Essays on Cointegration Analysis in the State Space Framework

Dissertation

zur Erlangung des Grades
eines Doktors der Naturwissenschaften

eingereicht an der Fakultät Statistik
der Technischen Universität Dortmund

vorgelegt von

Lukas Matuschek

Dortmund, Oktober 2021

Dissertation
eingereicht an der Fakultät Statistik
der Technischen Universität Dortmund
verteidigt am 16. Dezember 2020

Gutachter:
Prof. Dr. Martin Wagner, TU Dortmund
Prof. Dr. Walter Krämer, TU Dortmund
Prof. Dr. Dietmar Bauer, Universität Bielefeld

To Marta

Contents

1 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 8
1.3.1 The Parameterization in the MFI(1) Case 19
1.3.2 The Parameterization in the I(2) Case 24
1.3.3 The Parameterization in the General Case 27
1.4 The Topological Structure 29
1.4.1 The Closure of Δ_Γ in Δ_n 30
1.4.2 The Closure of M_Γ 33
1.5 Testing Commonly Used Hypotheses in the MFI(1) and I(2) Cases 35
1.5.1 The $MFI(1)$ Case 35
1.5.2 The $I(2)$ Case 41
1.6 Summary and Conclusions 43

2 Inference on Cointegrating Ranks and Spaces of Multiple Frequency I(1) Processes: A State Space Approach 47
2.1 Introduction 47
2.2 Definitions 49
2.2.1 Unit Root Processes and Cointegration 49
2.3 I(1) processes 54
2.3.1 Error Correction Representation in the State Space Framework 54
2.3.2 Concentration of the Gaussian Pseudo Log-Likelihood Function 56
2.3.3 Inference on Cointegrating Spaces 58
2.3.4 Inference on Cointegrating Ranks 59
2.4 MFI(1) processes 60
2.4.1 Error Correction Representation in the State Space Framework 60
2.4.2 Concentration of the Gaussian Pseudo Log-Likelihood Function 61
2.4.3 Inference on Cointegrating Ranks and Spaces 64
2.5 Simulation Results 66
2.6 Summary and Conclusion 69

3 Pseudo Maximum Likelihood Estimation and Inference for I(2) Processes: A State Space Approach 71
3.1 Introduction 71
3.2 I(2) Processes in the State Space Framework 72
3.2.1 The State Space Representation of I(2) Processes 73
3.2.2 Error Correction Models for I(2) Processes 75
3.3 Pseudo Maximum Likelihood Estimation 76
3.4 Tests for the State Space Unit Root Structure 81

3.5 Simulation Results 84
3.6 Summary and Conclusion 87

A Appendix to Chapter 1 **89**
A.1 Proofs of the Results of Section 1.3 89
A.1.1 Proof of Lemma 1 89
A.1.2 Proof of Lemma 2 90
A.1.3 Proof of Theorem 2 90
A.2 Proofs of the Results of Section 1.4 91
A.2.1 Proof of Theorem 3 91
A.2.2 Proof of Theorem 4 92
A.2.3 Proof of Theorem 5 92

B Appendix to Chapter 2 **95**
B.1 Proof of Theorem 6 95
B.2 Auxiliary Results for the SSECM 96
B.3 Asymptotic Results in the SSECM 99
B.4 Proof of Theorem 7 104
B.5 Proof of Theorem 8 109

C Appendix to Chapter 3 **113**
C.1 SSECM 113
C.2 PML Estimator 114
C.2.1 Preliminaries 114
C.2.2 Proof of Consistency 118
C.2.3 Derivation of the Asymptotic Distribution 127
C.3 Convergence Results 140
C.4 Proof of Theorem 12 148

Introduction

Cointegration analysis is by now a standard tool in multivariate time series analysis with application ranging from economics to climate science. The main idea of this concept is the assumption of underlying stochastic trends in a multivariate stochastic process. If the number of stochastic trends is smaller than the dimension of the multivariate process, the aim is to find linear combinations within the times series, that cancel out the stochastic trends. Knowledge about the number of the stochastic trends as well as their influence on the observable times series might then improve prediction performance or validate results from economic theory. Evidence for or against a common stochastic trend between, e. g., money supply and inflation or money supply and investment may be of high interest, e. g., from the point of view of a central bank.

Cointegration was first considered by Granger (1981), and further formalized by Soren Johansen and Katarina Juselius and their co-authors, see, e. g., the monographs Johansen (1995) and Juselius (2006), who introduced a vector error correction model (VECM) and, thus, chose the class of vector autoregressive (VAR) processes to model cointegrated processes. This choice may be too restrictive, since, e. g., Zellner and Palm (1974) show that VAR processes are not invariant with respect to marginalization (subsets of variables of a VAR process are not necessarily VAR processes). Similarly, VAR processes are not invariant with respect to aggregation. An invariant class with respect to these two operations are the vector autoregressive moving average (VARMA) processes. These processes, e. g., also occur as linearized solutions to dynamic stochastic general equilibrium models, compare Campbell (1994).

Note that a VARMA process can be approximated arbitrary closely by a VAR process, such that no changes should occur in the overall asymptotic results regarding cointegration analysis using VARMA processes as a model class. However, VARMA models might be more parsimonious in some cases, which together with the arguments presented above, was enough to justify further research in this area. The subsequent work led to the completion of this thesis, which consists of three chapters corresponding to three articles written in collaboration with my co-authors Professor Dietmar Bauer, Patrick de Matos Ribeiro and Professor Martin Wagner. In the analysis we employ the representation of VARMA processes by state space systems, see, e. g., Hannan and Deistler (1988). The chosen representation decomposes the VARMA process into underlying components including its stochastic trends, thus, allowing for an intuitive understanding of the cointegrating properties of the process.

Chapter 1 first focuses on theoretical results regarding the sets of transfer functions corresponding to VARMA systems with similar cointegrating properties, summarized in the so-called state space unit root structure. We develop and discuss different parameterizations for vector autoregressive moving average processes with arbitrary unit roots and (co)integration orders. The detailed analysis of the topological properties of the parameterizations – based upon the state space canonical form of Bauer and Wagner (2012) – is an essential input for establishing statistical and numerical properties of pseudo maximum likelihood estimators as well as, e. g., pseudo likelihood ratio tests based upon them. The general results are exemplified in detail for the empirically most relevant cases, the (multiple frequency or seasonal) I(1) and the I(2) case. For these two cases we also discuss the modeling of deterministic components in detail.

In Chapter 2 we show that the Johansen framework for testing hypotheses on the cointegrating ranks and spaces for seasonally integrated processes of the multiple frequency I(1)(MFI(1)) type can be extended to the class of VARMA processes by using a state space error correction representation for MFI(1) processes. The estimated cointegrating vectors are asymptotically mixed Gaussian and pseudo likelihood ratio tests of linear restrictions on the cointegrating spaces are χ^2 distributed. Also, pseudo likelihood ratio tests for the cointegrating ranks have the same distributions under the null hypothesis in the VARMA case as in the VAR case. Hence, no new tables for critical values are needed. In a simulation study our tests outperform the tests by Johansen and Schaumburg and the canonical variate analysis subspace tests in small samples in a MFI(1) setting considerably.

In Chapter 3 we develop estimation and inference techniques for I(2) cointegrated VARMA processes cast in state space format. In particular, we derive consistency as well as the asymptotic distributions of estimators maximizing the Gaussian pseudo likelihood function. As usual, the parameters corresponding to I(2) and I(1) stochastic trends are estimated super-consistently at rates T^2 and T respectively, whereas the parameters of the stationary components of the state are estimated at rate $T^{1/2}$. The limiting distributions of the parameters corresponding to the integrated components are mixtures of Brownian motions, the parameters of the stationary subsystem are asymptotically normally distributed. Furthermore, we discuss hypothesis tests for the state space unit root structure, leading to the well-known limiting distributions for VAR I(2) processes. Again, a small simulation study shows favorable results for small samples, with our test leading to better performance in determining these integer parameters.

All simulations have been performed in MATLAB. The code containing the respective procedures can be obtained from the author upon request.

Acknowledgments

Zuallererst danke ich Gott unserem Schöpfer, dass neben der leeren Menge auch anderes existiert, was sich zu erforschen lohnt, und meinem Herrn und Erlöser Jesus Christus, dass Er meine leeren Mengen durch Seine Gnade in Seinem Blut zu verwandeln weiß. Schließlich dem Heiligen Geist, Dem ich alles anvertraue, ohne Dem allem Nachsinnen jeder Sinn und das Ziel fehlt.
Ich danke Dir, Marta, für Deine Liebe und die fordernde Art, mit der Du diese Dissertation ermöglicht hast und gleichzeitig mir und unseren Kindern die Wärme schenkst, als emotionales Zentrum unseres Hauses, für Deine Schönheit.
Mein herzlicher Dank gilt meinen Koautoren. Ich danke meinem Professor Martin Wagner, für seine Geduld bei den Manuskripten, für sein strapaziertes Vertrauen und die Weisheit, dass Präzision und Sorgfalt bei der Arbeit mehr bringen als die besten Ideen und skizzierten Projekte. Wenn dies in dieser Dissertation noch nicht vollständig zum Tragen kommt, nehme ich dies doch aus dieser lehrreichen Zeit mit. Professor Dietmar Bauer danke ich für sein schnelles Mitdenken und das Niederschlagen einiger noch lückenhafter Beweise. Schließlich danke ich meinem Kollegen und Freund Patrick Ribeiro für die prägende Zeit am Schreibtisch und Telefon.
Ich danke meinen Freunden am Lehrstuhl, Rafael (mit dem die Zeit hier anfing), Fabian, Peter und Oliver, allen von der TU Dortmund und Uni Bielefeld, die diese fünf Jahre bereichert haben.
Mein Dank gebührt auch der DFG für die finanzielle Unterstützung im Rahmen der Projekte BA 5404/1-1 und WA 3427/1-1, in denen diese Dissertation entstanden ist.
Ich danke meinen Eltern für Ihre Sorge und Unterstützung und dafür, dass Sie die Anfänge meines Weges behütet haben. Meiner Schwester Ania, die auch fast von Anfang dabei ist, dazu meinem Großvater Johann, meinen Schwiegereltern, meiner Schwägerin Sarah, und Babcia Danusia. Bez waszej pomocy i waszego uśmiechu, cała praca byłaby dużo trudniejsza. Schließlich auch meinen drei schon geborenen Kindern: Amelka, Noe i Julka.

Chapter 1

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 ag-

[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".

gregation and in mixed-frequency settings. Third, the linearized solutions to dynamic stochastic general equilibrium (DSGE) models are typically VARMA rather than VAR processes, see, e. g., Campbell (1994). Fourth, a VARMA model may be a more parsimonious description of the data generating process (DGP) than a VAR model, with parsimony becoming more important with increasing dimension of the process.[3]

If one accepts the above arguments as a motivation for considering VARMA processes in cointegration analysis, it is convenient to move to the – essentially equivalent, see Hannan and Deistler (1988, 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.

[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.

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

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

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

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

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

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

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

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

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

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

Remark 1 *In the above discussion the parameters, θ_Σ say, describing the variance covariance matrix Σ of ε_t are not considered. These can be easily included, similarly to Φ by, e.g., parameterizing positive definite symmetric $s \times s$ matrices via their lower triangular Cholesky factor. This leads to a parameter space $\Theta_{p,\Phi,\Sigma} \subset \mathbb{R}^{s^2p+sm+\frac{s(s+1)}{2}}$. 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 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 $a_p \neq 0$:

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

[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.

Considering V_p^{OLS} allows for unconstrained optimization. It is well-known that for $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ as given above, the OLS estimator is consistent over the larger set V_p^{OLS}, i. e., without imposing non-explosiveness and also when specifying p too high. Alternatively, and closely related to OLS in the VAR case, the pseudo likelihood can be maximized over $\Theta_{p,\Phi}$. With this approach, maxima respectively suprema can occur at the boundary of the parameter space, i. e., maximization effectively has to consider $\overline{\Theta}_{p,\Phi}$. 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, 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 below. Consider first the I(1) case: As is well-known, the rank of the matrix $a(1)$ equals the dimension of the cointegrating space given in Definition 3 below – also referred to as the cointegrating rank. Therefore, determination of the rank of this matrix is of key importance. With the parameterization used so far, imposing a certain (maximal) rank on $a(1)$ implies complicated restrictions on the matrices $a_j,\ j = 1, \dots, p$. This in turn renders the correspondingly restricted optimization unnecessarily complicated and not conducive to develop tests for the cointegrating rank. It is more convenient to consider the so-called *vector error correction model* (VECM) representation of autoregressive processes, discussed in full detail in the monograph Johansen (1995). To this end let us first introduce the differencing operator at frequency $0 \leq \omega \leq \pi$

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

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

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

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

This constitutes a reparameterization, where $k(z) \in V_p$ is now represented by the matrices $(\alpha, \beta, \Gamma_1, \dots, \Gamma_{p-1})$ and a corresponding parameter vector $\boldsymbol{\theta}_a^{\text{VECM}} \in \Theta_{p,r}^{\text{VECM}}$. 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 $\alpha, \beta^*, \Gamma_1, \dots, \Gamma_{p-1}$, 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:

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

The dimension of the parameter vector $\boldsymbol{\theta}_a^{\text{VECM}}$ 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

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

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

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

[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 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 $(a(z), b(z))$ can realize the same transfer function $k(z)$. It is clear that

$$k(z) = a(z)^{-1} m(z)^{-1} m(z) b(z) = a(z)^{-1} b(z)$$

[11] Below Example 3 we clarify how these indices are related to the state space unit root structure defined in Bauer and Wagner (2012, Definition 2) and link these to the dimensions of the cointegrating spaces in Section 1.5.2.

for all non-singular polynomial matrices $m(z)$. Thus, the mapping π attaching the transfer function $k(z) = a(z)^{-1}b(z)$ to the pair of polynomial matrices $(a(z), b(z))$ is not *injective*.[12]

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

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):

[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.

Definition 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 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 4, where the "simple" unit root properties of blocks of the state vector are discussed.*

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

[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).

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

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

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

Definition 2 *A matrix $B = [b_{i,j}]_{i=1,\dots,c,j=1,\dots,s} \in \mathbb{C}^{c\times s}$ is positive upper triangular (p.u.t.) if there exist integers $1 \leq j_1 \leq j_2 \leq \cdots \leq j_c \leq s+1$, 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*

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

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

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

Theorem 1 *For every transfer function $k(z) \in M_n$ there exists a unique minimal (complex) state space realization $(\mathcal{A},\mathcal{B},\mathcal{C})$ 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) $\mathcal{A} := diag(\mathcal{A}_u, \mathcal{A}_\bullet) := diag(\mathcal{A}_{1,\mathbb{C}}, \dots, \mathcal{A}_{l,\mathbb{C}}, \mathcal{A}_\bullet)$, $\mathcal{A}_u \in \mathbb{C}^{n_u\times n_u}, \mathcal{A}_\bullet \in \mathbb{R}^{n_\bullet\times n_\bullet}$, where it holds for $k = 1,\dots,l$ that

– *for $0 < \omega_k < \pi$:*

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

– *for $\omega_k \in \{0,\pi\}$:*

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

with $J_k :=$

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

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

(ii) $\mathcal{B} := [\mathcal{B}_u', \mathcal{B}_\bullet']' := [\mathcal{B}_{1,\mathbb{C}}', \dots, \mathcal{B}_{l,\mathbb{C}}', \mathcal{B}_\bullet']'$ *and* $\mathcal{C} := [\mathcal{C}_u, \mathcal{C}_\bullet] := [\mathcal{C}_{1,\mathbb{C}}, \dots, \mathcal{C}_{l,\mathbb{C}}, \mathcal{C}_\bullet]$ *are partitioned accordingly. It holds for* $k = 1, \dots, l$ *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* $\mathcal{B}_k = [\mathcal{B}_{k,1}', \dots, \mathcal{B}_{k,h_k}']'$ *as* $\mathcal{B}_{k,h_k} = [\mathcal{B}_{k,h_k,1}', \dots, \mathcal{B}_{k,h_k,h_k}']'$*, with* $\mathcal{B}_{k,h_k,j} \in \mathbb{C}^{(d_j^k - d_{j-1}^k) \times s}$ *it holds that* $\mathcal{B}_{k,h_k,j}$ *is p.u.t. for* $d_j^k > d_{j-1}^k$ *for* $j = 1, \dots, h_k$ *and* $k = 1, \dots, l$*.*

(iv) For $k = 1, \dots, l$ *define* $\mathcal{C}_k = [\mathcal{C}_{k,1}, \mathcal{C}_{k,2}, \dots, \mathcal{C}_{k,h_k}], \mathcal{C}_{k,j} = [\mathcal{C}_{k,j}^G, \mathcal{C}_{k,j}^E]$*, with* $\mathcal{C}_{k,j}^E \in \mathbb{C}^{s \times (d_j^k - d_{j-1}^k)}$ *and* $\mathcal{C}_{k,j}^G \in \mathbb{C}^{s \times d_{j-1}^k}$ *for* $j = 1, \dots, h_k$*, with* $d_0^k := 0$*. Furthermore, define* $\mathcal{C}_k^E := [\mathcal{C}_{k,1}^E, \dots, \mathcal{C}_{k,h_k}^E] \in \mathbb{C}^{s \times d_{h_k}^k}$*. It holds that* $(\mathcal{C}_k^E)' \mathcal{C}_k^E = I_{d_{h_k}^k}$ *and* $(\mathcal{C}_{k,j}^G)' \mathcal{C}_{k,i}^E = 0$ *for* $1 \leq i \leq j$ *for* $j = 2, \dots, h_k$ *and* $k = 1, \dots, l$*.*

(v) $\lambda_{|max|}(\mathcal{A}_\bullet) < 1$ *and the* stable subsystem $(\mathcal{A}_\bullet, \mathcal{B}_\bullet, \mathcal{C}_\bullet)$ *of state dimension* $n_\bullet = n - n_u$ *is in echelon canonical form, cf. Hannan and Deistler (1988, Theorem 2.5.2).*

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

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

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

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

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:*

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

with

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

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

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

with

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

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

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* $\{x_{t,\mathbb{R}}\}_{t\in\mathbb{Z}}$, $x_{t,\mathbb{R}} := [x'_{t,u,\mathbb{R}}, x'_{t,\bullet}]' := [x'_{t,1,\mathbb{R}}, \dots, x'_{t,l,\mathbb{R}}, x'_{t,\bullet}]'$, *obtained from*

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

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 $k = 1, \dots, l$ *the sub-vectors* $x_{t,k,\mathbb{R}}$ *are further decomposed into* $x_{t,k,\mathbb{R}} := [(x_{t,k,\mathbb{R}}^1)', \dots, (x_{t,k,\mathbb{R}}^{h_k})']'$, *with* $x_{t,k,\mathbb{R}}^j \in \mathbb{R}^{2d_j^k}$ *if* $0 < \omega_k < \pi$ *and* $x_{t,k,\mathbb{R}}^j \in \mathbb{R}^{d_j^k}$ *if* $\omega_k \in \{0, \pi\}$ *for* $j = 1, \dots, h_k$ *and* $\mathcal{C}_{k,\mathbb{R}} := [\mathcal{C}_{k,1,\mathbb{R}}, \dots, \mathcal{C}_{k,h_k,\mathbb{R}}]$ *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* $j = 1, \dots, h_k$ *and* $k = 1, \dots, l$. *Furthermore, for* $j = 1, \dots, h_k$ *and* $k = 1, \dots, l$ *the processes* $\{x_{t,k,\mathbb{R}}^j\}_{t\in\mathbb{Z}}$ *are not cointegrated, as defined in Definition 3 below. For* $\omega_k = 0$, *the process* $\{x_{t,k,\mathbb{R}}^j\}_{t\in\mathbb{Z}}$ *is the* d_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 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 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"* $\mathcal{C}_u \mathcal{A}_u^t x_{1,u}$, *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 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 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 $\tilde{h}_{\tilde{k}} \leq h_k$.*

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

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

Remark 7 *(i) It is merely a matter of taste whether cointegrating spaces are defined in terms of their order* $(\Omega, \tilde{\Omega})$ *or their decrease* $\delta(\Omega, \tilde{\Omega}) := (\delta_1(\Omega, \tilde{\Omega}), \dots, \delta_l(\Omega, \tilde{\Omega}))$*, with* $\delta_k(\Omega, \tilde{\Omega})$ *as defined above. Specifying* Ω *and* $\delta(\Omega, \tilde{\Omega})$ *contains the same information as providing the order of (polynomial) cointegration.*

(ii) Notwithstanding the fact that CIVs and PCIVs in general may lead to changes of the integration orders at different unit root frequencies it may be of interest to "zoom in" on only one unit root frequency ω_k*, thereby leaving the potential reductions of the integration orders at other unit root frequencies unspecified. This allows to – entirely similarly as in Definition 3 – define cointegrating and polynomial cointegrating spaces of different orders at a single unit root frequency* ω_k*. Analogously one can also define cointegrating and polynomial cointegrating spaces of different orders for subsets of the frequencies in* $F(\Omega)$*.*

(iii) In principle the polynomial cointegrating spaces defined so far are infinite-dimensional as the polynomial degree is not bounded. However, since every polynomial vector $\beta(z)$ *can be written as* $\beta_0(z) + \beta_\Omega(z)\Delta_\Omega(z)$*, where by definition* $\{\Delta_\Omega y_t\}_{t\in\mathbb{Z}}$ *has empty unit root structure, it suffices to consider PCIVs of polynomial degree smaller than the polynomial degree of* $\Delta_\Omega(z)$*. This shows that it is sufficient to consider finite dimensional polynomial cointegrating spaces. When considering, as in item (ii), (polynomial) cointegration only for one unit root it similarly suffices to consider polynomials of maximal degree equal to* $h_k - 1$ *for real unit roots and* $2h_k - 1$ *for complex unit roots. Thus, in the I(2) case it suffices to consider polynomials of degree one.*

(iv) The argument about maximal relevant polynomial degrees given in item (iii) can be made more precise and combined with the decrease in Ω *achieved. Every polynomial vector* $\beta(z)$ *can be written as* $\beta_0(z) + \beta_{\omega_k,\delta_k}(z)\Delta_{\omega_k}^{\delta_k}(z)$ *for* $\delta_k = 1, \dots, h_k$*. By definition it holds that* $\{\Delta_{\omega_k}^{\delta_k} y_t\}_{t\in\mathbb{Z}}$ *has integration order* $h_k - \delta_k$ *at frequency* ω_k*. Thus, it suffices to consider PCIVs of polynomial degree smaller than* δ_k *for* $\omega_k \in \{0, \pi\}$ *or* $2\delta_k$ *for* $0 < \omega_k < \pi$ *when considering the polynomial cointegrating space at* ω_k *with decrease* δ_k*. In the MFI(1) case therefore, when considering only one unit root frequency, again only polynomials of degree one need to be considered. This space is often referred to in the literature as* dynamic *cointegration space.*

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

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

By Remark 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 $\beta'[\mathcal{C}_{k,1,\mathbb{R}}, \dots, \mathcal{C}_{k,j,\mathbb{R}}] = 0$ and $\beta'\mathcal{C}_{k,j+1,\mathbb{R}} \neq 0$ or equivalently $\beta'[\mathcal{C}_{k,1}, \dots, \mathcal{C}_{k,j}] = 0$ and $\beta'\mathcal{C}_{k,j+1} \neq 0$ (using the transformation to the complex matrices of the canonical form, as discussed in Remark 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 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 $[\begin{array}{cc} \beta_0' & \beta_1' \end{array}]'$ 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 $\max_{k=1,\dots,l}\|\beta(e^{i\omega_k})\|\delta_k(\Omega, \tilde{\Omega}) \neq 0$. 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 (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 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

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

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 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*

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

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 degreee 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 4 *A unit root process $\{y_t\}_{t\in\mathbb{Z}}$ with a canonical state space representation as given in Theorem 1 has* state space unit root structure

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

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

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

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

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

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

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

Example 4 *Consider the following state space system:*

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

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

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

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 2.4) 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 6 (Real Givens rotation) *The* real Givens rotation $R_{q,i,j}(\theta) \in \mathbb{R}^{q\times q}$, $\theta \in [0, 2\pi)$ *is defined as*

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

Remark 10 *Givens rotations allow 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 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 7 (Complex Givens rotation) *The* complex Givens rotation $Q_{q,i,j}(\varphi) \in \mathbb{C}^{q\times q}$, $\varphi := [\varphi_1, \varphi_2]' \in \Theta_{\mathbb{C}} := [0, \pi/2] \times [0, 2\pi)$, *is defined as*

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

Remark 12 *Complex Givens rotations allow 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*

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

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

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

The parameterization of $O_{s,d}$ is based on the combination of real Givens rotations, as given in Definition 6, that allow 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 $I_s^- := \mathrm{diag}(I_{s-1}, -1)$, since, compare Remark 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 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 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 demonstrates the usefulness of this parameterization.

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

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

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

$$\mathcal{C} = C_O(\boldsymbol{\theta}) = R_L(\boldsymbol{\theta}_L)' \left[\begin{array}{c} I_d \\ 0_{(s-d)\times d} \end{array}\right] R_R(\boldsymbol{\theta}_R).$$

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

(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 $\boldsymbol{\theta} \in \Theta_O^{\mathbb{R}}$ such that

$$\mathcal{C} = C_O(\boldsymbol{\theta}) = R_L(\boldsymbol{\theta}_L)' \left[\begin{array}{c} I_d \end{array}\right] R_R(\boldsymbol{\theta}_R) = R_R(\boldsymbol{\theta}_R).$$

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

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

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

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

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

As in the parametrization of the VAR I(1)-case in the VECM framework, where the restriction $\beta = [I_{s-d}, \beta^*]'$ can only be imposed when the upper $(s-d) \times (s-d)$ block of the true β_0 of the DGP is of full rank,

cf. Johansen (1995, 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 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 5 *Consider the matrix*

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

with $d = 2$ and $s = 3$. As discussed, the static cointegrating space is characterized by the left kernel of this matrix. The left kernel of a matrix in $\mathbb{R}^{3\times 2}$ with full rank two is given by a one-dimensional space, with the corresponding basis vector parameterized, when normalized to length one, by two free parameters. Thus, for the characterization of the static cointegrating space two parameters are required, which exactly coincides with the dimension of $\boldsymbol{\theta}_L$ given in Remark 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 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 $[\ 0 \quad \frac{1}{\sqrt{2}}\ ] R_{2,1,2}(\theta_{R,1})' = [\ r \quad 0\ ]$ for $r \geq 0$ and $\theta_{R,1} \in [0, 2\pi)$. In other words, find $r \geq 0$ and $\theta_{R,1} \in [0, 2\pi)$ such that $[\ 0 \quad \frac{1}{\sqrt{2}}\ ] = r[\ \cos(\theta_{R,1}) \quad \sin(\theta_{R,1})\ ]$, which leads to $r = \frac{1}{\sqrt{2}}$, $\theta_{R,1} = \frac{\pi}{2}$. Thus, the orthonormal matrix $R_R(\boldsymbol{\theta}_R)$ 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 $R_{3,1,3}(\theta_{L,1})[\ \frac{1}{\sqrt{2}}\ \ 0\ \ \frac{1}{\sqrt{2}}\ ]' = [\ 1\ \ 0\ \ 0\ ]'$, i. e., $[\ \frac{1}{\sqrt{2}}\ \ \frac{1}{\sqrt{2}}\ ]' = r[\ \cos(\theta_{L,1})\ \ \sin(\theta_{L,1})\ ]$. 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 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 $[c_{j,j}, \dots, c_{j,d}]$ in the j-th row of $\mathcal{C}^{(j)}$, to $[\tilde{c}_{j,j}, 0, \dots, 0]$. 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 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 $[c_{d-j,d-j}, c_{d+1,d-j}, \dots, c_{s,d-j}]'$. Using the algorithm described in Remark 12 transform this vector to $[\tilde{c}_{d-j,d-j}, 0, \dots, 0]'$ by left-multiplication of $\tilde{\mathcal{C}}^{(j)}$ 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 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*

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

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 15 *Note that using 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 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 $[\begin{array}{cc} \frac{1-i}{2} & \frac{1-i}{2} \end{array}]Q_{2,1,2}(\varphi_{R,1})' = [\begin{array}{cc} re^{i\varphi_a} & 0 \end{array}]$ with $r > 0$, transform the entries of $[\begin{array}{cc} \frac{1-i}{2} & \frac{1-i}{2} \end{array}]'$ into polar coordinates. The equation $[\begin{array}{cc} \frac{1-i}{2} & \frac{1-i}{2} \end{array}]' = [\begin{array}{cc} ae^{i\varphi_a} & be^{i\varphi_b} \end{array}]'$ has the solutions $a = b = \frac{1}{\sqrt{2}}$ and $\varphi_a = \varphi_b = \frac{7\pi}{4}$. Using the results of Remark 12, the parameters of the Givens rotation are $\varphi_{R,1,1} = tan^{-1}(\frac{b}{a}) = \frac{\pi}{4}$ and $\varphi_{R,1,2} = \varphi_a - \varphi_b = 0$. Right-multiplication of C with $Q_{2,1,2}\left(\left[\frac{\pi}{4}, 0\right]\right)'$ 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 general case. The dimensions of the parameter vectors of the respective blocks of the system matrices $(\mathcal{A}, \mathcal{B}, \mathcal{C})$ depend on the multi-index Γ, consisting of the state space unit root structure Ω_S, the structure indices p and the Kronecker indices $\alpha_\bullet$ 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 and Lemma 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 7 *Consider an MFI(1) process with* $\Omega_S = ((0, 2), (\frac{\pi}{2}, 2))$, $p = [1, 3, 1, 2, 1, 2]'$, $n_\bullet = 0$, *and system matrices*

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

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

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 5 and 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 $\Omega_s = ((0, d^1_1, d^1_2))$:

$$\mathcal{A} = \begin{bmatrix} I_{d^1_1} & I_{d^1_1} & 0 & 0 \\ 0 & I_{d^1_1} & 0 & 0 \\ 0 & 0 & I_{d^1_2 - d^1_1} & 0 \\ 0 & 0 & 0 & \mathcal{A}_\bullet \end{bmatrix}, \quad \mathcal{B} = \begin{bmatrix} \mathcal{B}_{1,1} \\ \mathcal{B}_{1,2,1} \\ \mathcal{B}_{1,2,2} \\ \mathcal{B}_\bullet \end{bmatrix}, \quad \mathcal{C} = \left[\begin{array}{cccc} \mathcal{C}^E_{1,1} & \mathcal{C}^G_{1,2} & \mathcal{C}^E_{1,2} & \mathcal{C}_\bullet \end{array} \right],$$

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 $(\mathcal{A}_\bullet, \mathcal{B}_\bullet, \mathcal{C}_\bullet)$ 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

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

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.

Corollary 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*

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

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. 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 $\mathcal{C}_{1,2}^E \in O_{s,d_2^1-d_1^1}(\mathcal{C}_{1,1}^E)$ there exists a vector $\tilde{\boldsymbol{\theta}} = [\tilde{\boldsymbol{\theta}}'_L, \tilde{\boldsymbol{\theta}}'_R]' \in \Theta_{O,d_2^1-d_1^1}^{\mathbb{R}}$ such that

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

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

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

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

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

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

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

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

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

Steps 1-4 of the algorithm discussed above Lemma 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 $v := [\pi, \dots, \pi]' \in \mathbb{R}^{(s-d_1^1)(s-d_1^1-1)/2}$. *Then a parameterization of* $O_{s,s-d_1^1}(\mathcal{C}^E_{1,1})$ *is given by*

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

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 uses the same arguments as the proof of Lemma 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

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

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 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 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 $O_{s,G}(\mathcal{C}^E_{1,1}, \mathcal{C}^E_{1,2})$ and thus a useful parameterization, since the matrix $\mathcal{C}^G_{1,2}$ is only multiplied with two constant invertible matrices. The entries of the matrix product are then collected in a parameter vector as shown in Corollary 2.

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

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

where $\boldsymbol{\theta}_L$ *denotes the parameter values corresponding to* $[\boldsymbol{\theta}'_L, \boldsymbol{\theta}'_R]' = C_O^{-1}(\mathcal{C}^E_{1,1})$ *as defined in Lemma 1 and* $\tilde{\boldsymbol{\theta}}_L$ *denotes the parameter values corresponding to* $[\tilde{\boldsymbol{\theta}}'_L, \tilde{\boldsymbol{\theta}}'_R]' = C^{-1}_{O,d_2^1-d_1^1}(\mathcal{C}^E_{1,2}; \mathcal{C}^E_{1,1})$ *as defined in Corollary 1. The set* $O_{s,G}(\mathcal{C}^E_{1,1}, \mathcal{C}^E_{1,2})$ *is closed and both* $C_{O,G}$ *as well as* $C^{-1}_{O,G}(\cdot)$ *- the parameterization of* $O_{s,G}(\mathcal{C}^E_{1,1}, \mathcal{C}^E_{1,2})$ *- 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 $\boldsymbol{\theta}_{B,p,1} := \left[b^1_{d^1-d^1_{h_1}+1, p^1_{d^1-d^1_{h_1}+1}}, \ldots, b^1_{d^1_1, p^1_{d^1_1}} \right]'$ 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 $\mathcal{C}^E_{1,1}$ as in the MFI(1) case and $\mathcal{C}^E_{1,2}$ as discussed in Corollary 1.
- $\boldsymbol{\theta}_{C,G} \in \Theta_{C,G} = \mathbb{R}^{d_{C,G}}$: The parameters for the matrix $\mathcal{C}^G_{1,2}$ as discussed in Corollary 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 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*

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

that $\boldsymbol{\theta}_{C,E} = [\boldsymbol{\theta}'_{C,E,1,1}, \boldsymbol{\theta}_{C,E,1,2}]' = \left[\left[\frac{\pi}{2}, \frac{7\pi}{4}\right], \left[\frac{7\pi}{4}\right]\right]'$ *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 and 2 and Corollaries 1 and 2 above. The extension of Corollaries 1 and 2 to complex matrices and to matrices which are orthogonal to a larger number of blocks of $\mathcal{C}_k$ 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 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 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*

  $$\boldsymbol{\theta}_{C,E} = [\boldsymbol{\theta}'_{C,E,1}, ..., \boldsymbol{\theta}'_{C,E,l}]' \in \Theta_{C,E} \subset \mathbb{R}^{d_{C,E}}, \quad \boldsymbol{\theta}_{C,E,k} = [\boldsymbol{\theta}'_{C,E,k,1}, \ldots, \boldsymbol{\theta}'_{C,E,k,h_k}]'$$

  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 and 2 and Corollary 1 and its extension to complex matrices).*

- *The parameter vector*

  $$\boldsymbol{\theta}_{C,G} = [\boldsymbol{\theta}'_{C,G,1}, ..., \boldsymbol{\theta}'_{C,G,l}]' \in \Theta_{C,G} = \mathbb{R}^{d_{C,G}}, \quad \boldsymbol{\theta}_{C,G,k} = [\boldsymbol{\theta}'_{C,G,k,2}, \ldots, \boldsymbol{\theta}'_{C,G,k,h_k}]'$$

  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 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 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 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 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 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 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

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

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.

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 is composed of blocks $\mathcal{C}^E_k$ that are sub-blocks of unitary (or orthonormal) matrices and blocks $\mathcal{C}_k^G$ that have to fulfill (recursive) orthogonality constraints. The corresponding sets have been shown to be closed in Lemmas 1 and 2 and Corollaries 1 and 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 sub-set of $\mathbb{R}^m$ for some m. For convergent sequences of systems with fixed Ω_S and p, limits of entries restricted to be positive may be zero. When this happens, two cases have to be distinguished. First, all p.u.t. sub-matrices still have full row rank. In this case the limiting system, $(\mathcal{A}_0, \mathcal{B}_0, \mathcal{C}_0)$ say, is still minimal and can be transformed to a system in canonical form $(\tilde{\mathcal{A}}_0, \tilde{\mathcal{B}}_0, \tilde{\mathcal{C}}_0)$ with *fewer* unrestricted entries in $\tilde{\mathcal{B}}_0$.

Second, if at least one of the row ranks of the p.u.t. blocks decreases in the limit, the limiting system is no longer minimal. Consequently, $(\tilde{\Omega}_S, \tilde{p}) < (\Omega_S, p)$ in the limit.
To illustrate this point consider again Example 4 with equation (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

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

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

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

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 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 9. The precise statement is given in Theorem 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 $\tilde{A} = \text{diag}(\tilde{J}_2, \tilde{A}_\bullet)$, where all eigenvalues of $\tilde{J}_2$ have unit modulus and $\lambda_{|\max|}(\tilde{A}_\bullet) < 1$.

The first case is well understood, compare Hannan and Deistler (1988, 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 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 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 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 3 provides a characterization of the transfer functions corresponding to systems in the closure of Δ_Γ. The conformable set $\mathcal{K}(\Gamma)$ plays a key role here, since it characterizes the set of all minimal systems that can be obtained as limits of convergent sequences from within the set

Δ_Γ. Conformable indices extend the matrix $\mathcal{A}_u$ corresponding to the unit root structure by the block $\tilde{J}_2$.

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

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

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 3 implies the existence of a generic neighborhood for the canonical form considered in this paper:

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

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

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

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

The closure of $M(\Omega_S, n_\bullet)$ contains also transfer functions that have a different state space unit root structure than Ω_S. Considering convergent sequences of state space realizations $(A_j, B_j, C_j)_{j\in\mathbb{N}}$ of transfer functions in $M(\Omega_S, n_\bullet)$, the state space unit root structure of

$$(A_0, B_0, C_0) := \lim_{j\to\infty}(A_j, B_j, C_j)$$

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 9 (ii). The implications of the other two cases are summarized in the following definition:

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

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

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

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

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

Finally, note that Theorem 5 also implies the following result relevant for the determination of the unit root structure, further discussed in Sections 1.5.1 and 1.5.2:

Corollary 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 (co-) integrated processes for quarterly data with state space unit root structure $((0, d_1^1), (\pi/2, d_1^2), (\pi, d_1^3))$. In general, assuming for notational simplicity $\omega_1 = 0$ and $\omega_l = \pi$, it holds that for $t > 0$ and $x_{1,u} = 0$

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

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 from considering the real valued canonical form discussed in Remark 4, with the matrices $\mathcal{A}_{k,\mathbb{R}}$ for $k = 2, \dots, l-1$, 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 $\Omega_S = ((\omega_1, d_1^1), \dots, (\omega_l, d_1^l))$. The number of stochastic trends is equal to d_1^1, the number of stochastic cycles at frequency ω_k is equal to $2d_1^k$ for $k = 2, \dots, l-1$ and equal to d_1^l if $k = l$, as discussed in Section 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 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 2.

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

Testing Hypotheses on the State Space Unit Root Structure

Using the generic sets of transfer functions M_{Γ_g} presented in Theorem 4, we can construct pseudo likelihood ratio tests for different hypotheses $H_0 : (\Omega_S, n_\bullet) = (\Omega_{S,0}, n_{\bullet,0})$ against chosen alternatives. Note, however, that by the results of Theorem 5 the null hypothesis includes all pairs $(\Omega_S, n_\bullet) \in \mathcal{A}(\Omega_{S,0}, n_{\bullet,0})$ as well as all pairs $(\Omega_S, n_\bullet)$ 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 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 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 different cases correspond to different types of hypotheses related to restrictions implied by economic theory.

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

(ii) $H_0' : \beta = [b, \varphi], \beta \in \mathbb{R}^{s\times r}, b \in \mathbb{R}^{s\times t}, \varphi \in \mathbb{R}^{s\times r-t}, 0 < t \leq r$: Some cointegrating relations are known.

(iii) $H_0^{''} : \beta = [H_1\varphi_1, \ldots, H_c\varphi_c], \beta \in \mathbb{R}^{s\times r}, H_j \in \mathbb{R}^{s\times t_j}, \varphi_j \in \mathbb{R}^{t_j\times r_j}, r_j \leq t_j \leq s$, 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, 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. From this one can derive a parameterization of the set of matrices $\mathcal{C}_1^r$ corresponding to H_0, analogously to Lemma 1. The difference of the number of free parameters under the null hypothesis and under the alternative is the difference between the number of free parameters in $\boldsymbol{\theta}_L \in [0, 2\pi)^{r(s-r)}$ 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 $R_R(\boldsymbol{\theta}_{R,\beta}) \in \mathbb{R}^{r\times r}$ as in Lemma 1 and

$$R_H(\boldsymbol{\theta}_H) := R_H\left((\boldsymbol{\theta}_{H_1}, \boldsymbol{\theta}_{H_2}, \boldsymbol{\theta}_{H_b})\right) := R_{H_1}(\boldsymbol{\theta}_{H_1}) R_{H_2}(\boldsymbol{\theta}_{H_2}) R_{H_b}(\boldsymbol{\theta}_{H_b}) \in \mathbb{R}^{\tilde{t}\times\tilde{t}}$$

is a product of Givens rotations corresponding to the entries in the blocks highlighted by bold font. The three matrices are defined as follows:

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

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

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

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. 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 $r_1 \le \tilde{t}_1$ and $r_2 \le \tilde{t}_2$.

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

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

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

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

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

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

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

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 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 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 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:

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. $H_2(r) \;:\; \Phi d_t = \mathcal{C}_1\tilde{\Phi}_0,$

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 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 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 $x_{0,u}$: The solution to the state space equations for $t > 0$ and $x_{1,u} \neq 0$ is given by:

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

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

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

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 $O_{s,d_2^1-d_1^1}([b, \mathcal{C}_{1,1}^E])$. 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., $\mathcal{C}_{1,1}^E$. Implementation of different types of hypotheses on $[\beta, \beta_1]$ proceeds as for similar hypotheses on β in the MFI(1) case at unit root one, replacing $[\mathcal{C}_{1,1}^E, \mathcal{C}_{1,2}^E]$ by $\mathcal{C}_{1,1}^E$.

The hypothesis of no minimum degree polynomial cointegrating relations implies the restriction $\mathcal{C}_{1,2}^G = 0$, compare Example 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

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

where the columns of $\tilde{\mathcal{C}}_{1,2}$ are a basis for the column space of $\mathcal{C}_{1,2}^G$, which does not necessarily have full column rank, and the columns of $\mathcal{C}_\perp$ span the orthocomplement of the column space of $[\mathcal{C}_{1,1}^E, \mathcal{C}_{1,2}^E, \tilde{\mathcal{C}}_{1,2}]$. 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) provides the very first step of this agenda by providing a *canonical form* for unit root processes in the state space framework, which is shown in that paper to be very convenient for cointegration analysis.

Based on the earlier canonical form paper this paper derives a state space model *parameterization* for VARMA processes with unit roots using the state space framework. The canonical form and a fortiori the parameterization based 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.

Chapter 2

Inference on Cointegrating Ranks and Spaces of Multiple Frequency I(1) Processes: A State Space Approach

2.1 Introduction

For empirical macroeconomic research two of the most important toolkits are dynamic stochastic general equilibrium (DSGE) models and cointegrated vector autoregressive (VAR) models.
The DSGE approach is firmly rooted in economic theory and mathematically consists of solving dynamic stochastic optimization problems with well-defined objective functions of the different agents (i.e., individuals, firms, the government, the central bank) as well as resource constraints. Typically this leads to a state space system and solutions of these systems of equations are in general vector autoregressive moving average (VARMA) processes, see, e. g., Campbell (1994).
The cointegrated VAR approach of Johansen on the other hand is a well developed method to analyze long-run relationships between economic time series, see, e. g., Johansen (1996). In this context one has to consider the different asymptotic properties of standard estimators, e. g., the OLS estimator, caused by the presence of stochastic trends. To gain insight on economic information one typically investigates the number of stochastic trends present in the data, which is even more of a challenge if seasonal dependencies in the data come into play.
Hylleberg, Engle, Granger and Yoo (1990) first proposed the analysis of seasonal (co-)integration. Different papers confirm the presence of seasonal unit roots in common macroeconomic time series, compare, e. g., Hylleberg, Jorgensen and Sorensen (1993). Although a commonly used approach when dealing with seasonal data is to perform seasonal adjustment, studies such as Lof and Franses (2001) suggest that this practice leads to a loss of valuable information, such that for instance the forecasting performance tends to be inferior to forecasts from seasonal cointegration models. To accomodate such seasonal cointegration properties, Johansen and Schaumburg (1999) extend the vector error correction model from I(1) processes to models for seasonal cointegration.
However, the VAR framework is not as flexible as the VARMA setting. Therefore, it is a natural question whether the procedures already developed for cointegration analysis in VAR models can be adapted to (pseudo) maximum likelihood estimation in a VARMA model or the – in a certain sense – equivalent state space framework. Apart from economic interest in handling DSGE models, one important motive for such a generalization is the fundamentalness of VARMA models. Since subsets of variables jointly characterized as a VAR process are in general VARMA processes, see, e. g., Zellner and Palm (1974), a state space framework appears to be the more appropriate choice in some situations. A further advantage lies in the identification of structural shocks, which cannot

necessarily be recovered from VAR models, but can be analyzed employing state space models. Finally the state space framework allows for a factor-like structure and potentially reduces the number of parameters in particular for processes with moderate or large dimension.
A recent development in the analysis of cointegrated state space systems was the introduction of the state space error correction model by Ribarits and Hanzon (2014), mimicking the error correction formulation of VAR processes. Earlier approaches by Lütkepohl and Claessen (1997) and Poskitt (2003) formulated the error correction model in the VARMA setting.
Bauer and Wagner (2012) examined state space systems involving unit roots, introducing a canonical form which highlights the cointegrating properties of the corresponding solution processes. Bauer, Matuschek, de Matos Ribeiro and Wagner (2020, Section 2) extend the discussion by an extensive comparison between VARMA and VAR cointegration models. They propose parameterizations which also include restrictions corresponding to different types of hypotheses and illustrate subset relations between sets with different so-called state space unit root structure, thus, setting a basis for applying likelihood ratio-type tests.
In the companion paper de Matos Ribeiro, Bauer, Matuschek and Wagner (2020) the authors derive the consistency of the pseudo maximum likelihood estimator for state space MFI(1) systems and provide the asymptotic distribution of the suitably standardized parameters, laying the ground work for statistical inference.
In this paper we combine these results to develop tests for hypotheses on cointegrating ranks at the unit roots and linear hypotheses on cointegrating relations by employment of Johansen-style (pseudo) likelihood ratio tests. It turns out that a suitable extension of the Ribarits and Hanzon (2014) error correction formulation to the MFI(1) situation can be used to obtain analogous results for state space processes as the ones presented by Johansen (1996) for I(1) and by Johansen and Schaumburg (1999) for MFI(1) VAR processes. The underlying idea in this respect is the formulation of the error correction representation which for state space processes takes the same form as in the VAR framework except for a different inclusion of the stationary regressors: In the VAR setting lags of (seasonal) differences are used while in the state space framework filtered versions of the differences take the same role. We link these representations to the pseudo likelihood ratio tests in the VARMA setting and show that going from VAR to VARMA models the traditional tools from the Johansen framework are applicable in the investigation. We especially focus on the treatment of complex unit roots, observing that simulation results indicate advantages over the standard VAR approach in seasonal data. The treatment of deterministics is stated only briefly for the sake of completeness.
This paper is structured as follows: In Section 2 the data generating process dealt with in this paper is investigated. This also includes a discussion of the parameterization used and the structure of the state space representation for unit root processes. In Section 3 the state space error correction model by Ribarits and Hanzon (2014) is used to develop tests on the cointegrating rank and the cointegrating space in the I(1) case. This contains already all the complexity needed for the MFI(1) case which is dealt with in Section 4. The rank tests are compared to those by Johansen and Schaumburg (1999) and the CCA subspace tests by Bauer and Buschmeier (2016) in a small simulation study in Section 5. Section 6 summarizes and concludes this paper. All proofs are relegated to the appendix.
Notation in the paper is as follows: 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 square matrix X we denote the spectral radius (i.e., the maximum of the modulus of its eigenvalues) by $\lambda_{|\max|}(X)$. With $\mathcal{R}(M)$ we denote the real part of a complex matrix $M \in \mathbb{C}^{k\times l}$ and with $\mathcal{I}(M)$ its imaginary part. For a $m \times n$ matrix X of full rank, with $n < m$, $X_\perp$ denotes an $m \times (m-n)$ matrix of full rank such that $X'X_\perp = 0$. For a matrix or vector X the matrix or vector with complex conjugated entries is denoted by $\overline{X}$. For finite sequences $\{a_t\}_{t=1,\dots,T}, \{b_t\}_{t=1,\dots,T}$, we define $\langle a_t, b_t\rangle := T^{-1}\sum_{t=1}^T a_t b_t'$. Convergence in distribution is denoted by $\stackrel{d}{\to}$ and convergence in probability by $\stackrel{p}{\to}$.

2.2 Definitions

2.2.1 Unit Root Processes and Cointegration

Most of the literature regarding integrated processes consider the so-called *vector error correction model* (VECM) representation of autoregressive processes, discussed for processes integrated of order one and two in full detail in the monograph Johansen (1996). Bauer et al. (2020, Section 2) contains a detailed discussion on the necessity of the VECM for a suitable parameterization of integrated processes in the VAR framework, again exemplifying the arguments for integrated processes of order one and two only. The same arguments also hold for seasonally integrated VAR processes.
In this paper we use the following differencing operators:

$$\begin{aligned} \mathbf{\Delta}_S &:= 1 - L^S \\ \mathbf{\Delta}_{S,k} &:= \frac{\bar{z}_k}{\prod_{j\neq k}(1-\bar{z}_j z_k)} \prod_{j \neq k} (1 - \bar{z}_j L) \end{aligned}$$

where

$$S \in \mathbb{N}, \quad z_k := \exp(i\omega_k), \quad \omega_k := \frac{k-1}{S} 2\pi \quad \text{for } k = 1, \dots, S.$$

For notational brevity, we omit the dependence on L in $\mathbf{\Delta}_S(L)$ and $\mathbf{\Delta}_{S,k}(L)$. Using this notation, the error correction representation of a seasonally integrated VAR process of order p is given by

$$\begin{aligned} \mathbf{\Delta}_S y_t &= \sum_{k=1}^{S} \Pi_k \mathbf{\Delta}_{S,k} y_{t-1} + \sum_{j=1}^{p-S} \Gamma_j \mathbf{\Delta}_S y_{t-j} + \varepsilon_t \qquad (2.1) \\ &= \sum_{k=1}^{S} \alpha_k \beta_k' \mathbf{\Delta}_{S,k} y_{t-1} + \sum_{j=1}^{p-S} \Gamma_j \mathbf{\Delta}_S y_{t-j} + \varepsilon_t, \end{aligned}$$

where $\Pi_1 \in \mathbb{R}^{s\times s}$, $\Pi_k \in \mathbb{C}^{s\times s}$ for $k = 2, \dots, S$, $\Gamma_j \in \mathbb{R}^{s\times s}$ for $j = 1, \dots, p-S$ and $\alpha_k, \beta_k \in \mathbb{C}^{s\times r_k}$ of full (column) rank r_k, for $k = 1, \dots, S$. As additional assumptions, impose $\Pi_k = \overline{\Pi}_{S+2-k}$ for all $k = 2, \dots, S$, such that the imaginary part of the right hand side of equation (2.1) is zero.[1] The integer r_k, equal to the rank of Π_k, is called *cointegrating rank* at frequency ω_k. Clearly the VECM can be transformed to a VAR(p) representation of the form

$$y_t + \sum_{j=1}^{p} a_i y_{t-i} = \varepsilon_t \qquad t \in \mathbb{Z},$$

with coefficients $a_j \in \mathbb{R}^{s\times s}$ dependent on Π_k, $k = 1, \dots, S$ and Γ_j, $j = 1, \dots, p-S$. Conversely it holds that $\Pi_k = -a(z_k)$ for $k = 1, \dots, S$, with $a(z) = I_s + \sum_{j=1}^{p} a_j z^j$.
Analogous to Bauer et al. (2020), for given restrictions on some of the ranks r_k, namely the ranks $\tilde{r}_j$ of $a(e^{i\tilde{\omega}_j})$, for $0 \leq \tilde{\omega}_1 < \dots < \tilde{\omega}_l \leq \pi$, we formally define different sets of transfer functions $a^{-1}(z)$ by

$$V(p, \{(\tilde{\omega}_1, \tilde{r}_1), \dots, (\tilde{\omega}_l, \tilde{r}_l)\}) := \left\{ a^{-1}(z) : \begin{array}{c} a(z) = I_s + \sum_{j=1}^{p} a_j z^j, a_p \neq 0, \\ \text{rank}(a(e^{i\tilde{\omega}_j})) \leq \tilde{r}_j < s, \text{ for } j = 1, \dots, l \end{array} \right\}.$$

Having established useful notation for the VECM we now proceed to the discussion of integrated VARMA processes. We refer to a process $\{y_t\}_{t\in\mathbb{Z}}$ as a VARMA process if we have integers p and

[1] Moreover, let $(\alpha_k)'_\perp (I_s - \sum_{l\neq k}^{S} \Pi_l \Delta_{S,l,k} z_k - \sum_{j=1}^{p-S} \Gamma_j \Delta_{S,k} z_k^j)(\beta_k)_\perp$ be of full rank, where $\Delta_{S,l,k} := \bar{z}_l (1 - \bar{z}_k z_l)^{-1} \prod_{j\neq k,l} (1 - \bar{z}_j z_l)^{-1} (1 - \bar{z}_j z_k)$ and $\Delta_{S,k} := \prod_{j \neq k} (1 - \bar{z}_j z_k)$. This ensures that $\{y_t\}_{t\in\mathbb{Z}}$ is not of higher integration order, compare again the discussion of integrated process of order one and two in Bauer et al. (2020, Section 2).

q, matrices $a_j \in \mathbb{R}^{s\times s}$, $j = 1,\dots,p$, $b_j \in \mathbb{R}^{s\times s}$, $j = 1,\dots,q$, $D \in \mathbb{R}^{s\times m}$, a deterministic process $\{s_t\}_{t\in\mathbb{Z}}$, $s_t \in \mathbb{R}^m$ and a white noise process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$, $\varepsilon_t \in \mathbb{R}^s$ with $\mathbb{E}(\varepsilon_t\varepsilon_t') = \Sigma > 0$ such that

$$y_t - Ds_t + \sum_{j=1}^{p} a_j(y_{t-j} - Ds_{t-j}) = \varepsilon_t + \sum_{j=1}^{q} b_j\varepsilon_{t-j}, \qquad t \in \mathbb{Z}. \tag{2.2}$$

Note that by this definition VARMA processes are not required to be stationary.
The pair $(a(z), b(z))$, where $a(z) = I_s + \sum_{j=1}^{p} a_j z^j$ and $b(z) = I_s + \sum_{j=1}^{q} b_j z^j$ are matrix polynomials, is called a VARMA system corresponding to the process $\{y_t\}_{t\in\mathbb{Z}}$. The function $k(z) := a(z)^{-1}b(z)$ is called the transfer function of the process.
It is well known – see, for instance, Hannan and Deistler (1988, Chapter 1) – that for given matrix polynomials $(a(z), b(z))$, (2.2) always has a solution. If $\det(a(z)) \neq 0$ for $|z| \leq 1$ the transfer function $k(z)$ has a convergent power series expansion in a disk containing the closed unit disk. In this case (2.2) has a unique stationary and causal solution $\tilde{y}_t = y_t - Ds_t = \sum_{j=0}^{\infty} K_j\varepsilon_{t-j}$ where K_j are the coefficients from the power series expansion of $k(z)$, compare, e. g., Hannan and Deistler (1988, pp. 9-11).
Defining the operator Δ_ω, we ar now ready to introduce multiple frequency I(1), short MFI(1), processes, compare Bauer and Wagner (2012, Definition 1):

Definition 12 • *The difference operator at frequency $0 \leq \omega \leq \pi$ is defined by*

$$\Delta_\omega := \begin{cases} 1 - 2\cos(\omega)L + L^2, & \text{for } 0 < \omega < \pi, \\ 1 - \cos(\omega)L, & \text{for } \omega \in \{0, \pi\}. \end{cases} \tag{2.3}$$

For notational brevity, we again omit the dependence on L in $\Delta_\omega(L)$.

- *The s-dimensional process $\{y_t\}_{t\in\mathbb{Z}}$ is called MFI(1) process with set of unit root frequencies $\Omega := \{\tilde{\omega}_1,\dots,\tilde{\omega}_l\}$ with $0 \leq \tilde{\omega}_1 < \tilde{\omega}_2 < \dots < \tilde{\omega}_l \leq \pi$, if it is a solution of the difference equation*

$$\Delta_\Omega(y_t - Ds_t) := \prod_{k=1}^{l} \Delta_{\tilde{\omega}_k}(y_t - Ds_t) = v_t, \quad t \in \mathbb{Z}, \tag{2.4}$$

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

- *An MFI(1) process with set of unit root frequencies $\Omega := \{0\}$ is called integrated process of order one, short I(1) process.*

- *A stationary VARMA process $\{y_t - Ds_t\}_{t\in\mathbb{Z}} = c(L)\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is said to have an empty set of unit root frequencies $\Omega_0 := \{\}$.*

- *An MFI(1) process with set of unit root frequencies $\Omega := \{\tilde{\omega}_1,\dots,\tilde{\omega}_l\}$ is called seasonally integrated, if there exists $S \in \mathbb{N}$ such that $e^{i\tilde{\omega}_k S} = 1$ for all $k = 1,\dots,l$. In this case it holds that $\Omega \subset \{\omega_1,\dots,\omega_S\}$.*

Remark 18 *As in de Matos Ribeiro et al. (2020) 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 $((\tilde{\omega}_1, 1),...,(\tilde{\omega}_l, 1))$. Since in this paper we only deal with MFI(1) processes the simpler notation of only listing the unit root frequencies suffices.*

Every VARMA system with $\det(a(z)) \neq 0$ for all z with $|z| \leq 1$ has a unique stationary solution on $\mathbb{Z}$ depending only on the transfer function $k(z)$ and the process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ – see, e. g., Hannan and Deistler (1988), who also show that this solution can be represented as a solution of a state space system

$$\begin{aligned} y_t &= Cx_t + Ds_t + \varepsilon_t \\ x_{t+1} &= Ax_t + B\varepsilon_t, \end{aligned} \tag{2.5}$$

where $A \in \mathbb{C}^{n\times n}$, $B \in \mathbb{C}^{n\times s}$, $C \in \mathbb{C}^{s\times n}$ and $D \in \mathbb{R}^{s\times m}$. This result is generalized to MFI(1) processes and more general unit root processes in Bauer and Wagner (2012, Theorem 1), compare Proposition 1 below. By Hannan and Deistler (1988, Theorem 1.2.2) the roots of $\det(a(z))$ are equal to the inverses of the non-zero eigenvalues of A. Consequently the condition $\det(a(z)) \neq 0$, $|z| \leq 1$ is equivalent to $\lambda_{|\max|}(A) < 1$. By the same argument the unit roots correspond to eigenvalues of A with modulus 1. The state space representation of a process $\{y_t\}_{t\in\mathbb{Z}}$ is not unique. There are two sources of non-uniqueness: First, there exist representations with different state dimension n. Second, for given minimal state dimension, the basis of the state space can still be chosen arbitrarily as the state is not directly observed. Technically, state space realizations (A, B, C) and $(\tilde{A}, \tilde{B}, \tilde{C})$ are called observationally equivalent if they describe the same transfer function $k(z) = \pi(A, B, C) = I_s + \sum_{j=1}^{\infty} CA^{j-1}Bz^j$. We call the state space realization (A, B, C) controllable if the controllability matrix $\mathcal{C} = [B, AB, \ldots, A^{n-1}B]$ is of full rank, and observable if the observability matrix $\mathcal{O}' = [C', A'C', \ldots, (A^{n-1})'C']$ is of full rank. The state space realization (A, B, C) is called minimal if it is both observable and controllable. If a state space realization is minimal, there is no observationally equivalent realization with a lower state dimension. The state dimension of a minimal realization is called the order of the state space system, see, e.g., Hannan and Deistler (1988, Chapter 2.3).
The following theorem, compare Bauer and Wagner (2012, Theorem 1), describes the relation between state space systems and MFI(1) processes.

Proposition 1 *Let $\{y_t\}_{t\in\mathbb{Z}}$ be an MFI(1) process as given in Definition 12. Then there exists a solution $\{y_{t,h}\}_{t\in\mathbb{Z}}$ of the homogeneous equation $\Delta_\Omega y_{t,h} = 0$, $t \in \mathbb{Z}$, such that the process $\{y_{t,p}\}_{t\in\mathbb{Z}} := \{y_t - y_{t,h}\}_{t\in\mathbb{Z}}$ is the particular solution given by*

$$\begin{aligned} y_{t,p} &= C_u x_{t,u} + C_\bullet x_{t,\bullet} + Ds_t + \varepsilon_t \\ x_{t,u} &= \begin{cases} \sum_{j=1}^{t-1} A_u^{j-1} B_u \varepsilon_{t-j}, \; t \geq 1 \\ -\sum_{j=t}^{0} A_u^{j-1} B_u \varepsilon_{t-j}, \; t < 1, \end{cases} \\ x_{t,\bullet} &= \sum_{j=1}^{\infty} A_\bullet^{j-1} B_\bullet \varepsilon_{t-j} \end{aligned} \tag{2.6}$$

to a minimal state space system (A, B, C) with $A = diag(A_u, A_\bullet)$, where all eigenvalues of A_u are simple (i.e., their algebraic and geometric multiplicity coincide) and have unit modulus and $\lambda_{|max|}(A_\bullet) < 1$ and $B = [B_u', B_\bullet']'$ and $C = [C_u, C_\bullet]$ are partitioned accordingly.
Conversely, every process $\{y_{t,p}\}_{t\in\mathbb{Z}}$ defined through (2.6) *for a minimal state space system (A, B, C) satisfying $\lambda_{|max|}(A) = 1$ and $\lambda_{|max|}(A - BC) \leq 1$ is an MFI(1) process.*

Thus, in this paper we consider multivariate data y_t, $t = 1, \ldots, T$, from a process $\{y_t\}_{t\in\mathbb{Z}}$ fulfilling the following assumptions:

Assumption 1 (DGP) *The seasonally integrated process $\{y_t\}_{t\in\mathbb{Z}}$ with a set of unit root frequencies $\Omega = \{\tilde{\omega}_1, \ldots, \tilde{\omega}_l\}$, with $\Omega \subset \{\omega_1, \ldots, \omega_S\}$ for an even integer S, is a solution of a state space system*

$$\begin{aligned} y_t &= \mathcal{C}x_t + Ds_t + \varepsilon_t \\ x_{t+1} &= \mathcal{A}x_t + \mathcal{B}\varepsilon_t, \end{aligned} \tag{2.7}$$

where $\mathcal{A} \in \mathbb{C}^{n\times n}$*,* $\mathcal{B} \in \mathbb{C}^{n\times s}$*,* $\mathcal{C} \in \mathbb{C}^{s\times n}$ *and* $D \in \mathbb{R}^{s\times m}$ *for a deterministic sequence* $s_t \in \mathbb{R}^m$ *containing a constant, a linear trend and seasonal dummies.*
(State Space System:) *The state space system* $(\mathcal{A}, \mathcal{B}, \mathcal{C})$ *is minimal and of the following form:*

- $\mathcal{A} = diag(\mathcal{A}_{1,\mathbb{C}}, ..., \mathcal{A}_{l,\mathbb{C}}, \mathcal{A}_\bullet)$ *where* $\mathcal{A}_{j,\mathbb{C}} = diag(e^{i\tilde{\omega}_j} I_{\tilde{c}_j}, e^{-i\tilde{\omega}_j} I_{\tilde{c}_j})$ *for* $0 < \tilde{\omega}_j < \pi$ *or* $\mathcal{A}_{j,\mathbb{C}} = e^{i\tilde{\omega}_j} I_{\tilde{c}_j}$ *for* $\tilde{\omega}_j \in \{0, \pi\}$*. All eigenvalues of* $\mathcal{A}_\bullet$ *are smaller than 1 in modulus.*

- $\mathcal{B} = [\mathcal{B}_{1,\mathbb{C}}', ..., \mathcal{B}_{l,\mathbb{C}}', \mathcal{B}_\bullet']', \mathcal{B}_{j,\mathbb{C}} = [\mathcal{B}_j', \overline{\mathcal{B}_j}']', \mathcal{B}_j \in \mathbb{C}^{\tilde{c}_j \times s}$ *for* $z_j \neq \pm 1$ *and* $\mathcal{B}_{j,\mathbb{C}} = \mathcal{B}_j \in \mathbb{R}^{\tilde{c}_j\times s}$ *else, where* $\mathcal{B}_j$ *for* $j = 1, \ldots, l$ *are positive upper triangular matrices.*

- $\mathcal{C} = [\mathcal{C}_{1,\mathbb{C}}, ..., \mathcal{C}_{l,\mathbb{C}}, \mathcal{C}_\bullet], \mathcal{C}_{j,\mathbb{C}} = [\mathcal{C}_j, \overline{\mathcal{C}_j}], \mathcal{C}_j \in \mathbb{C}^{s\times\tilde{c}_j}$ *for* $z_j \neq \pm 1$ *and* $\mathcal{C}_{j,\mathbb{C}} = \mathcal{C}_j \in \mathbb{R}^{s\times\tilde{c}_j}$ *else, where* $\mathcal{C}_j'\mathcal{C}_j = I_{\tilde{c}_j}$ *for* $j = 1, \ldots, l$*.*

- $x_t' = [x_{t,1,\mathbb{C}}', ..., x_{t,l,\mathbb{C}}', x_{t,\bullet}']', x_{t,j,\mathbb{C}} = [x_{t,j}', \overline{x_{t,j}}']', x_{t,j} \in \mathbb{C}^{\tilde{c}_j}$ *for* $z_j \neq \pm 1$ *and* $x_{t,j} \in \mathbb{R}^{\tilde{c}_j}$ *else such that* $x_{1,j} = 0, x_{1,\bullet} = \sum_{j=0}^{\infty} \mathcal{A}_\bullet^j \mathcal{B}_\bullet \varepsilon_{-j}$*.*

- $|\lambda_{max}(\mathcal{A} - \mathcal{B}\mathcal{C})| < 1$ *(strict minimum-phase assumption).*

(Deterministics:) *The deterministic term* Ds_t *is of the form*

- $Ds_t := d_1 + \sum_{k=2}^{\tilde{S}-1} \left[d_k^r \cos(\omega_k(t-1)) + d_k^i \sin(\omega_k(t-1))\right] + d_{\tilde{S}}(-1)^{t-1} + d_{S+1}t$ *where* $\tilde{S} := S/2 + 1$*,* $d_1, d_{\tilde{S}}, d_{S+1} \in \mathbb{R}^s$*, and* $d_k^r, d_k^i \in \mathbb{R}^s$*, for* $k = 2, \ldots, \tilde{S} - 1$*.*

(Noise process:) *The noise process* $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ *fulfills the following assumptions:*

- $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ *is a strictly stationary martingale difference sequence.*
- $\mathbb{E}(\varepsilon_t\varepsilon_t') = \mathbb{E}(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}) = \Sigma$*.*
- $\mathbb{E}(\|\varepsilon_t\|^4) < \infty$*.*

Here $\mathcal{F}_{t-1}$ *is the* σ*-algebra spanned by* $\{\varepsilon_j\}_{j\in\mathbb{Z}, j<t}$*.*

Collecting the integers $\tilde{c}_k$ together with their frequencies the state space unit root structure is defined as

$$\Omega_{\mathcal{S}} := \{(\tilde{\omega}_1, \tilde{c}_1), \ldots, (\tilde{\omega}_l, \tilde{c}_l)\}.$$

The transfer function associated with the state space system is given by $k(z) = \sum_{j=0}^{\infty} K_j z^j = I_s + zC(I_n - zA)^{-1}B$. We are now ready to define the set $M(\Omega_{\mathcal{S}}, n_\bullet)$ of transfer functions of given state space unit roots structure $\Omega_{\mathcal{S}}$ as

$$M(\Omega_{\mathcal{S}}, n_\bullet) := \left\{ k(z) : \begin{array}{c} k(z) = I_s + zC(I_n - zA)^{-1}B, \\ A \in \mathbb{R}^{n\times n}, B, C' \in \mathbb{R}^{n\times s}, \text{ as in Proposition 1} \\ \text{with } A_u \text{ corresponding to } \Omega_{\mathcal{S}} \text{ and } A_\bullet \in \mathbb{R}^{n_\bullet \times n_\bullet} \end{array} \right\}.$$

Defining $n_u(\Omega_{\mathcal{S}}) := \sum_{k=1}^{l} \delta_k c_k$ with $\delta_k = 1$ if $\tilde{\omega}_k \in \{0, \pi\}$ and $\delta_k = 2$ if $0 < \tilde{\omega}_k < \pi$, the order of a transfer function $k(z) \in M(\Omega_{\mathcal{S}}, n_\bullet)$ is equal to $n = n_u(\Omega_{\mathcal{S}}) + n_\bullet$.
Inserting $\varepsilon_t = y_t - Cx_t - Ds_t$ into the equation of the state x_{t+1}, the system given in (2.5) is alternatively expressed as

$$\begin{aligned} y_t &= Cx_t + Ds_t + \varepsilon_t \\ x_{t+1} &= \underline{A}x_t + B(y_t - Ds_t) \qquad \underline{A} := A - BC. \end{aligned}$$

Analogously, for the inverse transfer function $k^{-1}(z) = \sum_{j=0}^{\infty} K_j^- z^j$ it holds that $K_0^- = I_s$ and $K_j^- = -C\underline{A}^{j-1}B$ and, thus, $k^{-1}(z) = I_s - zC(I_n - z\underline{A})^{-1}B$. For a transfer function corresponding to a seasonally integrated process with unit root frequencies $\Omega \subset \{\omega_1, \ldots, \omega_S\}$, define $\Pi_k := -k^{-1}(e^{i\omega_k}) \in \mathbb{C}^{s\times s}$, $k = 1, \ldots, S$ and let r_k be the rank of Π_k, with $c_k := s - r_k$. The matrix Π_k is of full rank s if $\omega_k \notin \Omega$. If $\omega_k = \tilde{\omega}_j \in \Omega$, the rank of Π_k is equal to $r_k = s - \tilde{c}_j$ and, consequently, there exists a decomposition $\Pi_k = \alpha_k \beta_k'$ with $\alpha_k, \beta_k \in \mathbb{C}^{s \times r_k}$. The definition of Π_k dependent on the inverse transfer function $k^{-1}(z)$ corresponding to a VARMA process is such that it coincides with the matrices of the VECM in (2.1) in case of VAR processes and it holds that

$$V(p, \{(\tilde{\omega}_1, \tilde{r}_1), \ldots, (\tilde{\omega}_l, \tilde{r}_l)\}) \subset \overline{M}(\Omega_{\mathcal{S}}, n_\bullet)$$

with $\Omega_{\mathcal{S}} = \{(\tilde{\omega}_1, \tilde{c}_1), \ldots, (\tilde{\omega}_l, \tilde{c}_l)\}$ and $n_\bullet$ such that $\tilde{c}_j = s - \tilde{r}_j$ and $n_u(\Omega_{\mathcal{S}}) + n_\bullet \geq ps$. Moreover, the column space of β_k is orthogonal to the column space of the matrix $\mathcal{C}_k$ of the canonical form and the columns space α_k is orthogonal to the row space of $\mathcal{B}_k$, which follows by comparing (2.6) with the Granger representation of MFI(1) processes, compare Johansen and Schaumburg (1999, Theorem 3).

The column spaces of $\mathcal{C}_k$ and β_k are closely related to the cointegrating properties of the process $\{y_t\}_{t \in \mathbb{Z}}$:

Definition 13 *(i) An MFI(1) process $\{y_t\}_{t \in \mathbb{Z}}$ with set $\Omega = \{\omega_1, ..., \omega_l\}$ of unit root frequencies is called statically cointegrated at ω_k, if there exists a vector $\beta \in \mathbb{R}^s$ such that the set of unit root frequencies of the process $\{\beta' y_t\}_{t \in \mathbb{Z}}$ does not contain ω_k. In this case β is called a cointegrating vector (CIV) at ω_k.*
For an MFI(1) process the cointegrating space at frequency ω_k is spanned by all cointegrating vectors β_k at ω_k.

(ii) An MFI(1) process $\{y_t\}_{t \in \mathbb{Z}}$ with set $\Omega = \{\omega_1, ..., \omega_l\}$ is called dynamically cointegrated at ω_k for $0 < \omega_k < \pi$, if there exist vectors $\beta_0, \beta_1 \in \mathbb{R}^s$ such that the set of unit root frequencies of the process $\{\beta_0' y_t - \beta_1' y_{t-1}\}_{t \in \mathbb{Z}}$ does not contain ω_k. In this case $\beta(L) = \beta_0 + \beta_1 L$ is called a polynomial cointegrating vector (PCIV) of degree one at ω_k.
For an MFI(1) process the polynomial cointegrating space at frequency ω_k is spanned by all polynomial cointegrating vectors $\beta(L)$ of degree one at ω_k.

