Cov}(y_t, y_{t-j})$ only depend on the system via the transfer function $k(z)$ and not on the system matrices (A, B, C) directly.
Alternatively, e. g., in the situation where the multi-index Γ is not considered to be known, the *prediction error* method of estimation is used. In this approach state values at time $t = 1$ are ignored and instead all of them are set equal to zero, that is $x_{1,\bullet} = 0$. Consequently, here $P_\bullet(\theta, \sigma) = 0$. While this does not conform with the data generating process, the exponential decay $\mathcal{A}_\bullet^{t-1} x_{1,\bullet}$ implies that for the asymptotical analysis the error is negligible in our setting. Furthermore, in terms of computations the approach is favorable since here instead of running the Kalman filter only filtering with fixed system is needed.
In order to define the pseudo likelihood function we use $Y_T := [\ y_1' \quad \dots \quad y_T'\ ]' \in \mathbb{R}^{Ts}$ for denoting the stacked observations, $\mathbf{D}_T(\theta_D) \in \mathbb{R}^{Ts}$ is equal to $[\ \mathbf{d}_1(\theta_D)' \quad \dots \quad \mathbf{d}_T(\theta_D)'\ ]'$, where

$$\mathbf{d}_t(\theta_D) = D(\theta_D) s_t, \quad s_t = [1, \sin(\omega_2(t-1)), \cos(\omega_2(t-1)), ..., \cos(\omega_m(t-1)), \sin(\omega_m(t-1)), t]'$$

denotes the deterministic terms corresponding to θ_D given in (2.3). Note that under our assumptions $\mathbb{E}y_t = \mathbf{d}_t$ and, moreover, that there exists a matrix $\mathcal{S}$ such that $s_t = \mathcal{S}s_{t-1}$. Further, let $\Gamma_T(k(z), \Sigma)$ denote the variance matrix corresponding to $Y_T - \mathbf{D}_T(\theta_D)$, which according to the model is given by

$$\begin{aligned} \Gamma_T(k(z), \Sigma) &= \mathcal{T}_T(k(z))\,(I_T \otimes \Sigma)\,\mathcal{T}_T(k(z))' + \mathcal{O}_{T,\bullet} P_\bullet(\theta, \sigma) \mathcal{O}_{T,\bullet}', \\ \mathcal{T}_T(k(z)) &:= \begin{bmatrix} K_0 & 0 & \dots & 0 \\ K_1 & K_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ K_{T-1} & \dots & K_1 & K_0 \end{bmatrix}, \end{aligned}$$

where K_j are the coefficients of the power series expansion of the transfer function $k(z) = \sum_{j=0}^{\infty} K_j z^j$. The $j-th$ block row of the observability matrix $\mathcal{O}_{T,\bullet}$ of the stable subsystem consists of $\mathcal{C}_\bullet \mathcal{A}_\bullet^{j-1}$.
Using this notation we obtain $-2/T$ times the logarithm of the Gaussian pseudo likelihood function as (up to a constant)

$$\begin{aligned} &L_T(k(z), \Sigma, \theta_D; Y_T) = \\ &\frac{1}{T}\left(\log\det\Gamma_T(k(z), \Sigma) + (Y_T - \mathbf{D}_T(\theta_D))'\Gamma_T(k(z), \Sigma)^{-1}(Y_T - \mathbf{D}_T(\theta_D))\right). \end{aligned} \tag{2.4}$$

The formulas for the prediction error (PE) approach can be significantly simplified: First, in this case $\Gamma_T(k(z),\Sigma) = \mathcal{T}_T(k(z))\,(I_T \otimes \Sigma)\,\mathcal{T}_T(k(z))'$ such that

$$\det(\Gamma_T(k(z),\Sigma)) = \det(\mathcal{T}_T(k(z))\,(I_T \otimes \Sigma)\,\mathcal{T}_T(k(z))') = \det(\mathcal{T}_T(k(z)))^2 \det(\Sigma)^T = \det(\Sigma)^T.$$

Second, noting that due to the block triangular structure $\mathcal{T}_T(k(z))^{-1} = \mathcal{T}_T(k^{-1}(z))$ we define

$$\varepsilon_t(k(z),\theta_D) := k^{-1}(L)(y_t - \mathbf{d}_t(\theta_D)) = \sum_{j=0}^{t-1} \underline{K}_j (y_{t-j} - \mathbf{d}_{t-j}(\theta_D)),$$

where $\underline{K}_j = -C\underline{A}^{j-1}B, \underline{A} = A - BC$, denote the power series coefficients of the inverse transfer function $k^{-1}(z)$. Letting

$$\mathcal{T}^{-1}(k(z))(Y_T - \mathbf{D}_T(\theta_D)) = \mathcal{E}_T(k(z),\theta_D) = [\ \varepsilon_1(k(z),\theta_D)' \quad \dots \quad \varepsilon_T(k(z),\theta_D)'\ ]' \in \mathbb{R}^{Ts},$$

we obtain in this case that $-2/T$ times the logarithm of the Gaussian likelihood function simplifies to

$$\begin{aligned} L_{PE,T}(k(z),\Sigma,\theta_D;Y_T) &= \log\det\Sigma + \mathcal{E}_T(k(z),\theta_D)'\,(I_T \otimes \Sigma)^{-1}\,\mathcal{E}_T(k(z),\theta_D)/T \\ &= \log\det\Sigma + \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t(k(z),\theta_D)'\Sigma^{-1}\varepsilon_t(k(z),\theta_D) \end{aligned} \tag{2.5}$$

We obtain the pseudo maximum likelihood estimate and the prediction error estimate respectively by minimizing (2.4) over the set $\overline{M_\Gamma} \times \underline{\Sigma} \otimes \Theta_D$:

$$\begin{aligned} \left(\hat{k}(z),\hat{\Sigma},\hat{\theta}_D\right) &:= \arg \min_{(k(z)\in\overline{M_\Gamma},\Sigma\in\underline{\Sigma},\theta_D\in\Theta_D)} L_T(k(z),\Sigma,\theta_D;Y_T), \\ \left(\tilde{k}(z),\tilde{\Sigma},\tilde{\theta}_D\right) &:= \arg \min_{(k(z)\in\overline{M_\Gamma},\Sigma\in\underline{\Sigma},\theta_D\in\Theta_D)} L_{PE,T}(k(z),\Sigma,\theta_D;Y_T). \end{aligned}$$

Then, using the coordinate free consistency proof in the stationary case of Hannan and Deistler (1988, Section 4.2.), the following result can be shown, whose proof in connection with some useful lemmas is given in Appendix B.2.1:

Theorem 2.1 *Let $\{y_t\}_{t\in\mathbb{Z}}$ be a real valued MFI(1) process generated by a system of the form (2.2) with $\mathbf{d}_t$ of the form (2.3) and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ fulfilling Assumption 2.1, where all frequencies $\omega_1,\dots,\omega_l$ are rational multiples of 2π.*
Let $k_\circ(z) \in \overline{M_\Gamma}$.
Then, the pseudo maximum likelihood estimator $\hat{k}(z) = I_s + \sum_{j=1}^{\infty}\hat{K}_j z^j$ converges in probability to the true transfer function $k_\circ(z)$ with rate $T^{1/2}$, i. e., $T^\gamma\|\hat{K}_j - K_{j,\circ}\| \to 0$ in probability for all $j \in \mathbb{N}$ and all $0 < \gamma < 1/2$. Furthermore,

$$T^\gamma\|\hat{\Pi}_j C_{j,\circ}\| \to 0,\ j = 1,\dots,l$$

in probability for all $0 < \gamma < 1$ where $\hat{\Pi}_j := \hat{k}^{-1}(e^{i\omega_j})$.
For $\hat{\theta}_D$ the following results hold:

- $T^\gamma\|C'_{1,\circ,\perp}(\hat{d}_{m+1} - d_{m+1,\circ})\| \to 0$ *in probability for all* $0 < \gamma < 3/2$.
- $T^\gamma\|C'_{1,\circ}(\hat{d}_{m+1} - d_{m+1,\circ})\| \to 0$ *in probability for all* $0 < \gamma < 1/2$.
- $T^\gamma\left[\|C'_{1,\circ,\perp}(\hat{d}_{1,\circ} - d_{1,\circ})\| + \sum_{k=2}^{l-1}\|C'_{k,\circ,\perp}(\hat{d}^r_k + i\hat{d}^i_k - d^r_{k,\circ} - id^i_{k,\circ})\| + \|(C'_{l,\circ,\perp}(\hat{d}_{l,\circ} - d_{l,\circ})\|\right] \to 0$ *in probability for all* $0 < \gamma < 1/2$.
- $T^\gamma\|\hat{d}_k - d_{k,\circ}\| \to 0$ *in probability for* $k = l+1,\dots,m$ *and all* $0 < \gamma < 1/2$.

The same holds for all prediction error estimators.

Remark 2.3 *The restriction of the frequencies being rational multiples of 2π is not necessary. However, it greatly simplifies the proof since, as will be seen in the appendix, the theorem is in this case equivalent to the consistency of the pseudo maximum likelihood estimator in the I(1) case. Since the restriction is fulfilled in all relevant cases like quarterly or monthly data, it is therefore reasonable to use it.*

Remark 2.4 *Our result of convergence of the transfer function estimator in suitable sets of transfer functions mirrors the coordinate free consistency theorem in the stationary case of Hannan and Deistler (1988).*

Remark 2.5 *Inspecting the results with respect to the deterministic terms a number of facts stick out: First, for the parameters corresponding to the linear trend term td_{m+1} the directions in the cointegrating space converge faster than $1/T$ as usual. In the directions of the common trends, however, convergence is slower of order $1/\sqrt{T}$ roughly. This corresponds to the difference in the order of growth between the linear trend term and the stochastic common trends. The same distinction also holds for deterministic terms with unit root frequency: In the directions of the cointegrating spaces convergence as in the stationary case occurs, in the direction of the common trends, however, not even convergence holds. Here the non-identification with terms due to the initial state manifests itself.*
For all other frequencies the deterministic terms show the usual convergence speed in all directions. For the rest of the paper thus we use the following notation:

$$\theta_{d,b} := \begin{bmatrix} e'_{1,\perp} & \mathcal{R}(e_{2,\perp})' & \dots & \mathcal{I}(e_{l-1,\perp})' & e'_{l,\perp} & \frac{1}{2}(d^r_{l+1})' & \dots & \frac{1}{2}(d^i_m)' \end{bmatrix}',$$
$$\theta_d := \begin{bmatrix} \theta_{d,b} & e'_{m+1} & e'_{m+1,\perp} \end{bmatrix}',$$

such that $\theta_d \in \mathbb{R}^{n_d}$ where $n_d = (s-c_1) + \sum_{k=2}^{l-1} 2(s-c_k) + (s-c_l) + (m+1-l)s$. By the above construction there exists a matrix $\mathcal{P}(\mathcal{C}_u) \in \mathbb{R}^{n_d \times s(2m-1)}$ such that $\mathcal{P}(\mathcal{C}_{u,\circ})\theta_D = \theta_d$. It follows that $\mathcal{P}(\mathcal{C}_{u,\circ})\hat{\theta}_D$ is a consistent estimator for $\theta_{d,\circ} := \mathcal{P}(\mathcal{C}_{u,\circ})\theta_{D,\circ}$.

The theorem states that we achieve consistent estimation of the transfer function both with the PML and the PE approach, if the data generating transfer function is contained in the closure of set over which maximization is performed. Since M_Γ is open in $\overline{M_\Gamma}$ for each Γ, see Bauer et al. (2020, Theorem 5), and since for given correctly specified Γ the parametrization is continuous on a generic subset, see Bauer et al. (2020, Theorem 2), we obtain the following corollary:

Corollary 2.1 *Let $\{y_t\}_{t\in\mathbb{Z}}$ be as in Theorem 2.1, where $k_\circ$ is in M_Γ. Further, assume that $k_\circ$ is a point of continuity of the parameterization $\theta = \phi_\Gamma(\psi_\Gamma(k_\circ))$.*
Then, $\begin{bmatrix} \hat{\theta}' & \hat{\theta}'_d \end{bmatrix}', \hat{\theta} = \phi_\Gamma(\psi_\Gamma(\hat{k}(z)))$, and $\begin{bmatrix} \tilde{\theta}' & \hat{\theta}'_d \end{bmatrix}', \tilde{\theta} = \phi_\Gamma(\psi_\Gamma(\tilde{k}(z)))$, converge in probability to the true parameter vector $\begin{bmatrix} \theta'_\circ & \theta'_{d,\circ} \end{bmatrix}'$.

Therefore, in the case of a correctly specified multi-index (generically) we obtain consistent parameter estimators. In this case the true parameter vector is an interior point of the parameter space and we can also develop the asymptotic distribution following standard asymptotic procedures proceeding in two steps: The first step is the derivation of the asymptotic distribution of the score vector and the second step is to derive convergence of the suitably normalized Hessian of the log-likelihood function. The proof of the theorem is given in Appendix B.2.2:

Theorem 2.2 (Asymptotic Distribution) *Let the assumptions of Theorem 2.1 hold. Additionally, let $\begin{bmatrix} \theta'_\circ & \theta'_{D,\circ} \end{bmatrix}'$, be an interior point of $\Theta_\Gamma \otimes \Theta_D$ over which the pseudo likelihood function is maximized such that $\hat{\theta}$ denotes the PMLE. Assume that the model for the deterministic terms contains the deterministic terms included in the data generating process.*
Split $\hat{\theta} = \begin{bmatrix} \hat{\theta}'_u & \hat{\theta}'_{st} \end{bmatrix}'$ with $\theta_u := \theta_{C,L} \in \mathbb{R}^{n_u}$, and $\theta_{st} := \begin{bmatrix} \theta'_{C,D} & \theta'_{C,R} & \theta'_{B,f} & \theta'_{B,p} & \theta'_\bullet \end{bmatrix}'$ and let θ_d be partitioned as above. Then, the following asymptotic distribution holds for the components

of θ:
(A)

$$\sqrt{T}(\hat{\theta}_{st} - \theta_{st,\circ}) \xrightarrow{d} \mathcal{N}(0, V_{st}), \quad V_{st} = [\mathbb{E}\partial_a \varepsilon_t(\theta_\circ)' \Sigma_\circ^{-1} \partial_b \varepsilon_t(\theta_\circ)]^{-1}_{a,b},$$

where $\partial_a \varepsilon_t(\theta)$ denotes the derivative of $\varepsilon_t(\theta)$ with respect to the a-th component of θ_{st}.
(B) Reorder the remaining parameters $\theta_u = [\theta'_{u,1}, ..., \theta'_{u,l}]'$ where $\theta_{u,k}$ corresponds to $C_k(\theta)$ and θ_d as $\tau = [\tau'_1, \tau'_2, ..., \tau'_m]'$, where τ_k corresponds to all parameters for unit root (or pair of complex unit roots) z_k. That is,

$$\tau'_1 = [\theta'_{u,1}, e'_{1,\perp}, e'_{m+1}, e'_{m+1,\perp}], ..., \tau'_l = [\theta'_{u,l}, e'_{1,\perp}], \tau'_{l+1} = 0.5[(d^r_{l+1})', (d^i_{l+1})'], ..., \tau'_m = 0.5[(d^r_m)', (d^i_m)'].$$

Here it is understood that only parameters included in the model occur. Further, let $D_{T,k} = diag(TI_{u_k}, T^{1/2}I_{f_k}, T^{3/2}I_{g_k})$ denote the corresponding diagonal scaling matrix where u_k denotes the dimension of $\theta_{u,k}$, f_k the dimension of the parameters for the constant and seasonal dummies and g_k is non-zero only for $k = 1$ where $g_1 = s - c_1$ such that it denotes the dimension of $e_{m+1,\perp}$. Then, the vectors $\hat{\tau}_k$ are asymptotically uncorrelated. Their asymptotic distribution is given by:

$$D_{T,k}(\hat{\tau}_k - \tau_{k,\circ}) \xrightarrow{d} H_k^{-1} \mathrm{v}_k.$$

Here H_k denotes the limit of the suitably normalized entries of the Hessian obtained as the limit to $T^{-h(a,b)} \sum_{t=1}^T \partial_a \varepsilon_t(\theta_\circ)' \Sigma_\circ^{-1} \partial_b \varepsilon_t(\theta_\circ)$ and v_k denotes the limit to $T^{-g(a,b)} \sum_{t=1} T \partial_a \varepsilon_t(\theta_\circ)' \Sigma_\circ^{-1} \varepsilon_t$, where the normalization factors $h(a,b)$ and $g(a,b)$ depend on the entries a and b.
(C) For a corresponding to an entry in $\theta_{u,k}$ we have

$$\partial_a \varepsilon_t(\theta_\circ) = k^{-1}(L, \theta_\circ)(\partial_a C_k) x_{t,k}(\theta_\circ), \quad x_{k,t+1}(\theta_\circ) = \overline{z}_k x_{k,t}(\theta_\circ) + B_{k,\circ} \varepsilon_t$$

for $z_k = \pm 1$,

$$\partial_a \varepsilon_t(\theta_\circ) = -2\mathcal{R}\{k^{-1}(L, \theta_\circ)(\partial_a C_k) x_{t,k}(\theta_\circ)\}, \quad x_{k,t+1}(\theta_\circ) = \overline{z}_k x_{k,t}(\theta_\circ) + B_{k,\circ} \varepsilon_t$$

for $z_k \neq \pm 1$ and else

$$\partial_a \varepsilon_t(\theta_\circ) = -k^{-1}(L, \theta_\circ)(-\partial_a D(\theta_{d,\circ}) s_t(\theta)).$$

(D) Let $W_k(u)$ denote a (real or complex) Brownian motion with variance $\Sigma_\circ$. Then, $H_{k,*}$ depends on $B_{k,\circ} W_k(u)$, while $\mathrm{v}_{k,*}$ depends on $B_{k,\circ} W_k(u)$ and $\alpha'_{k,\circ} \Sigma_\circ^{-1} W_k(u)$. These two Brownian motions are independent.
Moreover, $D_{T,k}(\hat{\tau}_k - \tau_{k,\circ})$ is mixed Gaussian distributed with conditional (on $B_{k,\circ} W_k(u)$) variance H_k^{-1}.
(E) For $k = 1$ we obtain $T(\hat{\theta}_{u,1} - \theta_{u,1,\circ}) \xrightarrow{d} H_{1,*}^{-1} \mathrm{v}_{1,*}$ where

$$H_{1,*,a,b} = tr\left[(\partial_a C'_1) \beta'_{1,\circ} \alpha'_{1,\circ} \Sigma_\circ^{-1} \alpha_{1,\circ} \beta'_{1,\circ} (\partial_b C_1) Z_{1,*}\right], \quad Z_{1,*} = B_{1,\circ} \int_0^1 W_{1,*}(u) W_{1,*}(u)' B'_{1,\circ},$$

$$\mathrm{v}_{1,*,a} = tr\left[(\partial_a C'_1) \beta'_{1,\circ} \alpha'_{1,\circ} \Sigma_\circ^{-1} X_{1,*}\right], \quad X_{1,*} = \int_0^1 dW_1(u) W_{1,*}(u)' B'_{1,\circ}$$

where $W_{1,*}(u)$ depends on the specification of the deterministic terms: $W_{1,*}(u) = W_1(u)$, if no constant or linear trend is included in the model, $W_{1,*}(u) = W_1(u) - \int_0^1 W_1(v)dv$, if a constant, but no linear trend is included and

$$W_{1,*}(u) = W_1(u) - \int_0^1 W_1(v)dv - 12(u - 1/0.5) \int_0^1 (v - 0.5) W_1(v)dv,$$

if both a linear trend and a constant are included.
(F) The prediction error estimator shows the same asymptotic distribution.

Remark 2.6 *The results of Theorem 2.2 can easily be extended to the case of seasonal trends corresponding to the unit root frequencies. Since these trends are not relevant in most applications, this is omitted.*

In the special case of an $I(1)$ process with no deterministic terms (both in the data generating process (DGP) and in the model) with only one common trend, i. e., $c_1 = 1$, and hence $s-1$ cointegrating relations β normalized such that $c'\beta = I_{s-1}$ Johansen (1995) obtains in the Johansen setting of a finite order VAR model for estimators $\tilde{\beta}, c'\tilde{\beta} = I_{s-1}$ the asymptotic distribution (compare Theorem 13.3):

$$T(\tilde{\beta} - \beta) \xrightarrow{d} (I - \beta c')\mathcal{C}_{1,\circ} \left(\mathcal{B}_{1,\circ} \int_0^1 W(u)W(u)'du\mathcal{B}'_{1,\circ} \right)^{-1} \int_0^1 \mathcal{B}_{1,\circ} W(u) dW(u)' \Sigma_\circ^{-1} \alpha (\alpha' \Sigma_\circ^{-1} \alpha)^{-1}.$$

Note that $\hat{\beta}'\hat{\mathcal{C}}_1 = 0 = \beta'\mathcal{C}_{1,\circ}$ and thus

$$0 = T(\hat{\beta}'\hat{\mathcal{C}}_1 - \beta'\mathcal{C}_{1,\circ}) = T(\hat{\beta} - \beta)'\mathcal{C}_{1,\circ} + T\beta'(\hat{\mathcal{C}}_1 - \mathcal{C}_{1,\circ}) + o_P(1).$$

Therefore, the PML estimate of the cointegrating space is in this case closely related to the estimate of the vector $\mathcal{C}_{1,\circ}$. Moreover, if $c = \beta$ it follows that

$$T(\mathcal{C}_1(\hat{\theta}) - \mathcal{C}_{1,\circ}) \xrightarrow{d} -\beta(\alpha'\Sigma_\circ^{-1}\alpha)^{-1}\alpha'\Sigma_\circ^{-1} \int_0^1 W(u)dW(u)'\mathcal{B}'_{1,\circ} \left(\mathcal{B}_{1,\circ} \int_0^1 W(u)W(u)'du\mathcal{B}'_{1,\circ} \right)^{-1}$$

is sufficient for equality of the asymptotic distributions.
For $c_1 = 1$ the vectors $\theta_{C,D}$ and $\theta_{C,R}$ do not occur. Therefore, all parameters corresponding to the $\mathcal{C}_1$ are contained in $\theta_{C,L} \in \mathbb{R}^{s-1}$ in this case. We obtain for a derivative of the residuals with respect to an entry in $\mathcal{C}_1$

$$\partial\varepsilon_t(\theta_\circ, \theta_{D,\circ}) = \partial(k^{-1}(L;\theta))(y_t - \mathbf{d}_t(\theta_{D,\circ})) = (\partial\Pi_1(\theta_\circ))(\mathcal{C}_{1,\circ}x_{t,1}) + v_t,$$

where v_t collects terms dominated by $x_{t,1}$. As $\Pi_1(\theta)\mathcal{C}_1(\theta) = 0$, we obtain $(\partial\Pi_1)\mathcal{C}_{1,\circ} + \Pi_{1,\circ}\partial\mathcal{C}_1 = 0$. Thus, the dominant term equals

$$-\Pi_{1,\circ}(\partial\mathcal{C}_1)x_{t,1} = -\alpha\beta'(\partial\mathcal{C}_1)x_{t,1} = -\alpha\beta'\beta V_C x_{t,1},$$

as the derivatives of $\mathcal{C}_1$ lie in the orthocomplement of $\mathcal{C}_1$, denoted as β and normalized such that $\beta'\beta = I_{s-1}$. The derivative with respect to the j-th parameter for $\mathcal{C}_1$ then has the representation $\beta V_{C,j}$, where thus $V_C \in \mathbb{R}^{(s-1)\times(s-1)}$ is a regular matrix.
It follows that T^{-1} times the diagonal block of the Hessian corresponding to the parameters for $\mathcal{C}_1$ contains as its (i,j)-th entry

$$\operatorname{tr}\left[\Sigma_\circ^{-1}\alpha V_{C,i} T^{-1} \langle x_{t,1}, x_{t,1}\rangle V'_{C,j}\alpha'\right],$$

where the notation $\langle a_t, b_t\rangle = T^{-1}\sum_{t=1}^T a_t b'_t$ is used.
Theorem 2.2 shows that $T^{-1}\langle x_{t,1}, x_{t,1}\rangle \xrightarrow{d} Z_{1,\star} = \int_0^1 \mathcal{B}_{1,\circ}W(u)(\mathcal{B}_{1,\circ}W(u))'du$ for a Brownian motion W with variance Σ_0. The block of the appropriately normalized Hessian of the scaled negative log-likelihood function then converges to $Z_{1,\star}V'_C\alpha'\Sigma_\circ^{-1}\alpha V_C$.
The derivative of the scaled negative log-likelihood function with respect to the i-th parameter to $\mathcal{C}_1$ is of the form

$$-\operatorname{tr}\left[\Sigma_\circ^{-1}\alpha V_{C,i}\langle x_{t,1}, \varepsilon_t\rangle\right] \xrightarrow{d} -\operatorname{tr}\left[\Sigma_\circ^{-1}\alpha V_{C,i} v'_{1,\star}\right] = -V'_{C,i}\alpha'\Sigma_\circ^{-1}v_{1,\star}, \quad v_{1,\star} = \int_0^1 \mathcal{B}_{1,\circ}W(u)dW(u)'.$$

Thus, we obtain

$$\begin{aligned}
&T(\hat{\theta}_{C,L} - \theta_{C,L,\circ}) \xrightarrow{d} \\
&\quad - (V'_C\alpha'\Sigma_\circ^{-1}\alpha V_C)^{-1}V'_C\alpha'\Sigma_\circ^{-1}v_{1,\star}Z_{1,\star}^{-1} = \\
&\quad - V_C^{-1}(\alpha'\Sigma_\circ^{-1}\alpha)^{-1}\alpha'\Sigma_\circ^{-1}v_{1,\star}Z_{1,\star}^{-1} = \\
&\quad - V_C^{-1}(\alpha'\Sigma_\circ^{-1}\alpha)^{-1}\alpha'\Sigma_\circ^{-1} \int_0^1 W(u)dW(u)'\mathcal{B}'_{1,\circ} \left(\mathcal{B}_{1,\circ} \int_0^1 W(u)W(u)'du\mathcal{B}'_{1,\circ} \right)^{-1}.
\end{aligned}$$

Then, the result follows from

$$T(\mathcal{C}_1(\hat{\theta}) - \mathcal{C}_{1,\circ}) = T\beta V_C(\hat{\theta}_{C,L} - \theta_{C,L,\circ}) + o_P(1).$$

This has two implications: First, the estimation of the cointegrating space in the VAR framework and in the VARMA framework achieves the same asymptotic distribution in this case. Therefore, we do not pay a price asymptotically in terms of accuracy by using the larger VARMA model in case the true data generating process is an VAR(p). Secondly, in the VARMA context the long VAR framework also asymptotically achieves the same distribution. Therefore, we do not gain asymptotically from using the correct model class in terms of the estimation accuracy for the cointegrating space. This insight is true not only in the specific case above but also in general as is shown in the companion paper Matuschek et al. (2020). Furthermore, this is also visible in the simulations in the next section. For the other parameters and prediction, however, using the parsimonious true model is beneficial, as is also visible in the simulations to follow.
The derivation of the asymptotic distribution allows to test hypotheses on the parameters with tests similar to the Wald tests. The following corollary, which is proven in Appendix B.2.2, gives the test statistics and asymptotic distribution for these tests :

Corollary 2.2 (Wald-type test) *Let the assumptions of Theorem 2.2 hold.*
Let $D_T^\theta := diag(T^{1/2} I_{n_{st}}, D_T^\star)$ with $D_T^\star = diag(D_{T,1}, \dots, D_{T,l})$ using the matrices $D_{T,k}$ defined in Theorem 2.2, $n_{st} := n_\bullet + n_{C,R} + n_{C,D} + n_{B,f} + n_{B,p}$ and $n_\theta := n_{st} + n_u + n_d$. Consider p linearly independent restrictions collected in $H_0 : R\theta = r$, with $R \in \mathbb{R}^{p \times n_\theta}$ of full row rank p, $r \in \mathbb{R}^p$ and suppose that there exists a matrix D_T^R such that

$$\lim_{T\to\infty} D_T^R R (D_T^\theta)^{-1} = R^\infty,$$

where $R^\infty \in \mathbb{R}^{p \times n_\theta}$ has full row rank p. Then, it holds that the Wald-type statistic

$$\hat{W}_R := (R\hat{\theta} - r)'(R(\hat{Z})^{-1}R')^{-1}(R\hat{\theta} - r), \quad [\hat{Z}]_{ij} = T \cdot tr\left[\hat{\Sigma}^{-1}\left\langle \partial_i \varepsilon_t(\hat{\theta}), \partial_j \varepsilon_t(\hat{\theta})\right\rangle\right]$$

is asymptotically χ_p^2 distributed under the null hypothesis.

2.4 Simulation Results

In this section we compare the pseudo maximum likelihood estimator to the vector error correction model (VECM) approach in an VAR setting of Johansen and Schaumburg (1999) and the CCA subspace estimator of Bauer and Buschmeier (2016). The data generating processes used are of the form

$$y_t = \Pi_1 y_{t-1} + \Pi_2 y_{t-2} + \Pi_3 y_{t-3} + \Pi_4 y_{t-4} + \varepsilon_t + \lambda \varepsilon_{t-4}$$

with

$$\Pi_1 = \begin{bmatrix} \gamma & 0 \\ 0 & 0 \end{bmatrix}, \Pi_2 = \begin{bmatrix} -0.4 & 0.4-\gamma \\ 0 & 0 \end{bmatrix}, \Pi_3 = \begin{bmatrix} -\gamma & 0 \\ 0 & 0 \end{bmatrix}, \Pi_4 = \begin{bmatrix} 0.6-0.1\gamma & 0.4+\gamma \\ 0 & 1 \end{bmatrix}$$

and

$$\varepsilon_t \sim \text{i.i.d. } N\left(0, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right).$$

In the case $\lambda = 0$ the processes correspond to those used in Bauer and Buschmeier (2016) and are similar to the ones used in Cubadda and Omtzigt (2005). For $\gamma = 0.2$ and $\lambda = 0$ as well as $\lambda = 0.9$ we generate samples of sizes $T \in \{100, 500\}$ with initial values for the state set to zero. For the simulations we make 5000 replications. All systems have unit root frequencies $\Omega = (0, \frac{\pi}{2}, \pi)$.
Each algorithm estimates the complex cointegrating space at unit root $\frac{\pi}{2}$. We calculate the gap

between the true and the estimated cointegrating space, where the gap $d_g(M,N)$ between to linear subspaces $M, N \in \mathbb{C}^s$ is defined by

$$d_g(M,N) = \max\left(\sup_{x\in M,||x||=1} ||(I_s - Q)x||, \sup_{x\in N,||x||=1} ||(I_s - P)x||\right),$$

where P denotes the orthogonal projection on M, Q the orthogonal projection on N and $||\cdot||$ the Euclidean norm on $\mathbb{C}^s$. Additionally, we make predictions with each algorithm with forecasting horizon twelve, which corresponds in the case of quarterly data to a three years ahead prediction. We compare the average (over the twelve prediction horizons) of the root mean squared error. For the Johansen Schaumburg VECM procedure we minimize the AIC to choose the lag length k. Since we have four unit roots, we have a lower bound of four for k such that $\hat{k} = \max\{4, \hat{k}_{AIC}\}$. For the CCA subspace algorithm we choose $f = p = 2\hat{k}_{AIC}$. To choose the system order n, we minimize a singular value criterion, see Bauer (2001). The number of common cycles introduces a lower bound $n = 4$, such that $\hat{n} = \max\{4, \hat{n}_{SVC}\}$. Since the CCA subspace algorithm delivers a stable system as an estimate, in order to obtain a state space system with the true state space unit root structure we first estimate the matrices $\mathcal{C}_1^{\text{ECM}}$ (corresponding to unit root $z = 1$), $\mathcal{C}_2^{\text{ECM}}$ (corresponding to unit roots $z = \pm i$) and $\mathcal{C}_3^{\text{ECM}}$ (corresponding to $z = -1$) separately, using the state space error correction model described Matuschek et al. (2020). This provides estimates for the different cointegrating spaces. Next, we reestimate C by a regression in a linear regression model

$$y_t = \mathcal{C}\hat{x}_t + \varepsilon_t \qquad \hat{x}_{t+1} = \underline{\mathcal{A}}^{\text{CCA}}\hat{x}_t + \mathcal{B}^{\text{CCA}}y_t$$

under the restrictions

$$\begin{aligned}(I - \mathcal{C}(I - \underline{\mathcal{A}}^{\text{CCA}})^{-1}\mathcal{B}^{\text{CCA}})\mathcal{C}_1^{\text{ECM}} &= 0\\ (I - i\mathcal{C}(I - i\underline{\mathcal{A}}^{\text{CCA}})^{-1}\mathcal{B}^{\text{CCA}})\mathcal{C}_2^{\text{ECM}} &= 0\\ (I + \mathcal{C}(I + \underline{\mathcal{A}}^{\text{CCA}})^{-1}\mathcal{B}^{\text{CCA}})\mathcal{C}_3^{\text{ECM}} &= 0,\end{aligned}$$

where $\hat{x}_t$ is computed using $(\mathcal{A}^{\text{CCA}}, \mathcal{B}^{\text{CCA}}, \mathcal{C}^{\text{CCA}})$, which are the estimated system matrices from the CCA subspace procedure. Denoting the reestimated matrix by $\mathcal{C}^{\text{reg}}$, we also find a new estimate for the matrix $\mathcal{A}$ using the equality $\underline{\mathcal{A}} = \mathcal{A} - \mathcal{B}\mathcal{C}$. The resulting system is given by

$$(\mathcal{A}^{\text{reg}}, \mathcal{B}^{\text{CCA}}, \mathcal{C}^{\text{reg}}) = (\underline{\mathcal{A}}^{\text{CCA}} + \mathcal{B}^{\text{CCA}}\mathcal{C}^{\text{reg}}, \mathcal{B}^{\text{CCA}}, \mathcal{C}^{\text{reg}}).$$

This system is also used for the predictions.
For the pseudo maximum likelihood algorithm we use $(\mathcal{A}^{\text{CCA}}, \mathcal{B}^{\text{CCA}}, \mathcal{C}^{\text{CCA}})$ as a starting value for the PML algorithm maximizing over $\Theta_{\Gamma_0(n)}$, that is over all state space systems of fixed order with empty unit root structure, such that the resulting PML estimate is again a stable system. Applying again the reestimation of the matrix $\mathcal{C}$ and $\mathcal{A}$ under restrictions as described above, we use the reestimated system as a starting value for the PML algorithm over systems with the true unit root structure, that is maximizing over Θ_Γ. We use this final estimate to derive the complex cointegrating spaces and the predictions. In Figure 2.1 the density of the logarithm of the gap between true and estimated complex cointegrating space for the three algorithms for sample sizes $T = 100$ and $T = 500$ is compared for the VAR case $\lambda = 0$. In this case the Johansen Schaumburg algorithm outperforms the CCA subspace and the PML algorithm for sample size $T = 100$. For sample size $T = 500$ all algorithms perform comparably. This is not surprising, as the data generating process is a VAR process. The advantage for the Johansen Schaumburg procedure, however, disappears for sample size $T = 500$.
In the VARMA case with $\lambda = 0.9$, shown in Figure 2.2, the pseudo maximum likelihood algorithm performs best for sample size $T = 100$ and the Johansen Schaumburg procedure performs better than the CCA subspace algorithm. For sample size $T = 500$ the CCA subspace and the PML algorithm outperform the Johansen Schaumburg procedure. Since the MA-part of the data generating

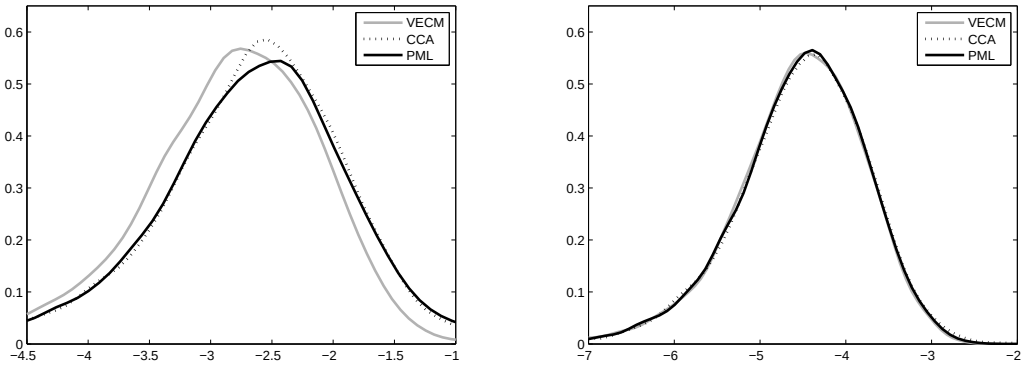

Figure 2.1: Density of the logarithm of the gap between estimated and true complex cointegrating space at unit root frequency $\frac{\pi}{2}$ for $T = 100$ (left) and $T = 500$ (right) in the VAR case

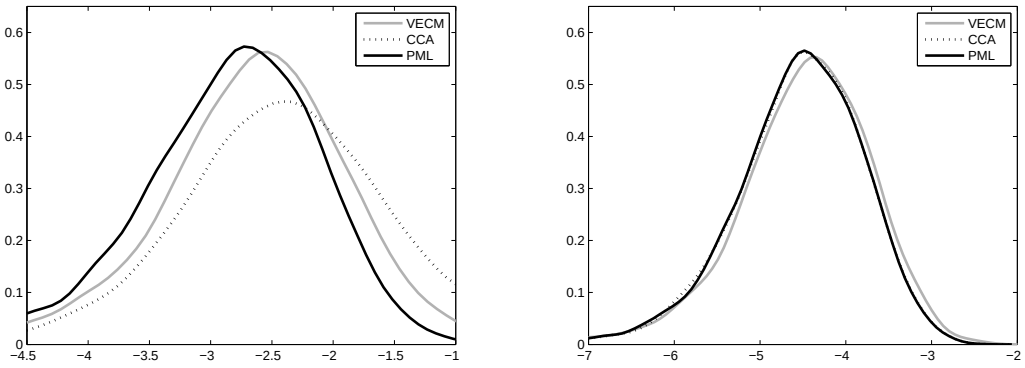

Figure 2.2: Density of the logarithm of the gap between estimated and true cointegrating space at unit root frequency $\frac{\pi}{2}$ for $T = 100$ (left) and $T = 500$ (right) in the VARMA case with $\lambda = 0.9$

process has zeros near the unit circle, the estimated lag length $\hat{k}$ increases which explains why the Johansen Schaumburg algorithm is not as good as the other algorithms in this case.

The results for the predictions are similar to the complex cointegrating space estimations. As seen in Figure 2.3, in the VAR-case $\lambda = 0$ the predictions of the Johansen Schaumburg algorithm have the smallest root mean square error for sample size $T = 100$ and there are no differences visible between the algorithms for sample size $T = 500$. The density of the root mean squared error of the predictions in the VARMA case with $\lambda = 0.9$ is shown in Figure 2.4. In this case for sample size $T = 100$ the PML algorithm is best and the Johansen Schaumburg approach is better than the CCA subspace algorithm. For sample size $T = 500$ the CCA subspace and the PML algorithm are better than the Johansen Schaumburg algorithm.

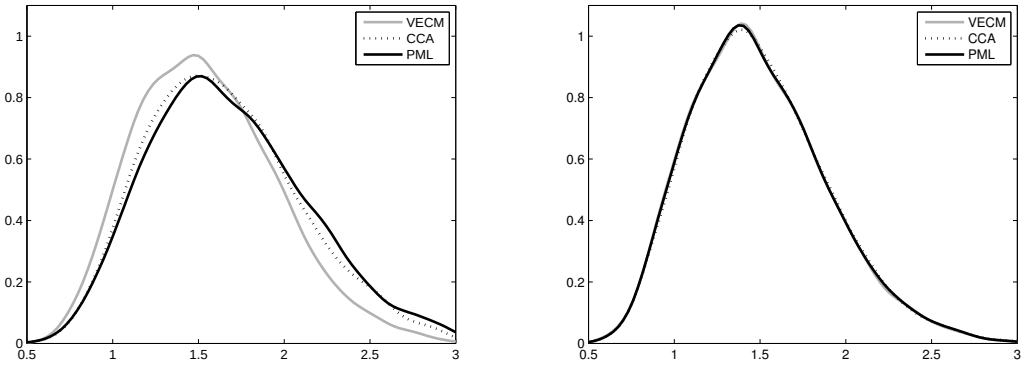

Figure 2.3: Density of the average (over the two dimensions) of the root mean squared errors for $T = 100$ (left) and $T = 500$ (right) for twelve steps ahead predictions in the VAR case.