Combining the equations in Proposition 1 and using the block structure of $(\mathcal{A}_u, \mathcal{B}_u, \mathcal{C}_u)$ we get

$$y_t = \sum_{k=1}^{l} \mathcal{C}_{k,\mathbb{C}} \left(\sum_{j=1}^{t-1} \mathcal{A}_{k,\mathbb{C}}^{j-1} \mathcal{B}_{k,\mathbb{C}} \varepsilon_{t-j} \right) + \mathcal{C}_\bullet \left(\sum_{j=1}^{t-1} \mathcal{A}_\bullet^{j-1} \mathcal{B}_\bullet \varepsilon_{t-j} + \right) + \mathcal{C} \mathcal{A}^{t-1} x_1 + D s_t. \tag{2.8}$$

Example 1 and 2 of Bauer et al. (2020) show that for real $\omega_k \in \{0, \pi\}$ the cointegrating space at ω_k is spanned by the columns of β_k which fulfill $\beta_k' \mathcal{C}_k = 0$ and for $0 < \omega_k < \pi$ the cointegrating space is spanned by the vectors $\gamma \in \mathbb{R}^s$ which fulfill $\gamma' \mathcal{C}_k = 0$. Moreover, there is a close relation between the columns of $\beta_k \in \mathbb{C}^s$ and the polynomial cointegrating vectors of degree one at ω_k. It holds that the space of PCIVs – considered as a subspace of $\mathbb{R}^{2s}$, identifying $\gamma(z) = \gamma_0 + \gamma_1 z$ with the vector $[\gamma_0', \gamma_1']'$ – is spanned by the columns of

$$\begin{bmatrix} I_s & 0_{s \times s} \\ -\mathcal{R}(z_k) I_s & \mathcal{I}(z_k) I_s \end{bmatrix} \begin{bmatrix} \mathcal{R}(\beta_k) & \mathcal{I}(\beta_k) \\ -\mathcal{I}(\beta_k) & \mathcal{R}(\beta_k) \end{bmatrix}.$$

For this reason the matrices $\mathcal{C}_k$ and β_k contain all information on the cointegrating spaces.

Bauer et al. (2020) provide a parameterization of the sets $M(\Omega_{\mathcal{S}}, n_\bullet)$. The parameterization partitions the set into subsets with one 'generic' subset open and dense in $M(\Omega_{\mathcal{S}}, n_\bullet)$. Furthermore, Bauer et al. (2020) define a parameterization of this open and dense subset of $M(\Omega_{\mathcal{S}}, n_\bullet)$ by establishing a real valued parameter space and a bijective mapping attaching parameters to transfer functions. Furthermore, Bauer et al. (2020) clarify the relations between sets of transfer functions for different sets $M(\Omega_{\mathcal{S}}, n_\bullet)$.

In the following let Θ_n denote the parameter space corresponding to $M(\{\}, n)$ and let $\Theta_n^{c,\omega}$ denote the parameter space corresponding to $M(\{(\omega, c)\}, n - c)$ if $\omega \in \{0, \pi\}$ and to $M(\{(\omega, c)\}, n - 2c)$ if $0 < \omega < \pi$, for a given unit root ω.

In this setting de Matos Ribeiro et al. (2020) investigate the asymptotic properties of the pseudo maximum likelihood estimator using the Gaussian likelihood function: Let the parameter vector $\theta \in \Theta$ correspond to a system $(\mathcal{A}(\theta), \mathcal{B}(\theta), \mathcal{C}(\theta))$ and the transfer function $k(z,\theta) = I_s + z\mathcal{C}(\theta)(I_n - z\mathcal{A}(\theta))^{-1}\mathcal{B}(\theta)$. Let S_σ denote a parameter space corresponding to the set of all positive definite symmetric matrices $\Sigma \in \mathbb{R}^{s\times s}$ and $\sigma \in S_\sigma$. Then the logarithm of the Gaussian likelihood function under the assumption that $x_1 = 0$ is given (up to a constant) as:

$$\begin{aligned}
\mathcal{L}_T(\theta, D, \sigma) &= -\frac{T}{2}\log|\Sigma(\sigma)| - \frac{1}{2}\sum_{t=1}^{T}\varepsilon_t(\theta, D)'\Sigma^{-1}(\sigma)\varepsilon_t(\theta, D)\\
\varepsilon_t(\theta, D) &= y_t - Ds_t - \mathcal{C}(\theta)x_t(\theta, D)\\
x_{t+1}(\theta, D) &= \underline{\mathcal{A}}(\theta)x_t(\theta, D) + \mathcal{B}(\theta)(y_t - Ds_t), \quad x_1(\theta, D) = 0,\ \underline{\mathcal{A}}(\theta) := \mathcal{A}(\theta) - \mathcal{B}(\theta)\mathcal{C}(\theta).
\end{aligned}$$

The estimators maximizing this function are shown to be weakly consistent in de Matos Ribeiro et al. (2020, Theorem 1):

Proposition 2 (Consistency of PML) *Let $\{y_t\}_{t\in\mathbb{Z}}$ be a real valued MFI(1) process generated by a system of the form (2.6) with state space unit root structure Ω_S and dimension $n_\bullet$ of the stable subsystem $(A_\bullet, B_\bullet, C_\bullet)$. Let $(k_\circ(z), D_\circ, \sigma_\circ)$ denote the transfer function and parameters corresponding to $\{y_t\}_{t\in\mathbb{Z}}$.*
Let $\phi : M \to \Theta$ be a parameterization of a set M of transfer functions and let $k_\circ(z) \in \overline{M}$.
Let $(\hat\theta, \hat D, \hat\sigma)$ denote the pseudo maximum likelihood estimator, i. e., the maximizer of $\mathcal{L}_T(\theta, D, \sigma)$, over $\Theta \times \mathbb{R}^{s\times(S+1)} \times S_\sigma$. Then $(k(z,\hat\theta), \hat D, \hat\sigma)$ is consistent for $(k_\circ(z), D_\circ, \sigma_\circ)$ and $k(z,\hat\theta) = I_s + \sum_{j=1}^{\infty} K_j(\hat\theta)z^j$ converges in probability to the true transfer function $k_\circ(z) = I_s + \sum_{j=1}^{\infty} K_{j,\circ}z^j$ with rate $T^{1/2}$, i.e. $T^\gamma\|K_j(\hat\theta) - K_{j,\circ}\| \to 0$ in probability for all $j \in \mathbb{N}$ and all $0 < \gamma < 1/2$. Furthermore,

$$T^\gamma\|k^{-1}(e^{i\tilde\omega_j}, \hat\theta)\mathcal{C}_{j,\circ}\| \to 0,\ j = 1, \ldots, l$$

in probability for all $0 < \gamma < 1$, where $\mathcal{C}_{j,\circ}$ is the respective subblock of the system matrix $\mathcal{C}_\circ$ corresponding to $k_\circ(z)$.

Note that here the pseudo log-likelihood function is defined for $x_1 = 0$ whereas the data generating process assumes that $x_{1,\bullet}$ corresponds to the stationary distribution of the stable part of the state. It can be shown that the difference in the corresponding pseudo log-likelihood functions is negligible in the sense that the asymptotic distributions of the corresponding estimators coincide, compare Theorem 2 in (de Matos Ribeiro et al., 2020).
Note that the crucial component of the Gaussian likelihood is $\varepsilon_t(\theta, D) = k^{-1}(L;\theta)(y_t - Ds_t)$ with the convention that $y_t - Ds_t = 0, t \le 0$. This representation indicates that a power series representation around the unit roots leads to similar expressions as in the VAR case. The main insight lies in the fact, that this power series development in the state space case only differs in the definition of the contribution by (seasonally) differenced terms compared to the VAR situation. This observation first has been made in the $I(1)$ case by Ribarits and Hanzon (2014) leading to their state space error correction model (SSECM). The next section, thus, first discusses the simpler case of I(1) processes, while Section 2.4 focuses on cointegration analysis at complex unit roots in SSECMs for seasonal cointegrating processes.

2.3 I(1) processes

2.3.1 Error Correction Representation in the State Space Framework

Ribarits and Hanzon (2014) developed a state space error correction model for I(1) processes. At the heart of their approach lies a Beveridge Nelson decomposition of the inverse transfer function $k^{-1}(z)$. We note that

$$k^{-1}(z) = k^{-1}(1)z + (1-z)\frac{k^{-1}(z) - k^{-1}(1)z}{1-z}$$

which for state space systems translates into (for example, by comparing the power series coefficients)

$$\begin{aligned} k^{-1}(z) = \sum_{j=0}^{\infty} K_j^- z^j &= I_s - zC\sum_{j=0}^{\infty}(\underline{A}z)^j B \\ &= \left(I_s - C(I_n - \underline{A})^{-1}B\right)z + (1-z)\left(I_s + zC(I_n-\underline{A})^{-1}\underline{A}\sum_{j=0}^{\infty}(\underline{A}z)^j B\right) \\ &= -\Pi z + \left(\sum_{i=0}^{\infty}\tilde{K}_i^- z^i\right)(1-z) = -\Pi z + \tilde{k}^-(z)\boldsymbol{\Delta}_1(z). \end{aligned}$$

Truncating the power series at power t we obtain for every state space system (A, B, C) that (using $y_0 = 0$)

$$\sum_{j=0}^{t-1} K_j^- y_{t-j} = -\Pi y_{t-1} + \sum_{i=0}^{t-1}\tilde{K}_i^- \boldsymbol{\Delta}_1 y_{t-i}.$$

This builds the cornerstone of the state space error correction model (SSECM): As will be shown in Theorem 6, for every state space system (A, B, C) the residuals

$$\varepsilon_t(A,B,C) = \sum_{j=0}^{t-1} K_j^-(y_{t-j} - Ds_{t-j})$$

where $K_0^- = I_s, K_j^- = -C\underline{A}^{j-1}B, j \in \mathbb{N}$ have the following representation:

$$\begin{aligned} \boldsymbol{\Delta}_1(y_t - Ds_t) &= \Pi(y_{t-1} - Ds_{t-1}) + Cv_t + \varepsilon_t(A,B,C) \\ v_{t+1} &= \underline{A}v_t - (I_n - \underline{A})^{-1}\underline{A}B\boldsymbol{\Delta}_1(y_t - Ds_t), \quad v_1 = x_1, \\ \Pi &= -I_s + C(I_n - \underline{A})^{-1}B. \end{aligned} \tag{2.9}$$

If the process $\{y_t\}_{t\in\mathbb{Z}}$ fulfills the assumptions stated in Assumption 1 with corresponding system $(A_\circ, B_\circ, C_\circ)$, then for $D = D_\circ$

$$\varepsilon_t(A_\circ, B_\circ, C_\circ) = \varepsilon_t.$$

Furthermore, let the deterministics be given as $Ds_t := [d_1, d_2][1, t]'$. For a system (A, B, C) assume that $-\Pi = k^{-1}(1) := -\alpha\beta'$, where $\alpha, \beta \in \mathbb{R}^{s\times r}$ and define $\Psi := \frac{\partial k^{-1}(z)}{\partial z}\Big|_{z=1} = -C(I_n - \underline{A})^{-2}B$. The state space error correction model (SSECM) then is given by

$$\boldsymbol{\Delta}_1 y_t = \Pi(y_{t-1} + d_1 + d_2 t) - \Psi d_2 + Cv_t + \tilde{\varepsilon}_t(A,B,C,D), \tag{2.10}$$

$$v_{t+1} = \underline{A}v_t - (I_n - \underline{A})^{-1}\underline{A}B\boldsymbol{\Delta}_1 y_t, \tag{2.11}$$

$$\begin{aligned} v_1 &= x_1 + (I_n - \underline{A})^{-1}Bd_1 - (I_n - \underline{A})^{-2}Bd_2 + (I_n - \underline{A})^{-1}Bd_2, \\ \Pi &= -I_s + C(I_n - \underline{A})^{-1}B, \end{aligned}$$

where again

$$\tilde{\varepsilon}_t(A_\circ, B_\circ, C_\circ, D_\circ) = \varepsilon_t.$$

Note that if $\underline{A}$ is nilpotent, the solutions of the system correspond to VAR processes as then $\underline{A}^n = 0$. Also in this case our approach differs from Johansen's VECM which considers the pseudo log-likelihood function conditional on the first $k-1$ values. Therefore, only $T-k+1$ equations are considered in the likelihood analysis in the Johansen framework while the above model considers

T equations but assumes a zero starting state.
Further, note that there are two ways to include deterministic terms: Either they are modelled as additive components to the process y_t, such that y_t is replaced by $y_t - Ds_t$ in the state space model as in (2.9), with deterministics present in both the output equation and the modified state equation. Alternatively, deterministics can be included as regressors in the error correction formulation, as is done in the Johansen framework. This corresponds to choosing v_1 independent of d_1 and d_2, since the influence of starting values v_1 on the output is negligible for large t, such that deterministics are only present in the output equation (2.10). Also note that the presence of a linear trend term in the solutions process when introducing the deterministics via regressors in the error correction representation is ascertained by including a restricted linear trend term of the form $\alpha\rho t$ in the error correction equation.
Note that – apart from the treatment of the starting values – the SSECM takes the same form as the VECM in the VAR framework: The right hand side essentially features the matrix Π multiplied by y_{t-1} plus a stationary process and potentially deterministics as regressors. The main difference to the Johansen framework lies in the fact that the regressors here depend on parameters that are also determining the matrix Π. The next subsections describes two different approaches to deal with these interdependencies.

2.3.2 Concentration of the Gaussian Pseudo Log-Likelihood Function

Consider for the moment the simpler case with no deterministic terms present neither in the data generating process nor in the model, that is $D = D_\circ = 0$. Then for the I(1) model given in Assumption 1 the Gaussian pseudo log-likelihood function for $x_1 = 0$ is equal to

$$\begin{aligned}
\mathcal{L}_T(\theta,\sigma) &= -\frac{T}{2}\log|\Sigma(\sigma)| - \frac{1}{2}\sum_{t=1}^{T}\varepsilon_t(\theta)'\Sigma^{-1}(\sigma)\varepsilon_t(\theta)\\
\varepsilon_t(\theta) &= \mathbf{\Delta}_1 y_t - \Pi(\theta)y_{t-1} - C(\theta)v_t(\theta)\\
v_{t+1}(\theta) &= \underline{A}(\theta)v_t(\theta) - (I_n - \underline{A}(\theta))^{-1}\underline{A}(\theta)B(\theta)\mathbf{\Delta}_1 y_t, \quad v_1(\theta) = 0.
\end{aligned}$$

Concentrating out the noise variance parameter σ we obtain (up to the constant $-Ts/2$):

$$L_T(\theta) = -\frac{T}{2}\log\left|\hat{\Sigma}_T(\theta)\right| = -\frac{T}{2}\log|\langle\varepsilon_t(\theta),\varepsilon_t(\theta)\rangle|\,.$$

Note that $v_t(\theta)$ depends on the parameter vector only via the pair $(\underline{A}, B)$. The equation used for estimation equals

$$\mathbf{\Delta}_1 y_t = \Pi(\theta)y_{t-1} + C(\theta)v_t(\theta) + \varepsilon_t(\theta).$$

where $\Pi(\theta) = -I_s + C(I - \underline{A})^{-1}B$ depends on $(\underline{A}, B)$ and C.
We are interested in the specification of the number of cointegrating relations as well as on inference on the cointegrating vectors. This information is encoded in the matrix $\Pi(\theta)$. Our aim is, therefore, to find an expression of L_T depending on the rank and the kernel of Π. Let us first define an expanded version of the likelihood function for general $\Pi \in \mathbb{R}^{s\times s}, C \in \mathbb{R}^{s\times n}$ not necessarily respecting the constraints $\Pi = \Pi(\theta)$, $C = C(\theta)$, as

$$L_T^{ex}(\Pi, C, \theta) := -\frac{T}{2}\log|\langle\mathbf{\Delta}_1 y_t - \Pi y_{t-1} - Cv_t(\theta), \mathbf{\Delta}_1 y_t - \Pi y_{t-1} - Cv_t(\theta)\rangle|\,.$$

Let $\hat{\theta}$ denote the pseudo maximum likelihood estimator of θ over a parameter space Θ. Note that the following relations hold

$$L_T(\hat{\theta}) = \max_{\theta\in\Theta} L_T(\theta) \leq \max_{\substack{[\Pi, C]\,\in\,\mathbb{R}^{s\times(s+n)}\\ \Pi = -I_s + C(I_n - \underline{A}(\hat{\theta}))^{-1}B(\hat{\theta})}} L_T^{ex}(\Pi, C, \hat{\theta}) \leq \max_{(\Pi,C)\in\mathbb{R}^{s\times(s+n)}} L_T^{ex}(\Pi, C, \hat{\theta}).$$

We will show that the solutions to these three problems are related. Solutions to the rightmost problem with the fewest restrictions will be called **unrestricted concentration approach**, while the solutions to the problem in the middle will be termed **restricted concentration approach**. Thus, consider the unrestricted concentration approach which maximizes $L_T^{ex}(\Pi, C, \theta)$ for given θ and subject to $rank(\Pi) = r$. Here as usual the matrix C can be concentrated out using simple regression techniques leading to a reduced model $R_{0,t}(\theta) = \Pi R_{1,t}(\theta) + \varepsilon_t^U(\theta)$, where

$$R_{0,t}(\theta) := \mathbf{\Delta}_1 y_t - \langle \mathbf{\Delta}_1 y_t, v_t(\theta)\rangle \langle v_t(\theta), v_t(\theta)\rangle^{-1} v_t(\theta)$$
$$R_{1,t}(\theta) := y_{t-1} - \langle y_{t-1}, v_t(\theta)\rangle \langle v_t(\theta), v_t(\theta)\rangle^{-1} v_t(\theta).$$

Next, rewriting $\Pi = \alpha\beta'$, $\alpha, \beta \in \mathbb{R}^{s\times r}$, concentrate out α for given β, where the maximizer is given by the corresponding OLS-estimator $\langle R_{0,t}(\theta), \beta' R_{1,t}(\theta)\rangle \langle \beta' R_{1,t}(\theta), \beta' R_{1,t}(\theta)\rangle^{-1}$. Using the notation of Johansen (1996), define:

$$S_{00}(\theta) := \langle R_{0,t}(\theta), R_{0,t}(\theta)\rangle \qquad S_{10}(\theta) := \langle R_{1,t}(\theta), R_{0,t}(\theta)\rangle =: S_{10}(\theta)'$$
$$S_{11}(\theta) := \langle R_{1,t}(\theta), R_{1,t}(\theta)\rangle \qquad S_{11,0}(\theta) := S_{11}(\theta) - S_{10}(\theta)S_{00}(\theta)^{-1}S_{01}(\theta).$$

Consequently, the unrestricted concentration approach leads to the maximization of the function

$$\begin{aligned} L_T^{ex,U}(\beta,\theta) :&= -\frac{T}{2}\log|S_{00}(\theta) - S_{01}(\theta)\beta(\beta' S_{11}(\theta)\beta)^{-1}\beta' S_{10}(\theta)| \\ &= -\frac{T}{2}\log\left(|S_{11}(\theta)|\frac{|\beta' S_{11,0}(\theta)\beta|}{|\beta' S_{11}(\theta)\beta|}\right). \end{aligned}$$

The solution of this reduced rank regression problem is well known in the I(1) literature.
The alternative restricted concentration approach takes the relation between Π, C and θ into account. This approach has been pioneered by Ribarits and Hanzon (2014). Note that $\alpha\beta' = -I_s + C(I_n - \underline{A}(\theta))^{-1}B(\theta)$ is equivalent to $I_s = C\underline{B}(\theta) - \alpha\beta'$ where $\underline{B}(\theta) := (I_n - \underline{A}(\theta))^{-1}B(\theta)$. This defines a linear restriction between the matrices C and α for given $\underline{B}(\theta)$ and β.
Note that if $\underline{B}(\theta)$ does not have full column rank, not all matrices β allow for solutions. This will be the case for example if $n < s$. A necessary assumption for θ and β to be compatible is for $\tilde{B}(\beta,\theta) := [\underline{B}(\theta)', -\beta]'$ to be of full (column) rank. Otherwise, there exists a vector $\gamma \neq 0$, $\gamma \in \mathbb{R}^s$, such that $\tilde{B}(\beta,\theta)\gamma = 0$. This in turn implies $\gamma = I_s\gamma = C\underline{B}(\theta)\gamma - \alpha\beta'\gamma = 0$, leading to a contradiction. If $\tilde{B}(\beta,\theta)$ is of full rank, $\underline{B}(\theta)(\beta)_\perp$ is of full rank and a solution $\check{C}$ for $C\underline{B}(\theta)(\beta)_\perp = (\beta)_\perp$ exists (which is not necessarily unique). Setting $\alpha = (\check{C}\underline{B}(\theta) - I_s)\beta(\beta'\beta)^{-1}$, the set of matrices C and α for given θ, β fulfilling the restrictions is not empty.
For true $\theta_\circ$ and $\Pi_\circ = \alpha_\circ\beta_\circ'$ the matrix $\tilde{B}(\beta_\circ, \theta_\circ)$ needs to be of full column rank for $I_s = C_\circ\underline{B}(\theta_\circ) - \alpha_\circ\beta_\circ'$ to hold. Note also that if $n \geq s$, the set of parameter vectors θ such that $\underline{B}(\theta)$ and, thus, $\tilde{B}(\beta,\theta)$ for arbitrary β has full column rank is generic[2]. Hence, a pair of estimates θ, β will allow for a solution fulfilling the restrictions with probability one.
In light of the above discussion, the problem to be solved is to maximize

$$L_T^{ex}(\alpha\beta', C, \theta)$$

subject to

$$I_s = \left[\begin{array}{cc} C & \alpha \end{array}\right]\tilde{B}(\beta,\theta), \qquad \text{where } \tilde{B}(\beta,\theta) := \left[\begin{array}{c} \underline{B}(\theta) \\ -\beta' \end{array}\right],$$

for given θ and $\beta \in \mathbb{B}(r,\theta) := \{\beta \in \mathbb{R}^{r\times s} : \tilde{B}(\beta,\theta) \text{ of full column rank}\} \subset \mathbb{R}^{r\times s}$. Solving for C and α, we arrive at a system of first order equations of the form

$$\left[\begin{array}{cc} \left[\begin{array}{cc} \hat{C} & \hat{\alpha} \end{array}\right] & \hat{\Lambda} \end{array}\right]\left[\begin{array}{cc} \langle V_t^{ex}(\beta,\theta), V_t^{ex}(\beta,\theta)\rangle & \tilde{B}(\beta,\theta) \\ \tilde{B}(\beta,\theta)' & 0 \end{array}\right] = \left[\begin{array}{cc} \langle \mathbf{\Delta}_1 y_t, V_t^{ex}(\beta,\theta)\rangle & I_s \end{array}\right],$$

[2] In the sense that the set of vectors θ such that $\underline{B}(\theta)$ has full row rank is an open and dense subset of the parameter set Θ using the parameterization of de Matos Ribeiro et al. (2020); this follows since the set of matrices B with full row rank is generic within the set of all matrices in $\mathbb{R}^{n\times s}$ and $\underline{A}$ is stable due to the assumptions such that $I_n - \underline{A}$ is non-singular.

where $\hat{\Lambda}$ is the Lagrange multiplier matrix and $V_t^{ex}(\beta,\theta) := [v_t(\theta)', (\beta' y_{t-1})']'$. Next, define the inverse

$$\begin{bmatrix} H_{11}(\beta,\theta) & H_{12}(\beta,\theta) \\ H_{21}(\beta,\theta) & H_{22}(\beta,\theta) \end{bmatrix} := \begin{bmatrix} \langle V_t^{ex}(\beta,\theta), V_t^{ex}(\beta,\theta)\rangle & \tilde{B}(\beta,\theta) \\ \tilde{B}(\beta,\theta)' & 0 \end{bmatrix}^{-1}.$$

Existence is due to Lemma 3 in the appendix. Using these definitions Ribarits and Hanzon (2014) obtain the following residuals after concentrating out C and α (respecting the linear restriction):

$$\varepsilon_t^R(\beta,\theta) := \mathbf{\Delta}_1 y_t - \left(\langle \mathbf{\Delta}_1 y_t, V_t^{ex}(\beta,\theta)\rangle H_{11}(\beta,\theta) - H_{21}(\beta,\theta)\right) V_t^{ex}(\beta,\theta),$$

which yields the following concentrated pseudo log-likelihood function

$$L_T^{ex,R}(\beta,\theta) := -\frac{T}{2}\log\left|\langle \varepsilon_t^R(\beta,\theta), \varepsilon_t^R(\beta,\theta)\rangle\right|.$$

2.3.3 Inference on Cointegrating Spaces

Let $\hat{\beta} := \hat{\beta}(\theta)$ denote a maximizer of the respective pseudo log-likelihood function given θ such that

$$L_T^{ex,U}(\hat{\beta},\theta) = \max_{\beta\in\mathbb{R}^{s\times r}} L_T^{ex,U}(\beta,\theta).$$

This maximizer is not unique, as $\hat{\beta}M$, where $M \in \mathbb{R}^{r\times r}$ is an invertible but otherwise arbitrary matrix, also maximizes the pseudo log-likelihood function. Define

$$\bar{\beta}_\circ := \beta_\circ(\beta_\circ'\beta_\circ)^{-1} \quad \tilde{\beta} := \tilde{\beta}(\theta) := \hat{\beta}(\theta)(\bar{\beta}_\circ'\hat{\beta}(\theta))^{-1}$$

such that $\beta_\circ'(\tilde{\beta} - \beta_\circ) = 0$. This introduces a normalization of $\hat{\beta}$, since $\hat{\beta}(\bar{\beta}_\circ'\hat{\beta})^{-1} = \hat{\beta}M(\bar{\beta}_\circ'\hat{\beta}M)^{-1}$. Analogously, for given θ, we define $\tilde{\beta}^R := \tilde{\beta}^R(\theta)$ as the normalized maximizer of the pseudo log-likelihood function under the restricted concentration step

$$L_T^{ex,R}(\beta,\theta) = -\frac{T}{2}\log\left|\langle \varepsilon_t^R(\beta,\theta), \varepsilon_t^R(\beta,\theta)\rangle\right|,$$

under the condition of $\tilde{B}(\beta,\theta)$ being of full column rank. Under appropriate assumptions, made precise in Theorem 7 below, the estimators $\tilde{\beta}(\hat{\theta}), \tilde{\beta}^R(\hat{\theta}) \in \mathbb{R}^{s\times r}$ are consistent and their asymptotic distribution is asymptotically mixed Gaussian:

$$\left.\begin{matrix} TC_{1,\circ}'(\tilde{\beta}(\hat{\theta}) - \beta_\circ) \\ TC_{1,\circ}'(\tilde{\beta}^R(\hat{\theta}) - \beta_\circ) \end{matrix}\right\} \xrightarrow{d} \left(\int_0^1 FF'\mathrm{d}u\right)^{-1}\int_0^1 F(\mathrm{d}V)'$$

where $F = B_{1,\circ}W$, $V = (\alpha_\circ'\Sigma^{-1}\alpha_\circ)^{-1}\alpha_\circ'\Sigma^{-1}W$, where W is a $s-r$-dimensional standard Brownian motion.
We discuss three different options to obtain inference on the cointegrating spaces:

1. (Pseudo-)Likelihood ratio testing in the unrestricted model using a PML estimator $\hat{\theta}$.
2. (Pseudo-)Likelihood ratio testing in the restricted model using a PML estimator $\hat{\theta}$.
3. (Pseudo)-likelihood ratio testing in the state space framework using two PML estimators $\hat{\theta}_0$ and $\hat{\theta}_1$ over appropriately chosen sets.

Let us discuss the second option in more detail. Considering inference on the cointegrating spaces, the above convergence result for $\tilde{\beta}^R(\hat{\theta})$ implies that pseudo likelihood ratio test statistics regarding linear hypotheses on the long-run coefficients – e. g., of the type $H_0 : \beta = H\varphi$ with $H \in \mathbb{R}^{s\times t}$, $\varphi \in \mathbb{R}^{t\times(s-c)}$, $t < s$ – are asymptotically χ^2, i. e., under the null hypothesis it holds that

$$\tau_T^c := 2\left(L_T^{ex,R}(\tilde{\beta}^R(\hat{\theta}_n^c), \hat{\theta}_n^c) - L_T^{ex,R}(\tilde{\beta}^H(\hat{\theta}_n^c), \hat{\theta}_n^c)\right) \xrightarrow{d} \chi_m^2,$$

where $\tilde{\beta}^H(\hat{\theta}_n^c)$ denotes the normalized maximizer of $L_T^{ex,R}(\beta, \hat{\theta}_n^c)$ under the restrictions of the null hypothesis and $\hat{\theta}_n^c$ is the pseudo maximum likelihood estimator over the parameter space Θ_n^c. Denote by $\Theta_n^h \subset \Theta_n^c$ the set of the parameter vectors in Θ_n^c fulfilling the restrictions of the null hypothesis. Under the null hypothesis, $\hat{\theta}_n^h := \operatorname{argmax}_{\theta\in\Theta_n^h} L_T(\theta) = \operatorname{argmax}_{\theta\in\Theta_n^c} L_T^{ex,R}(\tilde{\beta}^H(\theta), \theta)$ is also a consistent estimator of the true parameter vector, and it holds that

$$\tau_T^h := 2\left(L_T^{ex,R}(\tilde{\beta}^R(\hat{\theta}_n^h), \hat{\theta}_n^h) - L_T^{ex,R}(\tilde{\beta}^H(\hat{\theta}_n^h), \hat{\theta}_n^h)\right) \xrightarrow{d} \chi_m^2.$$

Both these variants of the second option imply that the same result holds for the third option, i. e., the pseudo likelihood ratio test statistic of the null hypothesis $H_0 : \theta \in \Theta_n^h$ against the alternative $H_1 : \theta \in \Theta_n^c \setminus \Theta_n^h$. Thus, under the null hypothesis

$$\begin{aligned}\tau_T^{LR} &:= 2\left(L_T^{ex,R}(\tilde{\beta}^R(\hat{\theta}_n^c), \hat{\theta}_n^c) - L_T^{ex,R}(\tilde{\beta}^H(\hat{\theta}_n^h), \hat{\theta}_n^h)\right)\\ &= 2\left(L_T(\hat{\theta}_n^c) - L_T(\hat{\theta}_n^h)\right) \xrightarrow{d} \chi_m^2,\end{aligned}$$

since $L_T^{ex,R}(\tilde{\beta}^R(\hat{\theta}_n^c), \hat{\theta}_n^c) > L_T^{ex,R}(\tilde{\beta}^R(\hat{\theta}_n^h), \hat{\theta}_n^h)$ and $L_T^{ex,R}(\tilde{\beta}^H(\hat{\theta}_n^c), \hat{\theta}_n^c) < L_T^{ex,R}(\tilde{\beta}^H(\hat{\theta}_n^h), \hat{\theta}_n^h)$ imply

$$\tau_T^h \leq \tau_T^{LR} \leq \tau_T^c.$$

To compute $L_T(\hat{\theta}_n^h)$, the reparameterizations presented in de Matos Ribeiro et al. (2020, Section 5.1.2) for different types of hypotheses on the cointegrating space at frequency zero can be directly implemented, avoiding computations of the SSECM. The degrees of freedom m are derived analogously to the results in the VECM model. It follows that each of the three approaches can be used for inference that is asymptotically equivalent to using the pseudo-likelihood ratio tests in the sense of jointly accepting or rejecting under the null hypothesis.

2.3.4 Inference on Cointegrating Ranks

The maximum of the pseudo log-likelihood function L_T can alternatively be described explicitly as a function dependent on r:

$$\max_{\beta\in\mathbb{R}^{s\times r}, \alpha\in\mathbb{R}^{s\times r}} L_T^{ex,U}(\alpha\beta', \theta) = -\frac{T}{2}\log\left(|S_{00}(\theta)|\prod_{i=1}^{r}(1-\lambda_i(\theta))\right),$$

where $\lambda_1(\theta) \geq \lambda_2(\theta) \geq \cdots \geq \lambda_s(\theta) \geq 0$ are the ordered solutions of

$$|\lambda S_{11}(\theta) - S_{10}(\theta)S_{00}(\theta)^{-1}S_{01}(\theta)| = 0.$$

For the restricted concentration step a closed form solution does not exist in general. In this case define

$$L_T^{rank}(r, \theta) := \max_{\beta\in\mathbb{B}(r,\theta)} -\frac{T}{2}\log\left|\left\langle\varepsilon_t^R(\beta, \theta), \varepsilon_t^R(\beta, \theta)\right\rangle\right|.$$

The likelihood ratio test statistics for the hypothesis $H(r) : \mathrm{rk}(\Pi) = r$ versus $H(s) : \mathrm{rk}(\Pi) = s$ as functions dependent on θ in the unrestricted and the restricted approach are given below:

$$\begin{aligned}-2\log Q_T^U(H(r)/H(s), \theta) &:= -T\sum_{i=r+1}^{s}\log(1-\lambda_i(\theta)),\\ -2\log Q_T^R(H(r)/H(s), \theta) &:= L_T^{rank}(s, \theta) - L_T^{rank}(r, \theta).\end{aligned}$$

Under appropriate assumptions on the choice of θ (see Theorem 8 below), both rank test statistics have limiting distributions which can be expressed in terms of an $s-r$-dimensional standard Brownian motion W as

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

Below it is shown that this convergence of the likelihood ratio test holds for a number of different estimators $\hat{\theta}$. In particular it holds for the pseudo maximum likelihood estimator $\hat{\theta}_n$ obtained by maximizing the likelihood over the parameter set Θ_n. Choosing this parameter space does not impose any restrictions on the rank of $k^{-1}(1)$ or equivalently on the number of eigenvalues at 1 for the matrix A in the state space realisation in the sense that the closure of the corresponding set of transfer functions contains all transfer functions corresponding to $I(1)$ processes of order at most n.

Alternatively we can use the pseudo maximum likelihood estimator $\hat{\theta}_n^{s-r,0}$ obtained from maximizing the pseudo likelihood over $\Theta_n^{s-r,0}$. Note that

$$L_T^{rank}(s,\hat{\theta}_n) = L_T(\hat{\theta}_n) \quad \text{and} \quad L_T^{rank}(r,\hat{\theta}_n^{s-r,0}) = L_T(\hat{\theta}_n^{s-r,0}),$$

Therefore,

$$\begin{aligned}
-2\log Q_T^R(H(r)/H(s),\hat{\theta}_n^{s-r,0}) &= -2\left(L_T^{rank}(r,\hat{\theta}_n^{s-r,0}) - L_T^{rank}(s,\hat{\theta}_n^{s-r,0})\right)\\
&\leq -2\left(L_T(\hat{\theta}_n^{s-r,0}) - L_T(\hat{\theta}_n)\right)\\
&\leq -2\left(L_T^{rank}(r,\hat{\theta}_n) - L_T^{rank}(s,\hat{\theta}_n)\right) = -2\log Q_T^R(H(r)/H(s),\hat{\theta}_n).
\end{aligned}$$

Since the left and the right hand side of this chain of inequalities converge to the same limiting distribution, the pseudo likelihood ratio test statistic

$$-2\left(L_T(\hat{\theta}_n^{s-r,0}) - L_T(\hat{\theta}_n)\right)$$

has the same asymptotic limit, similarly to the discussion at the end of the previous section.

If we include deterministic variables and use a SSECM the limiting distribution changes in accordance to the findings of Johansen (1996).

2.4 MFI(1) processes

2.4.1 Error Correction Representation in the State Space Framework

In this section the results for the I(1) case are extended to the MFI(1) case. This turns out to be notationally more complicated. However, the main ideas remain the same as in the I(1) case: First the SSECM is given which is subsequently used for the pseudo likelihood maximization relating the PML estimation to the framework of Johansen and Schaumburg (1999).

Thus, consider a seasonally integrated MFI(1) process $\{y_t\}_{t\in\mathbb{Z}}$ with its unit root frequencies contained in the set $\{\omega_k := \frac{k-1}{S}2\pi : k = 1,\ldots,\tilde{S} := S/2+1\}$ for some even integer S and recall $z_k := e^{i\omega_k}$. Let $\{y_t\}_{t\in\mathbb{Z}}$ be generated by a minimal state space system according to the assumptions stated in Assumption 1. Analogously to the $I(1)$ case we obtain an SSECM representation:

Theorem 6 (SSECM-MFI(1)) *For every state space system* (A,B,C)*, satisfying* $\det(I_n - \underline{A}^S) \neq 0$*, the residuals*

$$\varepsilon_t(A,B,C) = \sum_{j=0}^{t-1} K_j^-(y_{t-j} - Ds_{t-j})$$

with $K_0^- = I_s, K_j^- = -C\underline{A}^{j-1}B$, $j \in \mathbb{N}$, *have the following representation (using* $\tilde{y}_t := y_t - Ds_t$*):*

$$\begin{aligned}
\mathbf{\Delta}_S \tilde{y}_t &= \sum_{k=1}^{S} \Pi_k X_t^{(k)} + Cv_t + \varepsilon_t(A,B,C), \qquad X_t^{(k)} := \mathbf{\Delta}_{S,k}\tilde{y}_{t-1} \\
v_{t+1} &= \underline{A}v_t - (I_n - \underline{A}^S)^{-1}\underline{A}^S B \mathbf{\Delta}_S \tilde{y}_t, \quad v_1 = x_1, \\
\Pi_k &= \left(-I_s + z_k C(I_n - z_k\underline{A})^{-1}B\right).
\end{aligned}$$

If the process $\{y_t\}_{t\in\mathbb{Z}}$ *fulfills the assumptions stated in Assumption 1 with corresponding system* (A_0, B_0, C_0) *then for* $D = D_\circ$

$$\varepsilon_t(A_\circ, B_\circ, C_\circ) = \varepsilon_t.$$

Furthermore, let the deterministics be rewritten as $Ds_t = [d_1, \dots, d_S, d_{S+1}][s_{t,1}, \dots, s_{t,S}, t]'$ *where* $s_{t,k} = \overline{z_k}^{t-1}$ *and* $d_k := 1/2(d_k^r + id_k^i)$ *and* $d_{S+2-k} := 1/2(d_k^r - id_k^i)$ *for* $k = 2, \dots \tilde{S} - 1$*. Choose a factorization* $\Pi_k := -k^{-1}(z_k) = \alpha_k \beta_k'$*, where* $\alpha_k, \beta_k \in \mathbb{C}^{s\times r}$*, and define* $\Psi = \left.\frac{\partial k^{-1}(z)}{\partial z}\right|_{z=1} = -C(I_n - \underline{A})^{-2}B$*.*
The state space error correction model (SSECM-MFI(1)) is then given by

$$\begin{aligned}
\mathbf{\Delta}_S y_t &= \sum_{k=1}^{S} \Pi_k X_t^{(k)} + Cv_t + \zeta_1 + \sum_{k=2}^{S} \alpha_k \zeta_k' s_{t,k} + \alpha_1 \zeta_{S+1} t + \tilde{\varepsilon}_t(A,B,C,D) \qquad (2.12) \\
v_{t+1} &= \underline{A}v_t - (I_n - \underline{A}^S)^{-1}\underline{A}^S B \mathbf{\Delta}_S y_t, \\
v_1 &= x_1 - \sum_{k=1}^{S} z_k (I_n - z_k\underline{A})^{-1}Bd_k - (I_n - \underline{A})^{-2}Bd_{S+1} + (I_n - \underline{A})^{-1}Bd_{S+1}, \\
\Pi_k &= \left(-I_s + z_k C(I_n - z_k\underline{A})^{-1}B\right),
\end{aligned}$$

where $\zeta_1 := \Pi_1 d_1 + \Psi d_{S+1}$*,* $\alpha_k\zeta_k = \Pi_k d_k$*, for* $k = 2, \dots, S$*,* $\alpha_1\zeta_{S+1} = \Pi_1 d_{S+1}$ *and*

$$\tilde{\varepsilon}_t(A_\circ, B_\circ, C_\circ, D_\circ) = \varepsilon_t.$$

The theorem is proven in the appendix. The SSECM is a special case for $S = 1$.
If $\{y_t\}_{t\in\mathbb{Z}}$ is cointegrated at frequency ω_k this implies that the matrix Π_k is of reduced rank r_k such that it is common to write $\Pi_k = \alpha_k\beta_k'$ with two matrices $\alpha_k, \beta_k \in \mathbb{C}^{s\times r_k}$. Note that by definition of a seasonally integrated process not all frequencies ω_k for $k = 0, \dots, S$ need to be unit root frequencies. For those ω_k which are not unit root frequencies we have $r_k = s$. The real unit roots $z = \pm 1$ correspond to indices 1 and $\tilde{S}$.

2.4.2 Concentration of the Gaussian Pseudo Log-Likelihood Function

In the following we start with $D = D_\circ = 0$ to simplify the already complex notation. Consider the MFI(1) error correction representation

$$\begin{aligned}
\mathbf{\Delta}_S y_t &= \sum_{k=1}^{S} \Pi_k X_t^{(k)} + Cv_t(\theta) + \varepsilon_t(\theta) \\
&= \Pi_1 X_t^{(1)} + \sum_{k=2}^{\tilde{S}-1} 2\mathcal{R}(\Pi_k X_t^{(k)}) + \Pi_{\tilde{S}} X_t^{(\tilde{S})} + Cv_t(\theta) + \varepsilon_t(\theta)
\end{aligned}$$

Assume that we are interested in the cointegrating relations at frequency $\omega_0 = \frac{k_0-1}{S}2\pi$, $k_0 \in \mathbb{N}$ with $1 < k_0 < S/2+1$, such that $0 < \omega_0 < \pi$ and $e^{i\omega_0}$ is a complex unit root. In order to parallel the standard notation introduced by Johansen and Schaumburg (1999) we introduce real valued notation for the complex quantities:

Definition 14 *For a matrix $M \in \mathbb{C}^{k\times l}$ define*

$$[M]^{\mathbb{R}} := \begin{bmatrix} \mathcal{R}(M) & -\mathcal{I}(M) \\ \mathcal{I}(M) & \mathcal{R}(M) \end{bmatrix}$$

and note that for $N \in \mathbb{C}^{l\times m}$ the mapping also preserves multiplication $[MN]^{\mathbb{R}} = [M]^{\mathbb{R}}[N]^{\mathbb{R}}$. Furthermore for a vector $X \in \mathbb{C}^l$ we define $[X]^{\mathbb{R}}_v = [\mathcal{R}(X)', \mathcal{I}(X)']'$.

Using this notation we rewrite $\mathcal{R}(\Pi_{k_0} X_t^{(k_0)})$ as $[I_s, 0][\Pi_{k_0}]^{\mathbb{R}}[X_t^{(k_0)}]^{\mathbb{R}}_v$. It follows that

$$\mathbf{\Delta}_S y_t = \Pi_1 X_t^{(1)} + \sum_{k=2}^{\tilde{S}-1} [I_s, 0][\Pi_k]^{\mathbb{R}}[2X_t^{(k)}]^{\mathbb{R}}_v + \Pi_{\tilde{S}} X_t^{(\tilde{S})} + C v_t(\theta) + \varepsilon_t(\theta).$$

Defining

$$\begin{aligned} Z_{0,t} &:= \mathbf{\Delta}_S y_t, \quad Z_{1,t} := [2X_t^{(k_0)}]^{\mathbb{R}}_v, \quad \mathbf{\Pi}_{k_0} := [I_s, 0][\Pi_{k_0}]^{\mathbb{R}} \\ Z_{2,t} &:= [X_t^{(1)\prime}, X^{(\tilde{S})\prime}, (2[X_t^{(2)}]^{\mathbb{R}}_v)', \ldots, ([2X_t^{(k_0-1)}]^{\mathbb{R}}_v)', ([2X_t^{(k_0+1)}]^{\mathbb{R}}_v)', \ldots, ([2X_t^{(\tilde{S}-1)}]^{\mathbb{R}}_v)']', \\ \mathbf{\Pi}_{-k_0} &:= [\Pi_1, \Pi_{\tilde{S}}, \mathbf{\Pi}_2, \ldots, \mathbf{\Pi}_{k_0-1}, \mathbf{\Pi}_{k_0+1}, \ldots, \mathbf{\Pi}_{\tilde{S}-1}] \end{aligned}$$

and

$$V_t(\theta) := [Z_{2,t}', v_t(\theta)']', \qquad C_{-k_0} := [\mathbf{\Pi}_{-k_0}, C]$$

the residuals can be written as $\varepsilon_t(\theta) = Z_{0,t} - \mathbf{\Pi}_{k_0} Z_{1,t} - C_{-k_0} V_t(\theta)$. In this case the concentrated pseudo log-likelihood function (up to a constant and assuming $x_1 = 0$) is given by

$$L_T(\theta) = -\frac{T}{2} \log \left| \hat{\Sigma}_T(\theta) \right| = -\frac{T}{2} \log \left| \langle \varepsilon_t(\theta), \varepsilon_t(\theta) \rangle \right|.$$

Expand the pseudo log-likelihood function as in the I(1) case:

$$L_T^{ex}(\mathbf{\Pi}_{k_0}, C_{-k_0}, \theta) := -\frac{T}{2} \log \left| \langle Z_{0,t} - \mathbf{\Pi}_{k_0} Z_{1,t} - C_{-k_0} V_t(\theta), Z_{0,t} - \mathbf{\Pi}_{k_0} Z_{1,t} - C_{-k_0} V_t(\theta) \rangle \right|.$$

Again for the expanded pseudo log-likelihood function we have two possibilities to estimate C. We can employ an **unrestricted concentration approach**, which parallels Johansen's methods. We regress $Z_{0,t}$ and $Z_{1,t}$ on $V_t(\theta)$ where all involved quantities are real. Thus, we derive the following residuals:

$$\begin{aligned} R_{0,t}(\theta) &:= Z_{0,t} - \langle Z_{0,t}, V_t(\theta) \rangle \langle V_t(\theta), V_t(\theta) \rangle^{-1} V_t(\theta) \\ R_{1,t}(\theta) &:= Z_{1,t} - \langle Z_{1,t}, V_t(\theta) \rangle \langle V_t(\theta), V_t(\theta) \rangle^{-1} V_t(\theta). \end{aligned}$$

Assuming $\Pi_{k_0} = \alpha\beta'$ with $\alpha, \beta \in \mathbb{C}^{s\times r}$ of full rank (omitting the indices for $\alpha_{k_0}, \beta_{k_0}$), we continue to use the notation of Johansen and Schaumburg (1999) and introduce the following matrices:

$$\boldsymbol{\beta} := [\beta]^{\mathbb{R}} \quad \boldsymbol{\alpha} := [\alpha]^{\mathbb{R}} \quad \breve{\alpha} := [I_s, 0]\boldsymbol{\alpha}.$$

Consequently, $\mathbf{\Pi}_{k_0} = \breve{\alpha}\boldsymbol{\beta}'$. Concentrating out $\breve{\alpha}$ using the OLS estimator for given $\boldsymbol{\beta}$ we find that

$$L_T^{ex,U}(\boldsymbol{\beta}, \theta) = -\frac{T}{2} \log \left(|S_{00}(\theta)| \frac{|\boldsymbol{\beta}' S_{11,0}(\theta)\boldsymbol{\beta}|}{|\boldsymbol{\beta}' S_{11}(\theta)\boldsymbol{\beta}|} \right),$$

where we define the matrices in accordance to the approach of Johansen and Schaumburg (1999):

$$\begin{aligned} S_{00}(\theta) &:= \langle R_{0,t}(\theta), R_{0,t}(\theta) \rangle, & S_{10}(\theta) &:= \langle R_{1,t}(\theta), R_{0,t}(\theta) \rangle =: S_{10}(\theta)', \\ S_{11}(\theta) &:= \langle R_{1,t}(\theta), R_{1,t}(\theta) \rangle, & S_{11,0}(\theta) &:= S_{11}(\theta) - S_{10}(\theta) S_{00}^{-1}(\theta) S_{01}(\theta). \end{aligned}$$

Using a **restricted concentration approach** alternatively, we concentrate out C under the restrictions $\Pi_k = \left(-I_s + z_k C(I_n - z_k \underline{A}(\theta))^{-1} B(\theta)\right)$ for $0 \le k \le S$. Define

$$\begin{aligned} \underline{B}_k(\theta) &:= z_k(I - z_k\underline{A}(\theta))^{-1}B(\theta), \qquad \underline{B}_k^{\mathbb{R}}(\theta) := [I_n, 0][\underline{B}_k(\theta)]^{\mathbb{R}}, \qquad \text{for } k = 1, \ldots, \tilde{S} \\ \underline{B}_{-k}^{\mathbb{R}}(\theta) &:= \left[\underline{B}_1(\theta), \underline{B}_{\tilde{S}}(\theta), \underline{B}_2^{\mathbb{R}}(\theta), \ldots, \underline{B}_{k-1}^{\mathbb{R}}(\theta), \underline{B}_{k+1}^{\mathbb{R}}(\theta), \ldots, \underline{B}_{\tilde{S}-1}^{\mathbb{R}}(\theta)\right], \\ I^{\mathbb{R}} &:= [I_s, I_s, [I_s, 0], \ldots, [I_s, 0]] \in \mathbb{R}^{s \times s(S-2)}. \end{aligned}$$

The problem to be solved is to maximize

$$L_T^{ex}(\breve{\alpha}\boldsymbol{\beta}', C_{-k_0}, \theta) = L_T^{ex}(\breve{\alpha}\boldsymbol{\beta}', [\boldsymbol{\Pi}_{-k_0}, C], \theta)$$

subject to

$$J := \left[[I_s, 0] \quad I^{\mathbb{R}}\right] = \left[\begin{array}{ccc} \boldsymbol{\Pi}_{-k_0} & C & \breve{\alpha} \end{array}\right] G(\boldsymbol{\beta}, \theta), \qquad \text{where } G(\boldsymbol{\beta}, \theta) := \begin{bmatrix} 0 & I_{S-2} \\ \underline{B}_{k_0}^{\mathbb{R}}(\theta) & \underline{B}_{-k_0}^{\mathbb{R}}(\theta) \\ -\boldsymbol{\beta}' & 0 \end{bmatrix}$$

for given θ and $\boldsymbol{\beta} \in \mathbb{G}(r, \theta) : \{\boldsymbol{\beta} = [\beta]^{\mathbb{R}} : \beta \in \mathbb{C}^{s \times r}, G(\boldsymbol{\beta}, \theta) \text{ of full column rank}\}$. Note that as in the I(1) case, an additional condition is needed to ensure the existence of solutions, as in some cases θ introduces restrictions on the right kernel of the corresponding Π_{k_0}. Solving for $C_{k_0}^{ex} := \left[\begin{array}{ccc} \boldsymbol{\Pi}_{-k_0} & C & \breve{\alpha} \end{array}\right]$, we arrive at a quite involved system of first order conditions of the form

$$\left[C_{k_0}^{ex,R}(\boldsymbol{\beta}, \theta) \quad \Lambda\right] \begin{bmatrix} \langle V_t^{ex}(\boldsymbol{\beta}, \theta), V_t^{ex}(\boldsymbol{\beta}, \theta)\rangle & G(\boldsymbol{\beta}, \theta) \\ G(\boldsymbol{\beta}, \theta)' & 0 \end{bmatrix} = \left[-\langle Z_{0,t}, V_t^{ex}(\boldsymbol{\beta}, \theta)\rangle \quad J\right],$$

where $C_{k_0}^{ex,R}(\boldsymbol{\beta}, \theta) := \left[\begin{array}{cc} C_{-k_0}^R(\boldsymbol{\beta}, \theta) & \breve{\alpha}^R(\boldsymbol{\beta}, \theta) \end{array}\right]$ is the maximizer of the likelihood, Λ is the corresponding Lagrange multiplier matrix and $V_t^{ex}(\boldsymbol{\beta}, \theta) := [Z_{2,t}', v_t(\theta)', (\boldsymbol{\beta}' Z_{1,t})']'$. As before, in a next step we define the blocks of the inverse matrix of interest by

$$\begin{bmatrix} H_{11}(\boldsymbol{\beta}, \theta) & H_{12}(\boldsymbol{\beta}, \theta) \\ H_{21}(\boldsymbol{\beta}, \theta) & H_{22}(\boldsymbol{\beta}, \theta) \end{bmatrix} := \begin{bmatrix} \langle V_t^{ex}(\boldsymbol{\beta}, \theta), V_t^{ex}(\boldsymbol{\beta}, \theta)\rangle & G(\boldsymbol{\beta}, \theta) \\ G(\boldsymbol{\beta}, \theta)' & 0 \end{bmatrix}^{-1},$$

where existence follows from Lemma 3 of the appendix. Hence, we find matrices $C_{k_0}^{ex,R}(\boldsymbol{\beta}, \theta)$ equal to

$$C_{k_0}^{ex,R}(\boldsymbol{\beta}, \theta) = \langle Z_{0,t}, V_t^{ex}(\boldsymbol{\beta}, \theta)\rangle H_{11}(\boldsymbol{\beta}, \theta) - J H_{21}(\boldsymbol{\beta}, \theta),$$

which maximizes the pseudo log-likelihood function under the given restrictions. Using this notation let the corresponding residuals be defined as

$$\varepsilon_t^R(\boldsymbol{\beta}, \theta) := Z_{0,t} - C_{k_0}^{ex,R}(\boldsymbol{\beta}, \theta) V_t^{ex}(\boldsymbol{\beta}, \theta),$$

The logarithm of the Gaussian pseudo log-likelihood function up to a constant can be written as

$$L_T^{ex,R}(\boldsymbol{\beta}, \theta) = -\frac{T}{2} \log \left|\left\langle \varepsilon_t^R(\boldsymbol{\beta}, \theta), \varepsilon_t^R(\boldsymbol{\beta}, \theta)\right\rangle\right|.$$

Solutions to the problem of maximizing with respect to $\boldsymbol{\beta}$ have been given in Johansen and Schaumburg (1999) for the VAR setting. Note that in both concentration steps the rank of Π_{k_0} is considered to be restricted, but possible rank constraints on all other matrices Π_j, $j \neq k_0$, are not retained. It should be noted, however, that for most results below we only use consistency of $(\hat{\underline{A}}, \hat{B})$, which also holds for additional rank constraints on Π_j, $j \neq k_0$. Thus, inference can be achieved whether or not the system correctly specifies the unit roots present at other locations. We will see that asymptotically the estimation of the cointegrating spaces and ranks and the asymptotic results for the pseudo likelihood ratio tests are not affected by the specification at other unit roots. The effects of the inclusion of these specifications on the power properties of the tests are not further investigated.

2.4.3 Inference on Cointegrating Ranks and Spaces

Note that both the unrestricted and the restricted concentration approach do not lead to a representation of the pseudo log-likelihood function as a function of eigenvalues of a certain matrix, as in the case of real unit roots, since there are further restrictions placed on $\boldsymbol{\beta}$. Let $\hat{\boldsymbol{\beta}}$ denote a maximizer of the respective pseudo log-likelihood function $L_T^{ex,U}(\boldsymbol{\beta},\theta)$. Similarly let $\hat{\boldsymbol{\beta}}^R$ denote a maximizer of

$$L_T^{ex,R}(\boldsymbol{\beta},\theta) = -\frac{T}{2}\log\left|\langle \varepsilon_t^R(\boldsymbol{\beta},\theta), \varepsilon_t^R(\boldsymbol{\beta},\theta)\rangle\right|$$

in $\mathbb{G}(r,\theta)$. Let $\Pi_{k_0,\circ} = \alpha_{k_0,\circ}\beta_{k_0,\circ}'$ denote the true matrices of the SSECM corresponding to the data generating process and define $\boldsymbol{\beta}_\circ := [\beta_{k_0,\circ}]^{\mathbb{R}}$ and $\boldsymbol{\alpha}_\circ := [\alpha_{k_0,\circ}]^{\mathbb{R}}$. To construct a unique maximizer, consider the normalization

$$\bar{\boldsymbol{\beta}}_\circ := \boldsymbol{\beta}_\circ(\boldsymbol{\beta}_\circ'\boldsymbol{\beta}_\circ)^{-1} \quad \tilde{\boldsymbol{\beta}} := \hat{\boldsymbol{\beta}}(\bar{\boldsymbol{\beta}}_\circ'\hat{\boldsymbol{\beta}})^{-1}$$

such that $\boldsymbol{\beta}_\circ'(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_\circ) = 0$. Similarly, define $\tilde{\boldsymbol{\beta}}^R$. Let $\mathcal{B}_\omega := [\mathcal{B}_{k_0,\circ}]^{\mathbb{R}}$ and $\mathcal{C}_\omega := [\mathcal{C}_{k_0,\circ}]^{\mathbb{R}}$. The asymptotic distribution of $T\mathcal{C}_\omega'(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_\circ)$ and $T\mathcal{C}_\omega'(\tilde{\boldsymbol{\beta}}^R - \boldsymbol{\beta}_\circ)$ is then given by the following theorem.

Theorem 7 *Let $\{y_t\}_{t\in\mathbb{Z}}$ be a seasonally integrated MFI(1) process generated according to the assumptions stated in Assumption 1, and assume the true order n is known and $D = D_\circ = 0$ such that no deterministic terms are contained neither in the model nor the data generating process. Let $\hat{\theta}$ be the PML estimator over a suitable parameter space Θ fulfilling the assumption of Proposition 2.*
(I) The estimator $\tilde{\boldsymbol{\beta}}(\hat{\theta}) \in \mathbb{R}^{2s\times 2r_\omega}$ based on the unconstrained approach and the estimator $\tilde{\boldsymbol{\beta}}^R(\hat{\theta}) \in \mathbb{R}^{2s\times 2r_\omega}$ based on the constrained approach are consistent and their asymptotic distribution is mixed Gaussian:

$$\left.\begin{array}{l} T\mathcal{C}_\omega'(\tilde{\boldsymbol{\beta}}(\hat{\theta}) - \boldsymbol{\beta}_\circ) \\ T\mathcal{C}_\omega'(\tilde{\boldsymbol{\beta}}^R(\hat{\theta}) - \boldsymbol{\beta}_\circ) \end{array}\right\} \xrightarrow{d} \left(\int_0^1 \boldsymbol{F}\boldsymbol{F}'\,du\right)^{-1}\int_0^1 \boldsymbol{F}(d\boldsymbol{V})'$$

where $\boldsymbol{F} = \mathcal{B}_\omega\boldsymbol{W}$, $\boldsymbol{V} = (\boldsymbol{\alpha}_\circ'\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha}_\circ)^{-1}\boldsymbol{\alpha}_\circ'\boldsymbol{\Sigma}^{-1}\boldsymbol{W}$ and $\boldsymbol{W} = [\frac{1}{\sqrt{2}}(W_1 + iW_2)]^{\mathbb{R}}$, where W_1 and W_2 are two independent s-dimensional Brownian motions with variance Σ.
(II) Consider the null hypothesis $\beta = b$, $b \in \mathbb{C}^{s\times r}$, against the alternative $\beta \neq b$, $\beta \in \mathbb{C}^{s\times r}$. Let $\hat{\theta}_n^c$ be the pseudo maximum likelihood estimator over the parameter space $\Theta_n^{c,\omega}$ corresponding to $M(\{(\omega,c)\}, n-2c)$ and let $\Theta_n^b \subset \Theta_n^{c,\omega}$ denote the set of parameter vectors in $\Theta_n^{c,\omega}$ fulfilling the restriction of the null hypothesis. Define $\hat{\theta}_n^b := argmax_{\theta\in\Theta_n^b} L_T(\theta)$.
Under the null hypothesis, the asymptotic distribution of the pseudo likelihood ratio test statistic $\tau_T^{LR} := L_T(\hat{\theta}_n^c) - L_T(\hat{\theta}_n^b)$ for the hypothesis $\beta = b$ is $\chi^2_{2r(s-r)}$. The same holds for the test using differences in $L_T^{ex,U}$.
(III) Consider the null hypothesis of the identifying restrictions $\beta_k = (H_1\psi_1, \ldots, H_r\psi_r)$ for $\beta \in \mathbb{C}^{s\times r}$, against the alternative $\beta \neq (H_1\psi_1, \ldots, H_r\psi_r)$, $\beta \in \mathbb{C}^{s\times r}$. Let $\Theta_n^h \subset \Theta_n^c$ denote the set of parameter vectors in $\Theta_n^{c,\omega}$ fulfilling the restriction of the null hypothesis. Define $\hat{\theta}_n^h := argmax_{\theta\in\Theta_n^h} L_T(\theta)$.
Under the null hypothesis, the asymptotic distribution of the pseudo likelihood ratio test statistic $\tau_T^{LR} := L_T(\hat{\theta}_n^c) - L_T(\hat{\theta}_n^h)$ for the identifying restrictions $\beta_k = (H_1\psi_1, \ldots, H_r\psi_r)$ normalized by $\beta_i = b_i + H^i\phi_i$ is χ^2 with $2\sum_{i=1}^r(s - r - p_i - 1)$ degrees of freedom , where H_i is $s \times p_i$, provided β is identified. The same holds for the test using differences in $L_T^{ex,U}$.

A proof is given in the appendix. Again, the three different test statistics τ_T^c, τ_T^h and τ_T^{LR} presented in the I(1) case can be considered and the computations of τ_T^{LR} can be implemented by the reparameterizations discussed in Bauer et al. (2020, Section 5.1.2), with the corresponding number of degrees of freedom also given for different types of hypotheses. For the asymptotics of tests on the coefficients β_k at complex valued unit roots the number of degrees of freedom are slightly different than in the case of real unit roots, due to the different number of parameters for

complex matrices.
The result shows that the tests for hypotheses on the cointegrating space share the same asymptotic distribution with the VECM framework. The proof of Theorem 7 also shows that the test can be executed frequency by frequency assuming full rank of the matrices Π_k at other frequencies. Clearly, it is possible to include further specifications of the cointegrating ranks at other frequencies. The effects of these additional specifications on the power properties of the tests are not known to date. Further investigations are left for future research.
Finally, likelihood ratio-type test statistics for the hypothesis $H(r_\omega) : \mathrm{rk}(\Pi_\omega) = r_\omega$ versus $H(s) : \mathrm{rk}(\Pi_\omega) = s$, $\omega \in (0, \pi)$ in the SSECM-MFI(1) model using the unrestricted or the restricted concentration step for appropriate choice of θ are given below:

$$-2\log Q_T^U(H(r_\omega)/H(s),\theta) := -T \log \frac{|\tilde{\boldsymbol{\beta}}' S_{11,0}(\theta)\tilde{\boldsymbol{\beta}}||S_{11}(\theta)|}{|\tilde{\boldsymbol{\beta}}' S_{11}(\theta)\tilde{\boldsymbol{\beta}}||S_{11,0}(\theta)|},$$

$$-2\log Q_T^R(H(r_\omega)/H(s),\theta) := -T\left(\log\left|\left\langle \varepsilon_t^R(\tilde{\boldsymbol{\beta}}^R,\theta), \varepsilon_t^R(\tilde{\boldsymbol{\beta}}^R,\theta)\right\rangle\right| - \log\left|\left\langle \varepsilon_t^R(I_{2s},\theta), \varepsilon_t^R(I_{2s},\theta)\right\rangle\right|\right).$$

Under a suitably chosen θ, the following asymptotic results hold:

Theorem 8 *Let $\{y_t\}_{t\in\mathbb{Z}}$ be a seasonally integrated MFI(1) process generated according to the Assumption 1, and assume the true order is known. For $\omega \in (0,\pi)$ let the true cointegrating rank at frequency ω be $r_\omega = s - c$. Let $\hat{\theta}$ be the PML estimator over a suitable parameter space Θ fulfilling the assumption of Proposition 2. Let, further, $\hat{\theta}_n^{c,\omega}$ denote the PML estimator over $\Theta_n^{c,\omega}$ and $\hat{\theta}_n$ denote the PML estimator over Θ_n.*
(I) If no deterministics are present neither in the data generating process nor the model ($D = D_\circ = 0$), then the limiting distributions under the null hypothesis of the rank test statistics

$$-2\log Q_T^U(H(r_\omega)/H(s),\hat{\theta}),$$
$$-2\log Q_T^R(H(r_\omega)/H(s),\hat{\theta}),$$
$$-2(L_T(\hat{\theta}_n^{c,\omega}) - L_T(\hat{\theta}_n))$$

for the hypothesis $H(r_\omega) : rk(\Pi_\omega) = r_\omega$ against the alternative $H(s) : rk(\Pi_\omega) = s$ can be expressed in terms of two independent $s - r_\omega$-dimensional standard Brownian motions W_1, W_2 as

$$\frac{1}{2} tr\left(\int_0^1 (d\boldsymbol{W})\boldsymbol{W}'\left(\int_0^1 \boldsymbol{W}\boldsymbol{W}' du\right)^{-1}\int_0^1 \boldsymbol{W}(d\boldsymbol{W}')\right),$$

with $\boldsymbol{W} = [\frac{1}{\sqrt{2}}(W_1 + iW_2)]^{\mathbb{R}}$.
(II) If the data generating process and the model have deterministic components of the form $Ds_t = \sum_{k=1}^S d_k s_{t,k}$ or $Ds_t = \sum_{k=1}^S d_k s_{t,k} + d_{S+1}t$, then the limiting distributions of the pseudo likelihood ratio tests under the null hypothesis can be expressed in terms of two independent $s - r_\omega$-dimensional standard Brownian motions W_1, W_2 as

$$\frac{1}{2} tr\left(\int_0^1 (d\boldsymbol{W})\boldsymbol{H}'\left(\int_0^1 \boldsymbol{H}\boldsymbol{H}' du\right)^{-1}\int_0^1 \boldsymbol{H}(d\boldsymbol{W}')\right),$$

with $\boldsymbol{H} = [(\frac{1}{\sqrt{2}}(W_1 + iW_2)', 1)']^{\mathbb{R}}$ and $\boldsymbol{W} = [W_1 + iW_2]^{\mathbb{R}}$.

A proof is given in the appendix. A slight correction of Johansen and Schaumburg (1999) is necessary. The two rank test statistics given there for the cases including a constant restricted or unrestricted to the cointegrating space correspond to the two different variants given in Theorem 8(II). Both lead to the same asymptotic distribution and, therefore, the same critical values. The two different formulas for the asymptotic distributions given in Johansen and Schaumburg (1999) are in fact equal such that the distributions coincide, while the tables for the critical values differ. The critical values given in Johansen and Schaumburg (1999) for the setting of the unrestricted constant in the VECM equations are wrong and are equal to the ones found for the restricted constant. Table 2 should be used for the critical values of both variants of determinstics.

2.5 Simulation Results

In this section we compare the ranktests introduced in the previous section to the ranktests based on a VAR approximation in the Johansen-Schaumburg vector error correction model, compare Johansen and Schaumburg (1999), and the ranktests based on the CCA subspace algorithm introduced in Bauer and Buschmeier (2016) in a simulation study.
The data generating processes used are of the form

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + a_4 y_{t-4} + \varepsilon_t + \lambda \varepsilon_{t-4}$$

with

$$a_1 = \begin{bmatrix} \gamma & 0 \\ 0 & 0 \end{bmatrix}, a_2 = \begin{bmatrix} -0.4 & 0.4-\gamma \\ 0 & 0 \end{bmatrix}, a_3 = \begin{bmatrix} -\gamma & 0 \\ 0 & 0 \end{bmatrix}, a_4 = \begin{bmatrix} 0.6-.1\gamma & 0.4+\gamma \\ 0 & 1 \end{bmatrix}$$

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 $\lambda = \{0, 0.5, 0.9\}$ samples of sizes $140, 144 \ldots, 300$ are generated with initial values set to zero. The first 100 values are discarded, so that we have sample sizes $T = 40, 44, \ldots, 200$. For the simulations 10000 replications are made. All systems have state space unit root structure $((0, 1), (\frac{\pi}{2}, 1), (\pi, 1))$.
With the five algorithms described below the cointegrating rank at frequency $\frac{\pi}{2}$ is determined by testing the null hypothesis of $2s$ stochastic cycles, i.e., cointegrating rank zero against the alternative of fewer than $2s$ stochastic cycles. In case of rejection the null hypothesis of $2(s-1)$ stochastic cycles is tested against the alternative of fewer stochastic cylces. This procedure is continued until a null hypothesis is accepted or until there are zero stochastic cycles under the alternative. The hit rate, i.e., the percentage of correctly specified cointegrating ranks is compared to $1-\alpha$, where the nominal significance level α is chosen to be 0.05.
JS-VECM: For the Johansen Schaumburg procedure the lag length k is chosen by minimizing Akaikes information criterion. Since the data generating process computes quarterly data we have a lower bound of four for the lag length in the VECM such that $\hat{k} = \max = \{4, \hat{k}_{AIC}\}$.
CCA subspace: For CCA we choose $f = p = 2\hat{k}_{AIC}$. The system order n is chosen by minimizing a singular value criterion, see Bauer (2001).
LR-, R-, Q-statistics: For the pseudo maximum likelihood (PML) algorithms used for the restricted (R), the unrestricted (Q) SSECM test and the pseudo likelihood ratio test in the VARMA setting (LR) the output of the subspace algorithm is used as a starting value to find the maximizer $\hat{\theta}_n$ of the pseudo log-likelihood function over Θ_n. The R- and Q-statistics are then computed using $\hat{\theta}_n$ to maximize $L_T^{ex,U}(\boldsymbol{\beta}, \hat{\theta}_n)$ and $L_T^{ex,R}(\boldsymbol{\beta}, \hat{\theta}_n)$ with respect to $\boldsymbol{\beta}$ over $\mathbb{G}(r, \hat{\theta}_n)$. To maximize the likelihood over $\Theta_n^{c,\omega}$ the initial parameter vector $\hat{\theta}_{0,c} \in \Theta_n^{c,\omega}$ corresponding to $\hat{\theta}_n$ and $\tilde{\boldsymbol{\beta}}^R(\hat{\theta}_n)$ is used as a starting value.
In Figure 2.1 we see the hit rates for the VAR case $\lambda = 0$. Here and in the next figure we have excluded the CCA subspace procedure, as its low hit rates for small sample sizes distorts the diagram, making the differences between the other results indistinguishable. Among all other listed procedures the hit rate of the Q-statistic based on unrestricted optimization starts with the lowest hit rate at about 81%. The hit rates of the R-statistic using the restricted SSECM and the likelihood ratio test are lower than the results for Johansen Schaumburg test procedure for samples of size smaller than about 80. It is remarkable that the hit rates of the state-space-based tests for sample sizes larger than 120 are slightly higher than the expected and asymptotically valid $1-\alpha = 95\%$ level. The Johansen Schaumburg test on the other hand exhibit a hit rate lower than the 95% level for all observed sample sizes. Since the Johansen Schaumburg test was designed for VAR systems it is not surprising that the small sample properties of the Johansen Schaumburg procedure are better.

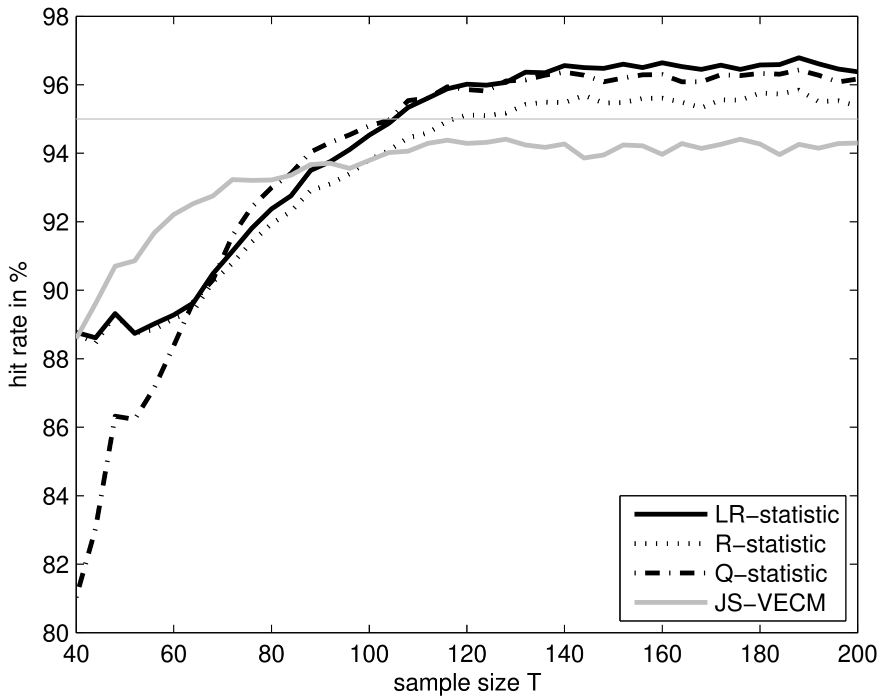

Figure 2.1: Hit rates of the ranktests for the VAR case

The hit rates for $\lambda = 0.5$ are displayed in Figure 2.2. The effect of the inclusion of an MA-part on the state-space-based tests is small. The hit rate of the Johansen Schaumburg test however drops from about 89% in the case $\lambda = 0$ to 75% for sample size $T = 40$. Also, the hit rate of the Johansen Schaumburg test only reaches a hit rate of 85% for sample sizes up to $T = 200$, missing the supposed 95% level by a wide margin. This can be explained by the fact that an MA-part increases the estimate of the lag length k. This in turn increases the number of parameters estimated by the Johansen Schaumburg procedure which affects the small sample properties. Again the state space-based procedures have hit rates higher than 95%. It is notable that the R-statistic using the restricted concentration step and the true pseudo likelihood ratio test achieve hit rates of about 89% even for sample size 40 which in the case of quarterly data corresponds to ten years of observations and the resulting hit rates are basically equal for all observed sample sizes. The Q-statistic based on the unrestricted SSECM is comparable to the Johansen procedure for $T = 40$ and reaches the performance of the other two state-space-based tests for sample sizes of $T = 90$ and higher.
In Figure 2.3 for the case of $\lambda = 0.9$ we observe results of the same kind. Again the restricted SSECM test and the pseudo likelihood ratio test have higher hit rates than the unrestricted one in small samples. This time the effect is observable for samples up to $T = 100$. Again the effects of the MA-polynomial of the true system on the state-space-based tests is small, the best two still reaching hit rates of about 88% for sample size $T = 40$. The hit rates of the Johansen Schaumburg test are even lower than in the case $\lambda = 0.5$. In this figure we have included the hit rates of the CCA subspace test, which has the lowest hit rate for samples up to about size 120. For sample sizes lower than $T = 90$ the hit rate lies below 20%. For larger samples it has about the same hit rates as the SSECM tests. Application of the CCA subspace procedure for $\lambda = 0$ or $\lambda = 0.5$ yield similar results.