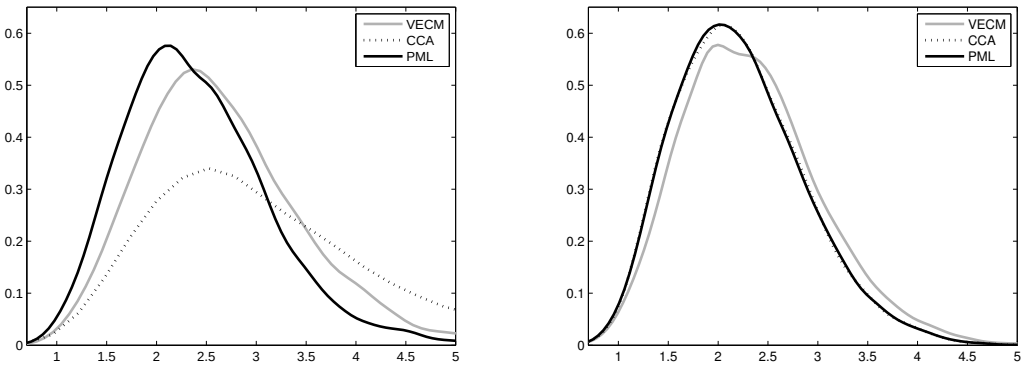

Figure 2.4: Density of the average (over the two dimensions) of the root mean squared errors for $T = 100$ (left) and $T = 500$ (right) for twelve steps ahead predictions in the VARMA case with $\lambda = 0.9$

2.5 Summary and Conclusion

This paper shows that transfer functions corresponding to MFI(1) processes can be consistently estimated by maximizing the pseudo likelihood function in the state space framework. As state space systems correspond in a certain sense to VARMA systems, this overcomes the limitation in the literature to VAR processes only. Consistent estimates can be obtained even if the unit root structure is unknown or ignored such that models corresponding to stationary processes are considered. The model allows for a deterministic constant, a linear trend and seasonal dummies. The cointegrating space for each unit root is estimated with rate T. If a linear trend is included in the model the corresponding coefficients in the direction of the cointegrating space at unit root frequency zero are estimated with rate $T^{3/2}$.
The asymptotic distribution of the parameters describing the cointegrating space is a mixture of Brownian motions. The Brownian motions become demeaned Brownian motions if a constant and seasonal dummies are included in the model. If a linear trend is included in the model, the asymptotic distribution of the parameters describing the cointegrating space at unit root frequency zero is a mixture of demeaned and detrended Brownian motions. The asymptotic distribution of the coefficients of the trend in the direction of the cointegrating space is also a mixture of Brownian motions. The other parameters are asymptotically normally distributed.
In the simulation study the pseudo maximum likelihood algorithm, as expected, estimates the cointegrating space more precisely and leads to better predictions in small samples than the Johansen Schaumburg approach in the VAR framework and the CCA subspace algorithm for a VARMA process, where the MA part has roots near the unit circle.
This indicates that state space modeling has the potential to improve estimation of cointegrating spaces as well as predictions in situations where in the VAR framework a large order is needed, while more parsimonious approximations in the VARMA framework can be found. The derivation of the asymptotic distribution in principle opens the door for the derivation of tests for hypotheses on the parameters provided the multi-index Γ is specified correctly. An investigation of methods for specifying this multi-index as well as pseudo likelihood ratio tests for cointegrating spaces is conducted in a companion paper showing that procedures analogous to the Johansen framework can also be used in the VARMA situation (Matuschek et al., 2020).

Stochastic Trends and Economic Fluctuations: Reconsidered From the State Space

3.1 Introduction

In his seminal paper Granger (1981) introduced the concept of cointegration. Since then the usage of models and methods that allow for and incorporate unit roots and cointegration has become a common practice for the modeling of econometric time series with applications in various fields.
There are non-parametric and parametric approaches for cointegration analysis. Non-parametric approaches as, e. g., Stock and Watson (1988) focus on the estimation and development of hypotheses tests for the cointegrating relationships and treat all other characteristics of the data generating process as nuisance parameters.
Parametric approaches perform cointegration analysis in a fully specified model class. The most prominent model class for cointegration analysis is the vector error correction model (VECM) using the vector autoregressive (VAR) framework, popularized by Johansen and his co-authors, see, e.g., the monograph Johansen (1995).
A less prominent parametric approach is the vector autoregressive moving average (VARMA) framework, see, e. g., Yap and Reinsel (1995) and Poskitt (2006). This approach can overcome some limitations of the VAR framework. First, unlike VARMA processes, VAR processes are not closed, as shown by Amemiya and Wu (1972), with respect to temporal aggregation and, as shown by Zellner and Palm (1974), with respect to marginalization. Second, as shown by Campbell (1994), the solutions of dynamic stochastic general equilibrium (DSGE) models are typically VARMA rather than VAR processes. Third, VARMA models may be more parsimonious than VAR models in some cases.
With a few early exceptions as, e. g., Aoki and Havenner (1989) and Aoki (1990) the state space framework, which is, as discussed below, equivalent to the VARMA framework, has not received a lot of attention. In a recent series of papers, however, the authors have developed several tools for cointegration analysis in the state space framework. A brief description will be given below for the I(1) case, see Bauer and Wagner (2012), Bauer et al. (2020), de Matos Ribeiro, Matuschek, Bauer and Wagner (2020) and Matuschek et al. (2020) for details.
The aim of this paper is to provide a tutorial for the application of these tools. To achieve this, we repeat the seminal analysis of postwar US economic data by King et al. (1991) using the state space framework, showing which economically relevant questions can be examined with these methods and explaining which code can be used for this analysis. The results are also compared to those obtained using the VECM. Additionally, we test the robustness of the VECM and the state space framework by repeating the analysis on a larger data set with economic data until 2018, not until 1988 as in King et al. (1991), and on the subsample of the data set from 1989 until 2018.
This paper is structured as follows: Section 3.2 describes cointegration analysis using the state space framework and the VECM and the methods used by King et al. (1991). The latter two are only described briefly, since they have been extensively covered elsewhere. Section 3.3 describes

the data sets used in the analysis. Section 3.4 compares the results of King et al. (1991) to those obtained using the VECM and the state space framework. Section 3.5 contains the results of the analysis of the longer data set using the VECM and the state space framework. Section 3.6 contains the analysis of the subsample from 1989 to 2018. Section 3.7 summarizes and concludes the paper.

3.2 Methods

3.2.1 The State Space Framework

An s-dimensional stochastic process $\{y_t\}_{t\in\mathbb{Z}}$ has a state space representation, if there are matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{s\times n}$, $C \in \mathbb{R}^{s\times n}$ and $\Phi \in \mathbb{R}^{s\times m}$, an n-dimensional stochastic process $\{x_t\}_{t\in\mathbb{Z}}$ called the state process, which is in general unobserved, an s-dimensional white noise process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ and an m-dimensional deterministic process $\{d_t\}_{t\in\mathbb{Z}}$ such that[23]

$$\begin{aligned} y_t &= Cx_t + \Phi d_t + \varepsilon_t \\ x_{t+1} &= Ax_t + B\varepsilon_t. \end{aligned} \tag{3.1}$$

Every ARMA process has a state space representation. Consider an ARMA process

$$y_t = \textstyle\sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{q} b_i \varepsilon_{t-i} + \varepsilon_t,$$

neglecting the deterministic process for simplicity. Choosing

$$A = \begin{pmatrix} a_1 & a_2 & \dots & a_{p-1} & a_p & b_1 & \dots & b_{q-1} & b_q \\ I_s & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & I_s & \dots & 0 & 0 & \vdots & & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & I_s & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & I_s & \dots & 0 & 0 \\ & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & I_s & 0 \end{pmatrix}, \quad B = \begin{pmatrix} I_s \\ 0 \\ \vdots \\ 0 \\ I_s \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad x_t = \begin{pmatrix} y_{t-1} \\ \vdots \\ y_{t-p} \\ \varepsilon_{t-1} \\ \vdots \\ \varepsilon_{t-q} \end{pmatrix}$$

$$C = \begin{pmatrix} a_1 & a_2 & \dots & a_p & b_1 & b_2 \dots & b_q \end{pmatrix}$$

we get (3.1). Conversely, every process, which has a state space representation, also has an ARMA representation. Like ARMA representations, the state space representation of a stochastic process is not unique. There are two sources of non-uniqueness: First, it is possible to introduce superfluous components in the state space representations without changing the data generating process. For example

$$\begin{aligned} y_t &= C_1 x_{t,1} + \Phi d_t + \varepsilon_t \\ x_{t+1} &= A_1 x_t + B_1 \varepsilon_t \end{aligned}$$

with $x_t \in \mathbb{R}^{n_1}$ and

$$\begin{aligned} y_t &= \begin{pmatrix} C_1 & 0 \end{pmatrix} \begin{pmatrix} x_{t,1} \\ x_{t,2} \end{pmatrix} + \Phi dt + \varepsilon_t \\ x_{t+1} &= \begin{pmatrix} A_1 & 0 \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} x_{t,1} \\ x_{t,2} \end{pmatrix} + \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} \varepsilon_t \end{aligned}$$

lead to the same $\{y_t\}_{t\in\mathbb{Z}}$ for all $n_2 \in \mathbb{N}$, $\{x_{t,2}\}_{t\in\mathbb{Z}}$ with $x_{t,2} \in \mathbb{R}^{n_2}$, $A_{21} \in \mathbb{R}^{n_2\times n_1}$, $A_{22} \in \mathbb{R}^{n_2\times n_2}$ and $B_2 \in \mathbb{R}^{n_2\times s}$. The reason is that due to the structure of the matrices there is no feedback from

[23] 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 , as shown by Aoki (1990, Chapter 7.1), every state space system has an innovation representation, this is no restriction.

$\{x_{t,2}\}_{t\in\mathbb{Z}}$ to $\{y_t\}_{t\in\mathbb{Z}}$ and $\{x_{t,1}\}_{t\in\mathbb{Z}}$. A state space representation is called minimal, if $\{y_t\}_{t\in\mathbb{Z}}$ has no state space representation with lower dimensional state process. Minimality in the state space framework is analogue to left coprimeness in the VARMA framework, cf. Hannan and Deistler (1988, Chapter 2). Second, minimality is not sufficient for uniqueness of the state space representation, because the basis of the state process can be freely chosen. Since for every regular matrix $T \in \mathbb{R}^{n\times n}$

$$\begin{aligned} y_t &= CT^{-1}Tx_t + \Phi d_t + \varepsilon_t \\ Tx_t &= TAT^{-1}Tx_t + TB\varepsilon_t \end{aligned}$$

is equivalent to (3.1), the matrices (A, B, C) and (TAT^{-1}, TB, CT^{-1}) are observationally equivalent, i. e., they generate the same stochastic process. By Hannan and Deistler (1988, Theorem 2.3.4) two minimal state space systems are observationally equivalent if and only if such a regular matrix T exists. The above discussion shows that the state space representation is not unique. To achieve uniqueness, a canonical form is necessary, placing restrictions on the system matrices. Many canonical forms exists, e. g., the echelon canonical form, cf. Hannan and Deistler (1988, Theorem 2.5.2). Bauer and Wagner (2012) developed a canonical form that separates the integrated and the stationary subsystem. For the I(1) case the system matrices in Bauer-Wagner canonical form are

$$A = \begin{pmatrix} I_{n_c} & 0 \\ 0 & A_\bullet \end{pmatrix}, \quad B = \begin{pmatrix} B_1 \\ B_\bullet \end{pmatrix}, \quad C = \begin{pmatrix} C_1 & C_\bullet \end{pmatrix},$$

where the absolute value of all eigenvalues of $A_\bullet$ is smaller than one, the matrices $(A_\bullet, B_\bullet, C_\bullet)$ with $A_\bullet \in \mathbb{R}^{n_\bullet\times n_\bullet}$, $B_\bullet \in \mathbb{R}^{n_\bullet\times s}$ and $C_\bullet \in \mathbb{R}^{s\times n_\bullet}$ are in echelon canonical form, $C_1'C_1 = I_{n_c}$ and $B_1 \in \mathbb{R}^{n_c\times s}$ is a positive upper triangular matrix of full rank, i. e., of the form

$$B_1 = \begin{pmatrix} 0 & \dots & 0 & b_{1j_1} & * & & \dots & & * \\ 0 & & \dots & & 0 & b_{2j_2} & * & \dots & * \\ \vdots & & & & & & & & \vdots \\ 0 & & & \dots & & 0 & & b_{n_cj_{n_c}} & * \end{pmatrix},$$

where $1 \le j_1 < j_2 < \dots < j_{n_c} \le s$, $b_{ij_i} > 0$ for $i = 1, \dots, n_c$ and $*$ denotes unrestricted entries. Here $(A_\bullet, B_\bullet, C_\bullet)$ is the stationary subsystem and $(A_1 = I_{n_c}, B_1, C_1)$ is the integrated subsystem. Bauer and Wagner (2012) show that $\beta \in \mathbb{R}^s$, $\beta \neq 0$ is a cointegrating vector if and only if $\beta' C_1 = 0$. Thus, C_1 spans the orthocomplement of the cointegrating space and the cointegrating vectors can be easily derived from the Bauer-Wagner canonical form. The dimension of the cointegrating space, the cointegrating rank, is $r = s - n_c$, where n_c is the number of common trends. As in the VARMA case no continuous parameterization of all state space systems of a certain state process dimension n exists, see Hazewinkel and Kalman (1976). Thus, we need to partition the model set into subsets where a continuous parameterization exists. Bauer et al. (2020) use a partitioning based on the canonical form of Bauer and Wagner (2012) where the relevant parameters are collected in a multi-index Γ. For the I(1) case Γ consists of the number of common trends n_c, the positions of the positive entries of B_1, $j_1, \dots, j_{n_c}$ and the Kronecker indices of the stable subsystem, see Hannan and Deistler (1988, Chapter 2.4). Bauer et al. (2020) discuss the number of free parameters for each subset in great detail. The number of parameters of the stable subsystem depends on the Kronecker indices. Obviously, A_1 does not contain any free parameters. For the calculation of the number of free parameters in C_1, consider a parameterization that places restrictions on the columns to ensure the normalization and the orthogonality to previous columns. Using this parameterization, the first column of the matrix C_1 contains $s-1$ free parameters, due to the normalization. The second column contains $s-2$ free parameters, due to the normalization and the orthogonality to the first column. By the same argument there are $s-i$ free parameters in the i-th column.[24] Thus, there are $n_c \cdot s - \frac{1}{2}n_c(n_c+1)$ free parameters in C_1. The number of

[24] Since Bauer et al. (2020) use Givens rotations and distinguish between the parameters that span the cointegrating space and the ones which choose a basis in the cointegrating space, the parameters are slightly more complicated. The number of parameters, however, is accurate.

free parameters in the i-th row in B_1 is $s+1-j_i$. Thus, B_1 contains $n_c(s+1)-\sum_{i=1}^{n_c} j_i$ free parameters. Consequently, the integrated subsystem contains $n_c(2s+1)-\frac{1}{2}n_c(n_c+1)-\sum_{i=1}^{n_c} j_i$ free parameters. In the case $j_i = i$ for $i = 1, \ldots, n_c$ the number of free parameters simplifies to $2n_c(s-n_c)$.
Additionally, Bauer et al. (2020, Theorem 4) show that the set of all transfer functions corresponding to I(1)-systems with state dimension n, c common trends, $j_i = i$ for $i = 1, \ldots, n_c$ and generic Kronecker indices of the stable subsystem is open and dense in the set of all transfer functions corresponding to I(1)-systems with state dimension n and c common trends. Bauer et al. (2020, Theorem 4) also show that the set of all stationary systems of state dimension n with generic Kronecker indices is open and dense in the set of all I(1) systems of state dimension n. These results provide the basis for consistent estimators for unknown Γ. Denoting the transfer function, i.e, the function relating the error sequence to the observed process with $k(z)$ de Matos Ribeiro et al. (2020) use the power series expansion of the inverse transfer function with coefficients $\underline{K}_j$ to define the residuals $\varepsilon_t(k(z,\theta),\theta_D) := \sum_{j=1}^{t-1} \underline{K}_j(y_{t-j} - \Phi(\theta_D)d_t)$, where θ is the parameter vector of the system matrices (A, B, C) and θ_D is the parameter vector containing the entries of Φ. de Matos Ribeiro et al. (2020, Section 3) show that if we set $x_1 = 0$ the logarithm of the likelihood function for given $y_1, \ldots, y_T$ is

$$L_T = -\frac{2}{T}\left(\log\det\Sigma + \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t(k(z,\theta),\theta_D)'\Sigma^{-1}\varepsilon_t(k(z,\theta),\theta_D)\right). \tag{3.2}$$

A necessary condition for consistency of the pseudo maximum likelihood estimator is the convergence of the sum in (3.2). Since, as discussed in Bauer and Wagner (2012), $\underline{K}_j = -C(A-BC)^{j-1}B$, this is the case if the absolute value of all eigenvalues of $A - BC$ is smaller than one. This condition is called the strict minimum phase condition. If this condition is fulfilled, de Matos Ribeiro et al. (2020, Theorem 1) show that the pseudo maximum likelihood estimator is consistent under relatively mild conditions on ε_t. The cointegrating space is estimated with rate T, where T is the sample size. For given multi-index Γ, de Matos Ribeiro et al. (2020, Theorem 2) also derive the asymptotic distribution.
For the determination of the cointegrating rank $r = s - c$ Matuschek et al. (2020) rewrite (3.1) to a state space error correction model (SSECM) first introduced by Ribarits and Hanzon (2014)

$$\begin{aligned} \Delta(y_t - \Phi d_t) &= \Pi(y_{t-1} - \Phi d_{t-1}) - Cv_t + \varepsilon_t \\ v_{t+1} &= \underline{A}v_t + (I_n - \underline{A})^{-1}\underline{A}B\Delta(y_t - \Phi d_t) \\ \Pi &= -I_s + C(I_n - \underline{A})^{-1}B, \end{aligned} \tag{3.3}$$

where $\Delta := 1 - L$ with the lag operator $L\{y_t\}_{t\in\mathbb{Z}} := L\{y_{t-1}\}_{t\in\mathbb{Z}}$, $\underline{A} := A - BC$ and $v_1 = x_1$. The matrix Π in the SSECM can be written as $\Pi = \alpha\beta'$ with α, $\beta \in \mathbb{R}^{s\times r}$ of full rank. The columns of β form a basis of the cointegrating space. Focusing on the case without deterministic terms, i. e., $\Phi = 0$ for simplicity concentrating Σ out of (3.2) the likelihood up to a constant is proportional to

$$\log\left(\left|\sum_{t=1}^{T}(\Delta y_t - \Pi y_{t-1} + Cv_t(\theta))(\Delta y_t - \Pi y_{t-1} + Cv_t(\theta))'\right|\right), \tag{3.4}$$

where $v_t(\theta)$ depends on the parameters in $\underline{A}$ and B. After having estimated $\underline{A}$ and B with the pseudo maximum likelihood estimator from (3.2) the authors maximize (3.4) under the constraint $\Pi = -I_s + C(I_n - \underline{A})^{-1}B$, using Lagrange multipliers. They then repeat the maximization under the additional constraint $\operatorname{rank}(\Pi) = r$. The test statistic is the difference of the log-likelihoods at these maxima and is

$$\operatorname{tr}\left(\int_0^1 (dW)W'\left(\int_0^1 WW'du\right)^{-1}\int_0^1 WdW'\right)$$

distributed, where W is a $s-r$ dimensional Brownian motion. If a deterministic linear trend or a deterministic constant are included the test statistic is

$$\operatorname{tr}\left(\int_0^1 (dW)H'\left(\int_0^1 HH'du\right)^{-1}\int_0^1 HdW'\right)$$

distributed with $H = (W', 1)'$.
After having determined the cointegrating rank, Matuschek et al. (2020) estimate the basis of the cointegrating space β by maximizing (3.4) under the constraints $\Pi = -I_s + C(I_n - \underline{A})^{-1}B$ and $\operatorname{rank}(\Pi) = r$. To test a hypothesis on the cointegrating space as, e. g., $\beta = H\phi$, the authors maximize (3.4) under this additional constraint. The pseudo likelihood ratio test taking the difference between the log-likelihoods is χ^2-distributed.

3.2.2 The Johansen Vector Error Correction Model

Johansen developed vector error correction models for integrated VAR processes of orders one and two, see Johansen (1995). In the I(1) case he rewrites the equation of a VAR process

$$y_t \quad = \quad \sum_{i=1}^{p} a_i y_{t-i} + \Phi d_t + \varepsilon_t,$$

to the form

$$\Delta y_t \quad = \quad \alpha\beta' y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \Phi d_t + \varepsilon_t, \tag{3.5}$$

where α and β are full rank matrices in $\mathbb{R}^{s\times r}$ and $\Gamma_i \in \mathbb{R}^{s\times s}$ and ε_t is white noise. Here, $r < s$ is the cointegrating rank, i. e., the number of linearly independent cointegrating vectors. The columns of β span the cointegrating space. The parameters are estimated by projecting Δy_t and y_{t-1} on the lagged differences. Then, the pseudo likelihood ratio test statistic for the cointegrating rank uses a generalized eigenvalue problem of the residual matrices. The eigenvectors corresponding to the r largest eigenvalues are the estimates of β and, thus, of the cointegrating vectors. Hypotheses on the cointegrating space can be tested with pseudo likelihood ratio test, which are χ^2-distributed under the null hypotheses.
This approach is more restrictive than the state space framework, since not every process that has a state space representation is a VAR process. Conversely, every VAR process has a state space representation. Consider the VAR process

$$y_t \quad = \quad \sum_{i=1}^{p} a_i y_{t-i} + \varepsilon_t,$$

neglecting the deterministic process for simplicity. Choosing

$$\begin{aligned} A \quad &= \quad \begin{pmatrix} a_1 & a_2 & \dots & a_{p-1} & a_p \\ I_s & 0 & \dots & 0 & 0 \\ 0 & I_s & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_s & 0 \end{pmatrix}, \quad B = \begin{pmatrix} I_s \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad x_t = \begin{pmatrix} y_{t-1} \\ \vdots \\ y_{t-p} \end{pmatrix} \\ C \quad &= \quad \begin{pmatrix} a_1 & a_2 & \dots & a_p \end{pmatrix} \end{aligned}$$

we get (3.1). VARMA processes can be written as VAR(∞) processes, see, e.g., Saikkonen (1992). So the results of the estimator are still valid for VARMA processes for $p \to \infty$. An approximation with a large lag order p still leads to consistent estimators and tests.

3.2.3 The Approach of Stock and Watson

For the DOLS estimator Stock and Watson (1993) assume that the time series $\{y_t\}_{t\in\mathbb{Z}}$ has a partitioning $y_t = (y_{t,1}, y_{t,2})$ such that

$$\begin{aligned} \Delta y_{t,1} \quad &= \quad u_{t,1} \\ y_{t,2} \quad &= \quad \mu + \theta y_{t,1} + u_{t,2}, \end{aligned} \tag{3.6}$$

where $u_t := (u_{t,1}, u_{t,2})$ is a stationary, linearly regular process. θ cannot be estimated with the OLS estimator, because $y_{t,1}$ and $u_{t,2}$ are correlated. Thus, $u_{t,2}$ must be replaced by $u_{t,2} - \mathbb{E}(u_{t,2}|\{u_{t,1}\})$. $\mathbb{E}(u_{t,2}|\{u_{t,1}\})$ depends potentially on both future and past lags of $y_{t,1}$ and is thus of the form

$\mathbb{E}(u_{t,2}|\{u_{t,1}\}) = \sum_{j=-\infty}^{\infty} d_j \Delta y_{t-j,1}$. This is approximated by a finite sum $\sum_{j=\underline{q}}^{\overline{q}} d_j \Delta y_{t-j,1}$. Inserting this in the second equation of (3.6) and using $v_{t,2} := u_{t,2} - \mathbb{E}(u_{t,2}|\{u_{t,1}\})$

$$y_{t,2} \quad = \quad \mu + \theta y_{t,1} + \sum_{j=\underline{q}}^{\overline{q}} d_j \Delta y_{t-j,1} + v_{t,2}, \tag{3.7}$$

follows. Since $\{v_{t,2}\}_{t\in\mathbb{Z}}$ and $\{u_{t,1}\}_{t\in\mathbb{Z}}$ are uncorrelated, the likelihood for given $\Delta y_{-\underline{q},1}, \ldots, \Delta y_{T+\overline{q},1}$ can be divided into a factor depending on the marginal distribution of $\{u_{t,1}\}_{t\in\mathbb{Z}}$ and a factor depending on μ, θ, $d_{-\underline{q}}, \ldots, d_{\overline{q}}$ and the marginal distribution of $\{v_{t,2}\}_{t\in\mathbb{Z}}$. If there are no cross restrictions, θ can be estimated with a OLS estimator. The OLS estimator in (3.7) is called dynamic OLS (DOLS) estimator. Unlike the parametric approaches in the previous subsections the DOLS estimator only estimates the cointegrating space and not the other parameters in the model. Additionally, Stock and Watson (1993, Section 4) derive the asymptotic distribution of the DOLS estimator.
The rank tests used in King et al. (1991) are from Stock and Watson (1988, Section 3). Stock and Watson (1988) assume that $\{\Delta y_t\}_{t\in\mathbb{Z}}$ has a MA representation

$$\Delta y_t = \mu + \sum_{j=0}^{\infty} C_j \varepsilon_{t-j}$$

with $\sum_{j=1}^{\infty} jC_j < \infty$, $C_0 = I_s$ and $C(z) := \sum_{j=0}^{\infty} C_j z^j$. If y_t is cointegrated, then $C(1)$ has reduced rank $k < s$. Choosing a matrix $\beta \in \mathbb{R}^{s\times s-k}$, whose columns span the cointegrating space, and a matrix $\beta_\perp \in \mathbb{R}^{s\times k}$, whose columns span the orthocomplement of the cointegrating space, Stock and Watson (1988) regress $W_t := \beta_\perp' y_t$ on W_{t-1}. Under the null hypothesis of cointegrating rank $s-k$ the matrix $\tilde{\Phi} := \sum W_t W_{t-1} \left(\sum W_{t-1} W_{t-1}\right)^{-1}$ has k unit eigenvalues in the case without deterministic terms. Under the alternative of cointegrating rank $s-m$ with $m < k$ the real part of $k-m$ eigenvalues is smaller than one. Thus, ordering the eigenvalues $\lambda_1, \ldots, \lambda_k$ by the real parts in decreasing order, the null hypothesis $\mathcal{R}(\lambda_{m+1}) = 1$ is tested against the alternative $\mathcal{R}(\lambda_{m+1}) < 1$. The distribution under the null is not straightforward. Stock and Watson (1988) assume that ΔW_t has a finite VAR representation. For known VAR polynomial Ξ and known $\beta_\perp$ the distribution of λ_{m+1} is a known function of Brownian motions. If $\beta_\perp$ and Ξ are unknown, $\beta_\perp$ is estimated by estimating the principal components of y_t and Ξ is estimated by a VAR(p) regression of $\Delta\hat{W}_t$ onto $\hat{W}_{t-1}$, where $\hat{W}_t = \hat{\beta}_\perp' y_t$ and $\hat{\beta}_\perp$ is the estimator of $\beta_\perp$. Then, the distribution of $\hat{\lambda}_{m+1}$, the eigenvalue of $\sum \hat{W}_t \hat{W}_{t-1} \left(\sum \hat{W}_{t-1} \hat{W}_{t-1}\right)^{-1}$, is also a function of Brownian motions. For the case that a deterministic constant and trend is included in the model, $\hat{W}_t$ is replaced by $\hat{W}_t - \hat{\beta}_1 - \hat{\beta}_2 t$, where $\hat{\beta}_1$ and $\hat{\beta}_2$ are the OLS estimates of the regression of $\hat{W}_t$ on a constant and t.
For known cointegrating rank the asymptotic distribution of the DOLS estimator allows Wald-type tests for hypotheses on the cointegrating space.

3.3 Data Set

The original data set as described in King et al. (1991, Section III) is from the homepage of the authors. It contains seasonally adjusted quarterly data from 1949:1 to 1988:4. The data consist of the logarithms of per capita real consumption expenditures (c), per capita gross private domestic fixed investment (i), and per capita "private"gross national product (y), which is the gross national product without the government purchases, the money supply (m), the implicit price deflator of the private GNP (Δp) and the three-month US treasury bill rate (R). Since an extension of the data set is not available, an analogous data set from 1949:1 to 2018:4 is used, obtained from the site of the Federal Reserve Bank of St. Louis. It consists of seasonally adjusted quarterly observations of the logarithm of the US per capita real consumption expenditure c in 2012 dollars, the logarithm of the US per capita gross private domestic fixed investment i in 2012 dollars, the logarithm of the US per capita total gross national product y less real total government purchases in 2012 dollars, the logarithm of the US saving deposits per capita m in 2012 dollars, the logarithm of the implicit price deflator of the US private GNP Δp and the three-month US treasury bill rate R in percent.

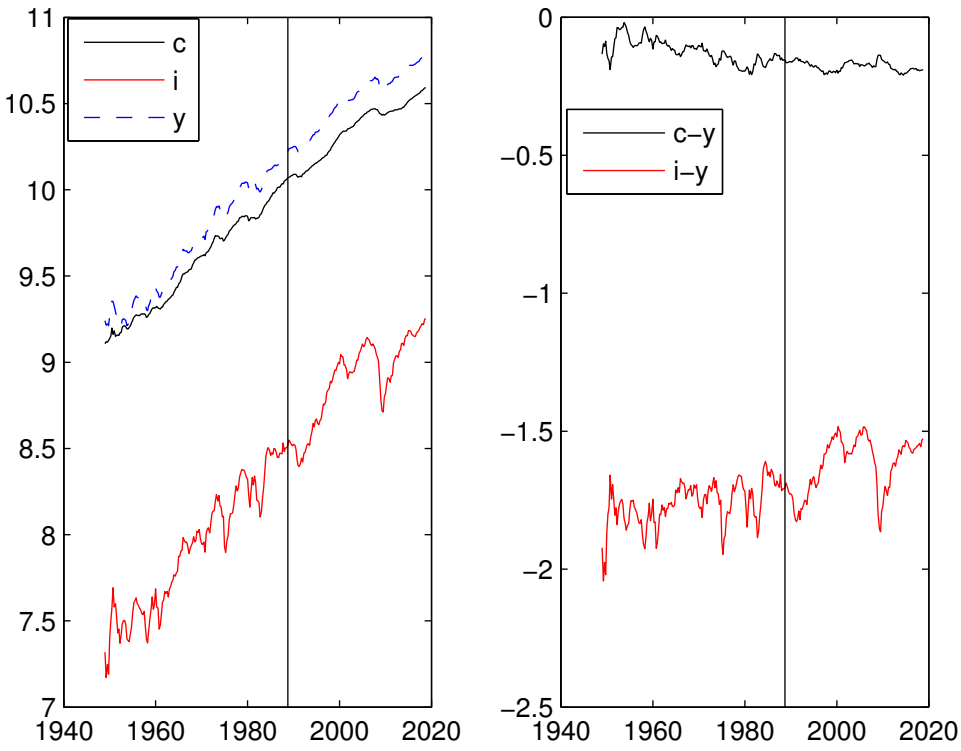

Figure 3.1: U.S. economic data and great ratios from 1949 to 2018

Figure 3.1 shows the variables c, i and y and the great ratios $c-y$ and $i-y$, which are stationary according to economic theory for the new data set.

3.4 Analysis of the Data Set of King et al. (1991)

3.4.1 Choice of System Parameters

For all methods some system parameters need to be set before beginning with the estimation. For the DOLS estimator the number of leads and lags is necessary, for the VECM the VAR order p of the system and for the state space framework the state dimension n. King et al. (1991) use $\overline{q} = \underline{q} = 5$, i. e., five leads and five lags. However, they do not explain this choice. To choose the VAR order of the VECM, we estimate systems of different lag lengths with a pseudo maximum likelihood estimator and compute the Bayesian information criterion (BIC). For the three-dimensional data set the BIC is minimal for $p = 2$. For the six-dimensional data set we do the same. Again, the BIC is minimal for $p = 2$. Table C.1 in the appendix contains the BIC for $p = 1, \ldots, 8$ for both data sets.
Similarly, to determine the state dimension, we compute the BIC for state space systems of order $n = 1, \ldots, 10$ and compare the BIC for the three-dimensional data set.[25] The BIC is minimal for $n = 3$. As for the three-dimensional data set we estimate models for different system orders and compute the BIC for the six-dimensional data set. We use system orders $n = 1, \ldots, 15$. The BIC is minimal for $n = 7$. Table C.2 in the appendix contains the BIC for the different system orders for both data sets.

3.4.2 Deterministic Components

Often, economic time series contain deterministic terms like a deterministic constant and a deterministic linear trend. For the economic analysis it is important to find out which of these deterministic components are present. Johansen (1995) distinguishes five cases of deterministic components:

1. The VECM equation (3.5) contains an unrestricted linear trend and constant. This leads to a linear trend in the cointegrating space and a quadratic trend in its orthocomplement.

2. The VECM equation (3.5) contains an unrestricted constant and a linear trend restricted to the cointegrating space. This leads to a linear trend in all components.

3. The VECM equation (3.5) contains a unrestricted constant. This leads to a constant in the cointegrating space and a linear trend in its orthocomplement.

4. The VECM equation (3.5) contains a constant restricted to the cointegrating space. This leads to a constant in all components.

5. The VECM equation (3.5) does not contain any deterministic components.

These five cases can be easily implemented in the state space framework. For the first case, include a deterministic linear trend in the state equation, i. e., the second equation of (3.1). For the second case, include a deterministic linear trend in the output equation, i. e., the second equation of (3.1). For the third case include a deterministic constant in the state equation and for the fourth case include a deterministic constant in the output equation.[26]
Figure 3.1 shows a linear trend in the variables, but not in the great ratios, which suggests that case three might describe our data best. King et al. (1991) sometimes include a constant and a trend and sometimes only a constant in their analysis. Testing hypotheses on the deterministic terms is complicated. In the cases two and four the restrictions depend on the dimension of the cointegrating space. This dimension is not known and the tests to determine it depend on the deterministic components. Both for the VECM and the state space framework we use the cointegrating rank predicted by economic theory and use pseudo likelihood ratio tests for the determination

[25] The `matlab` function for this is `optimize_estimate`.
[26] For the first case choose the option {`'AR',[0,2]`} for the variable `deter` in `optimize_estimate`, for the second {`'MA',[0,2]`}, for the third {`'AR',[0,1]`}, for the fourth {`'MA',[0,1]`} and for the fifth {`'AR',[0,0]`}.

Table 3.1: Test statistics and p-values of the hypotheses on the deterministic terms for VECM and state space framework for the three- and the six dimensional data set until 1988

Null, alternative	VECM: test statistic (p-value)	state space: test statistic (p-value)
	$s = 3$	
case 3, case 2	0.53 (0.767)	0.21 (0.900)
case 4, case 3	12.06 (< 0.001)	12.67 (< 0.001)
	$s = 6$	
case 3, case 2	4.35 (0.227)	3.47 (0.324)
case 4, case 3	19.13 (< 0.001)	28.19 (< 0.001)

of the deterministic components. Since Figure 3.1 clearly shows that there is no quadratic trend and there are deterministic components present, we neglect cases one and five. We test the null hypotheses that there is only a unrestricted constant against the alternative that there is a trend restricted to the cointegrating space. If the hypothesis is rejected, we conclude that there is a trend restricted to the cointegrating space. Otherwise, we test the null hypothesis that there is a constant restricted to the cointegrating space against the alternative of an unrestricted constant. For the three-dimensional data set consisting of c, i and y we use a cointegrating rank of two. Thus, for the first test the linear trend has two coefficients under the alternative, which are zero under the null. Consequently, the test statistic is χ^2_2 distributed under the null. For the second test, the constant has three coefficients under the alternative and only two under the null. Thus, under the null one coefficient is zero and the test statistic is χ^2_1 distributed under the null. Table 3.1 contains the test statistics and their p-values for both the VECM and the state space framework.[27] With both methods the first hypothesis is not rejected at 5%-level, but the second hypothesis is rejected. Thus, we use the third case for the modeling, which is consistent with the observations in Figure 3.1.
For the six-dimensional data set we use a three-dimensional cointegrating space. Thus, the test statistics are χ^2_3-distributed under the null for both hypotheses. The test statistics and their p-values for both the VECM and the state space framework are also shown in Table 3.1. Again, neither the pseudo likelihood ratio test using the VECM nor the one using the state space framework reject the null hypothesis of the first test at 5%-level. Again, the null hypothesis of the second test is rejected. Thus, we use the third case in both models.

3.4.3 Determination of the Cointegrating Rank

A question of economic interest is the determination of the cointegrating vectors. The first step for this is the determination of the cointegrating rank, i. e., the number of linearly independent cointegrating vectors. Johansen (1995, Chapter 11) developed a test for the cointegrating rank. This test does not test directly hypotheses on r, but the number of common trends, i. e., $s - r$. Using this procedure, we first test the null hypothesis of cointegrating rank $s - r = s$, i. e., $r = 0$ against the alternative of at least one cointegrating relation. Then, we test the null hypothesis $r \leq 1$ against the alternative $r \geq 2$. We continue this procedure until one hypothesis is rejected or until we test the hypothesis $r \leq s - 1$ against the alternative $r = s$. The distributions of the test statistics under the null hypotheses depend on the deterministic terms. Due to the structure of the tests the distribution of hypotheses as, e. g., $r \leq 1$ depends on s. The critical values for the third case are shown in Table 3.2. The rank tests of Matuschek et al. (2020) test the same hypotheses as the ones of Johansen (1995). The authors show that the test statistics of these tests also have the same distribution under the null hypotheses, such that the critical values of Table 3.2 are also asymptotically valid for these tests. As explained in Section 3.2.3, the rank tests of Stock and Watson (1988) test the null hypothesis of k common trends against the alternative of $m < k$

[27] After having calculated the first case with `optimize_estimate`, we calculate the other cases with `optimize_ione` and compare the critical values to calculate the test statistic.

Table 3.2: Critical values for the hypotheses on the cointegrating rank for different levels of significance α in the model with a unrestricted constant

Hypothesis	$H_0: s-r=6$	$H_0: s-r\geq 5$	$H_0: s-r\geq 4$	$H_0: s-r\geq 3$
$\alpha=0.1$	89.37	64.74	43.84	26.70
$\alpha=0.05$	93.92	68.68	47.21	29.38
$\alpha=0.01$	102.95	76.37	53.91	34.87

Hypothesis	$H_0: s-r\geq 2$	$H_0: s-r\geq 1$
$\alpha=0.1$	13.31	2.71
$\alpha=0.05$	15.35	3.84
$\alpha=0.01$	19.69	6.64

Table 3.3: p-values for the rank tests of Stock and Watson (1988) and test statistics for the rank tests in the VECM and the state space framework for the three-dimensional data set until 1988

Null, alternative	p-values of Stock Watson
$r=0,\ r=1$	0.21
$r=0,\ r=2$	<0.01

Null, alternative	test statistic VECM	test statistic state space
$r=0,\ r\geq 1$	51.08	51.48
$r\leq 1,\ r\geq 2$	16.53	1.22
$r\leq 2,\ r=3$	23.05	0.96

common trends. Table 3.3 shows the results of the rank tests for the three-dimensional data set.[28] The rank test of Stock and Watson (1988) rejects the null when tested against the alternative of cointegrating rank two. The tests of Johansen (1995) and Matuschek et al. (2020) both reject the hypotheses $r=0$ and $r\leq 1$ at 5%-level. The hypothesis $r\leq 2$ is not rejected at 5%-level. Thus, all three tests find a cointegrating rank of two, which is consistent with economic theory. Another possibility to examine the cointegrating rank in the state space framework is to look at the eigenvalues of the matrix A. As discussed above the number of common trends, i. e., $s-r$ coincides with the eigenvalues of A which are one. Additionally, the eigenvalues of $\underline{A}:=A-BC$ are of interest, because eigenvalues of $\underline{A}$ near the unit circle complicate the analysis and because, as discussed above, the strict minimum phase condition is necessary for pseudo maximum likelihood estimation. For this reason, we estimate the system matrices without imposing restrictions on the cointegrating space and examine the eigenvalues of A and $\underline{A}$, shown in Table 3.4. One eigenvalue of A is very close to one and the eigenvalues of $\underline{A}$ are not near the unit circle. This fits to the result of the rank tests that the cointegrating rank is two and there is thus one common trend.
For the six-dimensional data set King et al. (1991) do not show all the tests. They just state that the null hypothesis of cointegrating rank $r=6$ is rejected against the alternative $r=3$. The test statistics for the VECM and the state space framework are shown in Table 3.5. The test of Johansen (1995) rejects the hypothesis $r\leq 3$ at 5%-, but not at 1%-level. The hypothesis $r\leq 4$ then cannot be rejected at 10%-level. The test of Matuschek et al. (2020) rejects the hypothesis $r\leq 3$ at 1%-level. The hypothesis $r\leq 4$ is rejected at 5%, but not at 1%-level. The hypothesis $r\leq 5$ cannot be rejected at 5%-level. Thus, using the 5%-level King et al. (1991) find a cointegrating rank of three, the test of Johansen (1995) finds a cointegrating rank of four and the test of Matuschek et al. (2020) a cointegrating rank of five. Again, we take a look at the eigenvalues of A and $\underline{A}$ for the estimator without restrictions on the cointegrating rank for the state space framework, also shown in Table 3.4. One eigenvalue of A is 0.994 and thus very close to one, a second eigenvalue is 0.945. This fits to the results of the rank tests, since they find a cointegrating rank of five and thus only one common trend. The absolute value of two eigenvalues of $\underline{A}$ is 1.024. Thus, the minimum phase condition is not fulfilled.

[28] For the rank test we use the `matlab` function `ranktest_real`.

Table 3.4: Eigenvalues of the matrices A and $\underline{A}$ and their absolute values for the state space system estimated for the three- and six-dimensional data set until 1988

Eigenvalues of A	absolute value	Eigenvalues of $\underline{A}$	absolute value
		$s = 3$	
0.997	0.997	-0.401	0.401
0.854+0.087i	0.858	0.005	0.005
0.854-0.087i	0.858	0.232	0.232
		s=6	
0.467	0.467	0.945+0.394i	1.024
0.873+0.304i	0.924	0.945-0.394i	1.024
0.873-0.304i	0.924	-0.729	0.729
0.883+0.103i	0.889	0.289	0.289
0.883-0.103i	0.889	-0.414	0.414
0.937	0.937	-0.181	0.181
0.994	0.994	-0.030	0.030

Table 3.5: Test statistics for the rank tests in the VECM and the state space framework for the six-dimensional data set until 1988

Null, alternative	test statistic VECM	test statistic state space
$r = 0$, $r \geq 1$	151.85	349.63
$r \leq 1$, $r \geq 2$	98.68	159.13
$r \leq 2$, $r \geq 3$	54.27	88.57
$r \leq 3$, $r \geq 4$	30.88	51.20
$r \leq 4$, $r \geq 5$	11.04	18.16
$r \leq 5$, $r = 6$	1.28	2.01

3.4.4 Estimation of the Cointegrating Space

It is of economic interest to find cointegrating relations in economic time series. Thus, an estimator of the cointegrating space is useful. After having determined the cointegrating rank the DOLS estimator and the pseudo maximum likelihood estimators for the VECM and the state space framework can estimate the cointegrating space. To make the basis of the cointegrating space unique, we impose the condition that the upper $r \times r$-block is the identity matrix.
For the three-dimensional data set all rank tests found a cointegrating rank of two. Economic theory predicts a cointegrating space spanned by the two great ratios $\begin{pmatrix}1 & 0 & -1\end{pmatrix}'$ and $\begin{pmatrix}0 & 1 & -1\end{pmatrix}'$. We test with all three methods whether this is the cointegrating space.[29] Table 3.6 contains the results. While the estimates are not exactly identical, they are similar and the hypothesis that the cointegrating space coincides with the theoretic prediction is not rejected at 5%-level for all three methods. It is, however, notable that the p-values for the tests using the VECM and the state space framework are much higher than the one using DOLS.
For the six-dimensional data set the results of the rank tests were different for the different methods. Thus, King et al. (1991) estimate a three-dimensional cointegrating space, we estimate a four-dimensional cointegrating space using the VECM and a five-dimensional cointegrating space using the state space framework. Table 3.7 contains the estimates. We notice that the standard errors of the estimates for the state space framework are very large, in particular for the last entry.

[29] We use the `matlab` function `optimizeconstr` to estimate the system, where the cointegrating space is spanned by these vectors and calculate the test statistic by taking the difference between this log-likelihood and the one of the unrestricted estimator, the standard errors are computed with the function `standarderror`.

Table 3.6: Estimates of the cointegrating space with standard errors in brackets, test statistic of the test whether the cointegrating space is spanned by $(1, 0 - 1)'$ and $(0, 1, -1)'$ and p-value of the test for the three-dimensional data set until 1988

method	DOLS	VECM	state space
c	$\begin{pmatrix}1 & 0\end{pmatrix}$	$\begin{pmatrix}1 & 0\end{pmatrix}$	$\begin{pmatrix}1 & 0\end{pmatrix}$
i	$\begin{pmatrix}0 & 1\end{pmatrix}$	$\begin{pmatrix}0 & 1\end{pmatrix}$	$\begin{pmatrix}0 & 1\end{pmatrix}$
y	$\begin{pmatrix}-1.058 & -1.004\\ (0.026) & (0.038)\end{pmatrix}$	$\begin{pmatrix}-1.113 & -1.144\\ (0.101) & (0.188)\end{pmatrix}$	$\begin{pmatrix}-1.115 & -1.064\\ (0.091) & (0.201)\end{pmatrix}$
test statistic	4.96	1.06	1.19
p-value	0.08	0.59	0.55

Table 3.7: Estimates of the cointegrating space for the different methods with standard errors in brackets for the six-dimensional data set until 1988

method	DOLS			VECM			
c	1	0	0	1	0	0	0
i	0	1	0	0	1	0	0
$m-p$	0	0	1	0	0	1	0
y	−1.118	−1.120	−1.152	0	0	0	1
	(0.050)	(0.083)	(0.063)				
R	0.004	0.002	0.009	0.053	0.069	0.064	0.048
	(0.003)	(0.005)	(0.004)	(0.408)	(0.436)	(0.436)	(0.374)
Δp	0.004	0.006	0.002	−0.062	−0.071	−0.055	−0.061
	(0.003)	(0.004)	(0.003)	(0.329)	(0.351)	(0.351)	(0.301)

	state space				
c	1	0	0	0	0
i	0	1	0	0	0
$m-p$	0	0	1	0	0
y	0	0	0	1	0
R	0	0	0	0	1
Δp	−0.065	−0.073	−0.086	−0.063	0.646
	(1.301)	(1.134)	(0.876)	(1.224)	(37.175)

3.4.5 Testing Hypotheses on the Cointegrating Space

For the three-dimensional data set economic theory predicts that the cointegrating space is spanned by the two great ratios $c-y$ and $i-y$. For the six-dimensional data set several hypotheses can be considered. One hypothesis, which is not tested by King et al. (1991), is that the cointegrating space is spanned by the great ratios $c-y$, $i-y$ and $m-p-y$. We test this hypothesis in the first line of Table 3.8 for the VECM and the state space framework. Instead, King et al. (1991) test the hypothesis that the cointegrating vectors are the first two great ratios and a long-run money-demand relation (line two of Table 3.8). They also test the hypothesis that additionally to these three cointegrating vectors there are stationary real rates (line three of Table 3.8). Line four of Table 3.8 tests the hypothesis that the great ratios and real rates are cointegrated, combined with the money-demand cointegrating vector. Line five of Table 3.8 contains the tests for the stationary velocity model, which allows cointegration between the great ratios and real rates. The hypotheses considered in King et al. (1991) are for a three- or four-dimensional cointegrating space. In the models, where the cointegrating rank is larger than the one considered in the hypothesis, we test the hypothesis that the proposed space is a subspace of the cointegrating space.[30] Table 3.8 shows the hypotheses and their p-values. The hypotheses that the great ratios are contained in the cointegrating space is rejected by the VECM, but not by state space framework. The hypothesis that the first two great ratios and a long-run money-demand relation are contained in the cointegrating space and the hypothesis that the great ratios and real rates are cointegrated, combined with the money-demand cointegrating vector are not rejected at 5%-level for all three models. The hypothesis of the stationary velocity model is rejected by King et al. (1991) at 5%-level, but it is not rejected using the VECM or the state space framework. The hypothesis that additional to the first two great ratios and a long-run money-demand relation there are stationary real rates is rejected both by King et al. (1991) and using the VECM. For the state space framework it cannot be tested, because the cointegrating space is five-dimensional and for every five-dimensional cointegrating space the parameters β_y and β_R can be chosen in such a way that the space of second hypothesis is a subspace of the cointegrating space.

[30] For this we use the `matlab` function `optimize_hyptest`.

Table 3.8: Hypotheses on the cointegrating space and their p-values for the six-dimensional data set until 1988

Hypothesis	p-value King et al. (1991)	p-value VECM
$(c-y)$, $(i-y)$, $(m-p-y)$		< 0.01
$(c-y)$, $(i-y)$, $m-p-\beta_y y+\beta_R R$	0.08	0.12
$(c-y)$, $(i-y)$,$m-p-\beta_y y+\beta_R R$, $R-\Delta p$	< 0.01	< 0.01
$(c-y)-\varphi_1(R-\Delta p)$, $(i-y)-\varphi_2(R-\Delta p)$, $m-p-\beta_y+\beta_R R$	0.18	0.27
$(c-y)-\varphi_1(R-\Delta p)$, $(i-y)-\varphi_2(R-\Delta p)$, $m-p-y$	< 0.01	0.10

Hypothesis	p-value state space
$(c-y)$, $(i-y)$, $(m-p-y)$	0.55
$(c-y)$, $(i-y)$, $m-p-\beta_y y+\beta_R R$	0.59
$(c-y)-\varphi_1(R-\Delta p)$, $(i-y)-\varphi_2(R-\Delta p)$, $m-p-\beta_y+\beta_R R$	0.43
$(c-y)-\varphi_1(R-\Delta p)$, $(i-y)-\varphi_2(R-\Delta p)$, $m-p-y$	0.76

3.5 Full Sample Analysis

In this section we repeat the analysis from the last section with the larger data set until 2018 for the VECM and the state space framework. We then compare the results obtained with the two data sets.

3.5.1 Choice of System Parameters

Again, we start by determining the VAR order for the VECMs and the state dimension of the state space systems.
Again, we estimate VECMs of different VAR orders for the three-dimensional data set and compare the BIC. As for the data set until 1988, the BIC is minimal for $p = 2$. For the six-dimensional data set we do the same. This time the BIC is minimal for $p = 1$. Thus, we use the VAR order $p = 1$ for the analysis of the six-dimensional data set. Table C.1 in the appendix contains the BIC for $p = 1, \dots, 8$ for both data sets.
Similarly, to determine the state dimension we compute the BIC for state space systems of order $n = 1, \dots, 10$ and compare the BIC for the three-dimensional data set. This time the BIC is minimal for $n = 5$. Thus, compared to the data set until 1988 the state dimension increases. As for the three-dimensional data set we estimate models for different system orders for the six-dimensional data set and compute the BIC. We use system orders $n = 1, \dots, 15$. The BIC is minimal for $n = 8$. Thus, as for the three-dimensional data set, the state dimension increases compared to the data set until 1988. Table C.2 in the appendix contains the BIC for the different state dimensions for both data sets.

3.5.2 Deterministic Components

As for the data set until 1988, we use the cointegrating rank predicted by economic theory to estimate a model for both the VECM and the state space framework and then use pseudo likelihood ratio tests for the determination of the deterministic components. Again, we neglect cases one and five since Figure 3.1 clearly shows that there is no quadratic trend and there are deterministic components present. We test the null hypothesis that there is only an unrestricted constant against the alternative that there is a trend restricted to the cointegrating space. If the hypothesis is rejected, we conclude that there is a trend restricted to the cointegrating space. Otherwise, we test the null hypothesis that there is a constant restricted to the cointegrating space against the alternative of an unrestricted constant.
For the three-dimensional data set we use a cointegrating rank of two. Thus, the test statistic

Table 3.9: Test statistics and p-values of the hypotheses on the deterministic terms for VECM and state space framework for the three- and six-dimensional data set from 1949 to 2018

Null, alternative	VECM: test statistic (p-value)	state space: test statistic (p-value)
	$s = 3$	
case 3, case 2	3.61 (0.307)	10.55 (0.014)
case 4, case 3	8.42 (0.038)	13.14 (0.004)
	$s = 6$	
case 3, case 2	35.98 (< 0.001)	74.10 (< 0.001)
case 4, case 3	21.62 (< 0.001)	11.75 (0.008)

for the first test is χ^2_2 distributed under the null and the test statistic of the second test is χ^2_1 distributed under the null. Table 3.9 contains the test statistics and their p-values for both the VECM and the state space framework. As in the case of the data set until 1988, the first hypothesis is not rejected at 5%-level, but the second hypothesis is rejected for the VECM. In the state space framework, however, the first hypothesis is rejected at 5%-level. Thus, we use an unrestricted constant for the VECM and a restricted linear trend for the state space framework.
For the six-dimensional data set we use a three-dimensional cointegrating space. Thus, the test statistics are χ^2_3-distributed under the null for both hypotheses. The test statistics and their p-values for both the VECM and the state space framework are also in Table 3.9. This time both methods reject the first hypothesis at 5%-level. We thus use the second case, a linear trend restricted to the cointegrating space for the modeling.

3.5.3 Determination of the Cointegrating Rank

As in the case of the data set until 1988, we first test the null hypothesis of cointegrating rank $s - r = s$, i. e., $r = 0$ against the alternative of at least one cointegrating relation. Then, we test the null hypothesis $r \leq 1$ against the alternative $r \geq 2$. We continue this procedure until one hypothesis is not rejected or until we test the hypothesis $r \leq s - 1$ against the alternative $r = s$. Since the distributions of the test statistics under the null hypotheses depends on the deterministic terms, the critical values of the tests for the state space framework in the three-dimensional case and for the six-dimensional case for both models are not those of Table 3.2, but those shown in Table 3.10. Table 3.11 shows the results of the rank tests for the three-dimensional data set. Both tests reject the hypotheses $r = 0$ and $r \leq 1$ at 5%-level. The hypothesis $r \leq 2$ is not rejected at 5%-level. Thus, both tests find a cointegrating rank of two, which is consistent with economic theory. Again, we estimate the system matrices without imposing restrictions on the cointegrating space in the state space framework and look at the eigenvalues of A and $\underline{A}$, shown in Table 3.12. One eigenvalue of A is very close to one and the eigenvalues of $\underline{A}$ are not near the unit circle. This

Table 3.10: Critical values for the hypotheses on the cointegrating rank for different levels of significance α in the model with a linear trend restricted to the cointegrating space

Hypothesis	$H_0 : s - r = 6$	$H_0 : s - r \geq 5$	$H_0 : s - r \geq 4$	$H_0 : s - r \geq 3$
$\alpha = 0.1$	110.00	82.68	58.96	39.08
$\alpha = 0.05$	114.96	86.96	62.61	42.20
$\alpha = 0.01$	124.61	95.38	70.22	48.59

Hypothesis	$H_0 : s - r \geq 2$	$H_0 : s - r \geq 1$
$\alpha = 0.1$	22.95	10.56
$\alpha = 0.05$	25.47	12.39
$\alpha = 0.01$	30.65	16.39

Table 3.11: Test statistics for the rank tests in the VECM and the state space framework for the three- and the six-dimensional data set from 1949 to 2018

Null, alternative	VECM, $s = 3$	state space, $s = 3$	VECM, $s = 6$	state space, $s = 6$
$r = 0$, $r \geq 1$	55.84	77.68	344.94	548.63
$r \leq 1$, $r \geq 2$	16.16	35.49	127.09	192.14
$r \leq 2$, $r = 3$	2.00	2.95	59.18	79.28
$r \leq 3$, $r \geq 4$			28.66	28.92
$r \leq 4$, $r \geq 5$			12.95	11.97
$r \leq 5$, $r = 6$			4.68	2.78

Table 3.12: Eigenvalues of the matrices A and $\underline{A}$ and their absolute values for the state space system estimated for the three- and the six-dimensional data set from 1949 to 2018

Eigenvalues of A	absolute value	Eigenvalues of $\underline{A}$	absolute value
		$s = 3$	
0.973	0.973	-0.865	0.865
0.913+0.098i	0.918	0.394+0.244i	0.463
0.913-0.098i	0.918	0.394-0.244i	0.463
-0.897	0.897	-0.115	0.115
0.806	0.806	-0.003	0.003
		$s = 6$	
0.452	0452	-0.578	0.578
0.811+0.274i	0.856	0.608+0.222i	0.647
0.811-0.274i	0.856	0.608-0.222i	0.647
0.743	0.743	0.355+0.337i	0.489
0.936+0.070i	0.939	0.355-0.337i	0.489
0.936-0.070i	0.939	-0.232+0.041i	0.236
0.995+0.005i	0.995	-0.232-0.041i	0.236
0.995-0.005i	0.995	-0.048	0.048

fits the result of the rank tests that the cointegrating rank is two and there is thus one common trend.
For the six-dimensional data set the test statistics for the VECM and the state space framework are also shown in Table 3.11. The test of Johansen (1995) rejects the hypothesis $r \leq 2$ at 10%-, but not at 5%-level. The hypothesis $r \leq 3$ then cannot be rejected at 1%-level. The test of Matuschek et al. (2020) rejects the hypothesis $r \leq 2$ at 1%-level. The hypothesis $r \leq 3$ cannot be rejected at 5%-level. Thus, using the 5%-level the tests of Johansen (1995) find a cointegrating rank of two and the tests of Matuschek et al. (2020) a cointegrating rank of three. In both cases the results differ from the ones obtained for the data set until 1988. Again, we take a look at the eigenvalues of the estimates of A and $\underline{A}$ without restrictions on the cointegrating rank for the state space framework. The eigenvalues are also in Table 3.12. The two eigenvalues $0.995 \pm 0.005i$ are close to one. This rather indicates two common trends and thus a cointegrating rank of four. This time the strict minimum phase condition is fulfilled.

3.5.4 Estimation of the Cointegrating Space

Again, we estimate the cointegrating space with a pseudo maximum likelihood estimator for both the VECM and the state space framework. For the three-dimensional data set both rank tests found a cointegrating rank of two, which is consistent with economic theory. As for the data set until 1988, we test for both estimates whether $\begin{pmatrix} 1 & 0 & -1 \end{pmatrix}'$ and $\begin{pmatrix} 0 & 1 & -1 \end{pmatrix}'$ span the cointegrating space. Table 3.13 contains the results. The estimates differ from those for the data set until 1988. The hypothesis cannot be rejected for the VECM at 5%-level, although the p-value is lower than the one for the data set until 1988. For the state space framework the hypothesis is rejected at 5%-level, unlike in the analysis for the data set until 1988.
For the six-dimensional data set the results of the rank tests were different for the VECM and the state space framework. Consequently, we estimate a two-dimensional cointegrating space for the VECM and a three-dimensional cointegrating space for the state space framework. The estimates are also shown in Table 3.13. A comparison to the results for the data set until 1988 is not possible, because the dimensions of the cointegrating spaces differ. It is, however, notable that the standard errors of the estimates obtained with the VECM are quite large.

Table 3.13: Estimates of the cointegrating space with standard errors in brackets, test statistic of the test whether the cointegrating space is spanned by $(1, 0, -1)'$ and $(0, 1, -1)'$ and p-value of the test for the three-dimensional data set from 1949 to 2018 and estimates for the cointegrating space for the six-dimensional data set from 1949 to 2018

method	VECM	state space
	$s = 3$	
c	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -0.753 & -0.964 \\ (0.071) & (0.144) \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -0.701 & -1.066 \\ (0.043) & (0.145) \end{pmatrix}$
i		
y		
test statistic	1.69	15.50
p-value	0.4296	< 0.001
	$s = 6$	
c	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -0.068 & -5.088 \\ (0.274) & (4.690) \\ -1.087 & -3.776 \\ (0.152) & (2.599) \\ 0.010 & -0.159 \\ (0.008) & (0.141) \\ -0.010 & 0.502 \\ (0.025) & (0.420) \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1.059 & -1.728 & 0.897 \\ (0.072) & (0.170) & (0.239) \\ 0.009 & 0.013 & 0.024 \\ (0.001) & (0.003) & (0.005) \\ -0.004 & 0.005 & -0.079 \\ (0.001) & (0.003) & (0.005) \end{pmatrix}$
i		
$m - p$		
y		
R		
Δp		

3.5.5 Testing Hypotheses on the Cointegrating Space

For the state space framework we test the same hypotheses on the cointegrating space as for the data set until 1988. Table 3.14 shows the hypotheses and their p-values. All hypotheses are rejected at 1%-level. This is very different from the results for the first data set. For the VECM we have a two-dimensional cointegrating space. Thus, it does not make sense to test the hypotheses directly. We could test the hypotheses that the two-dimensional cointegrating space is a subspace of the spaces suggested by the hypotheses of King et al. (1991). This, however, would not help answering the question which of the cointegrating vectors suggested by the hypotheses are in fact cointegrating vectors. It makes thus more sense to test which of the cointegrating vectors suggested by the hypotheses are in the two-dimensional cointegrating space. Table 3.15 contains the p-values of the tests. All hypotheses are rejected at 1%-level. This may be a hint that the cointegrating rank estimated by the model is wrong.
All in all, we get very different results for the two data sets.

Table 3.14: Hypotheses on the cointegrating space and their p-values for the six-dimensional data set from 1949 to 2018

Hypothesis	p-value state space
$(c - y)$, $(i - y)$, $(m - p - y)$	< 0.001
$(c - y)$, $(i - y)$, $m - p - \beta_y y + \beta_R R$	< 0.001
$(c - y)$, $(i - y)$,$m - p - \beta_y y + \beta_R R$, $R - \Delta p$	< 0.001
$(c - y) - \varphi_1(R - \Delta p)$, $(i - y) - \varphi_2(R - \Delta p)$, $m - p - \beta_y + \beta_R R$	< 0.001
$(c - y) - \varphi_1(R - \Delta p)$, $(i - y) - \varphi_2(R - \Delta p)$, $m - p - y$	0.009

Table 3.15: p-values of the hypotheses that the following vectors are contained in the two-dimensional cointegrating space estimated with the VECM for the six-dimensional data set from 1949 to 2018

Hypothesis	p-value
$c - y$	< 0.001
$i - y$	< 0.001
$m - p - y$	< 0.001
$m - p - \beta_y y + \beta_R R$	< 0.001
$R - \Delta p$	< 0.001
$(c - y) - \varphi_1(R - \Delta p)$	< 0.001
$(i - y) - \varphi_2(R - \Delta p)$	< 0.001
$m - p - y$	< 0.001

3.6 Analysis of the Data Set From 1989 to 2018

Since the results for the full data set and the subset until 1988 were so different, we analyze the subset from 1989 to 2018 separately and compare the results to those obtained in the previous two sections.

3.6.1 Choice of System Parameters

Again, we start by determining the VAR order for the VECMs and the state dimension of the state space systems.
Again, we estimate VECMs of different VAR orders and compare the BIC. Unlike for the other two data sets the BIC is minimal for $p = 1$ in the three-dimensional case. For the six-dimensional data set we do the same. The BIC is also minimal for $p = 1$, which coincides with the results for the full data set, but differs from the VAR order for the data set until 1988. Table C.1 in the appendix contains the BIC for $p = 1, \dots, 8$ for both data sets.
Similarly, to determine the state dimension we compute the BIC for state space systems of order $n = 1, \dots, 10$ and compare the BIC for the three-dimensional data set. The BIC is minimal for $n = 3$. This coincides with the state dimension obtained for the data set until 1988, but differs from the dimension for the full data set. As for the three-dimensional data set we estimate models for different system orders for the six-dimensional data set and compute the BIC. We use system orders $n = 1, \dots, 15$. This time the BIC is minimal for $n = 5$. This is lower than the system orders for the other two six-dimensional data sets. Table C.2 in the appendix contains the BIC for the different system orders for both data sets.

3.6.2 Deterministic Components

As for the other data sets, we use the cointegrating rank predicted by economic theory to estimate a model for both the VECM and the state space framework and then use pseudo likelihood ratio tests for the determination of the deterministic components. Again, we neglect cases one and five since Figure 3.1 clearly shows that there is no quadratic trend and there are deterministic components present. We test the null hypotheses that there is only a unrestricted constant against the alternative that there is a trend restricted to the cointegrating space. If the hypothesis is rejected, we conclude that there is a trend restricted to the cointegrating space. Otherwise, we test the null hypothesis that there is a constant restricted to the cointegrating space against the alternative of an unrestricted constant.
For the three-dimensional data set we use a cointegrating rank of two. Thus, the test statistic for the first test is χ^2_2 distributed under the null and the test statistic of the second test is χ^2_1 distributed under the null. Table 3.16 contains the test statistics and their p-values for both the VECM and the state space framework. In both cases the first hypothesis is rejected at 5%-level.

Table 3.16: Test statistics and p-values of the hypotheses on the deterministic terms for VECM and state space framework for the three- and the six-dimensional data set from 1989 to 2018

Null, alternative	VECM: test statistic (p-value)	state space: test statistic (p-value)
	$s = 3$	
case 3, case 2	13.91 (< 0.001)	10.61 (0.005)
case 4, case 3	3.96 (0.047)	4.01 (0.045)
	$s = 6$	
case 3, case 2	3.21 (0.360)	52.06 (< 0.001)
case 4, case 3	25.10 (< 0.001)	409.18 (< 0.001)

For the state space framework this coincides with the result for the full data set, for the VECM this is different from the results of both data sets. Thus, we use a restricted deterministic linear trend for both the VECM and the state space framework.
For the six-dimensional data set we use again a three-dimensional cointegrating space. Thus, the test statistics are χ^2_3-distributed under the null for both hypotheses. Table 3.16 also contains the test statistics and their p-values for both the VECM and the state space framework. The state space framework rejects the first hypothesis at 5%-level, as in the case of the full data set. The VECM cannot reject the first hypothesis, but rejects the second at 5% level, as in the case of the data set until 1988. Thus, we use a restricted linear trend in the state space model and an unrestricted constant in the VECM.

3.6.3 Determination of the Cointegrating Rank

We use the same procedure for the rank tests as for the other two data sets. Due to the choice of the deterministic terms in the previous subsection, we use the critical values of Table 3.2 for the VECM of the six-dimensional data set and those of Table 3.10 for the other models. Table 3.17 shows the results of the rank tests for the three-dimensional data set. Both tests reject the hypotheses $r = 0$ at 5%-level. The hypothesis $r \leq 1$ is not rejected at 10%-level. Thus, both tests find a cointegrating rank of one. This is different from the other two data sets, where we have found a cointegrating rank of two with both methods. Again, we estimate the system matrices without imposing restrictions on the cointegrating space in the state space framework and look at the eigenvalues of A and $\underline{A}$, shown in Table 3.18. One pair of complex conjugated eigenvalues of A is very close to one and the eigenvalues of $\underline{A}$ are not near the unit circle. This fits the result of the rank tests that the cointegrating rank is one and there are thus two common trends.
The test statistics for the VECM and the state space framework for the six-dimensional data set are also shown in Table 3.17. The test of Johansen (1995) rejects the hypothesis $r \leq 2$ at 5%-, but not at 1%-level. The hypothesis $r \leq 3$ then cannot be rejected at 10%-level. The test of Matuschek et al. (2020) rejects the hypothesis $r \leq 2$ at 1%-level. The hypothesis $r \leq 3$ cannot be rejected at 10%-level. Thus, using the 5%-level, both tests find a cointegrating rank of three. The

Table 3.17: Test statistics for the rank tests in the VECM and the state space framework for the three- and the six-dimensional data set from 1989 to 2018

Null, alternative	VECM, $s = 3$	state space, $s = 3$	VECM, $s = 6$	state space, $s = 6$
$r = 0, r \geq 1$	59.00	53.78	201.55	592.64
$r \leq 1, r \geq 2$	14.32	11.32	119.66	166.11
$r \leq 2, r = 3$	4.73	3.54	53.50	75.25
$r \leq 3, r \geq 4$			18.08	19.72
$r \leq 4, r \geq 5$			6.04	6.38
$r \leq 5, r = 6$			0.52	0.56

Table 3.18: Eigenvalues of the matrices A and $\underline{A}$ and their absolute values for the state space system estimated for the three- and the six-dimensional data set from 1989 to 2018

Eigenvalues of A	absolute value	Eigenvalues of $\underline{A}$	absolute value
		$s = 3$	
0.959+0.051i	0.961	-0.212	0.212
0.959-0.051i	0.961	0.116	0.116
0.868	0.868	-0.000	0.000
		$s = 6$	
0.979	0.979	-0.564	0.564
0.937+0.075i	0.940	-0.292	0.292
0.937-0.075i	0.940	0.073+0.132i	0.150
0.918+0.113i	0.925	0.073-0.132i	0.150
0.918-0.113i	0.925	0.001	0.001

cointegrating rank is thus smaller than the one found in the data set until 1988. For the state space framework the cointegrating rank coincides with the one found for the full data set. Again, we take a look at the eigenvalues of the estimates of A and $\underline{A}$ without restrictions on the cointegrating rank for the state space framework. The eigenvalues are also in Table 3.18. The eigenvalues of A rather indicate one common trend and thus a cointegrating rank of five. Since all eigenvalues are greater than 0.9, however, the number of common trends is difficult to determine by looking at the eigenvalues. As for the full data set, the strict minimum phase condition is fulfilled.

3.6.4 Estimation of the Cointegrating Space

Again, we estimate the cointegrating space with a pseudo maximum likelihood estimator for both the VECM and the state space framework. For the three-dimensional data set both rank tests found a cointegrating rank of one. Thus, we do not test the hypothesis, whether $\begin{pmatrix}1 & 0 & -1\end{pmatrix}'$ and $\begin{pmatrix}0 & 1 & -1\end{pmatrix}'$ span the cointegrating space, but the hypothesis, whether the one-dimensional cointegrating space is contained in the space spanned by $\begin{pmatrix}1 & 0 & -1\end{pmatrix}'$ and $\begin{pmatrix}0 & 1 & -1\end{pmatrix}'$. Table 3.19 contains the results. A comparison to the cointegrating spaces of the other data sets is not possible, since the cointegrating rank is different. It is notable that the standard errors of the estimates are very large. The hypothesis that the cointegrating space is a subspace of the the space spanned by $(1, 0 - 1)'$ and $(0, 1, -1)'$ is rejected by both the VECM and the state space framework.
For the six-dimensional data set we estimate a three-dimensional cointegrating space for both the VECM and the state space framework. Table 3.19 also contains the estimates for the six-dimensional data set. The estimated cointegrating spaces are similar and similar to the cointegrating space estimated for the full data set using the state space framework, the only other three-dimensional cointegrating space.

Table 3.19: Estimates of the cointegrating space with standard errors in brackets, test statistic of the test whether the cointegrating space is a subspace of the the space spanned by $(1, 0 - 1)'$ and $(0, 1, -1)'$ and p-value of the test for the three-dimensional data set from 1989 to 2018 and estimates of the cointegrating space for the six-dimensional data set from 1989 to 2018

method	VECM			state space		
	$s = 3$					
c	1			1		
i	11.489			17.134		
	(33.018)			(78.287)		
y	-31.480			-44.922		
	(88.672)			(202.529)		
test statistic	20.10			16.29		
p-value	< 0.001			< 0.001		
	$s = 6$					
c	1	0	0	1	0	0
i	0	1	0	0	1	0
$m - p$	0	0	1	0	0	1
y	-1.269	-2.312	0.845	-1.247	-2.245	1.076
	(0.683)	(1.813)	(4.997)	(0.408)	(0.732)	(2.994)
R	0.005	-0.005	0.015	0.004	-0.022	0.023
	(0.022)	(0.059)	(0.164)	(0.012)	(0.022)	(0.091)
Δp	0.010	-0.049	-0.131	-0.007	-0.048	-0.160
	(0.033)	(0.087)	(0.240)	(0.007)	(0.012)	(0.050)

3.6.5 Testing Hypotheses on the Cointegrating Space

For both the VECM and the state space framework we test the same hypotheses on the cointegrating space as for the other two data sets. Table 3.20 shows the hypotheses and their p-values. As for the full data set, all hypotheses are rejected at 1%-level.

Table 3.20: Hypotheses on the cointegrating space and their p-values for the six-dimensional data set from 1989 to2018

Hypothesis	p-value VECM	p-value state space
$(c - y)$, $(i - y)$, $(m - p - y)$	< 0.001	< 0.001
$(c - y)$, $(i - y)$, $m - p - \beta_y y + \beta_R R$	< 0.001	< 0.001
$(c - y)$, $(i - y)$,$m - p - \beta_y y + \beta_R R$, $R - \Delta p$	< 0.001	< 0.001
$(c - y) - \varphi_1(R - \Delta p)$, $(i - y) - \varphi_2(R - \Delta p)$, $m - p - \beta_y + \beta_R R$	< 0.001	< 0.001
$(c - y) - \varphi_1(R - \Delta p)$, $(i - y) - \varphi_2(R - \Delta p)$, $m - p - y$	< 0.001	< 0.001

3.7 Summary and Conclusion

The use of VARMA processes, which can be analyzed using the state space framework, may be beneficial in cointegration analysis compared to the more restrictive VAR framework. For this reason the authors have provided tools for cointegration analysis using the state space framework in a series of papers focusing on theory. To illustrate the results of these papers, this paper provides a tutorial for the application of these results. This paper describes pseudo maximum likelihood analysis for I(1) processes possessing a state space representation with an eye on application. By repeating the analysis of US economic data from 1949 to 1988 of King et al. (1991) the authors demonstrate that all economically relevant questions examined by King et al. (1991) can also be analyzed using the state space framework. Additionally, this paper points out which `matlab` functions can be used for cointegration analysis in the state space framework. The economic analysis includes tests for the presence or absence of deterministic terms, tests for the determination of the cointegrating rank and pseudo likelihood ratio tests for hypotheses on the cointegrating space.
This paper compares the results of King et al. (1991) to those obtained using the VECM of Johansen (1995) and the state space framework. While the results for the three-dimensional data set are similar, the results for the six-dimensional data set differ. For example, employing the three different methods for the determination of the cointegrating rank leads to three different results for the six-dimensional data set. This indicates that the results of empirical applications with six or more variables and sample sizes of about two hundred should be interpreted more carefully than is commonly done.
Additionally, this paper tests the robustness of the VECM and the state space framework by repeating the analysis both on a longer data set until 2018 and on the subset of the data set from 1989 to 2018. For both the VECM and the state space framework the rank tests lead to a lower cointegrating rank and the pseudo likelihood ratio tests reject hypotheses that were not rejected for the data set until 1988 for both the full sample and the data set from 1989 to 2018. Structural breaks are one possible explanation for these differences. This could be investigated with monitoring methods. This is, however, beyond the scope of this paper.

Appendix of the First Chapter

A.1 Proofs of the Results of Section 1.3

A.1.1 Proof of Lemma 1.1

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

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

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 $i = 1, \ldots, d$, the entries of C are bounded.

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

(iii) The algorithm discussed above Lemma 1.1 maps every $C \in O_{s,d}$ to $[I_d, 0'_{s-d\times d}]'$. 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 $[I_d, 0'_{s-d\times d}]'$ with the transposed Givens rotations.

(iv) As discussed, $C_O^{-1}(\cdot)$ is obtained from a repeated application of the algorithm described in Remark 1.10. In each step two entries are transformed to polar coordinates. According to Amann and Escher (2008, Chapter 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 $\boldsymbol{\theta}$.
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 $\boldsymbol{\theta}_0$ for arbitrary $C_0 \in O_{s,d}$ such that $C_O(\boldsymbol{\theta}_0) = C_0$. 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 $\boldsymbol{\theta}_j \to \boldsymbol{\theta}_0$. Then, $C_O(\boldsymbol{\theta}_j) \to C_0$ because of the continuity of C_O. Since $C_O(\boldsymbol{\theta}_j)$ 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 quadratic 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 $C_O^{-1} : O_{s,s}^+ \to \Theta_O^{\mathbb{R}}$ and $C_O^{\pm}(\cdot)\big|_{O_{s,s}^+}$ 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 $C_O^{\pm}(\cdot)\big|_{O_{s,s}^-}$ is infinitely often differentiable with infinitely often differentiable inverse on an open subset of $O_{s,s}^-$, from which the result follows.□

A.1.2 Proof of Lemma 1.2

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

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

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 $i = 1, \ldots, d$, the entries of C are bounded.

(ii) By definition $C_U(\boldsymbol{\varphi})$ is a product of matrices whose elements are either constant or infinitely often differentiable functions of the elements of $\boldsymbol{\varphi}$.

(iii) The algorithm discussed above Lemma 1.2 maps every $C \in U_{s,d}$ to $[D_d(\boldsymbol{\varphi}_D), 0'_{s-d\times d}]'$ with $D_d(\boldsymbol{\varphi}_D) = \mathrm{diag}(e^{i\varphi_{D,1}}, \ldots, e^{i\varphi_{D,d}})$. 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 $[D_d(\boldsymbol{\varphi}_D), 0'_{s-d\times d}]'$ with the transposed Givens rotations.

(iv) The algorithms in Remark 1.12 and above Lemma 1.2 describe C_U^{-1} in detail. The determination of an element of $\boldsymbol{\varphi}_L$ or $\boldsymbol{\varphi}_R$ uses the transformation of two complex numbers into polar coordinates in step 2 of Remark 1.12, which according to Amann and Escher (2008, Chapter 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 1.12 uses the formulas $\varphi_1 = \tan^{-1}\left(\frac{b}{a}\right)$, which is infinitely often differentiable for $a > 0$, and $\varphi_2 = \varphi_a - \varphi_b \bmod 2\pi$, 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 $\boldsymbol{\varphi}_D$ 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, Chapter 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 $U_{s,d}$.□

A.1.3 Proof of Theorem 1.2

(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, Chapter 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 $(\mathcal{A}(\boldsymbol{\theta}), \mathcal{B}(\boldsymbol{\theta}), \mathcal{C}(\boldsymbol{\theta}))$ for given $\boldsymbol{\theta}$ and Γ is straightforward: $\mathcal{A}_u$ is uniquely determined by Γ. Since $\boldsymbol{\theta}_{B,p}$ contains the entries of $\mathcal{B}_u$ restricted to be positive and $\boldsymbol{\theta}_{B,f}$ contains the free parameters of $\mathcal{B}_u$, the mapping $\boldsymbol{\theta}_{B,p} \times \boldsymbol{\theta}_{B,f} \to \mathcal{B}_u$ is continuous. The mapping $\boldsymbol{\theta}_\bullet \to (\mathcal{A}_\bullet, \mathcal{B}_\bullet, \mathcal{C}_\bullet)$ is continuous, cf. Hannan and Deistler (1988, Theorem 2.5.3 (ii)). The mapping $\boldsymbol{\theta}_{C,E} \times \boldsymbol{\theta}_{C,G} \to \mathcal{C}_u$ consists of iterated applications of C_O, and C_U (compare Lemmas 1.1 and 1.2), 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 1.1 and 1.2) to general unit root structures and to complex matrices. The proof that these functions are differentiable is analogous to the proofs of Lemma 1.1 and Lemma 1.2.