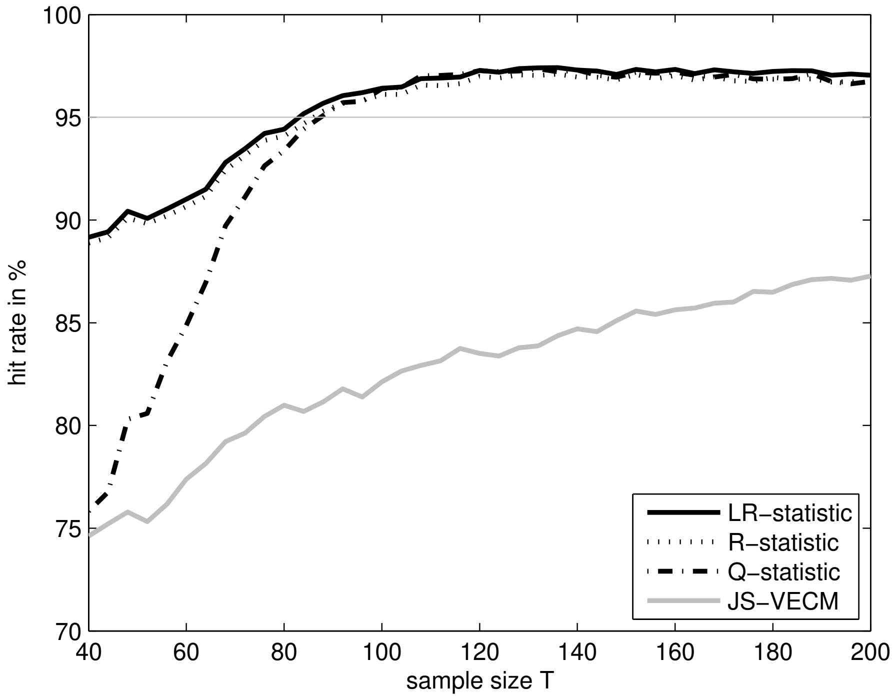

Figure 2.2: Hit rates of the ranktests for the VARMA case with $\lambda = 0.5$

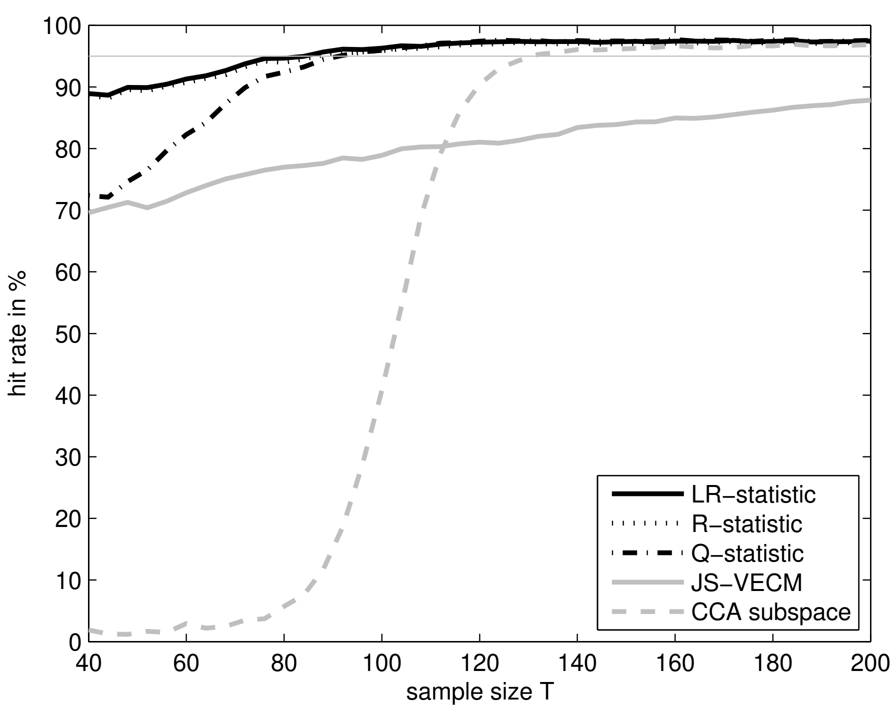

Figure 2.3: Hit rates of the ranktests for the VARMA case with $\lambda = 0.9$

2.6 Summary and Conclusion

In this paper the state space error correction representation for I(1) processes by Ribarits and Hanzon (2014) was extended from the I(1) to the MFI(1) case. This approach was used to derive pseudo likelihood ratio tests for the cointegrating rank and linear hypotheses on the cointegrating space. The asymptotic distribution of these tests was derived.

The main message in this respect is that in the state space framework the tools of the Johansen framework for VAR processes can be used with the only change that the stationary terms in the error correction representation depends on estimated quantities. This does not change the asymptotic inference, though. This implies for example that distributions from the VAR framework can be directly transferred to the state space framework. The same holds for hypothesis tests.

This opens the possibility of three different tests based on pseudo-likelihood ratios both for cointegrating spaces and cointegrating ranks: The unrestricted concentration step is easy to implement and leads to inference totally analogous to the VECM inference. The restricted concentration approach is more involved. Finally the results show that these tests are related to the pseudo-likelihood ratio tests in the state space setting leading to identical asymptotics under the null hypothesis.

A simulation study shows that these rank tests are able to correctly specify the cointegrating rank at complex unit roots more often than the tests based on the Johansen Schaumburg procedure and the CCA subspace algorithm for small samples for VARMA processes. It can be seen, moreover, that the unrestricted concentration step performs slightly worse than the other two approaches, while the performance of the other two methods is practically identically in the simulations. The CCA subspace procedure proofs useful in providing initial estimates, but inference directly in the CCA setting is not optimal.

Thus, this paper extends the toolbox for investigating MFI(1) processes from the VAR framework to the VARMA setting.

Chapter 3

Pseudo Maximum Likelihood Estimation and Inference for I(2) Processes: A State Space Approach

3.1 Introduction

Cointegration analysis for nominal macroeconomic time series seems to indicate that some first differences of common nominal variables such as inflation rates or money growth are well described as autoregressive processes integrated of order one, compare, e. g., King, Plosser, Stock and Watson (1991). Thus, the levels of these variables can be adequately described as I(2) processes.
For I(2) processes in a multivariate setting the VECM introduced by Johansen (1992) is the major workhorse for both the determination of the number of underlying stochastic trends as well as for inference on the different cointegrating spaces occuring in these models. Consequently, the model class is restricted to VAR processes, which for a number of reasons may lead to disadvantages, especially in case of I(2) processes. As discussed in Johansen (1997) there are different ways to parameterize the model to ensure that conditions on the rank of certain matrices are satisfied, which is necessary for the model to correspond to an I(2) process. As no explicit solution for the pseudo maximum likelihood estimator (PMLE) in this restricted setting is available, numerical algorithms are needed, an example being a switching algorithm introduced by Nielsen and Rahbek (2007).
A different approach for cointegration analysis is to employ a state space framework, effectively shifting the model class to VARMA processes. Recently Ribarits and Hanzon (2014) introduced the state space error correction model for I(1) processes, allowing for yet another parameterization using the state space framework. Tools to make the state space framework useful for cointegrated processes were developed by Bauer and Wagner (2012) who introduced a canonical form tailored to make the structure of the underlying stochastic trends visible. This is done by separating the state process into components with different integration orders at different unit roots. The authors of this paper then combined the available results, proposing several parameterizations and characterizing their properties, defining sets with different cointegration structures, and introducing a semi-order showing that some of these sets are included in the closures of others, compare (Bauer et al., 2020). After showing consistency of the PMLE for I(1) processes (de Matos Ribeiro et al., 2020), in this paper we extend the theory to processes integrated of order two.
Our aim is two-fold. First, we highlight the advantages of the state space framework for I(2) processes. It turns out that all information on the different cointegrating spaces is neatly encoded in a single matrix $\mathcal{C}_u$. The state space error correction model for I(2) processes then links the

VECM framework and its reduced rank matrices directly to the matrices of the state space system. Thus, the interconnected restrictions of the VECM for I(2) processes translate into a state space system with an already available parameterization. Finally, for given or estimated state x_t and given matrix $\mathcal{C}_u$ we derive an explicit solution maximizing the pseudo likelihood function of the regression $y_t = Cx_t + \varepsilon_t$ over the set of matrices C corresponding to $\mathcal{C}_u$. Maximizing the likelihood over the set of possible $\mathcal{C}_u$ is then another way to find the PMLE under certain rank restrictions, and also applicable in a VAR setting.

The second aim of this paper is to establish the stochastic properties of the PMLE, showing consistency, deriving the asymptotic distribution and proposing a test for linear hypothesis on the parameters. This lays the groundwork for cointegration analysis for VARMA I(2) processes, extending the tools available in the VECM also to the state space framework. As a final result we show that likelihood ratio statistics testing for the cointegration indices introduced by Paruolo (1996) exhibit the same asymptotic distribution as the corresponding tests in the VECM setting.

The paper is structured as follows. Section 2 contains the necessary definitions of the I(2) processes and cointegration properties considered in this paper. It also introduces the canonical form of the state space representation and the state space error correction model for I(2) processes. We discuss the relations between the different representations, and between the system matrices and the cointegrating spaces. The next section covers the PMLE over different sets of transfer functions, starting with the description of their respective parameter spaces, before the discussion of the consistency result and the asymptotic distribution of the parameter vector. Section 4 illustrates a way to derive explicit solutions for given matrix $\mathcal{C}_u$ of cointegrating properties. This is the basis for the subsequent Theorem dealing with the asymptotic distribution of the rank test statistics under the null. A short simulation study is discussed in Section 5, showing the advantages of using the state space approach to determine the cointegration indices, especially for small sample sizes. Finally, Section 6 summarizes the results. All proofs are relegated to the appendix.

Notation in this paper is as follows: I_s denotes the s-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 $\lambda_{\min}(X)$. L denotes the backshift operator such that $L(\{y_t\}_{t\in\mathbb{Z}}) = \{y_{t-1}\}_{t\in\mathbb{Z}}$ for a process $\{y_t\}_{t\in\mathbb{Z}}$, for brevity written as $Ly_t = y_{t-1}$. 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 a set Θ, $\bar{\Theta}$ denotes the closure of the set in its corresponding space. Convergence in distribution is denoted by $\stackrel{d}{\to}$ convergence in probability by $\stackrel{p}{\to}$. 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 $\{a_t\}_{t=1,\dots,T}, \{b_t\}_{t=1,\dots,T}$, $a_t \in \mathbb{R}^k$, $b_t \in \mathbb{R}^m$ for $t = 1,\dots,T$, we define $\langle a_t, b_t\rangle := T^{-1}\sum_{t=1}^T a_t b_t'$ analogously to Johansen and Nielsen (2018). For these sequences $a_t|_{b_t}$ denotes the residuals of the regression of a_t on b_t, i. e., it holds that $a_t|_{b_t} = a_t - \langle a_t, b_t\rangle \langle b_t, b_t\rangle^{-1} b_t$. For functions $f : [0,1] \to \mathbb{R}^k$, $g : [0,1] \to \mathbb{R}^m$, define $f(u)|_{g(u)} := f(u) - \int_0^1 f(u)g(u)'du \left(\int_0^1 g(u)g(u)'du\right)^{-1} g(u)$.

3.2 I(2) Processes in the State Space Framework

This paper deals with the same class of multivariate ARMA processes $\{y_t\}_{t\in\mathbb{Z}}$, $y_t \in \mathbb{R}^s$, as discussed in Bauer and Wagner (2012, p. 1316-1317). As done there, we refer to a stochastic process $\{y_t\}_{t\in\mathbb{Z}}$ as an ARMAX process, if there exist a deterministic process $\{\mathcal{D}_t\}_{t\in\mathbb{Z}}$, integers $p, q \geq 1$, matrices $A_j \in \mathbb{R}^{s\times s}$, $j = 1,\dots,p$ and $B_j \in \mathbb{R}^{s\times s}$, $j = 1,\dots,q$ and a white noise process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$, with $\mathbb{E}(\varepsilon_t\varepsilon_t') = \Sigma > 0$, such that

$$a(L)(y_t - \mathcal{D}_t) = y_t - \mathcal{D}_t + \sum_{j=1}^p A_j(y_{t-j} - \mathcal{D}_{t-j}) = \varepsilon_t + \sum_{j=1}^q B_j\varepsilon_{t-j} = b(L)\varepsilon_t, \quad t \in \mathbb{Z}, \qquad (3.1)$$

Defining the matrix polynomials $a(z) := I_s + \sum_{j=1}^p A_j z^j$ and $b(z) := I_s + \sum_{j=1}^q B_j z^j$ where $z \in \mathbb{C}$, the pair $(a(z), b(z))$ is called an ARMAX system corresponding to the stochastic process $\{y_t\}_{t\in\mathbb{Z}}$.

Defining the difference operator as

$$\mathbf{\Delta} = \mathbf{\Delta}(L) := 1 - L,$$

We are now ready to define I(k) processes:

Definition 15 *A stochastic process $\{y_t\}_{t\in\mathbb{Z}}, y_t \in \mathbb{R}^s$, is an integrated process of order k, if*

$$\mathbf{\Delta}^k(y_t - \mathcal{D}_t) = v_t, \quad t \in \mathbb{Z},$$

where $\{\mathcal{D}_t\}_{t\in\mathbb{Z}}$, $\mathcal{D}_t \in \mathbb{R}^s$, is deterministic and $\{v_t\}_{t\in\mathbb{Z}}, v_t \in \mathbb{R}^s$, is the solution of a vector autoregressive moving average (VARMA) system

$$a(L)v_t = b(L)\varepsilon_t, \quad t \in \mathbb{Z},$$

fulfilling $\det(a(z)) \neq 0$ for $|z| \leq 1$, $\det(b(z)) \neq 0$ for all $|z| < 1$ and $b(1) \neq 0$ and where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$, $\varepsilon_t \in \mathbb{R}^s$ is a white noise process with $\mathbb{E}(\varepsilon_t\varepsilon_t') = \Sigma > 0$.

3.2.1 The State Space Representation of I(2) Processes

Bauer and Wagner (2012, Theorem 2) show that every I(2) process $\{y_t\}_{t\in\mathbb{Z}}$ has a unique state space representation of the form

$$\begin{aligned} y_t &= \underbrace{\begin{bmatrix} \mathcal{C}_u & \mathcal{C}_\bullet \end{bmatrix}}_{\mathcal{C}} \begin{bmatrix} x_{t,u} \\ x_{t,\bullet} \end{bmatrix} + \mathcal{D}_t + h_t + \varepsilon_t \\ \begin{bmatrix} x_{t+1,u} \\ x_{t+1,\bullet} \end{bmatrix} &= \underbrace{\begin{bmatrix} \mathcal{A}_u & 0 \\ 0 & \mathcal{A}_\bullet \end{bmatrix}}_{\mathcal{A}} \begin{bmatrix} x_{t,u} \\ x_{t,\bullet} \end{bmatrix} + \underbrace{\begin{bmatrix} \mathcal{B}_u \\ \mathcal{B}_\bullet \end{bmatrix}}_{\mathcal{B}} \varepsilon_t \\ \begin{bmatrix} x_{1,u} \\ x_{1,\bullet} \end{bmatrix} &= \begin{bmatrix} 0 \\ \sum_{j=0}^{\infty} \mathcal{A}_\bullet^j \mathcal{B}_\bullet \varepsilon_{-j} \end{bmatrix} \end{aligned} \tag{3.2}$$

where $\mathcal{A} \in \mathbb{R}^{n\times n}$, $\mathcal{B} \in \mathbb{R}^{n\times s}$, $\mathcal{C} \in \mathbb{R}^{s\times n}$, $\{\mathcal{D}_t\}_{t\in\mathbb{Z}}$ is deterministic and $\{h_t\}_{t\in\mathbb{Z}}$ fulfills $\mathbf{\Delta}^2 h_t = 0$ for all $t \in \mathbb{Z}$. All eigenvalues of $\mathcal{A}_u$ are equal to one, $\lambda_{|\max|}(\mathcal{A}_\bullet) < 1$ 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 and the system matrices fulfill the following constraints

- The subsystem ($\mathcal{A}_u$, $\mathcal{B}_u$, $\mathcal{C}_u$) is of the form

  $$\mathcal{A}_u = \begin{bmatrix} I_{c_1} & I_{c_1} & 0 \\ 0 & I_{c_1} & 0 \\ 0 & 0 & I_{c_2} \end{bmatrix}, \quad \mathcal{B}_u := \begin{bmatrix} \mathcal{B}_1' & \mathcal{B}_E' \end{bmatrix}' := \begin{bmatrix} \mathcal{B}_1' & \mathcal{B}_2' & \mathcal{B}_3' \end{bmatrix}', \quad \mathcal{C}_u := \begin{bmatrix} \mathcal{C}_1 & \mathcal{C}_2 & \mathcal{C}_3 \end{bmatrix},$$

  where $\mathcal{C}_1, \mathcal{C}_2, \mathcal{B}_1', \mathcal{B}_2' \in \mathbb{R}^{s\times c_1}$ and $\mathcal{B}_3', \mathcal{C}_3 \in \mathbb{R}^{s\times c_2}$. Moreover $\mathcal{B}_2$ and $\mathcal{B}_3$ are positive upper triangular (p.u.t.) matrices, and $(\mathcal{C}_E)'\mathcal{C}_E = I_{c_1+c_2}$ and $(\mathcal{C}_E)'\mathcal{C}_2 = 0$, where $\mathcal{C}_E := [\mathcal{C}_1, \mathcal{C}_3]$.

- The state space representation of the (stable) subsystem ($\mathcal{A}_\bullet$, $\mathcal{B}_\bullet$, $\mathcal{C}_\bullet$) is in echelon canonical form.

Conversely every process $\{y_t\}_{t\in\mathbb{Z}}$ generated by a the state space system (3.2) satisfying $\lambda_{|\max|}(\mathcal{A}_\bullet) = 1$ and $\lambda_{|\max|}(\underline{\mathcal{A}} := \mathcal{A} - \mathcal{B}\mathcal{C}) \leq 1$ is an I(2) process.

Remark 19 *This result is analogous to Bauer and Wagner (2012) where the authors introduce the process $\{y_{t,h}\}_{t\in\mathbb{Z}}$ satisfying $\mathbf{\Delta}^2 y_{t,h} = 0$ instead of the two components $\{\mathcal{D}_t\}_{t\in\mathbb{Z}}$ and $\{h_t\}_{t\in\mathbb{Z}}$. Clearly we can decompose $\{y_{t,h}\}_{t\in\mathbb{Z}}$ into a stochastic part $y_{t,h} - \mathbb{E}(y_{t,h})$ corresponding to h_t in the above system and a deterministic part $\mathbb{E}(y_{t,h})$.*

Remark 20 *The above canonical form is associated to a multi-index* $\mathbf{\Gamma}$*, whose properties are examined in Bauer et al. (2020, Theorem 2). The multi-index contains the state dimension, the unit root indices* c_1 *and* c_2*, the position of the positive entries in* B_E *restricted due to the p.u.t. form and the Kronecker indices (see e.g. Hannan and Deistler (1988, Chapter 2.4) for a precise definition) of the stable subsystem. It is useful to redefine the state space unit root structure*[1] *for I(2) processes as* $\Omega_{\mathcal{S}} := (c_1, c_2)$*, with unit root indices* c_1, c_2*,* $c_1 + c_2 \leq s$*.*

Remark 21 *For any I(2) process with arbitrary unit root indices* c_1, c_2*,* $c_1 + c_2 \leq s$*, there also exists a unique state space representation in echelon canonical form. If no information on* c_1 *and* c_2 *is available, as is common in application, the echelon canonical form is, therefore, useful to get initial estimates for the system matrices, without imposing restrictions on* $\Omega_{\mathcal{S}}$*.*

The solution for $t > 0$ and $x_{1,u} = 0$ of the system in canonical form in this setting is given by

$$\begin{aligned} y_t &= \mathcal{C}_1 x_{t,1} + \mathcal{C}_2 x_{t,2} + \mathcal{C}_3 x_{t,3} + \mathcal{C}_\bullet x_{t,\bullet} + \mathcal{D}_t + h_t + \varepsilon_t \\ &= \mathcal{C}_1 \mathcal{B}_2 \sum_{k=1}^{t-1} \sum_{j=1}^{k} \varepsilon_{t-j} + (\mathcal{C}_1 \mathcal{B}_1 + \mathcal{C}_2 \mathcal{B}_2 + \mathcal{C}_3 \mathcal{B}_3) \sum_{j=1}^{t-1} \varepsilon_{t-j} \\ &\quad + \mathcal{C}_\bullet \sum_{j=1}^{t-1} \mathcal{A}_\bullet^{j-1} \mathcal{B}_\bullet \varepsilon_{t-j} + \mathcal{C}_\bullet \mathcal{A}_\bullet^{t-1} x_{1,\bullet} + \mathcal{D}_t + h_t + \varepsilon_t, \end{aligned} \tag{3.3}$$

which showcases the different integrated components, making the canonical from, therefore, especially suited for cointegration analysis. We define cointegrating vectors and spaces as follows.

Definition 16 *(i) An s-dimensional I(2) process* $\{y_t\}_{t\in\mathbb{Z}}$ *is called* cointegrated of order one*, if there exists a vector* $\beta \in \mathbb{R}^s, \beta \neq 0$*, such that* $\{\beta' y_t\}_{t\in\mathbb{Z}}$ *is an* $I(1)$ *process. In this case the vector* β *is a* cointegrating vector (CIV) of order one*.*

(ii) An s-dimensional I(2) process $\{y_t\}_{t\in\mathbb{Z}}$ *is called* cointegrated of order two*, if there exists a vector* $\beta \in \mathbb{R}^s, \beta \neq 0$*, such that* $\{\beta' y_t\}_{t\in\mathbb{Z}}$ *is a stationary process. In this case the vector* β *is a* cointegrating vector (CIV) of order two*.*

(iii) The span of all CIVs of order $k = 1, 2$ *is called* (static) cointegrating space of order $k = 1, 2$.

(iv) An s-dimensional I(2) process $\{y_t\}_{t\in\mathbb{Z}}$ *is called* polynomially cointegrated*, if there exists a vector polynomial* $\beta(z) = \beta_0 + \beta_1 z$*,* $\beta_k \in \mathbb{R}^s$*,* $k = 0, 1$*, such that* $\beta(L)'(\{y_t\}_{t\in\mathbb{Z}})$ *is stationary and* $\beta_0 + \beta_1 \neq 0$*. In this case the vector polynomial* $\beta(z)$ *is a* polynomial cointegrating vector (PCIV)*.*

(v) The span of all PCIVs is called polynomial cointegrating space*.*

Remark 22 *In the I(2) case, setting* $\mathcal{D}_t = h_t = x_{t,\bullet} = 0$ *for simplicity, we have*

$$\begin{aligned} y_t &= \mathcal{C}_1 x_{t,1} + \mathcal{C}_2 x_{t,2} + \mathcal{C}_3 x_{t,3} + \varepsilon_t, \\ x_{t+1,1} &= x_{t,1} + x_{t,2} + \mathcal{B}_1 \varepsilon_t, \\ x_{t+1,2} &= x_{t,2} + \mathcal{B}_2 \varepsilon_t, \\ x_{t+1,3} &= x_{t,3} + \mathcal{B}_3 \varepsilon_t \\ x_{t,e} &:= [(x_{t,1})', (x_{t,3})']', \quad x_{t,g} := x_{t,2}. \end{aligned}$$

The vector $\beta \in \mathbb{R}^s$ *is a CIV of order one if and only if* $\beta' \mathcal{C}_1 = 0$ *and* $\beta'[\mathcal{C}_2, \mathcal{C}_3] \neq 0$*.*
The vector $\beta \in \mathbb{R}^s$ *is a CIV of order two if and only if* $\beta'[\mathcal{C}_E, \mathcal{C}_2] = 0$ *and* $\beta \neq 0$*.*
The vector polynomial $\beta(z) = \beta_0 + \beta_1 z$*, with* $\beta_0, \beta_1 \in \mathbb{R}^s$ *is a PCIV of order two if and only if*

[1] The state space unit root structure as defined in Bauer et al. (2020) for general unit root processes also contains unit root frequencies. In the context of I(2) processes with only one unit root we omit the frequency in the definition of $\Omega_{\mathcal{S}}$ to shorten notation.

$[\beta_0', \beta_1'] \begin{bmatrix} \mathcal{C}_E & \mathcal{C}_1 + \mathcal{C}_2 \\ \mathcal{C}_E & \mathcal{C}_2 \end{bmatrix} = 0$ *and* $\beta_0 + \beta_1 \neq 0$.
If $\mathcal{C}_2 = 0$ *this equation is only fulfilled if already a CIV of order two exists. In this case the first condition implies* $\beta_0'\mathcal{C}_1 = 0$, $\beta_1'\mathcal{C}_1 = 0$ *and* $(\beta_0 + \beta_1)'\mathcal{C}_3 = 0$. *Consequently* $\beta_0 + \beta_1$ *is a CIV of order two and all PCIVs are trivial in the sense that they correspond to CIVs of order two.*
If $\mathcal{C}_2 \neq 0$ *however, PCIVs which are not trivial exist. This is because the only integrated contribution to* $\{\mathbf{\Delta} y_t\}_{t\in\mathbb{Z}}$ *equals* $\{\mathcal{C}_1 \mathbf{\Delta} x_{t,1}\}_{t\in\mathbb{Z}} = \{\mathcal{C}_1 x_{t-1,2} + \mathcal{C}_1\mathcal{B}_1\varepsilon_{t-1}\}_{t\in\mathbb{Z}}$. *Thus, in order for cointegration between* $\{\mathbf{\Delta} y_t\}_{t\in\mathbb{Z}}$ *and* $\{\gamma' y_t\}_{t\in\mathbb{Z}}$ *to be present, the latter must eliminate the contribution of* $\{x_{t,1}\}_{t\in\mathbb{Z}}$ *and* $\{x_{t,3}\}_{t\in\mathbb{Z}}$ *but contain some contribution of* $\{x_{t,2}\}_{t\in\mathbb{Z}}$, *which is only possible if* $\mathcal{C}_2 \neq 0$.

3.2.2 Error Correction Models for I(2) Processes

The transfer function of an ARMA system is $k(z) := a(z)^{-1}b(z)$. It is well known (see e.g. Hannan and Deistler (1988)) that the transfer function corresponding to a state space system has a power series expansion $k(z) = I_s + \sum_{j=1}^{\infty} \mathcal{C}\mathcal{A}^{j-1}\mathcal{B}z^j$ and thus converges absolutely for $z \in \mathbb{C}$ with $|z| < 1$, assuming $\lambda_{|\max|}(\mathcal{A}) \leq 1$. Thus, it has an equivalent representation $k(z) = I_s + z\mathcal{C}(I_n - z\mathcal{A})^{-1}\mathcal{B}$ on the complex unit disk. The corresponding inverse transfer function has a power series expansion $k^{-1}(z) = I_s - \sum_{j=1}^{\infty} \mathcal{C}\underline{\mathcal{A}}^{j-1}\mathcal{B}z^j$. The poles of the transfer function are the inverses of the non-zero eigenvalues of $\mathcal{A}$. Thus, the only pole on the closed unit disk for I(2) processes is $z = 1$. The zeros of $k^{-1}(z)$ are the poles of $k(z)$ which implies that the matrix $\Pi := -k^{-1}(1) = -I_s + \mathcal{C}(I_n - \underline{\mathcal{A}})^{-1}\mathcal{B}$ is of reduced rank, such that

$$\Pi = \alpha\beta' \quad \text{with } \alpha, \beta \in \mathbb{R}^{s\times(s-c_1+c_2)}.$$

Moreover, defining

$$\begin{aligned}\Gamma :&= k^{-1}(1) + \partial_z k^{-1}(z)|_{z=1} \\ &= \Pi - \mathcal{C}(I - \underline{\mathcal{A}})^{-2}B = -I_s - \mathcal{C}(I_n - \underline{\mathcal{A}})^{-2}\underline{\mathcal{A}}\mathcal{B},\end{aligned}$$

it holds that $\alpha_\perp'\Gamma\beta_\perp$ is also of reduced rank $c_2 < c_1 + c_2$, such that

$$\alpha_\perp'\Gamma\beta_\perp \quad =: \quad \xi\eta' \quad \text{with } \xi, \eta \in \mathbb{R}^{(c_1+c_2)\times c_2}. \tag{3.4}$$

These matrices together with their rank restrictions also occur in the VECM, compare Johansen (1997), which is given by

$$\mathbf{\Delta}^2 y_t \quad = \quad \alpha\beta' y_{t-1} - \Gamma\mathbf{\Delta} y_{t-1} + \sum_{j=1}^{p-2} \Psi_j \mathbf{\Delta}^2 y_{t-j} + \varepsilon_t. \tag{3.5}$$

Based on the matrices Π and Γ corresponding to an I(2) VAR process $\{y_t\}_{t\in\mathbb{Z}}$, 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 $r_0 = s - c_1 - c_2$ and $r_1 = c_2$ and $r_2 = c_1$. 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 $\{\beta' y_t\}_{t\in\mathbb{Z}}$ typically is $I(1)$ and cointegrates with $\{\beta_2'\mathbf{\Delta} y_t\}_{t\in\mathbb{Z}}$ to stationarity. Thus, there is a direct correspondence of these indices to the dimensions of the different cointegrating spaces – both static and polynomial – and the column spaces of β_2 and $\mathcal{C}_1$ and of β_1 and $\mathcal{C}_3$ coincide (compare Remark 22).
Finally let us also briefly consider the Granger type representation for I(2) VAR processes due to Johansen (1992, compare p. 195), which is of the form

$$y_t = C_2 \sum_{s=1}^{t}\sum_{j=1}^{s} \varepsilon_j + C_1 \sum_{j=1}^{t} \varepsilon_j + C^*(L)\varepsilon_t + A + Bt$$

where the following relations hold

$$C_2 = \beta_2(\alpha_2'\Theta\beta_2)^{-1}\alpha_2', \quad \Theta := \Gamma\beta(\beta'\beta)^{-1}(\alpha'\alpha)^{-1}\alpha'\Gamma + I_s - \textstyle\sum_{j=1}^{p-2}\Psi_j,$$
$$\beta_1'C_1 = ((\alpha'\alpha)^{-1}\alpha_1'\Gamma\beta_1(\beta_1'\beta_1)^{-1})^{-1}(\alpha_1'\alpha_1)^{-1}\alpha_1'(I_s - \Theta C_2).$$

Comparing the matrices $\beta_1'C_1$ and C_2 with the corresponding terms in (3.3), we see that the row spaces of α_2' and $\mathcal{B}_2$ and $[\alpha_1, \alpha_2]'$ and $\mathcal{B}_E$ coincide, the latter due to $\mathcal{B}_3 = \mathcal{C}_3'\beta_1(\beta_1'\beta_1)^{-1}\beta_1'C_1$ and $\alpha_1'(I_s - \Theta C_2)\alpha = 0$ which implies $\mathcal{B}_3\alpha = 0$.
Analogously to the VECM we introduce a state space error correction model (SSECM) for I(2) processes:

Theorem 9 (SSECM-I(2)) *Every state space state space representation with system matrices (A, B, C), satisfying $\det(I_n - \underline{A}) \neq 0$, is equivalently represented by the following system (using $\tilde{y}_t := y_t - \mathcal{D}_t - h_t$):*

$$\begin{aligned}
\mathbf{\Delta}^2\tilde{y}_t &= \Pi\tilde{y}_{t-1} + \Gamma\mathbf{\Delta}\tilde{y}_{t-1} + Cv_t + \varepsilon_t,\\
v_{t+1} &= \underline{A}v_t + (I_n - \underline{A})^{-2}\underline{A}^2B\mathbf{\Delta}^2\tilde{y}_t, \quad v_1 = x_1,\\
\Pi &= -I_s + C(I_n - \underline{A})^{-1}B,\\
\Gamma &= -I_s - C(I_n - \underline{A})^{-2}\underline{A}B.
\end{aligned}$$

Furthermore, let $\mathcal{D}_t + h_t := -[d, e][1, t]'$. The state space error correction model (SSECM-I(2)) is then given by

$$\begin{aligned}
\mathbf{\Delta}^2 y_t &= \Pi(y_{t-1} + d + e(t-1)) + \Gamma(\mathbf{\Delta}y_{t-1} + e) + Cv_t + \varepsilon_t \qquad (3.6)\\
v_{t+1} &= \underline{A}v_t + (I_n - \underline{A})^{-2}\underline{A}^2B\mathbf{\Delta}^2 y_t,\\
v_1 &= x_1 - (I_n - \underline{A})^{-1}B(d-e) - (I_n - \underline{A})^{-2}Be,
\end{aligned}$$

where $\Pi = -I_s + C(I_n - \underline{A})^{-1}B$ and $\Gamma = -I_s - C(I_n - \underline{A})^{-2}\underline{A}B$ as before.

Combining the SSECM with (3.3), for an I(2) process $\{y_t\}_{t\in\mathbb{Z}}$ generated by a system $(\mathcal{A}, \mathcal{B}, \mathcal{C})$, it holds that ε_t is equal to

$$\begin{aligned}
\varepsilon_t &= \mathbf{\Delta}^2 y_t - \Pi\tilde{y}_{t-1} - \Gamma\mathbf{\Delta}\tilde{y}_{t-1} - \sum_{i=1}^{t-1}\underline{\mathcal{A}}^{i-1}(I_n - \underline{\mathcal{A}})^{-2}\underline{\mathcal{A}}^2\mathcal{B}\mathbf{\Delta}^2\tilde{y}_{t-i}\\
&= -\Pi\mathcal{C}_1 x_{t-1,1} - \Pi\mathcal{C}_3 x_{t-1,3} - \big(\Pi\mathcal{C}_2 + \Gamma\mathcal{C}_1\big)x_{t-2,2} + v_t^d(\mathcal{A}, \mathcal{B}, \mathcal{C}),
\end{aligned}$$

where $v_t^d(\mathcal{A}, \mathcal{B}, \mathcal{C})$ does not contain integrated components. Since $\{x_{t,1}\}_{t\in\mathbb{Z}}$, $\{x_{t,2}\}_{t\in\mathbb{Z}}$ and $\{x_{t,3}\}_{t\in\mathbb{Z}}$ are all integrated processes, the first three terms are necessarily zero for the equality to hold, such that $v_t^d(\mathcal{A}, \mathcal{B}, \mathcal{C}) = \varepsilon_t$. Thus, the following relation hold between the subblocks of $\mathcal{C}_u$ and the matrices Π and Γ:

$$\begin{aligned}
\big(-I_s + \mathcal{C}(I_n - \underline{\mathcal{A}})^{-1}\mathcal{B}\big)\begin{bmatrix}\mathcal{C}_1 & \mathcal{C}_3\end{bmatrix} &= 0\\
\big(-I_s + \mathcal{C}(I_n - \underline{\mathcal{A}})^{-1}\mathcal{B}\big)\mathcal{C}_2 + \big(-I_s - \mathcal{C}(I_n - \underline{\mathcal{A}})^{-2}\underline{\mathcal{A}}\mathcal{B}\big)\mathcal{C}_1 &= 0,
\end{aligned}$$

or, in shorter notation, $\Pi\mathcal{C}_E = 0$ and $\Pi\mathcal{C}_2 + \Gamma\mathcal{C}_1 = 0$ with Π and Γ corresponding to $(\mathcal{A}, \mathcal{B}, \mathcal{C})$.

3.3 Pseudo Maximum Likelihood Estimation

Let $(\mathcal{A}_\circ, \mathcal{B}_\circ, \mathcal{C}_\circ)$ refer to the true system of the data generating process $\{y_t\}_{t\in\mathbb{Z}}$. We assume that these matrices satisfy the following:

Assumption 2 (Strict minimum phase assumption) *For the matrices $(\mathcal{A}_\circ, \mathcal{B}_\circ, \mathcal{C}_\circ)$ we have $\lambda_{|max|}(\underline{\mathcal{A}}_\circ) < 1$.*

For the estimation we consider two different sets over which we optimize the likelihood, depending on the available knowledge on the state space unit root structure $\Omega_{\mathcal{S}}$ of $\{y_t\}_{T\in\mathbb{Z}}$.

(I) In the first scenario we only assume to know the state dimension n. In this case we consider the set of transfer functions corresponding to the stationary processes with state dimension n, which we denote by $M_{n,\bullet}$. By Bauer et al. (2020, Theorem 5) the true transfer function is contained in the closure $\overline{M}_{n,\bullet}$, compare also Remark 21. Let $\mathbf{\Theta}_n$ denote the corresponding parameter space, using the echelon canonical form with generic Kronecker indices, compare Hannan and Deistler (1988, Chapter 2.4).

(II) In the second case we assume to know $\Omega_{\mathcal{S}} = (c_1, c_2)$. Thus, we consider the set $M_n(c_1, c_2)$ of transfer functions corresponding to the systems $(\mathcal{A}, \mathcal{B}, \mathcal{C})$ as in (3.2) with

$$\mathcal{A}_u = \begin{bmatrix} I_{c_1} & I_{c_1} & 0 \\ 0 & I_{c_1} & 0 \\ 0 & 0 & I_{c_2} \end{bmatrix}.$$

and state dimension $n \geq 2c_1 + c_2$.
Obviously not all entries in $(\mathcal{A}, \mathcal{B}, \mathcal{C})$ are free parameters. Since a single parameterization of $M_n(c_1, c_2)$ does not exist, it is partitioned into a number of parametrizable sets $M_{\mathbf{\Gamma}}$, such that $M_n(c_1, c_2) = \bigcup_{\mathbf{\Gamma}} M_{\mathbf{\Gamma}}$, where the multi-index $\mathbf{\Gamma}$ encodes information on the structure of the canonical form as described in Remark 20. There exists a generic multi-index $\mathbf{\Gamma}_g$ and a corresponding set $M_{\mathbf{\Gamma}_g} \subset M_n(c_1, c_2)$ such that $M_n(c_1, c_2) \subset \overline{M}_{\mathbf{\Gamma}_g}$ as discussed in detail in Bauer et al. (2020). Let $\mathbf{\Theta}_n^{c_1,c_2}$ denote the parameter space corresponding to $M_{\mathbf{\Gamma}_g}$.
For given $\mathbf{\Gamma}$ the parameter space $\mathbf{\Theta}_{\mathbf{\Gamma}} \subset \mathbb{R}^{c_\theta}$ is equal to $\mathbf{\Theta}_{\mathbf{\Gamma}} = \mathbf{\Theta}_E \times \mathbf{\Theta}_G \times \mathbf{\Theta}_{B,f} \times \mathbf{\Theta}_{B,p} \times \mathbf{\Theta}_\bullet$, such that a parameter vector $\theta \in \mathbf{\Theta}_{\mathbf{\Gamma}}$ is composed of

- the parameter vectors $\theta_E \in \mathbf{\Theta}_E \subset \mathbb{R}^{c_E}$ collecting parameters for the block columns of the unitary matrices $\mathcal{C}_1 = \mathcal{C}_1(\theta_{E,1})$ and $\mathcal{C}_3 = \mathcal{C}_3(\theta_{E,2})$, with $\theta_E = [\theta'_{E,1}, \theta'_{E,2}]'$ moreover $\theta_{E,k} = [\theta'_{k,L}, \theta'_{k,R}]'$, $k = 1, 2$, where $\theta_{k,L}$ contains the parameters characterizing the column space of $\mathcal{C}_1$ and $\mathcal{C}_3$ respectively.
- the parameter vector $\theta_G \in \mathbf{\Theta}_G \subset \mathbb{R}^{c_G}$ collecting parameters for the block columns of the matrix $\mathcal{C}_2 = \mathcal{C}_2(\theta_G)$.
- the parameter vector $\theta_{B,f} \in \mathbf{\Theta}_{B,f} = \mathbb{R}^{c_{B,f}}$ collecting the non-restricted entries in all $\mathcal{B}_k$,
- the parameter vector $\theta_{B,p} \in \mathbf{\Theta}_{B,p} = \mathbb{R}_+^{c_{B,p}}$ collecting the positive real entries in all $\mathcal{B}_k$ restricted due to the p.u.t. form,
- the parameter vector $\theta_\bullet \in \mathbf{\Theta}_\bullet \subset \mathbb{R}^{c_\bullet}$ 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 $\Lambda_\bullet$.

Concerning $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ we assume the following.

Assumption 3 *The errors $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are a strictly stationary martingale difference sequence satisfying:*

- $\mathbb{E}(\varepsilon_t|\mathcal{F}_{t-1}) = 0$.
- $\mathbb{E}(\varepsilon_t\varepsilon_t') = \mathbb{E}(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}) = \Sigma_\circ > 0$.
- $\mathbb{E}(\|\varepsilon_t\|^4) < \infty$.

where $\mathcal{F}_{t-1}$ is the σ-algebra spanned by $\{\varepsilon_j\}_{j\in\mathbb{Z},j<t}$.

We need to consider the parametrization of a set of variance matrices of dimension $s \times s$. Here the parameter vector is $\sigma \in \mathbb{R}^{s(s+1)/2}$ and collects the diagonal and sub-diagonal elements of the symmetric $s \times s$ matrix. Let $\mathbf{\Theta}_\Sigma \subset \mathbb{R}^{s(s+1)/2}$ denote the corresponding parameter space.

The starting values of the state equation are not part of the parameter space. Nevertheless we need to specify their stochastic properties in order to derive the likelihood function. The assumptions on their properties depend on the available information on the state space unit root structure (c_1, c_2) corresponding to $\{y_t\}_{t\in\mathbb{Z}}$.

(I) In case of no further knowledge on $\{y_t\}_{t\in\mathbb{Z}}$ we chose $x_1 = 0$ for all transfer functions in $\overline{M_{n,\bullet}}$ (*prediction error* method of estimation).

(II) For given true state space unit root structure (c_1, c_2), 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

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

As a last step let us specify the form of the deterministic process $\{\mathcal{D}_t\}_{t\in\mathbb{Z}}$ and the singular process $\{h_t\}_{t\in\mathbb{Z}}$. Since in practice only (a part of) one realization of $\{y_t\}_{t\in\mathbb{Z}}$ is available, the contribution of $\{h_t\}_{t\in\mathbb{Z}}$ can be accounted for by including non-zero starting values in addition to $d + et$, with $d, e \in \mathbb{R}^s$, among the components of $\{\mathcal{D}_t\}_{t\in\mathbb{Z}}$. Therefore, consider $h_t = 0$ and $\mathcal{D}_t = \mathcal{D}_t(d, e) := d + et$ with two parameter vectors d, e. Let $(d_\circ, e_\circ)$ be the pair corresponding to the true deterministic process $\{\mathcal{D}_{t,\circ}\}_{t\in\mathbb{Z}}$. Let $\mathbf{\Theta}_D = \mathbb{R}^s \times \mathbb{R}^s$ denote the corresponding parameter space.
In order to define the pseudo likelihood function we use $Y_T := [\ y_1' \ \ \dots \ \ y_T' \ ]' \in \mathbb{R}^{Ts}$ for denoting the stacked observations, $D_T(d, e) \in \mathbb{R}^{Ts}$ is equal to $[\ \mathcal{D}_1(d, e)' \ \ \dots \ \ \mathcal{D}_T(d, e)' \ ]'$, where $\mathcal{D}_t(d, e) = d + et = [d, e][1, t]' =: [d, e]s_t$, denotes the deterministic terms as a function of the vectors d and e. Note that under our assumptions $\mathbb{E}y_t = \mathcal{D}_t$ and that, moreover, there exists a matrix S such that $s_t = Ss_{t-1}$. Further let $\Gamma_T(k(z), \sigma)$ denote the variance matrix corresponding to $Y_T - D_T(d, e)$ which according to the model is given by

$$\Gamma_T(k(z), \sigma) = \mathcal{T}_T(k(z))\,(I_T \otimes \Sigma(\sigma))\,\mathcal{T}_T(k(z))' + \mathcal{O}_{T,\bullet}P_\bullet(\theta,\sigma)\mathcal{O}_{T,\bullet}',$$

$$\mathcal{T}_T(k(z)) := \begin{bmatrix} K_0 & 0 & \dots & 0 \\ K_1 & K_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ K_{T-1} & \dots & K_1 & K_0 \end{bmatrix},$$

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)

$$L_T(k(z), \sigma, d, e; Y_T) = \\ \frac{1}{T}\left(\log\det\Gamma_T(k(z),\sigma) + (Y_T - D_T(d,e))'\Gamma_T(k(z),\sigma)^{-1}(Y_T - D_T(d,e))\right). \tag{3.7}$$

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

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

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

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

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

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

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

$$\begin{aligned} L_{PE,T}(k(z),\sigma,d,e;Y_T) &= \log\det\Sigma(\sigma) + \mathcal{E}_T(k(z),d,e)'\left(I_T \otimes \Sigma(\sigma)\right)^{-1}\mathcal{E}_T(k(z),d,e)/T \\ &= \log\det\Sigma(\sigma) + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(k(z),d,e)'\Sigma(\sigma)^{-1}\varepsilon_t(k(z),d,e) \end{aligned} \tag{3.8}$$

We obtain the pseudo maximum likelihood estimate and the prediction error estimate respectively by minimizing (3.7) over the set $\overline{M} \times \mathbf{\Theta}_\Sigma \times \mathbb{R}^s \times \mathbb{R}^s$:

$$\begin{aligned} \left(\hat{k}(z),\hat{\sigma},\hat{d},\hat{e}\right) &:= \arg\min_{(k(z)\in\overline{M}_{\mathbf{\Gamma}},\sigma\in\mathbf{\Theta}_\Sigma,d,e\in\mathbb{R}^s)} L_T(k(z),\sigma,d,e;Y_T), \\ \left(\tilde{k}(z),\tilde{\sigma},\tilde{d},\tilde{e}\right) &:= \arg\min_{(k(z)\in\overline{M}_{n,\bullet},\sigma\in\mathbf{\Theta}_\Sigma,d,e\in\mathbb{R}^s)} L_{PE,T}(k(z),\sigma,d,e;Y_T). \end{aligned}$$

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. Its proof in connection with some useful lemmata is given in Appendix C.2.2:

Theorem 10 *Let $\{y_t\}_{t\in\mathbb{Z}}$ be an I(2) process generated by a system of the form* (3.2) *satisfying Assumption 2 with $\Omega_{\mathcal{S}} = (c_1,c_2)$, $\mathcal{D}_t = d_\circ + e_\circ t$ and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ fulfilling Assumption 3.*
Let $k_\circ(z) \in \overline{M}_{\mathbf{\Gamma}} \subset \overline{M}_n(c_1,c_2)$.
Then the pseudo maximum likelihood estimator $\hat{k}(z) = I + \sum_{j=1}^{\infty}\hat{K}_j z^j$ 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,

$$T^\gamma\|\hat{\Pi}\mathcal{C}_{1,\circ}\| \to 0,$$

in probability for all $0<\gamma<2$, where $\hat{\Pi} := \hat{k}^{-1}(1)$, and

$$T^\gamma\|\hat{\Pi}\mathcal{C}_{3,\circ}\| \to 0, \quad \text{and} \quad T^\gamma\|\hat{\Pi}\mathcal{C}_{2,\circ} + \hat{\Gamma}\mathcal{C}_{1,\circ}\| \to 0,$$

in probability for all $0<\gamma<1$, where $\hat{\Gamma} := -\hat{k}^{-1}(1) + \frac{\partial}{\partial z}\hat{k}^{-1}(z)|_{z=1}$. For $\hat{d}$ and $\hat{e}$ the following results hold

- *$T^\gamma\|(\hat{\Pi}(d_\circ - \hat{d}) + \hat{\Gamma}(e_\circ - \hat{e})\| \to 0$ in probability for all $0<\gamma<1/2$.*
- *$T^\gamma\|\hat{\Pi}(e_\circ - \hat{e})\| \to 0$ in probability for all $0<\gamma<3/2$.*

The same holds for all prediction error estimators.

The results of Bauer et al. (2020) imply that the likelihood $\mathcal{L}_T$ is a differentiable function of the parameters. This together with the consistency theorem 10 above implies:

Corollary 5 *Let $\{y_t\}_{t\in\mathbb{Z}}$ be as in Theorem 10 where $k_\circ$ is in $M(c_1,c_2)$. Further assume that $k_\circ$ is a point of continuity of the parameterization $\theta = \phi(\psi(k_\circ))$.*
Then $\hat{\theta}$ converges in probability to the true parameter vector $\theta_\circ$.

Remark 23 *Inspecting the results with respect to the deterministic terms, a number of facts stick out: First, the parameters corresponding to the linear trend term $e\cdot t$ converge as $T^{-3/2}$ in the directions orthogonal to the column space of $\mathcal{C}_{E,\circ}$, as usual in a regression model with stationary processes. In the directions of the I(1)-common trends spanned by the columns of $\mathcal{C}_{3,\circ}$, however, convergence is slower and of order $T^{-1/2}$. This corresponds to the difference in the order of growth between the linear trend term and the stochastic common trends. In the directions of the I(2)-common trends spanned by the columns of $\mathcal{C}_{1,\circ}$ not even convergence holds. The same distinction also holds for the constant term: In the directions orthogonal to the column space of $\mathcal{C}_{E,\circ}$ convergence as in the stationary case occurs, in the direction of the common trends, however, convergence does not occur.*

For the rest of the paper, thus, we use the following notation: Let $\Pi_\circ = \alpha_\circ \beta_\circ'$ such that the columns of $\beta_\circ$ are a basis for the orthogonal complement of $\mathcal{C}_{E,\circ}$. Define

$$\theta_{d,1} := \beta_\circ' e \qquad \theta_{d,2} := \begin{bmatrix} \theta_{d,2e}' & \theta_{d,2d}' \end{bmatrix}' := \begin{bmatrix} (\mathcal{C}_{3,\circ}' e)' & (\beta_\circ' d)' \end{bmatrix}'$$
$$\theta_d := \begin{bmatrix} \theta_{d,1}' & \theta_{d,2}' \end{bmatrix}',$$

such that $\theta_d \in \mathbb{R}^{n_d}$, where $n_d = s - 2c_1 - c_2$. By the above construction there exists a matrix $\mathcal{P}(\mathcal{C}_{u,\circ}) \in \mathbb{R}^{c_d \times 2s}$ such that $\mathcal{P}(\mathcal{C}_{u,\circ})[d', e']' = \theta_d$. It follows that $\mathcal{P}(\hat{\mathcal{C}}_{u,\circ})[\hat{d}', \hat{e}']'$ is a consistent estimator for $\theta_{d,\circ} := \mathcal{P}(\mathcal{C}_{u,\circ})[d_\circ', e_\circ']'$. Moreover, let $D_d(\theta_d) := [(\beta_\circ \theta_{d,2d}), (\mathcal{C}_{3,\circ}\theta_{d,2e} + \beta_\circ \theta_{d,1})]$.

The derivation of the asymptotic distribution of the consistent part of the pseudo maximum likelihood estimator follows the usual scheme. It proceeds in two steps and is based on linearization arguments around the true parameter values $(\theta_\circ, \theta_{d,\circ})$. 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 approach is inspired by Saikkonen (1995) and the proof of the theorem is given in Appendix C.2.3:

Theorem 11 (Asymptotic Distribution) *Let the assumptions of Theorem 10 hold and let the true parameter vector $[\ \theta_\circ' \quad d_\circ \quad e_\circ'\ ]'$ be an interior point of $\mathbf{\Theta_\Gamma} \times (\mathbb{R}^s \times \mathbb{R}^s)$ over which the pseudo likelihood function is maximized, with $\hat{\theta}$ and $\hat{\theta}_d$ denoting the PMLE. Assume that the model for the deterministic terms contains the deterministic terms included in the data generating process. Split*

$$\hat{\theta} = [\ \hat{\theta}_{st}' \quad \hat{\theta}_u'\ ]'$$
$$\text{with } \theta_{st} := [\ \theta_{1,R}' \quad \theta_{2,R}' \quad \theta_{B,f}' \quad \theta_{B,p}' \quad \theta_\bullet'\ ]' \text{ and } \theta_u := [\ \theta_{1,L}' \quad \theta_{2,L}' \quad \theta_G'\ ]'.$$

Then the following asymptotic distribution holds for the components of θ:
(A)

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

where $\partial_a \varepsilon_t(\theta)$ denotes the derivative of $\varepsilon_t(\theta)$ with respect to the a-th component of θ_{st}.
(B) Let

$$\hat{\theta}_\star := [\ \hat{\theta}_u' \quad \hat{\theta}_{d,2}' \quad \hat{\theta}_{d,1}'\ ]',$$

where it is understood that only parameters included in the model occur. Let c_u denote the dimension of θ_u, $c_{d,2}$ the dimension of $\theta_{d,2}$, $c_{d,1}$ the dimension of $\theta_{d,1}$ and define $c_{u,1} := c_1(s - c_1 - c_2)$. Then there exists a scaling matrix

$$D_T^{M\star} := diag(T^2 I_{c_{u,1}}, T I_{c_u - c_{u,1}}, T^{1/2} I_{c_{d,2}}, T^{3/2} I_{c_{d,1}}) diag(M, I_{c_u - c_1(s-c_1) + c_{d,2} + c_{d,1}}),$$

with non-singular $M \in \mathbb{R}^{c_1(s-c_1) \times c_1(s-c_1)}$, such that the asymptotic distribution of $\theta_\star$ is given by:

$$D_T^{M\star}(\hat{\theta}_\star - \theta_{\star,\circ}) \xrightarrow{d} H^{-1}\mathrm{v}.$$

Here H denotes the limit of the suitably normalized entries of the Hessian obtained as the limit to $T^{-h(a,b)} \sum_{t=1}^T \partial_a^M \varepsilon_t(\theta_\circ)' \Sigma_\circ^{-1} \partial_b^M \varepsilon_t(\theta_\circ)$ and v denotes the limit to $T^{-g(a,b)} \sum_{t=1} T \partial_a^M \varepsilon_t(\theta_\circ)' \Sigma_\circ^{-1} \varepsilon_t$ where the normalization factors $h(a,b)$ and $g(a,b)$ depend on the entries a and b. Here ∂_a^M denotes directional derivatives corresponding to the a-th column of $diag(M, I_{c_u - c_1(s-c_1) + c_{d,2} + c_{d,1}})$.
(C) For the index a corresponding to an entry in θ_u we have

$$\partial_a \varepsilon_t(\theta_\circ) = -k^{-1}(L, \theta_\circ)(\partial_a \mathcal{C}_k) x_{t,u}(\theta_\circ), \quad x_{u,t+1}(\theta_\circ) = \mathcal{A}_{u,\circ} x_{u,t}(\theta_\circ) + \mathcal{B}_{u,\circ} \varepsilon_t$$

and else

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

(D) Let $\boldsymbol{W}(u)$ denote a Brownian motion with variance $\Sigma_\circ$. Then H depends on $\mathcal{B}_{E,\circ}\boldsymbol{W}(u)$, while v depends on $\mathcal{B}_{E,\circ}\boldsymbol{W}(u)$ and $\alpha_\circ'\Sigma_\circ^{-1}\boldsymbol{W}(u)$. These two Brownian motions are independent. Moreover, $diag(T^2 I_{c_{u,1}}, T I_{c_u - c_{u,1}}) diag(M, I_{c_u - c_1(s-c_1)})(\hat{\theta}_u - \theta_{u,\circ})$ is mixed Gaussian distributed with conditional (on $\mathcal{B}_{E,\circ}\boldsymbol{W}(u)$) variance H^{-1}.

(E) If no constant or linear trend is included in the model, for the first $c_{u,1}$ components θ_1^M of $M\theta_{1,L} =: [(\theta_1^M)', (\theta_2^M)']'$, we obtain $T^2(\hat{\theta}_1^M - \theta_{1,\circ}^M) \xrightarrow{d} H_{1,*}^{-1}\mathrm{v}_{1,*}$ where

$$H_{1,*,a,b} = tr\left[(\partial_a^M \mathcal{C}_1')\beta_\circ\alpha_\circ'\Sigma_\circ^{-1}\alpha_\circ\beta_\circ'(\partial_b^M \mathcal{C}_1)\boldsymbol{Z}_{1,1*}\right], \quad \boldsymbol{Z}_{1,1*} = \int_0^1 \boldsymbol{G}_1(u)\boldsymbol{G}_1(u)' du,$$

$$\mathrm{v}_{1,*,a} = tr\left[(\partial_a^M \mathcal{C}_1')\beta_\circ\alpha_\circ'\Sigma_\circ^{-1}\boldsymbol{X}_{1*}\right], \quad \boldsymbol{X}_{1*} = \int_0^1 d\boldsymbol{W}(u)\boldsymbol{G}_1(u)'$$

where $\boldsymbol{G}_1(u) := \int_0^u \mathcal{B}_{2,\circ}\boldsymbol{W}(v)dv\big|_{\mathcal{B}_{E,\circ}\boldsymbol{W}(u)}$.

If a constant but no linear trend is included, we obtain $T^2(\hat{\theta}_1^M - \theta_{1,\circ}^M) \xrightarrow{d} H_{1,*}^{-1}\mathrm{v}_{1,*}$ with $\boldsymbol{G}_1(u)$ in $H_{1,*}^{-1}$ and $\mathrm{v}_{1,*}$ replaced by $\boldsymbol{G}_1(u)\big|_1 = \boldsymbol{G}_1(u) - \int_0^1 \boldsymbol{G}_1(v)dv$.

If both a linear trend and a constant are included, we obtain $T^2(\hat{\theta}_1^M - \theta_{1,\circ}^M) \xrightarrow{d} H_{1,*}^{-1}\mathrm{v}_{1,*}$ with $\boldsymbol{G}_1(u)$ in $H_{1,*}^{-1}$ and $\mathrm{v}_{1,*}$ replaced by $\boldsymbol{G}_1(u)\big|_{1,u} = \boldsymbol{G}_1(u) - \int_0^1 \boldsymbol{G}_1(v)dv - 12(u-\frac{1}{2})\int_0^1 (v-\frac{1}{2})\boldsymbol{G}_1(v)dv$. Detailed expressions for the asymptotic distribution of the whole parameter vector θ_u are given in Appendix C.2.3.

(F) The prediction error estimator shows the same asymptotic distribution.

The matrix $M \in \mathbb{R}^{c_1(s-c_1)\times c_1(s-c_1)}$ depends on the true matrix $\mathcal{C}_{u,\circ}$ and is, therefore, not available in application. In the Theorem it is used to separate the $c_{u,1}$-dimensional subspace, corresponding to components that are estimated with rate T^2 (a consequence of the presence of I(2) components within the state) from its orthogonal complement, whose corresponding components are estimated with rate T (linked to the remaining I(1) components within the state). The different orders of convergence occur within the parameter vector $\theta_{1,L}$, since the parameterization given in Bauer et al. (2020) is constructed without explicit dependence on a true underlying system. This approach is comparable to the normalization of β using the true matrix $\beta_\circ$ employed by Johansen in the I(1)-VECM and similar procedures in the I(2)-VECM. In practice, further specification of the matrix M is often not necessary, e. g., in the context of Wald-type tests of hypotheses on the parameters. The following corollary, which is proven in Appendix C.2.3, gives the test statistics and asymptotic distribution for these tests:

Corollary 6 (Wald-type test) *Let the assumptions of Theorem 11 hold. Let $D_T^\theta := diag(T^{1/2}I_{c_{st}}, D_T^{M\star})$, $c_{st} := c_\bullet + c_{B,f} + c_{B,p}$ and $c_\theta := c_{st} + c_u + c_d$. Consider p linearly independent restrictions collected in $H_0 : R\theta = r$, with $R \in \mathbb{R}^{p\times c_\theta}$ 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 $R^\infty \in \mathbb{R}^{p\times c_\theta}$ has full rank p. Then it holds that the Wald-type statistic

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

is asymptotically χ_p^2 distributed under the null hypothesis.

3.4 Tests for the State Space Unit Root Structure

The concentrated log-likelihood function in prediction error representation is up to a constant equal to the criterion function

$$\mathcal{L}_T^\varphi(\varphi) := -\frac{T}{2}\log\det\langle\varepsilon_t(\varphi), \varepsilon_t(\varphi)\rangle, \quad \varphi := [\ \theta' \quad \theta_d'\ ],$$

where $\varepsilon_t(\varphi) := y_t - C(\theta)x_t(\varphi) - \mathbf{d}_t(\varphi)$ with $x_{t+1}(\theta) := \underline{A}(\theta)x_t(\theta) + \mathcal{B}(\theta)y_t$ and $\mathbf{d}_{t+1}(\varphi) := \underline{A}(\theta)\mathbf{d}_t(\varphi) + \mathcal{B}(\theta)(-\mathcal{D}_d(\theta_d))$. We are interested in the specification of the state space unit root structure. This information is encoded in the matrix $\mathcal{C}_u$, $\Pi(\theta)$ and $\Gamma(\theta)$. Assume for the moment $\theta_d = 0$ such that $\mathbf{d}_t(\varphi) = 0$ to shorten notation. First let us define an expanded criterion function

$$\mathcal{L}_T^{ex}(C,\theta) := -\frac{T}{2}\log\det\langle y_t - Cx_t(\theta), y_t - Cx_t(\theta)\rangle .$$

Note that $x_t(\theta)$ depends on the parameter vector θ only via the pair $\underline{A}(\theta)$ and $\mathcal{B}(\theta)$. For given $\mathcal{C}_u$ and PML estimator $\hat{\theta}$ over $\mathbf{\Theta}$ define

$$\mathcal{L}_T^{\mathcal{C}}(\mathcal{C}_u,\hat{\theta}) := \max_{\substack{C\in\mathbb{R}^{s\times n}\,:\,\Pi(C,\hat{\theta})[\mathcal{C}_1,\mathcal{C}_3]=0\\ \Pi(C,\hat{\theta})\mathcal{C}_2+\Gamma(C,\hat{\theta})\mathcal{C}_1=0}} L_T^{ex}(C,\hat{\theta})$$

The FOC for the optimal $\hat{C}^c(\mathcal{C}_u)$ fulfilling the above restriction are given by

$$\begin{bmatrix}\hat{C}^c(\mathcal{C}_u) & \Lambda_1 & \Lambda_2\end{bmatrix}\begin{bmatrix}\langle\hat{x}_t,\hat{x}_t\rangle & \underline{\hat{B}}_1\mathcal{C}_E & \underline{\hat{B}}_1\mathcal{C}_2+\underline{\hat{B}}_2\mathcal{C}_1\\ (\underline{\hat{B}}_1\mathcal{C}_E)' & 0 & 0\\ (\underline{\hat{B}}_1\mathcal{C}_2+\underline{\hat{B}}_2\mathcal{C}_1)' & 0 & 0\end{bmatrix} = \begin{bmatrix}\langle y_t,\hat{x}_t\rangle & \mathcal{C}_E & \mathcal{C}_2+\mathcal{C}_1\end{bmatrix},$$

where Λ_1 and Λ_2 are Lagrange multipliers, $\hat{x}_t := x_t(\theta)$, $\underline{\hat{B}}_1 := \underline{B}_1(\hat{\theta}) := (I_n - \underline{A}(\hat{\theta}))^{-1}B(\theta)$ and $\underline{\hat{B}}_2 := \underline{B}_2(\hat{\theta}) := (I_n - \underline{A}(\hat{\theta}))^{-2}\underline{A}(\hat{\theta})B(\hat{\theta})$. Define

$$\underline{\hat{B}}(\mathcal{C}_u) := \begin{bmatrix}\underline{\hat{B}}_1\mathcal{C}_E & \underline{\hat{B}}_1\mathcal{C}_2+\underline{\hat{B}}_2\mathcal{C}_1\end{bmatrix}.$$

Note that a necessary condition for matrix appearing in the FOC to be invertible is $\underline{\hat{B}}(\mathcal{C}_u)$ being of full rank. Thus, define

$$\mathbb{U}(c_1,c_2,\theta) := \left\{\begin{array}{c}\mathcal{C}_u = [\mathcal{C}_1,\mathcal{C}_2,\mathcal{C}_3], \mathcal{C}_1,\mathcal{C}_2\in\mathbb{R}^{s\times c_1}, \mathcal{C}_3\in\mathbb{R}^{s\times c_2}:\\ (\mathcal{C}_E)'\mathcal{C}_E = I_{c_1+c_2}\text{with } \mathcal{C}_E=[\mathcal{C}_1,\mathcal{C}_2] \text{ and } (\mathcal{C}_2)'\mathcal{C}_E = 0\end{array}\right\}.$$

If the system corresponding to θ is controllable, the residuals corresponding to $\mathcal{C}_u \in \mathbb{U}(c_1,c_2,\hat{\theta})$ can be expressed using block matrix inversion:

$$\begin{aligned}\hat{\varepsilon}_t^c(\mathcal{C}_u) :&= y_t - \hat{C}^c(\mathcal{C}_u)\hat{x}_t\\ &= (y_t - \hat{C}^{\text{OLS}}\hat{x}_t)\\ &\quad + \left(\hat{C}^{\text{OLS}}\underline{\hat{B}}(\mathcal{C}_u) - [\mathcal{C}_E,\mathcal{C}_2+\mathcal{C}_1]\right)\left(\underline{\hat{B}}(\mathcal{C}_u)'\langle\hat{x}_t,\hat{x}_t\rangle^{-1}\underline{\hat{B}}(\mathcal{C}_u)\right)^{-1}\underline{\hat{B}}(\mathcal{C}_u)'\langle\hat{x}_t,\hat{x}_t\rangle^{-1}\hat{x}_t\\ &= \hat{\varepsilon}_t + [\hat{\Pi}\mathcal{C}_E,\hat{\Pi}\mathcal{C}_2+\hat{\Gamma}\mathcal{C}_1]\left(\underline{\hat{B}}(\mathcal{C}_u)'\langle\hat{x}_t,\hat{x}_t\rangle^{-1}\underline{\hat{B}}(\mathcal{C}_u)\right)^{-1}\underline{\hat{B}}(\mathcal{C}_u)'\langle\hat{x}_t,\hat{x}_t\rangle^{-1}\hat{x}_t,\end{aligned}$$

where $\hat{C}^{\text{OLS}}$ denotes the OLS estimator, $\hat{\varepsilon}_t := y_t - \hat{C}^{\text{OLS}}\hat{x}_t$ denotes the corresponding residuals, $\hat{\Pi} := -I_s + \hat{C}^{\text{OLS}}\underline{\hat{B}}_1$ and $\hat{\Gamma} := -I_s - \hat{C}^{\text{OLS}}\underline{\hat{B}}_2$.

Remark 24 *Note that if (A,B,C) is a state space representation of a VAR process, $\underline{A}$ is nilpotent. The system $(\underline{A}^V + B^V C^V, B^V, C^V)$, with*

$$\underline{A}^V := \begin{bmatrix}0_{s(p-1)\times s} & I_{s(p-1)}\\ 0_{s\times s} & 0_{s\times s(p-1)}\end{bmatrix}, \qquad B^V := \begin{bmatrix}0_{s(p-1)\times s}\\ -I_s\end{bmatrix}, \qquad C^V := \begin{bmatrix}C_p & \dots & C_1\end{bmatrix},$$

corresponds to the VAR system $y_t = \sum_{j=1}^p C_j y_{t-j} + \varepsilon_t$. Thus, for given $\mathcal{C}_u$ the matrix $\hat{C}^c(\mathcal{C}_u)$ computed using a state x_t defined through $x_{t+1} = \underline{A}^V x_t + B^V y_t$ can be immediately translates into the polynomial $a(L)$ of a VAR(p) process. Maximizing over $\mathbb{U}(c_1,c_2,\theta^V)$, with θ^V corresponding to $\underline{A}^V$,B^V is then a suitable way to compute the VAR polynomial maximizing the pseudo likelihood function for given (c_1,c_2). This approach might be advantageous as it omits the switching algorithm used to encompass the rank restrictions within the I(2)-VECM. Moreover, it is also extendable to processes of higher integration order.

Let $\tilde{\mathcal{C}}_u$ denote the maximizer of $\mathcal{L}_T^{\mathcal{C}}(\mathcal{C}_u, \hat{\theta})$ with respect to $\mathcal{C}_u \in \mathbb{U}(c_1, c_2, \hat{\theta})$, normalized such that the corresponding $\tilde{\mathcal{C}}_E$ is positive lower triangular. Thus, it holds that

$$\max_{\mathcal{C}_u \in \mathbb{U}(c_1,c_2,\hat{\theta})} \mathcal{L}_T^{\mathcal{C}}(\mathcal{C}_u, \hat{\theta}) = \mathcal{L}_T^{\mathcal{C}}(\tilde{\mathcal{C}}_u, \hat{\theta})$$

Let $\hat{\varepsilon}_t^c := y_t - \hat{C}^c(\tilde{\mathcal{C}}_u)\hat{x}_t$ denote the corresponding residuals for given c_1 and c_2. Then a rank test statistic can be defined as:

$$-2 \log Q(H(c_1, c_2)/H_\bullet, \hat{\theta}) := -T \log \det \left[\langle \hat{\varepsilon}_t^c, \hat{\varepsilon}_t^c \rangle \langle \hat{\varepsilon}_t, \hat{\varepsilon}_t \rangle^{-1} \right],$$

The rank test statistics has the following asymptotic distribution under the null hypothesis:

Theorem 12 *Let $\{y_t\}_{t\in\mathbb{Z}}$ be an I(2) process generated by a system of the form (3.2) satisfying Assumption 2 and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ fulfilling Assumption 3. Assume that the true order n is known. Let $\hat{\theta}$ be the PML estimator over a suitable parameter space $\mathbf{\Theta}$ fulfilling the assumption of Theorem 10. Let, further, $\hat{\varphi}_n^{c_1,c_2}$ denote the PML estimator over $\mathbf{\Theta}_n^{c_1,c_2} \times \mathbf{\Theta}_D$ and $\hat{\varphi}_n$ denote the PML estimator over $\mathbf{\Theta}_n \times \mathbf{\Theta}_D$.*

- *Let $\mathcal{D}_t = 0$. Under $H_0 = H(c_1, c_2)$, as $T \to \infty$, it holds that*

$$\begin{array}{rcl} -2 \log Q(H(c_1, c_2)/H_\bullet, \hat{\theta}) & \to & Q_r^\infty + Q_{r,s}^\infty, \\ -2(\mathcal{L}_T^{\varphi}(\hat{\varphi}_n^{c_1,c_2}) - \mathcal{L}_T^{\varphi}(\hat{\varphi}_n)) & \to & Q_r^\infty + Q_{r,s}^\infty, \end{array} \tag{3.9}$$

  where

$$Q_r^\infty := tr\left\{ \int_0^1 d\boldsymbol{B}(u)\boldsymbol{H}(u)' \left(\int_0^1 \boldsymbol{H}(u)\boldsymbol{H}(u)'du \right)^{-1} \int_0^1 \boldsymbol{H}(u)d\boldsymbol{B}(u)' \right\},$$

$$Q_{r,s}^\infty := tr\left\{ \int_0^1 d\boldsymbol{B}_1(u)\boldsymbol{B}_1(u)' \left(\int_0^1 \boldsymbol{B}_1(u)\boldsymbol{B}_1(u)'du \right)^{-1} \int_0^1 \boldsymbol{B}_1(u)d\boldsymbol{B}_1(u)' \right\},$$

  where $\boldsymbol{B} = (\boldsymbol{B}_1', \boldsymbol{B}_2')'$ is an $(c_1 + c_2)$-dimensional standard Brownian motion on the unit interval, $u \in [0, 1]$, $\boldsymbol{B}_1$ of dimension c_1 and $\boldsymbol{B}_2$ of dimension c_2 and

$$\boldsymbol{H}(u) := \begin{pmatrix} \int_0^u \boldsymbol{B}_1(v)dv \\ \boldsymbol{B}_2(u) \end{pmatrix}\Bigg|_{\boldsymbol{B}_1(u)}.$$

- *If $\mathcal{D}_t = d$, the asymptotic distributions of $-2(\mathcal{L}_T^{\varphi}(\hat{\varphi}_n^{c_1,c_2}) - \mathcal{L}_T^{\varphi}(\hat{\varphi}_n))$ under $H_0 = H(c_1, c_2)$, as $T \to \infty$, is as in (3.9) with $\boldsymbol{H}(u)$ in Q_r^∞ replaced by*

$$\boldsymbol{H}_d(u) := \begin{pmatrix} \int_0^u \boldsymbol{B}_1(v)dv \\ \boldsymbol{B}_2(u) \\ 1 \end{pmatrix}\Bigg|_{\boldsymbol{B}_1(u)}.$$

- *If $\mathcal{D}_t = d + et$, the asymptotic distribution of $-2(\mathcal{L}_T^{\varphi}(\hat{\varphi}_n^{c_1,c_2}) - \mathcal{L}_T^{\varphi}(\hat{\varphi}_n))$ under $H_0 = H(c_1, c_2)$, as $T \to \infty$, is as in (3.9) with $\boldsymbol{H}(u)$ in Q_r^∞ replaced by*

$$\boldsymbol{H}_e(u) := \begin{pmatrix} \int_0^u \boldsymbol{W}_1(v)dv \\ \boldsymbol{W}_2(u) \\ u \end{pmatrix}\Bigg|_{\boldsymbol{W}_{1d}}.$$

  and $Q_{r,s}^\infty$ replaced by

$$tr\left\{ \int_0^1 d\boldsymbol{B}_1(u)\boldsymbol{B}_{1d}(u)' \left(\int_0^1 \boldsymbol{B}_{1d}(u)\boldsymbol{B}_{1d}(u)'du \right)^{-1} \int_0^1 \boldsymbol{B}_{1d}(u)d\boldsymbol{W}_1(u)' \right\},$$

  with $\boldsymbol{W}_{1d} = (\boldsymbol{B}_1', 1)'$.