(iii) The definitions of $\boldsymbol{\theta}_{B,f}$ and $\boldsymbol{\theta}_{B,p}$ immediately imply that they depend continuously on $\mathcal{B}_u$. The parameter vector $\boldsymbol{\theta}_\bullet$ 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 Lemma 1.1 and Lemma 1.2.□

A.2 Proofs of the Results of Section 1.4

A.2.1 Proof of Theorem 1.3

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 1.9. 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 $\mathcal{C} = \begin{bmatrix} \tilde{\mathcal{C}} & C_1 \end{bmatrix} S'$ it holds that $(\mathcal{A}, \mathcal{B}, \mathcal{C}) \in \Delta_\Gamma$. Then, $(\mathcal{A}_j, \mathcal{B}_j, \mathcal{C}_j) = (\mathcal{A}, S \operatorname{diag}(I_{n_1}, j^{-1} I_{n_2}) S' \mathcal{B}, \mathcal{C}) \in \Delta_\Gamma$, where n_i is the number of rows of S_i for $i = 1, 2$ converges for $j \to \infty$ to $\left(\mathcal{A}, S \begin{bmatrix} \tilde{\mathcal{B}} \\ 0 \end{bmatrix}, \mathcal{C} \right) \in \overline{\Delta_\Gamma}$, which is observationally equivalent to $(\tilde{\mathcal{A}}, \tilde{\mathcal{B}}, \tilde{\mathcal{C}})$. Consequently, $\tilde{k} = \pi \left(\mathcal{A}, S \begin{bmatrix} \tilde{\mathcal{B}} \\ 0 \end{bmatrix}, \mathcal{C} \right) \in \pi(\overline{\Delta_\Gamma})$.

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

For the stable system we can separate the subsystem $(A_{j,s}, B_{j,s}, C_{j,s})$ remaining stable in the limit and the part with eigenvalues of A_j tending to the unit circle. As discussed in Section 1.4.1, $(A_{j,s}, B_{j,s}, C_{j,s})$ 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 $\alpha_\bullet$, 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 $(\bar{\Omega}_S, \bar{p})$ 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 $(\bar{\Omega}_S, \bar{p})$ that is smaller than the pair $(\tilde{\Omega}_S, \tilde{p})$ corresponding to an element of $\mathcal{K}(\Gamma)$.□

A.2.2 Proof of Theorem 1.4

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, has been discussed in Section 1.4.2. 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 $\beta_\bullet$ 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 (1.14) 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 $\pi(\mathcal{A}, \mathcal{B}, \mathcal{C}) = \lim_{j\to\infty} \pi(A_j, B_j, C_j) \in \overline{M_{\Gamma_{\alpha_{\bullet,g(n)}}}}$.□

A.2.3 Proof of Theorem 1.5

(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 $\alpha_\bullet$ 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 $k_i \to k_0$ 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 1.3. Comparing Definitions 1.10 and 1.11, we see $\bigcup_{\tilde\Gamma\in\mathcal{K}(\Gamma_g)} M_{\tilde\Gamma} \subset \bigcup_{(\tilde\Omega_S,\tilde n_\bullet)\in\mathcal{A}(\Omega_S,n_\bullet)} M(\tilde\Omega_S, \tilde n_\bullet)$. By the definition of the partial ordering (compare Definition 1.9) $\bigcup_{\tilde\Gamma\leq\Gamma_g} M_{\tilde\Gamma} \subset \bigcup_{(\tilde\Omega_S,\tilde n_\bullet)\leq(\Omega_S,n_\bullet)} M(\tilde\Omega_S, \tilde n_\bullet)$ holds. Together these two statements imply the second inclusion.

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

(a) If $M(\tilde\Omega_S, \tilde n_\bullet) \subset \overline{M(\Omega_S, n_\bullet)}$, then $\bigcup_{(\check\Omega_S,\check n_\bullet)\leq(\tilde\Omega_S,\tilde n_\bullet)} M(\check\Omega_S, \check n_\bullet) \subset \overline{M(\Omega_S, n_\bullet)}$.

(b) If $(\tilde\Omega_S, \tilde n_\bullet) \in \mathcal{A}(\Omega_S, n_\bullet)$, then $M(\tilde\Omega_S, \tilde n_\bullet) \subset \overline{M(\Omega_S, n_\bullet)}$.

For (a) note that for an arbitrary transfer function $\check k \in M(\check\Omega_S, \check n_\bullet)$ with $(\check\Omega_S, \check n_\bullet) \leq (\tilde\Omega_S, \tilde n_\bullet)$ there is a multi-index $\check\Gamma$ such that $\check k \in M_{\check\Gamma}$. By the definition of the partial ordering (compare Definition 1.9) we find a multi-index $\tilde\Gamma \geq \check\Gamma$ such that $M_{\check\Gamma} \subset M(\tilde\Omega_S, \tilde n_\bullet)$. By Theorem 1.3 and the continuity of π we have $M_{\check\Gamma} \subset \pi(\overline{\Delta_{\tilde\Gamma}}) \subset \overline{M_{\tilde\Gamma}}$. Since $\overline{M(\tilde\Omega_S, \tilde n_\bullet)} \subset \overline{M(\Omega_S, n_\bullet)}$ by assumption, $\check k \in \overline{M_{\tilde\Gamma}} \subset \overline{M(\tilde\Omega_S, \tilde n_\bullet)} \subset \overline{M(\Omega_S, n_\bullet)}$ which finishes the proof of (a).

With respect to (b) note that by Definition 1.11, $\mathcal{A}(\Omega_S, n_\bullet)$ contains transfer functions with two types of state space unit root structures. First, $\tilde{\mathcal{A}}_u$ corresponding to state space unit

root structure $\tilde{\Omega}_S$ may be of the form

$$S\,\tilde{\mathcal{A}}_u S' \;=\; \begin{bmatrix} \mathcal{A}_u & J_{12} \\ 0 & J_2 \end{bmatrix}. \tag{A.1}$$

Second, $\check{\mathcal{A}}_u$ corresponding to state space unit root structure $\check{\Omega}_S$ may be of the form (A.1), 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 $\overline{M(\tilde{\Omega}_S, \tilde{n}_\bullet)}$, where $\tilde{\Omega}_S$ is the state space unit root structure corresponding to $\tilde{\mathcal{A}}_u$ in (A.1). For this, consider the sequence

$$A_j = \left[\begin{array}{cc} 1 & j^{-1} \\ 0 & 1 \end{array}\right], \quad B_j = \left[\begin{array}{c} B_1 \\ B_2 \end{array}\right], C_j = \left[\begin{array}{cc} C_1 & C_2 \end{array}\right].$$

Clearly, every system (A_j, B_j, C_j) corresponds to an $I(2)$ process, while the limit for $j \to \infty$ 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 $\pi(\overline{\Delta_{\Gamma_g}})$, 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

$$\mathcal{A}_j = \left[\begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array}\right], \quad \mathcal{B}_j = \left[\begin{array}{c} B_1 \\ B_2/j \end{array}\right], \mathcal{C}_j = \left[\begin{array}{cc} C_1 & jC_2 \end{array}\right].$$

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 $\mathcal{A}(\tilde{\Omega}_S, \tilde{n}_\bullet)$, where $\check{\mathcal{A}}_u$ can be obtained by replacing off-diagonal entries of $\mathcal{A}_u$ with zero, can be reached from within $M(\tilde{\Omega}_S, \tilde{n}_\bullet)$.

To prove $\tilde{k} \in \overline{M(\Omega_S, n_\bullet)}$ in the first case, where the state space unit root structure is extended as visible in equation (A.1), consider the sequence:

$$\tilde{A}_j = \left[\begin{array}{cc} 1 & 1 \\ 0 & 1 - j^{-1} \end{array}\right], \quad \tilde{B}_j = \left[\begin{array}{c} B_1 \\ B_2 \end{array}\right], \quad \tilde{C}_j = \left[\begin{array}{cc} C_1 & C_2 \end{array}\right],$$

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

$$\tilde{\mathcal{A}}_j = \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 - j^{-1} \end{array}\right], \quad \tilde{\mathcal{B}}_j = \left[\begin{array}{c} B_1 + jB_2 \\ -jB_2 \end{array}\right], \quad \tilde{\mathcal{C}}_j = \left[\begin{array}{cc} C_1 & C_1 - C_2/j \end{array}\right].$$

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 $\tilde{\Omega}_S$ such that $(\tilde{\Omega}_S, \tilde{n}_\bullet) \in \mathcal{A}(\Omega_S, n_\bullet)$, satisfying

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

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

the approximating sequence of transfer functions $\tilde{k}_j \in M_{\Gamma_g(\Omega_S,n_\bullet)} \to k_0 \in M_{\Gamma_g(\tilde{\Omega}_S,\tilde{n}_\bullet)}$ 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 $M_{\alpha_\bullet}$.

Both results combined constitute the whole set of attainable state space unit root structures in Definition 1.11 and prove (b).

As follows from Corollary 1.4, $\overline{M(\Omega_S,n_\bullet)} = \overline{M_{\Gamma_g(\Omega_S,n_\bullet)}}$. Thus, (b) implies the inclusion $\bigcup_{(\tilde{\Omega}_S,\tilde{n}_\bullet)\in\mathcal{A}(\Omega_S,n_\bullet)} M(\tilde{\Omega}_S,\tilde{n}_\bullet) \subset \overline{M_{\Gamma_g}}$ 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 $k_j \in M_{\Gamma_g(\Omega_S,n_\bullet)}, k_j \to k_0$, 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

$$\operatorname{rank}(\lim_{j\to\infty} E_j) = \operatorname{rank}(E_0) \leq \lim\inf_{j\to\infty} \operatorname{rank}(E_j).$$

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

$$\operatorname{rank}((A_0 - z_k I_n)^t) \quad \leq \quad n - \sum_{r=1}^{t} d^k_{j,h_k-r+1},$$

since $k_j(z) \in M_{\Gamma_g(\Omega_s,n_\bullet)}$ implies that the number d^k_{j,h_k-r+1} 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 $\sum_{r=1}^{t} d^k_{j,h_k-r+1} \leq \sum_{r=1}^{t} d^k_{0,h_k-r+1}$ for $t = 1,2,...,n$. 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

$$S\mathcal{A}_0S' \quad = \quad \begin{bmatrix} \mathcal{A}_j & \tilde{J}_{12} \\ 0 & \tilde{J}_2 \end{bmatrix}$$

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 has been 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 $d^1_{j,1} = 1, d^1_{j,2} = 1$ for the approximating sequence and $d^1_{0,1} = 0, d^1_{0,2} = 2$ for the limit respectively. Summing these indices starting from the last one, one obtains $d^1_{j,2} = 1 \leq d^1_{0,2} = 2$ and $d^1_{j,1} + d^1_{j,2} = 2 \leq d^1_{0,1} + d^1_{0,2} = 2$.

Hence, the state space unit root structure corresponding to (A_0,B_0,C_0) must be attainable according to Definition 1.11. 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.□

Appendix of the Second Chapter

B.1 Preliminaries

Remark B.1 *Lukas Matuschek extends the results of the second chapter to the I(2) case in his dissertation. Since a proof of the results in the I(2) case focusing only on the differences to the proof in the MFI(1) case would be impossible to understand for everyone who has not read this proof first, Lukas Matuschek has copied the proof inserting the necessary adaptations. For this reason the passages in the two theses are very similar.*

In particular for the convergence of the appropriately scaled Hessian matrix in a neighborhood of the true system, we will use a concept of stochastic equicontinuity, as used also in Saikkonen (1995, Section 5). Instead of Saikkonen's Condition 3.1., compare Saikkonen (1993, p. 160), we will use the following uniform equicontinuity condition, that is later shown to hold for the required quantities:

Condition B.1 (USE - Uniform Stochastic Equicontinuity) *A sequence $X_n(\theta)$, $\theta \in \Theta$ is said to fulfill Condition USE, if for every sequence $\theta_n \to \theta$ and every $\epsilon > 0, \delta > 0$ and $\eta > 0$ there exists an integer $n(\epsilon, \eta, \delta)$ such that $\mathbb{P}\{\sup_{t \in B(\theta_n, \delta)} \|X_n(t) - X_n(\theta_n)\| > \epsilon\} \leq \eta\delta$ for $n \geq n(\epsilon, \eta, \delta)$.*

This condition ensures that the convergence is uniformly in the parameter space. In the special case of a compact parameter space we obtain the following consequence:

Lemma B.1 *Assume that $X_j(\theta)$, $\theta \in \Theta$ fulfills Condition USE, where Θ is compact. Further, assume that for each fixed $\theta \in \Theta$ the sequence $X_j(\theta) \to 0$ in probability for $j \to \infty$. Then, $\sup_{\theta \in \Theta} X_j(\theta) \to 0$ in probability for $j \to \infty$.*

PROOF: Fix $\epsilon > 0, \delta > 0$. Let $\theta_1, \ldots, \theta_k$ denote k points, such that $\cup_{i=1}^{k} B(\theta_i, \delta)$ covers Θ. Due to the assumed compactness of Θ a finite cover exists. Since $X_j(\theta)$ fulfills Condition USE, there exists for each i an integer $j_i(\epsilon, \eta, \delta)$ such that the probability that the error $\sup_{\theta' \in B(\theta_i, \delta)} \|X_j(\theta') - X_j(\theta_i)\| > \epsilon/2$ is smaller than $\eta\delta$ for $j \geq j_i(\epsilon, \eta, \delta)$. If $\eta = \epsilon/(2k\delta)$ is chosen,

$$\begin{aligned}
\mathbb{P}\{\sup_{\theta \in \Theta} \|X_j(\theta)\| > \epsilon\} &\leq \mathbb{P}\{\sup_{\theta \in \cup B(\theta_i, \delta)} \|X_j(\theta)\| > \epsilon\} \\
&\leq \sum_{i=1}^{k} \mathbb{P}\{\sup_{\theta \in B(\theta_i, \delta)} \|X_j(\theta) - X_j(\theta_i) + X_j(\theta_i)\| > \epsilon\} \\
&\leq \sum_{i=1}^{k} \mathbb{P}\{\sup_{\theta \in B(\theta_i, \delta)} \|X_j(\theta) - X_j(\theta_i)\| > \epsilon/2\} + \mathbb{P}\{\|X_j(\theta_i)\| > \epsilon/2\} \\
&\leq k\eta\delta + \sum_{i=1}^{k} \mathbb{P}\{\|X_j(\theta_i)\| > \epsilon/2\}
\end{aligned}$$

and this can be made arbitrarily small by choosing j large, since $k\eta\delta = \epsilon/2$ and $X_n(\theta_i) \to 0$ in probability. □

For the convergence results it is convenient to use complex processes. For the deterministic terms, for example, note that $\mathbf{d}_t(\theta_D) = Ds_t$ where

$$s_t = [\ 1 \quad \cos(\omega_2 t) \quad \sin(\omega_2 t) \quad \dots \quad \sin(\omega_{l-1} t) \quad (-1)^t \quad \cos(\omega_{l+1} t) \quad \dots \quad \sin(\omega_m t) \quad t\ ]'.$$

For the following analysis we use the transformation $s_{t,\mathbb{C}} = U^{-1} s_t$ where $U = \mathrm{diag}(1, I_{l-2} \otimes u, 1, I_{m-l} \otimes u, 1)$ with $u = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}$ to obtain

$$s_{t,\mathbb{C}} = [\ 1 \quad (z_2)^t \quad (\bar{z}_2)^t \quad \dots \quad (\bar{z}_{l-1})^t \quad (-1)^t \quad (z_{l+1})^t \quad \dots \quad (\bar{z}_m)^t \quad t\ ]'.$$

with components $s_t(z_k) = \overline{z_k} s_{t-1}(z_k)$. Note that in the following lemmas we refer to the complex quantities and accordingly the roots are allowed to vary in the interval $[0, 2\pi)$. By $x_t(z_k)$ we denote the state components corresponding to one potentially complex unit root z_k. This is not to be confused with $x_{t,k}$, which collects all components corresponding to pairs of complex conjugate unit roots.

Lemma B.2 *Let $\omega_1, \dots, \omega_{l_{2\pi}}$ denote $l_{2\pi}$ distinct frequencies in $[0, 2\pi)$ and let $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be a martingale difference sequence fulfilling the assumptions stated in Assumption 2.1 with non-singular innovation variance $\Sigma_\circ$. Further, let $x_{t+1}(z_k) = \overline{z}_k x_t(z_k) + B_{k,\circ}\varepsilon_t, x_1(z_k) = x_1, z_k = e^{i\omega_k}$ and $s_{t+1}(z_k) = \overline{z}_k s_t(z_k), s_1(z_k) = 1$ for $k = 1, \dots, l_{2\pi}$. Here, x_1 is a complex valued constant vector. Let $\delta_k = 1$ for $z_k = \pm 1$ and $\delta_k = \frac{1}{\sqrt{2}}$ else. Further, let $\delta_k W_k(u), k = 1, \cdots, l_{2\pi}$ denote the weak limit of $T^{-1/2} \sum_{t=1}^{[Tu]} \overline{z}_k^t \varepsilon_t$, where $[Tu]$ denotes the integer part of Tu. $W_k(u)$ and $W_j(u)$ are independent for $j \neq k$. $W_k(u) = W_k^r(u) + iW_k^i(u)$, where $W_k^r(u)$ denotes the real part and $W_k^i(u)$ the imaginary part of the random variable, which are independent real valued random walks with variance $\Sigma_\circ$.*
Then, the following statements hold:

(i) $\langle \varepsilon_t, \varepsilon_{t+j} \rangle \to \delta_{0,j}\Sigma_\circ$ *in probability.*

(ii) $\langle x_t(z_k), \varepsilon_t \rangle \xrightarrow{d} \delta_k^2 B_{k,\circ} \int_0^1 W_k(u) dW_k(u)' =: X(z_k)$,
$\langle s_t(z_k), \varepsilon_t \rangle \to 0$.

(iii) $T^{-1}\langle x_t(z_k), x_t(z_k) \rangle \xrightarrow{d} \delta_k^2 B_{k,\circ} \int_0^1 W_k(u) W_k(u)' du B_{k,\circ}' =: Z(z_k)$,
$\langle s_t(z_k), s_t(z_k) \rangle = 1$.

(iv) For $z_k \neq z_j$ it holds that $T^{-1}\langle x_t(z_k), x_t(z_j) \rangle \to 0$ in probability and
$\langle s_t(z_k), s_t(z_j) \rangle \to 0$.

(v) $T^{-1/2}\langle x_t(z_k), s_t(z_k) \rangle \xrightarrow{d} \delta_k \int_0^1 B_{k,\circ} W_k(u) du =: Y(z_k)$
and $T^{-1/2}\langle x_t(z_k), s_t(z_j) \rangle \to 0$ *in probability for $z_j \neq z_k$.*

(vi) $T^{-3/2}\langle t, x_t(1) \rangle \xrightarrow{d} \int_0^1 uW_1(u)' du B_{1,\circ}' =: V(1)$
and $T^{-3/2}\langle t, x_t(z_k) \rangle \to 0$ *in probability for $z_k \neq 1$.*

(vii) $T^{-1/2}\langle t, \varepsilon_t \rangle \xrightarrow{d} \int_0^1 u dW_1(u)' =: U(1)$.

PROOF: The proof of the items (i) to (v) is in many parts a direct consequence of results obtained in Johansen and Schaumburg (1999), Lemma 5, Theorem 6 and Corollary 7. One difference is the inclusion of starting values, which however does not influence the asymptotic behavior. This follows, since the initial effects – being solutions to the homogenous equation – are of the form $Ds_t(z_k)$.
Also the results concerning $s_t(z_j)$ are standard, except for the cross terms with $x_t(z_k)$. For $z_j \neq z_k$ the proof is analogous to the proof of equation (22) in Johansen and Schaumburg (1999) and for the case $z_j = z_k$ the continuous mapping theorem leads to the stated results. Items (vi) and (vii) are direct consequences of Lemma 1 (d) and (a) of Sims et al. (1990) respectively. □

In the expressions for the pseudo likelihood function, terms that can be represented as filtered versions of the observations y_t show up, where the filters depend upon the parameter values. Thus, it is necessary to understand the convergence properties of estimated sample covariances of expressions of the form $g(L,\theta)x_t(z_k) = \sum_{j=0}^{t-1} G_j(\theta)x_{t-j}(z_k)$, where $g(z,\theta) = \sum_{j=0}^{\infty} G_j(\theta)z^j$ denotes a family of stable transfer functions parametrized by the parameter vector θ. The notation here somewhat hides the fact that the summation is only performed for $t > 0$ or equivalently $x_t(z_k) = 0$, $t < 0$ is assumed. We will use this notation throughout the appendix. A family of transfer functions $g(z,\theta), \theta \in \Theta$ is called *uniformly stable*, if there exist constants $C < \infty, 0 < \rho < 1$, such that $\sup_{\theta\in\Theta} \|G_j(\theta)\| \leq C\rho^j$, i. e., the decay in the transfer function coefficients is exponential and uniform in the parameter set. For quantities of this form in the following lemma the asymptotic behavior is clarified and for each of the considered expressions *Condition USE* is established. The lemma parallels Theorem 4.2 in Saikkonen (1993, page 167), in which he establishes his Condition 3.1.

Lemma B.3 *Let $g(z;\theta) = \sum_{i=0}^{\infty} G_i(\theta)z^i = \pi(A_g, B_g, C_g)$, $k(z;\theta) = \sum_{i=0}^{\infty} K_i(\theta)z^i = \pi(A_k, B_k, C_k)$, $\theta \in \Theta$ be two uniformly stable families of rational transfer functions, of finite orders less or equal to n, where it is always assumed that the transfer functions are of the correct dimensions needed in the expressions below. Furthermore $(A_g(\theta), B_g(\theta), C_g(\theta))$ and $(A_k(\theta), B_k(\theta), C_k(\theta))$ are continuously differentiable functions of θ. Let $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be a martingale difference sequence fulfilling the assumptions stated in Assumption 2.1 with non-singular innovation variance $\Sigma_\circ$. Furthermore, let $\omega_1, \ldots, \omega_{l_{2\pi}}$ denote $l_{2\pi}$ distinct frequencies in $[0, 2\pi)$ (where $\omega_1 = 0$) and let $z_k = e^{i\omega_k}$. Furthermore, $x_t(z_k)$, $s_t(z_k)$, $U(1)$, $V(1)$, $X(z_k)$, $Y(z_k)$ and $Z(z_k)$ are as defined in Lemma B.2.*
The following asymptotic results hold for each fixed $\theta \in \Theta$.

(i) $\langle g(L;\theta)\varepsilon_t, k(L;\theta)\varepsilon_t\rangle \to \sum_{r=0}^{\infty} G_r(\theta)\Sigma_\circ K_r(\theta)'$ *in probability.*
$\langle g(L;\theta)s_t(z_k), k(L;\theta)\varepsilon_t\rangle \to 0.$

(ii) $\langle g(L;\theta)x_t(z_k), k(L;\theta)\varepsilon_t\rangle \xrightarrow{d}$

$$g(z_k;\theta)X(z_k)k(z_k;\theta)' - g(z_k;\theta)z_k B_{k,\circ}\Sigma_\circ \tilde{k}(0;\theta)' + \lim_{t\to\infty} \mathbb{E}\tilde{g}(L;\theta)B_{k,\circ}\varepsilon_{t-1}(k(L;\theta)\varepsilon_t)',$$

where $g(z;\theta) = g(z_k;\theta) + (1-\overline{z}_k z)\tilde{g}(z;\theta)$, $k(z;\theta) = k(z_k;\theta) + (1-\overline{z}_k z)\tilde{k}(z;\theta)$.

(iii) $T^{-1}\langle g(L;\theta)x_t(z_k), k(L;\theta)x_t(z_j)\rangle \xrightarrow{d} \delta_{k,j} g(z_k;\theta)Z(z_k)k(z_k;\theta)'$.

(iv) $\langle g(L;\theta)s_t(z_k), k(L;\theta)s_t(z_j)\rangle \to \delta_{k,j} g(z_k;\theta)k(z_k;\theta)'$.

(v) $T^{-1/2}\langle g(L;\theta)x_t(z_k), k(L;\theta)s_t(z_j)\rangle \to \delta_{k,j} g(z_k;\theta)Y(z_k)k(z_k;\theta)'$.

(vi) $T^{-1}\langle g(L;\theta)t, k(L;\theta)s_t(z_k)\rangle \to \delta_{k,1}\frac{1}{2}g(1;\theta)k(1;\theta)'$.
$T^{-2}\langle g(L;\theta)t, k(z;\theta)t\rangle \to \frac{1}{3}g(1;\theta)k(1;\theta)'$.

(vii) $T^{-3/2}\langle g(L;\theta)t, k(L;\theta)x_t(z_k)\rangle \xrightarrow{d} \delta_{k,1} g(1;\theta)V(1)k(1;\theta)'$.

(viii) $T^{-1/2}\langle g(L;\theta)t, k(L;\theta)\varepsilon_t\rangle \xrightarrow{d} g(1;\theta)U(1)k(1;\theta)$.

All sequences in items (i) to (viii) fulfill condition USE.

PROOF: The proof of the lemma rests upon the results established in Lemma B.2. Item (i) is standard and its proof is therefore omitted. Analogously to the well known decomposition for the case $z_k = 1$, decompose $g(z;\theta) = g(z_k;\theta) + (1-\overline{z}_k z)\tilde{g}(z;\theta)$ for each θ and $|z_k| = 1$. The assumed uniform stability of $g(z,\theta)$ implies that also $\tilde{g}(z;\theta) = \sum_{j=0}^{\infty} \tilde{G}_j(\theta)z^j$ is a uniformly stable family of transfer functions. Using the decomposition we obtain:

$$\begin{aligned} g(L;\theta)x_t(z_k) &= \sum_{i=0}^{t-1} G_i(\theta)x_{t-i}(z_k) \\ &= g(z_k;\theta)x_t(z_k) + \tilde{g}(L;\theta)(x_t(z_k) - \overline{z}_k x_{t-1}(z_k)) \\ &= g(z_k,\theta)x_t(z_k) + \tilde{g}(L;\theta)B_{k,\circ}\varepsilon_{t-1} + \tilde{G}_{t-1}(\theta)x_1(z_k) \end{aligned}$$

for $t \in \mathbb{N}$. Now, due to the fact that also $\tilde{G}_{t-1}(\theta)$ converges uniformly in $\theta \in \Theta$ at an exponential rate to zero, the last term in the above expression can be neglected. Then, item (ii) follows from

$$\langle x_t(z_k), k(L;\theta)\varepsilon_t\rangle = \langle x_t(z_k), k(z_k;\theta)\varepsilon_t\rangle + \langle x_t(z_k), \tilde{k}(L;\theta)(1-\overline{z}_k L)\varepsilon_t\rangle.$$

The first term converges to $X(z_k)k(z_k;\theta)'$ according to Lemma B.2. The second term is equal to

$$T^{-1}\sum_{t=1}^{T}\left[x_t(z_k)(\tilde{k}(L;\theta)\varepsilon_t)' - z_k x_t(z_k)(\tilde{k}(L;\theta)\varepsilon_{t-1})'\right] =$$
$$z_k\left(T^{-1}\sum_{t=1}^{T-1}\left[\overline{z}_k x_t(z_k) - x_{t+1}(z_k)\right](\tilde{k}(L;\theta)\varepsilon_t)'\right) + o_P(1),$$

where the $o_P(1)$ term is due to $T^{-1}x_T(z_k)(\tilde{k}(L;\theta)\varepsilon_T)'$. This term converges for $T\to\infty$ to $-z_k B_{k,\circ}\mathbb{E}\varepsilon_t(\tilde{k}(L;\theta)\varepsilon_t)' \to -z_k B_{k,\circ}\Sigma_\circ\tilde{k}(0;\theta)'$. Combining this with pre-multiplication of $x_t(z_k)$ with $g(z_k;\theta)$ then delivers the result. Item (iii) can be shown using similar arguments. The proof of (iv) and (v) follows from

$$g(L;\theta)s_t(z_k) = g(L;\theta)\overline{z}_k^t = \sum_{j=0}^{t-1}G_j(\theta)\overline{z}_k^{t-j} = \overline{z}_k^t\sum_{j=0}^{t-1}G_j(\theta)\overline{z}_k^{-j} = g(z_k;\theta)s_t(z_k) + o(1),$$

where the $o(1)$ term is of order $O(\rho^t)$.
The convergence results in (vi) are standard results for deterministic limits, compare Sims et al. (1990, Lemma 1(c)).
For the proof of (vii), note that

$$g(L;\theta)t = \sum_{j=0}^{t-1}G_j(\theta)(t-j) = \left(\sum_{j=0}^{t-1}G_j(\theta)\right)t - \sum_{j=0}^{t-1}jG_j(\theta)$$
$$= g(1;\theta)t - g^*(1;\theta) - \left(\sum_{j=t}^{\infty}G_j(\theta)\right)t + \sum_{j=t}^{\infty}jG_j(\theta),$$

where $g^*(z) = \sum_{j=1}^{\infty}jG_j(\theta)z^{j-1}$. Thus, $g(L;\theta)t$ essentially equals $g(1;\theta)t - g^*(1,\theta)$, where the difference is of order $O(t\rho^t)$. The result then follows from Lemma B.2 (vii). We compute

$$\lim_{T\to\infty}T^{-3/2}\langle g(L;\theta)t, k(L;\theta)x_t(z_k)\rangle = \lim_{T\to\infty}T^{-3/2}g(1;\theta)\langle t, x_t(z_k)\rangle(k(z_k;\theta))' + o_p(T^{-1})$$
$$\xrightarrow{d}\delta_{k,1}g(1;\theta)V(1)k(1;\theta)',$$

where we have used Lemma B.2 (vi) for the limit.
The proof of (viii) is similar. We compute

$$\lim_{T\to\infty}T^{-1/2}\langle g(L;\theta)t, k(L;\theta)\varepsilon_t\rangle = \lim_{T\to\infty}T^{-1/2}g(1;\theta)\langle t, \varepsilon_t\rangle k(1;\theta)' + o_p(T^{-1})$$
$$\xrightarrow{d}g(1;\theta)U(1)k(1;\theta)',$$

where we have used Lemma B.2 (vii) for the limit.
The fulfillment of Condition USE for the sequences considered in (i) to (viii) is left to be shown. For (i) the claim follows from standard arguments for stationary processes. The difference for two parameter vectors (remembering that Condition USE is concerned with the behavior for $\theta_n \to \theta$) can be decomposed in two parts: One part depends only upon the parameter vectors but not on ε_t, for which convergence to zero follows immediately due to continuity of the parametrization. The other part can be bounded by the estimation error from estimating sample covariances of stationary processes. This expression can be bounded uniformly in the lag, see Hannan and

Deistler (1988, Theorem 5.3.2). The same decomposition as just mentioned can also be applied to the terms appearing in the other items. Consider, e. g., $g(z_k;\theta)\langle x_t(z_k),\varepsilon_t\rangle k(z_k;\theta)'$, which is the product of three terms. Of these three terms, two are deterministic and depend continuously on the parameter vector, the third term is stochastic and independent of the parameter vector. This finishes the proof of the Lemma. □

B.2 Proofs of the Theorems

We start with the proof of Lemma 2.1.
PROOF: For $z_k \notin \{1,-1\}$ and for a vector $\gamma \in \mathbb{C}^s$, $\gamma \neq 0$ satisfying $\gamma' C_k = 0$ we have for the corresponding polynomial $\beta(z) = \beta_0 + \beta_1 z = \gamma(1 - z_k z) + \overline{\gamma}(1 - \overline{z}_k z)$

$$\begin{pmatrix}\beta_0' & \beta_1'\end{pmatrix}\begin{pmatrix}\mathcal{C}_k\mathcal{A}_k \\ \mathcal{C}_k\end{pmatrix} = \begin{pmatrix}(\gamma+\bar\gamma)' & (-\gamma z_k - \bar\gamma\bar z_k)'\end{pmatrix}\begin{pmatrix}C_k^r\cos\omega_k + C_k^i\sin\omega_k & -C_k^r\sin\omega_k + C_k^i\cos\omega_k \\ C_k^r & C_k^i\end{pmatrix}.$$

The first entry of this matrix is

$$\begin{aligned}&(\gamma+\bar\gamma)'(C_k^r\cos(\omega_k) + C_k^i\sin(\omega_k)) - (\gamma z_k + \bar\gamma\bar z_k)'C_k^r \\ &= (\gamma+\bar\gamma)'(C_k^r\cos(\omega_k) + C_k^i\sin(\omega_k)) - (\gamma(\cos(\omega_k) + i\sin(\omega_k)) + \bar\gamma(\cos(\omega_k) - i\sin(\omega_k)))'C_k^r \\ &= \left((\gamma+\bar\gamma)'C_k^i + i(\bar\gamma-\gamma)'C_k^r\right)\sin(\omega_k) = 2\mathcal{I}(\gamma' C_k)\sin(\omega_k) = 0\end{aligned}$$

and the second entry is

$$\begin{aligned}&(\gamma+\bar\gamma)'(-C_k^r\sin(\omega_k) + C_k^i\cos(\omega_k)) - (\gamma z_k + \bar\gamma\bar z_k)'C_k^i \\ &= (\gamma+\bar\gamma)'(-C_k^r\sin(\omega_k) + C_k^i\cos(\omega_k)) - (\gamma(\cos(\omega_k) + i\sin(\omega_k)) + \bar\gamma(\cos(\omega_k) - i\sin(\omega_k)))'C_k^i \\ &= -\left((\gamma+\bar\gamma)'C_k^r + i(\gamma-\bar\gamma)'C_k^i\right)\sin(\omega_k) = -2\mathcal{R}(\gamma' C_k)\sin(\omega_k) = 0.\end{aligned}$$

Since the polynomial evaluated at z_k is $\gamma(1-z_k^2) \neq 0$, the polynomial is a PCIV of order $(\Omega, \tilde\Omega)$. For a polynomial vector of degree one $\beta(L) = \beta_0 + \beta_1 L$ of order $(\Omega, \tilde\Omega)$ with $\tilde\Omega \subseteq \Omega \setminus \{\omega_k\}$ we find a vector $\gamma \in \mathbb{C}^s$ such that $\beta(L) = \gamma'(1 - z_k L) + \overline{\gamma}'(1 - \overline{z}_k L)$ by choosing $\mathcal{R}(\gamma) = \beta_0/2$ and $\mathcal{I}(\gamma) = (\beta_0\mathcal{R}(z_k) - \beta_1)/(2\mathcal{I}(z_k))$.
Since $\beta(L)$ is a PCIV of order $(\Omega, \tilde\Omega)$ with $\tilde\Omega \subseteq \Omega \setminus \{z_k\}$

$$0 = \begin{pmatrix}\beta_0' & \beta_1'\end{pmatrix}\begin{pmatrix}\mathcal{C}_k\mathcal{A}_k \\ \mathcal{C}_k\end{pmatrix} = \begin{pmatrix}2\mathcal{I}(\gamma' C_k)\sin(\omega_k) & -2\mathcal{R}(\gamma' C_k)\sin(\omega_k)\end{pmatrix}.$$

Since $\omega_k \in (0,\pi)$ by assumption we have $\sin(\omega_k) \neq 0$ and thus $\mathcal{R}(\beta' C_k) = \mathcal{I}(\beta' C_k) = 0$. □

Lemma B.4 *Let the MFI(1) process $\{y_t\}_{t\in\mathbb{Z}}$ be generated as in Theorem 2.1. Define the pseudo likelihood function and the prediction error criterion function*

$$\mathcal{L}_T(\theta,\theta_D,\Sigma;Y_T) = L_T(k(z;\theta),\Sigma,\theta_D;Y_T), \quad \mathcal{L}_{PE,T}(\theta,\theta_D,\Sigma,0;Y_T) = L_{PE,T}(k(z;\theta),\Sigma,\theta_D;Y_T),$$

where $k(z;\theta) = \pi(\mathcal{A}(\theta),\mathcal{B}(\theta),\mathcal{C}(\theta))$. Assume that the pseudo likelihood function $\mathcal{L}_T$ is maximized for given multi-index Γ over the parameters $\theta \in \Theta_\Gamma, \theta_D \in \Theta_D, \Sigma \in \Theta_\Sigma$, where all sets are compact such that $\inf_{\Sigma\in\Theta_\Sigma}\lambda_{min}(\Sigma) > 0, \sup_{\theta\in\Theta}\lambda_{|max|}(\underline{\mathcal{A}}(\theta)) < 1$.
Then,

$$\sup_{\theta\in\Theta_\Gamma,\theta_D\in\Theta_D,\Sigma\in\Theta_\Sigma}|\mathcal{L}_T(\theta,\theta_D,\Sigma;Y_T) - \mathcal{L}_{PE,T}(\theta,\theta_D,\Sigma;Y_T)| = o(T^{\epsilon-1})$$

for every $\epsilon > 0$. The same holds for the difference in the first and second derivatives.