In practice the test statistics $-2 \log Q(H(c_1, c_2)/H_\bullet, \hat{\theta})$ and $-2(\mathcal{L}_T^{\varphi}(\hat{\varphi}_n^{c_1,c_2}) - \mathcal{L}_T^{\varphi}(\hat{\varphi}_n))$ are nearly identical. Therefore, we only use $Q(H(c_1, c_2)/H_\bullet, \hat{\theta}_n)$ and denote it as the SS-LR test in the following simulation study.

3.5 Simulation Results

We compare the ranktest introduced in the previous section to the ranktests based on VECMs in a small simulation study.
The data generating processes used are of the form

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + a_4 y_{t-4} + \varepsilon_t + \lambda \varepsilon_{t-4} \tag{3.10}$$

with

$$a_1 = \begin{bmatrix} 2-2a & 0 \\ 0 & 0 \end{bmatrix}, a_2 = \begin{bmatrix} -1-a^2+4a & 0 \\ 0 & 0 \end{bmatrix}, a_3 = \begin{bmatrix} 2a^2-2a & 0 \\ 0 & 0 \end{bmatrix}, a_4 = \begin{bmatrix} -a^2 & 0 \\ 0 & 0 \end{bmatrix},$$

where $a = 0.9$, and

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

Clearly it holds that $y_{t,2} = \varepsilon_{t,2} + \lambda \varepsilon_{t-4,2}$, thus, the second component of y_t is stationary. The first component $y_{t,1}$ is integrated of order two, and it holds that $\mathbf{\Delta}^2(1+aL)^2 y_{t,1} = \varepsilon_{t,1} + \lambda\varepsilon_{t-4,1}$, such that a is the other root of the AR-polynomial. For $\lambda = 0$ the DGP is a VAR process. This quite simple processes already showcase some major difficulties found in multivariate cointegration analysis for I(2) processes.
For $\lambda = \{0, 0.5, 0.9\}$ samples of sizes $T = 50$ and $T = 100$ are generated with initial values set to zero. The simulation study is based on 10000 replications.
All processes have state space unit root structure $\Omega_{\mathcal{S}} = (1, 1)$. Note that the following relations hold, compare Bauer et al. (2020),

$$\overline{M}_{n,\bullet} = \overline{M}_n(0,0) \supset \overline{M}_n(0,1) \supset \overline{M}_n(1,0) \supset \overline{M}_n(0,2) \supset \overline{M}_n(1,1) \supset \overline{M}_n(2,0),$$

for sets of transfer functions of order $n \geq 4$. Thus, in application, one would usually start by testing the hypothesis $H(2,0)$ against $H_\bullet$. If this is rejected, continue with the next larger set, in this case $\overline{M}_n(1,1)$, testing $H(1,1)$ against $H_\bullet$, until a hypothesis is not rejected and is, therefore, a possible choice for the state space unit root structure or, analogously, the ranks of the respective matrices in the VECM.
We compare the rejection frequencies of three different tests, which have the same asymptotic distribution. Their large sample performance is, thus, nearly identical and we focus the presentation of the results on the small sample performance. The different tests statistics are computed as follows.
VECM-2S: The lag-length p is chosen by maximizing the Akaike Information criterion, with a lower bound of $p = 2$ to ensure the possibility of stochastic trends of order two. We first solve the reduced rank regression problem for an estimate of $\Pi = \alpha\beta'$ in the I(1)-VECM. In a second step, we find the maximizer of the likelihood under the rank restriction $\hat{\alpha}'_\perp \Gamma \hat{\beta}_\perp = \xi\eta'$ for given $\hat{\alpha}, \hat{\beta}$. The likelihood value of the two step approach and the likelihood value corresponding to the unrestricted OLS estimator in the I(2)-VECM is then used for a likelihood ratio test, as proposed in Johansen (1997).
VECM-LR: Consider both restrictions $\Pi = \alpha\beta'$ and $\alpha'_\perp \Gamma \beta_\perp = \xi\eta'$ and maximize the likelihood in an I(2)-VECM with respect to α, β, ξ, η using a switching algorithm. The likelihood ratio test statistics then compares the likelihood value of this restricted model with the likelihood value corresponding to the unrestricted OLS estimator. This test statistic and its implementation is due to Nielsen and Rahbek (2007).
SS-LR: For the pseudo maximum likelihood approach in the state space framework the output of the subspace algorithm is used as a starting value to find the maximizer $\hat{\theta}_n$ of the pseudo log-likelihood function over $\mathbf{\Theta}_n$. For the subspace algorithm we choose $f = p = 2\hat{k}_{AIC}$ and the system order n is chosen by minimizing a singular value criterion, see Bauer (2001), with a lower bound

of $n = 4$ to ensure the possibilty of two stochastic I(2)-trends. Next, compute the maximizer $\tilde{\mathcal{C}}_u$ of $L_T^{ex}(\mathcal{C}_u, \hat{\theta}_n)$ with respect to $\mathcal{C}_u$ over $\mathbb{U}(c_1, c_2, \hat{\theta}_n)$. This is used to compute the test statistic

$$-2 \log Q(H(c_1, c_2)/H_\bullet, \hat{\theta}_n) = -2(L^{ex}(\tilde{\mathcal{C}}_u, \hat{\theta}_n) - L(\hat{\theta}_n)).$$

For the case $\lambda = 0$, i. e., a VAR-DGP, the rejection rates are given in Table 3.1 for the three tests statistics and sample sizes $T = 50, 100$. All test statistics reject the hypothesis $H(2,0)$ in all cases.

H_0	$H(2,0)$	$H(1,1)$	$H(0,2)$	$\boldsymbol{H}(1,0)$	$H(0,1)$
			VECM-2S		
T=50	100.0	97.1	96.8	**79.4**	64.3
T=100	100.0	99.8	100.0	**80.3**	67.2
			VECM-LR		
T=50	100.0	96.6	96.8	**78.7**	64.3
T=100	100.0	99.8	100.0	**80.2**	67.2
			SS-LR		
T=50	100.0	99.8	99.6	**47.6**	38.6
T=100	100.0	99.9	99.9	**23.0**	22.9

Table 3.1: *Rejection rates of H_0 based on 10000 replications and a nominal 5% level with the DGP solving* (3.10) *with $\lambda = 0$. The tests are not calculated sequentially. The bold hypothesis and its corresponding frequencies indicate the correct model.*

The hypotheses $H(1,1)$ and $H(0,2)$ are rejected in 97% of the cases for the VECM-tests and in nearly 100% of the cases for the SS-test. Major differences can be seen with regard to the rejection rates for the true hypotheses $H(1,0)$ and $H(0,1)$. Here both VECM-tests falsely reject the null in 80% respectively 65% of the cases. The SS-LR rejects in 48% and 39% of the cases for sample size $T = 50$ and in 23% of the cases for sample size $T = 100$.
Clearly the SS-LR performs better than the VECM approaches, even though the DGP is a VAR process. This is the case since the true lag-length of the process is unknown and estimated using the AIC. The estimated lag-length is too low on average. Using the true lag-length $p = 4$, the rejection rates for $H_0 = H(1,0)$ are at 7% and 6%, thus, close to the nominal 5%-level, compare Table 3.2. As a downside, the false hypotheses are rejected in fewer cases for sample size $T = 50$.

H_0	$H(2,0)$	$H(1,1)$	$H(0,2)$	$\boldsymbol{H}(1,0)$	$H(0,1)$
			VECM-LR true p		
T=50	100.0	75.5	91.4	**6.6**	9.5
T=100	100.0	99.8	100.0	**5.7**	8.6

Table 3.2: *Rejection rates of H_0 based on 10000 replications and a nominal 5% level with the DGP solving* (3.10) *with $\lambda = 0$ for known lag-length $p = 4$. The tests are not calculated sequentially. The bold hypothesis and its corresponding frequencies indicate the correct model.*

For the case $\lambda = 0.9$, the rejection rates are given in Table 3.3, again for the three tests statistics and sample sizes $T = 50, 100$. Again, all test statistics reject the hypothesis $H(2,0)$ in all cases. The VECM-tests reject the hypotheses $H(1,1)$ and $H(0,2)$ in fewer cases than for $\lambda = 0$, depending on the sample size and the null, with the best rejection rate being 90%. The SS-LR rejects all false null hypotheses in nearly 100% of the cases. For the true hypotheses $H(1,0)$ and $H(0,1)$ both VECM-tests perform better than for $\lambda = 0$, falsely rejecting the null in around 40% of the cases for $T = 50$ and in around 15% of the cases for $T = 100$. The SS-LR rejects similarly as

H_0	$H(2,0)$	$H(1,1)$	$H(0,2)$	$\boldsymbol{H}(1,0)$	$H(0,1)$
			VECM-2S		
T=50	99.9	72.7	74.5	**45.4**	38.9
T=100	99.9	82.7	90.4	**14.6**	16.4
			VECM-LR		
T=50	99.9	65.0	74.5	**42.3**	38.9
T=100	99.9	79.0	90.4	**13.9**	16.4
			SS-LR		
T=50	100.0	99.9	99.9	**44.6**	38.8
T=100	99.9	99.8	99.9	**18.6**	19.5

Table 3.3: *Rejection rates of H_0 based on 10000 replications and a nominal 5% level with the DGP solving* (3.10) *with $\lambda = 0.9$. The tests are not calculated sequentially. The bold hypothesis and its corresponding frequencies indicate the correct model.*

the VECM-tests for sample size $T = 50$ and is slightly worse (19% of the cases) for sample size $T = 100$.
Thus, the VECM approaches exhibit a better performance, falsely rejecting the null hyptheses in fewer cases, due to a higher estimated lag-length taking account of the MA polynomial in the VARMA system of the DGP. Nevertheless, the rejection frequencies of false null hypotheses are also lower. The SS-LR rejects all false null hypotheses in practically all cases.
To compare the performance of both approaches it is, therefore, necessary to look at the resulting estimated state space unit root structure. The results are given in Table 3.4 for sample sizes $T = 50, 100, 150, 300$ for the VECM and the state space likelihood ratio tests. The VECM-LR

$\Omega_\mathcal{S}$	$(2,0)$	$(1,1)$	$(0,2)$	**(1,0)**	$(0,1)$	$(0,0)$
			VECM-LR			
T=50	0.1	34.8	1.3	23.8	5.9	34.2
T=100	0.1	20.6	0.8	65.1	2.8	10.6
T=150	0	6.8	0.2	84.2	3.7	5.1
T=300	0	2.3	0	89.7	3.5	4.5
			SS-LR			
T=50	0	0.1	0.1	54.6	10.7	34.5
T=100	0	0.3	0	80.6	4.1	15.1
T=150	0	0.5	0	87.4	2.4	9.8
T=300	0	1.4	0	89.4	2.9	6.2

Table 3.4: *Rates of estimated $\Omega_\mathcal{S} = (c_1, c_2)$ based on 10000 replications and a nominal 5% level with the DGP solving* (3.10) *with $\lambda = 0.9$. The tests are calculated sequentially. The correct state space unit root structure is highlighted in bold font.*

for $T = 50$ leads to an estimated $\Omega_\mathcal{S}$ of $(1,1)$ in 35% of the cases, while $(0,0)$ is chosen at nearly the same rate. Thus, a false state space unit root structure is chosen more often than the correct structure $(1,0)$, which exhibits a rate of only 24%. For higher sample sizes the most frequent choice of $\Omega_\mathcal{S}$ is the correct structure, with higher rates the higher the sample size. Note, however, that for $T = 100$ the structure $(1,1)$ is still chosen in every fifth case. The SS-LR rejects $H(1,1)$ in at least 98% of the cases. The most often estimated structure is the correct $\Omega_\mathcal{S}$, with a rate of 54% for $T = 50$, 81% for $T = 100$ and around 90% for $T = 300$. The false structure with the

highest rate is $(0,0)$ (35% for $T = 50$). Note that the rate of correct choices of both approaches is comparable for $T = 300$, while the SS-LR performs better in smaller samples, with a larger margin the smaller the sample size is.
Finally, consider the true order n of the transfer function corresponding to the DGP with $\lambda = 0.9$, which is equal to $n = 8$. Using this order leads to even better estimation of $\Omega_{\mathcal{S}}$ as shown in Table 3.5. For $T = 150, 300$ the rate of correct choices is higher than 94%, coming close to the

H_0	$H(2,0)$	$H(1,1)$	$H(0,2)$	$\boldsymbol{H}(1,0)$	$H(0,1)$	$H(0,0)$
			SS-LR true n			
T=50	0.1	17.5	0	60.1	6.8	14.9
T=100	0	4.3	0	87.5	2.6	5.5
T=150	0	0.2	0	94.0	2.2	3.7
T=300	0	0	0	94.7	2.0	3.3

Table 3.5: *Rates of estimated* $\Omega_{\mathcal{S}} = (c_1, c_2)$ *based on 10000 replications and a nominal 5% level with the DGP solving* (3.10) *with* $\lambda = 0.9$*. The tests are calculated sequentially. The correct state space unit root structure is highlighted in bold font.*

desired rate corresponding to the nominal 5% level of the tests. This shows that the performance of the tests for the state space unit root structure at small and intermediate sample sizes highly depends on the correct specification of the order for the SS-LR and, similarly, on the lag-length for the VECM-tests. Nevertheless, at least for the DGPs considered in this simulation study, the state space approach outperforms the VECM approaches for both VAR- and VARMA-DGPs, if both these integer quantities are estimated.

3.6 Summary and Conclusion

This paper shows that transfer functions corresponding to I(2) processes can be consistently estimated by maximizing the pseudo maximum likelihood function in the state space framework, thus, overcoming the limitation to the class of VAR processes in cointegration literature. The relations between the state space framework and the VECM and its Granger representation are fully characterized through the state space error correction model for I(2) processes. Consistent estimates are available even if the state space unit root structure and other relevant integers are not known. The model allows for a linear trend and a constant, covering the most relevant deterministics used in the literature. The inclusion of further deterministics was not pursued although it seems possible without major changes in the results.
The parameters determining the different cointegrating spaces are estimated with rate T and T^2 in analogy to the results for VAR processes, with their asymptotic distribution being a mixture of Brownian motions. Inclusion of a constant or a constant and a linear trend in the model and the DGP leads to demeaning or demeaning and detrending of some of the Brownian motions occurring in the asymptotic distribution. Hypotheses regarding linear restrictions on these parameters can be tested using Wald-type statistics, which are asymptotically χ^2.
Using likelihood ratio type approaches leads to test statistics for hypotheses on the state space unit root structure and consequently the cointegration indices. The asymptotic distribution of these test statistics under the null coincide with the respective distributions seen in VAR literature.
Finally, a simulation study shows good performance of the rank tests statistics for small sample sizes, outperforming the different tests based on the VECM, even if the DGP is itself a VAR process. The test performance highly depends on the estimation of the VAR order and state dimension of the respective models. This indicates that state space modeling has the potential to improve estimation of cointegrating spaces in situations where a large order is necessary in the VAR framework while a more parsimonious approximation is available using the VARMA framework. Another topic for further research might be monitoring of structural changes in cointe-

grating processes, possibly exploiting the advantages of the state space framework in small samples.

Appendix A

Appendix to Chapter 1

A.1 Proofs of the Results of Section 1.3

A.1.1 Proof of Lemma 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 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 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 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 $i = 1, \dots, d$, 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 2 maps every $C \in U_{s,d}$ to $[D_d(\boldsymbol{\varphi}_D), 0_{s-d\times d}']'$ with $D_d(\boldsymbol{\varphi}_D) = \operatorname{diag}(e^{i\varphi_{D,1}}, \dots, e^{i\varphi_{D,d}})$. Since $Q_{q,i,j}(\varphi)^{-1} = Q_{q,i,j}(\varphi)'$ for all q, i, j and φ, C can be obtained by multiplying $[D_d(\boldsymbol{\varphi}_D), 0_{s-d\times d}']'$ with the transposed Givens rotations.

(iv) The algorithms in Remark 12 and above Lemma 2 describe C_U^{-1} in detail. The determination of an element of $\boldsymbol{\varphi}_L$ or $\boldsymbol{\varphi}_R$ uses the transformation of two complex numbers into polar coordinates in step 2 of Remark 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 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 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 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 and 2) which are differentiable and thus continuous and iterated applications of the extensions of the mappings C_{O,d_2-d_1} and $C_{O,G}$ (compare Corollaries 1 and 2) to general unit root structures and to complex matrices. The proof that these functions are differentiable is analogous to the proofs of Lemma 1 and Lemma 2.

(iii) The definitions of $\boldsymbol{\theta}_{B,f}$ 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 and Lemma 2.

A.2 Proofs of the Results of Section 1.4

A.2.1 Proof of Theorem 3

For the first inclusion the proof can be divided into two parts, discussing the stable and the unstable subsystem separately. The result with regard to the stable subsystem is due to Hannan and Deistler (1988, Theorem 2.5.3 (iv)). For the unstable subsystem $(\tilde{\Omega}_S, \tilde{p}) \leq (\Omega_S, p)$ implies the existence of a matrix S as described in Definition 9. Partition $S = \begin{bmatrix} S_1 \\ S_2 \end{bmatrix}$ such that $S_1 p = p_1 \geq \tilde{p}$. Let $\tilde{k}$ be an arbitrary transfer function in $M_{\tilde{\Gamma}} = \pi(\Delta_{\tilde{\Gamma}})$ with corresponding state space realization $(\tilde{\mathcal{A}}, \tilde{\mathcal{B}}, \tilde{\mathcal{C}}) \in \Delta_{\tilde{\Gamma}}$. Then, we find matrices B_1 and C_1 such that for the state space realization given by $\mathcal{A} = S \begin{bmatrix} \tilde{\mathcal{A}} & \tilde{J}_{12} \\ 0 & \tilde{J}_2 \end{bmatrix} S'$, $\mathcal{B} = S \begin{bmatrix} \tilde{\mathcal{B}} \\ cB_1 \end{bmatrix}$ 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, $(\mathcal{A}_j, \mathcal{B}_j, \mathcal{C}_j) = (\mathcal{A}, S\,\mathrm{diag}(I_{n_1}, j^{-1} I_{n_2}) S' \mathcal{B}, \mathcal{C}) \in \Delta_\Gamma$, 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 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 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 3. Comparing Definitions 10 and 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 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.

$\bigcup_{(\tilde{\Omega}_S,\tilde{n}_\bullet)\in\mathcal{A}(\Omega_S,n_\bullet)} \bigcup_{(\check{\Omega}_S,\check{n}_\bullet)\leq(\tilde{\Omega}_S,\tilde{n}_\bullet)} M(\check{\Omega}_S, \check{n}_\bullet) \quad \subset \quad \overline{M_{\Gamma_g(\Omega_s,n_\bullet)}}$ 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 9) we find a multi-index $\tilde{\Gamma} \geq \check{\Gamma}$ such that $M_{\tilde{\Gamma}} \subset M(\tilde{\Omega}_S, \tilde{n}_\bullet)$. By Theorem 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 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 $\tilde{\Omega}_S$ may be of the form

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

Second, $\check{\mathcal{A}}_u$ corresponding to state space unit root $\check{\Omega}_S$ 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 $\pi(\overline{\Delta_{\Gamma_g(\Omega_S, n_\bullet)}})$, where the off-diagonal entry in A_j is restricted to equal one and hence the corresponding sequence of systems in the canonical form diverges to infinity. In a sense these systems correspond to "points at infinity": For the example given above we obtain the canonical form

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

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 $\tilde{\mathcal{A}}_j$ 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

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

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 as 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 11 and prove (b).

As follows from Corollary 4,

$$\overline{M(\Omega_S, n_\bullet)} = \overline{M_{\Gamma_g(\Omega_S,n_\bullet)}}.$$

Thus, (b) implies $\bigcup_{(\tilde{\Omega}_S,\tilde{n}_\bullet)\in\mathcal{A}(\Omega_S,n_\bullet)} M(\tilde{\Omega}_S,\tilde{n}_\bullet) \subset \overline{M_{\Gamma_g(\Omega_S,n_\bullet)}}$ and (a) adds the second union showing the subset inclusion.
It remains to show equality for the last set inclusion. Thus, we need to show that for $k_j \in M_{\Gamma_g(\Omega_S,n_\bullet)}, k_j \to k_0$, 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

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

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

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

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

$$S\mathcal{A}_0 S' \quad = \quad \begin{bmatrix} \mathcal{A}_j & \tilde{J}_{12} \\ 0 & \tilde{J}_2 \end{bmatrix}$$

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 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 B

Appendix to Chapter 2

B.1 Proof of Theorem 6

Proof: For every state space system (A, B, C) with corresponding transfer function $k(z) = I_s + zC(I_n - zA)^{-1}B$ the inverse transfer function is $k^{-1}(z) = I_s - zC(I_n - z\underline{A})^{-1}B$. Consider the power series

$$\tilde{k}^{-1}(z) = \mathbf{\Delta}_S(z)I_s - \sum_{k=1}^{S} \Pi_k \mathbf{\Delta}_{S,k}(z)z + C(I_n - \underline{A}^S)^{-1}\underline{A}^S \sum_{m=1}^{\infty} \underline{A}^{m-1}B\mathbf{\Delta}_S(z)z^m,$$

which equals $k^{-1}(z)$ as shown in what follows: for $m > S$ the m-th power series coefficient can be calculated as

$$\begin{aligned}\tilde{K}_m^- &= C(I_n - \underline{A}^S)^{-1}\underline{A}^S(\underline{A}^{m-1} - \underline{A}^{m-S-1})B \\ &= -C(I_n - \underline{A}^S)^{-1}(I_n - \underline{A}^S)\underline{A}^{m-1}B = -C\underline{A}^{m-1}B = K_m^-\end{aligned}$$

as required. Thus, $k^{-1}(z) - \tilde{k}^{-1}(z)$ is a polynomial of maximal degree S. Since by construction $\mathbf{\Delta}_{S,j}(z_j) = \overline{z}_j$ and $\mathbf{\Delta}_{S,k}(z_j) = 0$ for $j \neq k$ and $\mathbf{\Delta}_S(z_j) = 0$ for all $j = 1, \ldots, S$ we get

$$k^{-1}(z_j) - \tilde{k}^{-1}(z_j) = k^{-1}(z_j) + \Pi_j \mathbf{\Delta}_{S,j}(z_j)z_j = 0.$$

Additionally $k^{-1}(0) = \tilde{k}^{-1}(0) = I_s$. Thus, $k^{-1}(z) - \tilde{k}^{-1}(z)$ is a polynomial of degree S that is zero at $S + 1$ points. Hence, $k^{-1}(z) - \tilde{k}^{-1}(z) = 0$ and

$$\sum_{m=0}^{t-1} K_m^- \tilde{y}_{t-m} = \mathbf{\Delta}_S \tilde{y}_t - \sum_{k=1}^{S} \Pi_k \mathbf{\Delta}_{S,k} \tilde{y}_{t-1} + C(I_n - \underline{A}^S)^{-1}\underline{A}^S \sum_{m=1}^{t-1} \underline{A}^{m-1}B\mathbf{\Delta}_S \tilde{y}_{t-m}.$$

The representation then follows from setting

$$v_t := (I_n - \underline{A}^S)^{-1}\underline{A}^S \sum_{m=1}^{t-1} \underline{A}^{m-1}B\mathbf{\Delta}_S \tilde{y}_{t-m}, \quad v_1 := x_1.$$

The second result follows from $\tilde{y}_t = \varepsilon_t + \sum_{m=1}^{t-1} CA^{m-1}B\varepsilon_{t-m} + CA^{t-1}x_1$ and the fact that $\sum_{m=0}^{j} K_m K_{j-m}^- = 0, j > 0$.
With respect to deterministics note that for $s_{t,j} = \overline{z}_j^{t-1}$, $j = 1, \ldots, \tilde{S}$ we obtain

$$\begin{aligned}\sum_{m=0}^{t-1} K_m^- d_j s_{t-m,j} &= k^{-1}(z_j)d_j s_{t,j} - \sum_{m=t}^{\infty} K_m^- d_j \overline{z}_j^{t-m-1} \\ &= -\Pi_j d_j s_{t,j} + C\underline{A}^{t-1}(I_n - z_j\underline{A})^{-1}Bz_j d_j.\end{aligned}$$

For the linear trend we obtain

$$\begin{aligned}\sum_{m=0}^{t-1} K_m^- d_{S+1}(t-m) &= \sum_{m=0}^{\infty} K_m^- d_{S+1}(t-m) - \sum_{m=t}^{\infty} K_m^- d_{S+1}(t-m)\\ &= -\Pi_1 d_{S+1} t + \Psi d_{S+1} + C\underline{A}^{t-1} \sum_{m=t}^{\infty} \underline{A}^{m-t} B d_{S+1}(m-t)\\ &= -\Pi_1 d_{S+1} t + \Psi d_{S+1} + C\underline{A}^{t-1}(I_n - \underline{A})^{-2} d_{S+1} - C\underline{A}^{t-1}(I_n - \underline{A})^{-1} d_{S+1}.\end{aligned}$$

The above terms with factor $C\underline{A}^{t-1}$ are included in the SSECM through starting value v_1. ■

B.2 Auxiliary Results for the SSECM

Having introduced the residuals in the SSECM, we have used invertibility of a matrix appearing in the first order conditions. The necessary conditions are given in the following Lemma.

Lemma 3 *Let $\{y_t\}_{t\in\mathbb{Z}}$ be an MFI(1) process and let $(A(\theta), B(\theta), C(\theta))$ be a controllable state space system such that the process $\{v_t(\theta)\}_{t\in\mathbb{Z}}$ is defined as in Theorem 6.*

(I) Let $B_{\mathbb{R}}(\theta) := [\underline{B}_{k_0}^{\mathbb{R}}(\theta), \underline{B}_{-k_0}^{\mathbb{R}}(\theta)]$ be of full column rank. Then the matrix

$$\begin{bmatrix} \langle v_t(\theta), v_t(\theta)\rangle & B_{\mathbb{R}}(\theta) \\ B_{\mathbb{R}}(\theta)' & 0 \end{bmatrix}$$

is of full rank.

(II) Let $V_t^{ex}(\boldsymbol{\beta}, \theta) := [Z_{2,t}', v_t(\theta)', (\boldsymbol{\beta}' Z_{1,t})']'$ and $G(\boldsymbol{\beta}, \theta) := \begin{bmatrix} 0 & -I_{(S-2)s} \\ \underline{B}_{k_0}^{\mathbb{R}}(\theta) & \underline{B}_{-k_0}^{\mathbb{R}}(\theta) \\ -\boldsymbol{\beta}' & 0 \end{bmatrix}$ with $\boldsymbol{\beta} \in \mathbb{G}(r, \theta)$. Then the matrix

$$\begin{bmatrix} \langle V_t^{ex}(\boldsymbol{\beta}, \theta), V_t^{ex}(\boldsymbol{\beta}, \theta)\rangle & G(\boldsymbol{\beta}, \theta) \\ G(\boldsymbol{\beta}, \theta)' & 0 \end{bmatrix}$$

is of full rank.

Proof: To simplify notation, omit the dependence on θ in the following. (I) Consider vectors $\eta \in \mathbb{R}^n, \zeta \in \mathbb{R}^{Ss}$ satisfying

$$[\eta', \zeta'] \begin{bmatrix} \langle v_t, v_t\rangle & B_{\mathbb{R}} \\ B_{\mathbb{R}}' & 0 \end{bmatrix} = 0 \tag{B.1}$$

such that $\eta' \langle v_t, v_t\rangle + \zeta' B_{\mathbb{R}}' = 0, \eta' B_{\mathbb{R}} = 0$. Multiplying (B.1) by $[\eta', \zeta']'$ from the right yields $\eta' \langle v_t, v_t\rangle \eta + \zeta' B_{\mathbb{R}}' \eta + \eta' B_{\mathbb{R}} \zeta = 0$. Since $\eta' B_{\mathbb{R}} = 0$, it follows that $\eta' \langle v_t, v_t\rangle \eta = 0$, which holds true if and only if $\eta' v_t = 0$ for all $t = 1, \ldots, T$. Summarizing into a single equation for η, it holds that $\eta'[\langle v_t, v_t\rangle, B_{\mathbb{R}}] = 0$. Rewrite the SSECM equations

$$\begin{aligned}\boldsymbol{\Delta}_S \tilde{y}_t &= \sum_{k=1}^{S} \Pi_k X_t^{(k)} - C v_t + \varepsilon_t(A, B, C)\\ \Pi_k &= \big(-I_s + z_k C(I_n - z_k \underline{A})^{-1} B\big) = -I_s + C\underline{B}_k\end{aligned}$$

using the identity $\sum_{k=1}^{S} \boldsymbol{\Delta}_{S,k} \tilde{y}_{t-1} = \tilde{y}_{t-S}$ into

$$\tilde{y}_t = C\left(\sum_{k=1}^{S} \underline{B}_k X_t^{(k)} - v_t\right) + \varepsilon_t(A, B, C),$$

which implies $x_t = \sum_{k=0}^{S} \underline{B}_k X_t^{(k)} - v_t$. Consequently $\eta'[\langle v_t, v_t\rangle, B_{\mathbb{R}}] = 0$ implies $\eta' x_t = 0$, which is a contradiction to the controllability of the state space system (A, B, C). It follows that $\eta = 0$. Analogously $\zeta = 0$, since $\zeta \neq 0$ then implies $\zeta' B_{\mathbb{R}}' = 0$, which is a contradiction to the full column rank assumption on $B_{\mathbb{R}}$. Since both vectors are shown to be zero for (B.1) to hold, the matrix is indeed of full rank.

(II) The arguments proceed analogously. Here v_t is augmented by components of the form $(Z_{2,t})'$ and $\boldsymbol{\beta}' Z_{1,t}$ and $B_{\mathbb{R}}$ is extended to a matrix $\tilde{B}_{\mathbb{R}}$ accordingly. Again, assuming

$$\left[\tilde{\eta}', \tilde{\zeta}'\right] \begin{bmatrix} \langle V_t^{ex}(\boldsymbol{\beta}), V_t^{ex}(\boldsymbol{\beta})\rangle & G(\boldsymbol{\beta}) \\ G(\boldsymbol{\beta})' & 0 \end{bmatrix} = 0$$

for vectors $\tilde{\eta}' \in \mathbb{R}^{(S-2)s+n+r_a}, \tilde{\zeta} \in \mathbb{R}^{Ss}$, it follows that $\tilde{\eta}' \langle V_t^{ex}(\boldsymbol{\beta}), V_t^{ex}(\boldsymbol{\beta})\rangle \tilde{\eta} = 0$. Since $\langle Z_{1,t}, Z_{1,t}\rangle > 0$ and $\langle Z_{2,t}, Z_{2,t}\rangle > 0$, the vector $\tilde{\eta}$ is then of the form $\tilde{\eta} = [0_{(S-2)s}', \eta', 0_{r_a}']'$, which again implies $\eta' v_t = 0$. From here the proof proceeds as in (I). ■

Note that the assumption of $(A(\theta), B(\theta), C(\theta))$ being a controllable state space system is not a restriction in application, as the parameter space is usually chosen to consist of controllable systems in order to ensure identification. Moreover, any non-controllable system can be reduced in such a way that the resulting subsystem will be controllable and lead to the same residuals and consequently the same likelihood value as the original system.

As a next step let us present a different representation for the residuals in the SSECM, which will be used to prove consistency and derive the asymptotic distribution of the estimator $\boldsymbol{\beta}$ containing information on the cointegrating space. For this, let us briefly recall the derivation of the residuals $R_{0,t}(\theta), R_{1,t}(\theta)$ as well as $\varepsilon_t^R(\boldsymbol{\beta}, \theta)$ and define another auxiliary times series $R_{1,t}^R(\boldsymbol{\beta}, \theta)$. For the unrestricted concentration approach the residuals $R_{0,t}(\theta), R_{1,t}(\theta)$ are given by

$$\begin{aligned} R_{0,t}(\theta) &:= Z_{0,t} - \langle Z_{0,t}, V_t(\theta)\rangle \langle V_t(\theta), V_t(\theta)\rangle^{-1} V_t(\theta), \\ R_{1,t}(\theta) &:= Z_{1,t} - \langle Z_{1,t}, V_t(\theta)\rangle \langle V_t(\theta), V_t(\theta)\rangle^{-1} V_t(\theta), \end{aligned}$$

where $V_t(\theta) := [Z_{2,t}', v_t(\theta)']'$ contains stationary components and stochastic trends and cycles at frequencies other than the frequency of interest.

For the restricted approach first consider a decomposition of Π_{k_0} into two terms:

$$\Pi_{k_0} = [\alpha_\gamma, \alpha][\gamma, \beta]' = \alpha_\gamma \gamma' + \alpha \beta'$$

where $\alpha_\gamma, \gamma \in \mathbb{C}^{s\times(s-r)}$ and $\alpha, \beta \in \mathbb{C}^{s\times r}$, where γ and β are of full column rank, where additionally we assume $\gamma'\beta = 0$. Consider now the problem of maximizing

$$L_T^{ex}(\breve{\alpha}_\gamma \boldsymbol{\gamma}' + \breve{\alpha}\boldsymbol{\beta}', C_{-k_0}, \theta) = L_T^{ex}(\breve{\alpha}_\gamma \boldsymbol{\gamma}' + \breve{\alpha}\boldsymbol{\beta}', [\boldsymbol{\Pi}_{-k_0}, C], \theta)$$

subject to $J(\breve{\alpha}_\gamma) := \left[[I_s, 0] + \breve{\alpha}_\gamma \boldsymbol{\gamma}' \quad I^{\mathbb{R}}\right] = \left[\begin{array}{ccc} \boldsymbol{\Pi}_{-k_0} & C & \breve{\alpha} \end{array}\right] G(\boldsymbol{\beta}, \theta)$ for given θ and $\boldsymbol{\beta} \in \mathbb{G}(r, \theta)$. Solving for $C_{k_0}^{ex} = \left[\begin{array}{ccc} \boldsymbol{\Pi}_{-k_0} & C & \breve{\alpha} \end{array}\right]$, we arrive at a system of first order equations of the form

$$\begin{bmatrix} C_{k_0}^{ex,R}(\boldsymbol{\beta}, \theta, \breve{\alpha}_\gamma) & \Lambda \end{bmatrix} \begin{bmatrix} \langle V_t^{ex}(\boldsymbol{\beta}, \theta), V_t^{ex}(\boldsymbol{\beta}, \theta)\rangle & G(\boldsymbol{\beta}, \theta) \\ G(\boldsymbol{\beta}, \theta)' & 0 \end{bmatrix} = \begin{bmatrix} -\langle Z_{0,t} - \breve{\alpha}_\gamma \boldsymbol{\gamma}' Z_{1,t}, V_t^{ex}(\boldsymbol{\beta}, \theta)\rangle & J(\breve{\alpha}_\gamma) \end{bmatrix},$$

where Λ is the Lagrange multiplier matrix and $V_t^{ex}(\boldsymbol{\beta}, \theta) = [Z_{2,t}', v_t(\theta)', (\boldsymbol{\beta}' Z_{1,t})']'$. Using the blocks of the inverse matrix

$$\begin{bmatrix} H_{11}(\boldsymbol{\beta}, \theta) & H_{12}(\boldsymbol{\beta}, \theta) \\ H_{21}(\boldsymbol{\beta}, \theta) & H_{22}(\boldsymbol{\beta}, \theta) \end{bmatrix} := \begin{bmatrix} \langle V_t^{ex}(\boldsymbol{\beta}, \theta), V_t^{ex}(\boldsymbol{\beta}, \theta)\rangle & G(\boldsymbol{\beta}, \theta) \\ G(\boldsymbol{\beta}, \theta)' & 0 \end{bmatrix}^{-1},$$

we rewrite the first order equations to find $C^{ex,R}(\boldsymbol{\beta}, \theta, \breve{\alpha}_\gamma)$ equal to

$$C^{ex,R}(\boldsymbol{\beta}, \theta, \breve{\alpha}_\gamma) = -\langle Z_{0,t} - \breve{\alpha}_\gamma \boldsymbol{\gamma}' Z_{1,t}, V_t^{ex}(\boldsymbol{\beta}, \theta)\rangle H_{11}(\boldsymbol{\beta}, \theta) + J(\breve{\alpha}_\gamma) H_{21}(\boldsymbol{\beta}, \theta),$$

which maximizes the pseudo log-likelihood function under the given restrictions up to the choice of $\breve{\alpha}_\gamma$. Choosing $\breve{\alpha}_\gamma = 0$ results in $C_{k_0}^{ex,R}(\boldsymbol{\beta},\theta) = \left[\begin{array}{cc} C_{-k_0}^R(\boldsymbol{\beta},\theta) & \breve{\alpha}^R(\boldsymbol{\beta},\theta) \end{array}\right] := C^{ex,R}(\boldsymbol{\beta},\theta,0)$ as described in Section 2.4. Define the subblocks $H_{21,1}$ corresponding to the block $[I_s,0]+\breve{\alpha}_\gamma\boldsymbol{\gamma}'$ of $J(\breve{\alpha}_\gamma)$ and $H_{21,2}(\boldsymbol{\beta},\theta)$ corresponding to $I^{\mathbb{R}}$ of $H_{21}(\boldsymbol{\beta},\theta)$, recall the definition of $\varepsilon_t^R(\boldsymbol{\beta},\theta)$ and define $R_{1,t}^R(\boldsymbol{\beta},\theta)$

$$\begin{aligned}
\varepsilon_t^R(\boldsymbol{\beta},\theta) &:= Z_{0,t} - \left(\langle Z_{0,t}, V_t^{ex}(\boldsymbol{\beta},\theta)\rangle H_{11}(\boldsymbol{\beta},\theta) - J(0)H_{21}(\boldsymbol{\beta},\theta)\right) V_t^{ex}(\boldsymbol{\beta},\theta),\\
R_{1,t}^R(\boldsymbol{\beta},\theta) &:= Z_{1,t} - \left(\langle Z_{1,t}, V_t^{ex}(\boldsymbol{\beta},\theta)\rangle H_{11}(\boldsymbol{\beta},\theta) + H_{21,1}(\boldsymbol{\beta},\theta)\right) V_t^{ex}(\boldsymbol{\beta},\theta).
\end{aligned}$$

This leads to a reduced model of the form

$$\varepsilon_t^R(\boldsymbol{\beta},\theta) = \breve{\alpha}_\gamma\boldsymbol{\gamma}' R_{1,t}^R(\boldsymbol{\beta},\theta) + \varepsilon_t \tag{B.2}$$

Further, define $S_{11}^R(\boldsymbol{\beta},\theta) := \left\langle R_{1,t}^R(\boldsymbol{\beta},\theta), R_{1,t}^R(\boldsymbol{\beta},\theta)\right\rangle$ and $S_{1\varepsilon}^R(\boldsymbol{\beta},\theta) := \left\langle R_{1,t}^R(\boldsymbol{\beta},\theta), \varepsilon_t^R(\boldsymbol{\beta},\theta)\right\rangle =: S_{1\varepsilon}^R(\boldsymbol{\beta},\theta)'$. Setting $\breve{\alpha}_\gamma = 0$ results in an estimator $\breve{\alpha}^R(\boldsymbol{\beta},\theta)\boldsymbol{\beta}'$ corresponding to Π_{k_0} of reduced rank r. Estimating $\breve{\alpha}_\gamma$ by ordinary least squares leads to an estimated matrix Π_{k_0} of full rank with probability one. The relations between the residuals in the state space model and the SSECM corresponding to estimators of Π_{k_0} with different rank is summarized in the next lemma.

Lemma 4 *Let $\{y_t\}_{t\in\mathbb{Z}}$ be an MFI(1) process and let $(A(\theta), B(\theta), C(\theta))$ be a controllable state space system.*

(I) Define $C^{SS}(\theta) := \langle y_t, x_t(\theta)\rangle \langle x_t(\theta), x_t(\theta)\rangle^{-1}$, such that C^{SS} is the OLS estimator in the regression $y_t = Cx_t(\theta)+\varepsilon_t$ and let $\varepsilon_t^{SS}(\theta) := y_t - C^{SS}x_t(\theta)$ be the corresponding residuals. Then

$$\varepsilon_t^{SS}(\theta) = \varepsilon_t^R(I_{2s},\theta) = \varepsilon_t^R(\boldsymbol{\beta},\theta) - S_{\varepsilon 1}^R(\boldsymbol{\beta},\theta)\boldsymbol{\gamma}(\boldsymbol{\gamma}' S_{11}^R(\boldsymbol{\beta},\theta)\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}' R_{1,t}^R(\boldsymbol{\beta},\theta)$$

for arbitrary choice of $\boldsymbol{\beta}\in\mathbb{G}(r,\theta)$ and corresponding $\boldsymbol{\gamma} = [\gamma]^{\mathbb{R}}$, $\gamma := (\beta)_\perp \in \mathbb{C}^{s\times(s-r)}$. It follows that

$$\boldsymbol{\Pi}_{k_0}^{SS}(\theta)\boldsymbol{\gamma} = \breve{\alpha}^R(I_{2s})\boldsymbol{\gamma} = S_{\varepsilon 1}^R(\boldsymbol{\beta},\theta)\boldsymbol{\gamma}(\boldsymbol{\gamma}' S_{11}^R(\boldsymbol{\beta},\theta)\boldsymbol{\gamma})^{-1},$$

for $\boldsymbol{\Pi}_{k_0}^{SS}(\theta) := -[I_s,0] + C^{SS}(\theta)\underline{B}_{k_0}^{\mathbb{R}}(\theta)$.

(II) For $\boldsymbol{\beta}\in\mathbb{G}(r,\theta)$ and corresponding $\boldsymbol{\gamma}$, $\varepsilon_t^R(\boldsymbol{\beta},\theta)$ is equal to

$$\varepsilon_t^R(\boldsymbol{\beta},\theta) = \varepsilon_t^{SS}(\theta) + \boldsymbol{\Pi}_{k_0}^{SS}(\theta)\boldsymbol{\gamma}\left(\boldsymbol{\gamma}' N_T(\theta)\boldsymbol{\gamma}\right)^{-1}\boldsymbol{\gamma}'\underline{B}_{k_0}^{\mathbb{R}}(\theta)' \langle x_t(\theta), x_t(\theta)\rangle^{-1}\hat{x}_t(\theta),$$

where $N_T(\theta) := \underline{B}_{k_0}^{\mathbb{R}}(\theta)' \langle x_t(\theta), x_t(\theta)\rangle^{-1} \underline{B}_{k_0}^{\mathbb{R}}(\theta)$. Moreover,

$$\boldsymbol{\gamma}' S_{11}^R(\boldsymbol{\beta},\theta)\boldsymbol{\gamma} = \left(\boldsymbol{\gamma}' N_T(\theta)\boldsymbol{\gamma}\right)^{-1}.$$

Proof: To simplify notation, omit the dependence on θ in the following.
(I) Due to Theorem 6, the residuals ε_t^{SS} are equal to $\varepsilon_t^R(I_{2s})$, since estimating C for given regressors x_t using ordinary least squares leads to the same residuals as maximizing the likelihood of the SSECM under the restrictions $\Pi_k = -I_s + C(I_n - z_k\underline{A})^{-1}B$ for $k=1,\dots,S$ through the restricted concentration approach (but without restrictions on the rank of Π_{k_0}). It follows that $\boldsymbol{\Pi}_{k_0}^{SS}(\theta) = \breve{\alpha}^R(I_{2s})$.
The residuals $\varepsilon_t^R(I_{2s})$ can also be computed using the reduced model (B.2) such that

$$\varepsilon_t^R(I_{2s}) = \varepsilon_t^R(\boldsymbol{\beta}) - S_{\varepsilon 1}^R(\boldsymbol{\beta})\boldsymbol{\gamma}(\boldsymbol{\gamma}' S_{11}^R(\boldsymbol{\beta})\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}' R_{1,t}^R(\boldsymbol{\beta}),$$

for $\boldsymbol{\beta}\in\mathbb{G}(r,\theta)$ and corresponding $\boldsymbol{\gamma}$. It follows that $\boldsymbol{\Pi}_{k_0}^{SS} = \breve{\alpha}^R(\boldsymbol{\beta})\boldsymbol{\beta}' + S_{\varepsilon 1}^R(\boldsymbol{\beta})\boldsymbol{\gamma}(\boldsymbol{\gamma}' S_{11}^R(\boldsymbol{\beta})\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'$ for arbitrary choice of $\boldsymbol{\beta}\in\mathbb{G}(r,\theta)$ and corresponding $\boldsymbol{\gamma}$, which implies the above identity.

(II) Consider $y_t = Cx_t + \varepsilon_t$, and maximize

$$-T \log |\langle y_t - Cx_t, y_t - Cx_t \rangle|$$

subject to $(-[I_s, 0] + C\underline{B}_{k_0}^{\mathbb{R}})\boldsymbol{\gamma} = 0$, for given matrix $\boldsymbol{\gamma} = [\gamma]^{\mathbb{R}}$, $\gamma \in \mathbb{C}^{s\times(s-r)}$ such that $\underline{B}_{k_0}^{\mathbb{R}}\boldsymbol{\gamma}$ is of full rank. Note that $(-[I_s, 0] + C\underline{B}_{k_0}^{\mathbb{R}})\boldsymbol{\gamma} = 0$ if and only if $(-[I_s, 0] + C\underline{B}_{k_0}^{\mathbb{R}}) = \breve{\alpha}\boldsymbol{\beta}$ for a matrix $\boldsymbol{\beta} := [\beta]^{\mathbb{R}}$ with $\beta \in \mathbb{C}^{s\times r}$ of full rank and $\beta'\gamma = 0$. Moreover, $\underline{B}_{k_0}^{\mathbb{R}}\boldsymbol{\gamma}$ being of full rank is equivalent to $\boldsymbol{\beta} \in \mathbb{G}(r, \theta)$. The optimal $\tilde{C}(\boldsymbol{\gamma})$ fulfilling the restriction is given by

$$\begin{pmatrix} \tilde{C}(\boldsymbol{\gamma}) & \Lambda \end{pmatrix} = \begin{pmatrix} \langle y_t, x_t \rangle & [I_s, 0]\boldsymbol{\gamma} \end{pmatrix} \begin{pmatrix} \langle x_t, x_t \rangle & \underline{B}_{k_0}^{\mathbb{R}}\boldsymbol{\gamma} \\ \boldsymbol{\gamma}'(\underline{B}_{k_0}^{\mathbb{R}})' & 0 \end{pmatrix}^{-1}.$$

By Lemma 3(I) the inverse of the above matrix exists. Using block matrix inversion, the corresponding residuals are equal to

$$\begin{aligned} & y_t - \tilde{C}(\boldsymbol{\gamma})x_t \\ &\quad = (y - C^{SS}x_t) + (C^{SS}\underline{B}_{k_0}^{\mathbb{R}} - [I_s, 0])\boldsymbol{\gamma} \left(\boldsymbol{\gamma}'(\underline{B}_{k_0}^{\mathbb{R}})' \langle x_t, x_t \rangle^{-1} \underline{B}_{k_0}^{\mathbb{R}}\boldsymbol{\gamma}\right)^{-1} \boldsymbol{\gamma}'(\underline{B}_{k_0}^{\mathbb{R}})' \langle x_t, x_t \rangle^{-1} x_t \\ &\quad = \varepsilon_t^{SS} + \boldsymbol{\Pi}_{k_0}^{SS}\boldsymbol{\gamma} \left(\boldsymbol{\gamma}'(\underline{B}_{k_0}^{\mathbb{R}})' \langle x_t, x_t \rangle^{-1} \underline{B}_{k_0}^{\mathbb{R}}\boldsymbol{\gamma}\right)^{-1} \boldsymbol{\gamma}'(\underline{B}_{k_0}^{\mathbb{R}})' \langle x_t, x_t \rangle^{-1} x_t. \end{aligned}$$

Due to Theorem 6, these residuals are equal to $\varepsilon_t^R(\boldsymbol{\beta})$ for corresponding $\boldsymbol{\beta} \in \mathbb{G}(r, \theta)$, thus, giving an alternative derivation of the residuals in the restricted concentration approach. Together with the results in (I) it follows that

$$\boldsymbol{\gamma}' R_{1,t}^R(\boldsymbol{\beta}) = (\boldsymbol{\gamma}' N_T \boldsymbol{\gamma})^{-1} \boldsymbol{\gamma}'(\underline{B}_{k_0}^{\mathbb{R}})' \langle x_t, x_t \rangle^{-1} x_t,$$

which implies the last equality. ■

B.3 Asymptotic Results in the SSECM

Next, we analyze the asymptotic behavior of the product moments of $R_{0,t}(\theta)$,$R_{1,t}(\theta)$, $\varepsilon_t^R(\boldsymbol{\beta}, \theta)$ and $R_{1,t}^R(\boldsymbol{\beta}, \theta)$. The following Lemma will be fundamental for this section:

Lemma 5 *The following convergence results hold for processes satisfying $x_{t+1,k} = \bar{z}_k x_{t,k} + B_k \varepsilon_{t+1}$ with $z_k = e^{i\omega_k}$, $0 < \omega_k < \pi$, and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ denoting a white noise process with noise variance Σ as in the assumptions stated in Assumption 1.*

$$\begin{aligned} \frac{1}{T} \left\langle [2x_{t,k}]_v^{\mathbb{R}}, [2x_{t,k}]_v^{\mathbb{R}} \right\rangle &\xrightarrow{d} [B_k]^{\mathbb{R}} \int_0^1 \boldsymbol{W}\boldsymbol{W}' du \; ([B_k]^{\mathbb{R}})' \\ \left\langle [2x_{t,k}]_v^{\mathbb{R}}, \varepsilon_t \right\rangle &\xrightarrow{d} [B_k]^{\mathbb{R}} \int_0^1 \boldsymbol{W}\left(d\boldsymbol{W}'\right) \begin{bmatrix} I_s \\ 0 \end{bmatrix}, \\ \left\langle [2x_{t,k}]_v^{\mathbb{R}}, [2x_{t,m}]_v^{\mathbb{R}} \right\rangle &= O_p(1) \quad \text{for } k \neq m, \end{aligned}$$

where $\boldsymbol{W} = [\frac{1}{\sqrt{2}}(W_1 + iW_2)]^{\mathbb{R}}$ and W_1 and W_2 are two independent s-dimensional Brownian motions with variance Σ.

Proof: The first two results are proven by Johansen and Schaumburg, exploiting the notation of matrices in complex structure, compare Johansen and Schaumburg (1999, proof of Theorem 9). The last result concerning stochastic boundedness follows from the fact that

$$\left\langle [2x_{t,k}]^{\mathbb{R}}, [2x_{t,m}]^{\mathbb{R}} \right\rangle = \left\langle [2x_{t,k,m}]^{\mathbb{R}}, [2B_m\varepsilon_t]^{\mathbb{R}} \right\rangle + O_p(1)$$

with $x_{t,k,m} := \bar{z}_m x_{t,k,m} + x_{t,k}$. The process $[2x_{t,k,m}]^{\mathbb{R}}$ can then be expressed as the sum of two processes integrated at a single frequency. The second convergence result applied for both terms then implies that $\left\langle [2x_{t,k}]^{\mathbb{R}}, [2x_{t,m}]^{\mathbb{R}} \right\rangle$ is bounded in probability. Since the same holds true for $\left\langle [2x_{t,k}]^{\mathbb{R}}, [2\bar{x}_{t,m}]^{\mathbb{R}} \right\rangle$ combining both results ensures that $\left\langle [2x_{t,k}]^{\mathbb{R}}_v, [2x_{t,m}]^{\mathbb{R}}_v \right\rangle$ is also stochastically bounded. ■

Lemma 5 is the key component in the proof of the asymptotic distribution of the pseudo likelihood ratio tests. Since the proof proceeds analogously for real and complex unit roots we will only focus on the case of complex unit roots.

Before we proceed to the product moments, let us first discuss the dominant components in the $R_{0,t}(\theta)$,$R_{1,t}(\theta)$, $\varepsilon^R_t(\boldsymbol{\beta},\theta)$ and $R^R_{1,t}(\boldsymbol{\beta},\theta)$.

Lemma 6 *Let $\{y_t\}_{t\in\mathbb{Z}}$ be an MFI(1) process solving the SSECM equations as defined in Theorem 6 and let $\hat\theta$ be a PML estimator over Θ fulfilling the assumptions of Proposition 2. Let $\hat v_t := v_t(\hat\theta)$, $\hat R_{i,t} := R_{i,t}(\hat\theta)$, $\hat\varepsilon^R_t(\boldsymbol{\beta}) := \varepsilon^R_t(\boldsymbol{\beta},\hat\theta)$ and $\hat R^R_{1,t}(\boldsymbol{\beta}) := R^R_{1,t}(\boldsymbol{\beta},\hat\theta)$.*

(I) There exists matrices $M^U_{0,T}$ and $M^U_{1,T}$, bounded in probability, such that the following identities hold:

$$\hat R_{0,t} = Z_{0,t} - M^U_{0,T}[\frac{1}{T}(P^u Z_{2,t})', (P^{st}Z_{2,t})', \hat v_t']',$$
$$\hat R_{1,t} = Z_{1,t} - M^U_{1,T}[\frac{1}{T}(P^u Z_{2,t})', (P^{st}Z_{2,t})', \hat v_t']',$$

where $P^{st} := diag\left(\beta_1', \beta_{\tilde S}', [\beta_2']^{\mathbb{R}}, \dots, [\beta'_{k_0-1}]^{\mathbb{R}}, [\beta'_{k_0+1}]^{\mathbb{R}}, ..., [\beta'_{\tilde S-1}]^{\mathbb{R}}\right) \in \mathbb{R}^{(\sum_1^S r_k)\times s(S-2)}$,

and $P^u := diag\left(\gamma_1', \gamma_{\tilde S}', [\gamma_2']^{\mathbb{R}}, \dots, [\gamma'_{k_0-1}]^{\mathbb{R}}, [\gamma'_{k_0+1}]^{\mathbb{R}}, ..., [\gamma'_{\tilde S-1}]^{\mathbb{R}}\right) \in \mathbb{R}^{(sS-\sum_1^S r_k)\times s(S-2)}$,

for $\beta_k, \alpha_k \in \mathbb{R}^{s\times r_k}$ such that $\Pi_{k,\circ} = \alpha_k\beta_k'$ and $\gamma_k := (\beta_k)_\perp \in \mathbb{R}^{s\times(s-r_k)}$.

(II) Define $\boldsymbol{\beta}_\circ := [\beta_{k_0}]^{\mathbb{R}}$. There exists matrices $M^R_{0,T}$ and $M^R_{1,T}$ bounded in probability, such that the following identities hold:

$$\hat\varepsilon^R_t(\boldsymbol{\beta}_\circ) = Z_{0,t} - M^R_{0,T}[\frac{1}{T}(P^u Z_{2,t})', (P^{st}Z_{2,t})', \hat v_t', (\boldsymbol{\beta}_\circ' Z_{1,t})']',$$
$$\hat R^R_{1,t}(\boldsymbol{\beta}_\circ) = Z_{1,t} - M^R_{1,T}[\frac{1}{T}(P^u Z_{2,t})', (P^{st}Z_{2,t})', \hat v_t', (\boldsymbol{\beta}_\circ' Z_{1,t})']'.$$

Let $\beta_{(j)}$, for $j = 1,\dots,r$ denote the j-th column of β_{k_0}, define

$$\beta_{(-j)} := [\beta_{(1)}, \dots, \beta_{(j-1)}, \beta_{(j+1)}, \dots, \beta_{(r)}]$$

and let $\gamma_{(j)} \in \mathbb{C}^{s\times(s-1)}$ be orthogonal to $\beta_{(j)}$, such that $\beta_{(j)}'\gamma_{(j)} = 0$. Let $\boldsymbol{\beta}_{(j)} := [\beta_{(j)}]^{\mathbb{R}}$, $\boldsymbol{\beta}_{(-j)} := [\beta_{(-j)}]^{\mathbb{R}}$ and $\boldsymbol{\gamma}_{(j)} := [\gamma_{(j)}]^{\mathbb{R}}$. Then there exists matrices $M^{(j)}_{0,T}$, bounded in probability, such that

$$\boldsymbol{\beta}_{(j)}'\hat R^R_{1,t}(\boldsymbol{\gamma}_{(j)}) = \boldsymbol{\beta}_{(j)}' Z_{1,t} - M^{(j)}_{1,T}[\frac{1}{T}(P^u Z_{2,t})', \frac{1}{T}(\boldsymbol{\gamma}' Z_{1,t})', (P^{st}Z_{2,t})', \hat v_t', (\boldsymbol{\beta}_{(-j)}' Z_{1,t})']'.$$

Proof: Define $P^Z := [(P^u)', (P^{st})']'$, $P = \text{diag}(P^Z((P^Z)'P^Z)^{-1/2}, I_n)$, such that $P'P = I_{n+sS}$, and $D_T = \text{diag}(\frac{1}{T}I_{sS-\sum_1^S r_k}, I_{n+\sum_1^S r_k})$ and let $\hat V_t := V_t(\hat\theta)$. In case of an unrestricted concentration step it suffices to consider

$$\begin{aligned}
&\left\langle R_{0,t}, \hat V_t \right\rangle P'P \left\langle \hat V_t, \hat V_t \right\rangle^{-1} P' D_T^{-1} \\
&\quad = \left\langle R_{0,t}, P\hat V_t \right\rangle \left\langle D_T P\hat V_t, P\hat V_t \right\rangle^{-1} \\
&\quad = \left\langle R_{0,t}, P\hat V_t \right\rangle \begin{pmatrix} \frac{1}{T}\left\langle P^u Z_{2,t}, P^u Z_{2,t}\right\rangle & \frac{1}{T}\left\langle P^u Z_{2,t}, V_t^{st}\right\rangle \\ \left\langle V_t^{st}, P^u Z_{2,t}\right\rangle & \left\langle V_t^{st}, V_t^{st}\right\rangle \end{pmatrix}^{-1} := M^U_{0,T},
\end{aligned}$$

where $\hat{V}_t^{st} := [(P^{st}Z_{2,t})', \hat{v}_t']'$. Note that

$$\begin{pmatrix} \frac{1}{T}\langle P^u Z_{2,t}, P^u Z_{2,t}\rangle & \frac{1}{T}\left\langle P^u Z_{2,t}, \hat{V}_t^{st}\right\rangle \\ \left\langle \hat{V}_t^{st}, P^u Z_{2,t}\right\rangle & \left\langle \hat{V}_t^{st}, \hat{V}_t^{st}\right\rangle \end{pmatrix} \xrightarrow{d} \begin{bmatrix} \mathcal{K}_{1,1} & 0 \\ \mathcal{K}_{2,1} & \mathcal{K}_{2,2}\end{bmatrix} =: \mathcal{K},$$

using Lemma 5 and that $\mathcal{K}$ is invertible with probability one, since $\mathcal{K}_{1,1}$ is a random matrix invertible with probability one and $\mathcal{K}_{2,2}$ is determinstic and invertible. Here we also use that consistency of $\underline{\mathcal{A}}(\hat{\theta})$ and $\mathcal{B}(\theta)$, which implies $\langle \hat{v}_t, \hat{v}_t\rangle$ converges to $\mathrm{Var}(v_t(\theta_0))$. Consequently, $M_{0,T}^U$ converges to a random matrices $M_{0,\infty}^U$ and is, thus, bounded in probability. It follows that

$$\begin{aligned} R_{0,t} &= Z_{0,t} - \left(\left\langle Z_{0,t}, \hat{V}_t\right\rangle \left\langle \hat{V}_t, \hat{V}_t\right\rangle^{-1} P' D_T^{-1}\right) D_T P \hat{V}_t \\ &= Z_{0,t} - M_{0,T}^U D_T P \hat{V}_t. \end{aligned}$$

An analogous result holds for $R_{1,t}$ using

$$M_{1,T}^U := \left\langle R_{1,t}, \hat{V}_t\right\rangle \left\langle \hat{V}_t, \hat{V}_t\right\rangle^{-1} P' D_T^{-1},$$

together with the fact that $\langle R_{1,t}, \hat{V}_t\rangle$ is bounded in probability by the third result in Lemma 5. For the restricted concentration step consider similarly $P^{ex} := \mathrm{diag}(P, I_r)$, $D_T^{ex} := \mathrm{diag}(D_T, I_r)$, $\hat{V}_t^{ex}(\boldsymbol{\beta}_\circ) := V_t^{ex}(\boldsymbol{\beta}_\circ, \hat{\theta})$ such that

$$\begin{aligned} &\begin{bmatrix} P^{ex} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \langle \hat{V}_t^{ex}(\boldsymbol{\beta}_\circ), \hat{V}_t^{ex}(\boldsymbol{\beta}_\circ)\rangle & G(\boldsymbol{\beta}_\circ) \\ G(\boldsymbol{\beta}_\circ)' & 0 \end{bmatrix}^{-1} \begin{bmatrix} (P^{ex})'(D_T^{ex})^{-1} & 0 \\ 0 & I \end{bmatrix} \\ &\quad = \begin{bmatrix} \langle D_T^{ex} P^{ex} \hat{V}_t^{ex}(\boldsymbol{\beta}_\circ), P^{ex}\hat{V}_t^{ex}(\boldsymbol{\beta}_\circ)\rangle & D_T^{ex} P^{ex} G(\boldsymbol{\beta}_\circ) \\ (P^{ex} G(\boldsymbol{\beta}_\circ))' & 0 \end{bmatrix}^{-1} := \begin{bmatrix} \mathcal{H}_{11,T} & \mathcal{H}_{12,T} \\ \mathcal{H}_{21,T} & \mathcal{H}_{22,T}\end{bmatrix} := \mathcal{H}_T. \end{aligned}$$

Define $\hat{V}_t^{ex,st}(\boldsymbol{\beta}_\circ) := [(P^{st}Z_{2,t})', \hat{v}_t', (\boldsymbol{\beta}_\circ' Z_{1,t})']'$ and note that

$$\begin{aligned} &\begin{bmatrix} \langle D_T^{ex} P^{ex} \hat{V}_t^{ex}(\boldsymbol{\beta}_\circ), P^{ex}\hat{V}_t^{ex}(\boldsymbol{\beta}_\circ)\rangle & D_T^{ex} P^{ex} G(\boldsymbol{\beta}_\circ) \\ (P^{ex} G(\boldsymbol{\beta}_\circ))' & 0 \end{bmatrix} := \mathcal{G}_T \\ &\quad \xrightarrow{d} \begin{bmatrix} \mathcal{K}_{1,1} & 0 & 0 \\ \tilde{\mathcal{K}}_{2,1} & \tilde{\mathcal{K}}_{2,2} & \mathrm{diag}(P^{st}, I_{n+r}) G(\boldsymbol{\beta}_\circ) \\ (P^u G(\boldsymbol{\beta}_\circ))' & (\mathrm{diag}(P^{st}, I_{n+r}) G(\boldsymbol{\beta}_\circ))' & 0 \end{bmatrix} := \mathcal{G}, \end{aligned}$$

where $\tilde{\mathcal{K}}_{2,1}$ and $\tilde{\mathcal{K}}_{2,2}$ are the limits of $\langle \hat{V}_t^{ex,st}(\boldsymbol{\beta}_\circ), P^u Z_{2,t}\rangle$ respectively $\langle \hat{V}_t^{ex,st}(\boldsymbol{\beta}_\circ), \hat{V}_t^{ex,st}(\boldsymbol{\beta}_\circ)\rangle$. The matrices $\mathcal{G}_T$ are invertible with probability one, due to Lemma 3 and the fact that $\boldsymbol{\beta}_\circ \in G(r, \theta_\circ)$ for $\theta_\circ$ corresponding to the true system. The matrix $\mathcal{K}_{1,1}$ is invertible with probability one due to Lemma 5. The same arguments as in the proof of Lemma 3 imply that the matrix

$$\begin{bmatrix} \tilde{\mathcal{K}}_{2,2} & \mathrm{diag}(P^{st}, I_{n+r}) G(\boldsymbol{\beta}_\circ) \\ (\mathrm{diag}(P^{st}, I_{n+r}) G(\boldsymbol{\beta}_\circ))' & 0 \end{bmatrix}$$

is also invertible with probability one, since otherwise a vector η such that $\eta' \mathrm{Var}(x_t x_t') = 0$ would occur with positive probability, contradicting the fact that the estimated state space system is minimal with probability one. Since the true system is also minimal, $\mathcal{G}$ is invertible with probability one and $\mathcal{H}_T$ is bounded in probability. For $\varepsilon_t^R(\boldsymbol{\beta}_\circ)$ it follows that

$$\begin{aligned} \varepsilon_t^R(\boldsymbol{\beta}_\circ) &= Z_{0,t} - \left(\left\langle Z_{0,t}, \hat{V}_t^{ex}(\boldsymbol{\beta}_\circ)\right\rangle H_{11}(\boldsymbol{\beta}_\circ) - J H_{21}(\boldsymbol{\beta}_\circ)\right) \hat{V}_t^{ex}(\boldsymbol{\beta}_\circ) \\ &= Z_{0,t} - \left(\left\langle Z_{0,t}, P^{ex}\hat{V}_t^{ex}(\boldsymbol{\beta}_\circ)\right\rangle \mathcal{H}_{11,T} - J \mathrm{diag}((P^{ex})', I) \mathcal{H}_{21,T}\right) D_T^{ex} P^{ex} \hat{V}_t^{ex}(\boldsymbol{\beta}_\circ) \\ &= Z_{0,t} - M_{0,T}^R [\frac{1}{T}(P^u Z_{2,t})', (\hat{V}_t^{ex,st}(\boldsymbol{\beta}_\circ))']', \end{aligned}$$

again with

$$M_{0,T}^{R} := \left\langle Z_{0,t}, P\hat{V}_t^{ex}(\boldsymbol{\beta}_\circ)\right\rangle \mathcal{H}_{11,T} - J\text{diag}((P^{ex})', I)\mathcal{H}_{21,T}$$

bounded in probability. The results for $R_{1,t}^R(\boldsymbol{\beta}_\circ)$ and $\boldsymbol{\beta}_{(j)}' R_{1,t}^R(\boldsymbol{\gamma}_{(j)})$ follow analogously. ■

Lemma 7 *Let $\hat{\theta}$ denote the pseudo maximum likelihood estimator as in Proposition 2 and let $\hat{S}_{ij} := S_{ij}(\hat{\theta})$, $j = 0, 1$, $\hat{S}_{11}(\boldsymbol{\beta}) := S_{11}(\boldsymbol{\beta}, \hat{\theta})$ and $\hat{S}_{1\varepsilon}^R(\boldsymbol{\beta}) := S_{1\varepsilon}^R(\boldsymbol{\beta}, \hat{\theta}) =: \hat{S}_{1\varepsilon}^R(\boldsymbol{\beta})'$.*

(I) Defining $\boldsymbol{\alpha}_\circ := [\alpha_{k_0}]^{\mathbb{R}}$, $\breve{\alpha}_\circ := [I_s, 0]\boldsymbol{\alpha}_\circ$ and

$$\hat{S}_{1\varepsilon} := \langle \hat{R}_{1,t}, \hat{R}_{0,t} - \breve{\alpha}_\circ \boldsymbol{\beta}_\circ' \hat{R}_{1,t}\rangle =: \hat{S}_{\varepsilon 1}',$$

the following convergence results hold for the unrestricted concentration approach:

$$T^{-1}\mathcal{C}_\omega' \hat{S}_{11}\mathcal{C}_\omega, \xrightarrow{d} \mathcal{B}_\omega \int_0^1 \boldsymbol{W}(u)\boldsymbol{W}(u)'du\ \mathcal{B}_\omega',$$
$$\mathcal{C}_\omega' \hat{S}_{1\varepsilon} \xrightarrow{d} \mathcal{B}_\omega \int_0^1 \boldsymbol{W}(u)\big(d\boldsymbol{W}(u)\big)' \begin{bmatrix} I_s \\ 0 \end{bmatrix}.$$

(II) For the restricted concentration approach it holds that

$$\hat{C}_{k_0}^{ex,R}(\boldsymbol{\beta}_\circ) := -\langle Z_{0,t}, \hat{V}_t^{ex}(\boldsymbol{\beta}_\circ)\rangle H_{11}(\boldsymbol{\beta}_\circ, \hat{\theta}) + JH_{21}(\boldsymbol{\beta}_\circ, \hat{\theta}).$$

is a consistent estimator of $C_{k_0}^{ex}$. Moreover, the following convergence results hold:

$$T^{-1}\mathcal{C}_\omega' \hat{S}_{11}(\boldsymbol{\beta}_\circ)\mathcal{C}_\omega \xrightarrow{d} \mathcal{B}_\omega \int_0^1 \boldsymbol{W}(u)\boldsymbol{W}(u)'du\ \mathcal{B}_\omega'$$
$$\mathcal{C}_\omega' \hat{S}_{1\varepsilon}^R(\boldsymbol{\beta}_\circ) \xrightarrow{d} \mathcal{B}_\omega \int_0^1 \boldsymbol{W}(u)\big(d\boldsymbol{W}(u)\big)' \begin{bmatrix} I_s \\ 0 \end{bmatrix}.$$

Let $\tilde{\boldsymbol{\beta}}$ denote an estimate of the cointegrating space with the property that $T(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_\circ) = O_P(1)$.

(III) For the unrestricted concentration approach it holds that

$$\tilde{\alpha} := \hat{S}_{01}\tilde{\boldsymbol{\beta}}(\tilde{\boldsymbol{\beta}}'\hat{S}_{11}\tilde{\boldsymbol{\beta}})^{-1} \xrightarrow{p} \breve{\alpha}_\circ$$
$$\langle \tilde{\varepsilon}_t, \tilde{\varepsilon}_t\rangle \xrightarrow{p} \Sigma$$
$$(\tilde{\boldsymbol{\beta}}'\hat{S}_{11,0}\tilde{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{\beta}}'\hat{S}_{10}\hat{S}_{00}^{-1} \xrightarrow{p} \breve{\alpha}_\circ'\Sigma^{-1},$$

where $\tilde{\varepsilon}_t := \hat{R}_{0,t} - \tilde{\alpha}\tilde{\boldsymbol{\beta}}\hat{R}_{1,t}$.

(IV) For the restricted concentration step, it holds that $T(\hat{C}_{k_0}^{ex,R}(\tilde{\boldsymbol{\beta}}) - \hat{C}_{k_0}^{ex,R}(\boldsymbol{\beta}_\circ)) = O_p(1)$. Moreover,

$$\left\langle \hat{\varepsilon}_t^R(\tilde{\boldsymbol{\beta}}), \hat{\varepsilon}_t^R(\tilde{\boldsymbol{\beta}})\right\rangle \xrightarrow{p} \Sigma$$
$$T^{-1}(\mathcal{C}_\omega'\hat{S}_{11}(\boldsymbol{\beta}_\circ)\mathcal{C}_\omega - \mathcal{C}_\omega'\hat{S}_{11}(\tilde{\boldsymbol{\beta}})\mathcal{C}_\omega) \xrightarrow{p} 0,$$
$$\mathcal{C}_\omega'\hat{S}_{1\varepsilon}^R(\boldsymbol{\beta}_\circ) - \mathcal{C}_\omega'\left\langle \hat{R}_{1,t}^R(\tilde{\boldsymbol{\beta}}), \hat{\varepsilon}_t^R(\boldsymbol{\beta}_\circ)\right\rangle \xrightarrow{p} 0.$$

Proof: (I) The boundedness of $\hat{S}_{00}$, $\boldsymbol{\beta}_\circ'\hat{S}_{10}$ and $\boldsymbol{\beta}_\circ'\hat{S}_{11}$ in probability is a consequence of the dominating stationary components of $\hat{R}_{0,t}$ and $\boldsymbol{\beta}_\circ'\hat{R}_{1,t}$ since the remaining terms tend to zero, as discussed in Lemma 6.

To derive the limiting distribution of $T^{-1}\mathcal{C}_\omega'\hat{S}_{11}\mathcal{C}_\omega$, note that neither stationary nor integrated components at unit root frequencies other than ω_{k_0} contained in $\hat{R}_{1,t}$ are dominant, which is clear from Lemma 6. Thus, the limit of $T^{-1}\mathcal{C}_\omega'\hat{S}_{11}\mathcal{C}_\omega$ coincides with the limit of

$$T^{-1}\mathcal{C}_\omega'\left\langle [2\mathcal{C}_\omega x_{t,k_0}]_v^{\mathbb{R}}, [2\mathcal{C}_\omega x_{t,k_0}]_v^{\mathbb{R}}\right\rangle \mathcal{C}_\omega$$

for which Lemma 5 is applied. For the asymptotics of $\mathcal{C}_\omega'\hat{S}_{1\varepsilon}$ note that

$$\left\langle \mathcal{C}_\omega'\hat{R}_{1,t}, \hat{R}_{0,t} - \breve{\alpha}_\circ\boldsymbol{\beta}_\circ'\hat{R}_{1,t}\right\rangle = \left\langle \mathcal{C}_\omega'\hat{R}_{1,t}, \varepsilon_t\right\rangle + \left\langle \mathcal{C}_\omega'\hat{R}_{1,t}, C_{-k_0}(\theta_\circ)V_t(\theta_\circ) - \hat{C}_{-k_0}^U(\hat{\theta})V_t(\hat{\theta})\right\rangle \quad \text{(B.3)}$$

where $C_{-k_0}(\theta_\circ)$ is the coefficient matrix corresponding to the true parameter vector θ_0 and

$$\hat{C}_{-k_0}^U := \left\langle Z_{0,t} - \breve{\alpha}_\circ\boldsymbol{\beta}_\circ' Z_{1,t}, V_t(\hat{\theta})\right\rangle \left\langle V_t(\hat{\theta}), V_t(\hat{\theta})\right\rangle^{-1}$$

is a consistent estimator for $C_{-k_0}(\theta_\circ)$. Here the last term in (B.3) is $o_P(1)$ as

$$C_{-k_0}(\theta_\circ)V_t(\theta_\circ) - \hat{C}_{-k_0}^U V_t(\hat{\theta}) = \sum_{j=0}^{t-2}(K_j^-(\theta_\circ) - K_j^-(\hat{\theta}))(1-L^{\tilde{S}})y_{t-1-j} + \left(C_{-k_0}(\hat{\theta}) - \hat{C}_{-k_0}^U\right)V_t(\hat{\theta})$$

where $T^\nu(K_j^-(\theta_\circ) - K_j^-(\hat{\theta})) \to 0$ for $\nu < 0.5$ (in fact $\|K_j^-(\theta_\circ) - K_j^-(\hat{\theta})\| \leq \mu\rho^j\|\theta_\circ - \hat{\theta}\|$ for some $0 < \rho < 1$ uniformly in a small enough neighborhood of $\theta_\circ$) and $V_t(\hat{\theta})$ and $(1-L^{\tilde{S}})\{y_t\}_{t\in\mathbb{Z}}$ are stationary. Then the result follows by application of Lemma 5.

(II) To see that $\hat{C}_{k_0}^{ex,R}(\boldsymbol{\beta}_\circ)$ is consistent, consider first the OLS estimator $\hat{C}_{k_0}^{ex,U}(\boldsymbol{\beta}_\circ)$ in the expanded SSECM without taking the restrictions between C and Π_k, $k = 1,\dots,S$ into account. It holds that $\hat{C}_{k_0}^{ex,U}(\boldsymbol{\beta}_\circ) = \langle Z_{0,t}, \hat{V}_t^{ex}(\boldsymbol{\beta}_\circ)\rangle\langle \hat{V}_t^{ex}(\boldsymbol{\beta}_\circ), \hat{V}_t^{ex}(\boldsymbol{\beta}_\circ)\rangle^{-1}$ if $\hat{S}_{VV}(\boldsymbol{\beta}_\circ) := \langle \hat{V}_t^{ex}(\boldsymbol{\beta}_\circ), \hat{V}_t^{ex}(\boldsymbol{\beta}_\circ)\rangle$ is invertible. Since $k(z,\hat{\theta})$ converges to the true transfer function and the OLS estimator $C_{k_0}^{ex,U}(\boldsymbol{\beta}_\circ)$ at the true parameter value $\theta_\circ$ is consistent, it holds that also $\hat{C}_{k_0}^{ex,U}(\boldsymbol{\beta}_\circ)$ is consistent. Including linear restrictions of the form $C_{k,0}^{ex}G(\boldsymbol{\beta}_\circ,\hat{\theta}) = J$, the corresponding restricted least squares estimator $\hat{C}_{k_0}^{ex,R}(\boldsymbol{\beta}_\circ)$ can be written as

$$\begin{aligned}\hat{C}_{k_0}^{ex,R}(\boldsymbol{\beta}_\circ) &= \hat{C}_{k_0}^{ex,U}(\boldsymbol{\beta}_\circ) \\ &\quad - (\hat{C}_{k_0}^{ex,U}(\boldsymbol{\beta}_\circ)G(\boldsymbol{\beta}_\circ,\hat{\theta}) - J)\left(G(\boldsymbol{\beta}_\circ,\hat{\theta})'\hat{S}_{VV}(\boldsymbol{\beta}_\circ)^{-1}G(\boldsymbol{\beta}_\circ,\hat{\theta})\right)^{-1} G(\boldsymbol{\beta}_\circ,\hat{\theta})'\hat{S}_{VV}(\boldsymbol{\beta}_\circ)^{-1},\end{aligned}$$

if $\hat{S}_{VV}(\boldsymbol{\beta}_\circ)$ is invertible. Note that $(\hat{C}_{k_0}^{ex,R}(\boldsymbol{\beta}_\circ)G(\boldsymbol{\beta}_\circ,\hat{\theta}) - J) = o_p(1)$. Thus, $\hat{C}_{k_0}^{ex,R}(\boldsymbol{\beta}_\circ)$ is consistent. Slight adaptations of the arguments show consistency of $\hat{C}_{k_0}^{ex,R}(\boldsymbol{\beta}_\circ)$ also in the case of non-invertible $\hat{S}_{VV}(\boldsymbol{\beta}_\circ)$.

The derivation of the limiting distribution of $T^{-1}\mathcal{C}_\omega'\hat{S}_{11}(\boldsymbol{\beta}_\circ)\mathcal{C}_\omega$ follows analogously to the restricted approach, noting again that $T^{-1}\left\langle [2\mathcal{C}_{k_0}x_{t,k_0}]_v^{\mathbb{R}}, [2\mathcal{C}_{k_0}x_{t,k_0}]_v^{\mathbb{R}}\right\rangle$ is the dominant term due to Lemma 6(II) and applying Lemma 5. For the last result consider

$$\hat{\varepsilon}_t^R(\boldsymbol{\beta}_\circ) = \varepsilon_t + C_{k_0}^{ex}(\theta_\circ)V_t^{ex}(\boldsymbol{\beta}_\circ,\theta_\circ) - \hat{C}_{k_0}^{ex,R}(\boldsymbol{\beta}_\circ)V_t^{ex}(\boldsymbol{\beta}_\circ,\hat{\theta}),$$

where $C_{k_0}^{ex}(\theta_0)$ is the coefficient matrix corresponding to the true parameter vector θ_0. As in the unrestricted concentration approach we find

$$\begin{aligned}&C_{k_0}^{ex}(\theta_\circ)V_t^{ex}(\boldsymbol{\beta}_\circ,\theta_\circ) - \hat{C}_{k_0}^{ex,R}(\boldsymbol{\beta}_\circ)V_t^{ex}(\boldsymbol{\beta}_\circ,\hat{\theta}) \\ &\quad= \sum_{j=0}^{t-2}(K_j^-(\theta_\circ) - K_j^-(\hat{\theta}))(1-L^{\tilde{S}})y_{t-1-j} + \left(C_{k_0}^{ex}(\hat{\theta}) - \hat{C}_{k_0}^{ex,R}(\boldsymbol{\beta}_\circ)\right)V_t^{ex}(\boldsymbol{\beta}_\circ,\hat{\theta}),\end{aligned}$$

where $C^{ex}_{k_0}(\hat{\theta})$ is the coefficient matrix corresponding to $\hat{\theta}$ and, thus, a consistent estimator for $C^{ex}_{k_0}(\theta_\circ)$. Since $\{\boldsymbol{\beta}_\circ' Z_{1,t}\}_{t\in\mathbb{N}}$ is stationary and $\hat{\theta}$ is consistent, $C^{ex}_{k_0}(\hat{\theta}) - \hat{C}^{ex,R}_{k_0}(\boldsymbol{\beta}_\circ) = o_p(1)$ follows, such that the limits of $\mathcal{C}_\omega'\langle \hat{R}^R_{1,t}(\boldsymbol{\beta}_\circ), \hat{\varepsilon}^R_t(\boldsymbol{\beta}_\circ)\rangle$ and $\mathcal{C}_\omega'\langle \hat{R}^R_{1,t}(\boldsymbol{\beta}_\circ), \varepsilon_t\rangle$ coincide and Lemma 5 is applicable.

(III) For the first result, note that $\tilde{\boldsymbol{\beta}}'\hat{S}_{10} = (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_\circ)'\mathcal{C}_\omega\mathcal{C}_\omega'\hat{S}_{10} + \boldsymbol{\beta}_\circ'\hat{S}_{10}$ and $\tilde{\boldsymbol{\beta}}'\hat{S}_{11} = (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_\circ)'\mathcal{C}_\omega\mathcal{C}_\omega'\hat{S}_{11} + \boldsymbol{\beta}_\circ'\hat{S}_{11}$. Applying the results from (I) for the different parts, it follows that both matrices are bounded in probability. The convergence of $\tilde{\alpha}$ and $\langle\tilde{\varepsilon}_t, \tilde{\varepsilon}_t\rangle$ to $\breve{\alpha}_\circ$ and Σ follows from $T(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_\circ) = O_p(1)$, the consistency of $k(z,\hat{\theta})$ and the consistency of the OLS estimator for the true $R_{0,t}(\theta_\circ)$ and $\boldsymbol{\beta}_\circ' R_{1,t}(\theta_\circ)$. Similarly, the limits of $(\tilde{\boldsymbol{\beta}}'\hat{S}_{11,0}\tilde{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{\beta}}'\hat{S}_{10}\hat{S}_{00}^{-1}$ and $(\boldsymbol{\beta}_\circ' S_{11,0}\boldsymbol{\beta}_\circ)^{-1}\boldsymbol{\beta}_\circ' S_{10}S_{00}^{-1}$ coincide, where the latter converges to $\breve{\alpha}_\circ'\Sigma^{-1}$ due to Johansen and Schaumburg (1999, Lemma 8).

(IV) For the restricted concentration approach note that $T(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_\circ) = O_p(1)$ implies

$$\left\|\langle Z_{0,t}, V^{ex}_t(\tilde{\boldsymbol{\beta}},\hat{\theta})\rangle - \langle Z_{0,t}, V^{ex}_t(\boldsymbol{\beta}_\circ,\hat{\theta})\rangle\right\| = O_p(T^{-1})$$

and

$$\left\|\begin{bmatrix}\langle V^{ex}_t(\tilde{\boldsymbol{\beta}},\hat{\theta}), V^{ex}_t(\tilde{\boldsymbol{\beta}},\hat{\theta})\rangle & G(\tilde{\boldsymbol{\beta}},\hat{\theta}) \\ G(\tilde{\boldsymbol{\beta}},\hat{\theta})' & 0\end{bmatrix} - \begin{bmatrix}\langle V^{ex}_t(\boldsymbol{\beta}_\circ,\hat{\theta}), V^{ex}_t(\boldsymbol{\beta}_\circ,\hat{\theta})\rangle & G(\boldsymbol{\beta}_\circ,\hat{\theta}) \\ G(\boldsymbol{\beta}_\circ,\hat{\theta})' & 0\end{bmatrix}\right\| = O_p(T^{-1}),$$

using $G(\tilde{\boldsymbol{\beta}},\hat{\theta}) = G(\boldsymbol{\beta}_\circ,\hat{\theta}) + O_p(T^{-1})$, $\tilde{\boldsymbol{\beta}}'\langle Z_{1,t}, V_t(\hat{\theta})\rangle = \boldsymbol{\beta}_\circ'\langle Z_{1,t}, V_t(\hat{\theta})\rangle + O_p(T^{-1})$, due to the stochastic boundedness of $\langle Z_{1,t}, V_t(\hat{\theta})\rangle$, and

$$\begin{aligned}\tilde{\boldsymbol{\beta}}'\langle Z_{1,t}, Z_{1,t}\rangle\tilde{\boldsymbol{\beta}} &= \boldsymbol{\beta}_\circ'\langle Z_{1,t}, Z_{1,t}\rangle\boldsymbol{\beta}_\circ + (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_\circ)'\tilde{\mathcal{C}}_\omega\tilde{\mathcal{C}}_\omega'\langle Z_{1,t}, Z_{1,t}\rangle\boldsymbol{\beta}_\circ \\ &\quad + \boldsymbol{\beta}_\circ'\langle Z_{1,t}, Z_{1,t}\rangle\tilde{\mathcal{C}}_\omega\tilde{\mathcal{C}}_\omega'(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_\circ) + (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_\circ)'\tilde{\mathcal{C}}_\omega\tilde{\mathcal{C}}_\omega'\langle Z_{1,t}, Z_{1,t}\rangle\tilde{\mathcal{C}}_\omega\tilde{\mathcal{C}}_\omega'(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_\circ) \\ &= \boldsymbol{\beta}_\circ'\langle Z_{1,t}, Z_{1,t}\rangle\boldsymbol{\beta}_\circ + O_p(T^{-1}).\end{aligned}$$

It follows also that $H_{11}(\tilde{\boldsymbol{\beta}},\hat{\theta}) = H_{11}(\boldsymbol{\beta}_\circ,\hat{\theta}) + O_p(T^{-1})$ and $H_{21}(\tilde{\boldsymbol{\beta}},\hat{\theta}) = H_{21}(\boldsymbol{\beta}_\circ,\hat{\theta}) + O_p(T^{-1})$. The definition of $\hat{C}^{ex,R}_{k_0}(\boldsymbol{\beta}_\circ)$ then implies that $T(\hat{C}^{ex,R}_{k_0}(\tilde{\boldsymbol{\beta}}) - \hat{C}^{ex,R}_{k_0}(\boldsymbol{\beta}_\circ))$ is bounded in probability. For $\hat{\varepsilon}^R_t(\tilde{\boldsymbol{\beta}})$ it holds that

$$\begin{aligned}\hat{\varepsilon}^R_t(\tilde{\boldsymbol{\beta}}) &= \hat{\varepsilon}^R_t(\tilde{\boldsymbol{\beta}}) - \hat{\varepsilon}^R_t(\boldsymbol{\beta}_\circ) + \hat{\varepsilon}^R_t(\boldsymbol{\beta}_\circ) \\ &= -\hat{C}^{ex,R}_{k_0}(\tilde{\boldsymbol{\beta}})V^{ex}_t(\tilde{\boldsymbol{\beta}},\hat{\theta}) + \hat{C}^{ex,R}_{k_0}(\boldsymbol{\beta}_\circ)V^{ex}_t(\boldsymbol{\beta}_\circ,\hat{\theta}) + \hat{\varepsilon}^R_t(\boldsymbol{\beta}_\circ) \\ &= \left(-\hat{C}^{ex,R}_{k_0}(\tilde{\boldsymbol{\beta}}) + \hat{C}^{ex,R}_{k_0}(\boldsymbol{\beta}_\circ)\right)V^{ex}_t(\tilde{\boldsymbol{\beta}},\hat{\theta}) + \hat{C}^{ex,R}_{k_0}(\boldsymbol{\beta}_\circ)\left(V^{ex}_t(\boldsymbol{\beta}_\circ,\hat{\theta}) - V^{ex}_t(\tilde{\boldsymbol{\beta}},\hat{\theta})\right) + \hat{\varepsilon}^R_t(\boldsymbol{\beta}_\circ).\end{aligned}$$

The only non-zero component in $V^{ex}_t(\boldsymbol{\beta}_\circ,\hat{\theta}) - V^{ex}_t(\tilde{\boldsymbol{\beta}},\hat{\theta})$ is $(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_\circ)'\mathcal{C}_\omega Z_{1,t}$, such that the integrated component $\mathcal{C}_\omega Z_{1,t}$ is multiplied by $(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_\circ) = O_p(T^{-1})$. Thus, the dominant component in $\langle\hat{\varepsilon}^R_t(\tilde{\boldsymbol{\beta}}), \hat{\varepsilon}^R_t(\tilde{\boldsymbol{\beta}})\rangle$ is $\langle\hat{\varepsilon}^R_t(\boldsymbol{\beta}_\circ), \hat{\varepsilon}^R_t(\boldsymbol{\beta}_\circ)\rangle$, which converges to Σ due to the arguments in (II). The other convergence results follow similarly. ■

B.4 Proof of Theorem 7

We use the notation $[C]^{\mathbb{R}}$, where C is a complex valued matrix, to denote the corresponding real valued matrix of the form $\begin{bmatrix}\mathcal{R}(C) & -\mathcal{I}(C) \\ \mathcal{I}(C) & \mathcal{R}(C)\end{bmatrix}$. Such matrices are said to *have complex structure*. For the proof below we define an operator that adjusts a matrix to have complex structure.

Definition 17 *We define the following mapping for a general matrix $M \in \mathbb{R}^{2k\times 2l}$:*

$$[M]^{\mathbb{C}} := \frac{1}{2}\left(M - [i\cdot I_k]^{\mathbb{R}}\, M\, [i\cdot I_l]^{\mathbb{R}}\right).$$

Matrices $N \in \mathbb{R}^{2k\times 2l}$ satisfying $N = [N]^{\mathbb{C}}$ have complex structure.

Note that $\mathrm{tr}\,(M) = \mathrm{tr}\,([M]^{\mathbb{C}})$ and for all $N \in \mathbb{R}^{2l\times 2m}$ of complex structure it holds that $[MN]^{\mathbb{C}} = [M]^{\mathbb{C}}N$.

(I) **Asymptotic distribution of $\tilde{\boldsymbol{\beta}}$**
The proof follows loosely the ideas of Johansen and Schaumburg (1999). Consider first the **unrestricted concentration approach**, then the concentrated pseudo log-likelihood function up to a constant is given by

$$\begin{aligned} 2L_T^{ex,R}(\boldsymbol{\beta},\hat{\theta}) &= -T\log\frac{|\boldsymbol{\beta}'\hat{S}_{11,0}\boldsymbol{\beta}|}{|\boldsymbol{\beta}'\hat{S}_{11}\boldsymbol{\beta}|} - T\log|\hat{S}_{00}| \\ &= -T\log|I - (\boldsymbol{\beta}'\hat{S}_{11}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'\hat{S}_{10}\hat{S}_{00}^{-1}\hat{S}_{01}\boldsymbol{\beta}| - T\log|\hat{S}_{00}|, \end{aligned}$$

where $\boldsymbol{\beta}'\hat{S}_{00}\boldsymbol{\beta}$ and $(\boldsymbol{\beta}'\hat{S}_{00}\boldsymbol{\beta})^{-1}$ are bounded in probability for every $\boldsymbol{\beta}$, since the dominant components in $\hat{R}_{0,t}$ are stationary due to Lemma 6. For each $\boldsymbol{\beta}$, furthermore, standard arguments show that $\boldsymbol{\beta}'\hat{S}_{10}\hat{S}_{00}^{-1}\hat{S}_{01}\boldsymbol{\beta}$ converges in distribution for every $\boldsymbol{\beta}$. Now let $\check{\boldsymbol{\beta}}$ denote a maximizer of $L_T(\boldsymbol{\beta},\hat{\theta})$ for given $\hat{\theta}$ and let $I_{2s} = P_C + P_\perp$ denote the decomposition of the identity into the projection onto the column space of $\mathcal{C}_\omega$ and the orthocomplement (the space spanned by the columns of $\boldsymbol{\beta}_\circ$). Then

$$\check{\boldsymbol{\beta}}'\hat{S}_{11}\check{\boldsymbol{\beta}} = \check{\boldsymbol{\beta}}'P_C\hat{S}_{11}P_C\check{\boldsymbol{\beta}} + \check{\boldsymbol{\beta}}'P_C\hat{S}_{11}P_\perp\check{\boldsymbol{\beta}} + \check{\boldsymbol{\beta}}'P_\perp\hat{S}_{11}P_C\check{\boldsymbol{\beta}} + \check{\boldsymbol{\beta}}'P_\perp\hat{S}_{11}P_\perp\check{\boldsymbol{\beta}},$$

where $P_\perp\hat{S}_{11}P_\perp$ converges in probability, $P_\perp\hat{S}_{11}P_C$ converges in distribution and $P_C\hat{S}_{11}P_C/T$ converges in distribution to an almost surely non-singular matrix.
It follows that $\check{\boldsymbol{\beta}}'P_C \to 0$ is necessary for the first term to remain bounded. This implies that the second and third term tend to zero. The first term dominates the second and third term, if $\|\check{\boldsymbol{\beta}}'P_C\|^2T$ dominates $\|\check{\boldsymbol{\beta}}'P_C\|$. In such a situation the criterion function is smaller for $P_\perp\check{\boldsymbol{\beta}}$ than for $\check{\boldsymbol{\beta}}$ which, hence, cannot be a maximizer. This implies that $TP_C\check{\boldsymbol{\beta}}$ remains bounded. This shows that the normalized estimator $\tilde{\boldsymbol{\beta}}$ fulfills $T(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_\circ) = O_P(1)$ as assumed in Lemma 7(III).
Next differentiating L_T as a function of β we find the first order conditions to be fulfilled by the estimator $\tilde{\boldsymbol{\beta}}$, compare (33) of Johansen and Schaumburg (1999):

$$\begin{aligned} 0 &= \left[(\tilde{\boldsymbol{\beta}}'\hat{S}_{11}\tilde{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{\beta}}'\hat{S}_{11} - (\tilde{\boldsymbol{\beta}}'\hat{S}_{11,0}\tilde{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{\beta}}'\hat{S}_{11,0}\right]^{\mathbb{C}} \\ &= \left[(\tilde{\boldsymbol{\beta}}'\hat{S}_{11,0}\tilde{\boldsymbol{\beta}})^{-1}\left((\tilde{\boldsymbol{\beta}}'\hat{S}_{11,0}\tilde{\boldsymbol{\beta}})(\tilde{\boldsymbol{\beta}}'\hat{S}_{11}\tilde{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{\beta}}'\hat{S}_{11} - \tilde{\boldsymbol{\beta}}'\hat{S}_{11,0}\right)\right]^{\mathbb{C}} \\ &= \left[(\tilde{\boldsymbol{\beta}}'\hat{S}_{11,0}\tilde{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{\beta}}'\hat{S}_{10}\hat{S}_{00}^{-1}\left(\hat{S}_{01} - \hat{S}_{01}\tilde{\boldsymbol{\beta}}(\tilde{\boldsymbol{\beta}}'\hat{S}_{11}\tilde{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{\beta}}'\hat{S}_{11}\right)\right]^{\mathbb{C}}. \end{aligned} \tag{B.4}$$

The above is equivalent to

$$\left[(\tilde{\boldsymbol{\beta}}'\hat{S}_{11,0}\tilde{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{\beta}}'\hat{S}_{10}\hat{S}_{00}^{-1}\left\langle\tilde{\varepsilon}_t,\hat{R}_{1,t}\right\rangle\right]^{\mathbb{C}} = 0 \tag{B.5}$$

and implies $\left[(\tilde{\boldsymbol{\beta}}'\hat{S}_{11,0}\tilde{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{\beta}}'\hat{S}_{10}\hat{S}_{00}^{-1}\left\langle\tilde{\varepsilon}_t,\hat{R}_{1,t}\right\rangle\mathcal{C}_\omega\right]^{\mathbb{C}} = 0$. We first find the weak limit for the above matrix before it is complexified. From Lemma 7 we have $(\tilde{\boldsymbol{\beta}}'\hat{S}_{11,0}\tilde{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{\beta}}'\hat{S}_{10}\hat{S}_{00}^{-1} \xrightarrow{p} \breve{\alpha}_\circ'\Sigma^{-1}$. Since

$$\left\langle\tilde{\varepsilon}_t,\hat{R}_{1,t}\right\rangle = \hat{S}_{\varepsilon 1} + (\breve{\alpha}_\circ - \tilde{\alpha})\boldsymbol{\beta}'\hat{S}_{11} - \tilde{\alpha}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_\circ)'\hat{S}_{11}$$

we find

$$\left\langle \tilde{\varepsilon}_t, \hat{R}_{1,t} \right\rangle \mathcal{C}_\omega = \hat{S}_{\varepsilon 1}\mathcal{C}_\omega - \tilde{\alpha}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_\circ)'\mathcal{C}_\omega\mathcal{C}_\omega'\hat{S}_{11}\mathcal{C}_\omega + o_P(1).$$

Thus, the whole term considered in (B.4) can be written as

$$\left[\breve{\alpha}_\circ'\Sigma^{-1}\left(\hat{S}_{\varepsilon 1}\mathcal{C}_\omega - \tilde{\alpha}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_\circ)'\mathcal{C}_\omega\mathcal{C}_\omega'\hat{S}_{11}\mathcal{C}_\omega\right) + o_P(1)\right]^{\mathbb{C}}$$

and for its weak limit it holds that

$$\left[\breve{\alpha}_\circ'\Sigma^{-1}\left((I_s, 0)\int_0^1 (\mathrm{d}\boldsymbol{W})\boldsymbol{F}' - \breve{\alpha}_\circ\boldsymbol{B}_\infty'\int_0^1 \boldsymbol{F}\boldsymbol{F}'\mathrm{d}u\right)\right]^{\mathbb{C}} = 0,$$

where $\boldsymbol{F} = \boldsymbol{B}_\omega\boldsymbol{W}$ and $\boldsymbol{B}_\infty$ is the weak limit of $T\mathcal{C}_\omega'(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_\circ)$. Since $\boldsymbol{B}_\infty'\int_0^1 \boldsymbol{F}\boldsymbol{F}'\mathrm{d}u$ and $\int_0^1(\mathrm{d}\boldsymbol{W})\boldsymbol{F}'$ have complex structure, the first order conditions then imply

$$\left[\breve{\alpha}_\circ'\Sigma^{-1}(I_s, 0)\right]^{\mathbb{C}}\int_0^1 (\mathrm{d}\boldsymbol{W})\boldsymbol{F}' = [\breve{\alpha}_\circ'\Sigma^{-1}\breve{\alpha}_\circ]^{\mathbb{C}}\boldsymbol{B}_\infty'\int_0^1 \boldsymbol{F}\boldsymbol{F}'\mathrm{d}u$$

where $[\breve{\alpha}_\circ'\Sigma^{-1}(I_s, 0)]^{\mathbb{C}} = \frac{1}{2}\boldsymbol{\alpha}_\circ'\boldsymbol{\Sigma}^{-1}$ and $[\breve{\alpha}_\circ'\Sigma^{-1}\breve{\alpha}_\circ]^{\mathbb{C}} = \frac{1}{2}\boldsymbol{\alpha}_\circ'\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha}_\circ$ with

$$\boldsymbol{\Sigma} := \begin{bmatrix} \Sigma & 0 \\ 0 & \Sigma \end{bmatrix}.$$

This shows

$$\boldsymbol{B}_\infty = \left(\int_0^1 \boldsymbol{F}\boldsymbol{F}'\mathrm{d}u\right)^{-1}\int_0^1 \boldsymbol{F}(\mathrm{d}\boldsymbol{V}'),$$

where $\boldsymbol{V} = (\boldsymbol{\alpha}_\circ'\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha}_\circ)^{-1}\boldsymbol{\alpha}_\circ'\boldsymbol{\Sigma}^{-1}\boldsymbol{W}$.
From Lemma 4 it follows that for the **restricted concentration approach** the concentrated pseudo log-likelihood function is given by

$$\begin{aligned} 2L_T^{ex,R}(\boldsymbol{\beta}, \theta) &= -T\log\left|\left\langle \hat{\varepsilon}_t^R(\boldsymbol{\beta}), \hat{\varepsilon}_t^R(\boldsymbol{\beta}) \right\rangle\right| \\ &= -T\log\left|\langle \hat{\varepsilon}_t, \hat{\varepsilon}_t \rangle + \hat{\boldsymbol{\Pi}}\boldsymbol{\gamma}(\boldsymbol{\gamma}'\hat{N}_T\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\hat{\boldsymbol{\Pi}}'\right| \\ &= -T\log|\hat{\Sigma}| - T\log\left|I_s + \hat{\Sigma}^{-1}\hat{\boldsymbol{\Pi}}\boldsymbol{\gamma}(\boldsymbol{\gamma}'\hat{N}_T\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\hat{\boldsymbol{\Pi}}'\right| \end{aligned}$$

using orthogonality of $\hat{\varepsilon}_t := \varepsilon_t^{SS}(\hat{\theta})$ to $\hat{x}_t := x_t(\hat{\theta})$, and the notation $\hat{\boldsymbol{\Pi}} := \boldsymbol{\Pi}_{k_0}^{SS}(\hat{\theta})$, $\hat{\Sigma} := \langle \hat{\varepsilon}_t, \hat{\varepsilon}_t \rangle$ and $\hat{N}_T := \underline{B}_{k_0}^{\mathbb{R}}(\hat{\theta})'\langle \hat{x}_t, \hat{x}_t \rangle^{-1}\underline{B}_{k_0}^{\mathbb{R}}(\hat{\theta})$. It holds that $\boldsymbol{\gamma}'\hat{S}_{11}^R(\boldsymbol{\beta})\boldsymbol{\gamma}$ is equal to $(\boldsymbol{\gamma}'\hat{N}_T\boldsymbol{\gamma})^{-1}$ for every full rank matrix $\boldsymbol{\gamma} = [\gamma]^{\mathbb{R}}$, $\gamma \in \mathbb{C}^{s\times(s-r)}$, satisfying $\boldsymbol{\gamma}'\boldsymbol{\beta} = 0$, with $\boldsymbol{\beta} \in G(r, \hat{\theta})$.
The pseudo log-likelihood function is maximized if the term

$$\Lambda_T(\boldsymbol{\gamma}) := \log\left|I_s + \hat{\Sigma}^{-1}\hat{\boldsymbol{\Pi}}\boldsymbol{\gamma}(\boldsymbol{\gamma}'\hat{N}_T\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\hat{\boldsymbol{\Pi}}'\right|$$

is minimized by $\check{\boldsymbol{\gamma}}$ corresponding to $\check{\boldsymbol{\beta}}^R$. Note that by Lemma 6(II) and Lemma 7(II) $\|(\mathcal{C}_\omega'\hat{S}_{11}^R\mathcal{C}_\omega)^{-1}\|$ is in $O_p(T^{-1})$, while $\|\boldsymbol{\beta}_{(j)}'\hat{S}_{11}^R\boldsymbol{\beta}_{(j)}\|$ is bounded for $j = 1, \ldots, r$, which also translate to analogous results for $\hat{N}_T$, such that $\|\mathcal{C}_\omega'\hat{N}_T\mathcal{C}_\omega\| = O_p(T^{-1})$ and $\|\boldsymbol{\beta}_\circ'\hat{N}_T\boldsymbol{\beta}_\circ\|$ is bounded in probability. Using similar arguments, $\|\hat{\boldsymbol{\Pi}}\mathcal{C}_\omega\| = O_p(T^{-1})$ and $\|\hat{\boldsymbol{\Pi}}\boldsymbol{\beta}_\circ\| = O_p(1)$. Lemma 7(I) implies $\|\Lambda_T(\mathcal{C}_\omega)\|$ is in $O_p(T^{-1})$ at the true value $\mathcal{C}_\omega$ and, therefore, also at the optimum. From this again follows that $TP_C\check{\boldsymbol{\beta}}^R$ must remain bounded, using similar arguments as discussed above for the unrestricted concentration approach. Thus, the estimator $\tilde{\boldsymbol{\beta}}^R$ also satisfied $T(\tilde{\boldsymbol{\beta}}^R - \boldsymbol{\beta}_\circ) = O_P(1)$ as assumed

in Lemma 7(IV).
Further transformation of $L_T^{ex,R}(\boldsymbol{\beta},\theta)$ leads to

$$-T\log\left|\left\langle \hat{\varepsilon}_t^R(\boldsymbol{\beta}),\hat{\varepsilon}_t^R(\boldsymbol{\beta})\right\rangle\right| = -T\log\left(|\hat{\Sigma}|\frac{|\boldsymbol{\gamma}'(\hat{N}_T)\boldsymbol{\gamma}|}{|\boldsymbol{\gamma}'(\hat{N}_T+\hat{M}_T)\boldsymbol{\gamma}|}\right)$$

with $\hat{M}_T := \hat{\boldsymbol{\Pi}}'\hat{\Sigma}^{-1}\hat{\boldsymbol{\Pi}}$. Optimizing the pseudo log-likelihood over γ leads to the FOC at the optimum $\tilde{\mathcal{C}}_\omega$ (normalized as $\boldsymbol{\beta}$ using the true $\mathcal{C}_\omega$, i. e., for a maximizer $\hat{\mathcal{C}}_\omega$ define $\bar{\mathcal{C}}_\omega := \hat{\mathcal{C}}_\omega(\bar{\mathcal{C}}_\omega'\hat{\mathcal{C}}_\omega)^{-1}$ with $\bar{\mathcal{C}}_\omega := \mathcal{C}_\omega(\mathcal{C}_\omega'\mathcal{C}_\omega)^{-1}$)

$$\begin{aligned}0 &= \left[(\tilde{\mathcal{C}}_\omega'(\hat{N}_T+\hat{M}_T)\tilde{\mathcal{C}}_\omega)^{-1}\tilde{\mathcal{C}}_\omega'(\hat{N}_T+\hat{M}_T) - (\tilde{\mathcal{C}}_\omega'\hat{N}_T\tilde{\mathcal{C}}_\omega)^{-1}\tilde{\mathcal{C}}_\omega\hat{N}_T\right]^{\mathbb{C}}\\ &= \left[(\tilde{\mathcal{C}}_\omega'(\hat{N}_T+\hat{M}_T)\tilde{\mathcal{C}}_\omega)^{-1}\left(\tilde{\mathcal{C}}_\omega'\hat{M}_T - \tilde{\mathcal{C}}_\omega'\hat{M}_T\tilde{\mathcal{C}}_\omega(\tilde{\mathcal{C}}_\omega'\hat{N}_T\tilde{\mathcal{C}}_\omega)^{-1}\tilde{\mathcal{C}}_\omega\hat{N}_T\right)\right]^{\mathbb{C}}.\end{aligned}$$

Further, rewrite the second factor using

$$\begin{aligned}&\tilde{\mathcal{C}}_\omega'\hat{M}_T - \tilde{\mathcal{C}}_\omega'\hat{M}_T\tilde{\mathcal{C}}_\omega(\tilde{\mathcal{C}}_\omega'\hat{N}_T\tilde{\mathcal{C}}_\omega)^{-1}\tilde{\mathcal{C}}_\omega\hat{N}_T\\ &\quad= \tilde{\mathcal{C}}_\omega'\hat{\boldsymbol{\Pi}}'\hat{\Sigma}^{-1}\hat{\boldsymbol{\Pi}} - \tilde{\mathcal{C}}_\omega'\hat{\boldsymbol{\Pi}}'\hat{\Sigma}^{-1}\hat{\boldsymbol{\Pi}}\tilde{\mathcal{C}}_\omega(\tilde{\mathcal{C}}_\omega'\underline{B}_{k_0}^{\mathbb{R}}(\hat{\theta})'\left\langle \hat{x}_t,\hat{x}_t\right\rangle^{-1}\underline{B}_{k_0}^{\mathbb{R}}(\hat{\theta})\tilde{\mathcal{C}}_\omega)^{-1}\tilde{\mathcal{C}}_\omega'\underline{B}_{k_0}^{\mathbb{R}}(\hat{\theta})'\left\langle \hat{x}_t,\hat{x}_t\right\rangle^{-1}\underline{B}_{k_0}^{\mathbb{R}}(\hat{\theta})\\ &\quad= \tilde{\mathcal{C}}_\omega'\hat{\boldsymbol{\Pi}}'\hat{\Sigma}^{-1}\hat{\boldsymbol{\Pi}} - \tilde{\mathcal{C}}_\omega'\hat{\boldsymbol{\Pi}}'\hat{\Sigma}^{-1}(\hat{C}-\tilde{C}(\tilde{\mathcal{C}}_\omega))\underline{B}_{k_0}^{\mathbb{R}}(\hat{\theta})\\ &\quad= \tilde{\mathcal{C}}_\omega'\hat{\boldsymbol{\Pi}}'\hat{\Sigma}^{-1}\tilde{\boldsymbol{\Pi}},\end{aligned}$$

with $\hat{C} := C^{SS}(\hat{\theta})$ and $\tilde{\boldsymbol{\Pi}} := \tilde{C}(\tilde{\mathcal{C}}_\omega)\underline{B}_{k_0}^{\mathbb{R}}(\hat{\theta}) - [I_s, 0]$. Multiplying $\tilde{\boldsymbol{\beta}}$ from the right, and using $\tilde{\alpha} := \tilde{\boldsymbol{\Pi}}\tilde{\boldsymbol{\beta}}$ and $\hat{\alpha} := \hat{\boldsymbol{\Pi}}\boldsymbol{\beta}_\circ$, the above FOC imply

$$\left[\left(\tilde{\mathcal{C}}_\omega'(\hat{N}_T+\hat{M}_T)\tilde{\mathcal{C}}_\omega\right)^{-1}(\tilde{\mathcal{C}}_\omega-\mathcal{C}_\omega)'\boldsymbol{\beta}_\circ\hat{\alpha}'\hat{\Sigma}^{-1}\tilde{\alpha}\right]^{\mathbb{C}} = \left[-\left(\tilde{\mathcal{C}}_\omega'(\hat{N}_T+\hat{M}_T)\tilde{\mathcal{C}}_\omega\right)^{-1}\mathcal{C}_\omega'\hat{\boldsymbol{\Pi}}'\hat{\Sigma}^{-1}\tilde{\alpha}\right]^{\mathbb{C}}.$$

Note that $\mathcal{C}_\omega'\hat{\boldsymbol{\Pi}}'$ is equal to $(\mathcal{C}_\omega'\hat{S}_{11}^R(\boldsymbol{\beta}_\circ)\mathcal{C}_\omega)^{-1}\mathcal{C}_\omega'\hat{S}_{1\varepsilon}^R(\boldsymbol{\beta}_\circ)$, i. e., the OLS estimator in the reduced model of the restricted approach and recall the identity $(\tilde{\mathcal{C}}_\omega'\hat{N}_T\tilde{\mathcal{C}}_\omega)^{-1} = \tilde{\mathcal{C}}_\omega'\hat{S}_{11}^R(\tilde{\boldsymbol{\beta}})\tilde{\mathcal{C}}_\omega$. Therefore, $\tilde{\mathcal{C}}_\omega'\hat{M}_T\tilde{\mathcal{C}}_\omega$, which is $O_p(T^{-2})$ due to $\mathcal{C}_\omega'\hat{\boldsymbol{\Pi}}'$ being $O_p(T^{-1})$, is dominated by $\tilde{\mathcal{C}}_\omega'\hat{N}_T\tilde{\mathcal{C}}_\omega$, which is $O_p(T^{-1})$. Using also $T^{-1}\tilde{\mathcal{C}}_\omega'\hat{S}_{11}^R(\tilde{\boldsymbol{\beta}})\tilde{\mathcal{C}}_\omega = T^{-1}\mathcal{C}_\omega'\hat{S}_{11}^R(\tilde{\boldsymbol{\beta}})\mathcal{C}_\omega + o_p(1)$, thus,

$$\begin{aligned}&\left[\left(T^{-1}\tilde{\mathcal{C}}_\omega'\hat{S}_{11}^R(\tilde{\boldsymbol{\beta}})\tilde{\mathcal{C}}_\omega\right)T(\tilde{\mathcal{C}}_\omega-\mathcal{C}_\omega)'\boldsymbol{\beta}_\circ\hat{\alpha}'\hat{\Sigma}^{-1}\tilde{\alpha}\right]^{\mathbb{C}}\\ &\qquad= \left[-(T^{-1}\tilde{\mathcal{C}}_\omega'\hat{S}_{11}^R(\tilde{\boldsymbol{\beta}})\tilde{\mathcal{C}}_\omega)(T^{-1}\mathcal{C}_\omega'\hat{S}_{11}^R(\tilde{\boldsymbol{\beta}})\mathcal{C}_\omega)^{-1}\left\langle \mathcal{C}_\omega'\hat{R}_{1,t}(\boldsymbol{\beta}_\circ),\hat{\varepsilon}_t^R(\boldsymbol{\beta}_\circ)\right\rangle\hat{\Sigma}^{-1}\tilde{\alpha}\right]^{\mathbb{C}} + o_p(1),\end{aligned}$$

converges to

$$\left(\int_0^1 \boldsymbol{F}\boldsymbol{F}'\mathrm{d}u\right)T(\tilde{\mathcal{C}}_\omega-\mathcal{C}_\omega)'\boldsymbol{\beta}_\circ\left[\breve{\alpha}_\circ'\Sigma^{-1}\breve{\alpha}_\circ\right]^{\mathbb{C}} = -\int_0^1 \boldsymbol{F}(\mathrm{d}\boldsymbol{W}')\left[\Sigma^{-1}\breve{\alpha}_\circ\right]^{\mathbb{C}},$$

such that the limit of $T(\tilde{\mathcal{C}}_\omega-\mathcal{C}_\omega)'\boldsymbol{\beta}_\circ$ is $-\left(\int_0^1 \boldsymbol{F}\boldsymbol{F}'\mathrm{d}u\right)^{-1}\int_0^1 \boldsymbol{F}(\mathrm{d}\boldsymbol{V}')$. Finally it holds that $T\mathcal{C}_\omega'(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_\circ) = -T(\tilde{\mathcal{C}}_\omega-\mathcal{C}_\omega)'\boldsymbol{\beta}_\circ + O_p(T^{-1})$, which finishes the proof also for the restricted approach.

(II) **Distribution of pseudo likelihood ratio test for $\beta = b$:**
To shorten notation, define $[\hat{C}_{-k_0}^R(\boldsymbol{\beta}),\breve{\alpha}^R(\boldsymbol{\beta})] := \langle Z_{0,t},V_t^{ex}(\boldsymbol{\beta},\hat{\theta})\rangle H_{11}(\boldsymbol{\beta},\hat{\theta}) + JH_{21}(\boldsymbol{\beta},\hat{\theta})$ such that $\hat{\varepsilon}_t^R(\boldsymbol{\beta}) = Z_{0,t} - \breve{\alpha}^R(\boldsymbol{\beta})\boldsymbol{\beta}'Z_{1,t} - \hat{C}_{-k_0}^R(\boldsymbol{\beta})V_t(\hat{\theta})$. The test statistic is then equal to

$$\begin{aligned}2L_T^{ex,R}(\tilde{\boldsymbol{\beta}}^R,\hat{\theta}) - 2L_T^{ex,R}(\boldsymbol{b},\hat{\theta}) &= -T\log\left|\left\langle \hat{\varepsilon}_t^R(\boldsymbol{b}),\hat{\varepsilon}_t^R(\boldsymbol{b})\right\rangle^{-1}\left\langle \hat{\varepsilon}_t^R(\tilde{\boldsymbol{\beta}}^R),\hat{\varepsilon}_t^R(\tilde{\boldsymbol{\beta}}^R)\right\rangle\right|\\ &= -T\log\left|I_s + \left\langle \hat{\varepsilon}_t^R(\boldsymbol{b}),\hat{\varepsilon}_t^R(\boldsymbol{b})\right\rangle^{-1}\left(\left\langle \delta_t,\hat{\varepsilon}_t^R(\boldsymbol{b})\right\rangle + \left\langle \hat{\varepsilon}_t^R(\boldsymbol{b}),\delta_t\right\rangle + \left\langle \delta_t,\delta_t\right\rangle\right)\right|,\end{aligned}$$

where $\delta_t := \hat{\varepsilon}_t^R(\tilde{\boldsymbol{\beta}}^R) - \hat{\varepsilon}_t^R(\boldsymbol{b})$, such that

$$\begin{aligned}\delta_t &= (\breve{\alpha}^R(\boldsymbol{b})\boldsymbol{b}' - \breve{\alpha}^R(\tilde{\boldsymbol{\beta}}^R)(\tilde{\boldsymbol{\beta}}^R)')Z_{1,t} + (\hat{C}^R_{-k_0}(\boldsymbol{b}) - \hat{C}^R_{-k_0}(\tilde{\boldsymbol{\beta}}^R))V_t(\hat{\theta}) \\ &= (\breve{\alpha}^R(\boldsymbol{b}) - \breve{\alpha}^R(\tilde{\boldsymbol{\beta}}^R))\boldsymbol{b}'Z_{1,t} - \breve{\alpha}^R(\tilde{\boldsymbol{\beta}}^R)(\tilde{\boldsymbol{\beta}}^R - \boldsymbol{b})'\mathcal{C}_\omega\mathcal{C}_\omega' Z_{1,t} + (\hat{C}^R_{-k_0}(\boldsymbol{b}) - \hat{C}^R_{-k_0}(\tilde{\boldsymbol{\beta}}^R))V_t(\hat{\theta}).\end{aligned}$$

Note also that $(\tilde{\boldsymbol{\beta}}^R - \boldsymbol{b})$ in $O_p(T^{-1})$ under the null hypothesis implies $(\breve{\alpha}^R(\boldsymbol{b}) - \breve{\alpha}^R(\tilde{\boldsymbol{\beta}}^R))$ and $(\hat{C}^R_{-k_0}(\boldsymbol{b}) - \hat{C}^R_{-k_0}(\tilde{\boldsymbol{\beta}}^R))$ in $O_p(T^{-1})$, due to Lemma 7(IV). Note that $(\hat{C}^R_{-k_0}(\boldsymbol{b}) - \hat{C}^R_{-k_0}(\tilde{\boldsymbol{\beta}}^R))$ in $O_p(T^{-1})$ can be bounded even further using

$$\begin{aligned}\hat{\varepsilon}_t^R(\tilde{\boldsymbol{\beta}}^R) - \hat{\varepsilon}_t^R(\boldsymbol{b}) &= \left(\left\langle Z_{0,t}, \hat{V}_t^{ex}(\tilde{\boldsymbol{\beta}}^R)\right\rangle H_{11}(\tilde{\boldsymbol{\beta}}^R) - JH_{21}(\tilde{\boldsymbol{\beta}}^R)\right)(D_T^{ex}P^{ex})^{-1}[\frac{1}{T}(P^u Z_{2,t})', (\hat{V}_t^{ex,st}(\tilde{\boldsymbol{\beta}}^R))']' \\ &\quad - \left(\left\langle Z_{0,t}, \hat{V}_t^{ex}(\boldsymbol{b})\right\rangle H_{11}(\boldsymbol{b}) - JH_{21}(\tilde{\boldsymbol{\beta}}^R)\right)(D_T^{ex}P^{ex})^{-1}[\frac{1}{T}(P^u Z_{2,t})', (\hat{V}_t^{ex,st}(\boldsymbol{b}))']' \\ &=: M^R_{0,T}(\tilde{\boldsymbol{\beta}}^R)[\frac{1}{T}(P^u Z_{2,t})', (\hat{V}_t^{ex,st}(\tilde{\boldsymbol{\beta}}^R))']' - M^R_{0,T}(\boldsymbol{b})[\frac{1}{T}(P^u Z_{2,t})', (\hat{V}_t^{ex,st}(\boldsymbol{b}))']'.\end{aligned}$$

Since $(M^R_{0,T}(\tilde{\boldsymbol{\beta}}^R) - M^R_{0,T}(\boldsymbol{b}))$ is in $O_p(T^{-1})$, for the coefficient of the component $(P^u Z_{2,t})$ in δ_t it holds that

$$\left(\hat{C}^R_{-k_0}(\boldsymbol{b}) - \hat{C}^R_{-k_0}(\tilde{\boldsymbol{\beta}}^R)\right)P' \begin{bmatrix} I_{sS-\sum_1^S r_k} \\ 0 \end{bmatrix} = O_p(T^{-2}).$$

Thus, the dominant component in δ_t is $\breve{\alpha}^R(\tilde{\boldsymbol{\beta}}^R)(\tilde{\boldsymbol{\beta}}^R - \boldsymbol{b})'\mathcal{C}_\omega\mathcal{C}_\omega' Z_{1,t}$. Expanding the likelihood and extracting the dominant components leads to

$$\begin{aligned}&2L_T^{ex,R}(\tilde{\boldsymbol{\beta}}^R, \hat{\theta}) - 2L_T^{ex,R}(\boldsymbol{b}, \hat{\theta}) \\ &\quad = -T\text{tr}\left[\langle\hat{\varepsilon}_t(\boldsymbol{b}), \hat{\varepsilon}_t(\boldsymbol{b})\rangle^{-1}(\langle\delta_t, \hat{\varepsilon}_t(\boldsymbol{b})\rangle + \langle\hat{\varepsilon}_t(\boldsymbol{b}), \delta_t\rangle + \langle\delta_t, \delta_t\rangle)\right] + o_p(1) \\ &\quad = \text{tr}\left[\langle\hat{\varepsilon}_t(\boldsymbol{b}), \hat{\varepsilon}_t(\boldsymbol{b})\rangle^{-1}\breve{\alpha}^R(\tilde{\boldsymbol{\beta}}^R)\left(T(\tilde{\boldsymbol{\beta}}^R - \boldsymbol{b})'\mathcal{C}_\omega\right)\langle\mathcal{C}_\omega' Z_{1,t}, \hat{\varepsilon}_t(\boldsymbol{b})\rangle\right] \\ &\quad\quad + \text{tr}\left[\langle\hat{\varepsilon}_t(\boldsymbol{b}), \hat{\varepsilon}_t(\boldsymbol{b})\rangle^{-1}\langle\hat{\varepsilon}_t(\boldsymbol{b}), \mathcal{C}_\omega' Z_{1,t}\rangle\left(T\mathcal{C}_\omega'(\tilde{\boldsymbol{\beta}}^R - \boldsymbol{b})\right)\breve{\alpha}^R(\tilde{\boldsymbol{\beta}}^R)'\right] \\ &\quad\quad - \text{tr}\left[\langle\hat{\varepsilon}_t(\boldsymbol{b}), \hat{\varepsilon}_t(\boldsymbol{b})\rangle^{-1}\breve{\alpha}^R(\tilde{\boldsymbol{\beta}}^R)\left(T(\tilde{\boldsymbol{\beta}}^R - \boldsymbol{b})'\mathcal{C}_\omega\right)\frac{\langle\mathcal{C}_\omega' Z_{1,t}, \mathcal{C}_\omega' Z_{1,t}\rangle}{T}\left(T\mathcal{C}_\omega'(\tilde{\boldsymbol{\beta}}^R - \boldsymbol{b})\right)\breve{\alpha}^R(\tilde{\boldsymbol{\beta}}^R)'\right] \\ &\quad\quad + o_p(1)\end{aligned}$$

Application of the respective convergence results then implies

$$2L_T^{ex,R}(\tilde{\boldsymbol{\beta}}^R, \hat{\theta}) - 2L_T^{ex,R}(\boldsymbol{b}, \hat{\theta}) \xrightarrow{p} \text{tr}\left((\breve{\alpha}_\circ'\Sigma^{-1}\breve{\alpha}_\circ)\int_0^1(\text{d}\boldsymbol{V})\boldsymbol{F}'\left(\int_0^1 \boldsymbol{F}\boldsymbol{F}'\text{d}u\right)^{-1}\int_0^1 \boldsymbol{F}(\text{d}\boldsymbol{V})'\right).$$

Now notice that conditional upon $\boldsymbol{F}$ the matrix $\int_0^1(\text{d}\boldsymbol{V})\boldsymbol{F}'$ is Gaussian with mean zero and conditional variance $(\breve{\alpha}_\circ'\Sigma^{-1}\breve{\alpha}_\circ)^{-1} \otimes \int_0^1 \boldsymbol{F}\boldsymbol{F}'du$, such that conditionally on F the statistic is χ^2 with $2r(p-r)$ degrees of freedom. Consequently, this result also holds marginally.
Under the null hypothesis the above asymptotic result holds for both $\hat{\theta}_n^{c,\omega}$ and $\hat{\theta}_n^b$ as defined in the Theorem. Finally,

$$\begin{aligned}&2\left(L_T^{ex,R}(\tilde{\boldsymbol{\beta}}^R(\hat{\theta}_n^b), \hat{\theta}_n^b) - L_T^{ex,R}(\boldsymbol{b}, \hat{\theta}_n^b)\right) \\ &\quad < 2\left(L_T(\hat{\theta}_n^{c,\omega}) - L_T(\hat{\theta}_n^b)\right) < 2\left(L_T^{ex,R}(\tilde{\boldsymbol{\beta}}^R(\hat{\theta}_n^{c,\omega}), \hat{\theta}_n^{c,\omega}) - L_T^{ex,R}(\boldsymbol{b}, \hat{\theta}_n^{c,\omega})\right)\end{aligned}$$

implies that $2\left(L_T(\hat{\theta}_n^{c,\omega}) - L_T(\hat{\theta}_n^b)\right)$ follows the same distribution.
For the unrestricted concentration approach consider the expansion of $L_T^{ex,U}(\boldsymbol{\beta}_\circ, \hat{\theta})$ around its

minimizer $\tilde{\boldsymbol{\beta}}$ (where, thus, the first derivative is zero). The second order approximation of $2L_T^{ex,U}(\tilde{\boldsymbol{\beta}},\hat{\theta}) - 2L_T^{ex,U}(\boldsymbol{b},\hat{\theta})$ equals the difference between two terms, compare Johansen (1996, Lemma A.8). The first term is equal to

$$\begin{aligned} &T\mathrm{tr}\ \left((\tilde{\boldsymbol{\beta}}'\hat{S}_{11}\tilde{\boldsymbol{\beta}})^{-1}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_\circ)'(\hat{S}_{11}-\hat{S}_{11}\tilde{\boldsymbol{\beta}}(\tilde{\boldsymbol{\beta}}'\hat{S}_{11}\tilde{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{\beta}}'\hat{S}_{11})(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_\circ)\right) \\ &= T\mathrm{tr}\ \left((\tilde{\boldsymbol{\beta}}'\hat{S}_{11}\tilde{\boldsymbol{\beta}})^{-1}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_\circ)'\hat{S}_{11}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_\circ)\right) + O_p(T^{-1}) \end{aligned}$$

and the second term reduces to

$$\begin{aligned} &T\mathrm{tr}\ \left((\tilde{\boldsymbol{\beta}}'\hat{S}_{11,0}\tilde{\boldsymbol{\beta}})^{-1}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_\circ)'(\hat{S}_{11,0}-\hat{S}_{11,0}\tilde{\boldsymbol{\beta}}(\tilde{\boldsymbol{\beta}}'\hat{S}_{11,0}\tilde{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{\beta}}'\hat{S}_{11,0})(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_\circ)\right) \\ &= T\mathrm{tr}\ \left((\tilde{\boldsymbol{\beta}}'\hat{S}_{11,0}\tilde{\boldsymbol{\beta}})^{-1}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_\circ)'\hat{S}_{11}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_\circ)\right) + O_p(T^{-1}). \end{aligned}$$

Combining the results and using Johansen and Schaumburg (1999, Lemma 8), which implies that

$$(\tilde{\boldsymbol{\beta}}'\hat{S}_{11,0}\tilde{\boldsymbol{\beta}})^{-1} - (\tilde{\boldsymbol{\beta}}'\hat{S}_{11}\tilde{\boldsymbol{\beta}})^{-1} \xrightarrow{p} \breve{\alpha}_\circ'\Sigma^{-1}\breve{\alpha}_\circ$$

we find

$$2L_T^{ex}(\tilde{\boldsymbol{\beta}},\hat{\theta}) - 2L_T^{ex}(\boldsymbol{b},\hat{\theta}) = T\mathrm{tr}\ \left((\breve{\alpha}_\circ'\Sigma^{-1}\breve{\alpha}_\circ)(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_\circ)'\hat{S}_{11}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_\circ)\right) + O_p(T^{-1}).$$

From the previous convergence results we find that the limit of the above expression is given by

$$\mathrm{tr}\ \left((\breve{\alpha}_\circ'\Sigma^{-1}\breve{\alpha}_\circ)\int_0^1(\mathrm{d}\boldsymbol{V})\boldsymbol{F}'\left(\int_0^1\boldsymbol{F}\boldsymbol{F}'\mathrm{d}u\right)^{-1}\int_0^1\boldsymbol{F}(\mathrm{d}\boldsymbol{V})'\right).$$

From here the proof proceeds as for the restricted concentration approach.

(III) **Distribution of pseudo likelihood ratio test for linear hypotheses on** β
The developments for this pseudo likelihood ratio test are similar to the proof of Johansen (1996, Theorem 13.10.) and, hence, omitted.

B.5 Proof of Theorem 8

(I) **No deterministics**
The rank test statistic for the unrestricted approach in this case is given by

$$-2\log Q_T^U(H(r_\omega)/H(s)|\hat{\theta}) = -T\log\frac{|\hat{S}_{11}||\tilde{\boldsymbol{\beta}}'\hat{S}_{11,0}\tilde{\boldsymbol{\beta}}|}{|\hat{S}_{11,0}||\tilde{\boldsymbol{\beta}}'\hat{S}_{11}\tilde{\boldsymbol{\beta}}|}.$$

Using the identity

$$\begin{aligned} |[\tilde{\boldsymbol{\beta}},\tilde{\mathcal{C}}_\omega]'||\hat{S}_{11,0}||[\tilde{\boldsymbol{\beta}},\tilde{\mathcal{C}}_\omega]'| &= |[\tilde{\boldsymbol{\beta}},\tilde{\mathcal{C}}_\omega]'\hat{S}_{11,0}[\tilde{\boldsymbol{\beta}},\tilde{\mathcal{C}}_\omega]'| \\ &= |\tilde{\boldsymbol{\beta}}'\hat{S}_{11,0}\tilde{\boldsymbol{\beta}}||\tilde{\mathcal{C}}_\omega'(\hat{S}_{11,0}-\hat{S}_{11,0}\tilde{\boldsymbol{\beta}}(\tilde{\boldsymbol{\beta}}'\hat{S}_{11,0}\tilde{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{\beta}}\hat{S}_{11,0})\tilde{\mathcal{C}}_\omega| \\ &=: |\tilde{\boldsymbol{\beta}}'\hat{S}_{11,0}\tilde{\boldsymbol{\beta}}||\tilde{\mathcal{C}}_\omega'\hat{S}_{11,0\tilde{\boldsymbol{\beta}}}\tilde{\mathcal{C}}_\omega| \end{aligned}$$

and a similar one for the matrix $\hat{S}_{11}$, we find

$$-2\log Q_T^U(H(r_\omega)/H(s)|\hat{\theta}) = -T\log\frac{|\tilde{\mathcal{C}}_\omega'\hat{S}_{11,0\tilde{\boldsymbol{\beta}}}\tilde{\mathcal{C}}_\omega|}{|\tilde{\mathcal{C}}_\omega'\hat{S}_{11,\tilde{\boldsymbol{\beta}}}\tilde{\mathcal{C}}_\omega|} = -T\log\frac{|\frac{1}{T}\tilde{\mathcal{C}}_\omega'\hat{S}_{11,0\tilde{\boldsymbol{\beta}}}\tilde{\mathcal{C}}_\omega|}{|\frac{1}{T}\tilde{\mathcal{C}}_\omega'\hat{S}_{11,\tilde{\boldsymbol{\beta}}}\tilde{\mathcal{C}}_\omega|}.$$

For the restricted approach we have

$$-2\log Q_T^R(H(r_\omega)/H(s)|\hat{\theta}) = -T\left(\log\left|\left\langle\hat{\varepsilon}_t^R(\tilde{\boldsymbol{\beta}}^R),\hat{\varepsilon}_t^R(\tilde{\boldsymbol{\beta}}^R)\right\rangle\right| - \log\left|\langle\hat{\varepsilon}_t^R(I_{2s}),\hat{\varepsilon}_t^R(I_{2s})\rangle\right|\right)$$
$$= -T\log\left(|\hat{\Sigma}|\frac{|(\tilde{\mathcal{C}}_\omega^R)'\hat{N}_T\tilde{\mathcal{C}}_\omega^R|}{|(\tilde{\mathcal{C}}_\omega^R)'(\hat{N}_T+\hat{M}_T)\tilde{\mathcal{C}}_\omega^R|}\right)+\log|\hat{\Sigma}|,$$

compare the proof of Theorem 7. Using again the identities of Lemma 4 it follows that

$$-2\log Q_T^R(H(r_\omega)/H(s)|\hat{\theta}) = T\log\left|I_{s-r_\omega}+S_{1\varepsilon}^R(\tilde{\boldsymbol{\beta}}^R)\hat{\Sigma}^{-1}\hat{S}_{\varepsilon 1}^R(\tilde{\boldsymbol{\beta}}^R)((\tilde{\mathcal{C}}_\omega^R)'S_{11}^R(\tilde{\boldsymbol{\beta}}^R)\tilde{\mathcal{C}}_\omega^R)^{-1}\right|$$
$$= -T\log\frac{|\frac{1}{T}(\tilde{\mathcal{C}}_\omega^R)'\hat{S}_{11,0}^R(\tilde{\boldsymbol{\beta}}^R)\tilde{\mathcal{C}}_\omega^R|}{|\frac{1}{T}(\tilde{\mathcal{C}}_\omega^R)'\hat{S}_{11}^R(\tilde{\boldsymbol{\beta}}^R)\tilde{\mathcal{C}}_\omega^R|},$$

with $\hat{S}_{11,0}^R(\tilde{\boldsymbol{\beta}}^R) := \hat{S}_{11}^R(\tilde{\boldsymbol{\beta}}^R) - \hat{S}_{1\varepsilon}^R(\tilde{\boldsymbol{\beta}}^R)\hat{\Sigma}^{-1}\hat{S}_{\varepsilon 1}^R(\tilde{\boldsymbol{\beta}}^R)$, analogous to the unrestricted approach.
For the unrestricted approach, it follows from Lemma 7

$$T^{-1}\tilde{\mathcal{C}}_\omega'\hat{S}_{11,\tilde{\boldsymbol{\beta}}}\tilde{\mathcal{C}}_\omega := T^{-1}\tilde{\mathcal{C}}_\omega'(\hat{S}_{11}-\hat{S}_{11}\tilde{\boldsymbol{\beta}}(\tilde{\boldsymbol{\beta}}'\hat{S}_{11}\tilde{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{\beta}}'\hat{S}_{11})\tilde{\mathcal{C}}_\omega = T^{-1}\tilde{\mathcal{C}}_\omega'\hat{S}_{11}\tilde{\mathcal{C}}_\omega + o_p(1) \xrightarrow{d} \int_0^1 \boldsymbol{F}\boldsymbol{F}'\mathrm{d}u$$

and the same result holds for $\tilde{\mathcal{C}}_\omega'\hat{S}_{11,0\tilde{\boldsymbol{\beta}}}\tilde{\mathcal{C}}_\omega$ as well as for $(\tilde{\mathcal{C}}_\omega^R)'\hat{S}_{11}^R(\tilde{\boldsymbol{\beta}}^R)\tilde{\mathcal{C}}_\omega^R$ and $(\tilde{\mathcal{C}}_\omega^R)'\hat{S}_{11,0}^R(\tilde{\boldsymbol{\beta}}^R)\tilde{\mathcal{C}}_\omega^R$. Thus, the ratio in the above expression tends to 1 for both the restricted and the unrestricted approach.
Expanding the test statistic of the unrestricted approach leads to

$$-2\log Q_T^U(H(r_\omega)/H(s)|\hat{\theta}) = \mathrm{tr}\left((T^{-1}\tilde{\mathcal{C}}_\omega'\hat{S}_{11,\tilde{\boldsymbol{\beta}}}\tilde{\mathcal{C}}_\omega)^{-1}\tilde{\mathcal{C}}_\omega'\hat{S}_{10,\tilde{\boldsymbol{\beta}}}\hat{S}_{00,\tilde{\boldsymbol{\beta}}}^{-1}\hat{S}_{01,\tilde{\boldsymbol{\beta}}}\tilde{\mathcal{C}}_\omega\right)+o_p(1),$$

where $\hat{S}_{00,\tilde{\boldsymbol{\beta}}} := \hat{S}_{00}-\hat{S}_{01}\tilde{\boldsymbol{\beta}}'(\tilde{\boldsymbol{\beta}}'\hat{S}_{11}\tilde{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{\beta}}'\hat{S}_{10} = \langle\tilde{\varepsilon}_t,\tilde{\varepsilon}_t\rangle$ converges to Σ and

$$\begin{aligned}\tilde{\mathcal{C}}_\omega'\hat{S}_{10,\tilde{\boldsymbol{\beta}}} &:= \tilde{\mathcal{C}}_\omega'(\hat{S}_{10}-\hat{S}_{11}\tilde{\boldsymbol{\beta}}(\tilde{\boldsymbol{\beta}}'\hat{S}_{11}\tilde{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{\beta}}'\hat{S}_{10})\\ &= \tilde{\mathcal{C}}_\omega'\hat{S}_{10}-\tilde{\mathcal{C}}_\omega'\hat{S}_{11}\tilde{\boldsymbol{\beta}}\tilde{\alpha}'\\ &= \tilde{\mathcal{C}}_\omega'\left\langle\hat{R}_{1,t},\tilde{\varepsilon}_t\right\rangle \qquad = \tilde{\mathcal{C}}_\omega'\hat{S}_{1\varepsilon}-\tilde{\mathcal{C}}_\omega'\hat{S}_{11}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_\circ)\tilde{\alpha}'+o_p(1).\end{aligned}$$

Similarly, for the restricted concentration approach it holds that

$$-2\log Q_T^R(H(r_\omega)/H(s)|\hat{\theta}) = \mathrm{tr}\left(\left(T^{-1}(\tilde{\mathcal{C}}_\omega^R)'\hat{S}_{11}^R(\tilde{\boldsymbol{\beta}}^R)\tilde{\mathcal{C}}_\omega^R\right)^{-1}(\tilde{\mathcal{C}}_\omega^R)'\hat{S}_{1\varepsilon}^R(\tilde{\boldsymbol{\beta}}^R)\hat{\Sigma}^{-1}\hat{S}_{\varepsilon 1}^R(\tilde{\boldsymbol{\beta}}^R)\tilde{\mathcal{C}}_\omega^R\right)+o_p(1).$$

Further,

$$\begin{aligned}(\tilde{\mathcal{C}}_\omega^R)'\hat{S}_{1\varepsilon}^R(\tilde{\boldsymbol{\beta}}^R) &= (\tilde{\mathcal{C}}_\omega^R)'\left\langle\hat{R}_{1,t}(\tilde{\boldsymbol{\beta}}^R),\hat{\varepsilon}_t^R(\boldsymbol{\beta}_\circ)\right\rangle - (\tilde{\mathcal{C}}_\omega^R)'\left\langle\hat{R}_{1,t}(\tilde{\boldsymbol{\beta}}^R),\hat{\varepsilon}_t^R(\boldsymbol{\beta}_\circ)-\hat{\varepsilon}_t^R(\tilde{\boldsymbol{\beta}}^R)\right\rangle\\ &= (\tilde{\mathcal{C}}_\omega^R)'\left\langle\hat{R}_{1,t}(\tilde{\boldsymbol{\beta}}^R),\hat{\varepsilon}_t^R(\boldsymbol{\beta}_\circ)\right\rangle - (\tilde{\mathcal{C}}_\omega^R)'\left\langle\hat{R}_{1,t}(\tilde{\boldsymbol{\beta}}^R),\check{\alpha}^R(\tilde{\boldsymbol{\beta}}^R)(\tilde{\boldsymbol{\beta}}^R-\boldsymbol{\beta}_\circ)'\mathcal{C}_\omega\mathcal{C}_\omega'Z_{1,t}\right\rangle+o_p(1),\end{aligned}$$

where the $\hat{\varepsilon}_t^R(\boldsymbol{\beta}_\circ)-\hat{\varepsilon}_t^R(\tilde{\boldsymbol{\beta}}^R)$ is treated as δ_t in the proof of Theorem 7(II).
Note that $\tilde{\mathcal{C}}_\omega'$ respectively $(\tilde{\mathcal{C}}_\omega^R)'$ can be replaced by $\mathcal{C}_\omega'$ without changing the limit. From Theorem 7 the random matrices $\tilde{\mathcal{C}}_\omega'\hat{S}_{10,\tilde{\boldsymbol{\beta}}}$ and $(\tilde{\mathcal{C}}_\omega^R)'\hat{S}_{1\varepsilon}^R(\tilde{\boldsymbol{\beta}}^R)$ converge to

$$\begin{aligned}&\int_0^1\boldsymbol{F}(\mathrm{d}\boldsymbol{W})\begin{bmatrix}I_s\\0\end{bmatrix}-\int_0^1\boldsymbol{F}(\mathrm{d}\boldsymbol{W}')\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha}_\circ(\boldsymbol{\alpha}_\circ'\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha}_\circ)^{-1}\boldsymbol{\alpha}_\circ'\begin{bmatrix}I_s\\0\end{bmatrix}\\ &= \int_0^1\boldsymbol{F}(\mathrm{d}\boldsymbol{W}')(I_{2s}-\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha}_\circ(\boldsymbol{\alpha}_\circ'\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha}_\circ)^{-1}\boldsymbol{\alpha}_\circ')\begin{bmatrix}I_s\\0\end{bmatrix}\\ &= \int_0^1\boldsymbol{F}(\mathrm{d}\boldsymbol{W}')\boldsymbol{\alpha}_\perp((\boldsymbol{\alpha}_\perp)'\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha}_\perp)^{-1}(\boldsymbol{\alpha}_\perp)'\boldsymbol{\Sigma}\begin{bmatrix}I_s\\0\end{bmatrix},\end{aligned}$$

where $\boldsymbol{\alpha}_\perp = [\alpha_{k_0,\perp}]^{\mathbb{R}}$. Thus,

$$\left.\begin{array}{l}\tilde{\mathcal{C}}_\omega' \hat{S}_{10,\tilde{\boldsymbol{\beta}}} \hat{S}_{00,\tilde{\boldsymbol{\beta}}}^{-1} \hat{S}_{01,\tilde{\boldsymbol{\beta}}} \tilde{\mathcal{C}}_\omega \\ (\tilde{\mathcal{C}}_\omega^R)' \hat{S}_{1\varepsilon}^R(\tilde{\boldsymbol{\beta}}^R) \hat{\Sigma}^{-1} \hat{S}_{\varepsilon 1}^R(\tilde{\boldsymbol{\beta}}^R) \tilde{\mathcal{C}}_\omega^R\end{array}\right\} \xrightarrow{d} \int_0^1 \boldsymbol{F}(\mathrm{d}\boldsymbol{W}') M \int_0^1 (\mathrm{d}\boldsymbol{W}) \boldsymbol{F}',$$

where M is given by

$$M = \boldsymbol{\alpha}_\perp((\boldsymbol{\alpha}_\perp)'\boldsymbol{\Sigma}\boldsymbol{\alpha}_\perp)^{-1}(\boldsymbol{\alpha}_\perp)'\boldsymbol{\Sigma} \begin{bmatrix} I_s \\ 0 \end{bmatrix} \Sigma^{-1} \begin{bmatrix} I_s \\ 0 \end{bmatrix}' \boldsymbol{\Sigma}\boldsymbol{\alpha}_\perp((\boldsymbol{\alpha}_\perp)'\boldsymbol{\Sigma}\boldsymbol{\alpha}_\perp)^{-1}(\boldsymbol{\alpha}_\perp)'$$

such that

$$[M]^{\mathbb{C}} = \frac{1}{2}\boldsymbol{\alpha}_\perp((\boldsymbol{\alpha}_\perp)'\boldsymbol{\Sigma}\boldsymbol{\alpha}_\perp)^{-1}(\boldsymbol{\alpha}_\perp)'\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\alpha}_\perp((\boldsymbol{\alpha}_\perp)'\boldsymbol{\Sigma}\boldsymbol{\alpha}_\perp)^{-1}(\boldsymbol{\alpha}_\perp)' = \frac{1}{2}\boldsymbol{\alpha}_\perp((\boldsymbol{\alpha}_\perp)'\boldsymbol{\Sigma}\boldsymbol{\alpha}_\perp)^{-1}(\boldsymbol{\alpha}_\perp)'.$$

The asymptotic distribution for both the restricted and the unrestricted approach is then given by

$$\begin{aligned}\left.\begin{array}{l}-2\log Q_T^U(H(r_\omega)/H(s)|\hat{\theta}) \\ -2\log Q_T^R(H(r_\omega)/H(s)|\hat{\theta})\end{array}\right\} &\xrightarrow{d} \operatorname{tr}\left[\left(\int_0^1 \boldsymbol{F}\boldsymbol{F}'\mathrm{d}u\right)^{-1} \int_0^1 \boldsymbol{F}(\mathrm{d}\boldsymbol{W}')\ M \int_0^1 (\mathrm{d}\boldsymbol{W})\boldsymbol{F}'\right] \\ &= \operatorname{tr}\left[\left(\int_0^1 \boldsymbol{F}\boldsymbol{F}'\mathrm{d}u\right)^{-1} \int_0^1 \boldsymbol{F}(\mathrm{d}\boldsymbol{W}')\ [M]^{\mathbb{C}} \int_0^1 (\mathrm{d}\boldsymbol{W})\boldsymbol{F}'\right] \\ &= \frac{1}{2}\operatorname{tr}\left[\int_0^1 (\mathrm{d}\boldsymbol{B})\boldsymbol{B}' \left(\int_0^1 \boldsymbol{B}\boldsymbol{B}'\mathrm{d}u\right)^{-1} \int_0^1 \boldsymbol{B}(\mathrm{d}\boldsymbol{B}')\right],\end{aligned}$$

where $\boldsymbol{B} = (\boldsymbol{\alpha}_\perp' \boldsymbol{\Sigma}\boldsymbol{\alpha}_\perp)^{-1/2}\boldsymbol{\alpha}_\perp' \boldsymbol{W}_1$.
Note that the proof did not use any of the properties of Θ and, hence, the same asymptotic distribution holds if the assumptions of Proposition 2 are fulfilled. In general only consistent estimation of $(\hat{\underline{A}}, \hat{B})$ is needed.
These arguments show that the rank tests based on the SSECM lead to the well-known distributions. Finally also the pseudo likelihood ratio rank test is investigated. Recall $\hat{\theta}_n^{c,\omega}$ and $\hat{\theta}_n$ denoting the maximum of the pseudo log-likelihood function over the set of all systems with cointegrating rank of $r = s - c$ for unit root frequency ω_0 and the maximum over all stable systems respectively. Then the following relations hold:

$$L_T^{ex,R}(\tilde{\boldsymbol{\beta}}^R(\hat{\theta}_n), \hat{\theta}_n) \leq L_T(\hat{\theta}_n^{c,\omega}) \leq L_T^{ex,R}(I_{2s}, \hat{\theta}_n^{c,\omega}) \leq L_T(\hat{\theta}_n)$$

The first inequality holds since $\hat{\theta}_n^{c,\omega}$ maximizes the pseudo log-likelihood function over all systems with the specified structure. The second inequality holds because dropping the rank restriction can only increase the pseudo log-likelihood function. And the last holds due to $\hat{\theta}_n$ being the maximizer over all systems of order n. These inequalities imply

$$\begin{aligned}-2\log Q_T^R(H(r)/H(s), \hat{\theta}_n^{c,\omega}) &= -2\left(L_T(\hat{\theta}_n^{c,\omega}) - L_T^{ex,R}(I_{2s}, \hat{\theta}_n^{c,\omega})\right) \\ &\leq -2\left(L_T(\hat{\theta}_n^{c,\omega}) - L_T(\hat{\theta}_n)\right) \\ &\leq -2\left(L_T^{ex,R}(\tilde{\boldsymbol{\beta}}^R(\hat{\theta}_n), \hat{\theta}_n) - L_T(\hat{\theta}_n)\right) = -2\log Q_T^R(H(r)/H(s), \hat{\theta}_n).\end{aligned}$$

The evaluations above show that the limit of the left hand side and the one of the right hand side coincide. Thus, the bounds imply that also the pseudo likelihood ratio in the middle converges to the same limit. Moreover, also the essential term in these expressions is identical such that not only the limit is the same but in fact under the null hypothesis the difference of the test statistics converges to zero.
Note, furthermore, that for the test at one unit root the specification at the other unit roots does

not influence the limiting distribution under the null hypothesis.