PROOF: Let $P_1(\theta,\sigma)$ denote the variance-covariance matrix of x_1 according to the model given as $P_1(\theta,\sigma) = \mathrm{diag}(0, P_\bullet(\theta,\sigma))$. Moreover, define $\mathcal{O}_T' := [\ \mathcal{C}' \quad \mathcal{A}'\mathcal{C}' \quad \dots \quad (\mathcal{A}^{T-1})'\mathcal{C}'\ ]$. Then, the

covariance matrix of $Y_T - \mathbf{D}_T(\theta_D)$ corresponding to the parameter vector θ equals (omitting the arguments for notational simplicity)

$$\Gamma_T = \mathcal{T}_T(I_T \otimes \Sigma)\mathcal{T}_T' + \mathcal{O}_T P_1(\theta,\sigma)\mathcal{O}_T' \Rightarrow \Gamma_T^{-1} = (\mathcal{T}_T(I_T \otimes \Sigma)\mathcal{T}_T')^{-1}, -\underline{\mathcal{O}}_T X_T \underline{\mathcal{O}}_T'$$

where $\underline{\mathcal{O}}_T = (\mathcal{T}_T(I_T \otimes \Sigma)\mathcal{T}_T')^{-1}\mathcal{O}_T$ and $X_T = (P_1(\theta,\sigma)\mathcal{O}_T'\underline{\mathcal{O}}_T + I_n)^{-1}P_1(\theta,\sigma)$. It thus follows that $\|X_T\| \leq \lambda_{|\max|}(P_\bullet(\theta,\sigma))\mu$ for appropriate constant μ. Furthermore, the compactness assumption implies that

$$\mathcal{O}_T'\underline{\mathcal{O}}_T = \sum_{j=0}^{T-1}(\underline{\mathcal{A}}^j)'\mathcal{C}'\Sigma^{-1}\mathcal{C}\underline{\mathcal{A}}^j \to \sum_{j=0}^{\infty}(\underline{\mathcal{A}}^j)'\mathcal{C}'\Sigma^{-1}\mathcal{C}\underline{\mathcal{A}}^j < \infty$$

uniformly in the compact parameter space due to the uniform strict minimum-phase assumption. Therefore,

$$\Gamma_T^{-1} = (\tilde{\mathcal{T}}_T')^{-1}(I_{Ts} - \underline{\tilde{\mathcal{O}}}_T X_T \underline{\tilde{\mathcal{O}}}_T')(\tilde{\mathcal{T}}_T)^{-1},$$

where $\tilde{\mathcal{T}}_T = \mathcal{T}_T(I_T \otimes \Sigma^{1/2})$ and $\underline{\tilde{\mathcal{O}}}_T = (I_T \otimes \Sigma^{-1/2})\mathcal{T}_T^{-1}\mathcal{O}_T$. Note that the $Ts \times Ts$ matrix $I_{Ts} - \underline{\tilde{\mathcal{O}}}_T X_T \underline{\tilde{\mathcal{O}}}_T'$ has $Ts - n$ eigenvalues equal to 1 and a maximum of n eigenvalues equal to the inverses of the eigenvalues of $P_1(\theta,\sigma)\mathcal{O}_T'\underline{\mathcal{O}}_T + I_n$. These are bounded from below by 1 and have a uniform upper bound due to the assumptions on $P_1(\theta,\sigma)$ implied by the compactness of the parameter space.
Thus,

$$\log\det\Gamma_T(\theta, P_1(\theta,\sigma))^{-1} = -T\log\det\Sigma + O(1)$$

uniformly in the parameter space. It is easy to see that the same holds for the derivatives with respect to the parameter coordinates. Furthermore,

$$\underline{\mathcal{O}}_T'(Y_T - \mathbf{D}_T(\theta_D)) = \sum_{t=1}^{T}(\underline{\mathcal{A}}^{t-1})'\mathcal{C}'\Sigma^{-1}\varepsilon_t(\theta,\theta_D).$$

Here,

$$\begin{aligned}
&\sum_{t=1}^{T}(\underline{\mathcal{A}}^{t-1})'\mathcal{C}'\Sigma^{-1}\varepsilon_t(\theta,\theta_D) = \\
&\sum_{t=1}^{T}(\underline{\mathcal{A}}^{t-1})'\mathcal{C}'\Sigma^{-1}\left(\sum_{i=0}^{t-1}\underline{K}_i(y_{t-i} - \mathbf{d}_{t-i}(\theta_D))\right) = \\
&\sum_{t=1}^{T}(\underline{\mathcal{A}}^{t-1})'\mathcal{C}'\Sigma^{-1}\left(\sum_{i=0}^{t-1}\underline{K}_i(\mathcal{C}x_{t-i} + \varepsilon_{t-i} + \mathbf{d}_{t-i,\circ} - \mathbf{d}_{t-i}(\theta_D))\right) = \\
&\sum_{t=1}^{T}\left(\sum_{i=0}^{T-t}(\underline{\mathcal{A}}^{t+i-1})'\mathcal{C}'\Sigma^{-1}\underline{K}_i\right)(\mathcal{C}x_t + \varepsilon_t + \mathbf{d}_{t,\circ} - \mathbf{d}_t(\theta_D)) = \\
&\sum_{t=1}^{T}\left(\sum_{i=0}^{T-t}(\underline{\mathcal{A}}^{t+i-1})'\mathcal{C}'\Sigma^{-1}\underline{K}_i\right)\left(\sum_{j=0}^{t-1}K_{j,\circ}\varepsilon_{t-j} + \mathcal{C}_\circ\mathcal{A}_\circ^{t-1}x_1 + \mathbf{d}_{t,\circ} - \mathbf{d}_t(\theta_D)\right) = \\
&\sum_{t=1}^{T}\Big(\left(\sum_{i=0}^{T-t}(\underline{\mathcal{A}}^{t+i-1})'\mathcal{C}'\Sigma^{-1}\underline{K}_i\right)\sum_{j=0}^{t-1}K_{j,\circ}\varepsilon_{t-j} + \\
&\qquad\left(\sum_{i=0}^{T-t}(\underline{\mathcal{A}}^{t+i-1})'\mathcal{C}'\Sigma^{-1}\underline{K}_i\right)\left(\mathcal{C}_\circ\mathcal{A}_\circ^{t-1}x_1 + \mathbf{d}_{t,\circ} - \mathbf{d}_t(\theta_D)\right)\Big).
\end{aligned} \tag{B.1}$$

Now, the compactness assumptions imply

$$\|(\underline{\mathcal{A}}^{j-1})'\mathcal{C}'\Sigma^{-1}\| \leq \mu_A\rho^{j-1}, \quad \|\underline{K}_i\| \leq \mu_K\rho^i \Rightarrow \left\|\left(\sum_{i=0}^{T-t}(\underline{\mathcal{A}}^{t+i-1})'\mathcal{C}'\Sigma^{-1}\underline{K}_i\right)\right\| \leq \mu_A\mu_K\mu_P\rho^{t-1}.$$

Therefore, these coefficients show an exponential decrease uniformly in the parameter space. This implies that the terms due to x_1 and $\mathbf{d}_{t,\circ} - \mathbf{d}_t(\theta_D)$ both have bounded norm uniformly in T. For the first term in (B.1) note that this is of the form

$$\alpha(T) = \sum_{t=1}^{T} Q_t(\theta)\varepsilon_t,$$

where $\sup_{\theta,\Sigma} \|Q_t\| \leq \mu_Q \rho^t$. Considering $a(T) = (\sup_{\theta,\Sigma} |\alpha(T)| - \mathbb{E}|\sup_{\theta,\Sigma} \alpha(T)|)$ the arguments in Hannan and Deistler (1988, p. 106+107) show that this term is of order $o(T^\epsilon)$ for arbitrary small $\epsilon > 0$.
Again taking derivatives with respect to the parameter vector does not change the exponential decrease uniformly in the parameter set considered. Thus, the difference can be bounded also for the first and second derivatives. □

B.2.1 Proof of Theorem 2.1

We start by showing that consistency results for the I(1) case carry over to the MFI(1) case if all unit roots are rational multiples of 2π. This is shown in the following lemma.

Lemma B.5 *The results of Theorem 2.1 for the special case of I(1) processes imply the analogous results for MFI(1) processes where all unit root frequencies are rational multiples of 2π. The convergence rates of the various terms are the same in both cases.*

PROOF: We have to differentiate PMLE and prediction error estimators. We will deal with the PML case first.
$-2/T$ times the logarithm of the Gaussian likelihood function can be written as

$$\mathcal{L}_T(\theta, \theta_D, \Sigma; Y_T) = (\log\det\Gamma_T(\theta,\Sigma) + (Y_T - \mathbf{D}_T(\theta_D))'\Gamma_T(\theta,\Sigma)^{-1}(Y_T - \mathbf{D}_T(\theta_D)))/T,$$

where again $Y_T = [\ y_1' \quad \dots \quad y_T'\ ]' \in \mathbb{R}^{Ts}$ contains all observations,

$$\mathbf{D}_T(\theta_D) = [\ \mathbf{d}_1(\theta_D)' \quad \dots \quad \mathbf{d}_T(\theta_D)'\ ]'$$

and $\Gamma_T(\theta,\Sigma)$ denotes the corresponding variance matrix.
In the MFI(1) case with all roots z_j being rational multiples of 2π, there exists an integer S such that $z_j^S = 1$. Consequently, $\mathcal{A}^S = \mathrm{diag}(I_{n_c}, \mathcal{A}_\bullet^S)$. Stacking S consecutive observations of y_t into a vector $Y_{t,S} = [\ y_t' \quad y_{t+1}' \quad \dots \quad y_{t+S-1}'\ ]' \in \mathbb{R}^{Ss}$ and analogously defining $E_{t,S}$ using the residuals ε_t and $\mathbf{D}_{t,S}$ using the deterministic variables $\mathbf{d}_t(\theta_D)$ we obtain

$$\begin{aligned} Y_{t,S} - \mathbf{D}_{t,S} &= \begin{bmatrix} \mathcal{C} \\ \mathcal{C}\mathcal{A} \\ \vdots \\ \mathcal{C}\mathcal{A}^{S-1} \end{bmatrix} x_t + \begin{bmatrix} I_s & 0 & \dots & 0 \\ \mathcal{C}\mathcal{B} & I_s & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \mathcal{C}\mathcal{A}^{S-2}\mathcal{B} & \dots & \mathcal{C}\mathcal{B} & I_s \end{bmatrix} E_{t,S}, \\ x_{t+S} &= \mathcal{A}^S x_t + [\ \mathcal{A}^{S-1}\mathcal{B} \quad \mathcal{A}^{S-2}\mathcal{B} \quad \dots \quad \mathcal{B}\ ] E_{t,S}. \end{aligned} \tag{B.2}$$

This corresponds to an $I(1)$ process $\{Y_{j+mS,S}\}_{m\in\mathbb{Z}}$ with highly structured system matrices driven by the noise $\{E_{j+mS,S}\}_{m\in\mathbb{Z}}$ for each $j = 0, ..., S-1$. Therefore, the likelihood function for $T = SM$ can be seen as the likelihood function either for $\{y_t\}_{t\in\mathbb{Z}}$ or for $\{Y_{mS,S}\}_{m\in\mathbb{Z}}$.
Now, assume that consistency for the $I(1)$ case has been shown such that

$$\begin{aligned} \hat\theta_{SM} &\to \theta_\circ, \\ \mathcal{L}_{SM}(\hat\theta_{SM}, \hat\theta_{D,SM}, \hat\Sigma_{SM}; Y_T) &\to \mathcal{L}_\infty(\theta_\circ, \theta_{D,\circ}, \Sigma_\circ) \\ \left\langle \hat E_{Sm}(\hat\theta_{SM}, \hat\theta_{D,SM}), \hat E_{Sm}(\hat\theta_{SM}, \hat\theta_{D,SM}) \right\rangle &\to \mathbb{E}(E_{t,S}E_{t,S}') \end{aligned}$$

for $M \to \infty$. Further, let $\hat{\theta}_T, \hat{\theta}_{D,T}, \hat{\Sigma}_T$ denote the minimizing argument of $\mathcal{L}_T(\theta, \theta_D, \Sigma; Y_T)$ for $T \in \mathbb{N}$. Then, for $T = SM + j, 1 \leq j < S$, orthogonalizing the last j observations of y_t with respect to Y_{SN}, we obtain

$$\begin{aligned}
&\mathcal{L}_T(\theta, \theta_D, \Sigma; Y_T) = \\
&\quad (\log\det\Gamma_{SM}(\theta,\Sigma) + \log\det\Gamma_{T,22}(\theta,\Sigma) + (Y_{SM} - \mathbf{D}_{SM})'\Gamma_{SM}(\theta,\Sigma)^{-1}(Y_{SM} - \mathbf{D}_{SM}) + \\
&\quad \hat{E}_T(\theta,\theta_D)'\Gamma_{T,22}(\theta,\Sigma)^{-1}\hat{E}_T(\theta,\theta_D))/T = \\
&\quad \frac{SM}{T}\mathcal{L}_{SM}(\theta,\theta_D,\Sigma;Y_T) + \frac{\log\det\Gamma_{T,22}(\theta,\Sigma)}{T} + \hat{E}_T(\theta,\theta_D)'\Gamma_{T,22}(\theta,\Sigma)^{-1}\hat{E}_T(\theta,\theta_D)/T,
\end{aligned}$$

where

$$\hat{E}_T(\theta,\theta_D) = Y_{SM+1,T-SM} - \mathbf{D}_{SM+1,T-SM} - \beta_{T-SM,SM}(Y_{SM} - \mathbf{D}_{SM+1,T-SM})$$

and $\Gamma_{T,22}(\theta,\Sigma)$ denotes the corresponding variance matrix according to θ, θ_D, Σ.
For $\hat{\theta}_{SM}$ it follows that $\Gamma_{(S+1)M,22}(\hat{\theta}_{SM}, \hat{\Sigma}_{SM}) \to \mathbb{E}\left(E_{t,S}E_{t,S}'\right) < \infty$ and therefore, (since $\Gamma_{T,22}(\hat{\theta}_{SM}, \hat{\Sigma}_{SM})$ is a submatrix of $\Gamma_{(S+1)M,22}(\hat{\theta}_{SM}, \hat{\Sigma}_{SM})$)

$$(\log\det\Gamma_{T,22}(\hat{\theta}_{SM}, \hat{\Sigma}_{SM}))/T = o_P(1).$$

Furthermore, $\left\langle \hat{E}_{Sm}(\hat{\theta}_{SM}, \hat{\theta}_{D,SM}), \hat{E}_{Sm}(\hat{\theta}_{SM}, \hat{\theta}_{D,SM}) \right\rangle \to \mathbb{E}(E_{t,S}E_{t,S}')$ implies that

$$\hat{E}_T(\hat{\theta}_{SM}, \hat{\theta}_{D,SM})\hat{E}_T(\hat{\theta}_{SM}, \hat{\theta}_{D,SM})'/T = o_P(1).$$

Thus, $\mathcal{L}_T(\hat{\theta}_{SM}, \hat{\theta}_{D,SM}, \hat{\Sigma}_{SM}; Y_T) = \frac{SM}{T}\mathcal{L}_{SM}(\hat{\theta}_{SM}, \hat{\theta}_{D,SM}; Y_T) + o_P(1) \to \mathcal{L}_\infty(\theta_\circ, \theta_{D,\circ}, \Sigma_\circ)$. This implies

$$\mathcal{L}_T(\hat{\theta}_T, \hat{\theta}_{D,T}, \hat{\Sigma}_T; Y_T) \leq \mathcal{L}_T(\hat{\theta}_{SM}, \hat{\theta}_{D,SM}, \hat{\Sigma}_{SM}; Y_T) \to \mathcal{L}_\infty(\theta_\circ, \theta_{D,\circ}, \Sigma_\circ).$$

Using the developments of the proof for the I(1) case and Hannan and Deistler (1988) showing that $\log\det\Gamma_T(\hat{\theta}, \hat{\Sigma})/T \geq \log\det\hat{\Sigma}_T + o_P(1)$, this implies $\log\det\hat{\Sigma}_T \leq \log\det\Sigma_0 + o_P(1)$. Then, following the same arguments as Hannan and Deistler (1988) on the bottom of p. 121 and (4.2.25), see also below, it is straightforward to show that the smallest eigenvalue of $\hat{\Sigma}_T$ is bounded away from zero. This follows, since leaving out the last only partially observed time instant reduced the term $Q_T(\theta)$ while resulting in the same lower bound.
Consequently, $\Gamma_{T,22}(\hat{\theta}_T, \hat{\Sigma}_T) \geq I_{T-SM}\lambda_{\min}(\hat{\Sigma}_T)$, where $\lambda_{\min}(\hat{\Sigma}_T)$ is bounded from below by a constant $c > 0$ from above. Therefore, we obtain $\log\det\Gamma_{T,22}(\hat{\theta}_T, \hat{\Sigma}_T) \geq (T - SM)\log c$. Consequently,

$$\begin{aligned}
\frac{SM}{T}\mathcal{L}_{SM}(\hat{\theta}_T, \hat{\theta}_{D,T}, \hat{\Sigma}_T; Y_T) &\leq \mathcal{L}_T(\hat{\theta}_T, \hat{\theta}_{D,T}, \hat{\Sigma}_T; Y_T) - \frac{(T-SM)\log c}{T} \\
&\leq \mathcal{L}_T(\hat{\theta}_{SM}, \hat{\theta}_{D,SM}, \hat{\Sigma}_{SM}; Y_T) + o_P(1) \\
&\leq \frac{SM}{T}\mathcal{L}_{SM}(\hat{\theta}_{SM}, \hat{\theta}_{D,SM}, \hat{\Sigma}_{SM}; Y_T) + o_P(1) \to \mathcal{L}_\infty(\theta_\circ, \theta_{D,\circ}, \Sigma_\circ).
\end{aligned}$$

Thus, it follows that $\mathcal{L}_{SM}(\hat{\theta}_T, \hat{\theta}_{D,T}, \hat{\Sigma}_T; Y_T) \to \mathcal{L}_\infty(\theta_\circ, \theta_{D,\circ}, \Sigma_\circ)$ and hence we found a sequence such that the likelihood value converges to the minimum of the limiting criterion function. Thus, from identifiability of the parameters it follows that $\hat{\theta}_T \to \theta_\circ, \hat{\theta}_{D,T} \to \theta_{D,\circ}$ and thus consistency carries over from the $I(1)$ case to the MFI(1) case with unit roots frequencies which are rational multiples of 2π.
For the prediction error criterion function $\mathcal{L}_{PE,T}$ the arguments are almost unchanged (but simpler, as in this case the log det term equals $\log\det\Sigma$ directly) and hence omitted.
With respect to the convergence rates note that the matrix C and $\mathbf{D}_{t,S}$ are contained in the system representation for $Y_{t,S}$ in (B.2). Therefore, the convergence rates for the I(1) case in (B.2) imply the ones for the MFI(1) case. □

It is thus sufficient to prove Theorem 2.1 for the I(1) case. In this case the DGP is given by:

$$\begin{aligned} y_t &= C_1 x_{t,1} + C_\bullet x_{t,\bullet} + \varepsilon_t + \mathbf{d}_t, \\ x_{t+1,1} &= x_{t,1} + B_1 \varepsilon_t, \quad x_{1,1} = 0, \\ x_{t+1,\bullet} &= A_\bullet x_{t,\bullet} + B_\bullet x_{t,\bullet}, \quad x_{1,\bullet} = \sum_{j=0}^{\infty} A_\bullet^j B_\bullet \varepsilon_{-j}, \end{aligned}$$

where $C_1 \in \mathbb{R}^{s\times c}, c \leq s$ is such that $C_1'C_1 = I_c$. $(A_\bullet, B_\bullet, C_\bullet)$ denotes the stable subsystem of state order $n_\bullet = n - c$. Let $P_C = C_1C_1', P_\perp = I_s - P_C$ denote the projection onto the column space of C_1 and its orthocomplement respectively. Then, with $\mathbf{d}_t = Ds_t, s_t = [1, t]'$

$$\begin{aligned} \tilde{y}_t - \tilde{\mathbf{d}}_t &= (P_C\Delta + P_\perp)(y_t - \mathbf{d}_t) = C_1\Delta x_{t,1} + P_C C_\bullet \Delta x_{t,\bullet} + P_C\Delta\varepsilon_t + P_\perp C_\bullet x_{t,\bullet} + P_\perp\varepsilon_t \\ &= C_1 B_1 \varepsilon_{t-1} + P_C C_\bullet x_{t,\bullet} - P_C C_\bullet x_{t-1,\bullet} - P_C\varepsilon_{t-1} + P_\perp C_\bullet x_{t,\bullet} + \varepsilon_t \\ &= C_1(B_1\varepsilon_{t-1} - C_1'C_\bullet x_{t-1,\bullet} - C_1'\varepsilon_{t-1}) + C_\bullet x_{t,\bullet} + \varepsilon_t \end{aligned}$$

is the sum of a stationary process with a deterministic part

$$\tilde{\mathbf{d}}_t = (I - P_C L)Ds_t = Ds_t - P_C Ds_{t-1} = Ds_t - P_C D\mathcal{S}^{-1}s_t = \underbrace{\left(D - P_C D\mathcal{S}^{-1}\right)}_{\tilde{D}} s_t$$

and hence can be written as:

$$\tilde{y}_t = \tilde{k}(L)\varepsilon_t + (P_C\Delta + P_\perp)Ds_t = \tilde{k}(L)\varepsilon_t + Ds_t - P_C Ds_{t-1} = \tilde{k}(L)\varepsilon_t + \tilde{D}s_t,$$

where the solutions processes to the stationary transfer function $\tilde{k}(z)$ have the representation:

$$\tilde{y}_t - \tilde{D}s_t = C_1 z_t + C_\bullet x_{t,\bullet} + \varepsilon_t, \quad \begin{bmatrix} z_{t+1} \\ x_{t+1,\bullet} \end{bmatrix} = \begin{bmatrix} 0 & -C_1'C_\bullet \\ 0 & A_\bullet \end{bmatrix} \begin{bmatrix} z_t \\ x_{t,\bullet} \end{bmatrix} + \begin{bmatrix} B_1 - C_1' \\ B_\bullet \end{bmatrix} \varepsilon_t.$$

This follows, since $s_t = \mathcal{S}s_{t-1}$ for $\mathcal{S} := \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ by construction of the vector s_t of deterministic terms.
The transfer function $\tilde{k}(z)$ is stable, since $A_\bullet$ is. The zeros are the eigenvalues of

$$\begin{bmatrix} 0 & -C_1'C_\bullet \\ 0 & A_\bullet \end{bmatrix} - \begin{bmatrix} B_1 - C_1' \\ B_\bullet \end{bmatrix} [C_1 \quad C_\bullet] = \begin{bmatrix} I_c - B_1C_1 & -B_1C_\bullet \\ -B_\bullet C_1 & A_\bullet - B_\bullet C_\bullet \end{bmatrix}.$$

Note that this equals $A - BC$, where (A, B, C) denotes the original system in canonical form. Hence, invertibility of $k(z)$ implies invertibility of $\tilde{k}(z)$.
Therefore, the Gaussian likelihood function for $Y_T := [y_1', ..., y_T']'$ can be calculated using $\tilde{Y}_T := [y_1', \tilde{y}_2', ..., \tilde{y}_T']'$ which – being a linear invertible transformation of Y_T is also Gaussian distributed. Note that $y_1 = C_\bullet x_{1,\bullet} + \varepsilon_1 + Ds_1$ for $x_{1,1} = 0$. Then, $-2/T$ times the Gaussian likelihood function for Y_T equals:

$$\begin{aligned} L_T(k, D, \Sigma) &= \frac{1}{T}\left(\log|\Gamma_T(k,\Sigma)| + (Y_T - (I_T \otimes D)\mathbf{D}_T)'\Gamma_T(k,\Sigma)^{-1}(Y_T - (I_T \otimes D)\mathbf{D}_T)'\right) \\ &= \frac{1}{T}\left(\log|\Gamma_T(\tilde{k},\Sigma)| + (\tilde{Y}_T - (I_T \otimes \tilde{D})\mathbf{D}_T)'\Gamma_T(\tilde{k},\Sigma)^{-1}(\tilde{Y}_T - (I_T \otimes \tilde{D})\mathbf{D}_T)'\right), \end{aligned}$$

where the dependence of the covariance matrix Γ_T on the transfer function k or $\tilde{k}$ respectively and the noise variance Σ is emphasized, while the influence of the variance of the initial state $\mathrm{diag}(0, P_\bullet(\theta, \sigma))$ is neglected. Except for the inclusion of y_1 this is identical to the criterion function used in Hannan and Deistler (1988, Section 4.2).
The domain of the transfer function $\tilde{k}$ here is defined analogously to the sets Θ in Hannan and Deistler (1988, p. 110ff.): Let $\Theta \subset M_{n_\bullet} \times \underline{P} \times \underline{D} \times \underline{\Sigma}$ equal the product of the set of marginally stable

(having no poles within the closed unit disc) and minimum-phase (no zeros within the closed unit disc) transfer functions of order smaller or equal to n with the sets $\underline{P}$ (space of projector matrices in $\mathbb{R}^{s\times s}$ with rank equal to c), the set $\underline{D}$ (the vectorization of all $s \times d$ real matrices) and the set $\underline{\Sigma}$ (the set of all $s \times s$ symmetric positive definite matrices). This set is endowed with the product topology of the pointwise topology for $M_{n_\bullet}$, with the gap metric for projector matrices and with the Euclidean topology for the two sets of matrices. Then, $\overline{\Theta}$ denotes the corresponding closure, $\hat{\Theta} \subset \overline{\Theta}$ contains only stable transfer functions and $\Theta^* \subset \hat{\Theta}$ in addition strictly minimum-phase transfer functions without zeros on the unit circle.
Note that for $k(z)$ as in the theorem it follows that $\tilde{k}(z) \in M_{n_\bullet}$ (see above). Furthermore, given P_C there is a 1-1 mapping between the k and $\tilde{k}$:

$$\tilde{k}(z) = (P_C\Delta + P_\perp)k(z) \Rightarrow k(z) = (P_C\Delta + P_\perp)^{-1}\tilde{k}(z).$$

As seen above under the assumption of stable invertibility of the true transfer function $k(z)$, it follows that $\tilde{k}(z)$ is stably invertible. It further follows that the parameters for the transfer function k can be partitioned into a set that parameterizes the column space of C_1 relating to $P_C = C_1C_1'$ and the remaining ones relating to $\tilde{k}$. This conforms with the parameterization suggested in Bauer et al. (2020).
The proof of Theorem 2.1 is based on slightly adapting (and punctually slightly extending) the arguments of Hannan and Deistler (1988, Section 4.2., p. 110 ff) (called HD henceforth). Therefore, we also use the notation of HD referring to the quadruple (k, P_C, D, Σ) as θ in this section. Let us stress here again that θ thus here is no real valued parameter. As in HD, Section 4.2. our proof is coordinate independent using only the transfer functions and not the particular form of their realizations.
The pseudo maximum likelihood estimate $\hat{\theta}$ is obtained by minimizing $L_T(\theta) = \tilde{L}_T(\tilde{k}, P_C, D, \Sigma) = L_T(k, D, \Sigma)$ over Θ.
Now follow the proof of Theorem 4.2.1. of HD. As on p. 112 we define $u(t, \theta) = \tilde{y}_t - \tilde{D}s_t, t = 2, .., T$, where $u(1, \theta) = y_1 - \tilde{D}s_1$ and $u_T(\theta) = \{u(t, \theta)\}_{t=1,...,T}$. Note that here $u(1, \theta)$ deviates from the 'regular' definition $\tilde{y}_1 - \tilde{D}s_1 = u(1, \theta) - P_C y_0$. Using this, we define

$$L_T(\theta) = T^{-1} \log |\tilde{\Gamma}_T(\theta)| + T^{-1}u_T(\theta)\tilde{\Gamma}_T(\theta)^{-1}u_T(\theta),$$

where the dependence on $\theta = (k, P_C, D, \Sigma)$ is stressed. Here, $\tilde{\Gamma}_T(\theta) = \tilde{\Gamma}_T(\tilde{k}, \Sigma)$. Note that $L_T(\theta)$ depends on k only via $\tilde{k}$ and thus can be seen as a function of $\tilde{\theta} = (\tilde{k}, P_C, \Sigma, \tilde{D})$. Consequently, maximizing L_T over $\theta \in \Theta$ is equivalent to maximizing the corresponding function

$$\tilde{L}_T(\tilde{\theta}) = L_T(\theta) = T^{-1} \log |\tilde{\Gamma}_T(\tilde{\theta})| + T^{-1}u_T(\tilde{\theta})'\tilde{\Gamma}_T(\tilde{\theta})^{-1}u_T(\tilde{\theta}),$$

where $\theta = (k, P_C, D, \Sigma)$ maps onto $\tilde{\theta} = (\tilde{k}, P_C, \tilde{D}, \Sigma) = \Lambda(\theta)$ over the corresponding set $\tilde{\Theta} = \Lambda(\Theta) \subset \overline{\Theta}$.
Then, using the arguments of HD on p. 112, it follows that $\tilde{L}_T$ is finite on $\hat{\Theta}$ (as $\tilde{k}$ is stable there). This follows since all entries are bounded and the matrix $\tilde{\Gamma}_T(\theta)$ is non-singular. Otherwise there would be a relation defined via vectors $\alpha_j, j = 0, ..., T-1$ such that

$$\sum_{j=0}^{T-1} \alpha_j'(y_{T-j} - Ds_{T-j}) = 0 = \sum_{j=0}^{T-1} \alpha_j'(\varepsilon_{T-j} + Cx_{T-j}).$$

If $\alpha_0 \neq 0$, then the term $\alpha_0'\varepsilon_T$ would be contained in the linear span of the past of ε_t and x_t, which contradicts the white noise assumption.Therefore, $\alpha_0 = 0$. But then $\alpha_1 = 0$ follows. Continuing the recursion shows that $\alpha_j = 0, \forall j$.
Consider $\tilde{\Gamma}_T(\tilde{\theta})$ in more depth: It is defined as the variance of

$$u_T(\theta) = \tilde{Y}_{1,T}(\tilde{\theta}) + \begin{pmatrix} P_C y_0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \Rightarrow \tilde{\Gamma}_T(\tilde{\theta}) = \begin{bmatrix} V(y_1) & \mathrm{Cov}(y_1, \tilde{Y}_{2,T}(\tilde{\theta})) \\ \mathrm{Cov}(\tilde{Y}_{2,T}(\tilde{\theta}), y_1) & \Gamma_{T-1}(\tilde{\theta}) \end{bmatrix},$$

where $\tilde{Y}_{i,T}(\tilde{\theta}) = [\tilde{y}_t(\theta) - \tilde{D}s_t]_{t=i,\ldots,T}$. Here, $y_1 = C_\bullet x_{1,\bullet} + \varepsilon_1 + Ds_1$. This implies, using the block matrix inversion, that

$$\tilde{\Gamma}_T(\tilde{\theta})^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & \Gamma_{T-1}(\tilde{\theta})^{-1} \end{bmatrix} + \begin{bmatrix} I_s \\ -\Gamma_{T-1}(\tilde{\theta})^{-1}\mathrm{Cov}(\tilde{Y}_{2,T}(\tilde{\theta}), y_1) \end{bmatrix} V_\pi(\tilde{\theta})^{-1} \begin{bmatrix} I_s & -\mathrm{Cov}(y_1, \tilde{Y}_{2,T}(\tilde{\theta}))\Gamma_{T-1}(\tilde{\theta})^{-1} \end{bmatrix}, \tag{B.3}$$

where $V_\pi(\tilde{\theta}) := V(y_1) - \mathrm{Cov}(y_1, \tilde{Y}_{2,T}(\tilde{\theta}))\Gamma_{T-1}(\tilde{\theta})^{-1}\mathrm{Cov}(\tilde{Y}_{2,T}(\tilde{\theta})), y_1)$. Now, for stable $\tilde{\theta}$ it follows that $V_\pi(\tilde{\theta}) \geq I_s\lambda_{min}(\Sigma)$.
It follows that

$$u_T(\tilde{\theta})'\tilde{\Gamma}_T(\theta)^{-1}u_T(\tilde{\theta})/T \geq u_{2:T}(\tilde{\theta})'\Gamma_{T-1}(\tilde{\theta})^{-1}u_{2:T}(\tilde{\theta})/T,$$

where $u_{2:T}(\tilde{\theta}) = \{u_t(\tilde{\theta})\}_{t=2,\ldots,T} = \tilde{Y}_{2,T}(\tilde{\theta})$ is used in order to be closer to the notation in HD. Note that the right hand side term has exactly the same form as the second term of the log-likelihood function dealt with in Chapter 4 of HD. Further, note that the difference between these two terms equals

$$y_{1,\pi}(\tilde{\theta})'V_\pi(\tilde{\theta})^{-1}y_{1,\pi}(\tilde{\theta})/T,$$

where this equation defines $y_{1,\pi}(\tilde{\theta})$.
It follows that the criterion function to be considered equals

$$\tilde{L}_T(\tilde{\theta}) \geq \log|\tilde{\Gamma}_T(\tilde{\theta})| + u_{2:T}(\tilde{\theta})'\Gamma_{T-1}(\tilde{\theta})^{-1}u_{2:T}(\tilde{\theta})/T = \log|\tilde{\Gamma}_T(\tilde{\theta})| + \tilde{Q}_T(\tilde{\theta}).$$

In the following we will use the arguments of HD to deal with these two terms.

The log det term

In this subsection the asymptotic behavior of $\log|\tilde{\Gamma}_T(\tilde{\theta})|$ is investigated. This term is relatively easy to deal with, since it is not influenced by the data or by $\tilde{D}$. Using the definitions above we obtain

$$\begin{bmatrix} I_s & -\mathrm{Cov}(y_1, \tilde{Y}_{2,T}(\tilde{\theta}))\Gamma_{T-1}(\tilde{\theta})^{-1} \\ 0 & I \end{bmatrix} \tilde{\Gamma_T}(\tilde{\theta}) \begin{bmatrix} I_s & 0 \\ -\Gamma_{T-1}(\tilde{\theta})^{-1}\mathrm{Cov}(\tilde{Y}_{2,T}(\tilde{\theta}), y_1) & I \end{bmatrix} = \begin{bmatrix} V_\pi(\tilde{\theta}) & 0 \\ 0 & \Gamma_{T-1}(\tilde{\theta}) \end{bmatrix}.$$

Therefore, the determinant is the product of $|V_\pi(\tilde{\theta})|$ and $|\Gamma_{T-1}(\tilde{\theta})|$. From $V_\pi(\theta) \leq V(y_1)$ we see that

$$\frac{1}{T}\log\det\tilde{\Gamma}_T(\tilde{\theta}) = \frac{1}{T}\log|V_\pi(\tilde{\theta})| + \frac{1}{T}\log|\Gamma_{T-1}(\tilde{\theta})| \leq \frac{1}{T}\log|\Gamma_{T-1}(\tilde{\theta})| + \frac{1}{T}\log|V(y_1)|.$$

The behavior of $T^{-1}\log|\Gamma_{T-1}(\tilde{\theta})|$ follows as in HD, Lemma 4.2.2., p. 116: $T^{-1}\log|\Gamma_{T-1}(\tilde{\theta})| \geq \log|\Sigma|$ and $\lim_{T\to\infty} T^{-1}\log|\Gamma_{T-1}(\tilde{\theta})| = \log|\Sigma|$ for $\tilde{\theta} \in \hat{\Theta}$. For $\theta_j \to \tilde{\theta}_0 \in \Theta - \hat{\Theta}$, such that $\tilde{k}_0$ contains a pole on the unit circle, where $\theta_j \in \hat{\Theta}$, we have $\log|\Gamma_{T-1}(\theta_j)| \to \infty$ as $\lambda_{|max|}(P_\bullet) \to \infty$. For $\tilde{\theta} \in \hat{\Theta}$ we have

$$0 < V_\pi(\tilde{\theta}) \leq V(y_1) < \infty$$

such that $\lim_{T\to\infty} T^{-1}\log|\tilde{\Gamma}_T(\tilde{\theta})| = \log|\Sigma|$.
For $\theta_j \to \tilde{\theta}_0 \in \Theta - \hat{\Theta}$ we have $V_\pi(\theta_j) \geq \lambda_{min}(\Sigma)I_s$, which hence must hold also in the limit. Consequently, in this case $T^{-1}\log|\tilde{\Gamma}_T(\theta_j)| \to \infty$. Thus, we obtain the same asymptotic behavior as in HD.

The quadratic term Q_T

The second component of the criterion function is the term

$$\begin{aligned} Q_T(\tilde{\theta}) &= T^{-1}u_T(\tilde{\theta})'\tilde{\Gamma}_T(\tilde{\theta})^{-1}u_T(\tilde{\theta}) \\ &= T^{-1}y_{1,\pi}(\tilde{\theta})'V_\pi(\tilde{\theta})^{-1}y_{1,\pi}(\tilde{\theta}) + T^{-1}u_{2:T}(\tilde{\theta})'\Gamma_{T-1}(\tilde{\theta})^{-1}u_{2:T}(\tilde{\theta}) \\ &\geq T^{-1}u_{2:T}(\tilde{\theta})'\Gamma_{T-1}(\tilde{\theta})^{-1}u_{2:T}(\tilde{\theta}) =: \tilde{Q}_T(\tilde{\theta}), \end{aligned}$$

using the block matrix inversion. HD define the function $Q(\tilde{\theta})$ as the limit of $\tilde{Q}_T$ (on Θ^*). Here, (suppressing the dependence of $\tilde{k}(e^{-i\omega})$ on the frequency ω)

$$Q(\tilde{\theta}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}[(\tilde{k}\Sigma\tilde{k}^*)^{-1}(\tilde{k}_0\Sigma_0\tilde{k}_0^*))]d\omega.$$

HD add terms related to exogenous inputs potentially including deterministic terms. We will, however, deal differently with them here.
The next step in HD is crucial for avoiding problems with non-invertible transfer functions in $\overline{\Theta}$. In order to avoid problems with the term involving $(\tilde{k}\Sigma\tilde{k}^*)^{-1}$ due to zeros of $\tilde{k}$ on the unit circle, in $Q(\tilde{\theta})$ for $\tilde{k}(z) = N(z)/c(z)$, HD introduce a regularization term such that

$$\begin{aligned} \phi_\eta(\omega;\tilde{\theta}) &= 2\pi|c(e^{i\omega})|^2\{N(e^{i\omega})\Sigma N(e^{i\omega})^* + \eta I_s\}^{-1} \\ &\leq 2\pi|c(e^{i\omega})|^2\{N(e^{i\omega})\Sigma N(e^{i\omega})^*\}^{-1} = f_u^{-1}(\omega), \quad \forall\eta > 0, \end{aligned}$$

as a replacement for $(\tilde{k}\Sigma\tilde{k}^*)^{-1}$ in the definition of $\Gamma_T(\tilde{\theta})$. As the covariances are functions of the spectrum for stationary processes, HD state that the covariance matrix $\Gamma_T(\tilde{\theta})$ can be written as $\Gamma_T(f_u)$, where $f_u = \tilde{k}\Sigma\tilde{k}^*$. Moreover, they show that for $\phi_\eta^{-1} \geq f_u$ it holds that $\Gamma_T(\phi_\eta^{-1}) \geq \Gamma_T(f_u)$ and thus $\Gamma_T(\phi_\eta^{-1})^{-1} \leq \Gamma_T(f_u)^{-1}$, see HD, (4.2.18) on p. 119.
Then, HD consider the regularized version

$$\tilde{Q}_{T,\eta}(\tilde{\theta}) := u_{2:T}(\tilde{\theta})'\Gamma_{T-1}(\phi_\eta^{-1})^{-1}u_{2:T}(\tilde{\theta})/T \leq \tilde{Q}_T(\tilde{\theta}) := u_{2:T}(\tilde{\theta})'\Gamma_{T-1}(\tilde{\theta})^{-1}u_{2:T}(\tilde{\theta})/T.$$

Lemma 4.2.3. of HD shows that (i) $\tilde{Q}_T(\theta) \to Q(\theta), \theta \in \Theta^*$ and (ii) $\tilde{Q}_{T,\eta}(\theta) \to Q_\eta(\theta)$ uniformly in $\Theta_{c_1c_2c_3} \cap \hat{\Theta}$ where c_1, c_2 denotes constants bounding the eigenvalues of $0 < c_1 I_s \leq \Sigma \leq c_2 I_s$ for $\theta \in \Theta_{c_1c_2c_3}$ and c_3 bounds the entries of certain matrix polynomials (see HD, p. 118) corresponding to the transfer functions. These restrictions are sufficient to make $\Theta_{c_1c_2c_3}$ a compact set.
The intersection with $\hat{\Theta}$, wherein all transfer functions are stable, is not necessary for the argument. The only place, where stability enters the proof is in the strict lower bound of $P_{2,i}$ in the first display on p. 121. However, as $\Theta_{c_1c_2c_3}$ potentially also contains transfer functions with unit roots, the arguments also need to extend to the case where $P_{2,i}(\omega) = 0$ for some ω.
The regularization leads to a uniform lower bound of the eigenvalues of $\Gamma_T(\phi_\eta^{-1})$ on $\Theta_{c_1c_2c_3}$ such that the inverse can be uniformly bounded.
With respect to the pointwise convergence of $\tilde{Q}_T(\theta)$ in our setting we use the arguments of HD. For easier notation in the following we relabel the sample size by replacing $T-1$ with T, start indexing at $t = 2$ and using θ in place of $\tilde{\theta}$ for the remainder of this proof.
The proof of Lemma 4.2.3. proceeds by bounding the spectrum $(\tilde{k}\Sigma\tilde{k}^*)^{-1}$ below and above by spectra, P say, corresponding to autoregressive processes with lag length M. This is possible for $\theta \in \Theta^*$ which contains only stable and invertible transfer functions. Now for these autoregressive processes one finds matrices $\mathcal{C}$ such that (see HD, p. 119, bottom display)

$$\Gamma_T(P^{-1})^{-1} = \mathcal{C}'\begin{pmatrix} \Gamma_M^{-1} & 0 & \dots & 0 \\ 0 & \Sigma_P^{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \Sigma_P^{-1} \end{pmatrix}\mathcal{C}.$$

Here $\mathcal{C}$ is a lower triangular matrix, whose first M block rows are identical with the identity matrix while the remaining ones contain the autoregressive coefficients C_j:

$$\mathcal{C} = \begin{bmatrix} I_M & & & \vdots & & \\ \cdots & \cdots & \cdots & . & & \\ C_M & C_{M-1} & ... & C_0 & 0 & \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots \\ & & & & & \vdots \\ 0 & ... & C_M & C_{M-1} & ... & C_0 \end{bmatrix}.$$

It follows that

$$u_T(\theta)'\Gamma_T(P^{-1})^{-1}u_T(\theta) = u_M(\theta)'\Gamma_M^{-1}u_M(\theta)/T + T^{-1}\sum_{t=M+1}^{T}(\sum_{j=0}^{M}C_j u(t-j,\theta))'\Sigma_P^{-1}(\sum_{j=0}^{M}C_j u(t-j,\theta)).$$

The first term clearly tends to zero as $T \to \infty$, since M is fixed, whereas the second can be rewritten as

$$\text{tr}\left(\Sigma_P^{-1}[C_M, C_{M-1}, ..., C_0]\left[T^{-1}\sum_{t=M+1}^{T}U_{t,M}(\theta)U_{t,M}(\theta)'\right][C_M, C_{M-1}, ..., C_0]'\right).$$