(II) **Including deterministics**
The derivation of the asymptotic distribution of the various rank test statistics proceeds analogously and is, therefore, only briefly sketched here. The necessary adjustments are the inclusion of the deterministic term. Thus, define

$$\begin{aligned}\boldsymbol{\beta}_{\mathcal{D}} &:= [\beta_{\mathcal{D}}]^{\mathbb{R}}, \quad \beta_{\mathcal{D}} := [\beta_{k_0}', \xi_{k_0}']' \\ \boldsymbol{\beta}_{\mathcal{D},\circ} &:= [\beta_{\mathcal{D},\circ}]^{\mathbb{R}}, \quad \beta_{\mathcal{D},\circ} := [\beta_{k_0,\circ}', \xi_{k_0,\circ}']', \\ Z_{1,t}^{\mathcal{D}} &:= [[(X_t^{(k_0)})', s_{t,k_0}]']_v^{\mathbb{R}}, \\ V_t^{\mathcal{D}}(\theta) &:= [V_t(\theta), ([s_{t,1}]_v^{\mathbb{R}})', \dots, ([s_{t,k_0-1}]_v^{\mathbb{R}})', ([s_{t,k_0+1}]_v^{\mathbb{R}})', \dots, ([s_{t,\tilde{S}}]_v^{\mathbb{R}})', t].\end{aligned}$$

From this, $\hat{R}_{0,t}^{\mathcal{D}}$,$\hat{R}_{1,t}^{\mathcal{D}}$, $\hat{S}_{11}^{\mathcal{D}}, \hat{S}_{1\varepsilon}^{\mathcal{D}}$ and $\tilde{\boldsymbol{\beta}}_{\mathcal{D}}$ can be derived. Moreover, define

$$\mathcal{C}_{\mathcal{D}} = \begin{bmatrix} \mathcal{C}_{k_0,\circ} \\ T^{1/2} c_\xi \end{bmatrix}^{\mathbb{R}} \qquad \mathcal{B}_{\mathcal{D}} = \begin{bmatrix} \mathcal{B}_{k_0,\circ} \\ c_\xi \end{bmatrix}^{\mathbb{R}},$$

with c_ξ chosen such that $\boldsymbol{\beta}_{\mathcal{D},\circ}' \mathcal{C}_{\mathcal{D}} = 0$.
For the restricted approach define $V_t^{\mathcal{D},ex}(\boldsymbol{\beta}_{\mathcal{D}}, \theta) = [V_t^{\mathcal{D}}(\theta)', (\boldsymbol{\beta}_{\mathcal{D}}' Z_{1,t}^{\mathcal{D}})']'$. Using the blocks of the inverse matrix of interest

$$\begin{bmatrix} H_{11}^{\mathcal{D}}(\boldsymbol{\beta}_{\mathcal{D}}, \theta) & H_{12}^{\mathcal{D}}(\boldsymbol{\beta}_{\mathcal{D}}, \theta) \\ H_{21}^{\mathcal{D}}(\boldsymbol{\beta}_{\mathcal{D}}, \theta) & H_{22}^{\mathcal{D}}(\boldsymbol{\beta}_{\mathcal{D}}, \theta) \end{bmatrix} := \begin{bmatrix} \langle V_t^{\mathcal{D},ex}(\boldsymbol{\beta}_{\mathcal{D}}, \theta), V_t^{\mathcal{D},ex}(\boldsymbol{\beta}_{\mathcal{D}}, \theta) \rangle & G_{\mathcal{D}}(\boldsymbol{\beta}_{\mathcal{D}}, \theta) \\ G_{\mathcal{D}}(\boldsymbol{\beta}_{\mathcal{D}}, \theta)' & 0 \end{bmatrix}^{-1},$$

with $G_{\mathcal{D}}(\boldsymbol{\beta}_{\mathcal{D}}, \theta)$ defined accordingly by replacing $\boldsymbol{\beta}$ with $\boldsymbol{\beta}_{\mathcal{D}}$. Define the subblocks $H_{21,1}$ corresponding to the block $[I_s, 0] + \breve{\alpha}_\gamma \boldsymbol{\gamma}$ of $J(\breve{\alpha}_\gamma)$ and $H_{21,2}^{\mathcal{D}}(\boldsymbol{\beta}_{\mathcal{D}}, \theta)$ corresponding to $I^{\mathbb{R}}$ of $H_{21}^{\mathcal{D}}(\boldsymbol{\beta}_{\mathcal{D}}, \theta)$, and define

$$\begin{aligned}\varepsilon_t^{\mathcal{D},R}(\boldsymbol{\beta}_{\mathcal{D}}, \theta) &:= Z_{0,t} - \left(\left\langle Z_{0,t}, V_t^{\mathcal{D},ex}(\boldsymbol{\beta}_{\mathcal{D}}, \theta) \right\rangle H_{11}^{\mathcal{D}}(\boldsymbol{\beta}_{\mathcal{D}}, \theta) - J(0) H_{21}^{\mathcal{D}}(\boldsymbol{\beta}_{\mathcal{D}}, \theta) \right) V_t^{\mathcal{D},ex}(\boldsymbol{\beta}_{\mathcal{D}}, \theta), \\ R_{1,t}^{\mathcal{D},R}(\boldsymbol{\beta}_{\mathcal{D}}, \theta) &:= Z_{1,t}^{\mathcal{D}} - \left(\left\langle Z_{1,t}^{\mathcal{D}}, V_t^{\mathcal{D},ex}(\boldsymbol{\beta}_{\mathcal{D}}, \theta) \right\rangle H_{11}^{\mathcal{D}}(\boldsymbol{\beta}_{\mathcal{D}}, \theta) + H_{21,1}^{\mathcal{D}}(\boldsymbol{\beta}_{\mathcal{D}}, \theta) \right) V_t^{\mathcal{D},ex}(\boldsymbol{\beta}_{\mathcal{D}}, \theta),\end{aligned}$$

which are then used to define

$$\begin{aligned}\hat{S}_{11}^{\mathcal{D},R}(\boldsymbol{\beta}_{\mathcal{D}}) &:= \left\langle R_{1,t}^{\mathcal{D},R}(\boldsymbol{\beta}_{\mathcal{D}}, \hat{\theta}), R_{1,t}^{\mathcal{D},R}(\boldsymbol{\beta}_{\mathcal{D}}, \hat{\theta}) \right\rangle \\ \hat{S}_{1\varepsilon}^{\mathcal{D},R}(\boldsymbol{\beta}_{\mathcal{D}}) &:= \left\langle R_{1,t}^{\mathcal{D},R}(\boldsymbol{\beta}_{\mathcal{D}}, \hat{\theta}), \varepsilon_t^{\mathcal{D},R}(\boldsymbol{\beta}_{\mathcal{D}}, \hat{\theta}) \right\rangle\end{aligned}$$

for given $\hat{\theta}$. For these the following lemma holds.

Lemma 8 *Let $\{y_t\}_{t\in\mathbb{Z}}$ be a seasonally integrated MFI(1) process generated according to the assumptions stated in Assumption 1, and assume the true order n is known. Let $\hat{\theta}$ be the PML estimator over a suitable parameter space Θ fulfilling the requirements of Proposition 2. The following convergence results hold:*

$$\left.\begin{aligned} T^{-1}\mathcal{C}_{\mathcal{D}}'\hat{S}_{11}^{\mathcal{D}}\mathcal{C}_{\mathcal{D}} \\ T^{-1}\mathcal{C}_{\mathcal{D}}'\hat{S}_{11}^{\mathcal{D},R}(\boldsymbol{\beta}_{\mathcal{D},\circ})\mathcal{C}_{\mathcal{D}} \end{aligned}\right\} \xrightarrow{d} \mathcal{B}_{\mathcal{D}} \int_0^1 \boldsymbol{H}(u)\boldsymbol{H}(u)'du\ \mathcal{B}_{\mathcal{D}}'$$

$$\left.\begin{aligned} \mathcal{C}_{\mathcal{D}}'\hat{S}_{1\varepsilon}^{\mathcal{D}} \\ \mathcal{C}_{\mathcal{D}}'\hat{S}_{1\varepsilon}^{\mathcal{D},R}(\boldsymbol{\beta}_{\mathcal{D},\circ}) \end{aligned}\right\} \xrightarrow{d} \mathcal{B}_{\mathcal{D}} \int_0^1 \boldsymbol{H}(u)\big(d\boldsymbol{W}(u)'\big) \begin{bmatrix} I_s \\ 0 \end{bmatrix}.$$

where $\boldsymbol{H} = [\frac{1}{\sqrt{2}}((W_1 + iW_2)', 1)']^{\mathbb{R}}$ and W_1 and W_2 are two independent s-dimensional Brownian motions with variance Σ.

From here the derivation follows closely the arguments presented in the proof of the case without deterministic terms, with the dominant component in $\mathcal{C}_{\mathcal{D}}' R_{1,t}^{\mathcal{D}}$ equal to $[((2\mathcal{C}_{k_0,\circ} x_{t,k_0})', 2T^{1/2} c_\xi s_{t,k}')']_v^{\mathbb{R}}$. The asymptotic results of Theorem 8(II) are then a consequence of applying Lemma 8 on the dominant components of the expansion of the pseudo likelihood ratio test statistics.

Appendix C

Appendix to Chapter 3

C.1 SSECM

Proof of Theorem 9: Consider a state space system (A, B, C), with $\lambda_{|\max|}(A) \leq 1$, $\lambda_{|\max|}(\underline{A}) \leq 1$ and $I_s - \underline{A} = I_s - (A - BC)$ invertible. The corresponding transfer function is equal to $k(z) = I_s + zC(I_n - zA)^{-1}B$, while the inverse transfer function is given by $k^{-1}(z) = I_s - zC(I_n - z\underline{A})^{-1}B$. Consider the power series

$$\tilde{k}^{-1}(z) = \mathbf{\Delta}^2(z)I_s - \Pi z - \Gamma\mathbf{\Delta}(z)z - C(I_n - \underline{A})^{-2}\underline{A}^2 \sum_{m=1}^{\infty} \underline{A}^{m-1}B\mathbf{\Delta}^2(z)z^m,$$

with $\Pi = -I_s + C(I_n - \underline{A})^{-1}B$ and $\Gamma = -I_s - C(I_n - \underline{A})^{-2}\underline{A}B$, which equals $k^{-1}(z)$ as shown in what follows: for $m > 2$ the m-th power series coefficient can be calculated as

$$\begin{aligned}
\tilde{K}_m^- &= -C(I_n - \underline{A})^{-2}\underline{A}^2(\underline{A}^{m-1} - 2\underline{A}^{m-1-1} + \underline{A}^{m-2-1})B \\
&= -C(I_n - \underline{A})^{-2}(I_n - \underline{A})^2\underline{A}^{m-1}B = -C\underline{A}^{m-1}B = K_m^-
\end{aligned}$$

as required. For $m = 0, 1, 2$ it holds that

$$\begin{aligned}
\tilde{K}_0^- &= I_s, \\
\tilde{K}_1^- &= -2I_s + I_s - C(I_n - \underline{A})^{-1}B + I_s + C(I_n - \underline{A})^{-2}\underline{A}B - C(I_n - \underline{A})^{-2}\underline{A}^2B \\
&= -CB, \\
\tilde{K}_2^- &= I_s - I_s - C(I_n - \underline{A})^{-2}\underline{A}B + 2C(I_n - \underline{A})^{-2}\underline{A}^2B - C(I_n - \underline{A})^{-2}\underline{A}^3B \\
&= -C\underline{A}B.
\end{aligned}$$

Thus, for $t > 0$ and $y_0 = y_{-1} = 0$

$$\sum_{m=0}^{t-1} K_m^- \tilde{y}_{t-m} = \mathbf{\Delta}^2 I_s - \Pi\tilde{y}_{t-1} - \Gamma\mathbf{\Delta}\tilde{y}_{t-1} - C(I_n - \underline{A})^{-2}\underline{A}^2 \sum_{m=1}^{t-1} \underline{A}^{m-1}B\mathbf{\Delta}^2\tilde{y}_{t-m},$$

The representation then follows from setting

$$v_t := (I_n - \underline{A})^{-2}\underline{A}^2 \sum_{m=1}^{t-1} \underline{A}^{m-1}B\mathbf{\Delta}^2\tilde{y}_{t-m}, \quad v_1 := x_1.$$

With respect to deterministics note that for the constant we obtain

$$\begin{aligned}
\sum_{m=0}^{t-1} K_m^- d &= k^{-1}(1)d - \sum_{m=t}^{\infty} K_m^- d \\
&= -\Pi d + C\underline{A}^{t-1}(I_n - \underline{A})^{-1}Bd.
\end{aligned}$$

For the linear trend we obtain

$$\begin{aligned}\sum_{m=0}^{t-1} K_m^- e(t-m) &= \sum_{m=0}^{\infty} K_m^- e(t-m) - \sum_{m=t}^{\infty} K_m^- e(t-m) \\ &= -\Pi e t + (\Pi - \Gamma)e + C\underline{A}^{t-1} \sum_{m=t}^{\infty} \underline{A}^{m-t} B e(m-t) \\ &= -\Pi e(t-1) - \Gamma e + C\underline{A}^{t-1}(I_n - \underline{A})^{-2} B e - C\underline{A}^{t-1}(I_n - \underline{A})^{-1} B e.\end{aligned}$$

The above terms with factor $C\underline{A}^{t-1}$ are included in the SSECM through the starting value v_1. ■

C.2 PML Estimator

C.2.1 Preliminaries

The proof of consistency of the PML estimator for I(2) processes reiterates the arguments in the analogous proof in the MFI(1) case given in de Matos Ribeiro et al. (2020). To make the proof more readable we recount it almost literally, as focusing the discussion only on the differences of the proves would necessitate the reader to have both articles at hand.
The main ideas follow a similar line of thought as illustrated in the example given in Saikkonen (1995, Section 5). The key property in Saikkonen's work is the continuous convergence of certain quantities, which has also been developed in Saikkonen (1993). 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 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. For a compact space one obtains the following consequence, compare de Matos Ribeiro et al. (2020, Lemma 2):

Lemma 9 *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 $j \to \infty$.*

For the readers convenience we list the relevant convergence results from Lemma 1 of Sims et al. (1990).

Lemma 10 *Let $\{\varepsilon_t\}_{t \in \mathbb{N}}$ be a martingale difference sequence satisfying $\mathbb{E}(\varepsilon_t|\mathcal{F}_{t-1}) = 0$, $\mathbb{E}(\varepsilon_t \varepsilon_t') = \mathbb{E}(\varepsilon_t \varepsilon_t'|\mathcal{F}_{t-1}) = \Sigma_\circ > 0$ and $\mathbb{E}(\|\varepsilon_t\|^4) < \infty$. Further let $\boldsymbol{W}(u)$ denote the weak limit of $T^{-1/2}\sum_{t=1}^{[Tu]} \varepsilon_t$, where $[Tu]$ denotes the integer part of Tu, such that $\boldsymbol{W}(u)$ is a Brownian motion with variance*

$\Sigma_\circ$. *Then*

$$
\begin{array}{lll}
(i) & \left\langle \sum_{j=1}^{t-1} \varepsilon_j, \varepsilon_t \right\rangle & \xrightarrow{d} \int_0^1 \boldsymbol{W}(u) d\boldsymbol{W}(u)' \\
(ii) & T^{-1} \left\langle \sum_{j=1}^{t-1} \varepsilon_j, \sum_{j=1}^{t-1} \varepsilon_j \right\rangle & \xrightarrow{d} \int_0^1 \boldsymbol{W}(u) \boldsymbol{W}(u)' du \\
(iii) & T^{-1} \left\langle \sum_{k=1}^{t-1} \sum_{j=1}^{k-1} \varepsilon_j, \varepsilon_t \right\rangle & \xrightarrow{d} \int_0^1 \left(\int_0^u \boldsymbol{W}(v) dv \right) d\boldsymbol{W}(u)' \\
(iv) & T^{-2} \left\langle \sum_{k=1}^{t-1} \sum_{j=1}^{k-1} \varepsilon_j, \sum_{j=1}^{t-1} \varepsilon_j \right\rangle & \xrightarrow{d} \int_0^1 \left(\int_0^u \boldsymbol{W}(v) dv \right) \boldsymbol{W}(u)' du \\
(v) & T^{-3} \left\langle \sum_{k=1}^{t-1} \sum_{j=1}^{k-1} \varepsilon_t, \sum_{k=1}^{t-1} \sum_{j=1}^{k-1} \varepsilon_t \right\rangle & \xrightarrow{d} \int_0^1 \left(\int_0^u \boldsymbol{W}(v) dv \right) \left(\int_0^u \boldsymbol{W}(v) dv \right)' du \\
(vi) & T^{1/2} \langle 1, \varepsilon_t \rangle & \xrightarrow{d} \int_0^1 1 d\boldsymbol{W}(u)' \\
(vii) & T^{-1/2} \left\langle 1, \sum_{j=1}^{t-1} \varepsilon_j \right\rangle & \xrightarrow{d} \int_0^1 \boldsymbol{W}(u)' du \\
(viii) & T^{-3/2} \left\langle 1, \sum_{k=1}^{t-1} \sum_{j=1}^{k-1} \varepsilon_j \right\rangle & \xrightarrow{d} \int_0^1 \left(\int_0^u \boldsymbol{W}(v) dv \right)' du \\
(ix) & T^{-1/2} \langle t, \varepsilon_t \rangle & \xrightarrow{d} \int_0^1 u d\boldsymbol{W}(u)' \\
(x) & T^{-3/2} \left\langle t, \sum_{j=1}^{t-1} \varepsilon_j \right\rangle & \xrightarrow{d} \int_0^1 u \boldsymbol{W}(u)' du \\
(xi) & T^{-5/2} \left\langle t, \sum_{k=1}^{t-1} \sum_{j=1}^{k-1} \varepsilon_j \right\rangle & \xrightarrow{d} \int_0^1 u \left(\int_0^u \boldsymbol{W}(v) dv \right)' du
\end{array}
$$

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 $g(L,\theta)x_t = \sum_{j=0}^{t-1} G_j(\theta)x_{t-j}$, 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 and x_t is either integrated of order one, i. e. $x_t = \sum_{j=1}^{t-1} \varepsilon_j$ or of order two, i. e. $x_t = \sum_{k=1}^{t-1}\sum_{j=1}^{k-1} \varepsilon_j$. The notation here somewhat hides the fact that the summation is only performed for $t > 0$ or equivalently $x_t = 0$, $t < 0$ 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 11 *Let $g(z;\theta) = \sum_{j=0}^{\infty} G_j(\theta)z^j, k(z;\theta) = \sum_{j=0}^{\infty} K_j(\theta)z^j, \theta \in \Theta$ be two uniformly stable families of rational transfer functions, of finite McMillan degrees less or equal to n, where it is always assumed that the transfer functions are of the correct dimensions. Let $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be a martingale difference sequence fulfilling the assumptions of Lemma 10 with non-singular innovation variance $\Sigma_\circ$.*
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.*
$\langle g(L;\theta)1, k(L;\theta)\varepsilon_t\rangle \to 0.$

(ii) $\langle g(L;\theta)\sum_{j=1}^{t-1}\varepsilon_j, k(L;\theta)\varepsilon_t\rangle \xrightarrow{d}$

$$g(1;\theta)\left(\int_0^1 \boldsymbol{W}(u)d\boldsymbol{W}(u)'\right)k(1;\theta)' - g(1;\theta)\Sigma_\circ\tilde{k}(0;\theta)' + \lim_{t\to\infty}\mathbb{E}\tilde{g}(L;\theta)\varepsilon_{t-1}(k(L;\theta)\varepsilon_t)'$$

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

(iii) $T^{-1}\langle g(L;\theta)\sum_{j=1}^{t-1}\varepsilon_j, k(L;\theta)\sum_{j=1}^{t-1}\varepsilon_j\rangle \xrightarrow{d} g(1;\theta)\left(\int_0^1 \boldsymbol{W}(u)\boldsymbol{W}(u)'du\right)k(1;\theta)'$.

(iv) $\langle g(L;\theta)1, k(L;\theta)1\rangle \to g(1;\theta)k(1;\theta)'$.

(v) $T^{-1/2}\langle g(L;\theta)\sum_{j=1}^{t-1}\varepsilon_j, k(L;\theta)1\rangle \to g(1;\theta)\left(\int_0^1 \boldsymbol{W}(u)du\right)k(1;\theta)'$.

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

(vii) $T^{-3/2}\langle g(L;\theta)t, k(L;\theta)\sum_{j=1}^{t-1}\varepsilon_j\rangle \xrightarrow{d} g(1;\theta)\left(\int_0^1 u\boldsymbol{W}(u)'du\right)k(1;\theta)'$.

(viii) $T^{-1/2}\langle g(L;\theta)t, k(L;\theta)\varepsilon_t\rangle \xrightarrow{d} g(1;\theta)\left(\int_0^1 ud\boldsymbol{W}(u)'\right)k(1;\theta)$.

(ix) $T^{-1}\langle g(L;\theta)\sum_{k=1}^{t-1}\sum_{j=1}^{k-1}\varepsilon_j, k(L;\theta)\varepsilon_t\rangle \xrightarrow{d} g(1;\theta)\left(\int_0^1(\int_0^u \boldsymbol{W}(v)dv)d\boldsymbol{W}(u)'\right)k(1;\theta)$.

(x) $T^{-2}\langle g(L;\theta)\sum_{k=1}^{t-1}\sum_{j=1}^{k-1}\varepsilon_j, k(L;\theta)\sum_{j=1}^{t-1}\varepsilon_j\rangle \xrightarrow{d} g(1;\theta)\left(\int_0^1(\int_0^u \boldsymbol{W}(v)dv)\boldsymbol{W}(u)'du\right)k(1;\theta)$.

(xi) $T^{-3}\langle g(L;\theta)\sum_{k=1}^{t-1}\sum_{j=1}^{k-1}\varepsilon_j, k(L;\theta)\sum_{k=1}^{t-1}\sum_{j=1}^{k-1}\varepsilon_j\rangle \xrightarrow{d}$

$$g(1;\theta)\left(\int_0^1(\int_0^u \boldsymbol{W}(v)dv)(\int_0^u \boldsymbol{W}(v)dv)'du\right)k(1;\theta).$$

(xii) $T^{-3/2}\langle g(L;\theta)\sum_{k=1}^{t-1}\sum_{j=1}^{k-1}\varepsilon_j, k(L;\theta)1\rangle \xrightarrow{d} g(1;\theta)\left(\int_0^1(\int_0^u \boldsymbol{W}(v)dv)du\right)k(1;\theta)$
$T^{-5/2}\langle g(L;\theta)\sum_{k=1}^{t-1}\sum_{j=1}^{k-1}\varepsilon_j, k(L;\theta)t\rangle \xrightarrow{d} g(1;\theta)\left(\int_0^1(\int_0^u \boldsymbol{W}(v)dv)udu\right)k(1;\theta)$.

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

Proof: For (i)-(viii) compare de Matos Ribeiro et al. (2020, Lemma 4). The proof rests upon the results established in Lemma 10.
Decompose $g(z;\theta) = g(1;\theta) + (1-z)\tilde{g}(z;\theta)$, where 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:

$$\begin{aligned} g(L;\theta)\sum_{k=1}^{t-1}\sum_{j=1}^{k-1}\varepsilon_j &= \sum_{i=0}^{t-1}G_i(\theta)\sum_{k=1}^{t-1}\sum_{j=1}^{k-1}\varepsilon_j \\ &= g(1;\theta)\sum_{k=1}^{t-1}\sum_{j=1}^{k-1}\varepsilon_j + \tilde{g}(L;\theta)(\sum_{k=1}^{t-1}\sum_{j=1}^{k-1}\varepsilon_j - \sum_{k=1}^{t-2}\sum_{j=1}^{k-1}\varepsilon_j) \\ &= g(1;\theta)\sum_{k=1}^{t-1}\sum_{j=1}^{k-1}\varepsilon_j + \tilde{g}(L;\theta)\sum_{j=1}^{t-2}\varepsilon_j \end{aligned}$$

for $t \in \mathbb{N}$. Then item (ix) follows from

$$\begin{aligned} &T^{-1}\langle\sum_{k=1}^{t-1}\sum_{j=1}^{k-1}\varepsilon_j, k(L;\theta)\varepsilon_t\rangle \\ &\quad = T^{-1}\langle\sum_{k=1}^{t-1}\sum_{j=1}^{k-1}\varepsilon_j, k(1;\theta)\varepsilon_t\rangle + T^{-1}\langle\sum_{k=1}^{t-1}\sum_{j=1}^{k-1}\varepsilon_j, \tilde{k}(L;\theta)(1-L)\varepsilon_t\rangle. \end{aligned}$$

The first term converges to $\int_0^1 \boldsymbol{W}(u)d\boldsymbol{W}(u)'\, k(1;\theta)'$ according to Lemma 10. The second term is equal to

$$\begin{aligned}
&T^{-2}\sum_{t=1}^{T}\left[\sum_{k=1}^{t-1}\sum_{j=1}^{k-1}\varepsilon_j(\tilde{k}(L;\theta)\varepsilon_t)' - \sum_{k=1}^{t-1}\sum_{j=1}^{k-1}\varepsilon_j(\tilde{k}(L;\theta)\varepsilon_{t-1})'\right] \\
&= T^{-2}\sum_{t=1}^{T-1}\left[\sum_{k=1}^{t-1}\sum_{j=1}^{k-1}\varepsilon_j - \sum_{k=1}^{t}\sum_{j=1}^{k-1}\varepsilon_j\right](\tilde{k}(L;\theta)\varepsilon_t)' + o_P(1) \\
&= -T^{-2}\sum_{t=1}^{T-1}\sum_{j=1}^{t-1}\varepsilon_j(\tilde{k}(L;\theta)\varepsilon_t)' + o_P(1),
\end{aligned}$$

where the $o_P(1)$ term is due to $T^{-2}\sum_{k=1}^{T-1}\sum_{j=1}^{k-1}\varepsilon_j(\tilde{k}(L;\theta)\varepsilon_T)'$. This term converges to zero due to (ii). Combining this with pre-multiplication of $\sum_{k=1}^{t-1}\sum_{j=1}^{k-1}\varepsilon_j$ with $g(1;\theta)$ then delivers the result. Item (x) and (xi) can be shown using the same approach. The proof of (xii) follows from

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

where the $o(1)$ term is of order $O(\rho^t)$, and

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

where $g^*(z) = \sum_{j=1}^{\infty} jG_j(\theta)z^{j-1}$. The difference of $g(L;\theta)t$ and $g(1;\theta)t - g^*(1;\theta)$ is of order $O(t\rho^t)$. The result then follows from Lemma 10.
The fulfillment of Condition USE for the sequences considered in (ix) to (xii) is left to be shown, but proceeds as for the other terms. 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. Consider, e.g., $g(1;\theta)\langle\sum_{k=1}^{t-1}\sum_{j=1}^{k-1}\varepsilon_j, \varepsilon_t\rangle k(1;\theta)'$, 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. ■

Lemma 12 *Let the I(2) process $\{y_t\}_{t\in\mathbb{Z}}$ be generated as in Theorem 10. Define the pseudo likelihood function and the prediction error criterion function*

$$\begin{aligned}
\mathcal{L}_T(\theta,\sigma,d,e;Y_T) &= L_T(k(z;\theta),\sigma,d,e;Y_T), \\
\mathcal{L}_{PE,T}(\theta,\sigma,d,e;Y_T) &= L_{PE,T}(k(z;\theta),\sigma,d,e;Y_T),
\end{aligned}$$

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 $\boldsymbol{\Gamma}$ over the parameters $\theta \in \boldsymbol{\Theta}_{\boldsymbol{\Gamma}}$, $\sigma \in \boldsymbol{\Theta}_\Sigma$, $(d,e) \in \boldsymbol{\Theta}_D$, where all sets are compact such that $\inf_{\Sigma\in\boldsymbol{\Theta}_\Sigma}\lambda_{min}(\Sigma) > 0$, $\sup_{\theta\in\boldsymbol{\Theta}_{\boldsymbol{\Gamma}}}\lambda_{|max|}(\underline{\mathcal{A}}(\theta)) < 1$ and $\sup_{(d,e)\in\boldsymbol{\Theta}_D}(\|d\|+\|e\|) < \infty$.
Then

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

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

The proof proceeds as in the MFI(1) case, compare de Matos Ribeiro et al. (2020, Lemma 5).

C.2.2 Proof of Consistency

Consider a state space system given by:

$$\begin{aligned} y_t &= \mathcal{C}_1 x_{t,1} + \mathcal{C}_2 x_{t,2} + \mathcal{C}_3 x_{t,3} + \mathcal{C}_\bullet x_{t,\bullet} + \varepsilon_t + \mathcal{D}_t, \\ x_{t+1,1} &= x_{t,1} + x_{t,2} + \mathcal{B}_1 \varepsilon_t, \quad x_{1,1} = 0, \\ x_{t+1,2} &= \qquad\quad x_{t,2} + \mathcal{B}_2 \varepsilon_t, \quad x_{1,2} = 0, \\ x_{t+1,3} &= \qquad\quad x_{t,3} + \mathcal{B}_3 \varepsilon_t, \quad x_{1,3} = 0, \\ x_{t+1,\bullet} &= \mathcal{A}_\bullet x_{t,\bullet} + \mathcal{B}_\bullet x_{t,\bullet}, \quad x_{1,\bullet} = \textstyle\sum_{j=0}^{\infty} \mathcal{A}_\bullet^j \mathcal{B}_\bullet \varepsilon_{1-j}, \end{aligned}$$

Let $P_1 := \mathcal{C}_1 \mathcal{C}_1'$ and $P_{1\perp} := I_s - P_1$ denote the projection onto the column space of $\mathcal{C}_1$ and its ortho-complement respectively. Then with $\mathcal{D}_t = Ds_t$, with $D := [d, e]$ and $s_t = [1, t]'$, we find that

$$\begin{aligned} \check{y}_t - \check{\mathcal{D}}_T :&= (P_1 \boldsymbol{\Delta} + P_{1\perp})(y_t - \mathcal{D}_t) \\ &= \mathcal{C}_1 \boldsymbol{\Delta} x_{t,1} + P_1 \mathcal{C}_\bullet \boldsymbol{\Delta} x_{t,\bullet} + P_1 \boldsymbol{\Delta} \varepsilon_t + \mathcal{C}_2 x_{t,2} + \mathcal{C}_3 x_{t,3} + P_{1\perp} \mathcal{C}_\bullet x_{t,\bullet} + P_{1,\perp} \varepsilon_t \\ &= \mathcal{C}_1 (x_{t-1,2} + \mathcal{B}_1 \varepsilon_{t-1} - \mathcal{C}_1' \mathcal{C}_\bullet x_{t-1,\bullet} - \mathcal{C}_1' \varepsilon_{t-1}) + \mathcal{C}_2 x_{t,2} + \mathcal{C}_3 x_{t,3} + \mathcal{C}_\bullet x_{t,\bullet} + \varepsilon_t \end{aligned}$$

is the sum of a stationary process with a deterministic part

$$\check{\mathcal{D}}_T = (P_1 \boldsymbol{\Delta} + P_{1\perp}) Ds_t = Ds_t - P_1 Ds_{t-1} = Ds_t - P_1 D\mathcal{S}^{-1} s_t = \left(D - P_1 D \mathcal{S}^{-1}\right) s_t =: \check{D} s_t.$$

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 deterministics. Hence, it can be written as

$$\check{y}_t = \check{k}(L)\varepsilon_t + \check{D} s_t,$$

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

$$\begin{aligned} \check{y}_t - \check{D} s_t &= \mathcal{C}_1 \check{x}_{t,1} + \mathcal{C}_2 x_{t,2} + \mathcal{C}_3 x_{t,3} + \mathcal{C}_\bullet x_{t,\bullet} + \varepsilon_t, \\ \begin{bmatrix} \check{x}_{t+1,1} \\ x_{t+1,2} \\ x_{t+1,3} \\ x_{t+1,\bullet} \end{bmatrix} &= \begin{bmatrix} 0 & I & 0 & -\mathcal{C}_1' \mathcal{C}_\bullet \\ 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & \mathcal{A}_\bullet \end{bmatrix} \begin{bmatrix} \check{x}_{t,1} \\ x_{t,2} \\ x_{t,3} \\ x_{t,\bullet} \end{bmatrix} + \begin{bmatrix} \mathcal{B}_1 - \mathcal{C}_1' \\ \mathcal{B}_2 \\ \mathcal{B}_3 \\ \mathcal{B}_\bullet \end{bmatrix} \varepsilon_t. \end{aligned}$$

The transfer function $\check{k}(z)$ corresponds to process integrated of order one. The zeros are the eigenvalues of

$$\begin{bmatrix} 0 & I & 0 & -\mathcal{C}_1' \mathcal{C}_\bullet \\ 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & \mathcal{A}_\bullet \end{bmatrix} - \begin{bmatrix} \mathcal{B}_1 - \mathcal{C}_1' \\ \mathcal{B}_2 \\ \mathcal{B}_3 \\ \mathcal{B}_\bullet \end{bmatrix} [\mathcal{C}_1 \quad \mathcal{C}_2 \quad \mathcal{C}_3 \quad \mathcal{C}_\bullet] = \mathcal{A} - \mathcal{B}\mathcal{C},$$

where $(\mathcal{A}, \mathcal{B}, \mathcal{C})$ denotes the original system in canonical form. Hence, invertibility of $k(z)$ implies invertibility of $\check{k}(z)$. Note that we can further rewrite $\check{y}_t - \check{D} s_t$ as

$$\check{y}_t - \check{D} s_t = \mathcal{C}_1 (-\mathcal{B}_2 \varepsilon_{t-1} + \mathcal{B}_1 \varepsilon_{t-1} - \mathcal{C}_1' \mathcal{C}_\bullet x_{t-1,\bullet} - \mathcal{C}_1' \varepsilon_{t-1}) + (\mathcal{C}_1 + \mathcal{C}_2) x_{t,2} + \mathcal{C}_3 x_{t,3} + \mathcal{C}_\bullet x_{t,\bullet} + \varepsilon_t,$$

where the common I(1)-trends enter the process $\check{y}_t - \check{D} s_t$ through the matrix $[\mathcal{C}_1 + \mathcal{C}_2, \mathcal{C}_3]$. Thus, let $P_2 := [\mathcal{C}_1 + \mathcal{C}_2, \mathcal{C}_3] \left([\mathcal{C}_1 + \mathcal{C}_2, \mathcal{C}_3]'[\mathcal{C}_1 + \mathcal{C}_2, \mathcal{C}_3]\right)^{-1} [\mathcal{C}_1 + \mathcal{C}_2, \mathcal{C}_3]'$ and $P_{2\perp} = I_s - P_2$. It follows that

$$\tilde{y}_t = (P_2 \boldsymbol{\Delta} + P_{2\perp}) \check{y}_t = (P_2 \boldsymbol{\Delta} + P_{2\perp}) \check{k}(L) \varepsilon_t + (P_1 \boldsymbol{\Delta} + P_{1\perp}) \check{D} s_t = \tilde{k}(L) \varepsilon_t + \tilde{D} s_t.$$

The transfer function $\tilde{k}(z)$ is stable and $\tilde{D} := (\check{D} - P_2 \check{D} \mathcal{S}^{-1}) = D - (P_1 + P_2) D \mathcal{S}^{-1} + P_2 P_1 D \mathcal{S}^{-2}$. The zeros of $\tilde{k}(z)$ are the eigenvalues of

$$\begin{bmatrix} 0 & -\check{\mathcal{C}}_1' \check{\mathcal{C}}_\bullet \\ 0 & \check{\mathcal{A}}_\bullet \end{bmatrix} - \begin{bmatrix} \check{\mathcal{B}}_1 - \check{\mathcal{C}}_1' \\ \check{\mathcal{B}}_\bullet \end{bmatrix} [\check{\mathcal{C}}_1 \quad \check{\mathcal{C}}_\bullet] = \begin{bmatrix} I_c - \check{\mathcal{B}}_1 \check{\mathcal{C}}_1 & -\check{\mathcal{B}}_1 \check{\mathcal{C}}_\bullet \\ -\check{\mathcal{B}}_\bullet \check{\mathcal{C}}_1 & \check{\mathcal{A}}_\bullet - \check{\mathcal{B}}_\bullet \check{\mathcal{C}}_\bullet \end{bmatrix} = \check{\mathcal{A}} - \check{\mathcal{B}} \check{\mathcal{C}},$$

where $(\check{\mathcal{A}}, \check{\mathcal{B}}, \check{\mathcal{C}})$ denotes the system corresponding to $\check{k}(z)$ and the I(1) process $\check{y}_t - \check{D}s_t$ in canonical form. Hence, invertibility of $\check{k}(z)$ implies invertibility of $\tilde{k}(z)$.
Therefore, the Gaussian likelihood function for $Y_T := [y_1', ..., y_T']'$ can be calculated using $\tilde{Y}_T := [y_1', y_2', \tilde{y}_3', ..., \tilde{y}_T']'$, which – being a linear invertible transformation of Y_T – is also Gaussian distributed. Note that $y_1 = \mathcal{C}_\bullet x_{1,\bullet} + \varepsilon_1 + Ds_1$ and $y_2 = \sum_{j=1}^3 \mathcal{C}_j\mathcal{B}_j\varepsilon_1 + \mathcal{C}_\bullet x_{2,\bullet} + \varepsilon_2 + Ds_2$ for $x_{1,u} = 0$. Then $-2/T$ times the Gaussian likelihood function for Y_T equals:

$$\begin{aligned} L_T(k, D, \Sigma) &= \frac{1}{T}\left(\log|\Gamma_T(k,\Sigma)| + (Y_T - (I_T \otimes D)D_T)'\Gamma_T(k,\Sigma)^{-1}(Y_T - (I_T \otimes D)\mathbf{D}_T)'\right) \\ &= \frac{1}{T}\left(\log|\Gamma_T(\tilde{k},\Sigma)| + (\tilde{Y}_T - (I_T \otimes \tilde{D})D_T)'\Gamma_T(\tilde{k},\Sigma)^{-1}(\tilde{Y}_T - (I_T \otimes \tilde{D})D_T)'\right), \end{aligned}$$

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 $\text{diag}(0, P_\bullet(\theta,\sigma))$ is neglected. Except for the inclusion of y_1, y_2 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 $\Theta \subset M_{n,\bullet} \times \underline{P^{c_1}} \times \underline{P^{c_1+c_2}} \times \underline{D} \times \underline{\Sigma}$ 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 $\underline{P^{c_1}}$ and $\underline{P^{c_1+c_2}}$ (space of projector matrices in $\mathbb{R}^{s\times s}$ with rank equal to c_1 respectively $c_1 + c_2$), the set $\underline{D}$ (the vectorization of all $s \times 2$ 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_1 and P_2, there is a 1-1 mapping between the k and $\tilde{k}$:

$$\tilde{k}(z) = (P_2\mathbf{\Delta} + P_{2\perp})(P_1\mathbf{\Delta} + P_{1\perp})k(z) \Rightarrow k(z) = (P_1\mathbf{\Delta} + P_{1\perp})^{-1}(P_2\mathbf{\Delta} + P_{2\perp})^{-1}\tilde{k}(z).$$

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 $\mathcal{C}_1$ and $[\mathcal{C}_1 + \mathcal{C}_2, \mathcal{C}_3]$, relating to $P_1 = \mathcal{C}_1\mathcal{C}_1'$ and $P_2 = [\mathcal{C}_1 + \mathcal{C}_2, \mathcal{C}_3]([\mathcal{C}_1 + \mathcal{C}_2, \mathcal{C}_3]'[\mathcal{C}_1 + \mathcal{C}_2, \mathcal{C}_3])^{-1}[\mathcal{C}_1 + \mathcal{C}_2, \mathcal{C}_3]'$, and the remaining ones relating to $\tilde{k}$. This conforms with the parameterization suggested in Bauer et al. (2020), which also considers the parameters corresponding to $\mathcal{C}_u$ and the ones corresponding to $\mathcal{C}_\bullet$, $\mathcal{A}$, and $\mathcal{B}$ separately.
The proof of Theorem 10 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 quintuple (k, P_1, P_2, D, Σ) as θ in this section. 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 $L_T(\theta) = \tilde{L}_T(\tilde{k}, P_1, P_2, \tilde{D}, \Sigma) = L_T(k, D, \Sigma)$ over Θ.
Now follow the proof of Theorem 4.2.1. of HD. As on p. 112 we define $u(t,\theta) = \tilde{y}_t - \tilde{D}s_t, t = 3, .., T$, where $u(1,\theta) = y_1 - \tilde{D}s_1$, $u(2,\theta) = y_2 - (P_1 + P_2)y_1 - \tilde{D}s_2$ and $u_T(\theta) = \{u(t,\theta)\}_{t=1,...,T}$. Note that here $u(1,\theta)$ and $u(2,\theta)$ deviates from the 'regular' definition $\tilde{y}_1 - \tilde{D}s_1 = u(1,\theta) - (P_1 + P_2)y_0 + P_2P_1y_{-1}$ and $\tilde{y}_2 - \tilde{D}s_2 = u(2,\theta) + P_2P_1y_0$. 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 $\theta = (k, P_1, P_2, D, \Sigma)$ 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 $\tilde{\theta} = (\tilde{k}, P_1, P_2, \tilde{D}, \Sigma)$.

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 $\theta = (k, D, \Sigma)$ maps onto $\tilde{\theta} = (\tilde{k}, \tilde{D}, \Sigma) = \Lambda(\theta)$ 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. Consider $\tilde{\Gamma}_T(\tilde{\theta})$ in more depth: It is defined as the variance of

$$u_T(\theta) = \tilde{Y}_{1,T}(\tilde{\theta}) + \begin{pmatrix} (P_1+P_2)y_0 - P_2P_1y_{-1} \\ -P_2P_1y_0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\Rightarrow \tilde{\Gamma}_T(\tilde{\theta}) = \begin{bmatrix} V(y_{1,2}) & Cov(y_{1,2}, \tilde{Y}_{3,T}(\tilde{\theta})) \\ Cov(\tilde{Y}_{3,T}(\tilde{\theta}), y_{1,2}) & \Gamma_{T-2}(\tilde{\theta}) \end{bmatrix},$$

where $\tilde{Y}_{i,T}(\tilde{\theta}) = [\tilde{y}_t(\theta) - \tilde{D}s_t]_{t=i,\dots,T}$. Here $y_{1,2} = [(\mathcal{C}_\bullet x_{1,\bullet} + \varepsilon_1 - Ds_1)', (\mathcal{C}_u x_{2,u} + \mathcal{C}_\bullet x_{2,\bullet} + \varepsilon_2 - Ds_2)']'$. This implies, using the block matrix inversion, that

$$\begin{aligned}\tilde{\Gamma}_T(\tilde{\theta})^{-1} &= \begin{bmatrix} 0 & 0 \\ 0 & \Gamma_{T-2}(\tilde{\theta})^{-1} \end{bmatrix} \\ &+ \begin{bmatrix} I_s \\ -\Gamma_{T-2}(\tilde{\theta})^{-1}Cov(\tilde{Y}_{3,T}(\tilde{\theta}), y_{1,2}) \end{bmatrix} V_\pi(\tilde{\theta})^{-1} \begin{bmatrix} I_s & -Cov(y_{1,2}, \tilde{Y}_{3,T}(\tilde{\theta}))\Gamma_{T-2}(\tilde{\theta})^{-1} \end{bmatrix},\end{aligned} \tag{C.1}$$

where $V_\pi(\tilde{\theta}) := V(y_{1,2}) - Cov(y_{1,2}, \tilde{Y}_{3,T}(\tilde{\theta}))\Gamma_{T-2}(\tilde{\theta})^{-1}Cov(\tilde{Y}_{3,T}(\tilde{\theta})), y_{1,2})$. Now, for stable $\tilde{\theta}$ it follows that $V_\pi(\tilde{\theta}) \geq I_{2s}\lambda_{min}(\Sigma)$.
It follows that

$$u_T(\tilde{\theta})'\tilde{\Gamma}_T(\theta)^{-1}u_T(\tilde{\theta})/T \geq u_{3:T}(\tilde{\theta})'\Gamma_{T-2}(\tilde{\theta})^{-1}u_{3:T}(\tilde{\theta})/T,$$

where $u_{3:T}(\tilde{\theta}) = \{u_t(\tilde{\theta})\}_{t=3,\dots,T} = \tilde{Y}_{3,T}(\tilde{\theta})$ 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 dealt with in Chapter 4 of HD. Further, note that the difference between these two terms equals

$$y_{1,2,\pi}(\tilde{\theta})'V_\pi(\tilde{\theta})^{-1}y_{1,2,\pi}(\tilde{\theta})/T,$$

where this equation defines $y_{1,2,\pi}(\tilde{\theta})$.
It follows that the criterion function to be considered equals

$$\tilde{L}_T(\tilde{\theta}) \geq \log|\tilde{\Gamma}_T(\tilde{\theta})| + u_{3:T}(\tilde{\theta})'\Gamma_{T-2}(\tilde{\theta})^{-1}u_{3:T}(\tilde{\theta})/T = \log|\tilde{\Gamma}_T(\tilde{\theta})| + \tilde{Q}_T(\tilde{\theta}).$$

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

$$\begin{aligned}&\begin{bmatrix} I_s & -Cov(y_{1,2}, \tilde{Y}_{3,T}(\tilde{\theta}))\Gamma_{T-2}(\tilde{\theta})^{-1} \\ 0 & I \end{bmatrix} \tilde{\Gamma}_T(\tilde{\theta}) \begin{bmatrix} I_s & 0 \\ -\Gamma_{T-2}(\tilde{\theta})^{-1}Cov(\tilde{Y}_{3,T}(\tilde{\theta}), y_{1,2}) & I \end{bmatrix} \\ &= \begin{bmatrix} V_\pi(\tilde{\theta}) & 0 \\ 0 & \Gamma_{T-2}(\tilde{\theta}) \end{bmatrix}.\end{aligned}$$

Therefore, the determinant is the product of $|V_\pi(\tilde\theta)|$ and $|\Gamma_{T-2}(\tilde\theta)|$. From $V_\pi(\theta) \le V(y_{1,2})$ we see that

$$\frac{1}{T}\log\det\tilde\Gamma_T(\tilde\theta) = \frac{1}{T}\log|V_\pi(\tilde\theta)| + \frac{1}{T}\log|\Gamma_{T-2}(\tilde\theta)| \le \frac{1}{T}\log|\Gamma_{T-2}(\tilde\theta)| + \frac{1}{T}\log|V(y_{1,2})|.$$

The behavior of $T^{-1}\log|\Gamma_{T-2}(\tilde\theta)|$ follows as in HD, Lemma 4.2.2., p. 116: $T^{-1}\log|\Gamma_{T-2}(\tilde\theta)| \ge \log|\Sigma|$ and $\lim_{T\to\infty} T^{-1}\log|\Gamma_{T-2}(\tilde\theta)| = \log|\Sigma|$ 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 $\log|\Gamma_{T-2}(\theta_j)| \to \infty$ as $\lambda_{|max|}(P_\bullet) \to \infty$. For $\tilde\theta \in \hat\Theta$ we have

$$0 < V_\pi(\tilde\theta) \le V(y_{1,2}) < \infty,$$

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 $V_\pi(\theta_j) \ge \lambda_{min}(\Sigma)I_s$, which hence must also hold 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

$$\begin{aligned} Q_T(\tilde\theta) &= T^{-1}u_T(\tilde\theta)'\tilde\Gamma_T(\tilde\theta)^{-1}u_T(\tilde\theta) \\ &= T^{-1}y_{1,2,\pi}(\tilde\theta)'V_\pi(\tilde\theta)^{-1}y_{1,2,\pi}(\tilde\theta) + T^{-1}u_{3:T}(\tilde\theta)\Gamma_{T-2}(\tilde\theta)^{-1}u_{3:T}(\tilde\theta) \\ &\ge T^{-1}u_{3:T}(\tilde\theta)'\Gamma_{T-2}(\tilde\theta)^{-1}u_{3:T}(\tilde\theta) =: \tilde Q_T(\tilde\theta), \end{aligned}$$

using the block matrix inversion. HD define the function $Q(\tilde\theta)$ as the limit of $\tilde Q_T$ (on Θ^*). Here

$$Q(\tilde\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}[(\tilde k\Sigma\tilde k^*)^{-1}(\tilde k_0\Sigma_0\tilde k_0^*))]d\omega.$$

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

$$\begin{aligned} \phi_\eta(\omega;\tilde\theta) &= 2\pi|c(e^{i\omega})|^2\{N(e^{i\omega})\Sigma N(e^{i\omega})^* + \eta I_s\}^{-1} \le 2\pi|c(e^{i\omega})|^2\{N(e^{i\omega})\Sigma N(e^{i\omega})^*\}^{-1} = f_u^{-1}(\omega), \\ &\forall\eta > 0, \end{aligned}$$

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 $\phi_\eta^{-1} \ge f_u$ it holds that $\Gamma_T(\phi_\eta^{-1}) \ge \Gamma_T(f_u)$ and thus $\Gamma_T(\phi_\eta^{-1})^{-1} \le \Gamma_T(f_u)^{-1}$, see HD, (4.2.18) on p. 119.
Then HD consider the regularized version

$$\tilde Q_{T,\eta}(\tilde\theta) := u_{3:T}(\tilde\theta)'\Gamma_{T-2}(\phi_\eta^{-1})^{-1}u_{3:T}(\tilde\theta)/T \le \tilde Q_T(\tilde\theta) := u_{3:T}(\tilde\theta)'\Gamma_{T-2}(\tilde\theta)^{-1}u_{3:T}(\tilde\theta)/T.$$

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 denote constants bounding the eigenvalues of $0 < c_1 I_s \le \Sigma \le c_2 I_s$ 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 $T-2$ with T, start indexing at $t=3$ 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 C such that (see HD, p. 119, bottom display)

$$\Gamma_T(P^{-1})^{-1} = C' \begin{bmatrix} \Gamma_M^{-1} & 0 & \dots & 0 \\ 0 & \Sigma_P^{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \Sigma_P^{-1} \end{bmatrix} C.$$

Here 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:

$$C = \begin{bmatrix} I_M & & & \vdots & & \\ \dots & \dots & \dots & . & & \\ C_M & C_{M-1} & \dots & C_0 & 0 & \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots \\ & & & & & \vdots \\ 0 & \dots & C_M & C_{M-1} & \dots & C_0 \end{bmatrix}.$$

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

$$\operatorname{tr}\left(\Sigma_P^{-1}[C_M, C_{M-1}, ..., C_0]\left[T^{-1}\sum_{t=M+1}^{T}U_{t,M}(\theta)U_{t,M}(\theta)'\right][C_M, C_{M-1}, ..., C_0]'\right).$$

Therefore, we need to investigate $U_{t,M}(\theta)$ in more depth: Here

$$\begin{aligned} u(t,\theta) &= \tilde{y}_t - \tilde{D}s_t = y_t - (P_1+P_2)y_{t-1} + P_2P_1y_{t-2} - \tilde{D}s_t \\ &= (\tilde{D}_0 - \tilde{D})s_t + \mathcal{C}_\circ x_t - (P_1+P_2)\mathcal{C}_\circ x_{t-1} + P_2P_1\mathcal{C}_\circ x_{t-2} + \varepsilon_t - (P_1+P_2)\varepsilon_{t-1} + P_2P_1\varepsilon_{t-2} \\ &= (\tilde{D}_0 - \tilde{D})s_t + \varepsilon_t - (P_1+P_2)\varepsilon_{t-1} + P_2P_1\varepsilon_{t-2} \\ &\quad + \mathcal{C}_\circ((\mathcal{A}_\circ)^2x_{t-2} + \mathcal{A}_\circ\mathcal{B}_\circ\varepsilon_{t-2} + \mathcal{B}_\circ\varepsilon_{t-1}) - (P_1+P_2)\mathcal{C}_\circ(\mathcal{A}_\circ x_{t-2} + \mathcal{B}_\circ\varepsilon_{t-2}) + P_2P_1\mathcal{C}_\circ x_{t-2} \\ &= (\tilde{D}_\circ - \tilde{D})s_t + \varepsilon_t + (\mathcal{C}_\circ\mathcal{B}_\circ - (P_1+P_2))\varepsilon_{t-1} + (\mathcal{C}_\circ\mathcal{A}_\circ\mathcal{B}_\circ - (P_1+P_2)\mathcal{C}_\circ\mathcal{B}_\circ + P_2P_1)\varepsilon_{t-2} \\ &\quad + (\mathcal{C}_\circ(\mathcal{A}_\circ)^2 - (P_1+P_2)\mathcal{C}_\circ\mathcal{A}_\circ + P_2P_1\mathcal{C}_\circ)x_{t-2} \\ &= \varphi_t(\theta) + v_{t,\bullet}(\theta) + v_{t,u}(\theta). \end{aligned}$$

where

$$
\begin{aligned}
\varphi_t(\theta) &:= (\tilde{D}_\circ - \tilde{D})s_t \\
v_{t,\bullet}(\theta) &:= \varepsilon_t + (\mathcal{C}_\circ\mathcal{B}_\circ - (P_1+P_2))\varepsilon_{t-1} + (\mathcal{C}_\circ\mathcal{A}_\circ\mathcal{B}_\circ - (P_1+P_2)\mathcal{C}_\circ\mathcal{B}_\circ + P_2P_1)\varepsilon_{t-2} \\
&\quad + (\mathcal{C}_{\bullet,\circ}(\mathcal{A}_{\bullet,\circ})^2 - (P_1+P_2)\mathcal{C}_{\bullet,\circ}\mathcal{A}_{\bullet,\circ} + P_2P_1\mathcal{C}_{\bullet,\circ})x_{t-2,\bullet} \\
v_{t,u}(\theta) &:= (I_s - (P_1+P_2) + P_2P_1)\mathcal{C}_{u,\circ}x_{t-2,u} + (2I_s - P_1 - P_2)\mathcal{C}_{1,\circ}x_{t-2,2} \\
&= P_{2\perp}P_{1\perp}\mathcal{C}_{u,\circ}x_{t-2,u} + (P_{2\perp}+P_{1\perp})\mathcal{C}_{1,\circ}x_{t-2,2} \\
&= \begin{bmatrix} P_{2\perp}P_{1\perp}\mathcal{C}_{1,\circ} & P_{2\perp}P_{1\perp}\mathcal{C}_{2,\circ} + (P_{2\perp}+P_{1\perp})\mathcal{C}_{1,\circ} & P_{2\perp}P_{1\perp}\mathcal{C}_{3,\circ}\end{bmatrix} x_{t-2,u} := Px_{t-2,u}
\end{aligned}
$$

Thus, for every $\theta \in \hat{\Theta}$ the process $u_t(\theta)$ contains three components: a deterministic part $\varphi_t(\theta)$ dominated by $(\tilde{D}_\circ - \tilde{D})$, a stationary component (denoted with $v_{t,\bullet}(\theta)$ above) and the term $v_{t,u}(\theta)$ which is integrated of order two if $P_{2\perp}P_{1\perp}\mathcal{C}_{1,\circ} \neq 0$, integrated of order one if $P_{2\perp}P_{1\perp}\mathcal{C}_{1,\circ} = 0$ and $P_{2\perp}P_{1\perp}\mathcal{C}_{2,\circ} + (P_{2\perp}+P_{1\perp})\mathcal{C}_{1,\circ} \neq 0$ or $P_{2\perp}P_{1\perp}\mathcal{C}_{3,\circ} \neq 0$ and zero else. Therefore,

$$
\begin{aligned}
U_{t,M}(\theta) &= \begin{pmatrix} u(t-M,\theta) \\ u(t-M+1,\theta) \\ \vdots \\ u(t,\theta) \end{pmatrix} = \begin{pmatrix} \varphi_{t-M}(\theta) + v_{t-M,\bullet}(\theta) + Px_{t-M-2,u} \\ \varphi_{t-M+1}(\theta) + v_{t-M+1,\bullet}(\theta) + Px_{t-M-1,u} \\ \vdots \\ \varphi_t(\theta) + v_{t,\bullet}(\theta) + Px_{t-2,u} \end{pmatrix} \\
&= \begin{pmatrix} \varphi_{t-M}(\theta) \\ \varphi_{t-M+1}(\theta) \\ \vdots \\ \varphi_t(\theta) \end{pmatrix} + \begin{pmatrix} v_{t-M,\bullet}(\theta) \\ v_{t-M+1,\bullet}(\theta) + P(x_{t-M-1} - x_{t-M-2}) \\ \vdots \\ v_{t,\bullet}(\theta) + P(x_{t-2} - x_{t-M-2}) \end{pmatrix} + \begin{pmatrix} P \\ \vdots \\ P \end{pmatrix} x_{t-M-2,u} \\
&= D_{t,M}(\theta) + V_{t,M}(\theta) + \begin{pmatrix} P \\ \vdots \\ P \end{pmatrix} x_{t-M-2,u} \\
&= V_{t,M}(\theta) + D(\theta;M)s_t + P_M x_{t-M-2,u}.
\end{aligned}
$$

Define $z_{t,M} := [s_t', x_{t-M-2,u}']'$. In the following, let 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 $D_{t,M}(\theta)^\pi = 0, x_{t-M-1,1}^\pi = 0$. 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

$$
\langle V_{t,M}(\theta)^\pi, V_{t,M}(\theta)^\pi\rangle = \langle V_{t,M}(\theta), V_{t,M}(\theta)\rangle + o_P(1) \tag{C.2}
$$

as regressing out integrated processes, the constant, seasonal terms and a linear trend from stationary processes leads to negligible terms. If no deterministics are present, the negligible term is also $o(1)$.
It is now easy to verify that for $\tilde{D} = \tilde{D}_\circ$ and $P = 0$ 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 the term $Q_T(\theta)$ it holds that

$$\lim \inf_{T\to\infty} Q_T(\theta) \geq Q(\theta).$$

For $\tilde{\theta}_\circ = \Lambda(\theta_\circ)$ the additional terms due to $\tilde{D} - \tilde{D}_0$ and P are zero and, hence, $Q_T(\tilde{\theta}_\circ) \to s$. 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

$$\lim \inf_{T\to\infty} \inf_{\Theta_{c_1c_2c_3}} (Q_T(\theta) - Q_\eta(\theta)) \geq 0.$$

This follows since by taking the liminf all non-negative terms can be neglected. It is simple to verify that the convergence in (C.2) is uniform in the parameter set as $v_{t,\bullet}(\theta)$ only depends on the parameter vector via P_1 and P_2 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 $R(z) = adj(b(z))a(z) = \sum_{j=0}^r R_j z^j$ (where $\tilde{k}(z) = a^{-1}(z)b(z)$) can be bounded such that

$$\sum_{j=0}^r \|R_j\|_{Fr}^2 \leq c_3.$$

To show this, first note that

$$\lim \sup L_T(\hat{\theta}) \leq \log\det \Sigma_\circ + s \quad \text{a.s.}$$

as $L_T(\theta_\circ) \to \log\det\Sigma_\circ + s$ a.s. This holds true as the log det term converges to $\log\det\Sigma_\circ$ and for the Q_T term we have shown $\tilde{Q}_T(\tilde{\theta}_\circ) \to Q(\tilde{\theta}_\circ) = s$. The fact that $V_\pi(\tilde{\theta}_\circ) > 0$ then shows $L_T(\theta_0) \to \log\det\Sigma_\circ + s$.
This implies $\log|\hat{\Sigma}| \leq \log|\Sigma_\circ| + s$ 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

$$\begin{aligned}\tilde{Q}_T(\theta) &= u_T(\theta)'\Gamma_T(\theta)^{-1}u_T(\theta)/T \\ &\geq \text{tr}\left[\Sigma^{-1}[R_r, R_{r-1}, ..., R_0]\left[T^{-1}\sum_{t=r+1}^T U_{t,r}(\theta)U_{t,r}(\theta)'\right][R_r, R_{r-1}, ..., R_0]'\right]\gamma^{-2}.\end{aligned}$$

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, as it is related to $\langle V_t(\theta)^\pi, V_t(\theta)^\pi\rangle$ whose main component is

$$\begin{aligned}v_{t,\bullet}(\theta) := \varepsilon_t &+ (\mathcal{C}_\circ\mathcal{B}_\circ - (P_1+P_2))\varepsilon_{t-1} + (\mathcal{C}_\circ\mathcal{A}_\circ\mathcal{B}_\circ - (P_1+P_2)\mathcal{C}_\circ\mathcal{B}_\circ + P_2P_1)\varepsilon_{t-2} \\ &+ (\mathcal{C}_{\bullet,\circ}(\mathcal{A}_{\bullet,\circ})^2 - (P_1+P_2)\mathcal{C}_{\bullet,\circ}\mathcal{A}_{\bullet,\circ} + P_2P_1\mathcal{C}_{\bullet,\circ})x_{t-2,\bullet},\end{aligned}$$

where $x_{t-r-2,u}$ and s_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 $\lambda_j(\hat{\Sigma})$)

$$\operatorname{tr}(\hat{\Sigma}^{-1}) = \sum_{j=1}^{s} \frac{1}{\lambda_j(\hat{\Sigma})} \leq (\log\det\Sigma_\circ + s)\gamma^2/c < \infty$$

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 $\log|\hat{\Sigma}| \leq \log|\Sigma_\circ| + s$ 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

$$u_T(\hat{\theta})'\Gamma_T(\hat{\theta})^{-1}u_T(\hat{\theta})/T \geq \operatorname{tr}\left[\hat{\Sigma}^{-1}\left(\sum_{j=0}^{r}\hat{R}_j\hat{R}_j'\right)\right] c\gamma^{-2} \geq \operatorname{tr}\left[\sum_{j=0}^{r}\hat{R}_j\hat{R}_j'\right] c\gamma^{-2}/c_2.$$

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, 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)

$$\begin{aligned}\lim\inf_{T\to\infty} L_T(\hat{\theta}_T) &\geq \lim\inf_{T\to\infty}(\log|\hat{\Sigma}_T| + \tilde{Q}_T(\hat{\theta}_T)) \geq \sup_{\eta>0}\lim\inf_{T\to\infty}(\log|\hat{\Sigma}_T| + \tilde{Q}_{T,\eta}(\hat{\theta}_T)) \quad \text{a.s.} \\ &\geq \inf_{\theta\in\Theta_{c_1c_2c_3}}\left(\log|\Sigma| + \sup_{\eta>0} Q_\eta(\theta)\right) = \log|\Sigma_\circ| + s.\end{aligned}$$

Then as in HD, p. 125, for every sequence $\hat{\theta}_T \to \theta$ (choosing a subsequence if necessary) it holds that

$$\lim\inf_{T\to\infty} L_T(\theta_\circ) \geq \lim\inf_{T\to\infty} L_T(\hat{\theta}_T) \geq \log|\Sigma_\circ| + s.$$

Then $L_T(\hat{\theta}_T) \to L(\theta) = L(\theta_\circ) = \log|\Sigma_\circ| + s$ follows. Recall that θ is of the form (k, P_1, P_2, D, Σ). This shows that $\Sigma = \Sigma_\circ$ and $\tilde{k} = \tilde{k}_\circ$ is the unique limit, since $L(\theta)$ depends on θ only via Σ and $\tilde{k}$ but not on P_1,P_2 or D. Convergence of P_1, P_2 and D can be shown as in Lemma 4 of de Matos Ribeiro et al. (2020). An analogous approach proving the results presented in Theorem 10 will be discussed in subsection C.2.2. Note that from the above results $L_T(\hat{\theta}_T) \to L(\theta_\circ)$ and, hence, for large enough T it holds that $L_T(\hat{\theta}_T) \leq c$ a.s. for a constant $c < \log|\Sigma_\circ| + s + \epsilon$ for every $\epsilon > 0$.

The prediction error criterion function

For the prediction error criterion function

$$L_{PE,T}(k(z), \Sigma, d, e; Y_T) = \log|\Sigma| + u_T(\theta)'\Gamma_{T,PE}(\theta)^{-1}u_T(\theta)/T$$

it follows that

$$\Gamma_{T,PE}(\tilde{\theta}) = \underbrace{\begin{bmatrix} I_s & 0 & \dots & 0 \\ K_1 & I_s & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ K_{T-1} & \dots & K_1 & I_s \end{bmatrix}}_{\mathcal{T}_T(\theta)} (I_T \otimes \Sigma) \begin{bmatrix} I_s & 0 & \dots & 0 \\ K_1 & I_s & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ K_{T-1} & \dots & K_1 & I_s \end{bmatrix}'.$$

Consequently, $|\Gamma_{T,PE}(\tilde{\theta})| = |\Sigma|^T$ in this case such that $T^{-1}\log|\Gamma_T(\tilde{\theta})| = \log|\Sigma|$ holds. Note that this does not depend on the poles of $\tilde{k}$ and, hence, the first part of the criterion function does not diverge to infinity at the stability boundary of $\hat{\Theta}$.
Following the proof above, note that the log det term in this case is simpler such that we can directly obtain the result $\log|\Gamma_{T,PE}(\tilde{\theta})|/T \to \log|\Sigma|$.
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 $\Theta_{c_1c_2c_3}$ for large enough T. The result then follows from

$$L_{PE,T}(k_\circ(z),\sigma_\circ,d_\circ,e_\circ;Y_T) \to \log|\Sigma_\circ| + s,$$

see Lemma 9 for a proof.

Super-consistency for certain parts of the parameter vector

Let $\Pi(k) := -k^{-1}(1)$ and $\Gamma(k) := -k^{-1}(1) + \frac{\partial}{\partial z}k^{-1}(z)|_{z=1}$ for a transfer function $k(z)$. Defining $\tilde{k}$ through $k^{-1}(z) = \tilde{k}(z)(1-z)^2 - \Pi(k)z - \Gamma(k)z(1-z)$, we find the representation

$$\begin{aligned}\varepsilon_t(k,d,e) &= \tilde{k}(L)\boldsymbol{\Delta}^2(y_t - d - et) - \Pi(k)(y_{t-1} - d - e(t-1)) \\ &\quad - \Gamma(k)\boldsymbol{\Delta}(y_{t-1} - d - e(t-1)) \\ &= -\Pi(k)\mathcal{C}_{1,\circ}x_{t-1,1} - \Pi(k)\mathcal{C}_{3,\circ}x_{t-1,3} - \big(\Pi(k)\mathcal{C}_{2,\circ} + \Gamma(k)\mathcal{C}_{1,\circ}\big)x_{t-2,2} \\ &\quad - \Pi(k)(d_\circ - d) - \Gamma(k)(e_\circ - e) - \Pi(k)(e_\circ - e)(t-1) + \upsilon_t(k,d,e),\end{aligned}$$

where

$$\begin{aligned}\upsilon_t(k,d,e) :&= \tilde{k}(L)\boldsymbol{\Delta}^2[\mathcal{C}_{u,\circ}x_{t,u} + \mathcal{C}_{\bullet,\circ}x_{t,\bullet} + \varepsilon_t] - \Pi(k)[\mathcal{C}_{2,\circ}\mathcal{B}_{2,\circ}\varepsilon_{t-1} + \mathcal{C}_{\bullet,\circ}x_{t,\bullet} + \varepsilon_t] \\ &\quad - \Gamma(k)[\mathcal{C}_{1,\circ}\mathcal{B}_{1,\circ}\varepsilon_{t-1} + \mathcal{C}_{2,\circ}\mathcal{B}_{2,\circ}\varepsilon_{t-1} + \mathcal{C}_{3,\circ}\mathcal{B}_{3,\circ}\varepsilon_{t-1} + \boldsymbol{\Delta}(\mathcal{C}_{\bullet,\circ}x_{t,\bullet} + \varepsilon_t)] \\ &\quad - \textstyle\sum_{m=t}^{\infty}\underline{K}_m(d - e(t-m)).\end{aligned}$$

By the strict minimum-phase assumption $k^{-1}(z)$ and, thus, also $\tilde{k}(z)$ are uniformly stable and since ε_t, $x_{t,\bullet}$ are (asymptotically) stationary by definition, it follows that $\upsilon_t(k,d,e)$ is asymptotically stationary. The proof for the super-consistency results listed in Theorem 10 now proceeds as in Lemma 4 of de Matos Ribeiro et al. (2020). For this define $w_t := [x'_{t-1,1}, x'_{t-2,2}, x'_{t-1,3}, 1, t-1]'$ and

$$\begin{aligned}\Psi := \Psi(k,d,e) :&= \begin{bmatrix}\Psi_u(k) & \Psi_d(d,e)\end{bmatrix} \\ \Psi_u(k) :&= \begin{bmatrix}\Pi(k)\mathcal{C}_{1,\circ} & \Pi(k)\mathcal{C}_{2,\circ} + \Gamma(k)\mathcal{C}_{1,\circ} & \Pi(k)\mathcal{C}_{3,\circ}\end{bmatrix} \\ \Psi_d(d,e) :&= \begin{bmatrix}(\Pi(k)(d_\circ - d) + \Gamma(k)(e_\circ - e) & \Pi(k)(e_\circ - e)\end{bmatrix}\end{aligned}$$

and let $\hat{\Psi} := \Psi(\hat{k},\hat{d},\hat{e})$. The consistency proof implies

$$\left\langle \varepsilon_t(\hat{k},\hat{d},\hat{e}), \varepsilon_t(\hat{k},\hat{d},\hat{e})\right\rangle = \left\langle \upsilon_t(\hat{k},\hat{d},\hat{e}) - \hat{\Psi}w_t, \upsilon_t(\hat{k},\hat{d},\hat{e}) - \hat{\Psi}w_t\right\rangle < c.$$

Consider $D_T^w = \mathrm{diag}(T^{-3/2}I_{c_1}, T^{-1/2}I_{c_1+c_2}, 1, T^{-1})$ such that $D_T^w\langle w_t, w_t\rangle D_T^w$ converges to a positive definite matrix with probability one. The above implies

$$\left\langle \hat{\Psi}w_t, \hat{\Psi}w_t\right\rangle - \left\langle \hat{\Psi}w_t, \upsilon_t(\hat{k},\hat{d},\hat{e})\right\rangle \left\langle \upsilon_t(\hat{k},\hat{d},\hat{e}), \upsilon_t(\hat{k},\hat{d},\hat{e})\right\rangle^{-1} \left\langle \upsilon_t(\hat{k},\hat{d},\hat{e}), \hat{\Psi}w_t\right\rangle < c,$$

from which $\langle \hat{\Psi}w_t, \hat{\Psi}w_t\rangle < \tilde{c}$ follows since $\upsilon_t(\hat{k},\hat{d},\hat{e})$ is asymptotically stationary such that in a regression of w_t on $\upsilon_t(\hat{k},\hat{d},\hat{e})$ the variance of the fitted values tends to zero. Therefore, $\langle \hat{\Psi}w_t, \hat{\Psi}w_t\rangle <$

$\tilde{c}$ implies $\|\hat{\Psi}(D_T^w)^{-1}\| = O_p(1)$. Assume $\|\langle \upsilon_t(\hat{k},\hat{d},\hat{e}), w_t\rangle (D_T^w)\| = o_p(\|\hat{\Psi}(D_T^w)^{-1}\|)$ and, consequently, $\|\langle \upsilon_t(\hat{k},\hat{d},\hat{e}), \hat{\Psi} w_t\rangle\| = o_p(\|\langle \hat{\Psi} w_t, \hat{\Psi} w_t\rangle\|)$, which implies

$$\begin{aligned}
&\Big\langle \varepsilon_t(\hat{k},\hat{d},\hat{e}), \varepsilon_t(\hat{k},\hat{d},\hat{e})\Big\rangle \\
&\quad = \Big\langle \upsilon_t(\hat{k},\hat{d},\hat{e}), \upsilon_t(\hat{k},\hat{d},\hat{e})\Big\rangle - \Big\langle \upsilon_t(\hat{k},\hat{d},\hat{e}), \hat{\Psi} w_t\Big\rangle - \Big\langle \upsilon_t(\hat{k},\hat{d},\hat{e}), \hat{\Psi} w_t\Big\rangle + \Big\langle \hat{\Psi} w_t, \hat{\Psi} w_t\Big\rangle \\
&\quad = \Big\langle \upsilon_t(\hat{k},\hat{d},\hat{e}), \upsilon_t(\hat{k},\hat{d},\hat{e})\Big\rangle + \Big\langle \hat{\Psi} w_t, \hat{\Psi} w_t\Big\rangle - o_p(\|\langle \hat{\Psi} w_t, \hat{\Psi} w_t\rangle\|) \\
&\quad > \Big\langle \upsilon_t(\hat{k},\hat{d},\hat{e}), \upsilon_t(\hat{k},\hat{d},\hat{e})\Big\rangle
\end{aligned}$$

for large T. This is a contradiction to the optimality of the PML estimator, since in this case the choice $\Psi = 0$ would lead to a higher likelihood value. Thus, $\langle \hat{\Psi} w_t, \hat{\Psi} w_t\rangle = O_p(\langle \upsilon_t(\hat{k},\hat{d},\hat{e}), \hat{\Psi} w_t\rangle)$, which implies for the different components

$$\begin{aligned}
\Big\langle \upsilon_t(\hat{k},\hat{d},\hat{e}), \Pi(\hat{k})\mathcal{C}_{1,\circ} x_{t-1,1}\Big\rangle \quad & = O_p(\Big\langle \Pi(\hat{k})\mathcal{C}_{1,\circ} x_{t-1,1}, \Pi(\hat{k})\mathcal{C}_{1,\circ} x_{t-1,1}\Big\rangle) \\
& \Rightarrow T^\gamma \|\Pi(\hat{k})\mathcal{C}_{1,\circ}\| = o_p(1) \text{ for } \gamma < 2, \\
\Big\langle \upsilon_t(\hat{k},\hat{d},\hat{e}), (\Pi(\hat{k})\mathcal{C}_{2,\circ} + \Gamma(\hat{k})\mathcal{C}_{1,\circ}) x_{t-2,2}\Big\rangle \quad & = O_p(\Big\langle (\Pi(\hat{k})\mathcal{C}_{2,\circ} + \Gamma(\hat{k})\mathcal{C}_{1,\circ}) x_{t-2,2}, \Pi(\hat{k})\mathcal{C}_{2,\circ} x_{t-2,2}\Big\rangle \\
& \qquad + O_p(\Big\langle (\Pi(\hat{k})\mathcal{C}_{2,\circ} + \Gamma(\hat{k})\mathcal{C}_{1,\circ}) x_{t-2,2}, \Gamma(\hat{k})\mathcal{C}_{1,\circ} x_{t-2,2}\Big\rangle) \\
& \Rightarrow T^\gamma \|(\Pi(\hat{k})\mathcal{C}_{2,\circ} + \Gamma(\hat{k})\mathcal{C}_{1,\circ})\| = o_p(1) \text{ for } \gamma < 1, \\
\Big\langle \upsilon_t(\hat{k},\hat{d},\hat{e}), \Pi(\hat{k})\mathcal{C}_{3,\circ} x_{t-1,3}\Big\rangle \quad & = O_p(\Big\langle \Pi(\hat{k})\mathcal{C}_{3,\circ} x_{t-1,3}, \Pi(\hat{k})\mathcal{C}_{3,\circ} x_{t-1,3}\Big\rangle) \\
& \Rightarrow T^\gamma \|\Pi(\hat{k})\mathcal{C}_{3,\circ}\| = o_p(1) \text{ for } \gamma < 1, \\
\Big\langle \upsilon_t(\hat{k},\hat{d},\hat{e}), \Pi(\hat{k})(d_\circ - \hat{d}) + \Gamma(\hat{k})(e_\circ - \hat{e})\Big\rangle \quad & = O_p(\Big\langle \Pi(\hat{k})(d_\circ - \hat{d}) + \Gamma(\hat{k})(e_\circ - \hat{e}), \Pi(\hat{k})(d_\circ - \hat{d})\Big\rangle \\
& \qquad + O_p(\Big\langle \Pi(\hat{k})(d_\circ - \hat{d}) + \Gamma(\hat{k})(e_\circ - \hat{e}), \Gamma(\hat{k})(e_\circ - \hat{e})\Big\rangle) \\
& \Rightarrow T^\gamma \|\Pi(\hat{k})(d_\circ - \hat{d}) + \Gamma(\hat{k})(e_\circ - \hat{e})\| = o_p(1) \\
& \qquad \text{for } \gamma < 1/2, \\
\Big\langle \upsilon_t(\hat{k},\hat{d},\hat{e}), \Pi(\hat{k})(e_\circ - \hat{e})(t-1)\Big\rangle \quad & = O_p(\Big\langle \Pi(\hat{k})(e_\circ - \hat{e})(t-1), \Pi(\hat{k})(e_\circ - \hat{e})(t-1)\Big\rangle) \\
& \Rightarrow T^\gamma \|\Pi(\hat{k})(e_\circ - \hat{e})\| = o_p(1) \text{ for } \gamma < 3/2.
\end{aligned}$$

Alternatively, it holds that

$$\hat{\Psi}\mathrm{diag}(T^2 I_{c_1}, T I_{c_1+c_2}, T^{1/2}, T^{3/2}) T^{-\epsilon} = o_p(1)$$

for every $\epsilon > 0$, proving the results given in Theorem 10.

C.2.3 Derivation of the Asymptotic Distribution

Let us repeat at this point, that a specific parameterization is necessary in order to derive the asymptotic distribution, whereas the consistency proof has been independent of the chosen parameterization. Furthermore, for the applicability of linearization techniques, we assume that the parameters are introduced in such a way that the true parameter vector is defined as

$$\varphi_\circ := [\ \theta'_{u,\circ} \quad \theta'_{d,1,\circ} \quad \theta'_{d,2,\circ} \quad \theta'_{st,\circ}\ ]',$$

where $\theta_{u,\circ}$ collects the parameter vector corresponding to $\mathcal{C}_u = [\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3]$ by the relations

$$\begin{aligned}
\mathcal{C}_1(\theta_{E,1}) : &= R_{1,L}(\theta_{1,L})' \begin{bmatrix} I_{c_1} \\ 0_{(s-c_1)\times c_1} \end{bmatrix} R_{1,R}(\theta_{1,R}), \\
\mathcal{C}_3(\theta_{E,2}; \mathcal{C}_1(\theta_{E,1})) : &= R_{1,L}(\theta_{1,L})' \begin{bmatrix} 0_{c_1\times c_2} \\ R_{L,2}(\theta_{L,2})' \begin{bmatrix} I_{c_2} \\ 0_{(s-c_1-c_2)\times c_2} \end{bmatrix} \end{bmatrix} R_{2,R}(\theta_{2,R}), \\
\mathcal{C}_2(\theta_G; \mathcal{C}_E(\theta_E)) : &= R_{1,L}(\theta_{1,L})' \begin{bmatrix} 0_{c_1\times c_1} \\ R_{2,L}(\theta_{2,L})' \begin{bmatrix} 0_{c_2\times c_1} \\ \Lambda(\theta_G) \end{bmatrix} \end{bmatrix} R_{1,R}(\theta_{1,R}),
\end{aligned}$$

where $R_{1,L}(\theta_{1,L}), R_{2,L}(\theta_{2,L}) \in \mathbb{R}^{s\times s}$ and $R_{1,R}(\theta_{1,R}) \in \mathbb{R}^{c_1\times c_1}$, $R_{2,R}(\theta_{2,R}) \in \mathbb{R}^{c_1\times c_1}$ are orthonormal matrices and products of Givens rotations and $\Lambda(\theta_G) \in \mathbb{R}^{s-c_1-c_2\times c_1}$. For more details on the parameterization see Bauer et al. (2020). Note that we slightly adapted the factorization of $\mathcal{C}_2$ presented there by additionally multiplying $R_{1,R}(\theta_{1,R})$ from the right, in order to simplify some calculations.
By $\theta_{d,1,\circ}$ and $\theta_{d,2,\circ}$ denote the parameters for the deterministic terms (if it is included). Further,

$$\begin{aligned}\theta_{st,\circ} &:= [\ \theta'_{1,R,\circ} \quad \theta'_{2,R,\circ} \quad \theta'_{B,f,\circ} \quad \theta'_{B,p,\circ} \quad \theta'_{\bullet,\circ}\ ]' \\ \theta_{u,\circ} &:= [\ \theta'_{1,L,\circ} \quad \theta'_{2,L,\circ} \quad \theta'_{G,\circ}\ ]' \\ \theta_{\star,\circ} &:= [\ \theta'_{u,\circ} \quad \theta'_{d,1,\circ} \quad \theta'_{d,2,\circ}\ ]'\end{aligned}$$

Then in the theorem it is assumed that $\varphi_\circ$ is an interior point of the parameter set. This requires that the multi-index $\mathbf{\Gamma}$ is specified correctly.
Let the corresponding parameter estimator minimizing the scaled negative pseudo likelihood function L_T be denoted as $\hat{\varphi}$. Lemma 12 shows that the difference between the minima of L_T and the prediction error function $L_{PE,T}$ is negligible for the asymptotic distribution. Thus, we use the prediction error form which is easier to investigate. Furthermore, we concentrate out the parameters for Σ such that the criterion function equals

$$\mathcal{L}_{PE,T}(\varphi; Y_T) = \log\det\langle \varepsilon_t(\varphi), \varepsilon_t(\varphi)\rangle$$

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 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 $\|\underline{K}_j(\varphi_\circ)\| \le \mu_K \rho_0^j$ 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 $\varphi_\circ$.
Define the scaling matrices

$$\begin{aligned}\tilde{D}_T &:= \operatorname{diag}(T^{-1}I_{c_{u,1}}, I_{c_u - c_{u,1}}, T^{1/2}I_{c_{d,2}}, T^{-1/2}I_{c_{d,1}}, T^{1/2}I_{c_{st}}), \\ \tilde{D}_T^M &:= \tilde{D}_T\tilde{M}, \quad \tilde{M} := \operatorname{diag}(M, I_{c_u - c_1(s-c_1) + c_{d,2} + c_{d,1} + c_{st}}), \\ D_T^M &:= T\tilde{D}_T^{-1}\tilde{M}.\end{aligned}$$

The matrix $M \in \mathbb{R}^{c_1(s-c_1)\times c_1(s-c_1)}$ separates components with different orders of convergence within the parameter vector $\theta_{1,L}$. Let $\mathcal{C}_\perp \in \mathbb{R}^{s\times s-c_1-c_2}$ denote a non-singular matrix whose columns are orthogonal to the column space of $[\mathcal{C}_1, \mathcal{C}_3]$. In order for all $c_1(s-c_1)$ components of $\theta_{1,L}$ to be estimated with rate T^2, $T^\gamma\|[\hat{\mathcal{C}}_3, \hat{\mathcal{C}}_\perp]'\mathcal{C}_{1,\circ}]\| \to 0$ would need to hold in probability for all $0 < \gamma < 2$. By the results of Theorem 10 this is only the case if $c_2 = 0$, i. e., if there are no I(1)-common trends other than $\mathcal{C}_{2,\circ}x_{t,2}$. Consequently, the presence of I(1)-common components $\mathcal{C}_{3,\circ}x_{t,3}$ reduces the convergence rate in some components of $\theta_{1,L}$.
The derivation of the asymptotic distribution given in Theorem 11 proceeds in three steps:

1. Show that $\tilde{D}_T^M\partial\mathcal{L}_T(\varphi_\circ; Y_T)$ converges in distribution.
2. Show that $\tilde{D}_T^M\partial^2\mathcal{L}_T(\bar{\varphi}_T; Y_T)(D_T^M)^{-1}$ 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 $d, 2d$ we denote differentiation with respect to the entries in $\theta_{d,2d}$, with $d, 2e$ differentiation with respect to the entries in $\theta_{d,2e}$ and with $d, 1$ differentiation with respect to the entries in $\theta_{d,1}$. Furthermore, we 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.

Lemma 13 *Define $\tilde{k}(z, \theta)$ through $k^{-1}(z; \theta) = \tilde{k}(z, \theta)(1 - z)^2 - \Pi(\theta)z - \Gamma(\theta)z(1 - z)$ with $\Pi(\theta) := -k^{-1}(1; \theta)$ and $\Gamma(\theta) := -k^{-1}(1; \theta) + \frac{\partial}{\partial z} k^{-1}(z; \theta)|_{z=1}$. The derivatives of $\varepsilon_t(\varphi)$ with respect to the different parameters in θ_{st} evaluated at $\varphi_\circ$ are given by*

- $\partial_{i,\mathcal{A}_\bullet} \varepsilon_t(\varphi) = dk_{i,A_\bullet}(L, \theta) x_{t,\bullet}(\theta)$, *where* $dk_{i,\mathcal{A}_\bullet}(z, \theta) = -\mathcal{C}z(I - z\underline{\mathcal{A}})^{-1} \begin{bmatrix} 0_{c\times(n-c)} \\ \partial_i(\mathcal{A}_\bullet) \end{bmatrix}$,
- $\partial_{i,\mathcal{B}} \varepsilon_t(\varphi) = dk_{i,\mathcal{B}_\bullet}(L, \theta)\varepsilon_t(\varphi)$, *where* $dk_{i,\mathcal{B}}(z, \theta) = -\mathcal{C}z(I - z\underline{\mathcal{A}})^{-1}\partial_i\mathcal{B}$,
- $\partial_{i,\mathcal{C}_\bullet} \varepsilon_t(\varphi) = k^{-1}(L; \theta)(-\partial_i \mathcal{C}_\bullet)x_{t,\bullet}(\theta)$.
- $\partial_{i,R_{1,R}} \varepsilon_t(\varphi) = dk_{i,R_{1,R}}(L, \theta)\varepsilon_t(\varphi)$
  $= -\Pi(\theta)\partial_{i,R_{1,R}} \mathcal{C}_2\mathcal{B}_2\varepsilon_{t-1} - \Gamma(\theta)\partial_{i,R_{1,R}} \mathcal{C}_u\mathcal{B}_u\varepsilon_{t-2} - \tilde{k}(L; \theta)(1 - L)^2\partial_{i,R_{1,R}} \mathcal{C}_u x_{t,u}(\theta)$.
- $\partial_{i,R_{2,R}} \varepsilon_t(\varphi) = dk_{i,R_{2,R}}(L, \theta)\varepsilon_t(\varphi)$
  $= -\Gamma(\theta)\partial_{i,R_{2,R}} \mathcal{C}_3\mathcal{B}_3\varepsilon_{t-2} - \tilde{k}(L; \theta)(1 - L)^2\partial_{i,R_{2,R}} \mathcal{C}_u x_{t,u}(\theta)$.

All the above processes are asymptotically stationary.
The derivatives of $\varepsilon_t(\varphi)$ with respect to parameters in θ_u are given by

- $\partial_{i,R_{1,L}} \varepsilon_t(\varphi) = k^{-1}(L; \theta)(-\partial_{i,R_{1,L}} \mathcal{C}_u)x_{t,u}(\theta)$.
- $\partial_{i,R_{2,L}} \varepsilon_t(\varphi) = k^{-1}(L; \theta)(-\partial_{i,R_{2,L}} \mathcal{C}_u)x_{t,u}(\theta)$.
- $\partial_{i,\Lambda} \varepsilon_t(\varphi) = k^{-1}(L; \theta)(-\partial_{i,\Lambda} \mathcal{C}_u)x_{t,u}(\theta)$.

Finally the derivatives of $\varepsilon_t(\varphi)$ with respect to parameters in θ_d are given by

- $\partial_{i,D} \varepsilon_t(\varphi) = k^{-1}(L; \theta)(-\partial_{i,D} D_d(\theta_d))s_t(\theta)$.

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 $\mathcal{C}_u$. The partial derivatives are:

$$\begin{aligned} \partial_i \varepsilon_t(\varphi) &= -(\partial_i \mathcal{C}_u)x_{t,u}(\theta) - \mathcal{C}(\partial_i x_t(\theta)) \\ \partial_i x_{t+1}(\theta) &= \underline{\mathcal{A}}\partial_i x_t(\theta) - \mathcal{B}(\partial_i \mathcal{C}_u)x_{t,u}(\theta). \end{aligned}$$

These components of the score are filtered versions of $x_{t,k}(\theta)$. A possibly non-minimal representation of the filter is given by $dk_i(z, \theta) = -\partial_i\mathcal{C}_u + z\mathcal{C}(I - z\underline{\mathcal{A}})^{-1}\mathcal{B}\partial_i\mathcal{C}_u = k^{-1}(z; \theta)(-\partial_i\mathcal{C}_u)$. It holds that

$$\begin{aligned} \partial_{i,R_{1,R}} \mathcal{C}_u &= \begin{bmatrix} \mathcal{C}_1 R'_{1,R} \left(\partial_{i,R_{1,R}} R_{1,R}\right) & \mathcal{C}_2 R'_{1,R} \left(\partial_{i,R_{1,R}} R_{1,R}\right) & 0 \end{bmatrix}, \\ \partial_{i,R_{2,R}} \mathcal{C}_u &= \begin{bmatrix} 0 & 0 & \mathcal{C}_3 R'_{2,R} \left(\partial_{i,R_{2,R}} R_{2,R}\right) \end{bmatrix}, \\ \partial_{i,R_{1,L}} \mathcal{C}_u &= \left(\partial_{i,R_{1,L}} R_{1,L}\right) R'_{1,L} \begin{bmatrix} \mathcal{C}_1 & \mathcal{C}_2 & \mathcal{C}_3 \end{bmatrix}, \\ \partial_{i,R_{2,L}} \mathcal{C}_u &= \begin{bmatrix} 0 & 0 & \partial_{i,R_{2,L}} \mathcal{C}_3 \end{bmatrix}, \\ \partial_{i,\Lambda} \mathcal{C}_u &= \begin{bmatrix} 0 & \partial_{i,\Lambda} \mathcal{C}_2 & 0 \end{bmatrix}, \end{aligned}$$

again omitting the dependencies on the different components of $\theta_{E,1}$, $\theta_{E,2}$ and θ_G for ease of notation. Decomposing

$$k^{-1}(z;\theta) = \tilde{k}(z,\theta)(1-z)^2 - \Pi(\theta)z - \Gamma(\theta)z(1-z),$$

the derivatives with respect to different components of $\mathcal{C}_u$ can be further analyzed. It follows that the derivatives with respect to $\theta_{i,R}$ can be further simplified:
$\theta_{1,R}$:

$$\begin{aligned} k^{-1}(z;\theta)(-\partial_{i,R_{1,R}}\mathcal{C}_u x_{t,u}(\theta)) = &-\left(\Pi(\theta)\mathcal{C}_2 + \Gamma(\theta)\mathcal{C}_1\right) R_{1,R}'\left(\partial_{i,R_{1,R}} R_{1,R}\right) x_{t-1,2}(\theta) \\ &- \Pi(\theta)\partial_{i,R_{1,R}}\mathcal{C}_2\mathcal{B}_2\varepsilon_{t-1} - \Gamma(\theta)(\partial_{i,R_{1,R}}\mathcal{C}_1\mathcal{B}_1 + \partial_{i,R_{1,R}}\mathcal{C}_2\mathcal{B}_2)\varepsilon_{t-2} \\ &- \tilde{k}(L;\theta)(1-L)^2\partial_{i,R_{1,R}}\mathcal{C}_u x_{t,u}(\theta) \end{aligned}$$

where the first term is equal to zero.
$\theta_{2,R}$:

$$k^{-1}(z;\theta)(-\partial_{i,R_{2,R}}\mathcal{C}_u x_{t,u}(\theta)) = -\Gamma(\theta)\partial_{i,R_{2,R}}\mathcal{C}_3\mathcal{B}_3\varepsilon_{t-2} - \tilde{k}(L;\theta)(1-L)^2\partial_{i,R_{2,R}}\mathcal{C}_u x_{t,u}(\theta).$$

Note that both $dk_{i,R_{1,R}}(z,\theta)$ and $dk_{i,R_{2,R}}(z,\theta)$ are stable transfer functions.
If the derivative is taken with respect to $\theta_{1,L}$, $\theta_{2,L}$ or θ_G the terms $\Pi\partial_{i,R_{1,L}}\mathcal{C}_1$, $\Pi\partial_{i,R_{2,L}}\mathcal{C}_3$ and $\Pi\partial_{i,\Lambda}\mathcal{C}_2$ do not vanish, such that $\partial_{i,R_{1,L}}\varepsilon_t(\varphi)$ and $\partial_{i,R_{1,L}}\varepsilon_t(\varphi)$ contain integrated components of order one or two.
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 14 *Let the assumptions of Theorem 11 hold. Then the following statements hold true:*

- *For the derivative with respect to θ_{st},*

$$T^{1/2}\partial_{st}\mathcal{L}_T(\varphi_\circ) \xrightarrow{d} \mathcal{N}(0, V_{st}),$$

where V_{st} denotes the asymptotic variance matrix.

- *Let $\Pi_\circ := \Pi(\theta_\circ)$ and $\Gamma_\circ := \Gamma(\theta_\circ)$ and the scaling factor $D_{T,u,i}$, such that $D_{T,u,i} = 1$ if $\Pi_\circ(\partial_i\mathcal{C}_1) = 0$ and $D_{T,u,i} = 1/T$ else. Then*

$$D_{T,u,i}\left(\partial_u\mathcal{L}_T(\varphi_\circ)\right)_i \xrightarrow{d} \begin{cases} 2tr\left[-\Sigma_\circ^{-1}\Pi_\circ\partial_i\mathcal{C}_1\boldsymbol{X}_1\right] & \text{if } D_{T,u,i} \neq 1, \\ 2tr\left[-\Sigma_\circ^{-1}\partial_i\Psi_u\boldsymbol{X}_u\right] & \text{if } D_{T,u,i} = 1, \end{cases}$$

where

$$\partial_i\Psi_u := \begin{bmatrix}\Pi_\circ\partial_i\mathcal{C}_1 & (\Pi_\circ\partial_i\mathcal{C}_2 + \Gamma_\circ\partial_i\mathcal{C}_1) & \Pi_\circ\partial_i\mathcal{C}_3\end{bmatrix}, \quad \boldsymbol{X}_u := [\boldsymbol{X}_1', \boldsymbol{X}_2', \boldsymbol{X}_3']',$$

with $\partial_i\mathcal{C}_j := \partial_{u,i}\mathcal{C}_{j,\circ}$, $j = 1,2,3$, and the limits

$$\begin{aligned} \boldsymbol{X}_1 &:= \int_0^1 \left(\int_0^u \mathcal{B}_{2,\circ}\boldsymbol{W}(v)dv\right) d\boldsymbol{W}(u)', \\ \boldsymbol{X}_2 &:= \int_0^1 \mathcal{B}_{2,\circ}\boldsymbol{W}(u)d\boldsymbol{W}(u)', \quad X_3 := \int_0^1 \mathcal{B}_{3,\circ}\boldsymbol{W}(u)d\boldsymbol{W}(u)', \end{aligned}$$

where $\boldsymbol{W}$ is the weak limit of $T^{-1/2}\sum_{t=1}^{[Tu]}\varepsilon_t$.

- *For $\theta_{d,2}$ one obtains with respect to $\theta_{d,2d} = \beta_\circ' d$*

$$T^{1/2}\partial_{d,2d}\mathcal{L}_T(\varphi_\circ) \xrightarrow{d} -2\alpha_\circ'\Sigma_\circ^{-1}\boldsymbol{W}(1) =: \mathrm{v}_{d,2,d},$$

and for derivatives with respect to $\theta_{d,2e} = \mathcal{C}_{3,\circ}' e$

$$T^{1/2}\partial_{d,2e}\mathcal{L}_T(\varphi_\circ) \xrightarrow{d} -2(\Gamma_\circ\mathcal{C}_{3,\circ})'\Sigma_\circ^{-1}\boldsymbol{W}(1) =: \mathrm{v}_{d,2,e}.$$

- *For $\theta_{d,1} = \beta_\circ' e$ it holds that*

$$T^{-1/2}\partial_{d,1}\mathcal{L}_T(\varphi_\circ) \xrightarrow{d} -2\alpha_{1,\circ}'\Sigma_\circ^{-1}\boldsymbol{U}(1) =: \mathrm{v}_{e\perp},$$

where $\alpha_\circ := \Pi_\circ[\tilde{\mathcal{C}}_{2,\circ}\mathcal{C}_{\perp,\circ}]$ *and* $\boldsymbol{U}(1) := \int_0^1 ud\boldsymbol{W}(u)$.

- *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:

$$\begin{aligned}\partial_i\mathcal{L}_T(\varphi_\circ) = \partial_i\left(\log\det\langle\varepsilon_t(\varphi_\circ),\varepsilon_t(\varphi_\circ)\rangle\right) &= tr[\langle\varepsilon_t(\varphi_\circ),\varepsilon_t(\varphi_\circ)\rangle^{-1}\,2\langle\partial_i\varepsilon_t(\varphi_\circ),\varepsilon_t(\varphi_\circ)\rangle] \\ &= tr[\Sigma_\circ^{-1}2\langle\partial_i\varepsilon_t(\varphi_\circ),\varepsilon_t(\varphi_\circ)\rangle] + o_P(1).\end{aligned}$$

Let us start with the coordinates of $\theta_{st} = \begin{bmatrix} \theta_{1,R}' & \theta_{2,R}' & \theta_{B,f}' & \theta_{B,p}' & \theta_\bullet' \end{bmatrix}'$. For every component of θ_{st} Lemma 13 above establishes asymptotic stationarity. Asymptotic normality for $T^{1/2}\langle\partial_i\varepsilon_t(\varphi_\circ),\varepsilon_t(\varphi_\circ)\rangle$ 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 results hold jointly in all coordinates of θ_{st}.
The representation in Lemma 13 allows for the application of Lemma 10 and Lemma 11 to obtain

$$\begin{aligned}D_{T,u,i}\partial_{u,i}\mathcal{L}_T(\varphi_\circ) &= D_{T,u,i}\mathrm{tr}[\Sigma_\circ^{-1}2\langle k^{-1}(L;\theta_\circ)(-\partial_{u,i}\mathcal{C}_{u,\circ})x_{t,u}(\theta_\circ),\varepsilon_t\rangle] + o_P(1) \\ &= D_{T,u,i}\mathrm{tr}[\Sigma_\circ^{-1}2\langle\Pi_\circ\partial_i\mathcal{C}_1x_{t-1,1} + (\Pi_\circ\partial_i\mathcal{C}_2 + \Gamma_\circ\partial_i\mathcal{C}_1)x_{t-2,2} + \Pi_\circ\partial_i\mathcal{C}_3x_{t-1,3},\varepsilon_t\rangle] \\ &\quad + o_P(1) \\ &= D_{T,u,i}\mathrm{tr}[\Sigma_\circ^{-1}2\langle\partial_i\Psi_u x_{t-2,u},\varepsilon_t\rangle] + o_P(1) \\ &\xrightarrow{d} \begin{cases} 2\mathrm{tr}\left[-\Sigma_\circ^{-1}\Pi_\circ\partial_i\mathcal{C}_1\boldsymbol{X}_1\right], & \text{if } D_{T,u,i}\neq 1 \\ 2\mathrm{tr}\left[-\Sigma_\circ^{-1}((\Pi_\circ\partial_i\mathcal{C}_2 + \Gamma_\circ\partial_i\mathcal{C}_1)\boldsymbol{X}_2 + \Pi_\circ\partial_i\mathcal{C}_3\boldsymbol{X}_3)\right] & \text{if } D_{T,u,i} = 1.\end{cases}\end{aligned}$$

The next step is to derive the asymptotic distribution of the score components corresponding to θ_d. With respect to components of $\beta_\circ' d$ it holds that

$$\partial_{d,2d}\mathcal{L}_T(\varphi_\circ) = -2T^{-1}\textstyle\sum_{t=1}^T(k^{-1}(L;\theta_\circ)\beta_\circ)'\Sigma_\circ^{-1}\varepsilon_t + o_P(T^{-1/2}).$$

Note that

$$k^{-1}(L;\theta_\circ)\beta_\circ = (\alpha_\circ\beta_\circ')\beta_\circ - \mathcal{C}_\circ\underline{\mathcal{A}}_\circ^t(I-\underline{\mathcal{A}}_\circ)^{-1}\mathcal{B}_\circ\beta_\circ,$$

where the second component exhibits exponential decay. Therefore, it follows that

$$\begin{aligned}T^{1/2}\partial_{d,2d}\mathcal{L}_T(\varphi_\circ) &= -\alpha_\circ'\Sigma_\circ^{-1}2T^{-1/2}\textstyle\sum_{t=1}^T\varepsilon_t + o_P(T^{-1/2}) \\ &\xrightarrow{d} -\alpha_\circ'\Sigma_\circ^{-1}\boldsymbol{W}(1).\end{aligned}$$

For the derivatives with respect to $\mathcal{C}_{3,\circ}e$ and $\beta_\circ' e$ it holds that

$$\begin{aligned}\partial_{d,2e}\mathcal{L}_T(\varphi_\circ) &= -2T^{-1}\textstyle\sum_{t=1}^T(k^{-1}(L;\theta_\circ)\mathcal{C}_{3,\circ}t)'\Sigma_\circ^{-1}\varepsilon_t + o_P(T^{-1/2}), \\ \partial_{d,1}\mathcal{L}_T(\varphi_\circ) &= -2T^{-1}\textstyle\sum_{t=1}^T(k^{-1}(L;\theta_\circ)\beta_\circ t)'\Sigma_\circ^{-1}\varepsilon_t + o_P(T^{-1/2}).\end{aligned}$$

Using

$$k^{-1}(L;\theta_\circ)t = \textstyle\sum_{j=0}^{t-1}\underline{K}_j(t-j) = (\sum_{j=0}^{t-1}\underline{K}_j)t - \sum_{j=0}^{t-1}\underline{K}_j j = \Pi_\circ t + \Gamma_\circ + o_P(\rho^t),$$

we get for the derivatives with respect to $\beta_\circ' e$

$$\begin{aligned}T^{-1/2}\partial_{d,1}\mathcal{L}_T(\varphi_\circ) &= -2\alpha_\circ'\Sigma_\circ^{-1}T^{-3/2}\textstyle\sum_{t=1}^T t\varepsilon_t + o_P(T^{-1/2}) \\ &\xrightarrow{d} -2\alpha_\circ'\Sigma_\circ^{-1}\boldsymbol{U}(1),\end{aligned}$$

where we have used Lemma 10 (xi) for the convergence of the first summand. In the directions of $\mathcal{C}_{3,\circ}$ the first term vanishes because $\Pi_{1,\circ}\mathcal{C}_{3,\circ} = 0$. Consequently, we get

$$\begin{aligned} T^{1/2}\partial_e \mathcal{L}_T(\varphi_\circ) &= -2(\Gamma_\circ \mathcal{C}_{3,\circ})'\Sigma_\circ^{-1} T^{-1/2} \textstyle\sum_{t=1}^T \varepsilon_t + o_P(T^{-1/2}) \\ &\stackrel{d}{\to} -2(\Gamma_\circ \mathcal{C}_{3,\circ})'\Sigma_\circ^{-1}\boldsymbol{W}(1). \end{aligned}$$

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 14 in the discussion we have to distinguish with respect to which parameter components θ_u, θ_{st}, $\theta_{d,2}$ and $\theta_{d,1}$ 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 15 *Under the conditions of Theorem 11 one obtains* $\tilde{D}_T \partial^2 \mathcal{L}_T(\bar{\varphi}_T) D_T^{-1} \stackrel{d}{\to} Z$ *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 $\varphi = [\ \theta_u' \quad \theta_{d,2}' \quad \theta_{d,1}' \quad \theta_{st}'\ ]'$*. Then*

$$\tilde{D}_T^M \partial^2 L_T(\varphi_T)(D_T^M)^{-1} \quad \stackrel{d}{\to} \quad \begin{bmatrix} Z_\star & 0 \\ 0 & Z_{st} \end{bmatrix} = \begin{bmatrix} Z_u & Y_{u,2d}' & Y_{u,2e}' & Y_{u,1}' & 0 \\ Y_{u,2d} & Z_{d,2d} & Y_{2d,2e}' & Y_{2d,1}' & 0 \\ Y_{u,2e} & Y_{2d,2e} & Z_{d,2e} & Y_{2e,1}' & 0 \\ Y_{u,1} & Y_{2d,1} & Y_{2e,1} & Z_{d,1} & 0 \\ 0 & 0 & 0 & 0 & Z_{st} \end{bmatrix}$$

For typical indices i,j *(not the same for all the expressions below) the respective entries are of the form:*

$$[Z_u]_{i,j} \quad = \quad \begin{cases} 2tr\left[(\partial_j^M \Psi_{u,1})'\Sigma_\circ^{-1}\partial_i^M \Psi_{u,1} \boldsymbol{Z}_{1,1}\right], & \text{if } i,j \le c_{u,1} \\ 2tr\left[(\partial_j^M \Psi_{u})'\Sigma_\circ^{-1}\partial_i^M \Psi_{u,1} \boldsymbol{Z}_{1}\right], & \text{if } i \le c_{u,1}, j > c_{u,1} \\ 2tr\left[(\partial_j^M \Psi_{u,1})'\Sigma_\circ^{-1}\partial_i^M \Psi_{u} \boldsymbol{Z}_{1}'\right], & \text{if } i > c_{u,1}, j \le c_{u,1} \\ 2tr\left[(\partial_j^M \Psi_{u})'\Sigma_\circ^{-1}\partial_i^M \Psi_{u} \boldsymbol{Z}\right], & \text{if } i,j > c_{u,1} \end{cases}$$

where $\partial_j^M \Psi_u := \sum_{k=1}^{c_u} M_{jk}\partial_k \Psi_u$*, for an orthogonal matrix* $M \in \mathbb{R}^{c_u \times c_u}$*, where the first* $c_{u,1} = c_1(s - c_1 - c_2)$ *rows of* M *are such that the matrices* $\partial_j^M \Psi_{u,1} := \sum_{k=1}^{c_u} M_{jk}\partial_k \Psi_u [I_{c_1}, 0_{c_1 \times (c_1+c_2)}]'$ *are linearly independent, and* $\partial_j^M \Psi_{u,1} = 0$ *for all* $j > c_{u,1}$*, and where* $\boldsymbol{Z}_{1,1}$*,* $\boldsymbol{Z}_1$ *and* $\boldsymbol{Z}$ *are defined through*

$$T^{-3}\langle x_{t,1}, x_{t,1}\rangle \stackrel{d}{\to} \boldsymbol{Z}_{1,1}, \quad T^{-3/2}\langle x_{t,1}, D_T^u x_{t,u}\rangle \stackrel{d}{\to} \boldsymbol{Z}_1, \quad \langle D_T^u x_{t,u}, D_T^u x_{t,u}\rangle \stackrel{d}{\to} \boldsymbol{Z},$$

with $D_T^u := diag(T^{-3/2} I_{c_1}, T^{-1/2} I_{c_1+c_2})$*. For the blocks corresponding to deterministic components* *we have*

$$\begin{aligned}
[Y_{u,2d}]_{i,j} &= -2e_i'(\Pi_\circ \mathcal{C}_{\perp,\circ})'\Sigma_\circ^{-1}\partial_j^M \Psi_u \boldsymbol{Y}_2,\\
[Y_{u,2e}]_{i,j} &= -2e_i'(\Gamma_\circ \mathcal{C}_{3,\circ})'\Sigma_\circ^{-1}\partial_j^M \Psi_u \boldsymbol{Y}_2,\\
[Y_{u,1}]_{i,j} &= -2e_i'\alpha_{1,\circ}'\Sigma_\circ^{-1}\partial_j^M \Psi_u \boldsymbol{Y}_1,\\
Z_{2d} &= 2\alpha_\circ'\Sigma_\circ^{-1}\alpha_\circ.\\
Z_{2e} &= 2(\Gamma_\circ \mathcal{C}_{3,\circ})'\Sigma_\circ^{-1}\Gamma_\circ \mathcal{C}_{3,\circ},\\
Z_1 &= \frac{2}{3}\alpha_\circ'\Sigma_\circ^{-1}\alpha_\circ,\\
Y_{1,2e} &= \alpha_\circ'\Sigma_\circ^{-1}\Gamma_\circ \mathcal{C}_{3,\circ},\\
Y_{1,2d} &= \alpha_\circ'\Sigma_\circ^{-1}\alpha_\circ,\\
Y_{2e,2d} &= 2(\Gamma_\circ \mathcal{C}_{3,\circ})'\Sigma_\circ^{-1}\alpha_\circ,
\end{aligned}$$

where $\boldsymbol{Y}_2$ *is the limit of* $\langle D_T^u x_{t,u}, 1\rangle$ *and* $\boldsymbol{Y}_1$ *is the limit of* $T^{-1}\langle D_T^u x_{t,u}, t\rangle$.
It follows that $Z_u - Y_D' Z_D^{-1} Y_D$ *with* $Y_D = [\ Y_{u,2d}' \quad Y_{u,2e}' \quad Y_{u,1}'\ ]'$ *and*

$$Z_D := \begin{bmatrix} Z_{d,2d} & Y_{2d,2e}' & Y_{2d,1}' \\ Y_{2d,2e} & Z_{d,2e} & Y_{2e,1}' \\ Y_{2d,1} & Y_{2e,1} & Z_{d,1} \end{bmatrix}$$

has the same structure as Z_u*, where in the expression* $\Sigma_\circ^{-1/2}\partial_i^M\Psi_u$ *has to be replaced by*

$$\partial_i^M \Psi_{u,2\Sigma} := \left[P_{\Sigma^{-1/2}\alpha}\Sigma_\circ^{-1/2}(\Pi_\circ\partial_i^M\mathcal{C}_2 + \Gamma_\circ\partial_i^M\mathcal{C}_1) \quad P_{\Sigma^{1/2}\alpha_\perp}\Sigma_\circ^{-1/2}\Gamma_\circ\partial_i^M\mathcal{C}_1) \quad \Sigma_\circ^{-1/2}\Pi_\circ\partial_i^M\mathcal{C}_3\right]$$

where $P_{\Sigma^{-1/2}\alpha} := \Sigma_\circ^{-1/2}\alpha_\circ(\alpha_\circ'\Sigma_\circ^{-1}\alpha_\circ)^{-1}\alpha_\circ'\Sigma_\circ^{-1/2}$ *and* $P_{\Sigma^{1/2}\alpha_\perp} := I_s - P_{\Sigma^{-1/2}\alpha}$ *and* $\boldsymbol{Z}_{1,1}$, $\boldsymbol{Z}_1$ *and* $\boldsymbol{Z}$ *have to be replaced by*

$$\begin{aligned}
\boldsymbol{Z}_{1,1}^d &:= \textstyle\int_0^1 (\int_0^u \mathcal{B}_2 \boldsymbol{W}(v)dv)|_1(\int_0^u \mathcal{B}_2\boldsymbol{W}(v)dv)'du,\\
\boldsymbol{Z}_1^d &:= \textstyle\int_0^1 (\int_0^u \mathcal{B}_2 \boldsymbol{W}(v)dv)|_1 \boldsymbol{F}_d' du \quad \boldsymbol{F}_d := [(\mathcal{B}_2\boldsymbol{W}(u)|_1)', (\mathcal{B}_2\boldsymbol{W}(u))', (\mathcal{B}_3\boldsymbol{W}(u)|_1)']',\\
\boldsymbol{Z}^d &:= \textstyle\int_0^1 \boldsymbol{F}_d\boldsymbol{F}_d' du
\end{aligned}$$

if there is no linear trend term in the model. If a linear trend is present in the model, $\boldsymbol{Z}_{1,1}$, $\boldsymbol{Z}_1$ *and* $\boldsymbol{Z}$ *have to be replaced by*

$$\begin{aligned}
\boldsymbol{Z}_{1,1}^{de} &:= \textstyle\int_0^1 (\int_0^u \mathcal{B}_2 \boldsymbol{W}(v)dv)|_{1,u}(\int_0^u \mathcal{B}_2\boldsymbol{W}(v)dv)'du,\\
\boldsymbol{Z}_1^{de} &:= \textstyle\int_0^1 (\int_0^u \mathcal{B}_2 \boldsymbol{W}(v)dv)|_{1,u} \boldsymbol{F}_{de}' du \quad \boldsymbol{F}_{de} := [(\mathcal{B}_2\boldsymbol{W}(u)|_{1,u})', (\mathcal{B}_2\boldsymbol{W}(u)|_1)', (\mathcal{B}_3\boldsymbol{W}(u)|_{1,u})']',\\
\boldsymbol{Z}^{de} &:= \textstyle\int_0^1 \boldsymbol{F}_{de}\boldsymbol{F}_{de}' du.
\end{aligned}$$

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:

$$\begin{aligned}
\partial_{i,j}^2 \mathcal{L}_T(\bar\varphi_T) &= \partial_i \left(\mathrm{tr}[\langle\bar\varepsilon_t,\bar\varepsilon_t\rangle^{-1} 2\langle\partial_j\bar\varepsilon_t,\bar\varepsilon_t\rangle]\right)\\
&= -\mathrm{tr}\left[\langle\bar\varepsilon_t,\bar\varepsilon_t\rangle^{-1}(\langle\partial_i\bar\varepsilon_t,\bar\varepsilon_t\rangle + \langle\bar\varepsilon_t,\partial_i\bar\varepsilon_t\rangle)\langle\bar\varepsilon_t,\bar\varepsilon_t\rangle^{-1} 2\langle\partial_j\bar\varepsilon_t,\bar\varepsilon_t\rangle\right]\\
&\quad + \mathrm{tr}\left[\langle\bar\varepsilon_t,\bar\varepsilon_t\rangle^{-1} 2\left\langle(\partial_{i,j}^2\bar\varepsilon_t),\bar\varepsilon_t\right\rangle\right] + \mathrm{tr}\left[\langle\bar\varepsilon_t,\bar\varepsilon_t\rangle^{-1} 2\langle(\partial_j\bar\varepsilon_t),(\partial_i\bar\varepsilon_t)\rangle\right]\\
&= -\mathrm{tr}\left[\Sigma_\circ^{-1}(\langle\partial_i\bar\varepsilon_t,\bar\varepsilon_t\rangle + \langle\bar\varepsilon_t,\partial_i\bar\varepsilon_t\rangle)\Sigma_\circ^{-1}2\langle\partial_j\bar\varepsilon_t,\bar\varepsilon_t\rangle\right]\\
&\quad + \mathrm{tr}\left[\Sigma_\circ^{-1}2\left\langle(\partial_{i,j}^2\bar\varepsilon_t),\bar\varepsilon_t\right\rangle\right] + \mathrm{tr}\left[\Sigma_\circ^{-1}2\langle(\partial_j\bar\varepsilon_t),(\partial_i\bar\varepsilon_t)\rangle\right] + o_P(1), \qquad \text{(C.3)}
\end{aligned}$$

for every sequence $\bar{\varphi} \to \varphi_\circ$, using the short notation $\bar{\varepsilon}_t := \varepsilon_t(\bar{\varphi}_T), \partial_i\bar{\varepsilon}_t := \partial_i\varepsilon_t(\bar{\varphi}_T), \partial^2_{i,j}\bar{\varepsilon}_t := \partial^2_{i,j}\varepsilon_t(\bar{\varphi}_T)$. This follows from the USE condition, see Lemma 9 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 10 matrix blocks have to be dealt with (taking into account symmetry of the Hessian). The blocks are partitioned according to how often differentiation takes place with respect to a component of θ_u, $\theta_{d,2}$, $\theta_{d,1}$ 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 $\theta_{d,1}$ an additional scaling factor T^{-1} is introduced and for each derivative with respect to an entry in θ_u an additional scaling factor of $T^{-1/2}$ or $T^{-3/2}$ is introduced, which results in the proper scaling factor for each of the terms to obtain convergence in distribution.

In the above expression (C.3) 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 10 and 11 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 $\theta_{d,1}$) 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 (C.3). Hence, we obtain:

$$\partial^2_{i,j}\mathcal{L}_T(\bar{\varphi}_T) = \text{tr}\left[\Sigma_\circ^{-1}2\langle(\partial_i\varepsilon_t(\bar{\varphi}_T)), (\partial_j\varepsilon_t(\bar{\varphi}_T))\rangle\right] + o_P(T^{-N_u/2-N_{u,i}-N_{e_\perp}}), \tag{C.4}$$

where N_u counts the number of times differentiation takes place with respect to an element of θ_u, $N_{u,1}$ counts the number of times differentiation takes place with respect to an element θ_1^M as defined in Theorem 11 and $N_{d,1}$ counts the number of times differentiation takes place with respect to an element of $\theta_{d,1}$. Recall the definition of D_T:

$$\begin{aligned}
\tilde{D}_T &:= \text{diag}(T^{-1}I_{c_{u,1}}, I_{c_u-c_{u,1}}, T^{1/2}I_{c_{d,2}}, T^{-1/2}I_{c_{d,1}}, T^{1/2}I_{c_{st}}) \\
\tilde{D}_T^M &:= \tilde{D}_T\tilde{M} \quad \tilde{M} := \text{diag}(M, I_{c_u-c_{u,1}+c_{d,2}+c_{d,1}+c_{st}}) \\
D_T^M &:= T\tilde{D}_T^{-1}\tilde{M},
\end{aligned}$$

where the matrix $M \in \mathbb{R}^{c_u \times c_u}$ separates the I(1) and I(2) components. It follows that

$$\tilde{D}_T^M\partial^2\mathcal{L}_T(\bar{\varphi}_T; Y_T)(D_T^M)^{-1} = T^{-1}\tilde{D}_T\left(\tilde{M}\,\partial^2\mathcal{L}_T(\bar{\varphi}_T; Y_T)\,\tilde{M}'\right)\tilde{D}_T.$$

Note that left and right multiplication of the Hessian by $\tilde{M}$ and $\tilde{M}^{-1}$ corresponds to a change from the partial derivatives with respect to φ to a basis of certain directional derivatives, for which we use the notation ∂^M in the following. Note, however, that only the partial derivatives with respect to θ_u are affected by the change of the basis. Starting from equation (C.4) we now 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 $\varphi_\circ$.

$i \sim \theta_{st}, j \sim \theta_\star$: If differentiation takes place once with respect to an entry of θ_{st} and once with respect to an entry of $\theta_\star$, convergence of $(\tilde{D}_T^M \partial^2 \mathcal{L}_T(\bar{\varphi}_T; Y_T)(D_T^M)^{-1})_{i,j} \to 0$ follows. In this case define

$$\begin{aligned}
\partial_j^M \Psi &:= \begin{bmatrix} \partial_j^M \Psi_u & \partial_j^M \Psi_d \end{bmatrix}, \\
\partial_j^M \Psi_d &:= \begin{bmatrix} \left(-\alpha_\circ \partial_j^M \theta_{d,2d} - \Gamma_\circ [\mathcal{C}_{3,\circ}, \beta_\circ]' \partial_j^M [\theta_{d,2e}', \theta_{d,1}']' \right) & -\alpha_\circ \partial_j^M \theta_{d,1} \end{bmatrix} \\
\partial_j^M \Psi &:= \begin{cases} \sum_{k=1}^{c_1(s-c_1)} M_{jk} \partial_k \Psi & \text{if } j \leq c_{u,1} \\ \partial_j \Psi & \text{if } j > c_{u,1} \end{cases}
\end{aligned}$$

such that the dominant component in $(\tilde{D}_T^M \partial^2 \mathcal{L}_T(\bar{\varphi}_T; Y_T)(D_T^M)^{-1})_{i,j}$ is equal to

$$T^{-1}(\tilde{D}_T)_{i,i}(\tilde{D}_T)_{j,j} \text{tr}\left[\Sigma_\circ^{-1} 2 \langle \partial_i \varepsilon_t(\varphi_\circ), \partial_j^M \Psi w_t \rangle\right].$$

The process $\partial_i \varepsilon_t(\varphi_\circ)$ is asymptotically stationary, while, e. g., for directional derivatives $j \leq c_{u,1}$ the process $\partial_j^M \Psi w_t$ contains I(2) components. In this case $T^{-1}(\tilde{D}_T)_{i,i}(\tilde{D}_T)_{j,j} = T^{-2}$, such that $T^{-2}\text{tr}\left[\Sigma_\circ^{-1} 2 \langle \partial_i \varepsilon_t(\varphi_\circ), \partial_j^M \Psi w_t \rangle\right] \to 0$ according to Lemma 11.
$i \sim \theta_\star, j \sim \theta_\star$: Finally, examine the cases of differentiating twice with respect to an entry in $\theta_\star$. The dominant component in $(\tilde{D}_T^M \partial^2 \mathcal{L}_T(\bar{\varphi}_T; Y_T)(D_T^M)^{-1})_{i,j}$ is equal to

$$T^{-1}(\tilde{D}_T)_{i,i}(\tilde{D}_T)_{j,j} \text{tr}\left[\Sigma_\circ^{-1} 2 \langle \partial_j^M \Psi w_t, \partial_j^M \Psi w_t \rangle\right].$$

Consider again the directional derivatives $i, j \leq c_{u,1}$ corresponding to I(2) processes. Since $T^{-1}(\tilde{D}_T)_{i,i}(\tilde{D}_T)_{j,j} = T^{-3}$, Lemma 10 implies

$$T^{-3}\text{tr}\left[\Sigma_\circ^{-1} 2 \langle \partial_j^M \Psi w_t, \partial_j^M \Psi w_t \rangle\right] \to 2\text{tr}\left[(\partial_j^M \Psi_{u,1})' \Sigma_\circ^{-1} \partial_i^M \Psi_{u,1} \boldsymbol{Z}_{1,1}\right].$$

The other terms follow similarly by application of Lemma 10.
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 $\theta_{C,R}$, $\theta_{B,f}$ and $\theta_{B,p}$ have to be analyzed. The proof is indirect: If the sub-block of Z_{st} corresponding to $\theta_{C,R}$, $\theta_{B,f}$ and $\theta_{B,p}$ were singular, there would exist a vector $x = [\begin{array}{ccc} x_1 & \dots & x_v \end{array}]'$ such that

$$0 = \sum_{r,s=1}^{v} x_r x_s \text{tr}[\Sigma_\circ^{-1} \mathbb{E} \partial_s \varepsilon_t(\varphi_\circ) \partial_r \varepsilon_t(\varphi_\circ)'] = \text{tr}\left[\Sigma_\circ^{-1} \mathbb{E} \sum_{r=1}^{v} x_r \partial_r \varepsilon_t(\varphi_\circ) \left(\sum_{s=1}^{v} x_s \partial_s \varepsilon_t(\varphi_\circ) \right)' \right]$$

denoting the components of θ_{st} corresponding to $\theta_{C,R}$, $\theta_{B,f}$ 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 $K_{j,u} = j\mathcal{C}_1 \mathcal{B}_2 + \mathcal{C}_1 \mathcal{B}_1 + \mathcal{C}_2 \mathcal{B}_2 + \mathcal{C}_3 \mathcal{B}_3$. Thus, linear dependence of the derivatives with respect to θ_{st} implies

$$\sum_r x_r \partial_r [\mathcal{C}_1[\mathcal{B}_2, \mathcal{B}_1], \mathcal{C}_2 \mathcal{B}_2, \mathcal{C}_3 \mathcal{B}_3] = 0 \tag{C.5}$$

since both the I(2) and (1) components need to vanish separately, and $\mathcal{C}_k,\ k = 1,2,3$, span different spaces such that no linear dependance between matrices $\partial_r \mathcal{C}_k \mathcal{B}_k$ for different k occur. In the following we show that (C.5) implies $x_r = 0$.
We start with the derivatives with respect to $\theta_{2,R}$ and $\theta_{B_3} = [\theta_{B_3,p}, \theta'_{B_3,f}]'$ and consider the product $\mathcal{C}_3\mathcal{B}_3$ or $R_{2,R}\mathcal{B}_3$. We show the independence by induction over c_2. In the case $c_2 = 1$, there are s parameters in $\theta_{2,R}$ and θ_{B_3}, and all parameters correspond to the entries of $\mathcal{B}_3$ while $R_{2,R} = 1$, see Bauer et al. (2020, Section 3.1) for details. The coefficients x_r of a linear combination of the derivatives must be zero, in order for the sum to be zero, showing that the derivatives are linearly independent.
Suppose the linear independence has been shown for $c_2 = g$. To show the statement for $c_2 = g+1$, write

$$R_{2,R}(\theta_{2,R}) = \prod_{i=1}^{c_2-1}\prod_{j=1}^{c_2-i} R_{c_2,i,i+j}(\theta_{i(i-1)/2+j}) = \prod_{j=1}^{c_2-1} R_{c_2,1,1+j}(\theta_j) \prod_{i=2}^{c_2-1}\prod_{j=1}^{c_2-i} R_{c_2,i,i+j}(\theta_{i(i-1)/2+j}),$$

where $R_{c_2,i,i+j}(\theta_{i(i-1)/2+j})$ is a real Givens rotation, see Bauer et al. (2020, Definition 6). Note that this corresponds to the form of R_R in Bauer et al. (2020, Lemma 1) up to a reordering, which simplifies the proof. Clearly, the entries in the first column of $R_{2,R}\mathcal{B}_3$ with non-zero entries only depend on the parameters $\theta_1 \dots, \theta_{c_2-1}$ and the first non-zero entry in the first row of $\mathcal{B}_3$. Since the first columns of the derivatives with respect to these parameters are orthogonal to each other, the coefficients of these derivatives in (C.5) must be zero. Since the matrix $\prod_{j=1}^{c_2-1} R_{c_2,1,1+j}(\theta_j)$ is of full rank, the derivatives of

$$\prod_{j=1}^{c_2-1} R_{c_2,1,1+j}(\theta_j) \prod_{i=2}^{c_2-1}\prod_{j=1}^{c_2-i} R_{c_2,i,i+j}(\theta_{i(i-1)/2+j})\mathcal{B}_3$$

with respect to the other parameters are independent if and only if the derivatives of

$$\prod_{i=2}^{c_2-1}\prod_{j=1}^{c_2-i} R_{c_2,i,i+j}(\theta_{i(i-1)/2+j})\mathcal{B}_3 = \begin{bmatrix} 1 & 0 \\ 0 & \prod_{i=2}^{c_2-1}\prod_{j=1}^{c_2-i} R_{c_2-1,i-1,i+j-1}(\theta_{i(i-1)/2+j}) \end{bmatrix}\mathcal{B}_3$$

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, the coefficient of the derivative with respect to b_{12} must be zero. By an analogous argument the coefficients corresponding to $b_{13}, \dots, b_{1s}$ are zero. 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 $(c_2 - 1) \times (s - 1)$ block. This follows from the induction hypothesis, which finishes the proof.
An analogous argument for the product $\mathcal{C}_1[\mathcal{B}_1, \mathcal{B}_2]$ respectively $R_{1,R}[\mathcal{B}_1, \mathcal{B}_2]$ shows the linear independence of derivatives with respect to $\theta_{1,R}$, $\theta_{B,p}$ and $\theta_{B,f}$. Combining both, the linear independence of all derivatives with respect to θ_{st} follows.
Next consider 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 (C.3) is

$$\begin{aligned}(\tilde{D}_T^M \partial^2 \mathcal{L}_T(\bar{\varphi}_T; Y_T)(D_T^M)^{-1})_{i,j} &= T^{-1}(\tilde{D}_T)_{i,i}(\tilde{D}_T)_{j,j}\mathrm{tr}\left[\Sigma_\circ^{-1} 2\langle \partial_j^M \Psi_u x_{t,u}, \partial_j^M \Psi_u x_{t,u}\rangle\right] + o_P(1) \\ &= \mathrm{tr}\left[\partial_j^M \tilde{\Psi}'_u \Sigma_\circ^{-1} 2 \partial_j^M \tilde{\Psi}_u D_T^u \langle x_{t,u}, x_{t,u}\rangle D_T^u\right] + o_P(1),\end{aligned}$$

where $\partial_j^M \tilde{\Psi} := [\partial_j^M \tilde{\Psi}_{u,1}, 0_{s\times(c_1+c_2)}]$ for $j \le c_{u,1}$ and $\partial_j^M \tilde{\Psi} = \partial_j^M \Psi$ for $j > c_{u,1}$. Recall that $D_T^u = \mathrm{diag}(T^{-3/2} I_{c_1}, T^{-1/2} I_{c_1+c_2})$. This can be further rewritten as

$$\begin{aligned}&\tilde{D}_T^M \partial^2 \mathcal{L}_T(\bar{\varphi}_T; Y_T)(D_T^M)^{-1} = \Phi'_T \Phi_T + o_P(1), \\ &\Phi_T := [\ \phi_1, \ \ \dots, \ \ \phi_{c_u}\ ] \quad \phi_j := \mathrm{vec}\left(\Sigma_\circ^{-1/2} \partial_j^M \tilde{\Psi}(T^{-1} D_T^u \langle x_{t,u}, x_{t,u}\rangle D_T^u)^{-1/2}\right).\end{aligned}$$

Note that Z_u is invertible if and only if the limit of $\Phi_T'\Phi_T$ is invertible, which holds if and only if Φ_T is of full column rank, at least asymptotically. Since $T^{-1}D_T^u\langle x_{t,u}, x_{t,u}\rangle D_T^u$ is invertible in probability and converges to the random matrix Z, which is also invertible in probability, it is enough to show that $\psi_j := \text{vec}\left(\partial_j^M\tilde{\Psi}\right)$ are linearly independent, due to the identity

$$\phi_j = \left((D_T^u\langle x_{t,u}, x_{t,u}\rangle D_T^u)^{-1/2} \otimes \Sigma_\circ^{-1/2}\right)\text{vec}(\psi_j).$$

Note first that ψ_j, $j \le c_{u,1}$, is orthogonal to ψ_j, $j > c_{u,1}$, thus, separating the proof of linear independence into two parts.
For $j = 1, \ldots, c_{u,1}$ the linear independence follows from the fact that the derivatives of the product $(\mathcal{C}_{1,\circ})_\perp(\mathcal{C}_{1,\circ})_\perp' R_{1,L}(\theta_{1,L})$ with respect to $\theta_{1,L}$ are $c_1(s-c_1)$ linearly independent matrices. Thus, there exists a orthogonal matrix $M \in \mathbb{R}^{c_u \times c_u}$ of linear combinations, transforming the set of derivatives into two different sets, such that the set $\sum_{k=1}^{c_u} M_{j,k}(\beta_\circ\beta_\circ')\partial_k\mathcal{C}_1$, $j \le c_{u,1}$ is a basis for the set of all matrices $N \in \mathbb{R}^{s\times c_1}$ satisfying $[\mathcal{C}_{1\circ}, \mathcal{C}_{3\circ}]'N = 0$, and the set of $\sum_{k=1}^{c_u} M_{j,k}(\mathcal{C}_{3\circ}\mathcal{C}_{3\circ}')\partial_k\mathcal{C}_1$, $c_{u,1} < j \le c_u$ is a basis for the set of all matrices $N \in \mathbb{R}^{s\times c_1}$ satisfying $[\mathcal{C}_{1\circ}, \beta_\circ]'N = 0$. Multiplication with $\Pi_\circ$ from the left then proves the linear independence of the vectors ψ_j, $j \le c_{u,1}$.
For $j = 1, \ldots, c_{u,1}$, consider the submatrix $[\ \Pi_\circ\partial_i^M\mathcal{C}_2 + \Gamma_\circ\partial_i^M\mathcal{C}_1 \quad \Pi_\circ\partial_i^M\mathcal{C}_3\ ]$. Chose $[\partial_i^M\mathcal{C}_2, \partial_i^M\mathcal{C}_3] = [\partial_i\mathcal{C}_2, \partial_i\mathcal{C}_3]$ for the derivatives with respect to θ_G and $\theta_{2,L}$. Thus, for derivatives with respect to θ_G the above submatrix is equal to $[\ \Pi_\circ\partial_i\mathcal{C}_2 \quad 0\ ]$, while for derivatives with respect to $\theta_{2,L}$ the submatrix is equal to $[\ 0 \quad \Pi_\circ\partial_i\mathcal{C}_3\ ]$. Since the columns of $\partial_i\mathcal{C}_2$ are in the column space of $\beta_\circ$ by construction, it follows that the matrices $[\ \Pi_\circ\partial_i\mathcal{C}_2 \quad 0\ ]$ form a basis for the set of all matrices $[N, 0_{s\times c_2}]$, $N \in \mathbb{R}^{s\times c_1}$ satisfying $[\mathcal{C}_{1\circ}, \mathcal{C}_{3\circ}]'N = 0$. Similarly, the matrices $[\ 0 \quad \Pi_\circ\partial_i\mathcal{C}_3\ ]$ form a basis for the set of all matrices $[0_{s\times c_1}, N]$, $N \in \mathbb{R}^{s\times c_2}$ satisfying $[\mathcal{C}_{1\circ}, \mathcal{C}_{3\circ}]'N = 0$. This is due to the properties of the parameterization based on Givens rotation, ensuring by construction that $(\beta_\circ\beta_\circ')\partial_i\mathcal{C}_3$ are linearly independent, since the columns of $\partial_i\mathcal{C}_3$ are orthogonal to $\mathcal{C}_{1,\circ}$ and cannot lie in the span of $\mathcal{C}_{3,\circ}$. Note that these two sets of linearly independent matrices can be now used to reduce the set of $[\ \Pi_\circ\partial_i^M\mathcal{C}_2 + \Gamma_\circ\partial_i^M\mathcal{C}_1 \quad \Pi_\circ\partial_i^M\mathcal{C}_3\ ]$ into $[\ (\alpha_{\perp,\circ}\alpha_{\perp,\circ}'\Gamma_\circ\partial_i^M\mathcal{C}_1 \quad 0\ ]$, by regressing out the matrices corresponding to derivatives with respect to θ_G and $\theta_{2,L}$. Thus, as the last step, consider the set of matices $[\ (\alpha_{\perp,\circ}\alpha_{\perp,\circ}'\Gamma_\circ\partial_j^M\mathcal{C}_1 \quad 0\ ]$ for $c_{u,1} < j \le s(s-c_1)$ corresponding to derivatives with respect to $\theta_{1,L}$. Using $\alpha_{\perp,\circ}'\Gamma_\circ = \xi_\circ\eta_\circ' = \xi_\circ\mathcal{C}_3'$ and the fact that $\sum_{k=1}^{c_u} M_{j,k}(\mathcal{C}_{3\circ}\mathcal{C}_{3\circ}')\partial_k\mathcal{C}_1$, $c_{u,1} < j \le c_u$ is a basis for the set of all matrices $N \in \mathbb{R}^{s\times c_1}$ satisfying $[\mathcal{C}_{1\circ}, \beta_\circ]'N = 0$, it follows that the reduced vectors are linearly independent, which implies that the set of vectors $\psi_j := \text{vec}\left(\partial_j^M\tilde{\Psi}\right)$, $c_{u,1} < j \le c_u$ are jointly linearly independent.
In consequence, Φ_T is of full column rank, such that $\Phi_T'\Phi_T$ is invertible.
Finally, let us consider $Z_u - Y_D'Z_D^{-1}Y_D$. For $t = 1, \ldots, T$, $i = 1, \ldots, c_u$ define

$$\tau_{u,i,t} := -\Sigma_\circ^{-1/2}\partial_i^M\Psi_u' D_T^u x_{t,u}(\theta_\circ) \qquad \tau_{u,t} := [\ \tau_{u,1,t} \quad \ldots \quad \tau_{u,c_u,t}\ ].$$

We see that

$$\begin{aligned}\text{diag}(T^{-1}I_{c_{u,1}}, I_{c_u-c_{u,1}})M_u\partial_u^2\mathcal{L}_T(\bar{\varphi}_T)M_u^{-1}\text{diag}(T^{-2}I_{c_{u,1}}, T^{-1}I_{c_u-c_{u,1}})\\ = \sum_{t=1}^T \tau_{u,t}'\tau_{u,t} + o_P(1) \to Z_u,\end{aligned}$$

with $M_u := \text{diag}(M, I_{c_u - c_1(s-c_1)})$. Analogously, define for $t = 1, \ldots, T$

$$\tau_{2d,t} := -T^{-1/2}\Sigma_\circ^{-1/2}\alpha_\circ \qquad \partial_{d,2d}^2\mathcal{L}_T(\bar{\varphi}_T) = \sum_{t=1}^T \tau_{2d,t}'\tau_{2d,t} + o_P(1) \to Z_{2d}$$

$$\tau_{2e,t} := -T^{-1/2}\Sigma_\circ^{-1/2}\Gamma_\circ C_{1,\circ} \qquad \partial_{d,2e}^2\mathcal{L}_T(\bar{\varphi}_T) = \sum_{t=1}^T \tau_{2e,t}'\tau_{2e,t} + o_P(1) \to Z_{2e}$$

$$\tau_{1,t} := -T^{-3/2}\Sigma_\circ^{-1/2}\alpha_\circ t \qquad T^{-2}\partial_{d,1}^2\mathcal{L}_T(\bar{\varphi}_T) = \sum_{t=1}^T \tau_{1,t}'\tau_{1,t} + o_P(1) \to Z_1$$

and analogous expressions for the sample covariances between the deterministic terms.
Let us first deal with the case where no linear trend is present. To show the invertibility of $Z_u - Y_D' Z_D^{-1} Y_D$, we first investigate the invertibility of the corresponding sample covariance matrices and later take the limit of these quantities. We find

$$\begin{aligned}
&\sum_{t=1}^T \tau_{u,t}'\tau_{u,t} - \underbrace{\left(\sum_{t=1}^T \tau_{u,t}'\tau_{2d,t}\right)}_{\to Y_D'} \underbrace{\left(\sum_{t=1}^T \tau_{2d,t}'\tau_{2d,t}\right)^{-1}}_{\to Z_D^{-1}} \underbrace{\left(\sum_{t=1}^T \tau_{2d,t}'\tau_{u,k,t}\right)}_{\to Y_D} \\
&= \sum_{t=1}^T \tau_{u,t}'\tau_{u,t} - \left(\sum_{t=1}^T \tau_{u,t}'\right) \Sigma_\circ^{-1/2}\alpha_\circ(\alpha_\circ'\Sigma_\circ^{-1}\alpha_\circ)^{-1}\alpha_\circ'\Sigma_\circ^{-1/2} \left(\sum_{t=1}^T \tau_{u,t}\right) \\
&= \sum_{t=1}^T \tau_{u,t}' P_{\Sigma^{1/2}\alpha_\perp}\tau_{u,t} + \sum_{t=1}^T \tau_{u,t}' P_{\Sigma^{-1/2}\alpha}\tau_{u,t} - \left(\sum_{t=1}^T \tau_{u,t}'\right) P_{\Sigma^{-1/2}\alpha} \left(\sum_{t=1}^T \tau_{u,t}\right),
\end{aligned}$$

where $P_{\Sigma^{-1/2}\alpha}$ is the projection on the column space of $\Sigma_\circ^{-1/2}\alpha_\circ$, and $P_{\Sigma^{1/2}\alpha_\perp}$ is the projection on its orthogonal complement. Note that for $P_{\Sigma^{-1/2}\alpha}\tau_{u,t}$ the regression on $\tau_{2d,t}$ corresponds to regressing out the constant. Thus, defining

$$x_{t,d} := \left[(x_{t,1} - \bar{x}_1)' \quad x_{t,2}' \quad (x_{t,3} - \bar{x}_3)'\right]'$$

where $\bar{x}_j = T^{-1}\sum_{t=1}^T x_{t,j}$ for $j = 1, 2, 3$, and using

$$\partial_i^M \Psi_{u,2\Sigma} := \left[P_{\Sigma^{-1/2}\alpha}\Sigma_\circ^{-1/2}(\Pi_\circ \partial_i^M \mathcal{C}_2 + \Gamma_\circ \partial_i^M \mathcal{C}_1) \quad P_{\Sigma^{1/2}\alpha_\perp}\Sigma_\circ^{-1/2}\Gamma_\circ\partial_i^M\mathcal{C}_1) \quad \Sigma_\circ^{-1/2}\Pi_\circ\partial_i^M\mathcal{C}_3\right],$$

it follows that $Z_u - Y_D' Z_D^{-1} Y_D$ is given by

$$\left[Z_u - Y_D' Z_D^{-1} Y_D\right]_{i,j} = \begin{cases} 2\text{tr}\left[(\partial_j^M\Psi_{u,1})'\Sigma_\circ^{-1}\partial_i^M\Psi_{u,1}\boldsymbol{Z}_{1,1}^d\right], & \text{if } i,j \leq c_{u,1} \\ 2\text{tr}\left[(\partial_j^M\Psi_{u,2\Sigma})'\Sigma_\circ^{-1/2}\partial_i^M\Psi_{u,1}\boldsymbol{Z}_1^d\right], & \text{if } i \leq c_{u,1}, j > c_{u,1} \\ 2\text{tr}\left[(\partial_j^M\Psi_{u,1})'\Sigma_\circ^{-1/2}\partial_i^M\Psi_{u,2\Sigma}(\boldsymbol{Z}_1^d)'\right], & \text{if } i > c_{u,1}, j \leq c_{u,1} \\ 2\text{tr}\left[(\partial_j^M\Psi_{u,2\Sigma})'\partial_i^M\Psi_{u,2\Sigma}\boldsymbol{Z}^d\right], & \text{if } i,j > c_{u,1} \end{cases} \tag{C.6}$$

where $\boldsymbol{Z}_{1,1}^d$, $\boldsymbol{Z}_1^d$ and $\boldsymbol{Z}^d$ occur as the respective limits of

$$T^{-3}\langle x_{t,1} - \bar{x}_1, x_{t,1}\rangle \xrightarrow{d} \boldsymbol{Z}_{1,1}^d, \quad T^{-2}\langle x_{t,1} - \bar{x}_1, x_{t,d}\rangle \xrightarrow{d} \boldsymbol{Z}_1^d, \quad T^{-1}\langle x_{t,d}, x_{t,d}\rangle \xrightarrow{d} \boldsymbol{Z}^d.$$

This matrix is positive with probability one, which follows from the same arguments used for Z_u. The case with the linear trend can be dealt analogously. Note that the additional regression on the constant vector $\tau_{2e,t}$ leads to a correction for a constant for the whole vector τ_u. Adding $\tau_{1,t}$ as another additional regressor then leads to a linear detrending of $P_{\Sigma^{-1/2}\alpha}\tau_{u,t}$. It follows that $\boldsymbol{Z}_{1,1}^d$, $\boldsymbol{Z}_1^d$ and $\boldsymbol{Z}^d$ in the above expression needs to be replaced by $\boldsymbol{Z}_{1,1}^{de}$, $\boldsymbol{Z}_1^{de}$ and $\boldsymbol{Z}^{de}$, which are the limits of

$$\begin{aligned}
&T^{-3}\langle x_{t,1} - \bar{x}_1 - \acute{x}_1(t-\bar{t}), x_{t,1}\rangle \xrightarrow{d} \boldsymbol{Z}_{1,1}^{de}, \quad T^{-2}\langle x_{t,1} - \bar{x}_1 - \acute{x}_1(t-\bar{t}), x_{t,de}\rangle \xrightarrow{d} \boldsymbol{Z}_1^{de}, \\
&T^{-1}\langle x_{t,de}, x_{t,de}\rangle \xrightarrow{d} \boldsymbol{Z}^{de},
\end{aligned}$$

where

$$\begin{aligned}
\acute{x}_j &:= \left(\textstyle\sum_{t=1}^T (x_{t,j} - \bar{x}_j)t\right)\left(\textstyle\sum_{t=1}^T (t-\bar{t})^2\right)^{-1}, \quad \bar{t} = T^{-1}\sum_{t=1}^T t, \\
x_{t,de} &:= \left[(x_{t,2} - \bar{x}_2 - \acute{x}_1(t-\bar{t}))' \quad (x_{t,2} - \bar{x}_2)' \quad (x_{t,3} - \bar{x}_3 - \acute{x}_1(t-\bar{t}))'\right]'.
\end{aligned}$$

Thus, the inclusion of deterministics corresponds to replacing some of the Brownian motions by a demeaned and detrended Brownian motion. Again it follows 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 lemmata, 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 $\sqrt{T}(\hat{\theta}_{st} - \theta_{st,\circ}) \stackrel{d}{\to} \mathcal{N}(0, Z_{st}^{-1} V_{st} Z_{st}^{-1})$. Therefore (A) and (B) holds.
(C) has been shown in Lemma 11, while (D) is contained in the results of Lemma 13 and Lemma 15.
With respect to (E) note that if no deterministic terms are included in the estimation it follows, that the sub-block of $\hat{\theta}_u$ has a limiting distribution of the form $Z_u^{-1}\mathrm{v}_u$, where v_u is a random vector whose entries are given by

$$(\mathrm{v}_u)_j = \begin{cases} \mathrm{tr}[\Sigma_\circ^{-1}\partial_j^M \Psi_u \boldsymbol{X}_1] & \text{if } j \le c_{u,1} \\ \mathrm{tr}[\Sigma_\circ^{-1/2}\partial_j^M \Psi_{u,2\Sigma}[\boldsymbol{X}_2, \boldsymbol{X}_3]] & \text{if } j > c_{u,1} \end{cases}.$$

When a constant is included in the estimation, Z_u is replaced by $Z_u - Y_D' Z_D^{-1} Y_D$, compare (C.6) in Lemma 15, and v_u is replaced by

$$(\mathrm{v}_\star)_j = \begin{cases} \mathrm{tr}[\Sigma_\circ^{-1}\partial_j^M \Psi_u \boldsymbol{X}_1^d] & \text{if } j \le c_{u,1}, \\ \mathrm{tr}[\Sigma_\circ^{-1/2}\partial_j^M \Psi_{u,2\Sigma} \boldsymbol{X}_2^d] & \text{if } j > c_{u,1} \end{cases},$$

where $\boldsymbol{X}_1^d$ is the limit of $T^{-1}\langle x_{t,1} - \bar{x}_1, \varepsilon_t\rangle$ and equal to

$$\begin{aligned} &\mathcal{B}_{2,\circ}\int_0^1 \left(\int_0^u \boldsymbol{W}(v)dv - \int_0^1(\int_0^v \boldsymbol{W}(w)dw)dv\right) d\boldsymbol{W}(u)' \\ &\quad = \mathcal{B}_{2,\circ}\int_0^1 \left(\int_0^u \boldsymbol{W}(v)dv\big|_1\right) d\boldsymbol{W}(u)', \end{aligned}$$

and $\boldsymbol{X}_2^d$ is the limit of $\langle x_{t,d}, \varepsilon_t\rangle$.
If a linear trend is included Z_u is replaced by $Z_u - Y_D' Z_D^{-1} Y_D$ as in (C.6) (where Z_D and Y_D now also contain elements corresponding to the linear trend term) and v_u is replaced by a random vector defined as $\mathrm{v}_\star$ with X_1^d replaced by $\boldsymbol{X}_1^e$ defined as the limit of $T^{-1}\langle x_{t,1} - \bar{x}_1 - \acute{x}_1(t-\bar{t}), \varepsilon_t\rangle$, which is equal to

$$\begin{aligned} &\mathcal{B}_{2,\circ}\int_0^1 \left(\int_0^u \boldsymbol{W}(v)dv - \int_0^1(\int_0^v \boldsymbol{W}(w)dw)dv - 12\left(u-\tfrac{1}{2}\right)\int_0^1\left(v-\tfrac{1}{2}\right)(\int_0^v \boldsymbol{W}(w)dw)dv\right) d\boldsymbol{W}(u)' \\ &\quad = \mathcal{B}_{2,\circ}\int_0^1 \left(\int_0^u \boldsymbol{W}(v)dv\big|_{1,u}\right) d\boldsymbol{W}(u)' \end{aligned}$$

and $\boldsymbol{X}_2^d$ is replaced by $\boldsymbol{X}_2^{de}$ denoting the limit of $\langle x_{t,de}, \varepsilon_t\rangle$.
To derive the asymptotic distribution of θ_1^M another block inversion of the Hessian is necessary to derive the corresponding block of the Hessian. Following the same arguments as in Lemma 15 concerning the block corresponding to the deterministics it follows that $Z_{1,1}$ has to be replaced by the limit of $\left\langle x_{t,1}, x_{t,1}|_{x_{t,2},x_{t,3}}\right\rangle$ if no deterministics are present, by the limit of $\left\langle x_{t,1}, x_{t,1}|_{x_{t,2},x_{t,3},1}\right\rangle$ if there is a constant but no linear trend and by the limit of $\left\langle x_{t,1}, x_{t,1}|_{x_{t,2},x_{t,3},1,t}\right\rangle$ if both a constant and a linear trend occur. By Lemma 10 the above converge to $Z_{1,1\star}$ as presented in (E).
Analogously $\mathrm{v}_{1\star}$ is of the form

$$\mathrm{v}_{1\star} = \mathrm{tr}[\Sigma_\circ^{-1}\partial_j^M \Psi_u \boldsymbol{X}_{1\star}]$$

with $\boldsymbol{X}_{1\star}$ being the limit of $\left\langle x_{t,1}|_{x_{t,2},x_{t,3}}, \varepsilon_t\right\rangle$ if no deterministics are present. It is replaced by the limit of $\left\langle x_{t,1}|_{x_{t,2},x_{t,3},1}, \varepsilon_t\right\rangle$ if there is a constant but no linear trend and by the limit of $\left\langle x_{t,1}|_{x_{t,2},x_{t,3},1,t}, \varepsilon_t\right\rangle$ if both a constant and a linear trend occur. Again Lemma 10 provides the limiting distributions as given in (E).
Finally (F) is immediate from Lemma 12, which concludes the proof of Theorem 11.