Therefore, we need to investigate $U_{t,M}(\theta)$ in more depth: Here,

$$\begin{aligned} u(t,\theta) &= \tilde{y}_t - \tilde{D}s_t = y_t - P_C y_{t-1} - \tilde{D}s_t = (\tilde{D}_0 - \tilde{D})s_t + C_0 x_t - P_C C_0 x_{t-1} + \varepsilon_t - P_C\varepsilon_{t-1} \\ &= (\tilde{D}_0 - \tilde{D})s_t + \varepsilon_t - P_C\varepsilon_{t-1} + C_0(A_0 x_{t-1} + B_0\varepsilon_{t-1}) - P_C C_0 x_{t-1} \\ &= (\tilde{D}_0 - \tilde{D})s_t + \varepsilon_t + (C_0 B_0 - P_C)\varepsilon_{t-1} + (C_0 A_0 - P_C C_0)x_{t-1} \\ &= \underbrace{(\tilde{D}_0 - \tilde{D})s_t}_{\varphi_t(\theta)} + \underbrace{\varepsilon_t + (C_0 B_0 - P_C)\varepsilon_{t-1} + (C_{0,\bullet}A_{0,\bullet} - P_C C_{0,\bullet})x_{t-1,\bullet}}_{v_{t,\bullet}(\theta)} + \underbrace{(I_s - P_C)C_{0,1}x_{t-1,1}}_{v_{t,u}(\theta)}. \end{aligned}$$

Therefore, for every $\theta \in \hat{\Theta}$ the process $u_t(\theta)$ contains three components: a deterministic part $\varphi_t(\theta)$ dominated by $(\tilde{D}_0 - \tilde{D})$, a stationary component (denoted with $v_{t,\bullet}(\theta)$ above) and the term $v_{t,u}(\theta) := (I_s - P_C)C_{0,1}x_{t-1,1} = P_\perp C_{0,1}x_{t-1,1}$, which is integrated if $P_\perp C_{0,1} \neq 0$ and zero else. Therefore,

$$\begin{aligned} U_{t,M}(\theta) &= \begin{pmatrix} u(t-M,\theta) \\ u(t-M+1,\theta) \\ \vdots \\ u(t,\theta) \end{pmatrix} = \begin{pmatrix} \varphi_{t-M}(\theta) + v_{t-M,\bullet}(\theta) + P_\perp C_{0,1}x_{t-M-1,1} \\ \varphi_{t-M+1}(\theta) + v_{t-M+1,\bullet}(\theta) + P_\perp C_{0,1}x_{t-M,1} \\ \vdots \\ \varphi_t(\theta) + v_{t,\bullet}(\theta) + P_\perp C_{0,1}x_{t-1,1} \end{pmatrix} \\ &= \begin{pmatrix} \varphi_{t-M}(\theta) \\ \varphi_{t-M+1}(\theta) \\ \vdots \\ \varphi_t(\theta) \end{pmatrix} + \begin{pmatrix} v_{t-M,\bullet}(\theta) \\ v_{t-M+1,\bullet}(\theta) + P_\perp C_{0,1}B_{0,1}\varepsilon_{t-M-1} \\ \vdots \\ v_{t,\bullet}(\theta) + P_\perp C_{0,1}B_{0,1}\sum_{j=0}^{M-1}\varepsilon_{t-j} \end{pmatrix} + \begin{pmatrix} P_\perp C_{0,1} \\ \vdots \\ P_\perp C_{0,1} \end{pmatrix} x_{t-M-1,1} \\ &= D_{t,M}(\theta) + V_{t,M}(\theta) + \begin{pmatrix} P_\perp C_{0,1} \\ \vdots \\ P_\perp C_{0,1} \end{pmatrix} x_{t-M-1,1} \\ &= V_{t,M}(\theta) + D(\theta;M)s_t + P_M x_{t-M-1,1} \\ &= V_{t,M}(\theta) + [D(\theta;M), P_M]\begin{pmatrix} s_t \\ x_{t-M-1,1} \end{pmatrix} = V_{t,M}(\theta) + \Psi_M(\theta)z_{t,M}. \end{aligned}$$

Here, the last equation defines $\Psi_M(\theta)$ and $z_{t,M}$. Let in the following the superscript $.^\pi$ denote the residuals of a regression onto $z_{t,M}$ with the corresponding fitted values denoted as $.^z$ such that

$$U_{t,M}(\theta) = U_{t,M}(\theta)^\pi + U_{t,M}(\theta)^z = V_{t,M}(\theta)^\pi + U_{t,M}(\theta)^z,$$

since $D_{t,M}(\theta)^\pi = 0, x^\pi_{t-M-1,1} = 0$. It follows that

$$\langle U_{t,M}(\theta), U_{t,M}(\theta)\rangle = \langle V_{t,M}(\theta)^\pi, V_{t,M}(\theta)^\pi\rangle + \langle U_{t,M}(\theta)^z, U_{t,M}(\theta)^z\rangle \geq \langle V_{t,M}(\theta)^\pi, V_{t,M}(\theta)^\pi\rangle .$$

Furthermore, for fixed M we have

$$\langle V_{t,M}(\theta)^\pi, V_{t,M}(\theta)^\pi\rangle = \langle V_{t,M}(\theta), V_{t,M}(\theta)\rangle + o_P(1), \tag{B.4}$$

as regressing out integrated processes, the constant, seasonal terms and a linear trend from stationary processes leads to negligible terms. If no deterministic terms are present, the negligible term is also $o(1)$.
It is now easy to verify that for $\tilde{D} = \tilde{D}_0$ and $P_\perp C_{0,1} = 0$ the term $\tilde{Q}_T(\theta)$ converges to $Q(\theta)$, since the second moments $\langle V_{t,M}(\theta), V_{t,M}(\theta)\rangle$ converge in this case and $U_{t,M}(\theta) = V_{t,M}(\theta)$ holds then.
In the general case we obtain

$$\begin{aligned} Q_T(\theta) \geq \tilde{Q}_T(\theta) &\geq \operatorname{tr}\left(\Sigma_P^{-1}[C_M, C_{M-1}, ..., C_0]\,\langle U_{t,M}(\theta), U_{t,M}(\theta)\rangle\,[C_M, C_{M-1}, ..., C_0]'\right) \\ &\geq \operatorname{tr}\left(\Sigma_P^{-1}[C_M, C_{M-1}, ..., C_0]\,\langle V_{t,M}(\theta)^\pi, V_{t,M}(\theta)^\pi\rangle\,[C_M, C_{M-1}, ..., C_0]'\right) \\ &= \operatorname{tr}\left(\Sigma_P^{-1}[C_M, C_{M-1}, ..., C_0]\,[\langle V_{t,M}(\theta), V_{t,M}(\theta)\rangle + o_P(1)]\,[C_M, C_{M-1}, ..., C_0]'\right). \end{aligned}$$

Jointly we obtain that for fixed $\theta \in \Theta^*$ for term $Q_T(\theta)$ it holds that

$$\liminf_{T\to\infty} Q_T(\theta) \geq Q(\theta).$$

For $\tilde{\theta}_0 = \Lambda(\theta_0)$ the additional terms due to $\tilde{D} - \tilde{D}_0$ and $P_\perp C_{0,1}$ are zero and hence in this case $Q_T(\tilde{\theta}_0) \to s$.
Replacing $\tilde{Q}_T(\theta)$ with the corresponding $\tilde{Q}_{T,\eta}(\theta)$ and noting that $Q_T(\theta) \geq \tilde{Q}_T(\theta) \geq \tilde{Q}_{T,\eta}(\theta)$ for all $\eta > 0, \theta \in \overline{\Theta}$, we obtain uniformly in $\theta \in \Theta_{c_1c_2c_3}$ (a compact space) that

$$\liminf_{T\to\infty} \inf_{\Theta_{c_1c_2c_3}} (Q_T(\theta) - Q_\eta(\theta)) \geq 0.$$

This follows, since by taking the liminf all non-negative terms can be neglected. It is simple to verify that the convergence in (B.4) is uniform in the parameter set, as $v_{t,\bullet}(\theta)$ only depends on the parameter vector via P_C which varies in a compact set. The remaining arguments are as in HD, p. 119-121. In particular we only have to investigate a finite number of spectra P with a corresponding finite number of lag lengths M, since the set $\Theta_{c_1c_2c_3}$ is compact. Then, the convergence results are standard.
This implies that for each $\eta > 0$ the function $Q_T(\theta)$ stays uniformly in $\Theta_{c_1c_2c_3}$ above $Q_\eta(\theta)$ and hence also above $\sup_{\eta>0} Q_\eta(\theta)$.

Restriction to a compact set $\Theta_{c_1c_2c_3}$

A central step in HD on p. 121 is to show that the PML estimator is inside $\Theta_{c_1c_2c_3}$ a. s. for T large enough. That is, the eigenvalues of $\hat{\Sigma}$ are bounded from below and above and the coefficients of the polynomial $R(z) = \operatorname{adj}(b(z))a(z) = \sum_{j=0}^r R_j z^j$ (where $\tilde{k}(z) = a^{-1}(z)b(z)$) can be bounded such that

$$\sum_{j=0}^r \|R_j\|^2_{Fr} \leq c_3.$$

To show this, first note that

$$\limsup L_T(\hat{\theta}) \leq \log\det\Sigma_0 + s \quad \text{a. s.},$$

as $L_T(\theta_0) \to \log\det\Sigma_0 + s$ a. s.. This can be shown, as in this case the log det term converges to $\log\det\Sigma_0$ and for the Q_T term we have shown $\tilde{Q}_T(\tilde{\theta}_0) \to Q(\tilde{\theta}_0) = s$. The fact that $V_\pi(\tilde{\theta}_0) > 0$ then shows $L_T(\theta_0) \to \log\det\Sigma_0 + s$.
This implies $\log|\hat{\Sigma}| \leq \log|\Sigma_0| + s$ a. s. for T large enough. Next, we use the arguments on the bottom of p. 121 and (4.2.25) of HD to infer

$$\mathcal{A}\Gamma_T(\theta)\mathcal{A}' \leq \gamma^2(I_{T-r} \otimes \Sigma), \theta \in \hat{\Theta}$$

for some constant $0 < \gamma < \infty$, where $\mathcal{A}$ denotes the matrix A defined in line 2 of p. 122 of HD. As in HD it follows that

$$\begin{aligned}\tilde{Q}_T(\theta) &= u_T(\theta)'\Gamma_T(\theta)^{-1}u_T(\theta)/T \\ &\geq \operatorname{tr}\left[\Sigma^{-1}[R_r, R_{r-1}, ..., R_0]\left[T^{-1}\sum_{t=r+1}^{T} U_{t,r}(\theta)U_{t,r}(\theta)'\right][R_r, R_{r-1}, ..., R_0]'\right]\gamma^{-2}.\end{aligned}$$

Using the arguments above, it follows that the smallest eigenvalue of

$$\left[T^{-1}\sum_{t=r+1}^{T} U_{t,r}(\theta)U_{t,r}(\theta)'\right] \geq \left[T^{-1}\sum_{t=r+1}^{T} V_{t,r}(\theta)^\pi(V_{t,r}(\theta)^\pi)'\right]$$

can be bounded from below a. s. for T large enough by a constant c, because it is related to $\langle V_t(\theta)^\pi, V_t(\theta)^\pi\rangle$, whose main component is

$$v_{t,\bullet}(\theta) = \varepsilon_t + (C_0B_0 - P_C)\varepsilon_{t-1} + (C_{0,\bullet}A_{0,\bullet} - P_CC_{0,\bullet})x_{t-1,\bullet},$$

where $x_{t-r-1,1}$ and $\mathbf{d}_t$ are regressed out. Noting that $R_0 = I_s$ it follows that

$$u_T(\theta)'\Gamma_T(\theta)^{-1}u_T(\theta)/T \geq \operatorname{tr}\left[\Sigma^{-1}[R_r, R_{r-1}, ..., I_s][R_r, R_{r-1}, ..., I_s]'\right]c\gamma^{-2} \geq \operatorname{tr}(\Sigma^{-1})c\gamma^{-2}.$$

Consequently, (letting the eigenvalues of $\hat{\Sigma}$ be denoted as $\lambda_j(\hat{\Sigma})$)

$$\operatorname{tr}(\hat{\Sigma}^{-1}) = \sum_{j=1}^{s}\frac{1}{\lambda_j(\hat{\Sigma})} \leq (\log\det\Sigma_0 + s)\gamma^2/c < \infty$$

is bounded a. s. for large enough T. This implies that the smallest eigenvalue of $\hat{\Sigma}$ is bounded from below. Consequently, also the largest eigenvalue of $\hat{\Sigma}$ is bounded, since $\log|\hat{\Sigma}| \leq \log|\Sigma_0| + s$ a. s. for T large enough.
Furthermore, also the third restriction of $\Theta_{c_1c_2c_3}$ is valid a. s. for large enough T as the lower bound on the eigenvalues of $\hat{\Sigma}$ implies

$$u_T(\hat{\theta})'\Gamma_T(\hat{\theta})^{-1}u_T(\hat{\theta})/T \geq \operatorname{tr}\left[\hat{\Sigma}^{-1}[\sum_{j=0}^{r}\hat{R}_j\hat{R}_j']\right]c\gamma^{-2} \geq \operatorname{tr}[\sum_{j=0}^{r}\hat{R}_j\hat{R}_j']c\gamma^{-2}/c_2.$$

Therefore, it follows that for large enough T a. s. $\widehat{\tilde{\theta}_T} \in \Theta_{c_1c_2c_3}$.
We obtain that for large enough T a. s. $\widehat{\tilde{\theta}_T} \in \Theta_{c_1c_2c_3} \cap \hat{\Theta}$, as the criterion function is infinite for transfer functions $\tilde{k}$ with unit roots in this case, since in this case $\lambda_{|max|}(P_\bullet(\theta_j)) \to \infty$ for $\theta_j \to \theta \in \overline{\Theta} - \hat{\Theta}$.

Putting the pieces together

Using the uniform convergence of $\tilde{Q}_{T,\eta}(\theta)$ to $\tilde{Q}_\eta(\theta)$ on $\Theta_{c_1c_2c_3}$, we have (using Lemma 4.2.1. of HD for the last equation)

$$\begin{aligned}\liminf_{T\to\infty} L_T(\hat{\theta}_T) &\geq \liminf_{T\to\infty}(\log|\hat{\Sigma}_T| + \tilde{Q}_T(\hat{\theta}_T)) \geq \sup_{\eta>0}\liminf_{T\to\infty}(\log|\hat{\Sigma}_T| + \tilde{Q}_{T,\eta}(\hat{\theta}_T)) \quad \text{a. s.} \\ &\geq \inf_{\theta\in\Theta_{c_1c_2c_3}}\left(\log|\Sigma| + \sup_{\eta>0}Q_\eta(\theta)\right) = \log|\Sigma_0| + s.\end{aligned}$$

Then, as in HD, p. 125, for every sequence $\hat{\theta}_T \to \theta$ (choosing a subsequence if necessary) it holds that

$$\liminf_{T\to\infty} L_T(\theta_0) \geq \liminf_{T\to\infty} L_T(\hat{\theta}_T) \geq \log|\Sigma_0| + s.$$

Then, $L_T(\hat{\theta}_T) \to L(\theta) = L(\theta_0) = \log|\Sigma_0| + s$ follows, where $\theta = (\tilde{k}, P_C, \Sigma, \tilde{D})$. This shows that $\Sigma = \Sigma_0$ and $\tilde{k} = \tilde{k}_0$, since $L(\theta)$ depends on θ only via Σ and $\tilde{k}$ but not on P_C or $\tilde{D}$.
It follows that $L_T(\hat{\theta}_T) \to L(\theta_0)$ and hence for large enough T it holds that $L_T(\hat{\theta}_T) \leq c$ a. s. for a constant $c < \log|\Sigma_0| + s + \epsilon$ for every $\epsilon > 0$.
Inspecting the proof then shows that

$$\tilde{Q}_T^z(\hat{\theta}_T) := \operatorname{tr}\left(\Sigma_P^{-1}[C_M, C_{M-1}, ..., C_0]\left\langle U_{t,M}(\hat{\theta}_T)^z, U_{t,M}(\hat{\theta}_T)^z\right\rangle [C_M, C_{M-1}, ..., C_0]'\right) < c \quad \text{a. s.}$$

This upper bound implies the following convergence results:

Lemma B.6 *Under the assumptions of Theorem 2.1 let $D_T = diag(T^{-1/2}I_c, D_{T,d})$, where $D_{T,d} = diag(T^{\gamma_{c+1}}, ..., T^{\gamma_K})$, where the reals γ_j are such that (where $z_{t,0} := [x'_{t,1}, s'_t]' \in \mathbb{R}^K$)*

$$D_T \langle z_{t,M}, z_{t,M}\rangle D_T \xrightarrow{d} Z > 0$$

for some non-degenerate random variable Z.
Then, the upper bound $\tilde{Q}_T^z(\hat{\theta}) \leq c$ for some $c \in \mathbb{R}$ implies $(I_s - P_C)\hat{P}_C = o_P(T^{1-\epsilon})$.
Moreover, we obtain $C'_{1,\circ,\perp}(\hat{d}_1 - d_{1,\circ})T^{+1/2-\epsilon} \xrightarrow{p} 0, C'_{1,\circ}(\hat{d}_{m+1} - d_{m+1,\circ})T^{+1/2-\epsilon} \xrightarrow{p} 0, C'_{1,\circ,\perp}(\hat{d}_{m+1} - d_{m+1,\circ})T^{+3/2-\epsilon} \xrightarrow{p} 0$ and $(\hat{d}_j - d_{j,\circ})T^{+1/2-\epsilon} \xrightarrow{p} 0, 1 < j < m+1$.

Note that here only the unit root $z_1 = 1$ is present. Therefore, the deterministic terms corresponding to all other unit roots are estimated consistently, while the constant d_1 is only estimated consistently in the cointegrating space.
PROOF: Recall that $U_{t,M}(\theta)^z$ denotes the vector $U_{t,M}(\theta)$ projected onto $z_{t,M} := [x'_{t-M-1,1}, s'_t]' \in \mathbb{R}^K$ for suitable K. Then, the probability that

$$\lambda_{|min|}((\log T)D_T \langle z_{t,M}, z_{t,M}\rangle D_T) > 0$$

tends to 1. If no deterministic terms are contained, then Bauer (2009, Lemma 4) shows that this also holds a. s.. Otherwise, the lower bound only can be shown to hold in probability. Thus, in

$$\langle U_{t,M}(\theta)^z, U_{t,M}(\theta)^z\rangle = \langle U_{t,M}(\theta), z_{t,M}\rangle D_T(D_T \langle z_{t,M}, z_{t,M}\rangle D_T)^{-1} D_T \langle z_{t,M}, U_{t,M}(\theta)\rangle$$

the essential term is

$$\langle U_{t,M}(\theta), z_{t,M}\rangle = \langle V_{t,M}(\theta), z_{t,M}\rangle + \Psi_M(\theta)\langle z_{t,M}, z_{t,M}\rangle,$$

such that

$$\begin{aligned}&\langle U_{t,M}(\theta), z_{t,M}\rangle D_T(D_T \langle z_{t,M}, z_{t,M}\rangle D_T)^{-1} = \\ &\quad \langle V_{t,M}(\theta), z_{t,M}\rangle D_T(D_T \langle z_{t,M}, z_{t,M}\rangle D_T)^{-1} + \Psi_M(\theta)D_T^{-1}.\end{aligned}$$

Letting $\tilde{\Psi}_M(\theta) := \Psi_M(\theta)D_T^{-1}$ and $\tilde{V}_{t,M}(\theta) := \langle V_{t,M}(\theta), z_{t,M}\rangle D_T(D_T \langle z_{t,M}, z_{t,M}\rangle D_T)^{-1}$, we obtain

$$\langle U_{t,M}(\theta)^z, U_{t,M}(\theta)^z\rangle = (\tilde{\Psi}_M(\theta) + \tilde{V}_{t,M}(\theta))(D_T \langle z_{t,M}, z_{t,M}\rangle D_T)(\tilde{\Psi}_M(\theta) + \tilde{V}_{t,M}(\theta))'.$$

Now, the lower bound on the eigenvalues of $(\log T)(D_T \langle z_{t,M}, z_{t,M}\rangle D_T)$ jointly with the upper bound stated in the lemma implies that

$$\|[C_M, C_{M-1}, ..., C_0]\left(\tilde{\Psi}_M(\hat{\theta}) + \tilde{V}_{t,M}(\hat{\theta})\right)\|/(\log T)^{1/2} = O_P(1),$$

where the bound is almost sure, if no linear trend term is present. Thus, investigate this more closely:

$$[C_M, C_{M-1}, ..., C_0][D(\hat{\theta}; M), P_M] = [\sum_{j=0}^{M} C_j(\tilde{D}_\circ - \tilde{D})\mathcal{S}^{-j}, (\sum_{j=0}^{M} C_j)\hat{P}_\perp C_{0,1}].$$

Here, C_j denotes the impulse response to the inverse transfer function $\tilde{k}_0^{-1}(z)$, such that the term $\sum_{j=0}^{M} C_j$ converges to $\tilde{k}_0^{-1}(1)$ for $M \to \infty$. Thus, M can be chosen large enough to ensure the non-singularity of this term without loss of generality. With respect to $\sum_{j=0}^{M} C_j(\tilde{D}_\circ - \tilde{D})\mathcal{S}^{-j}$, examine $\mathcal{S}$ in more depth for $\omega_1 = 0, 0 < \omega_2 < \pi, l = m = 2$:

$$s_t = \mathcal{S}s_{t-1} \Rightarrow \mathcal{S} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\omega_2) & \sin(\omega_2) & 0 \\ 0 & -\sin(\omega_2) & \cos(\omega_2) & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \Rightarrow \mathcal{S}^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\omega_2) & -\sin(\omega_2) & 0 \\ 0 & \sin(\omega_2) & \cos(\omega_2) & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}.$$

For $l > 2$ or $m > l$ additional sine and cosine blocks are added. Changing to complex quantities we obtain

$$Ds_t = D_{\mathbb{C}} s_{t,\mathbb{C}}, s_{t,\mathbb{C}} = \mathcal{S}_{\mathbb{C}} s_{t-1,\mathbb{C}}, \quad \mathcal{S}_{\mathbb{C}}^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & z_2 & 0 & 0 \\ 0 & 0 & \overline{z}_2 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}.$$

Thus, we obtain for the column of $\sum_{j=0}^{M} C_j(\tilde{D}_\circ - \tilde{D})\mathcal{S}^{-j}$ corresponding to the complex root z_k in the complex representation $\sum_{j=0}^{M} C_j(\tilde{D}_{\mathbb{C},k,\circ} - \tilde{D}_{k,\mathbb{C}})z_k^{-j} \to k_\circ^{-1}(\overline{z}_k)(\tilde{D}_{\mathbb{C},k,\circ} - \tilde{D}_{k,\mathbb{C}})$.
Since $\tilde{k}_0^{-1}(1)$ is non-singular and it can be shown that

$$[C_M, C_{M-1}, ..., C_0] \left\langle V_{t,M}(\hat{\theta}), z_{t,M} \right\rangle D_T (\log T)^{1/2} ((\log T) D_T \langle z_t, z_t \rangle D_T)^{-1} = o_P(1)$$

implies that

$$\| [C_M, C_{M-1}, ..., C_0] \left(\tilde{\Psi}_M(\theta) \right) \| / (\log T)^{1/2} = O_P(1),$$

which in turn implies that $(\tilde{D}_1 - \tilde{D}_{1,\circ})(\log T)^{-1} = o_P(1)$, $(\tilde{D}_{m+1} - D_{m+1,\circ})T(\log T)^{-1} = o_P(1)$, $\hat{P}_\perp C_{0,1} T^{1/2} (\log T)^{-1} = o_P(1)$, where $\tilde{D}_j$ denotes the j-th column of $\tilde{D}$. If no linear trend is present, the last term is $o(1)$.
Now, note that $U_{t,M}(\theta) = V_{t,M}(\theta) + \Psi_M(\theta) z_{t,M}$, where we have shown that

$$\left\langle \Psi_M(\hat{\theta}) z_{t,M}, \Psi_M(\hat{\theta}) z_{t,M} \right\rangle = \tilde{\Psi}_M(\hat{\theta}) (D_T \langle z_{t,M}, z_{t,M} \rangle D_T) \tilde{\Psi}_M(\hat{\theta})' = O_P(1).$$

Furthermore,

$$\left\langle V_{t,M}(\hat{\theta}), \Psi_M(\hat{\theta}) z_{t,M} \right\rangle = \left\langle V_{t,M}(\hat{\theta}), z_{t,M} \right\rangle D_T (\Psi_M(\hat{\theta}) D_T^{-1})' = \left\langle V_{t,M}(\hat{\theta}), z_{t,M} \right\rangle D_T \tilde{\Psi}_M(\hat{\theta})' = o_P(1),$$

as $\left\langle V_{t,M}(\hat{\theta}), z_{t,M} \right\rangle D_T$ tends to zero, while the remaining factor is bounded. Now, assume that $\left\| \left\langle V_{t,M}(\hat{\theta}), z_{t,M} \right\rangle D_T \right\| = o_P(\|\tilde{\Psi}_M(\hat{\theta})\|)$. Then,

$$\left\langle V_{t,M}(\hat{\theta}), z_{t,M} \right\rangle \Psi_M(\hat{\theta})' = o_P(\tilde{\Psi}_M(\hat{\theta}) (D_T \langle z_{t,M}, z_{t,M} \rangle D_T) \tilde{\Psi}_M(\hat{\theta})')$$

and the squared terms $\left\langle \Psi_M(\hat{\theta}) z_{t,M}, \Psi_M(\hat{\theta}) z_{t,M} \right\rangle$ dominates, such that

$$\begin{aligned} &\left\langle U_{t,M}(\hat{\theta}), U_{t,M}(\hat{\theta}) \right\rangle = \\ &\quad \left\langle V_{t,M}(\hat{\theta}), V_{t,M}(\hat{\theta}) \right\rangle + \left\langle \Psi_M(\hat{\theta}) z_{t,M}, \Psi_M(\hat{\theta}) z_{t,M} \right\rangle + o_P\left(\left\langle \Psi_M(\hat{\theta}) z_{t,M}, \Psi_M(\hat{\theta}) z_{t,M} \right\rangle \right) > \\ &\quad \left\langle V_{t,M}(\hat{\theta}), V_{t,M}(\hat{\theta}) \right\rangle \end{aligned}$$

for large T contradicting the optimality property (note that $\Psi_M(\theta_0) = 0$).
It follows that the order of convergence of the elements of $\tilde{\Psi}_M(\hat{\theta}) = \Psi_M(\hat{\theta})D_T^{-1}$ must be of order equal to or smaller than the corresponding ones of $\left\langle V_{t,M}(\hat{\theta}), z_t \right\rangle D_T$.
Consider the component $x_{t-M-1,1}$ of $z_{t,M}$ first: Then, $\left\langle V_{t,M}(\hat{\theta}), x_{t-M-1,1} \right\rangle = O(\log T)$ such that the columns of $\Psi_M(\hat{\theta})$ corresponding to this regressor is of order $o_P((\log T)^2 T^{-1}) = o_p(T^{-1+\epsilon})$.
If $s_t = [1, t]'$, then $(\tilde{D}_1 - \tilde{D}_{1,\circ})$ corresponding to the constant is of order $O_P(T^{-1/2}(\log T)) = o_P(T^{-1/2+\epsilon})$, for $(\tilde{D}_{m+1} - \tilde{D}_{m+1,\circ})$ corresponding to the linear trend we obtain $O_P(T^{-3/2}(\log T)) = o_P(T^{-3/2+\epsilon})$.
Finally, if no linear trend term is present, then the orders are also almost sure.□

The lemma implies the consistency results of the required order also for the entries in $\hat{P}_C = \hat{C}_1 \hat{C}_1'$ and the matrices $\hat{D}$ corresponding to the deterministic terms.

The prediction error criterion function

For the prediction error criterion function

$$L_{PE,T}(k(z), \Sigma, \theta_D; Y_T) = \log|\Sigma| + u_T(\theta)' \Gamma_{T,PE}(\theta)^{-1} u_T(\theta)/T$$

it follows that

$$\Gamma_{T,PE}(\tilde{\theta}) = \underbrace{\begin{bmatrix} I_s & 0 & ... & 0 \\ K_1 & I_s & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ K_{T-1} & ... & K_1 & I_s \end{bmatrix}}_{\mathcal{T}_T(\theta)} (I_T \otimes \Sigma) \begin{bmatrix} I_s & 0 & ... & 0 \\ K_1 & I_s & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ K_{T-1} & ... & K_1 & I_s \end{bmatrix}'.$$

Consequently, $|\Gamma_{T,PE}(\tilde{\theta})| = |\Sigma|^T$ in this case such that $T^{-1} \log|\Gamma_T(\tilde{\theta})| = \log|\Sigma|$ here. Note that this does not depend on the poles of $\tilde{k}$ and hence in this case the first part of the criterion function does not diverge to infinity at the stability boundary of $\hat{\Theta}$.
Following the proof above, it is noted that the log det term in this case is simpler such that we can directly obtain the result $\log|\Gamma_{T,PE}(\tilde{\theta})|/T \to \log|\Sigma|$ in this case.
For the second part $Q_T(\theta)$ of the criterion function note that

$$\Gamma_{T,PE}(k, \Sigma) \leq \Gamma_T(k, \Sigma) \Rightarrow \Gamma_{T,PE}(k, \Sigma)^{-1} \geq \Gamma_T(k, \Sigma)^{-1}.$$

Therefore, it follows that the second term of the criterion function can be bounded from below by the arguments given above. This immediately implies that the prediction error estimator lies in $\Theta_{c_1 c_2 c_3}$ for large enough T. The result then follows from

$$L_{PE,T}(k_\circ(z), \Sigma_\circ, \theta_{D,\circ}; Y_T) \to \log|\Sigma_\circ| + s,$$

see Lemma B.4 for a proof.

B.2.2 Proof of Theorem 2.2 (Asymptotic Distribution) and Corollary 2.2

Let us repeat also at this point, that in order to derive the asymptotic distribution, the results are based on using a specific parameterization, whereas the consistency proof has been parameterization free. This approach allows us to embed the problem in some $\mathbb{R}^{d_1}$, for an appropriate d_1. We will use complex matrices in the proof, although we are concerned only with real valued processes. The relationships between complex and real valued representation in the canonical form have been discussed in Bauer and Wagner (2012).

For the applicability of linearization techniques we furthermore assume that the parameters are introduced in such a way that the true parameter vector

$$\varphi_\circ = [\ \theta'_{u,\circ} \quad e'_{m+1,\perp,\circ} \quad e'_{m+1,\circ} \quad \theta'_{d,b,\circ} \quad \theta'_{st,\circ}\ ]',$$

where $\theta_{u,\circ}$ collects the parameter vector corresponding to the vectors $\theta_{L,k}$ in the parameterization for $z_k \in \{-1, 1\}$

$$C_k(\theta_k) := R_L(\theta_{L,k})' \begin{bmatrix} I_{c_k} \\ 0_{(s-c_k)\times c_k} \end{bmatrix} R_R(\theta_{R,k}),$$

where $R_L(\theta_{L,k}) \in \mathbb{R}^{s\times s}$ and $R_R(\theta_{C,R,k}) \in \mathbb{R}^{c_k\times c_k}$ are orthonormal matrices, while for $z_k \neq \pm 1$ we use

$$C_k(\theta_k) := Q_L(\theta_{L,k})' \begin{bmatrix} I_{c_k} \\ 0_{(s-c_k)\times c_k} \end{bmatrix} Q_R(\theta_{C,R,k}) D_{c_k}(\theta_{C,D,k}),$$

where $Q_L(\theta_{L,k}) \in \mathbb{C}^{s\times s}$ and $D_{c_k}(\theta_{C,D,k}), Q_R(\theta_{C,R,k}) \in \mathbb{C}^{c_k\times c_k}$ are unitary matrices and $D_{c_k}(\theta_{C,D,k})$ moreover is a diagonal matrix. R_L, R_R, Q_L and Q_R here are products of Givens rotations. For more details on the parameterization see Bauer et al. (2020).
Furthermore, $e_{m+1,\perp,\circ}$ and $e_{m+1,\circ}$ denote the parameters for the linear trend term (if it is included), while $\theta_{d,b,\circ}$ collects all other parameters due to the deterministic terms.
Further,

$$\theta_{st,\circ} = \left[\ \theta'_{C,D,\circ} \quad \theta'_{C,R,\circ} \quad \theta'_{B,f,\circ} \quad \theta'_{B,p,\circ} \quad \theta'_{\bullet,\circ}\ \right]'.$$

Then, in the theorem it is assumed that $\varphi_\circ$ is an interior point of the parameter set. This requires that the multi-index Γ is specified correctly.
Let the corresponding parameter estimator minimizing the scaled negative pseudo likelihood function L_T be denoted as $\hat{\varphi}$. The proof of Lemma B.4 shows that asymptotically the difference between minimizing L_T and the prediction error function $L_{PE,T}$ is negligible for the asymptotic distribution. Thus, we will in the following use the prediction error form which is easier to investigate. Furthermore, we will concentrate out the parameters for Σ such that the criterion function equals

$$\mathcal{L}_{PE,T}(\varphi; Y_T) = \log\det\left\langle \varepsilon_t(\varphi), \varepsilon_t(\varphi)\right\rangle / T,$$

where

$$\varepsilon_t(\varphi) = y_t - D(\varphi)s_t - \sum_{j=1}^{t-1} \underline{K}_j(\varphi)(y_{t-j} - D(\varphi)s_{t-j}).$$

In the following we will omit the subscript 'PE' for notational convenience.
Here, $\underline{K}_j(\varphi)$ denotes the impulse response corresponding to the inverse transfer function $k^{-1}(z)$. Note that $\|\underline{K}_j(\varphi_\circ)\| \leq \mu_K \rho_0^j$ for some $0 < \rho_0 < 1$, due to the strict minimum-phase assumption for the data generating system.
Due to the consistency result it follows that for T large enough, the probability that the estimate $\hat{\varphi}$ is contained in Θ_ϵ (an open neighborhood of $\varphi_\circ$) tends to 1, where the exponential decrease of the impulse response sequence holds uniformly in Θ_ϵ.
Thus, a necessary condition for a minimum is a zero first derivative and we obtain from the mean value theorem

$$\partial\mathcal{L}_T(\hat{\varphi}; Y_T) = 0 = \partial\mathcal{L}_T(\varphi_\circ; Y_T) + \partial^2\mathcal{L}_T(\bar{\varphi}_T; Y_T)[\hat{\varphi} - \varphi_\circ],$$

where $\bar{\varphi}_T$ denotes an intermediate point between $\hat{\varphi}$ and $\varphi_\circ$, not necessarily the same in each row. Let $\tilde{D}_T = \operatorname{diag}(I, T^{-1/2}I, T^{1/2}I)$, where the sizes of the three blocks of $\tilde{D}_T$ are equal to the dimensions of the parameter vectors θ_u, $e_{m+1,\perp}$ and $[\ e'_{m+1} \quad \theta'_{d,b} \quad \theta'_{st}\ ]'$ respectively. Further, let $D_T = \operatorname{diag}(TI, T^{3/2}I, T^{1/2}I)$. Then the proof of the theorem proceeds in three steps:

1. Show that $\tilde{D}_T\partial\mathcal{L}_T(\varphi_\circ; Y_T)$ converges in distribution.

2. Show that $\tilde{D}_T \partial^2 \mathcal{L}_T(\bar{\varphi}_T; Y_T) D_T^{-1}$ converges in distribution to a random matrix Z.

3. Show that $\mathbb{P}\{Z > 0\} = 1$.

Let us start with the first item, i. e., with establishing the asymptotic properties of the score vector. Denote with $\partial_i f(\varphi_\circ)$ the partial derivative of a function f with respect to the i-th component of the parameter vector φ, evaluated at the point $\varphi = \varphi_\circ$. With subscript $i = st$ we denote the subvector of $\partial_i f(\varphi_\circ)$ for all i, such that the component θ_i is contained in θ_{st}. With subscript u we denote differentiation with respect to the entries in θ_u, with subscript $e_\perp$ we denote differentiation with respect to the entries in $e_{m+1,\perp}$, with e differentiation with respect to the entries in e_{m+1} and with db differentiation with respect to the entries in $\theta_{d,b}$. Furthermore we will use the notation $\partial_{i,H} f(\varphi)$ for differentiation with respect to the i-th component of φ which corresponds to the matrix H.
Here and also below the matrices $(\mathcal{A}, \mathcal{B}, \mathcal{C})$ correspond to the canonical representation of a system described by the parameter vector θ. We omit the dependency on the parameter θ for simplicity of notation. Recall that the matrix $\mathcal{C} = [\ C_1 \quad \dots \quad C_l \quad \mathcal{C}_\bullet\ ]$ is partitioned according to the blocks of different unit roots (or pairs of complex roots).

Lemma B.7 *The derivatives of $\varepsilon_t(\varphi)$ with respect to the different parameters in θ_{st} are given by*

- $\partial_{i,\mathcal{A}_\bullet} \varepsilon_t(\varphi) = dk_{i,A_\bullet}(L,\theta) x_{t,\bullet}(\theta)$, *where* $dk_{i,\mathcal{A}_\bullet}(z,\theta) = -\mathcal{C} z (I - z\underline{\mathcal{A}})^{-1} \begin{bmatrix} 0_{c\times(n-c)} \\ \partial_i(\mathcal{A}_\bullet) \end{bmatrix}$,
- $\partial_{i,\mathcal{B}_\bullet} \varepsilon_t(\varphi) = dk_{i,\mathcal{B}_\bullet}(L,\theta) \varepsilon_t(\varphi)$, *where* $dk_{i,\mathcal{B}_\bullet}(z,\theta) = -\mathcal{C} z (I - z\underline{\mathcal{A}})^{-1} \begin{bmatrix} 0_{c\times s} \\ \partial_i(\mathcal{B}_\bullet) \end{bmatrix}$,
- $\partial_{i,\mathcal{B}_u} \varepsilon_t(\varphi) = dk_{i,\mathcal{B}_u}(L,\theta) \varepsilon_t(\varphi)$, *where* $dk_{i,\mathcal{B}_u}(z,\theta) = -\mathcal{C} z (I - z\underline{\mathcal{A}})^{-1} \begin{bmatrix} \partial_i(\mathcal{B}_u) \\ 0_{(n-c)\times s} \end{bmatrix}$,
- $\partial_{i,\mathcal{C}_\bullet} \varepsilon_t(\varphi) = dk_{i,\mathcal{C}_\bullet}(L,\theta) x_{t,\bullet}(\theta)$, *where* $dk_{i,\mathcal{C}_\bullet}(z,\theta) = k^{-1}(z;\theta)(-\partial_i \mathcal{C}_\bullet)$,
- $\partial_{i,\theta_{C,D}} \varepsilon_t(\varphi) = 2\mathcal{R}\left\{ dk_{i,\theta_{C,D}}(L,\theta) \varepsilon_{t-1}(\varphi) \right\}$ *for* $z_k \neq \pm 1$,
  where $dk_{i,\theta_{C,D}}(z,\theta) = \tilde{k}(z;\theta)(-\partial_{i,\theta_{C,D}} C_k) B_k$, *where* $\tilde{k}(z;\theta)$ *is defined through* $k^{-1}(z;\theta) = \tilde{k}(z,\theta)(1 - z\bar{z}_k) + k^{-1}(z_k,\theta) z$,
- $\partial_{i,\theta_{C,R}} \varepsilon_t(\varphi) = dk_{i,\theta_{C,R}}(L,\theta) \varepsilon_{t-1}(\varphi)$ *if* $z_k = \pm 1$, *where* $dk_{i,\theta_{C,R}}(z,\theta) = \tilde{k}(z;\theta)(-\partial_{i,\theta_{C,R}} C_k) B_k$,
  where $k^{-1}(z;\theta) = \tilde{k}(z,\theta)(1 - z\bar{z}_k) + k^{-1}(z_k,\theta) z$,
- $\partial_{i,\theta_{C,R}} \varepsilon_t(\varphi) = 2\mathcal{R}\left\{ dk_{i,\theta_{C,R}}(L,\theta) \varepsilon_{t-1}(\varphi) \right\}$ *if* $z_k \neq \pm 1$,
  where $dk_{i,\theta_{C,R}}(z,\theta) = \tilde{k}(z;\theta)(-\partial_{i,\theta_{C,R}} C_k) B_k$.

All the above processes are asymptotically stationary.
The derivatives of $\varepsilon_t(\varphi)$ with respect to parameters in θ_u are given by

- $\partial_{i,\theta_L} \varepsilon_t(\varphi) = dk_{i,\theta_L}(L,\theta) x_{t,k}(\theta)$ *if* $z_k = \pm 1$, *where* $dk_{i,\theta_L}(z,\theta) = k^{-1}(z;\theta)(-\partial_{i,\theta_L} C_k)$,
- $\partial_{i,\theta_L} \varepsilon_t(\varphi) = 2\mathcal{R}\left\{ dk_{i,\theta_L}(L,\theta) x_{t,k}(\theta) \right\}$ *if* $z_k \neq \pm 1$, *where* $dk_{i,\theta_L}(z,\theta) = k^{-1}(z;\theta)(-\partial_{i,\theta_L} C_k)$.

Finally, the derivatives of $\varepsilon_t(\varphi)$ with respect to parameters in θ_d are given by

- $\partial_{i,D} \varepsilon_t(\varphi) = -k^{-1}(L;\theta) \partial_{i,D} D(\theta_d) s_t(\theta)$.

PROOF: The results follow from taking the derivative of the inverse transfer function. As an example, let us analyze derivation with respect to a parameter corresponding to the matrix C_k. The partial derivatives are:

$$\begin{aligned} \partial_i \varepsilon_t(\varphi) &= -(\partial_i \mathcal{C}_k) x_{t,k}(\theta) - \mathcal{C}(\partial_i x_t(\theta)), \\ \partial_i x_{t+1}(\theta) &= \underline{\mathcal{A}} \partial_i x_t(\theta) - \mathcal{B}(\partial_i \mathcal{C}_k) x_{t,k}(\theta). \end{aligned}$$

These components of the score are filtered version of $x_{t,k}(\theta)$. A possibly non-minimal representation of the filter is given by $dk_i(z,\theta) = -\partial_i \mathcal{C}_k + z\mathcal{C}(I - z\underline{\mathcal{A}})^{-1}\mathcal{B}\partial_i \mathcal{C}_k = k^{-1}(z;\theta)(-\partial_i \mathcal{C}_k)$. Decomposing

$$k^{-1}(z;\theta) = \tilde{k}(z,\theta)(1 - \overline{z}_k z) + k^{-1}(z_k,\theta)z$$

and noting that $k^{-1}(z_k,\theta)C_k = 0$, one sees that the second term cancels for vectors in the column space of C_k.
Consider now the derivative $\partial_i C_k$, which is complex valued if $z_k \neq \pm 1$. If the derivative is taken with respect to $\theta_{C,L}$, it follows that the column space of $-\partial_{i,\theta_{C,L}} C_k$ does not lie in the column space of C_k. If the derivative is taken with respect to $\theta_{C,D}$, then $-\partial_{i,\theta_{C,D}} C_k$ lies in the column space of C_k, such that

$$k^{-1}(z;\theta)(-\partial_{i,\theta_{C,D}} C_k) \quad = \quad \tilde{k}(z,\theta)(1 - \overline{z}_k z)(-\partial_{i,\theta_{C,D}} C_k).$$

Define $x_{t+1}(z_k,\theta) = \overline{z}_k x_t(z_k,\theta) + B_k(\theta)\varepsilon_t$, thus, $x_{t,k}(\theta) = [2\mathcal{R}\{x_t(z_k,\theta)\}', -2\mathcal{I}\{x_t(z_k,\theta)\}']'$ for $z_k \neq \pm 1$. Using

$$k^{-1}(z;\theta)(-\partial_i \mathcal{C}_k)x_{t,k}(\theta) = \mathcal{R}\left\{k^{-1}(z;\theta)(-\partial_i C_k)x_t(z_k,\theta)\right\}$$

and applying $k^{-1}(z;\theta)(-\partial_{i,\theta_{C,D}} C_k)$ (and $k^{-1}(z;\theta)(-\partial_{i,\theta_{C,D}} \overline{C}_k)$) to $x_t(z_k,\theta)$ (and $x_t(\overline{z}_k,\theta)$), implies the result for derivatives with respect to $\theta_{C,D}$. If the derivative is taken with respect to $\theta_{C,R}$, then $-\partial_{i,\theta_{C,R}} C_k$ also lies in the column space of C_k and the same argument holds.
Derivatives with respect to the other parameters are derived analogously. □
In the following lemma the asymptotic behavior of the score is summarized. In this lemma and the rest of the document the dependence of the prediction error criterion function on Y_T is omitted for notational simplicity.

Lemma B.8 *Let the assumptions of Theorem 2.2 hold. Then the following statements hold true:*

- $$\sqrt{T}\partial_{st}\mathcal{L}_T(\varphi_\circ) \xrightarrow{d} \mathcal{N}(0, V_{st}),$$

  where V_{st} denotes the asymptotic variance matrix.

- $$(\partial_u \mathcal{L}_T(\varphi_\circ))_i \xrightarrow{d} -2\delta_k \mathcal{R}\left\{tr\left[(\Sigma_\circ)^{-1}\Pi_{k,\circ}(\partial_{u,i} C_{k,\circ})X(z_k)\right]\right\} =: (\mathrm{v}_u)_i,$$

  with $X(z_k)$ as defined in Lemma B.2 and $\Pi_k = k^{-1}(z_k)$.

- *For a column $C_{k,\circ,\perp}d^k$ of D corresponding to $z_k = \pm 1$, one obtains*

  $$\sqrt{T}\partial_{db,k}\mathcal{L}_T(\varphi_\circ) \xrightarrow{d} -2\alpha'_{k,\circ}\Sigma_\circ^{-1}W_k(1) =: \mathrm{v}_{db,k},$$

  where we have used the decomposition $\Pi_{k,\circ} = \alpha_{k,\circ}\beta'_{k,\circ}$. If $z_k \neq \pm 1$, the vector of derivatives with respect to the real part of the k-th column $C_{k,\circ,\perp}d^k, k \leq l$, of D or the real part of $d^k, k > l$, respectively, and with respect to the imaginary part of the same column converges to

  $$\sqrt{T}\partial_{db,k}\mathcal{L}_T(\varphi_\circ) = \begin{bmatrix} -4\mathcal{R}\left\{\alpha'_{k,\circ}\Sigma_\circ^{-1}W_k(1)\right\} \\ -4\mathcal{I}\left\{\alpha'_{k,\circ}\Sigma_\circ^{-1}W_k(1)\right\} \end{bmatrix} =: \mathrm{v}_{db,k},$$

  where $W_k(1) = W_k^r(1) + iW_k^i(1)$ is as defined in Lemma B.2. Hence the covariance-matrix V_D is block diagonal, where the diagonal blocks are given by the covariance matrices of the two parts of the vector given above, taking into account the uncorrelatedness of $W_k^r(1)$ and $W_k^i(1)$. If $k > l$, we here use $\alpha_{k,\circ} = I_s$.

- *For the deterministic linear trends we find*

$$\frac{1}{\sqrt{T}}\partial_{e_\perp}\mathcal{L}_T(\varphi_\circ) \xrightarrow{d} -2\alpha'_{1,\circ}\Sigma_\circ^{-1}U(1) =: \mathrm{v}_{e\perp},$$
$$\sqrt{T}\partial_e\mathcal{L}_T(\varphi_\circ) \xrightarrow{d} -2(\Gamma_{1,\circ}C_{1,\circ})'\Sigma_\circ^{-1}W_1(1) =: \mathrm{v}_e,$$

  where $\Gamma_{1,\circ}$ *is the derivative of* $k_\circ(z)^{-1}$ *with respect to* z *evaluated at* 1.

- *All convergence results hold jointly.*

PROOF: In order to establish the asymptotic properties of the score, the partial derivatives of $\mathcal{L}_T(\varphi)$ are required. These can be derived from the system equations:

$$\begin{aligned}\partial_i\mathcal{L}_T(\varphi_\circ) = \partial_i\left(\log\det\langle\varepsilon_t(\varphi_\circ),\varepsilon_t(\varphi_\circ)\rangle\right) &= \mathrm{tr}[\langle\varepsilon_t(\varphi_\circ),\varepsilon_t(\varphi_\circ)\rangle^{-1}2\langle\partial_i\varepsilon_t(\varphi_\circ),\varepsilon_t(\varphi_\circ)\rangle] \\ &= \mathrm{tr}[\Sigma_\circ^{-1}2\langle\partial_i\varepsilon_t(\varphi_\circ),\varepsilon_t(\varphi_\circ)\rangle] + o_P(1).\end{aligned}$$

Let us start with the coordinates of $\theta_{st} = \left[\begin{array}{ccccc}\theta'_{C,D} & \theta'_{C,R} & \theta'_{B,f} & \theta'_{B,p} & \theta'_\bullet\end{array}\right]'$. For every component of θ_{st} Lemma B.7 above establishes asymptotic stationarity. Asymptotic normality for $\sqrt{T}\langle\partial_i\varepsilon_t(\varphi_\circ),\varepsilon_t(\varphi_\circ)\rangle$ then follows from well established theory for stationary processes, see, e. g., Hannan and Deistler (1988, Lemma 4.3.4 ff). It is straightforward to show that the result holds jointly in all coordinates of θ_{st}.
The representation in Lemma B.7 allows for the application of Lemma B.2 and Lemma B.3 to obtain

$$\begin{aligned}\partial_{u,i}\mathcal{L}_T(\varphi_\circ) &= \mathcal{R}\left\{\mathrm{tr}[\Sigma_\circ^{-1}2\langle k^{-1}(L;\theta_\circ)(-\partial_{u,i}C_{k,\circ})x_{t,k}(\theta_\circ),\varepsilon_t\rangle]\right\} + o_P(1) \\ &\xrightarrow{d} 2\mathcal{R}\left\{\mathrm{tr}\left[\Sigma_\circ^{-1}\Pi_{k,\circ}(-\partial_i C_k)X(z_k)\right]\right\}, \qquad \text{where } \partial_i C_k := \partial_{u,i}C_{k,\circ}.\end{aligned}$$

The next step is to derive the asymptotic distribution of the score components corresponding to $\theta_{d,b}$. The matrix $D_\mathbb{C}$ is parameterized with real parameters using $D = \mathcal{R}(D) + i\mathcal{I}(D)$, where both $\mathcal{R}(D)$ and $\mathcal{I}(D)$ are unconstrained, except for the restriction, that the columns of $\mathcal{I}(D)$ corresponding to $z_k = \pm 1$ are zero. Consider a specific element of this part of the score vector, corresponding to component i, say, of $\theta_{d,b}$, which corresponds to entry (a,k) in $\mathcal{R}(D)$. Because of the restriction to real valued output processes, only the real part of the derivative has to be investigated:

$$\partial_{db,k}\mathcal{L}_T(\varphi_\circ) = -\mathcal{R}\left\{\frac{2}{T}\sum_{t=1}^{T}(k^{-1}(L;\theta_\circ)C_{k,\circ,\perp}s_{t,k})'\Sigma_\circ^{-1}\varepsilon_t\right\} + o_P(T^{-1/2}),$$

where $\{s_{t,k}\}_{t\in\mathbb{N}}$ denotes a deterministic process satisfying $s_{t,k} = \overline{z}_k^t$ for all $t \in \mathbb{N}$. Note that

$$k^{-1}(L;\theta_\circ)C_{k,\circ,\perp}s_{t,k} = \Pi_{k,\circ}C_{k,\circ,\perp}s_{t,k} - \mathcal{C}_\circ(z_k\underline{\mathcal{A}}_\circ)^t(I - z_k\underline{\mathcal{A}}_\circ)^{-1}\mathcal{B}_\circ C_{k,\circ,\perp},$$

where the second component exhibits exponential decay. Therefore, it follows that

$$\begin{aligned}\sqrt{T}\partial_{db,k}\mathcal{L}_T(\varphi_\circ) &= -\mathcal{R}\left\{(\Pi_{k,\circ}C_{k,\perp})'\Sigma_\circ^{-1}\frac{2}{\sqrt{T}}\sum_{t=1}^{T}z_k^{t-1}\varepsilon_t\right\} + o_P(T^{-1/2}) \\ &\xrightarrow{d} -2\delta_k\ \mathcal{R}\left\{(\Pi_{k,\circ}C_{k,\circ,\perp})'\Sigma_\circ^{-1}W_k(1)\right\} \\ &= -2\delta_k\ \mathcal{R}\left\{\alpha'_{k,\circ}\Sigma_\circ^{-1}W_k(1)\right\},\end{aligned}$$

where we have used $\Pi_{k,\circ} = \alpha_{k,\circ}\beta'_{k,\circ}$, $\beta_{k,\circ} = C_{k,\circ,\perp}$ and $C'_{k,\circ,\perp}C_{k,\circ,\perp} = I_{r_k}$ in the last step. If $k > l$, such that the deterministic term does not correspond to a unit root, we choose $\alpha_{k,\circ} = \beta_{k,\circ} = I_s$. Note finally that if the derivative is with respect to the (a,k)-th entry in $\mathcal{I}(D)$, in the above equation $\mathcal{R}$ has to be replaced with $\mathcal{I}$.

Now, only the asymptotic distribution of the score components corresponding to $e_{m+1,\perp}$ and e_{m+1} are left to be derived. In both cases we have for entry a of d_{m+1}

$$\partial_{d_{m+1}}\mathcal{L}_T(\varphi_\circ) = -\frac{2}{T}\sum_{t=1}^{T}(k^{-1}(L;\theta_\circ)t)'\Sigma_\circ^{-1}\varepsilon_t + o_P(T^{-1/2}).$$

Using

$$k^{-1}(L;\theta_\circ)t = \sum_{j=0}^{t-1}\underline{K}_j(t-j) = (\sum_{j=0}^{t-1}\underline{K}_j)t - \sum_{j=0}^{t-1}\underline{K}_j j = \Pi_{1,\circ}t + \Gamma_{1,\circ} + o_P(\rho^t),$$

we get for the entries in the direction of $C_{1,\circ,\perp}$

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\partial_{e_\perp}\mathcal{L}_T(\varphi_\circ)\\
&= -2(\Pi_{1,\circ}C_{1,\circ,\perp})'\Sigma_\circ^{-1}T^{-3/2}\sum_{t=1}^{T}t\varepsilon_t - 2(\Gamma_{1,\circ}C_{1,\circ,\perp})'\Sigma_\circ^{-1}T^{-3/2}\sum_{t=1}^{T}\varepsilon_t + o_P(T^{-3/2})\\
&\xrightarrow{d} -2(\Pi_{1,\circ}C_{1,\circ,\perp})'\Sigma_\circ^{-1}U(1)\\
&= -2\alpha_{1,\circ}'\Sigma_\circ^{-1}U(1),
\end{aligned}$$

where we have used Lemma B.3 (viii) for the convergence of the first summand. In the directions of $C_{1,\circ}$ the first term vanishes because $\Pi_{1,\circ}C_{1,\circ} = 0$. Consequently, we get

$$\begin{aligned}
\sqrt{T}\partial_e\mathcal{L}_T(\varphi_\circ) &= -2(\Gamma_{1,\circ}C_{1,\circ})'\Sigma_\circ^{-1}T^{-1/2}\sum_{t=1}^{T}\varepsilon_t + o_P(T^{-1/2})\\
&\xrightarrow{d} -2(\Gamma_{1,\circ}C_{1,\circ})'\Sigma_\circ^{-1}W(1)
\end{aligned}$$

by the same arguments as in the analysis of the entries of $\theta_{d,b}$. This concludes the proof of the lemma. □

After having established the (asymptotic) properties of the score vector, the next step is the analysis of the asymptotic behavior of the Hessian. As in Lemma B.8 in the discussion we have to distinguish with respect to which parameter components θ_u, θ_{st}, $e_{m+1,\perp}$, e_{m+1} or $\theta_{d,b}$ differentiation takes place. In addition to the previous lemma, we also have to consider the cross terms, where differentiation takes place, e. g., once with respect to an entry in θ_u and once with respect to an entry in θ_{st}.

Lemma B.9 *Under the conditions of Theorem 2.2 one obtains* $\tilde{D}_T\partial^2\mathcal{L}_T(\bar{\varphi}_T)D_T^{-1} \xrightarrow{d} Z$ *for each sequence* $\bar{\varphi}_T \to \varphi_\circ$.
In case that no deterministic terms are included in the true data generating process and the model (i. e., $\hat{D} = D_\circ = 0$*),* $Z = diag(Z_\star, Z_{st})$ *is block diagonal. It holds that* $Z_{st} > 0$ *is a constant matrix and* $Z_\star$ *a random matrix, for which* $\mathbb{P}\{Z_\star > 0\} = 1$ *holds.*
If the deterministic terms are included in the model, the following asymptotic distribution is obtained: Here, again $\varphi = [\ \theta_u' \quad e_{m+1,\perp}' \quad e_{m+1}' \quad \theta_{d,b}' \quad \theta_{st}'\ ]'$*. Then*

$$\tilde{D}_T\partial^2 L_T(\varphi_T)D_T^{-1} \xrightarrow{d} \begin{bmatrix} Z_\star & 0 \\ 0 & Z_{st} \end{bmatrix} = \begin{bmatrix} Z_u & Y_{e_\perp}' & Y_e' & Y_{db}' & 0 \\ Y_{e_\perp} & Z_{e_\perp} & Y_{e_\perp,e}' & Y_{e_\perp,db}' & 0 \\ Y_e & Y_{e_\perp,e} & Z_e & Y_{e,db}' & 0 \\ Y_{db} & Y_{e_\perp,db} & Y_{e,db} & Z_{db} & 0 \\ 0 & 0 & 0 & 0 & Z_{st} \end{bmatrix}$$

Z_u, Z_{db} and Y_{db} are block-diagonal, with the diagonal blocks corresponding to different unit roots. For typical indices i, j (not the same for all the expressions below) corresponding to the same unit

root $z_k = \pm 1$ the respective entries are of the form:

$$\begin{aligned}
[Z_{u,k}]_{i,j} &= 2\mathcal{R}\left\{tr\left[\partial_j C_k' \Pi_{k,\circ}' \Sigma_\circ^{-1} \Pi_{k,\circ} \partial_i C_k Z(z_k)\right]\right\}, \\
[Y_{db,k}]_{i,j} &= -2\mathcal{R}\left\{e_i' \alpha_{k,\circ}' \Sigma_\circ^{-1} \Pi_{k,\circ} \partial_i C_k Y(z_k)\right\}, \\
Z_{db,k} &= 2\alpha_{k,\circ}' \Sigma_\circ^{-1} \alpha_{k,\circ}.
\end{aligned}$$

For the entries corresponding to $z_k \neq \pm 1$ the respective entries are of the form:

$$\begin{aligned}
[Z_{u,k}]_{i,j} &= 4\mathcal{R}\left\{tr\left[\partial_j \mathcal{C}_k' \Pi_{k,\circ}' \Sigma_\circ^{-1} \Pi_{k,\circ} \partial_i C_k Z(z_k)\right]\right\}, \\
Z_{db,k} &= 4\begin{bmatrix} \alpha_{k,\circ}' \Sigma_\circ^{-1} \alpha_{k,\circ} & 0 \\ 0 & \alpha_{k,\circ}' \Sigma_\circ^{-1} \alpha_{k,\circ} \end{bmatrix}, \\
[Y_{db,k}]_{i,j} &= -4\begin{bmatrix} \mathcal{R}\left\{e_i' \alpha_{k,\circ}' \Sigma_\circ^{-1} \Pi_{k,\circ} \partial_j C_k Y(z_k)\right\} \\ \mathcal{I}\left\{e_i' \alpha_{k,\circ}' \Sigma_\circ^{-1} \Pi_{k,\circ} \partial_j C_k Y(z_k)\right\} \end{bmatrix}.
\end{aligned}$$

Only the block corresponding to the unit root with $z_k = 1$ of $Y_{e_\perp}$, Y_e, $Y_{e_\perp,db}$ and $Y_{e,db}$ is different from zero. For these blocks we have

$$\begin{aligned}
Z_{e_\perp} &= \frac{2}{3}\alpha_{1,\circ}' \Sigma_\circ^{-1} \alpha_{1,\circ}, \\
Z_e &= 2(\Gamma_{1,\circ} C_{1,\circ})' \Sigma_\circ^{-1} \Gamma_{1,\circ} C_{1,\circ}, \\
Y_{e_\perp,e} &= \alpha_{1,\circ}' \Sigma_\circ^{-1} \Gamma_{1,\circ} C_{1,\circ}, \\
Y_{e_\perp,db} &= \alpha_{1,\circ}' \Sigma_\circ^{-1} \alpha_{1,\circ}, \\
Y_{e,db} &= 2(\Gamma_{1,\circ} C_{1,\circ})' \Sigma_\circ^{-1} \alpha_{1,\circ}, \\
[Y_{e_\perp}]_{i,j} &= -2e_i' \alpha_{1,\circ}' \Sigma_\circ^{-1} \Pi_{1,\circ} \partial_j C_1 V(1), \\
[Y_e]_{i,j} &= -2e_i' \alpha_{1,\circ}' \Sigma_\circ^{-1} \Pi_{1,\circ} \partial_j C_1 Y(1).
\end{aligned}$$

It follows that $Z_u - Y_D' Z_D^{-1} Y_D$ with $Y_D = [\ Y_{e_\perp}' \quad Y_e' \quad Y_{db}'\ ]'$ and

$$Z_D := \begin{bmatrix} Z_{e_\perp} & Y_{e_\perp,e}' & Y_{e_\perp,db}' \\ Y_{e_\perp,e} & Z_e & Y_{e,db}' \\ Y_{e_\perp,db} & Y_{e,db} & Z_{db} \end{bmatrix}$$

has the same structure as Z_u, where in the expression $Z(z_k)$ has to be replaced by $Z(z_k) - Y(z_k)Y(z_k)'$ for all z_k if there is no linear trend term in the model and for $z_k \neq 1$ else. If a linear trend is present in the model, $Z(1)$ has to be replaced by $Z(1) - 12V(1)V(1)' + 6Y(1)V(1)' + 6V(1)Y(1)' - 4Y(1)Y(1)'$. Further, $Z_{st} > 0$ and $\mathbb{P}\{Z_u > 0\} \to 1$ respectively $\mathbb{P}\{Z_u - Y_D' Z_D^{-1} Y_D > 0\} \to 1$.

PROOF: In the proof first convergence of the various parts is shown and in a final step the non-singularity of Z_u is established. First note that:

$$\begin{aligned}
&\partial_{i,j}^2 \mathcal{L}_T(\bar{\varphi}_T) = \partial_i \left(\mathrm{tr}[\langle \varepsilon_t(\bar{\varphi}_T), \varepsilon_t(\bar{\varphi}_T)\rangle^{-1} 2\langle \partial_j \varepsilon_t(\bar{\varphi}_T), \varepsilon_t(\bar{\varphi}_T)\rangle]\right) = \\
&\quad - \mathrm{tr}\Big[\langle \varepsilon_t(\bar{\varphi}_T), \varepsilon_t(\bar{\varphi}_T)\rangle^{-1} \\
&\qquad (\langle \partial_i \varepsilon_t(\bar{\varphi}_T), \varepsilon_t(\bar{\varphi}_T)\rangle + \langle \varepsilon_t(\bar{\varphi}_T), \partial_i \varepsilon_t(\bar{\varphi}_T)\rangle) \langle \varepsilon_t(\bar{\varphi}_T), \varepsilon_t(\bar{\varphi}_T)\rangle^{-1} 2 \langle \partial_j \varepsilon_t(\bar{\varphi}_T), \varepsilon_t(\bar{\varphi}_T)\rangle \Big] \\
&\qquad + \mathrm{tr}\left[\langle \varepsilon_t(\bar{\varphi}_T), \varepsilon_t(\bar{\varphi}_T)\rangle^{-1} 2 \left\langle (\partial_{i,j}^2 \varepsilon_t(\bar{\varphi}_T)), \varepsilon_t(\bar{\varphi}_T)\right\rangle \right] \\
&\qquad + \mathrm{tr}\left[\langle \varepsilon_t(\bar{\varphi}_T), \varepsilon_t(\bar{\varphi}_T)\rangle^{-1} 2 \langle (\partial_j \varepsilon_t(\bar{\varphi}_T)), (\partial_i \varepsilon_t(\bar{\varphi}_T))\rangle \right] = \\
&\quad - \mathrm{tr}\left[\Sigma_\circ^{-1} (\langle \partial_i \varepsilon_t(\bar{\varphi}_T), \varepsilon_t(\bar{\varphi}_T)\rangle + \langle \varepsilon_t(\bar{\varphi}_T), \partial_i \varepsilon_t(\bar{\varphi}_T)\rangle) \Sigma_\circ^{-1} 2 \langle \partial_j \varepsilon_t(\bar{\varphi}_T), \varepsilon_t(\bar{\varphi}_T)\rangle \right] \\
&\qquad + \mathrm{tr}\left[\Sigma_\circ^{-1} 2 \left\langle (\partial_{i,j}^2 \varepsilon_t(\bar{\varphi}_T)), \varepsilon_t(\bar{\varphi}_T)\right\rangle \right] + \mathrm{tr}\left[\Sigma_\circ^{-1} 2 \langle (\partial_j \varepsilon_t(\bar{\varphi}_T)), (\partial_i \varepsilon_t(\bar{\varphi}_T))\rangle \right] + o_P(1) \qquad \text{(B.5)}
\end{aligned}$$

for every sequence $\bar{\varphi} \to \varphi_\circ$. This follows from the USE condition, see Lemma B.1 in combination with $\langle \varepsilon_t(\varphi_\circ), \varepsilon_t(\varphi_\circ)\rangle \to \Sigma_\circ$.
According to the partitioning of φ in five sub-vectors in total 15 matrix blocks (taking into account symmetry of the Hessian) have to be dealt with. The blocks are partitioned according to how often differentiation takes place with respect to a component of θ_u, $e_{m+1,\perp}$, e_{m+1}, $\theta_{d,b}$ and θ_{st}.
The multiplication of the Hessian with D_T and $\tilde{D}_T$ has the following effect: For each derivative with respect to an entry in $e_{m+1,\perp}$ an additional scaling factor T^{-1} is introduced and for each derivative with respect to an entry in θ_u an additional scaling factor $T^{-1/2}$ is introduced, which results in the proper scaling factor for each of the terms to obtain convergence in distribution.
In the above expression (B.5) the variable $\varepsilon_t(\varphi_T)$ appears, in the first and second term to be precise. This variable has to be evaluated at the point $\bar{\varphi}_T$. Due to the assumptions $\bar{\varphi}_T$ converges to $\varphi_\circ$. Hence, applying a mean value expansion again $\varepsilon_t(\bar{\varphi}_T) = \varepsilon_t + \partial\varepsilon_t(\tilde{\varphi})(\bar{\varphi}_T - \varphi_\circ)$, for suitable intermediate value $\tilde{\varphi}$, it follows that both mentioned terms converge to 0. Look for example at the second term with essential term

$$\left\langle (\partial^2_{i,j}\varepsilon_t(\bar{\varphi}_T)), \varepsilon_t(\bar{\varphi}_T)\right\rangle = \left\langle (\partial^2_{i,j}\varepsilon_t(\bar{\varphi}_T)), \varepsilon_t\right\rangle + \sum_{l=1}^{\dim(\varphi)} \left\langle (\partial^2_{i,j}\varepsilon_t(\bar{\varphi}_T)), \partial_l\varepsilon_t(\tilde{\varphi})\right\rangle (\bar{\varphi}_{l,T} - \varphi_{l,\circ}).$$

Lemmas B.2 and B.3 show that for this term for all possible combinations of differentiation (including the necessary normalization if differentiation occurs with respect to an entry of θ_u or $e_{m+1,\perp}$), that the first term of the above equation converges to 0. Due to the established *condition USE* this convergence is uniformly in a compact neighborhood of $\varphi_\circ$. Analogous considerations deliver convergence of the second term to 0 as well. Here, the terms $(\partial^2_{i,j}\varepsilon_t(\varphi))\partial_l\varepsilon_t(\tilde{\varphi})'$ converge to random variables, post-multiplying with $(\bar{\varphi}_{l,T} - \varphi_{l,\circ})$ then delivers the result. Similar considerations also apply to the first term of equation (B.5). Hence, we obtain:

$$\partial^2_{i,j}\mathcal{L}_T(\bar{\varphi}_T) = \operatorname{tr}\left[\Sigma_\circ^{-1} 2\langle (\partial_i\varepsilon_t(\bar{\varphi}_T)), (\partial_j\varepsilon_t(\bar{\varphi}_T))\rangle\right] + o_P(T^{N_u/2+N_{e_\perp}}), \tag{B.6}$$

where N_u counts the number of times differentiation takes place with respect to an element of θ_u and $N_{e_\perp}$ counts the number of times differentiation takes place with respect to an element of $e_{m+1,\perp}$. Now, starting from equation (B.6), we analyze the asymptotic behavior of the derivatives.
$i \sim \theta_{st}, j \sim \theta_{st}$: If differentiation takes place twice with respect to an entry of θ_{st}, then all quantities in the above equation are asymptotically stationary, see also the previous lemma. In this case convergence to a constant matrix follows, using uniform convergence in a compact neighborhood of $\varphi_\circ$.
$i \sim \theta_{st}, j \sim \theta_{d,b}$: If differentiation takes place once with respect to an entry of θ_{st} and once with respect to an entry of $\theta_{d,b}$, convergence of $\partial^2_{i,j}\mathcal{L}_T(\bar{\varphi}_T) \to 0$ follows.
$i \sim \theta_{d,b}, j \sim \theta_{d,b}$: If differentiation takes place twice with respect to an entry in $\theta_{d,b}$, the relevant term is given by $\operatorname{tr}\left[\Sigma_\circ^{-1} 2\langle (k^{-1}(L;\bar{\theta})\partial_i D s_t), (k^{-1}(L;\bar{\theta})\partial_j D s_t)\rangle\right]$. This directly implies that the asymptotic entry in the (limit of the) Hessian is only non-zero, if both entries of d_b with respect to which differentiation takes place correspond to elements in the same column of D, k say. For $z_k \neq \pm 1$ the corresponding limit block in the Hessian is in this case given by

$$\begin{bmatrix} 4\alpha'_{k,\circ}\Sigma_\circ^{-1}\alpha_{k,\circ} & 0 \\ 0 & 4\alpha'_{k,\circ}\Sigma_\circ^{-1}\alpha_{k,\circ} \end{bmatrix}.$$

For $z_k = \pm 1$ the block is equal to $2\alpha'_{k,\circ}\Sigma_\circ^{-1}\alpha_{k,\circ}$.
$i \sim \theta_{st}, j \sim \theta_u$: If differentiation takes place once with respect to an entry of θ_{st} and once with respect to an entry of θ_u, the corresponding entry is the sum of a product of a stationary process with an integrated process (integrated of order 1 at the corresponding unit root). The normalization factor $T^{-3/2}$, with which these elements are scaled, then implies convergence to zero of the scaled quantity.
$i \sim \theta_u, j \sim \theta_u$: Next, we examine the case of differentiating twice with respect to an entry in θ_u. Note in this respect first that if in differentiating twice with respect to entries of θ_u, the two parameters correspond to different unit roots, it holds that $T^{-1}\partial^2_{i,j}\mathcal{L}_T(\bar{\varphi}_T) \to 0$. This follows from

observing that $\partial_i \varepsilon_t(\varphi) = k^{-1}(L;\theta)(-\partial_i C_k) x_{t,k}(\theta)$, where we assume that the entry with respect to which differentiation takes place corresponds to the pair of complex conjugate unit roots $z_k, \overline{z}_k$ or to $z_k = \pm 1$ and $x_{t,k}(\theta)$ denotes as before in Lemma B.8 the state vector in the real valued canonical form. The above expression for the partial derivative can directly be investigated using Lemma B.3, item (iii). The lemma provides similarly the result for the case that both i and j correspond to the same unit root or pair of unit roots:

$$T^{-1}\partial^2_{i,j}\mathcal{L}_T(\bar{\varphi}_T) \xrightarrow{d} 2\delta_k^2 \mathrm{tr}\left\{\mathcal{R}\left\{\left[\partial_j C_k' \Pi_{k,\circ}' \Sigma_\circ^{-1} \Pi_{k,\circ} \partial_i C_k Z(z_k)\right]\right\}\right\}.$$

$i \sim \theta_u, j \sim \theta_{d,b}$: Now, consider first differentiation with respect to an entry of θ_u and then with respect to an entry of $\theta_{d,b}$. Here we have to distinguish two cases, whether j corresponds to an element of the j-th column of $\mathcal{R}(D)$ or to an element in the j-th column of $\mathcal{I}(D)$ (which is corresponding to $s^1_{t,j} = z_j^{t-1}$). Suppose for the moment that it corresponds to an element of $\mathcal{R}(D)$, and let i denote a component of θ_u that corresponds to unit root z_k. Then, we obtain

$$T^{-1/2}\partial^2_{i,j}\mathcal{L}_T(\bar{\varphi}_T) \xrightarrow{d} -2\ \mathcal{R}\left\{e_i' \alpha_{k,\circ}' \Sigma_\circ^{-1} \Pi_{k,\circ} \partial_i C_k Y(z_k)\right\}$$

and zero else. The expression for an element of $\theta_{d,b}$ corresponding to an entry in the j-th column of $\mathcal{I}(D)$ is the same, except for replacing $\mathcal{R}$ by $\mathcal{I}$.
$i \sim e_{m+1,\perp}, j \sim e_{m+1,\perp}$: If differentiation takes place with respect to two entries of $e_{m+1,\perp}$, the dominating term is

$$\begin{aligned} T^{-2}\partial^2_{i,j}\mathcal{L}_T(\bar{\varphi}_T) &= \frac{2}{T^3}\sum_{t=1}^{T} e_i' \alpha_{1,\circ}' \Sigma_\circ^{-1} \alpha_{1,\circ} e_j t^2 \\ &\to \frac{2}{3} e_i' \alpha_{1,\circ}' \Sigma_\circ^{-1} \alpha_{1,\circ} e_j, \end{aligned}$$

where the convergence follows from the second statement of Lemma B.3 (vi).
$i \sim e_{m+1,\perp}, j \sim \theta_{st}$: If differentiation takes place with respect to an entry of $e_{m+1,\perp}$ and an entry of θ_{st}, the dominating term is of the form $T^{-1}\langle t, \tilde{\varepsilon}_t\rangle$ where $\tilde{\varepsilon}_t$ is a stationary process. Since by Lemma B.3 (viii) such a sum converges in distribution when the scaling factor is $T^{-3/2}$, these terms go to zero.
$i \sim e_{m+1,\perp}, j \sim \theta_{d,b}$: If differentiation takes place with respect to an entry of $e_{m+1,\perp}$ and an entry of $\theta_{d,b}$, the dominating term is

$$\begin{aligned} T^{-2}\partial^2_{i,j}\mathcal{L}_T(\bar{\varphi}_T) &= \frac{2}{T^2}\sum_{t=1}^{T} e_i' \alpha_{1,\circ}' \Sigma_\circ^{-1} \alpha_{1,\circ} e_j t s_t(z_k) \\ &\to e_i' \alpha_{1,\circ}' \Sigma_\circ^{-1} \alpha_{1,\circ} e_j \delta_{k,1}, \end{aligned}$$

where the convergence follows from the first statement of Lemma B.3 (vi).
$i \sim e_{m+1,\perp}, j \sim \theta_u$: If differentiation takes place with respect to an entry of $e_{m+1,\perp}$ and an entry of θ_u, the dominating term is

$$\begin{aligned} T^{-3/2}\partial^2_{i,j}\mathcal{L}_T(\bar{\varphi}_T) &= -\frac{1}{T^{5/2}}\sum_{t=1}^{T} e_i' \alpha_{1,\circ}' \Sigma_\circ^{-1} \Pi_{1,\circ} \partial_j C_1 x_{t,k}(\theta_\circ) t \\ &\to -2 e_i' \alpha_{1,\circ}' \Sigma_\circ^{-1} \Pi_{1,\circ} \partial_j C_1 \int_0^1 u W_1(u) du B_{1,\circ}', \end{aligned}$$

where we have used Lemma B.3 (vii) for the convergence.
$i \sim e_{m+1}, j \sim e_{m+1}$: If differentiation takes place with respect to two entries of e_{m+1}, the dominating term is

$$\begin{aligned} \partial^2_{i,j}\mathcal{L}_T(\bar{\varphi}_T) &= \frac{1}{T}\sum_{t=1}^{T} 2 e_i' (\Gamma_{1,\circ} C_{1,\circ})' \Sigma_\circ^{-1} \Gamma_{1,\circ} C_{1,\circ} e_j \\ &\to 2 e_i' (\Gamma_{1,\circ} C_{1,\circ})' \Sigma_\circ^{-1} \Gamma_{1,\circ} C_{1,\circ} e_j. \end{aligned}$$

$i \sim e_{m+1}, j \sim e_{m+1,\perp}$: If differentiation takes place with respect to one entry of e_{m+1} and one entry of $e_{m+1,\perp}$, the dominating term is

$$\begin{aligned}\frac{1}{T}\partial^2_{i,j}\mathcal{L}_T(\bar{\varphi}_T) &= \frac{1}{T^2}\sum_{t=1}^{T} 2e_i'(\Gamma_{1,\circ}C_{1,\circ})'\Sigma_\circ^{-1}\alpha_{1,\circ}e_j t \\ &\to e_i'(\Gamma_{1,\circ}C_{1,\circ})'\Sigma_\circ^{-1}\alpha_{1,\circ}e_j,\end{aligned}$$

where we have used the first statement of Lemma B.3 (vi) for the convergence.
$i \sim \theta_u, j \sim e_{m+1}$: By the same argument as in the case of differentiation with respect to an entry of θ_u and an entry of $\theta_{d,b}$ in the case of differentiation with respect to an entry of θ_u and an entry of e_{m+1} we have

$$T^{-1/2}\partial^2_{i,j}\mathcal{L}_T(\bar{\varphi}_T) \xrightarrow{d} 2e_i'(\Gamma_{1,\circ}C_{1,\circ})'\Sigma_\circ^{-1}\Pi_{1\circ}\partial_j C_1 Y(1),$$

if the entry of θ_u corresponds to the unit root frequency zero and convergence to zero else.
$i \sim \theta_{d,b}, j \sim e_{m+1}$: If differentiation takes place with respect to an entry of e_{m+1} and an entry of $\theta_{d,b}$, the relevant term is

$$\frac{2}{T}\sum_{t=1}^{T}\alpha_{1,\circ}'s_{t,k}\Sigma_\circ^{-1}\Gamma_{1,\circ}C_{1,\circ} \to \delta_{k,1}2\alpha_{1,\circ}'\Sigma_\circ^{-1}\Gamma_{1,\circ}C_{1,\circ}.$$

$i \sim \theta_{st}, j \sim e_{m+1}$: In the case where derivatives are taken with respect to an entry of e_{m+1} and an entry of θ_{st}, the dominating term is of the form

$$\frac{2}{T}\sum_{t=1}^{T}(\Gamma_{1,\circ}C_{1,\circ})'\Sigma_\circ^{-1}\tilde{\varepsilon}_t$$

for some asymptotically stationary process $\tilde{\varepsilon}_t$. This term goes to zero by the second statement of Lemma B.3 (i).
It remains to analyze the non-singularity properties of Z. In the case $D_\circ = \hat{D} = 0$ the block-diagonality of the asymptotic Hessian implies that it is sufficient to treat the blocks Z_u and Z_{st} separately. If deterministic terms are present it is sufficient to investigate Z_{st}, Z_D and $Z_u - Y_D'Z_D^{-1}Y_D$.
Consider the block Z_{st} corresponding to θ_{st} first: This block converges in fact to a constant matrix, i. e., asymptotic non-singularity is shown, if the limiting matrix is non-singular. For the part of θ_{st} corresponding to the parameters for $k_\bullet(z)$ this follows again from standard theory for stationary processes. Since the parameters of the stable and the unstable part of the transfer function are independent of each other, we can consider the unstable part alone.
Thus, only the derivatives corresponding to $\theta_{C,D}$, $\theta_{C,R}$, $\theta_{B,f}$ and $\theta_{B,p}$ have to be analyzed. The proof is indirect: If the sub-block of Z_{st} corresponding to $\theta_{C,D}$, $\theta_{C,R}$, $\theta_{B,f}$ and $\theta_{B,p}$ were singular, there existed a vector $x = [\ x_1 \quad \dots \quad x_v\ ]'$ such that

$$0 = \sum_{r,s=1}^{v} x_r x_s \mathrm{tr}[\Sigma_\circ^{-1}\mathbb{E}\partial_s\varepsilon_t(\varphi_\circ)\partial_r\varepsilon_t(\varphi_\circ)'] = \mathrm{tr}\left[\Sigma_\circ^{-1}\mathbb{E}\sum_{r=1}^{v} x_r\partial_r\varepsilon_t(\varphi_\circ)\left(\sum_{s=1}^{v} x_s\partial_s\varepsilon_t(\varphi_\circ)\right)'\right],$$

denoting the components of θ_{st} corresponding to $\theta_{C,D}$, $\theta_{C,R}$, $\theta_{B,f}$ and $\theta_{B,p}$ with $1, \dots, v$ for some integer v. This implies that

$$\sum_r x_r\partial_r\varepsilon_t(\varphi_\circ) = -k_0^{-1}(L)\sum_r x_r\partial_r k(z;\theta)\varepsilon_t$$

is equal to zero and thus that the filters for generating the score are linearly dependent.
The coefficients of the unstable part of the transfer function are of the form $K_{j,u} = \sum_{k=1}^{l}\mathcal{C}_k\mathcal{A}_k^{j-1}\mathcal{B}_k$ with $\mathcal{C}_k$ and $\mathcal{B}_k$ of full rank. Thus, linear dependence between the derivatives with respect to parameters corresponding to different unit roots cannot happen. Thus, linear dependence implies

$\sum_r x_r \partial_r \mathcal{C}_k \mathcal{B}_k = 0$. In the following we will show that this implies $x_r = 0$.
We start with the case of a real unit root z_k. We show the independence by induction over c_k. In the case $c_k = 2$, $s = 2$ there are four parameters in θ_{st}, the three parameters in $\mathcal{B}_k = \begin{bmatrix} b_{11} & b_{12} \\ 0 & b_{22} \end{bmatrix}$ and the parameter of the Givens rotation in $\mathcal{C}_k = R_R = \begin{bmatrix} \cos(\omega) & \sin(\omega) \\ -\sin(\omega) & \cos(\omega) \end{bmatrix}$, see Bauer et al. (2020, Section 3.1) for details. The four derivatives are

$$\begin{aligned}
\frac{\partial(\mathcal{C}_k\mathcal{B}_k)}{\partial\omega} &= \begin{bmatrix} -b_{11}\sin(\omega) & -b_{12}\sin(\omega) + b_{22}\cos(\omega) \\ -b_{11}\cos(\omega) & -b_{12}\cos(\omega) - b_{22}\sin(\omega) \end{bmatrix}, \\
\frac{\partial(\mathcal{C}_k\mathcal{B}_k)}{\partial b_{11}} &= \begin{bmatrix} \cos(\omega) & 0 \\ -\sin(\omega) & 0 \end{bmatrix}, \\
\frac{\partial(\mathcal{C}_k\mathcal{B}_k)}{\partial b_{12}} &= \begin{bmatrix} 0 & \cos(\omega) \\ 0 & -\sin(\omega) \end{bmatrix}, \\
\frac{\partial(\mathcal{C}_k\mathcal{B}_k)}{\partial b_{22}} &= \begin{bmatrix} 0 & \sin(\omega) \\ 0 & \cos(\omega) \end{bmatrix}.
\end{aligned}$$

The first columns of the first two matrices are orthogonal to each other and different from zero, since $b_{11} > 0$, the first columns of the other two matrices are zero. Thus, the coefficients x_r of the first two matrices of a linear combination of the derivatives, which is zero, must be zero. Since the second columns of the last two matrices are orthogonal and different from zero, their coefficients must also be zero and the four matrices are linearly independent. This argument is easily extended to the case $c_k = 2$, $s > 2$ as well as to the case $c_k = 1$.
Suppose the linear independence has been shown for $c_k = g$. To show the statement for $c_k = g+1$, write

$$R_R(\theta_{C,R,k}) = \prod_{i=1}^{c_k-1}\prod_{j=1}^{c_k-i} R_{c_k,i,i+j}(\theta_{i(i-1)/2+j}) = \prod_{j=1}^{c_k-1} R_{c_k,1,1+j}(\theta_j) \prod_{i=2}^{c_k-1}\prod_{j=1}^{c_k-i} R_{c_k,i,i+j}(\theta_{i(i-1)/2+j}),$$

where $R_{c_k,i,i+j}(\theta_{i(i-1)/2+j})$ is a real Givens rotation, see Bauer et al. (2020, Definition 6). Note that this is not in the form of R_R in Bauer et al. (2020, Lemma 1). It is, however, just a reordering of the Givens rotations. Since the order is arbitrary, this change does not affect the properties of the parameterization. Since the matrix R_L is of full rank, it is sufficient to show the linear independence of $R_R\mathcal{B}_k$. It is easily seen that the entries in the first column of $R_R\mathcal{B}_k$ with non-zero entries only depend on the parameters $\theta_1 \ldots, \theta_{c_k-1}$ and the first non-zero entry in the first row of $\mathcal{B}_k$. Since the first columns of the derivatives to these parameters are orthogonal to each other, the coefficients of these derivatives in a linear combination, which is zero, must be zero. Since the matrix $\prod_{j=1}^{c_k-1} R_{c_k,1,1+j}(\theta_j)$ is of full rank, the derivatives of

$$\prod_{j=1}^{c_k-1} R_{c_k,1,1+j}(\theta_j) \prod_{i=2}^{c_k-1}\prod_{j=1}^{c_k-i} R_{c_k,i,i+j}(\theta_{i(i-1)/2+j})\mathcal{B}_k$$

with respect to the other parameter are independent if and only if the derivatives of

$$\prod_{i=2}^{c_k-1}\prod_{j=1}^{c_k-i} R_{c_k,i,i+j}(\theta_{i(i-1)/2+j})\mathcal{B}_k = \begin{bmatrix} 1 & 0 \\ 0 & \prod_{i=2}^{c_k-1}\prod_{j=1}^{c_k-i} R_{c_k-1,i-1,i+j-1}(\theta_{i(i-1)/2+j}) \end{bmatrix}\mathcal{B}_k$$

are linearly independent. Then, consider the second column of this matrix. The derivative with respect to b_{12} is null outside the first row, while the derivative with respect to any other parameter is zero in the first row. Thus, clearly, the coefficient corresponding to b_{12} must be zero. By an analogous argument the coefficients corresponding to $b_{13}, \ldots, b_{1s}$ are zero. Thus, to show that the coefficients of the derivatives in the linear combination with respect to the other parameters are zero, it is sufficient to consider the lower right $(c_k - 1) \times (s - 1)$ block. This is a direct consequence

of the induction hypothesis. This finishes the proof for real unit roots.
For a complex unit root the linear independence is also shown by induction over c_k. The part of the matrix $Q_R D_{c_k}$ corresponding to θ_{st} is with a reordering of the parameters of Bauer et al. (2020, Lemma 2) of the form

$$Q_R(\theta_{C,R,k}) D_{c_k}(\theta_{C,D,k}) = \left(\prod_{i=1}^{c_k-1} \prod_{j=1}^{c_k-i} Q_{c_k,i,i+j}(\theta_{R,i(i-1)/2+j}) \right) \operatorname{diag}(\theta_{D,1}, \ldots, \theta_{D,c_k}),$$

where $Q_{c_k,i,i+j}$ are complex Givens rotations, compare Bauer et al. (2020, Definition 7), where again the ordering of the multiplications is changed to simplify the proof. In the case $c_k = 1$, $s = 2$ it must be shown that the derivatives of the matrix $\left[\begin{array}{cc} (Q_R D_{c_k} \mathcal{B}_k)' & (\overline{Q}_R \overline{D}_{c_k} \overline{\mathcal{B}}_k)' \end{array} \right]'$ with $\mathcal{B}_k = \left[\begin{array}{cc} b_{11} & \mathcal{R}(b_{12}) + i\mathcal{I}(b_{12}) \end{array} \right]$, see Bauer et al. (2020, Section 3.1), with respect to the different parameters are linearly independent for an inner point of the parameter space. Since $c_k = 1$, the matrix $Q_R(\theta_{C,R,k})$ does not exist. Thus, we have

$$\left[\begin{array}{c} (D_{c_k} \mathcal{B}_k) \\ (\overline{D}_{c_k} \overline{\mathcal{B}}_k) \end{array} \right] = \left[\begin{array}{cc} e^{i\theta_D} b_{11} & e^{i\theta_D}(\mathcal{R}(b_{12}) + i\mathcal{I}(b_{12})) \\ e^{-i\theta_D} b_{11} & e^{-i\theta_D}(\mathcal{R}(b_{12}) - i\mathcal{I}(b_{12})) \end{array} \right].$$

The derivatives of this matrix are

$$\begin{aligned}
\frac{\partial}{\partial \theta_D} &= \left[\begin{array}{cc} ie^{i\theta_D} b_{11} & ie^{i\theta_D}(\mathcal{R}(b_{12}) + i\mathcal{I}(b_{12})) \\ -ie^{-i\theta_D} b_{11} & -ie^{-i\theta_D}(\mathcal{R}(b_{12}) - i\mathcal{I}(b_{12})) \end{array} \right], \\
\frac{\partial}{\partial b_{11}} &= \left[\begin{array}{cc} e^{i\theta_D} & 0 \\ e^{-i\theta_D} & 0 \end{array} \right], \\
\frac{\partial}{\partial \mathcal{R}(b_{12})} &= \left[\begin{array}{cc} 0 & e^{i\theta_D} \\ 0 & e^{-i\theta_D} \end{array} \right], \\
\frac{\partial}{\partial \mathcal{I}(b_{12})} &= \left[\begin{array}{cc} 0 & ie^{i\theta_D} \\ 0 & -ie^{-i\theta_D} \end{array} \right],
\end{aligned}$$

which are clearly linearly independent. The extension to the case $c_k = 1$, $s > 2$ is straightforward. The first two columns of the derivatives with respect to the additional parameters are zero, such that the proof of the linear independence of the other parameters remains unchanged. It is then easy to see that the derivatives with respect to the additional parameters are linearly independent. Assume again that the linear independence has been shown for $c_k = g$. To show it for $c_k = g + 1$, we write

$$\begin{aligned}
Q_R(\theta_{C,R,k}) D_{c_k}(\theta_{C,D,k}) &= \left(\prod_{i=1}^{c_k-1} \prod_{j=1}^{c_k-i} Q_{c_k,i,i+j}(\theta_{R,i(i-1)/2+j}) \right) \operatorname{diag}(\theta_{D,1}, \ldots, \theta_{D,c_k}) \\
&= \left(\prod_{j=1}^{c_k-i} Q_{c_k,1,1+j}(\theta_{R,j}) \right) \operatorname{diag}(\theta_{D,1}, 1, \ldots, 1) \\
&\quad \left(\prod_{i=2}^{c_k-1} \prod_{j=1}^{c_k-i} Q_{c_k,i,i+j}(\theta_{R,i(i-1)/2+j}) \right) \operatorname{diag}(1, \theta_{D,2}, \ldots, \theta_{D,c_k}),
\end{aligned}$$

where we have used that $Q_{c_k,i,i+j} = \operatorname{diag}(1, Q_{c_k-1,i,i+j})$ for $i \geq 2$ and thus, the first entry of the diagonal matrix commutes with those Givens rotations. The entries in the first column of $\left[\begin{array}{cc} (Q_R D_{c_k} \mathcal{B}_k)' & (\overline{Q}_R \overline{D}_{c_k} \overline{B}_k)' \end{array} \right]'$ only depend on $\theta_{D,1}$, $\theta_{R,1}, \ldots, \theta_{R,c_k-1}$ and the first non-zero entry of the first row of $\mathcal{B}_k$. Note that $\theta_{R,j}$ consists of two parameters for every j. It is easily seen that the first columns of the derivatives with respect to the first non-zero entry of the first row of $\mathcal{B}_k$ and to the first parameter of $\theta_{R,j}$ are orthogonal to the first columns of the derivatives with respect to the

other parameters, their coefficients in any linear combination of derivatives must be zero. Since for an inner point of the parameter vector the first column of the derivative with respect to $\theta_{D,1}$ has only non-zero entries and there are entries in the first columns where the derivatives with respect to the second parameter of $\theta_{R,j}$ are zero for all j, the coefficient of the derivative with respect to $\theta_{D,1}$ is zero. Since the derivatives with respect to the second parameter of $\theta_{R,j}$ have non-zero entries at different positions in the first column, these derivatives are linearly independent. Thus, the coefficients of the derivatives with respect to $\theta_{D,1}, \theta_{R,1}, \ldots, \theta_{R,c_k-1}$ and the first non-zero entry of the first row of $\mathcal{B}_k$ are all zero. Since the matrix $\left(\prod_{j=1}^{c_k-i} Q_{c_k,1,1+j}(\theta_{R,j})\right) \operatorname{diag}(\theta_{D,1}, 1, \ldots, 1)$ is of full rank, the derivatives of

$$\left(\prod_{j=1}^{c_k-i} Q_{c_k,1,1+j}(\theta_{R,j})\right) \operatorname{diag}(\theta_{D,1}, 1, \ldots, 1)$$
$$\left(\prod_{i=2}^{c_k-1}\prod_{j=1}^{c_k-i} Q_{c_k,i,i+j}(\theta_{R,i(i-1)/2+j})\right) \operatorname{diag}(1, \theta_{D,2}, \ldots, \theta_{D,c_k})$$

with respect to the other parameters are linearly independent if and only if the derivatives of

$$\left(\prod_{i=2}^{c_k-1}\prod_{j=1}^{c_k-i} Q_{c_k,i,i+j}(\theta_{R,i(i-1)/2+j})\right) \operatorname{diag}(1, \theta_{D,2}, \ldots, \theta_{D,c_k})$$

are linearly independent. The derivatives with respect to the remaining parameters in the first row of $\mathcal{B}_k$ have non-zero entries only in rows one and $c_k + 1$. Thus, to show that the coefficients of the derivatives with respect to the other parameters, it is sufficient to consider the lower right $(c_k-1)\times(s-1)$ block of $Q_R D_{c_k} \mathcal{B}_k$ and of $\overline{Q}_R \overline{D}_{c_k} \overline{B}_k$. This is a direct consequence of the induction hypothesis. It is easy to see that the derivatives with respect to the entries of the first row of $\mathcal{B}_k$ are linearly independent. This finishes the proof.

Next, consider the same kind of argument for the block Z_u corresponding to θ_u. It has been shown above that if differentiation takes place twice with respect to an entry in θ_u, the essential term in equation (B.5) is

$$\begin{aligned} T^{-1}\partial^2_{i,j}\mathcal{L}_T(\bar{\varphi}_T) &= \frac{1}{T}\sum_{t=1}^{T} \operatorname{tr}[\Sigma^{-1}\partial_i\varepsilon_t(\varphi_\circ)\partial_j\varepsilon_t(\varphi_\circ)'] + o_P(1) \\ \partial_i\varepsilon_t(\varphi) &= dk_{i,C_k}(L,\theta_\circ)x_{t,k}(\theta_\circ) = -k(L;\theta_\circ)^{-1}\partial_i C_k x_{t,k}(\theta_\circ). \end{aligned}$$

The resulting matrix is block-diagonal, with the blocks corresponding to the different unit roots z_k. Therefore, again, the crucial fact to prove is the linear independence of the filters $dk_i(z,\theta_\circ)$ for all coordinates in θ_u corresponding to $z_k = e^{i\omega_k}$ for all k. Remember that $k^{-1}(e^{i\omega_k};\theta_\circ)C_{k,\circ} = 0$. Therefore, a necessary condition is that the derivative of C_k does not lie in the space spanned by the columns of C_k, and that the set of partial derivatives with respect to all different parameters is linearly independent. This is ensured by the specific parameterization of C_k as described in Bauer et al. (2020).

Finally, let us consider $Z_u - Y_D' Z_D^{-1} Y_D$. Define

$$\tau_{u,k,i,t} := -\Sigma_\circ^{-1/2}\Pi_{k,\circ}\partial_i\mathcal{C}_k x_{t,k}(\theta_\circ) \qquad \tau_{u,k,t} := [\ \tau'_{u,k,1,t} \quad \cdots \quad \tau'_{u,k,n_k,t}\ ]',$$

where n_k denotes the number of parameters for C_k. We see that

$$T^{-1}\partial^2_u\mathcal{L}_T(\bar{\varphi}_T) = \frac{1}{T^2}\sum_{t=1}^{T}\tau'_{u,k,t}\tau_{u,k,t} + o_P(1) \to Z_{u,k}.$$

Analogously, we define

$$\tau_{b,k,t} := -\Sigma_\circ^{-1/2}\alpha_{k,\circ}s_{k,t} \qquad \partial_{db}^2\mathcal{L}_T(\bar{\varphi}_T) = \frac{1}{T}\sum_{t=1}^{T}\tau'_{b,k,t}\tau_{b,k,t} + o_P(1) \to Z_{db,k}$$

$$\tau_{e,t} := -\Sigma_\circ^{-1/2}\Gamma_{1,\circ}C_{1,\circ} \qquad \partial_{e}^2\mathcal{L}_T(\bar{\varphi}_T) = \frac{1}{T}\sum_{t=1}^{T}\tau'_{e,t}\tau_{e,t} + o_P(1) \to Z_{e}$$

$$\tau_{e_\perp,t} := -\Sigma_\circ^{-1/2}\alpha_{1,\circ}t \qquad T^{-2}\partial_{e_\perp}^2\mathcal{L}_T(\bar{\varphi}_T) = \frac{1}{T^3}\sum_{t=1}^{T}\tau'_{e_\perp,t}\tau_{e_\perp,t} + o_P(1) \to Z_{e_\perp}$$

and analogous expressions for the sample covariances between the deterministic terms. These expressions can be simplified by noting that

$$\frac{1}{T}\sum_{t=1}^{T}\tau'_{b,k,t}\tau_{b,k,t} = \left(\frac{1}{T}\sum_{t=1}^{T}s'_{k,t}s_{k,t}\right)\alpha'_{k,\circ}\Sigma_\circ^{-1}\alpha_{k,\circ}, \quad \frac{1}{T}\sum_{t=1}^{T}s'_{k,t}s_{k,t} = 1,$$

where $\alpha'_{k,\circ}\Sigma_\circ^{-1}\alpha_{k,\circ} > 0$.
Above and in the following $Z_{u,k}$, $Z_{db,k}$, $Z_{D,k}$ and $Y_{D,k}$ denotes the blocks in Z_u, Z_{db}, Z_D and Y_D corresponding to the unit root z_k.
Let us first deal with the case, where no linear trend is present or the block corresponding to a unit root other than $z_k = 1$ is considered. To show the invertibility of $Z_{u,k} - Y'_{D,k}Z_{D,k}^{-1}Y_{D,k}$, we first investigate the invertibility of the corresponding sample covariance matrices and later take the limit of these quantities. We find

$$\frac{1}{T^2}\sum_{t=1}^{T}\tau'_{u,k,t}\tau_{u,k,t} - \underbrace{\frac{1}{T^{\frac{3}{2}}}\left(\sum_{t=1}^{T}\tau'_{u,k,t}\tau_{b,k,t}\right)}_{\to Y'_{D,k}}\underbrace{\left(\frac{1}{T}\sum_{t=1}^{T}\tau'_{b,k,t}\tau_{b,k,t}\right)^{-1}}_{\to Z_{D,k}^{-1}}\underbrace{\frac{1}{T^{\frac{3}{2}}}\left(\sum_{t=1}^{T}\tau'_{b,k,t}\tau_{u,k,t}\right)}_{\to Y_{D,k}} =$$

$$\frac{1}{T^2}\sum_{t=1}^{T}\tau'_{u,k,t}\tau_{u,k,t} - \left(\frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^{T}\tau'_{u,k,t}s_{k,t}\right)\Sigma_\circ^{-1/2}\alpha_{k,\circ}(\alpha'_{k,\circ}\Sigma_\circ^{-1}\alpha_{k,\circ})^{-1}\alpha'_{k,\circ}\Sigma_\circ^{-1/2}\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^{T}s_{k,t}\tau_{u,k,t}\right).$$

Note that, since $\tau_{u,k,t}$ lies in the span of $\Sigma_\circ^{-1/2}\alpha_{k,\circ}$ (since $\Pi_k = \alpha_{k,\circ}M_k(C_k^\perp)'$ for some matrix $M_k \in \mathbb{C}^{(s-c_k)\times(s-c_k)}$), the regression on $\tau_{b,k,t}$ is enough to regress out the subspace spanned by $s_{k,t}$ in $\tau_{u,k,t}$. Thus, the above expression is equal to the sample variance of the residuals of a regression of $\tau_{u,k,t}$ onto $s_{k,t}$. Taking the limit, $Z_{u,k}$ and $Z_{u,k} - Y'_{D,k}Z_{D,k}^{-1}Y_{D,k}$ have the same structure, where only $Z(z_k)$ is replaced by

$$\begin{aligned}
Z(z_k) - Y(z_k)Y(z_k)' &= \delta_k^2 B_k\int_0^1 W_k(u)W_k(u)'duB'_k - \delta_k^2 B_k\int_0^1 W_k(u)du\int_0^1 W'_k(u)duB'_k \\
&= \delta_k^2 B_k\left(\int_0^1 W_k(u)W_k(u)'du - \int_0^1 W_k(u)du\int_0^1 W_k{}'(u)du\right)B'_k \\
&= \delta_k^2 B_k\int_0^1\left(W_k(u) - \int_0^1 W_k(v)dv\right)\left(W_k(u) - \int_0^1 W_k(v)dv\right)'duB'_k.
\end{aligned}$$

It follows immediately that this matrix is positive with probability one.
By an analogous argument the case with the linear trend can be dealt with. Note that the regression on the constant vector $\tau_{e,t}$ does not change the residual compared to the above result, since τ_u was already corrected for a constant. Thus, we only have to add $\tau_{e_\perp,t}$ as an additional regressor. For this, we define $\tau_{D,t} := \begin{bmatrix} \tau_{b,1,t} & \tau_{e_\perp,t} \end{bmatrix}$ and $\sum_{t=1}^{T}\tau_{D,t}\tau'_{D,t} \to Z_{\tilde{D}}$, where $Z_{\tilde{D}}$ is given by

$$\alpha'_{1,\circ}\Sigma_\circ^{-1}\alpha_{1,\circ} \otimes \begin{bmatrix} \frac{2}{3} & 1 \\ 1 & 2 \end{bmatrix}.$$

Thus, we have

$$Z_{\tilde{D}}^{-1} = [\alpha_{1,\circ}'\Sigma_\circ^{-1}\alpha_{1,\circ}]^{-1} \otimes \begin{bmatrix} 6 & -3 \\ -3 & 2 \end{bmatrix}.$$

$Y_{\tilde{D}}$ is defined correspondingly. From this it follows by elementary calculations, that $Z_{u,1}$ and $Z_{u,1} - Y_{\tilde{D}}'Z_{\tilde{D}}^{-1}Y_{\tilde{D}}$ and thus also $Z_{u,1} - Y_{D,1}'Z_{D,1}^{-1}Y_{D,1}$ have again the same structure, where $Z(1)$ has to be replaced by $Z(1) - 12V(1)V(1)' + 6Y(1)V(1)' + 6V(1)Y(1)' - 4Y(1)Y(1)'$. This corresponds to replacing the Brownian motion by a demeaned and detrended Brownian motion. Again it follows immediately that this matrix is positive definite with probability one. This implies the invertibility of the Hessian and finishes the proof. □

Combining the results of the previous two lemmas, the asymptotic distributions of $T(\hat{\theta}_\star - \theta_{\star,\circ})$ and $\sqrt{T}(\hat{\theta}_{st} - \theta_{st,\circ})$ follow immediately.
In any case $\sqrt{T}(\hat{\theta}_{st} - \theta_{st,\circ}) \xrightarrow{d} \mathcal{N}(0, Z_{st}^{-1}V_{st}Z_{st}^{-1})$. Therefore, (A) holds.
(B) follows from Lemma B.9 showing that the Hessian asymptotically is block diagonal, where the blocks correspond to all parameters for one unit root.
(C) has been shown in Lemma B.3, while (D) is contained in the results of Lemma B.7 and Lemma B.9.
With respect to (E) note that if no deterministic terms are included in the estimation it follows, that the sub-block of $\hat{\theta}_u$ corresponding to the unit root z_k has a limiting distribution of the form $Z_{u,k}^{-1}\mathrm{v}_k$, where $Z_{u,k}$ is the block of Z_u corresponding to the unit root z_k and v_k is a random vector, whose entries are given by

$$(\mathrm{v}_k)_i = 2\delta_k \mathrm{tr}[\Sigma_\circ^{-1}(\Pi_{k,\circ}\partial_i C_k X(z_k))].$$

Stacking the vectors v_k into a single vector $\mathrm{v}_u = \begin{bmatrix} \mathrm{v}_1' & \dots & \mathrm{v}_l' \end{bmatrix}'$, we get the vector from Theorem 2.2. When a constant and seasonal dummies are included in the estimation, $Z(z_k)$ has to be replaced by $Z(z_k) - Y(z_k)Y(z_k)'$ and $X(z_k)$ has to be replaced by

$$\begin{aligned} X(z_k) - Y(z_k)W_k(1)' &= \delta_k^2 B_{k,\circ}\int_0^1 W_k(u)d(W_k(u))' - \delta_k^2 B_{k,\circ}\int_0^1 W_k(u)du\int_0^1 1dW_k'(u)du \\ &= \delta_k^2\int_0^1 B_{k,\circ}\left(W_k(u) - \int_0^1 W_k(v)dv\right)d(W_k(u))'. \end{aligned}$$

If a linear trend is included, $Z(1)$ has to be replaced by $Z(1) - 12V(1)V(1)' + 6Y(1)V(1)' + 6V(1)Y(1)' - 4Y(1)Y(1)'$ and $X(1)$ has to be replaced by

$$\int_0^1 B_{1,\circ}\left(W_1(u) - \int_0^1 W_1(v)dv - 12\left(u - \frac{1}{2}\right)\int_0^1\left(v - \frac{1}{2}\right)W_1(v)dv\right)dW_k(u)',$$

as follows from straightforward computations. This finally concludes the proof of Theorem 2.2.
To prove Corollary 2.2 let $Z = \mathrm{diag}(Z_\star, Z_{st})$ and note that

$$\begin{aligned} \hat{W}_R &= (R\hat{\theta} - r)'(R\hat{Z}^{-1}R')^{-1}(R\hat{\theta} - r) \\ &= ((D_T^R R(D_T^\theta)^{-1})D_T^\theta(\hat{\theta} - \theta_\circ))'(D_T^R R(D_T^\theta)^{-1} \\ &\quad (D_T^\theta\hat{Z}^{-1}D_T^\theta)(D_T^\theta)^{-1}R'D_T^R)^{-1}(D_T^R R(D_T^\theta)^{-1})D_T^\theta(\hat{\theta} - \theta_\circ) \\ &\xrightarrow{d} \begin{bmatrix} Z_\star^{-1}\mathrm{v}_\star' & \mathrm{v}_{st}' \end{bmatrix}(R^\infty)'(R^\infty Z^{-1}(R^\infty)')^{-1}R^\infty\begin{bmatrix} Z_\star^{-1}\mathrm{v}_\star' & \mathrm{v}_{st}' \end{bmatrix}', \end{aligned}$$

where we have used $D_T^\theta(\hat{\theta} - \theta_\circ) \to \begin{bmatrix} Z_\star^{-1}\mathrm{v}_\star' & \mathrm{v}_{st}' \end{bmatrix}'$ and $\hat{Z} \xrightarrow{d} Z$ because of Lemma B.9. Now, since v_{st} is asymptotically normally distributed with variance Z_{st} and $\mathrm{v}_\star$ conditionally upon $\begin{bmatrix} B_{1,\circ}W_1' & \dots & B_l W_{l,\circ}' \end{bmatrix}'$ is asymptotically normally distributed with variance $Z_\star$, the test statistic conditionally upon $\begin{bmatrix} B_{1,\circ}W_1' & \dots & B_{l,\circ}W_l' \end{bmatrix}'$ is asymptotically χ_p^2 distributed which implies that the same result holds marginally.
Finally, (F) is immediate from Lemma B.4.

Appedix of the Third Chapter

C.1 Supplementary Material

Table C.1: BIC for VAR systems of different lag lengths for the different data sets.

lag length	1	2	3	4	5	6	7	8
			1949	-	1988			
$s = 3$	-26.40	-26.69	-26.47	-26.36	-26.10	-25.86	-25.65	-25.42
$s = 6$	-35.20	-35.49	-34.68	-33.99	-33.21	-32.68	-31.96	-31.30
			1949	-	2018			
$s = 3$	-26.39	-26.65	-26.42	-26.21	-26.00	-25.68	-25.47	-25.27
$s = 6$	-35.53	-35.44	-34.46	-33.44	-32.19	-31.16	-30.14	-29.12
			1989	-	2018			
$s = 3$	-29.05	-29.03	-28.38	-27.81	-27.13	-26.51	-25.93	-25.52
$s = 6$	-40.72	-39.16	-36.75	-34.50	-32.17	-29.94	-27.87	-25.77

Table C.2: BIC for state space systems of different system orders for the different data sets.

			1949	-	1988			
system order	1	2	3	4	5	6	7	8
$s = 3$	-23.97	-25.59	-26.53	-26.45	-26.36	-26.38	-26.16	-25.99
$s = 6$	-28.77	-29.43	-31.55	-34.82	-35.42	-35.46	-35.53	-35.07
system order	9	10	11	12	13	14	15	
$s = 3$	-25.94	-25.68						
$s = 6$	-34.92	-34.99	-34.70	-34.45	-34.71	-33.93	-33.48	

			1949	-	2018			
system order	1	2	3	4	5	6	7	8
$s = 3$	-22.21	-25.32	-25.22	-26.93	-26.97	-26.91	-26.91	-26.81
$s = 6$	-24.43	-28.98	-27.24	-34.50	-35.02	-35.49	-36.56	-36.69
system order	9	10	11	12	13	14	15	
$s = 3$		-26.69	-23.50					
$s = 6$	-36.63	-36.36	-36.36	-36.00	-35.95	-36.03	-36.20	

			1989	-	2018			
system order	1	2	3	4	5	6	7	8
$s = 3$	-26.16	-28.26	-29.19	-29.02	-28.89	-28.95	-28.76	-28.30
$s = 6$	-32.91	-36.87	-39.35	-41.35	-41.94	-41.72	-41.44	-41.34
system order	9	10	11	12	13	14	15	
$s = 3$	-28.64	-28.49						
$s = 6$	-41.06	-40.68	-40.72	-40.74	-40.13	-40.37	-40.14	

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Das erste Kapitel dieser Dissertation ist auch Teil der Dissertation von Lukas Matuschek, die ebenfalls an der Fakultät Statistik der TU Dortmund eingereicht wurde. Davon abgesehen wurde diese Dissertation weder in der gegenwärtigen noch in einer anderen Fassung an der TU Dortmund oder an einer anderen Hochschule im Zusammenhang mit einer staatlichen oder akademischen Prüfung vorgelegt. Ich erkläre, dass ich bisher kein Promotionsverfahren erfolglos beendet habe und dass keine Aberkennung eines bereits erworbenen Doktorgrades vorliegt.

Patrick de Matos Ribeiro

Literaturverzeichnis

Amann, H. and Escher, J. (2008). *Analysis III*, Birkhäuser Basel, Basel.

Amemiya, T. and Wu, R. Y. (1972). The Effect of Aggregation on Prediction in the Autoregressive Model, *Journal of the American Statistical Association* **67**(339): 628–632.

Aoki, M. (1990). *State Space Modeling of Time Series*, Springer, New York.

Aoki, M. and Havenner, A. (1989). A Method for Approximate Representation of Vector Valued Time Series and its Relation to Two Alternatives, *Journal of Econometrics* **42**(2): 181–199.

Aoki, M. and Havenner, A. (1991). State Space Modeling of Multiple Time Series, *Econometric Reviews* **10**(1): 1–59.

Aoki, M. and Havenner, A. (1997). *Applications of Computer Aided Time Series Modelling*, Lecture Notes in Statistics, Springer, New York.

Bauer, D. (2001). Order Estimation for Subspace Procedures, *Automatica* **37**(10): 1561–1573.

Bauer, D. (2009). Almost Sure Bounds on the Estimation Error for OLS Estimators when the Regressors Include Certain MFI(1) Processes, *Econometric Theory* **25**(2): 571–582.

Bauer, D. and Buschmeier, R. (2016). Asymptotic Properties of Subspace Methods for the Estimation of Seasonally Cointegrated Models, *mimeo* .

Bauer, D., Matuschek, L., de Matos Ribeiro, P. and Wagner, M. (2020). A Parameterization of MFI(1) and I(2) Processes: Structure Theory and Hypothesis Testing, *Econometrics* **8**(4).

Bauer, D. and Wagner, M. (2003). On Polynomial Cointegration in the State Space Framework, *mimeo* .

Bauer, D. and Wagner, M. (2007). Autoregressive Approximations of Multiple Frequency I(1) Processes, *mimeo* .

Bauer, D. and Wagner, M. (2012). A State Space Canonical Form for Unit Root Processes, *Econometric Theory* **28**(6): 1313–1349.

Campbell, J. Y. (1994). Inspecting the Mechanism: An Analytical Approach to the Stochastic Growth Model, *Journal of Monetary Economics* **33**(3): 463–506.

Chatelin, F. (1993). *Eigenvalues of Matrices*, John Wiley & Sons, New York.

Cubadda, G. (2001). Complex Reduced Rank Models For Seasonally Cointegrated Time Series, *Oxford Bulletin of Economics and Statistics* **63**(4): 497–511.

Cubadda, G. and Omtzigt, P. (2005). Small-Sample Improvements in the Statistical Analysis of Seasonally Cointegrated Systems, *Computational statistics & data analysis* **49**(2): 333–348.

de Matos Ribeiro, P., Matuschek, L., Bauer, D. and Wagner, M. (2020). Pseudo Maximum Likelihood Parameter Estimation for Multiple Frequency I(1) Processes: A State Space Approach, *mimeo* .

Engle, R. F. and Granger, C. W. (1987). Cointegration and Error Correction: Representation, Estimation and Testing, *Econometrica* **55**(2): 251–276.

Golub, G. H. and van Loan, C. F. (1996). *Matrix Computations, 3.ed.*, The Johns Hopkins University Press, Baltimore.

Granger, C. W. (1981). Some Properties of Time Series Data and Their Use in Econometric Model Specification, *Journal of Econometrics* **16**(1): 121–130.

Gregoir, S. (1999). Multivariate Time Series with Various Hidden Unit Roots, Part I, *Econometric Theory* **15**(4): 435–468.

Hannan, E. J. and Deistler, M. (1988). *The Statistical Theory of Linear Systems*, John Wiley & Sons, New York.

Hazewinkel, M. and Kalman, R. E. (1976). Invariants, Canonical Forms and Moduli for Linear, Constant, Finite Dimensional, Dynamical Systems, *in* G. Marchesin and S. K. Mitter (eds), *Mathematical Systems Theory*, Springer-Verlag, Berlin, chapter 4, pp. 48–60.

Jensen, A. N. (2013). The Nesting Structure of the Cointegrated Vector Autoregressive Models, *QED Conference 2013*, Vienna, Austria.

Johansen, S. (1991). Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models, *Econometrica* **59**(6): 1551–1580.

Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Auto-Regressive Models*, Oxford University Press, Oxford.

Johansen, S. (1997). Likelihood Analysis of the I(2) Model, *Scandinavian Journal of Statistics* **24**(4): 433–462.

Johansen, S. (2006). Statistical Analysis of Hypotheses on the Cointegrating Relations in the I(2) Model, *Journal of Econometrics* **132**(1): 81–115.

Johansen, S. and Nielsen, M. O. (2018). The Cointegrated Vector Autoregressive Model with General Deterministic Terms, *Journal of Econometrics* **202**(2): 214–229.

Johansen, S. and Schaumburg, E. (1999). Likelihood Analysis of Seasonal Cointegration, *Journal of Econometrics* **88**(2): 301–339.

Juselius, K. (2006). *The Cointegrated VAR Model: Methodology and Applications*, Oxford University Press, Oxford.

King, R. G., Plosser, C. I., Stock, James, H. and Watson, M. W. (1991). Stochastic Trends and Economic Fluctuations, *American Economic Review* **81**(4): 819–840.

Lee, H. S. (1992). Maximum Likelihood Inference on Cointegration and Seasonal Cointegration, *Journal of Econometrics* **54**(1–3): 1–47.

Lewis, R. and Reinsel, G. C. (1985). Prediction of Multivariate Time Series by Autoregressive Model Fitting, *Journal of Multivariate Analysis* **16**(3): 393–411.

Lütkepohl, H. and Claessen, H. (1997). Analysis of Cointegrated VARMA Processes, *Journal of Econometrics* **80**(2): 223–229.

Matuschek, L., de Matos Ribeiro, P., Bauer, D. and Wagner, M. (2020). Inference on Cointegrating Ranks and Spaces of Multiple Frequency I(1) Processes: A State Space Approach, *mimeo* .

Otto, M. (2011). *Rechenmethoden für Studierende der Physik im ersten Jahr*, Spektrum Akademischer Verlag, Heidelberg.

Paruolo, P. (1996). On the Determination of Integration Indices in I(2) Systems, *Journal of Econometrics* **72**(1-2): 313–356.

Paruolo, P. (2000). Asymptotic Efficiency of the Two Stages Estimator in I(2) Systems, *Econometric Theory* **16**(4): 524–550.

Poskitt, D. S. (2006). On the Identification and Estimation of Nonstationary and Cointegrated ARMAX Systems, *Econometric Theory* **22**(6): 1138–1175.

Rahbek, A., Kongsted, H. C. and Jorgensen, C. (1999). Trend Stationarity in the I(2) Cointegration Model, *Journal of Econometrics* **90**(2): 265–289.

Ribarits, T. and Hanzon, B. (2014). The State-Space Error Correction Model: Definition, Estimation and Model Selection, *Available at SSRN: https://ssrn.com/abstract=2516529* .

Saikkonen, P. (1992). Estimation and Testing of Cointegrated Systems by an Autoregressive Approximation, *Econometric Theory* **8**(1): 1–27.

Saikkonen, P. (1993). Continuous Weak Convergence and Stochastic Equicontinuity Results for Integrated Processes with an Application to the Estimation of a Regression Mode, *Econometric Theory* **9**(2): 155–188.

Saikkonen, P. (1995). Problems with the Asymptotic Theory of Maximum Likelihood Estimation in Integrated and Cointegrated Systems, *Econometric Theory* **11**(5): 888–911.

Saikkonen, P. and Luukkonen, R. (1997). Testing Cointegration in Infinite Order Vector Autoregressive Processes, *Journal of Econometrics* **81**(1): 93–126.

Sims, C. A., Stock, J. H. and Watson, M. W. (1990). Inference in Linear Time Series Models with Some Unit Roots, *Econometrica* **58**(1): 113–144.

Stock, J. H. and Watson, M. W. (1988). Testing for Common Trends, *Journal of the American Statistical Association* **83**(404): 1097–1107.

Stock, J. H. and Watson, M. W. (1993). A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems, *Econometrica* **61**(4): 783–820.

Wagner, M. (2018). Estimation and Inference for Cointegrating Regressions, *Oxford Research Encyclopedia of Economics and Finance* .

Wagner, M. and Hlouskova, J. (2009). The Performance of Panel Cointegration Methods: Results from a Large Scale Simulation Study, *Econometric Reviews* **29**(2): 182–223.

Yap, S. F. and Reinsel, G. C. (1995). Estimation and Testing for Unit Roots in a Partially Nonstationary Vector Autoregressive Moving Average Model, *Journal of the American Statistical Association* **90**(429): 253–267.

Zellner, A. and Palm, F. C. (1974). Time Series Analysis and Simultaneous Equation Econometric Models, *Journal of Econometrics* **2**(1): 17–54.