To prove Corollary 6 let $Z = \mathrm{diag}(Z_\star, Z_{st})$ and note that

$$\hat{W}_R = (R\hat{\theta} - r)'(R\hat{Z}^{-1}R')^{-1}(R\hat{\theta} - r)$$
$$= ((D_T^R R (D_T^\theta)^{-1}) D_T^\theta(\hat{\theta} - \theta_\circ))'(D_T^R R (D_T^\theta)^{-1}(D_T^\theta \hat{Z}^{-1} D_T^\theta)(D_T^\theta)^{-1} R' D_T^R)^{-1}(D_T^R R (D_T^\theta)^{-1}) D_T^\theta(\hat{\theta} - \theta_\circ)$$
$$\stackrel{d}{\to} [\ Z_\star^{-1}\mathrm{v}_\star' \quad \mathrm{v}_{st}'\ ](R^\infty)'(R^\infty Z^{-1}(R^\infty)')^{-1} R^\infty [\ Z_\star^{-1}\mathrm{v}_\star' \quad \mathrm{v}_{st}'\ ]',$$

where we have used $D_T^\theta(\hat{\theta} - \theta_\circ) \to [\ Z_\star^{-1}\mathrm{v}_\star' \quad \mathrm{v}_{st}'\ ]'$ and $\hat{Z} \stackrel{d}{\to} Z$ because of Lemma 15. Since v_{st} is asymptotically normally distributed with variance Z_{st} and $\mathrm{v}_\star$ conditionally upon $\mathcal{B}_{E,\circ}\boldsymbol{W}(u)$ is asymptotically normally distributed with variance $Z_\star$, the test statistic conditionally upon $\mathcal{B}_{E,\circ}\boldsymbol{W}(u)$ is asymptotically χ_p^2 distributed which implies that the same result holds marginally.

C.3 Convergence Results

Let $\underline{\hat{A}}, \hat{B}$ be chosen by a pseudo maximum likelihood approach. An estimate of the state $\hat{x}_t$ is then given by $\hat{x}_{t+1} = \underline{\hat{A}}\hat{x}_t + \hat{B}y_t$. Minimizing the likelihood with respect to C is then equivalent to a regression in the model.

$$y_t = C\hat{x}_t + \varepsilon_t,$$

which is the starting point in the derivation of the asymptotic distribution of the rank test statistic. Let us introduce the following notation:

- Let $\hat{\varepsilon}_t$ denote the residuals of the (unrestricted) regression of y_t on $\hat{x}_t$.
- Let $\hat{\varepsilon}_t^c$ denote the residuals from the restricted regression and let $(\hat{\mathcal{A}}, \hat{\mathcal{B}}, \hat{\mathcal{C}})$ be the corresponding estimated system transformed into the canonical form for systems corresponding to I(2) processes.
- The restricted regression introduces a partitioning of the state $\hat{x}_t$ into estimated stochastic trends and estimated stationary components. Let $\hat{x}_{t,u}$ denote the state space components corresponding to the stochastic trends, $\hat{x}_{t,e} := [\hat{x}_{t,1}', \hat{x}_{t,3}]'$ the I(2) component and the I(1) component that does not sum up to an I(2) trend. Let $\hat{x}_{t,g} := \hat{x}_{t,2}$ denote the state space components corresponding to the $I(1)$ components of the states which sum up to the $I(2)$ component of the state, and $\hat{x}_{t,\bullet}$ the components corresponding to the stationary part of the state.
- Let $\hat{\alpha}_\perp := \hat{\mathcal{B}}_E'(\hat{\mathcal{B}}_E\hat{\mathcal{B}}_E')^{-1/2} = [\hat{\mathcal{B}}_2', \hat{\mathcal{B}}_3'](\hat{\mathcal{B}}_E\hat{\mathcal{B}}_E')^{-1/2}$ be such that $\hat{\alpha}_\perp'\hat{\Pi} = \hat{\mathcal{B}}_E\hat{\Pi} = 0$, where $\hat{\Pi} := -I + \hat{\mathcal{C}}(I - \underline{\hat{\mathcal{A}}})^{-1}\hat{\mathcal{B}}$, and $\hat{\alpha}$ the orthogonal complement of $\hat{\alpha}_\perp$ in $\mathbb{R}^s$ with $\hat{\beta}$ such that $\hat{\Pi} = \hat{\alpha}\hat{\beta}'$. Let $\hat{\gamma}_\perp := \hat{\mathcal{B}}_2'(\hat{\mathcal{B}}_2\hat{\mathcal{B}}_2')^{-1/2}$ be the such that $\hat{\gamma}_\perp'\hat{\Gamma}\hat{\mathcal{C}}_1 = \hat{\mathcal{B}}_2\hat{\Gamma}\hat{\mathcal{C}}_1 = 0$, where $\hat{\Gamma} := -I + \hat{\mathcal{C}}(I - \underline{\hat{\mathcal{A}}})^{-2}\underline{\hat{\mathcal{A}}}\hat{\mathcal{B}}$, and let $\hat{\gamma}$, with $\hat{\gamma}'\hat{\gamma} = I_{c_2}$, be a matrix whose column space is the orthogonal complement in the space spanned by the columns of $\hat{\alpha}_\perp$.

The next Lemma then introduces three necessary conditions for the residuals $\hat{\varepsilon}_t^c$ to correspond to a maximizer of the restricted regression.

Lemma 16 *The following equalities hold:*

$$\begin{array}{lll} (i) & \langle \hat{\varepsilon}_t^c, \hat{x}_{t,\bullet} \rangle & = 0 \\ (ii) & \hat{\alpha}' \langle \hat{\varepsilon}_t^c, \hat{\varepsilon}_t^c \rangle^{-1} \langle \hat{\varepsilon}_t^c, \hat{x}_{t,u} \rangle & = 0 \\ (iii) & \hat{\gamma}' \langle \hat{\varepsilon}_t^c, \hat{\varepsilon}_t^c \rangle^{-1} \langle \hat{\varepsilon}_t^c, \hat{x}_{t,g} \rangle & = 0 \end{array}$$

The above imply

$$\hat{\varepsilon}_t^c = \hat{\varepsilon}_t + \hat{\alpha}_\perp \left\langle \hat{\varepsilon}_t^c, \hat{x}_{t,e}|_{\hat{x}_{t,g}} \right\rangle \left\langle \hat{x}_{t,e}|_{\hat{x}_{t,g}}, \hat{x}_{t,e}|_{\hat{x}_{t,g}} \right\rangle^{-1} \hat{x}_{t,e}|_{\hat{x}_{t,g}} + \hat{\gamma}_\perp \langle \hat{\varepsilon}_t^c, \hat{x}_{t,g} \rangle \langle \hat{x}_{t,g}, \hat{x}_{t,g} \rangle^{-1} \hat{x}_{t,g}$$

Proof: Let $\hat{C}$ be the estimator of the restricted regression and let $(\hat{\mathcal{A}}, \hat{\mathcal{B}}, \hat{\mathcal{C}})$ be the corresponding system in canonical form. We will argue that if either (i) or (ii) or (iii) does not hold, then we can find a matrix $\tilde{C}$ (not necessarily in canonical form), satisfying the restrictions, such that

$$\det \left\langle y_t - \hat{C}\hat{x}_t, y_t - \hat{C}\hat{x}_t \right\rangle > \det \left\langle y_t - \tilde{C}\hat{x}_t, y_t - \tilde{C}\hat{x}_t \right\rangle ,$$

which is a contradiction to $\hat{C}$ being the estimator of the restricted regression.

(i) Consider the model $\hat{\varepsilon}_t^c = C_{(i)}\hat{x}_{t,\bullet} + \tilde{\varepsilon}_t^c$ such that the regression is performed only on the regressors in $\hat{x}_{t,\bullet}$. If $\langle \hat{\varepsilon}_t^c, \hat{x}_{t,\bullet} \rangle \neq 0$, we clearly have $\det \langle \hat{\varepsilon}_t^c, \hat{\varepsilon}_t^c \rangle > \det \langle \hat{\tilde{\varepsilon}}_t^c, \hat{\tilde{\varepsilon}}_t^c \rangle$ for the OLS residuals $\hat{\tilde{\varepsilon}}_t^c$ of the above restricted regression or equivalently

$$\det \left\langle y_t - \hat{\mathcal{C}}\hat{x}_t, y_t - \hat{\mathcal{C}}\hat{x}_t \right\rangle > \det \left\langle y_t - \tilde{C}\hat{x}_t, y_t - \tilde{C}\hat{x}_t \right\rangle$$

with $\tilde{C} = \hat{\mathcal{C}} + \begin{bmatrix} 0_{s\times n_u} & \hat{C}_{(i)} \end{bmatrix}$, where $\hat{C}_{(i)}$ denotes the OLS estimator from the above regression. What is left is to show that $\tilde{\mathcal{C}}$ fulfills the desired restriction. Using $\tilde{A} = \underline{\hat{\mathcal{A}}} + \hat{\mathcal{B}}\tilde{\mathcal{C}} = \hat{\mathcal{A}} + \hat{\mathcal{B}} \begin{bmatrix} 0 & \hat{C}_{(i)} \end{bmatrix}$ we see that $\tilde{A}$ is of the form

$$\begin{bmatrix} \mathcal{A}_u & \star \\ 0 & \star \end{bmatrix},$$

where $\star$ indicates entries which are not further specified. Thus, there exists a regular matrix $\tilde{M}$ such that $\tilde{A} = \tilde{M} \begin{bmatrix} \mathcal{A}_u & 0 \\ 0 & \star \end{bmatrix} \tilde{M}^{-1}$, which implies that $\tilde{C}$ indeed fulfills the rank restrictions. This contradicts the assumption of $\hat{\mathcal{C}}$ being the estimator of the restricted regression.

(ii) For the restricted optimization we have to minimize $\det \langle \hat{\varepsilon}_t^c, \hat{\varepsilon}_t^c \rangle$. We note that

$$\det \langle \hat{\varepsilon}_t^c, \hat{\varepsilon}_t^c \rangle = \det N \langle \hat{\varepsilon}_t^c, \hat{\varepsilon}_t^c \rangle N' = \det \langle N\hat{\varepsilon}_t^c, N\hat{\varepsilon}_t^c \rangle$$

for any matrix N with $\det N = 1$. Choose $N = [\hat{\alpha}_\perp, \hat{\alpha}]'$, such that $\det N = 1$ holds. Then the likelihood splits into two components.

$$\begin{aligned} &\det \langle \hat{\varepsilon}_t^c, \hat{\varepsilon}_t^c \rangle \\ &\quad = \det \langle \hat{\alpha}_\perp' \hat{\varepsilon}_t^c, \hat{\alpha}_\perp' \hat{\varepsilon}_t^c \rangle \det \left(\hat{\alpha}' \langle \hat{\varepsilon}_t^c, \hat{\varepsilon}_t^c \rangle \hat{\alpha} - \hat{\alpha}' \langle \hat{\varepsilon}_t^c, \hat{\varepsilon}_t^c \rangle \hat{\alpha}_\perp (\hat{\alpha}_\perp' \langle \hat{\varepsilon}_t^c, \hat{\varepsilon}_t^c \rangle \hat{\alpha}_\perp)^{-1} \hat{\alpha}_\perp' \langle \hat{\varepsilon}_t^c, \hat{\varepsilon}_t^c \rangle \hat{\alpha} \right) \\ &\quad = \det \langle \hat{\alpha}_\perp' \hat{\varepsilon}_t^c, \hat{\alpha}_\perp' \hat{\varepsilon}_t^c \rangle \det \langle \hat{\alpha}' \hat{\varepsilon}_{\perp,t}, \hat{\alpha}' \hat{\varepsilon}_{\perp,t} \rangle , \end{aligned}$$

where $\hat{\alpha}'\hat{\varepsilon}_{\perp,t} := \hat{\alpha}'\hat{\varepsilon}_t^c - \hat{\alpha}' \langle \hat{\varepsilon}_t^c, \hat{\varepsilon}_t^c \rangle \hat{\alpha}_\perp (\hat{\alpha}_\perp' \langle \hat{\varepsilon}_t^c, \hat{\varepsilon}_t^c \rangle \hat{\alpha}_\perp)^{-1} \hat{\alpha}_\perp' \hat{\varepsilon}_t^c$. These decompositions also play a role in the VECM setting, compare Johansen (1996, Lemma 10.1) for different (asymptotic) relations between these empirical variances. Note that

$$\hat{\varepsilon}_{\perp,t} = \hat{\alpha} \left(\hat{\alpha}' \langle \hat{\varepsilon}_t^c, \hat{\varepsilon}_t^c \rangle^{-1} \hat{\alpha} \right)^{-1} \hat{\alpha}' \langle \hat{\varepsilon}_t^c, \hat{\varepsilon}_t^c \rangle^{-1} \hat{\varepsilon}_t^c$$

since the residuals of the regression of $\hat{\varepsilon}_t^c$ on $\hat{\alpha}_\perp \hat{\varepsilon}_t^c$ are equal to the projection of $\hat{\varepsilon}_t^c$ on $\hat{\alpha}' \langle \hat{\varepsilon}_t^c, \hat{\varepsilon}_t^c \rangle^{-1} \hat{\varepsilon}_t^c$. If $\langle \hat{\varepsilon}_t^c, \hat{\varepsilon}_t^c \rangle^{-1} \langle \hat{\varepsilon}_t^c, \hat{x}_{t,u} \rangle \neq 0$, we can find a matrix $\hat{C}_{(ii)} \neq 0$ by regression in the model $\hat{\alpha}'\hat{\varepsilon}_{\perp,t} = C_{(ii)}\hat{x}_{t,u} + \hat{\alpha}'\tilde{\varepsilon}_{\perp,t}$ such that $\det \langle \hat{\alpha}'\hat{\varepsilon}_{\perp,t}, \hat{\alpha}'\hat{\varepsilon}_{\perp,t} \rangle > \det \langle \hat{\alpha}'\hat{\tilde{\varepsilon}}_{\perp,t}, \hat{\alpha}'\hat{\tilde{\varepsilon}}_{\perp,t} \rangle$, where $\hat{\tilde{\varepsilon}}_{\perp,t}$ denotes the OLS residuals in the auxiliary model. Thus,

$$\det \left\langle y_t - \hat{\mathcal{C}}\hat{x}_t, y_t - \hat{\mathcal{C}}\hat{x}_t \right\rangle > \det \left\langle y_t - \tilde{C}\hat{x}_t, y_t - \tilde{C}\hat{x}_t \right\rangle$$

with $\tilde{C} := \hat{\mathcal{C}} + [\hat{\alpha}\hat{C}_{(ii)}, 0_{s\times(n-2c_1-c_2)}]$ and $\hat{C}_{(ii)} := \hat{\alpha}' \langle \hat{\varepsilon}_t^c, \hat{x}_{t,u} \rangle \langle \hat{x}_{t,u}, \hat{x}_{t,u} \rangle^{-1}$ being the OLS estimator. Note that

$$\hat{\alpha}_\perp'(-I + \tilde{C}(I - \underline{\hat{\mathcal{A}}})^{-1}\hat{\mathcal{B}}) = \hat{\alpha}_\perp'\hat{\Pi} + \hat{\alpha}_\perp'\hat{\alpha}\hat{C}_{(ii)}(I - \underline{\hat{\mathcal{A}}})^{-1}\hat{\mathcal{B}} = 0$$

such that $rank(\tilde{\Pi}) \leq s - c_1 - c_2$ holds. Moreover,

$$\hat{\gamma}_\perp'(-I + \tilde{C}(I - \underline{\hat{\mathcal{A}}})^{-2}\underline{\hat{\mathcal{A}}}\hat{\mathcal{B}})\hat{\mathcal{C}}_E = \hat{\gamma}_\perp'\hat{\Pi}\hat{\mathcal{C}}_E + \hat{\gamma}_\perp'\alpha\hat{C}_{(ii)}(I - \underline{\hat{\mathcal{A}}})^{-2}\underline{\hat{\mathcal{A}}}\hat{\mathcal{B}}\hat{\mathcal{C}}_E = 0$$

such that $rank(\hat{\mathcal{B}}_E\tilde{\Gamma}\hat{\mathcal{C}}_E) \leq c_2$ holds. Thus, $\tilde{C}$ also fulfills the restrictions contradicting the assumption of an optimal choice of $\hat{\mathcal{C}}$.

(iii) Here we proceed similarly to (ii). To minimize $\det \langle \hat{\varepsilon}_t^c, \hat{\varepsilon}_t^c \rangle$ focus on $\det \langle \hat{\alpha}'\hat{\varepsilon}_{\perp,t}, \hat{\alpha}'\hat{\varepsilon}_{\perp,t} \rangle$ as defined in (ii). Decompose $\hat{\alpha}_\perp$ into $[\hat{\gamma}, \hat{\gamma}_\perp]$. Then the determinant splits into two components.

$$\det \langle \hat{\alpha}'\hat{\varepsilon}_{\perp,t}, \hat{\alpha}'\hat{\varepsilon}_{\perp,t} \rangle = \det \langle \hat{\gamma}_\perp'\hat{\varepsilon}_{\perp,t}, \hat{\gamma}_\perp'\hat{\varepsilon}_{\perp,t} \rangle \det \left\langle \hat{\gamma}'\hat{\varepsilon}^g_{\perp,t}, \hat{\gamma}'\hat{\varepsilon}^g_{\perp,t} \right\rangle,$$

where $\gamma'\hat{\varepsilon}^g_{\perp,t} := \hat{\gamma}'\hat{\varepsilon}_{\perp,t} - \hat{\gamma}' \langle \hat{\varepsilon}_{\perp,t}, \hat{\varepsilon}_{\perp,t} \rangle \hat{\gamma}_\perp \left(\hat{\gamma}_\perp' \langle \hat{\varepsilon}_{\perp,t}, \hat{\varepsilon}_{\perp,t} \rangle \hat{\gamma}_\perp\right)^{-1} \hat{\gamma}_\perp'\hat{\varepsilon}_{\perp,t}$. Again

$$\hat{\varepsilon}^g_{\perp,t} = \hat{\gamma} \left(\hat{\gamma}' \langle \hat{\varepsilon}_t^c, \hat{\varepsilon}_t^c \rangle^{-1} \hat{\gamma}\right)^{-1} \hat{\gamma}' \langle \hat{\varepsilon}_t^c, \hat{\varepsilon}_t^c \rangle^{-1} \hat{\varepsilon}_t^c$$

since the residuals of the regression of $\hat{\varepsilon}_{\perp,t}$ on $\hat{\gamma}_\perp'\hat{\varepsilon}^c_{\perp,t}$ are equal to the projection of $\hat{\varepsilon}_t^c$ on $\hat{\gamma}' \langle \hat{\varepsilon}_t^c, \hat{\varepsilon}_t^c \rangle^{-1} \hat{\varepsilon}_t^c$. If $\langle \hat{\varepsilon}_t^c, \hat{\varepsilon}_t^c \rangle^{-1} \langle \hat{\varepsilon}_t^c, \hat{x}_{t,g} \rangle \neq 0$, we can find a matrix $\hat{C}_{(iii)} \neq 0$ by regression in the model $\hat{\gamma}'\hat{\varepsilon}^g_{\perp,t} = C_{(iii)}\hat{x}_{t,g} + \hat{\gamma}'\tilde{\varepsilon}^g_{\perp,t}$ such that $\det\langle \hat{\gamma}'\hat{\varepsilon}^g_{\perp,t}, \hat{\gamma}'\hat{\varepsilon}^g_{\perp,t} \rangle > \det\langle \hat{\gamma}'\hat{\tilde{\varepsilon}}^g_{\perp,t}, \hat{\gamma}'\hat{\tilde{\varepsilon}}^g_{\perp,t} \rangle$, where $\hat{\tilde{\varepsilon}}^g_{\perp,t}$ denotes the OLS residuals in the auxiliary model. Thus,

$$\det \left\langle y_t - \hat{\mathcal{C}}\hat{x}_t, y_t - \hat{\mathcal{C}}\hat{x}_t \right\rangle > \det \left\langle y_t - \tilde{C}\hat{x}_t, y_t - \tilde{C}\hat{x}_t \right\rangle$$

with $\tilde{C} := \hat{\mathcal{C}} + [0_{s\times c_1}, \gamma\hat{C}_{(iii)}, 0_{s\times(n-2c_1)}]$ and $\hat{C}_{(iii)} := \hat{\gamma}' \left\langle \hat{\varepsilon}^g_{\perp,t}, \hat{x}_{t,g} \right\rangle \langle \hat{x}_{t,g}, \hat{x}_{t,g} \rangle^{-1}$ being the OLS estimator. Using $\tilde{A} = \underline{\hat{\mathcal{A}}} + \hat{\mathcal{B}}\tilde{C}$ we see that $\tilde{A}$ is of the form

$$\begin{bmatrix} I_{d_1} & \star & 0 & 0 \\ 0 & I_{d_1} & 0 & 0 \\ 0 & \star & I_{d_2} & 0 \\ 0 & \star & 0 & \mathcal{A}_\bullet \end{bmatrix},$$

where $\star$ indicates entries which are not further specified. It follows that the geometric multiplicity of the eigenvalues of $\tilde{A}$ is greater or equal to $c_1 + c_2$. Moreover, $\tilde{A}$ can be transformed into canonical form $\tilde{\mathcal{A}} = M^{-1}\tilde{A}M$ with M being a block diagonal matrix of the form $\mathrm{diag}(M_1, M_2, M_3, M_\bullet)$. This ensures that the column space of $\hat{\mathcal{C}}_E$ is a subspace of the column space of the subblock $\tilde{\mathcal{C}}_E$ of $\tilde{\mathcal{C}} = \tilde{C}M$. Note that

$$\hat{\gamma}_\perp'(-I + \tilde{C}(I - \underline{\hat{\mathcal{A}}})^{-2}\underline{\hat{\mathcal{A}}}\hat{\mathcal{B}})\hat{\mathcal{C}}_E = \hat{\gamma}_\perp'\hat{\Pi}\hat{\mathcal{C}}_E + \hat{\gamma}_\perp'\hat{\gamma}\hat{C}_{(iii)}(I - \underline{\hat{\mathcal{A}}})^{-2}\underline{\hat{\mathcal{A}}}\hat{\mathcal{B}}\hat{\mathcal{C}}_E = 0$$

such that $rank(\hat{\mathcal{B}}_E\tilde{\Gamma}\hat{\mathcal{C}}_E) \leq c_2$ holds. Thus, $\tilde{C}$ also fulfills the restrictions contradicting the assumption of an optimal choice of $\hat{\mathcal{C}}$.

■

Next, let us discuss some properties of $\hat{x}_t$, which we summarize in the following Lemma

Lemma 17 *Let $\{y_t\}_{t\in\mathbb{Z}}$ be an I(2) process generated by a system of the form (3.2) and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ fulfilling Assumption 3. Assume the true order n is known. Let $\hat{\theta}$ be a PML estimator of θ over a suitable parameter space $\mathbf{\Theta}$ fulfilling the assumption of Theorem 10 and define $\hat{x}_t^{\mathcal{E}} := \hat{x}_t^{\mathcal{E}}(\hat{\theta}) := \underline{\mathcal{A}}^{\mathcal{E}}(\hat{\theta}) + \mathcal{B}^{\mathcal{E}}(\hat{\theta})y_t$, and $\varepsilon_t(\hat{\theta}) := y_t - \mathcal{C}^{\mathcal{E}}(\hat{\theta})\hat{x}_t^{\mathcal{E}}$, where $\underline{\mathcal{A}}^{\mathcal{E}}(\hat{\theta})$, $\mathcal{B}^{\mathcal{E}}(\hat{\theta})$ and $\mathcal{C}^{\mathcal{E}}(\hat{\theta})$ are assumed to be in echelon canonical form. Let $x_t^{\mathcal{E}} = \underline{\mathcal{A}}_\circ^{\mathcal{E}} + \mathcal{B}_\circ^{\mathcal{E}}y_t$ denote the true state corresponding to the true* *matrices* $\underline{\mathcal{A}}_\circ^{\mathcal{E}}$ *and* $\mathcal{B}_\circ^{\mathcal{E}}$ *in echelon canonical form. Then there exists a sequence of matrices* G_T *with* $G_T \to G$ *such that the following results hold:*

$$\begin{aligned} &(i) && D_T^x\Big(\Big\langle Gx_t^{\mathcal{E}}, Gx_t^{\mathcal{E}}\Big\rangle - \Big\langle G_T\hat{x}_t^{\mathcal{E}}, G_T\hat{x}_t^{\mathcal{E}}\Big\rangle\Big)D_T^x && \to 0, \\ &(ii) && D_T^x\Big\langle G_T\hat{x}_t^{\mathcal{E}}, G_T\hat{x}_t^{\mathcal{E}}\Big\rangle D_T^x && \to Z_{xx}, \text{ with } \mathbb{P}\{Z_{xx} > 0\} \to 1 \\ &(iii) && \Big\langle \varepsilon_t(\hat{\theta}), G_T\hat{x}_t^{\mathcal{E}}\Big\rangle D_T^x && \to 0, \end{aligned}$$

where $D_T^x = diag(T^{-3/2}I_{c_1}, T^{-1/2}I_{c_1+c_2}, I_{n_\bullet})$*.*

Proof: Define $g(z,\theta) := \sum_{j=0}^{\infty}(z\underline{\mathcal{A}}^{\mathcal{E}}(\theta))^{j-1}\mathcal{B}^{\mathcal{E}}(\theta)$ and decompose $g(z;\theta) = g(1;\theta) + (1-z)\breve{g}(z;\theta) = g(1;\theta) + (1-z)\breve{g}(1;\theta) + (1-z)^2\tilde{g}(z;\theta)$, where stability of $\breve{g}(z;\theta)$ and $\tilde{g}(z;\theta)$ holds if $g(z;\theta)$ is stable. Let us first analyze the different components of

$$\begin{aligned} g(L;\hat{\theta})y_t &= g(1;\theta)y_t + \breve{g}(1;\theta)\mathbf{\Delta}y_t + \tilde{g}(z;\theta)\mathbf{\Delta}^2 y_t \\ &= g(1;\hat{\theta})\mathcal{C}_{1,\circ}x_{t,1} + \breve{g}(1;\theta)\mathcal{C}_{1,\circ}\mathbf{\Delta}x_{t,1} + g(1;\hat{\theta})\mathcal{C}_{2,\circ}x_{t,2} + g(1;\hat{\theta})\mathcal{C}_{3,\circ}x_{t,3} \\ &\quad \breve{g}(1;\hat{\theta})\mathbf{\Delta}(\mathcal{C}_{2,\circ}x_{t,2} + \mathcal{C}_{3,\circ}x_{t,3}) + g(1;\hat{\theta})\mathcal{C}_{\bullet,\circ}x_{t,\bullet} + \breve{g}(1;\hat{\theta})\mathbf{\Delta}\mathcal{C}_{\bullet,\circ}x_{t,\bullet} + \tilde{g}(z;\theta)\mathbf{\Delta}^2 y_t. \end{aligned}$$

Since $g(z;\hat{\theta}) \to g(z;\theta_\circ)$ uniformly due to the consistency of the PML estimator $\hat{\theta}$ and $g(z;\theta_\circ)$ is stable, it follows that $g(1;\hat{\theta})\mathcal{C}_{1,\circ}x_{t,1}$ is the only component integrated of order two. Let $G_1(\theta) := g(1;\theta_\circ)\mathcal{C}_{1,\circ} \in \mathbb{R}^{n\times c_1}$. Then by Lemma 11

$$T^{-3}\langle g(L;\hat{\theta})\mathcal{C}_{1,\circ}x_{t,1}, g(L;\hat{\theta})\mathcal{C}_{1,\circ}x_{t,1}\rangle \to G_1(\theta_\circ)\boldsymbol{Z}_{1,1}G_1(\theta_\circ)'$$

and analogously for $\langle g(L;\theta_\circ)\mathcal{C}_{1,\circ}x_{t,1}, g(L;\theta_\circ)\mathcal{C}_{1,\circ}x_{t,1}\rangle$ such that

$$T^{-3}\langle g(L;\hat{\theta})y_t, g(L;\hat{\theta})y_t\rangle - T^{-3}\,\langle g(L;\theta_\circ)y_t, g(L;\theta_\circ)y_t\rangle \to 0.$$

Similarly, define $G_2(\theta) := P_{G,1\perp}(\breve{g}(1;\theta)\mathcal{C}_{1,\circ} + g(1;\hat{\theta})\mathcal{C}_{2,\circ})$ and $P_{G,1\perp}G_3(\theta) := P_{G,1\perp}g(1;\hat{\theta})\mathcal{C}_{3,\circ}$, with $P_{G,1\perp} := I_n - G_1(\theta)(G_1(\theta)'G_1(\theta))^{-1}G_1(\theta)'$. Then again due to Lemma 11

$$\begin{aligned} &T^{-1}\langle[G_2(\hat{\theta}), G_3(\hat{\theta})]'g(L;\hat{\theta})y_t, [G_2(\hat{\theta}), G_3(\hat{\theta})]'g(L;\hat{\theta})y_t\rangle \\ &\quad - T^{-1}\,\langle[G_2(\theta_\circ), G_3(\theta_\circ)]'g(L;\theta_\circ)y_t, [G_2(\theta_\circ), G_3(\theta_\circ)]'g(L;\theta_\circ)y_t\rangle \to 0. \end{aligned}$$

Similarly for the stationary part, define $G_\perp(\theta)$ orthogonal to the column space of $[G_1(\hat{\theta}), G_2(\hat{\theta}), G_3(\hat{\theta})]$. Then

$$\begin{aligned} &T^{-1}\langle G_\perp(\hat{\theta})'g(L;\hat{\theta})y_t, G_\perp(\hat{\theta})'g(L;\hat{\theta})y_t\rangle \\ &\quad - T^{-1}\,\langle G_\perp(\theta_\circ)'g(L;\theta_\circ)y_t, G_\perp(\theta_\circ)'g(L;\theta_\circ)y_t\rangle \to 0, \end{aligned}$$

and similarly for the different cross terms. It follows that

$$D_T^x\,\langle G(\theta_\circ)'x_t^{\mathcal{E}}, G(\theta_\circ)'x_t^{\mathcal{E}}\rangle'\,D_T^x - D_T^x\,\Big\langle G(\hat{\theta})'\hat{x}_t^{\mathcal{E}}, G(\hat{\theta})'\hat{x}_t^{\mathcal{E}}\Big\rangle\,D_T^x \to 0.$$

which proves (i) for $G_T = G(\hat{\theta}) := [G_1(\theta), G_2(\theta), G_3(\theta), G_\perp(\theta)]$.
Further, note that multiplication by $G(\theta_\circ)'$ leads to a change of basis of the state vector in the echelon canonical form to a state corresponding to the canonical form for systems of I(2) processes as in (3.2), i. e.,

$$G(\theta_\circ)'x_t^{\mathcal{E}} = [(\tilde{G}_1x_{t,1})', (\tilde{G}_2x_{t,2})', (\tilde{G}_3x_{t,3})', (\tilde{G}_\bullet x_{t,\bullet})']',$$

where the non-singular matrices $\tilde{G}_1 \in \mathbb{R}^{c_1\times c_1}$, $\tilde{G}_2 \in \mathbb{R}^{c_2\times c_2}$ and $\tilde{G}_\bullet \in \mathbb{R}^{c_\bullet\times c_\bullet}$ still occur. This is due to the fact that the above basis change does not ensure the correct basis of the subspaces spanned by the new state components $G_1(\theta_\circ)'x_t^{\mathcal{E}}$, $G_2(\theta_\circ)'x_t^{\mathcal{E}}$, $G_3(\theta_\circ)'x_t^{\mathcal{E}}$ and $G_\perp(\theta_\circ)'x_t^{\mathcal{E}}$. Consequently, $G(\theta_\circ)$ is invertible, implying also asymptotic invertibility for $G(\hat\theta)$. It follows that

$$D_T^x \left\langle G_T\hat{x}_t^{\mathcal{E}}, G_T\hat{x}_t^{\mathcal{E}}\right\rangle D_T^x \to \operatorname{diag}(\tilde G_1,\tilde G_2,\tilde G_3,\tilde G_4)^{-1}\begin{bmatrix}\boldsymbol{Z} & 0\\ 0 & Z_{x_\bullet x_\bullet}\end{bmatrix}(\operatorname{diag}(\tilde G_1,\tilde G_2,\tilde G_3,\tilde G_4)^{-1})',$$

where $\boldsymbol{Z}$ is the limit of $D_T^u\left\langle x_{t,u}, x_{t,u}\right\rangle D_T^u$, for which $\mathbb{P}\{\boldsymbol{Z}>0\}\to 1$ holds, and $Z_{x_\bullet x_\bullet} = \mathbb{E}(x_{t,\bullet}x_{t,\bullet}')>0$, proving (ii).
For (iii) note that

$$\begin{aligned}
&T^{-1}\langle k^{-1}(L;\hat\theta)k(L;\theta_\circ)\varepsilon_t, g(L;\hat\theta)\mathcal{C}_{1,\circ}x_{t,1}\rangle\\
&\quad= T^{-1}\langle \varepsilon_t + \upsilon_t(\hat\theta) - \varepsilon_t - \Psi_u(\hat\theta)w_{t,u}, g(L;\hat\theta)\mathcal{C}_{1,\circ}x_{t,1}\rangle\\
&\quad= T^{-1}\langle \varepsilon_t, g(L;\hat\theta)\mathcal{C}_{1,\circ}x_{t,1}\rangle + o_p(1) \to \boldsymbol{X}_1' G_1(\theta_\circ)',
\end{aligned}$$

where

$$\begin{aligned}
\upsilon_t(\theta) :&= \tilde k(L;\theta)\boldsymbol{\Delta}^2[\mathcal{C}_{u,\circ}x_{t,u} + \mathcal{C}_{\bullet,\circ}x_{t,\bullet} + \varepsilon_t] - \Pi(\theta)[\mathcal{C}_{2,\circ}\mathcal{B}_{2,\circ}\varepsilon_{t-1} + \mathcal{C}_{\bullet,\circ}x_{t,\bullet} + \varepsilon_t]\\
&\quad - \Gamma(\theta)[\mathcal{C}_{1,\circ}\mathcal{B}_{1,\circ}\varepsilon_{t-1} + \mathcal{C}_{2,\circ}\mathcal{B}_{2,\circ}\varepsilon_{t-1} + \mathcal{C}_{3,\circ}\mathcal{B}_{3,\circ}\varepsilon_{t-1} + \boldsymbol{\Delta}(\mathcal{C}_{\bullet,\circ}x_{t,\bullet} + \varepsilon_t)]\\
&=: k^\upsilon(L,\theta)\varepsilon_t\\
\Psi_u(\theta) :&= \begin{bmatrix}\Pi(\theta)\mathcal{C}_{1,\circ} & \Pi(\theta)\mathcal{C}_{2,\circ} + \Gamma(\theta)\mathcal{C}_{1,\circ} & \Pi(\theta)\mathcal{C}_{3,\circ}\end{bmatrix}\\
w_{t,u} :&= [x_{t-1,1}', x_{t-2,2}', x_{t-1,3}']'
\end{aligned}$$

and where $T^{-1}\langle \upsilon_t(\hat\theta)-\varepsilon_t, g(1;\hat\theta)\mathcal{C}_{1,\circ}x_{t,1}\rangle = T^{-1}\langle (k_t^\upsilon(L;\hat\theta)-I_s)\varepsilon_t, g(1;\hat\theta)\mathcal{C}_{1,\circ}x_{t,1}\rangle$ tends to zero due to consistency of $k^\upsilon(z;\hat\theta)\to k^\upsilon(z;\theta_\circ) = I_s$ and $T^{-1}\langle \Psi_u(\hat\theta)w_{t,u}, g(1;\hat\theta)\mathcal{C}_{1,\circ}x_{t,1}\rangle$ tends to zero due to the super-consistency results for $\Psi_u(\hat\theta)$, compare Theorem 10. Thus,

$$T^{-3/2}\langle k^{-1}(L;\hat\theta)k(L;\theta_\circ)\varepsilon_t, g(1;\hat\theta)\mathcal{C}_{1,\circ}x_{t,1}\rangle \to 0.$$

An analogous result holds for the terms integrated of order one, while for the stationary terms in $\hat x_t$ convergence to zero holds since $G_\perp(\hat\theta)\hat x_t$ and ε_t are uncorrelated. Thus, (iii) follows. ■
Some general results for different estimators of C in this model are summarized in the next lemma.

Lemma 18 *Let $\{y_t\}_{t\in\mathbb{Z}}$ be an I(2) process generated by a system of the form (3.2) and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ fulfilling Assumption 3. Let $\hat x_t := x_t^{\mathcal{E}}(\hat\theta)$ be defined as in Lemma 17. Let C be the limit of $C(\hat\theta)\xrightarrow{p} C$. Then the following consistency results hold.*

(i) The OLS-estimator $\hat C^{OLS} := \langle y_t, \hat x_t\rangle\langle \hat x_t, \hat x_t\rangle^{-1}$ from the regression $y_t = C\hat x_t + \varepsilon_t$ is a consistent estimator for C.

(ii) Let $\tilde{\mathcal{C}}_u$ be a (normalized) maximizer of the function $L_T^{\mathcal{C}}(\mathcal{C}_u,\hat\theta)$ over $\mathbb{U}(c_1,c_2,\hat\theta)$, thus, satisfying

$$\max_{\mathcal{C}_u\in\mathbb{U}(c_1,c_2,\hat\theta)} L_T^{\mathcal{C}}(\mathcal{C}_u,\hat\theta) = L_T^{\mathcal{C}}(\tilde{\mathcal{C}}_u,\hat\theta).$$

The corresponding estimator $\hat C := \hat C(\tilde{\mathcal{C}}_u)$ is consistent for C.

(iii) Furthermore,

$$T^\gamma\|\hat\Pi^{OLS}\mathcal{C}_{1,\circ}\|\to 0,$$

in probability for all $0<\gamma<2$, where $\hat\Pi := -I_s - \hat C^{OLS}(I_s - \underline{\hat A})^{-1}\hat B$, and

$$T^\gamma\|\hat\Pi^{OLS}\mathcal{C}_{2,\circ}\|\to 0,\quad \text{and}\quad T^\gamma\|\hat\Pi^{OLS}\mathcal{C}_{3,\circ} + \hat\Gamma^{OLS}\mathcal{C}_{1,\circ}\|\to 0,$$

in probability for all $0<\gamma<1$, where $\hat\Gamma^{OLS} := -I_s + \hat C^{OLS}(I_s - \underline{\hat A})^{-2}\underline{\hat A}\hat B$. The same holds true using $\hat\Pi := -I_s - \hat C(I_s - \underline{\hat A})^{-1}\hat B$ and $\hat\Gamma := -I_s + \hat C(I_s - \underline{\hat A})^{-2}\underline{\hat A}\hat B$.

Proof: (i) The proof follows the arguments of the proof of Theorem A1 of Johansen (1997). First, note that the OLS-estimator maximizes the normalized concentrated likelihood function

$$f_T(C) = \frac{|\langle y_t - C\hat{x}_t, y_t - C\hat{x}_t\rangle|}{|\langle y_t - C(\hat{\theta})\hat{x}_t, y_t - C(\hat{\theta})\hat{x}_t\rangle|},$$

which can be further rewritten into

$$f_T(C) = |\Sigma(\hat{\theta})|^{-1}|M_{\hat{\varepsilon}\hat{\varepsilon}} + (C - C(\hat{\theta}) - M_{\hat{x}\hat{x}}^{-1}M_{\hat{x}\hat{\varepsilon}})' M_{\hat{x}\hat{x}}(C - C(\hat{\theta}) - M_{\hat{x}\hat{x}}^{-1}M_{\hat{x}\hat{\varepsilon}})|,$$

where

$$\Sigma(\hat{\theta}) := \left\langle y_t - C(\hat{\theta})\hat{x}_t, y_t - C(\hat{\theta})\hat{x}_t \right\rangle, \quad M_{\hat{\varepsilon}\hat{\varepsilon}} := \left\langle y_t - \hat{C}^{\text{OLS}}\hat{x}_t, y_t - \hat{C}^{\text{OLS}}\hat{x}_t \right\rangle,$$
$$M_{\hat{x}\hat{x}} := \langle \hat{x}_t, \hat{x}_t\rangle, \quad M_{\hat{x}\hat{\varepsilon}} := \left\langle \hat{x}_t, y_t - C(\hat{\theta})\hat{x}_t \right\rangle.$$

Choosing a scaling matrix $D_T := D_T^x G(\hat{\theta})'$ with $G(\hat{\theta}) \to G$ as in Lemma 17 it follows that $D_T M_{\hat{x}\hat{x}} D_T' = D_T^x G' M_{xx} G D_T^x + o_p(1)$, with $M_{xx} := \langle x_t, x_t\rangle$. Moreover, $D_T^x G' M_{xx} G D_T$ converges to a positive definite limit, compare Lemma 17 (ii).
Further, note that $\Sigma(\hat{\theta}) \to \Sigma$ due to Theorem 10. By Lemma 17 (iii) it holds that

$$M_{\hat{x}\hat{\varepsilon}}' D_T' (D_T M_{\hat{x}\hat{x}} D_T')^{-1} D_T M_{\hat{x}\hat{\varepsilon}} = o_p(1),$$

which also implies $M_{\hat{\varepsilon}\hat{\varepsilon}} = \Sigma(\hat{\theta}) - M_{\hat{x}\hat{\varepsilon}}' D_T' (D_T M_{\hat{x}\hat{x}} D_T')^{-1} D_T M_{\hat{x}\hat{\varepsilon}} \to \Sigma$.
Using $|A + x'Bx| \geq |A|(1 + \frac{\lambda_{|min|}(B)}{\lambda_{|max|}(A)}|x|^2)$, which holds for symmetric and positive definite matrices A and B and a vector x of matching dimensions, we get a lower bound of

$$f_T(C) \geq \frac{|M_{\hat{\varepsilon}\hat{\varepsilon}}|}{|\Sigma(\hat{\theta})|}\left(1 + \frac{\lambda_{|min|}(D_T M_{\hat{x}\hat{x}} D_T)}{\lambda_{|max|}(M_{\hat{\varepsilon}\hat{\varepsilon}})}(|D_T^{-1}(C - C(\hat{\theta}))| - |(D_T M_{\hat{x}\hat{x}} D_T)^{-1} D_T M_{\hat{x}\hat{\varepsilon}}|)^2\right),$$

Now, for any $\eta > 0$ and any $\delta > 0$, there exists a constant a, an integer T_0 and a set A_η with $\mathbb{P}(A_\eta) > 1 - \eta$, such that for $T < T_0$ and an outcome in A_η, we have $\lambda_{|min|}(D_T M_{\hat{x}\hat{x}} D_T) \geq a$, $\lambda_{|max|}(M_{\hat{\varepsilon}\hat{\varepsilon}}) \leq c$, $|(D_T M_{\hat{x}\hat{x}} D_T)^{-1} D_T M_{\hat{x}\hat{\varepsilon}}| \leq \delta/2$, $|M_{\hat{\varepsilon}\hat{\varepsilon}}|/|\Sigma(\hat{\theta})| \geq 1 - \frac{a\delta^2/8c}{1 + a\delta^2/4c}$, which implies that

$$f_T(C) \geq 1 + \frac{a\delta^2}{8c}$$

holds for all C such that $|D_T^{-1}(C - C(\hat{\theta}))| \geq \delta$. Hence,

$$\mathbb{P}(|D_T^{-1}(C - C(\hat{\theta}))| \geq \delta) \geq \mathbb{P}(A_\eta) > 1 - \eta$$

for $T \geq T_0$, such that $D_T^{-1}(\hat{C}^{\text{OLS}} - C(\hat{\theta})) \xrightarrow{p} 0$.
(ii) Since $\hat{\theta}$ is consistent, it follows that $\mathcal{C}_{u,\circ}$ is in $\mathbb{U}(c_1, c_2, \hat{\theta})$ with probability one. The estimator $\hat{C}_\circ := \hat{C}(\mathcal{C}_{u,\circ})$ can then be computed using the explicit formula:

$$\hat{C}(\mathcal{C}_{u,\circ}) := \hat{C}^{\text{OLS}} - [\hat{\Pi}\mathcal{C}_{E,\circ}, \hat{\Pi}\mathcal{C}_{2,\circ} + \hat{\Gamma}\mathcal{C}_{1,\circ}]\left(\underline{\hat{B}}(\mathcal{C}_{u,\circ})' \langle \hat{x}_t, \hat{x}_t\rangle^{-1} \underline{\hat{B}}(\mathcal{C}_{u,\circ})\right)^{-1} \underline{\hat{B}}(\mathcal{C}_{u,\circ})' \langle \hat{x}_t, \hat{x}_t\rangle^{-1}.$$

Note that

$$\left(\underline{\hat{B}}(\mathcal{C}_{u,\circ})' \langle \hat{x}_t, \hat{x}_t\rangle^{-1} \underline{\hat{B}}(\mathcal{C}_{u,\circ})\right)^{-1} \underline{\hat{B}}(\mathcal{C}_{u,\circ})' \langle \hat{x}_t, \hat{x}_t\rangle^{-1} \cdot \underline{\hat{B}}(\mathcal{C}_{u,\circ}) = I_{2c_1+c_2},$$
$$\left(\underline{\hat{B}}(\mathcal{C}_{u,\circ})' \langle \hat{x}_t, \hat{x}_t\rangle^{-1} \underline{\hat{B}}(\mathcal{C}_{u,\circ})\right)^{-1} \underline{\hat{B}}(\mathcal{C}_{u,\circ})' \langle \hat{x}_t, \hat{x}_t\rangle^{-1} \cdot \langle \hat{x}_t, \hat{x}_t\rangle \underline{\hat{B}}(\mathcal{C}_{u,\circ})_\perp = 0_{(2c_1+c_2)\times(n-2c_1+c_2)},$$

where the columns of $\left[\underline{\hat{B}}(\mathcal{C}_{u,\circ}), \langle \hat{x}_t, \hat{x}_t\rangle \underline{\hat{B}}(\mathcal{C}_{u,\circ})_\perp\right]$ form a basis of $\mathbb{R}^n$ with probability one. Thus, $\left(\underline{\hat{B}}(\mathcal{C}_{u,\circ})' \langle \hat{x}_t, \hat{x}_t\rangle^{-1} \underline{\hat{B}}(\mathcal{C}_{u,\circ})\right)^{-1} \underline{\hat{B}}(\mathcal{C}_{u,\circ})' \langle \hat{x}_t, \hat{x}_t\rangle^{-1}$ is bounded while $[\hat{\Pi}\mathcal{C}_{E,\circ}, \hat{\Pi}\mathcal{C}_{2,\circ} + \hat{\Gamma}\mathcal{C}_{1,\circ}]$ converges to zero, which implies consistency of $\hat{C}_\circ := \hat{C}(\mathcal{C}_{u,\circ})$.

Existence and consistency of the estimator $\hat{C}$ follows from the arguments given in (i), using $\hat{C}_\circ$ instead of $C(\hat{\theta})$ and $\hat{C}$ instead of $\hat{C}^{\text{OLS}}$. It follows that the corresponding estimator for $\mathcal{C}_u$ in canonical form is consistent for $\mathcal{C}_{u,\circ}$, as well as its normalized $\tilde{\mathcal{C}}_u$ for $\tilde{\mathcal{C}}_{u,\circ}$.
(iii) The super-consistency results hold using the same arguments as in subsection C.2.2. ■

As a last auxiliary step, we summarize the asymptotic behavior of different quantities in the following lemma:

Lemma 19 *Let $\hat{\varepsilon}_t^c$ denote the residuals from the restricted regression and let $D_T^e := diag(T^{-1}I_{c_1}, I_{c_2})$ It holds that*

(i) $\frac{1}{T}\langle \hat{x}_{t,g}, \hat{x}_{t,g}\rangle \to \int_0^1 \boldsymbol{W}_1(u)\boldsymbol{W}_1(u)'du,$

(ii) $\frac{1}{T}D_T^e \langle \hat{x}_{t,e}, \hat{x}_{t,e}\rangle D_T^e \to \int_0^1 \boldsymbol{F}(u)\boldsymbol{F}(u)'du,$

(iii) $\frac{1}{T}\langle \hat{x}_{t,g}, \hat{x}_{t,e}\rangle D_T^e \to \int_0^1 \boldsymbol{W}_1(u)\boldsymbol{F}(u)'du,$

(iv) $\frac{1}{T}D_T^e \left\langle \hat{x}_{t,e}|_{\hat{x}_{t,g}}, \hat{x}_{t,e}|_{\hat{x}_{t,g}}\right\rangle D_T^e \to \int_0^1 \boldsymbol{G}(u)\boldsymbol{G}(u)'du,$

(v) $\langle x_{t,g}, \hat{\mathcal{B}}_{2,\circ}\hat{\varepsilon}_t^c\rangle \to \int_0^1 \boldsymbol{W}_1(u)d\boldsymbol{W}_1(u)',$

(vi) $D_T^e\langle \hat{x}_{t,e}|_{\hat{x}_{t,g}}, \hat{\mathcal{B}}_{E,\circ}\hat{\varepsilon}_t^c\rangle \to \int_0^1 \boldsymbol{G}(u)d\boldsymbol{W}(u)',$

up to multiplication by regular matrices. Here $\boldsymbol{W}_1 = \mathcal{B}_{2,\circ}\Sigma_\circ^{1/2}\boldsymbol{B}$ and $\boldsymbol{W}_2 = \mathcal{B}_{3,\circ}\Sigma_\circ^{1/2}\boldsymbol{B}$ with an s-dimensional standard Brownian motion $\boldsymbol{B}$ on the unit interval, $u \in [0,1]$. Furthermore,

$$\boldsymbol{F}(u) := \begin{pmatrix} \int_0^u \boldsymbol{W}_1(v)dv \\ \boldsymbol{W}_2(u) \end{pmatrix} \qquad \boldsymbol{G}(u) := \begin{pmatrix} \int_0^u \boldsymbol{W}_1(v)dv \\ \boldsymbol{W}_2(u) \end{pmatrix}\Bigg|_{\boldsymbol{W}_1(u)}.$$

Proof: Recall that $(\hat{\mathcal{A}}, \hat{\mathcal{B}}, \hat{\mathcal{C}})$ denotes the system corresponding to $\hat{C}$ from the restricted regression and $\hat{\theta}$ transformed into canonical form and let $(\hat{\mathcal{A}}_\circ, \hat{\mathcal{B}}_\circ, \hat{\mathcal{C}}_\circ)$ denote the true system in canonical form. To prove (i) decompose $\hat{x}_{t-1,g}$ into:

$$\begin{aligned}
\hat{x}_{t-1,g} &= \hat{\mathcal{C}}_1'\hat{\mathcal{C}}_1\hat{x}_{t-1,g} \\
&= \hat{\mathcal{C}}_1'\mathcal{C}_{1,\circ}x_{t-1,g} + \hat{\mathcal{C}}_1'(\boldsymbol{\Delta}y_t - \mathcal{C}_{1,\circ}x_{t-1,g}) - \hat{\mathcal{C}}_1'(\boldsymbol{\Delta}y_t - \hat{\mathcal{C}}_{1,\circ}\hat{x}_{t-1,g}) \\
&= \hat{\mathcal{C}}_1'\mathcal{C}_{1,\circ}x_{t-1,g} + \hat{\mathcal{C}}_1'g_\circ(L)\varepsilon_t - \hat{\mathcal{C}}_1'\hat{g}(L)(\hat{v}_t - \hat{\Psi}_u w_{t,u}) \\
&= \hat{\mathcal{C}}_1'\mathcal{C}_{1,\circ}x_{t-1,g} + \hat{\mathcal{C}}_1'\left(g_\circ(L)\varepsilon_t - \hat{g}(L)\hat{v}_t\right) + \hat{\mathcal{C}}_1'\hat{g}(L)\hat{\Psi}_u w_{t,u},
\end{aligned}$$

where

$$g_\circ(z) := (1-z) + \mathcal{C}_{1,\circ}\mathcal{B}_{1,\circ}z + \mathcal{C}_{2,\circ}\mathcal{B}_{2,\circ}z + \mathcal{C}_{3,\circ}\mathcal{B}_{3,\circ}z + \mathcal{C}_{\bullet,\circ}\sum_{j=1}^\infty \mathcal{A}_{\bullet,\circ}^{j-1}(1-z)z^j\mathcal{B}_{\bullet,\circ}$$

$$\hat{g}(z) := (1-z) + \hat{\mathcal{C}}_1\hat{\mathcal{B}}_1 z + \hat{\mathcal{C}}_2\hat{\mathcal{B}}_2 z + \hat{\mathcal{C}}_3\hat{\mathcal{B}}_3 z + \hat{\mathcal{C}}_\bullet\sum_{j=1}^\infty \hat{\mathcal{A}}_\bullet^{j-1}(1-z)z^j\hat{\mathcal{B}}_\bullet$$

are stable transfer functions,

$$\begin{aligned}
\hat{v}_t := {}& \hat{\mathcal{C}}(I_n - \hat{\underline{\mathcal{A}}})^{-2}\hat{\underline{\mathcal{A}}}^2\sum_{j=1}^\infty \hat{\underline{\mathcal{A}}}^{j-1}\mathcal{B}\boldsymbol{\Delta}^2(z)z^j\boldsymbol{\Delta}^2[\mathcal{C}_{u,\circ}x_{t,u} + \mathcal{C}_{\bullet,\circ}x_{t,\bullet} + \varepsilon_t] \\
&- \hat{\Pi}[\mathcal{C}_{2,\circ}\mathcal{B}_{2,\circ}\varepsilon_{t-1} + \mathcal{C}_{\bullet,\circ}x_{t,\bullet} + \varepsilon_t] \\
&- \hat{\Gamma}[\mathcal{C}_{1,\circ}\mathcal{B}_{1,\circ}\varepsilon_{t-1} + \mathcal{C}_{2,\circ}\mathcal{B}_{2,\circ}\varepsilon_{t-1} + \mathcal{C}_{3,\circ}\mathcal{B}_{3,\circ}\varepsilon_{t-1} + \boldsymbol{\Delta}(\mathcal{C}_{\bullet,\circ}x_{t,\bullet} + \varepsilon_t)] \\
=: {}& \hat{k}^v(L)\varepsilon_t
\end{aligned}$$

is stationary and

$$\hat{\Psi}_u := \begin{bmatrix} \hat{\Pi}\mathcal{C}_{1,\circ} & \hat{\Pi}\mathcal{C}_{2,\circ} + \hat{\Gamma}\mathcal{C}_{1,\circ} & \hat{\Pi}\mathcal{C}_{3,\circ} \end{bmatrix}.$$

The dominant component is $\hat{\mathcal{C}}_1'\mathcal{C}_{1,\circ}x_{t-1,g}$, which is integrated of order one, while $\hat{\mathcal{C}}_1'g_\circ(L)\varepsilon_t$ and $\hat{g}(L)\hat{\upsilon}_t$ are stationary. Note that there is another non-stationary component of the form $\hat{g}(L)\hat{\Psi}_u w_{t,u}$. It holds that $\langle \varepsilon_t, \hat{\Psi}_u w_{t,u}\rangle$ and $\frac{1}{T}\langle x_{t-1,2}, \hat{\Psi}_u w_{t,u}\rangle$ as well as $\langle \hat{\Psi}_u w_{t,u}, \hat{\Psi}_u w_{t,u}\rangle$ are in $o_p(1)$, which holds also if we replace $\hat{\Psi}_u w_{t,u}$ by the filtered series $g(L;\theta)\hat{\Psi}_u w_{t,u}$. Similarly, all terms $\langle\cdot,\cdot\rangle$ containing the stationary component tend to zero, due to the convergence of the transfer function $\hat{g}(z)\hat{k}^\upsilon(z) \to g_\circ(z)$. Thus, noting also that $\hat{\mathcal{C}}_1'\mathcal{C}_{1,\circ}$ converges to I_s, it follows that

$$\begin{aligned} \frac{1}{T}\langle \hat{x}_{t,g}, \hat{x}_{t,g}\rangle &= \frac{1}{T}\langle \hat{\mathcal{C}}_1'\mathcal{C}_{1,\circ}x_{t,g}, \hat{\mathcal{C}}_1'\mathcal{C}_{1,\circ}x_{t,g}\rangle + o_p(1) \\ &= \frac{1}{T}\langle x_{t,g}, x_{t,g}\rangle + o_p(1) \to \int_0^1 \boldsymbol{W}_1(u)\boldsymbol{W}_1(u)'du. \end{aligned}$$

Similarly, decompose $\hat{x}_{t,e}$ into

$$\hat{x}_{t,e} = \hat{\mathcal{C}}_E'\mathcal{C}_{E,\circ}x_{t,e} + \hat{\mathcal{C}}_E'\mathcal{C}_{2,\circ}x_{t,g} + \hat{\mathcal{C}}_E'(y_t - \mathcal{C}_{E,\circ}x_{t,e} - \mathcal{C}_{2,\circ}x_{t,g}) - \hat{\mathcal{C}}_E'\hat{\varepsilon}_t^c - \hat{\mathcal{C}}_E'\hat{\mathcal{C}}_\bullet\hat{x}_{t,\bullet},$$

where we have used that $\hat{\mathcal{C}}_E'\hat{\mathcal{C}}_{2,\circ} = 0$. Since $\hat{\Pi}\mathcal{C}_E = \hat{\alpha}\hat{\beta}'\mathcal{C}_E$ tends to zero as T^{-1}, the same holds for $\hat{\mathcal{C}}_E'\mathcal{C}_2$, ensuring that all terms $\langle\cdot,\cdot\rangle$ containing $\hat{\mathcal{C}}_E'\mathcal{C}_2x_{t,g}$ vanish. The dominating term in $\hat{x}_{t,e}$ is, therefore, equal to $\hat{\mathcal{C}}_E'\mathcal{C}_Ex_{t,e}$ directly implying (ii) and (iii), while (iv) follows by application of (i)-(iii).
To show (v) we first consider

$$\langle \hat{x}_{t,g} - \hat{\mathcal{C}}_1'\mathcal{C}_1x_{t,g}, \hat{\mathcal{B}}_2\hat{\varepsilon}_t^c\rangle = \langle \hat{\mathcal{C}}_1'(g(L,\theta_\circ) - g(L,\hat{\theta})k^\upsilon(z;\hat{\theta}))\varepsilon_{t+1} + \hat{\Psi}_u w_{t+1,u}, \hat{\mathcal{B}}_2\hat{\varepsilon}_t^c\rangle.$$

Since $\hat{g}(z)\hat{k}^\upsilon(z) \to g_\circ(z)$ due to convergence of the PML estimator together with Lemma 18 and the super-consistency of $\hat{\Psi}_u \to 0$, this term is in $o_p(1)$ and it follows that

$$\begin{aligned} \langle \hat{x}_{t,g}, \hat{\mathcal{B}}_2\hat{\varepsilon}_t^c\rangle &= \langle \hat{\mathcal{C}}_1'\mathcal{C}_1x_{t,g}, \hat{\mathcal{B}}_2\hat{\varepsilon}_t^c\rangle + o_p(1) \\ &= \langle x_{t,g}, \hat{\mathcal{B}}_2\hat{\varepsilon}_t^c\rangle + o_p(1) \\ &= \langle x_{t,g}, \mathcal{B}_{2,\circ}\varepsilon_t\rangle + \langle x_{t,g}, \hat{\mathcal{B}}_2(\hat{\varepsilon}_t^c - \varepsilon_t)\rangle + o_p(1). \end{aligned}$$

What is left to show is that the last term is indeed also $o_p(1)$. For this note that $\mathcal{B}_2(\hat{\varepsilon}_t^c - \varepsilon_t)$ may contain integrated components of different integration orders. The term $\hat{\mathcal{B}}_2\hat{\Pi}\mathcal{C}_{1,\circ}x_{t-1,1}$ is integrated of order two. However, this component vanishes since $\hat{\mathcal{B}}_2\hat{\Pi} = 0$. For the other integrated components in $\hat{\mathcal{B}}_2(\hat{\varepsilon}_t^c - \varepsilon_t)$ we find $\hat{\mathcal{B}}_2\hat{\Pi}\mathcal{C}_{3,\circ}x_{t-1,3} = 0$ and $\hat{\mathcal{B}}_2(\hat{\Pi}\mathcal{C}_{2,\circ} + \hat{\Gamma}\mathcal{C}_{1,\circ})\boldsymbol{\Delta}x_{t-1,1} = O_p(T^{-1})$, the latter due to $\hat{\mathcal{B}}_2\hat{\Gamma}\mathcal{C}_{1,\circ} = \hat{\mathcal{B}}_2\hat{\Gamma}(\hat{\mathcal{C}}_E\hat{\mathcal{C}}_E' + I_s - \hat{\mathcal{C}}_E\hat{\mathcal{C}}_E')\mathcal{C}_1 = \hat{\mathcal{B}}_2\hat{\Gamma}(I_s - \hat{\mathcal{C}}_E\hat{\mathcal{C}}_E')\mathcal{C}_{1,\circ} = O_p(T^{-2})$, using that $\hat{\beta}'\mathcal{C}_{1,\circ} = O_p(T^{-2})$. Thus,

$$\langle \hat{x}_{t,g}, \hat{\mathcal{B}}_2\hat{\varepsilon}_t^c\rangle = \langle x_{t,g}, \mathcal{B}_{2,\circ}\varepsilon_t\rangle + \langle x_{t,g}, \hat{\mathcal{B}}_2(\upsilon_t - \varepsilon_t)\rangle + o_p(1) \to \int_0^1 \boldsymbol{W}_1(u)d\boldsymbol{W}_1(u)'.$$

To prove (vi) note that the stationary component in $\hat{x}_{t,e}|_{\hat{x}_{t,g}}$ according to the above decomposition is equal to $\mathcal{C}_E'(h_\circ(L) - \hat{h}(L)\hat{k}^\upsilon(L))\varepsilon_t$, with $h_\circ(z) := I_s + \mathcal{C}_{\bullet,\circ}\sum_{j=1}^\infty \mathcal{A}_{\bullet,\circ}^{j-1}z^j\mathcal{B}_{\bullet,\circ}$ and $\hat{h}(z) := I_s + \hat{\mathcal{C}}_\bullet\sum_{j=1}^\infty \hat{\mathcal{A}}_\bullet^{j-1}z^j\hat{\mathcal{B}}_\bullet$. This term as well as the term $\hat{\mathcal{C}}_E'\hat{h}(L)\hat{\Psi}_u w_{t,u}$ can be treated as in the case of the corresponding components of $\hat{x}_{t,g}$. Thus, consider

$$\begin{aligned} D_T^e\langle \hat{x}_{t,e}|_{\hat{x}_{t,g}}, \hat{\mathcal{B}}_E\hat{\varepsilon}_t^c\rangle &= D_T^e\langle (\hat{\mathcal{C}}_E'\mathcal{C}_{E,\circ}x_{t,e} + \hat{\mathcal{C}}_E'\mathcal{C}_{2,\circ}x_{t,g})|_{x_{t,g}}, \hat{\mathcal{B}}_E\hat{\varepsilon}_t^c\rangle + o_p(1) \\ &= D_T^e\langle x_{t,e}|_{x_{t,g}}, \hat{\mathcal{B}}_E\hat{\varepsilon}_t^c\rangle + o_p(1) \\ &= D_T^e\langle x_{t,e}|_{x_{t,g}}, \mathcal{B}_{E,\circ}\varepsilon_t\rangle + D_T^e\langle x_{t,e}|_{x_{t,g}}, \hat{\mathcal{B}}_E(\hat{\varepsilon}_t^c - \varepsilon_t)\rangle + o_p(1) \end{aligned}$$

The convergence of the last term follows by considering the different integrated components. It follows that $D_T^e\langle x_{t,e}|_{x_{t,g}}, \hat{\mathcal{B}}_E\hat{\Pi}\mathcal{C}_{1,\circ}x_{t-1,1}\rangle$ and $D_T^e\langle x_{t,e}|_{x_{t-1,g}}, \hat{\mathcal{B}}_E\hat{\Pi}\mathcal{C}_{3,\circ}x_{t,3}\rangle$ are equal to zero by the arguments given in (v). Using (i) and (ii) one can also replace the regression on $\hat{x}_{t,g}$ by a regression on the true state component $x_{t,g}$. It, therefore, holds that

$$\begin{aligned} D_T^e\langle x_{t,e}|_{x_{t,g}}, \hat{\mathcal{B}}_E\hat{\Gamma}\mathcal{C}_{1,\circ}x_{t-2,1}\rangle &= D_T^e\langle x_{t,e}|_{x_{t,g}}, \hat{\mathcal{B}}_E\hat{\Gamma}\mathcal{C}_1 x_{t-2,g}\rangle + o_p(1) \\ &= D_T^e\langle x_{t,e}, \hat{\mathcal{B}}_E\hat{\Gamma}\mathcal{C}_1 x_{t-2,g}|_{x_{t,g}}\rangle + o_p(1) = o_p(1). \end{aligned}$$

All in all

$$D_T^e\left\langle \hat{x}_{t,e}|_{\hat{x}_{t,g}}, \hat{\mathcal{B}}_{E,\circ}\hat{\varepsilon}_t^c\right\rangle = \left\langle x_{t,e}|_{x_{t,g}}, \mathcal{B}_E\varepsilon_t\right\rangle + \left\langle x_{t,e}|_{x_{t,g}}, \hat{\mathcal{B}}_E(v_t - \varepsilon_t)\right\rangle + o_p(1) \to \int_0^1 \boldsymbol{G}(u)d\boldsymbol{W}(u),$$

which finishes the proof also for (vi). ■

C.4 Proof of Theorem 12

Proof: The logarithm of the likelihood ratio is given by

$$\begin{aligned} -2\log Q(H(c_1,c_2)/H_\bullet,\hat{\theta}) &= -T\log\det(\langle\hat{\varepsilon}_t^c,\hat{\varepsilon}_t^c\rangle\langle\hat{\varepsilon}_t,\hat{\varepsilon}_t\rangle^{-1}) \\ &= T\log\det(\langle\hat{\varepsilon}_t,\hat{\varepsilon}_t\rangle\langle\hat{\varepsilon}_t^c,\hat{\varepsilon}_t^c\rangle^{-1}) \\ &= T\log\det\left[(\langle\hat{\varepsilon}_t^c,\hat{\varepsilon}_t^c\rangle + \langle\hat{\varepsilon}_t^c-\hat{\varepsilon}_t,\hat{\varepsilon}_t^c-\hat{\varepsilon}_t\rangle)\langle\hat{\varepsilon}_t^c,\hat{\varepsilon}_t^c\rangle^{-1}\right] \\ &= T\log\det\left[I_s + \langle\hat{\varepsilon}_t^c-\hat{\varepsilon}_t,\hat{\varepsilon}_t^c-\hat{\varepsilon}_t\rangle\langle\hat{\varepsilon}_t^c,\hat{\varepsilon}_t^c\rangle^{-1}\right]. \end{aligned}$$

Since the second term tends to zero, a Taylor expansion of the likelihood leads to the following representation

$$-2\log Q(H(c_1,c_2)/H_\bullet,\hat{\theta}) = T\cdot\mathrm{tr}\left(\langle\hat{\varepsilon}_t^c-\hat{\varepsilon}_t,\hat{\varepsilon}_t^c-\hat{\varepsilon}_t\rangle\langle\hat{\varepsilon}_t^c,\hat{\varepsilon}_t^c\rangle^{-1}\right) + o_p(1).$$

Lemma 16 (i) implies that the difference $\hat{\varepsilon}_t^c - \hat{\varepsilon}_t$ is given by $\langle\hat{\varepsilon}_t^c,\hat{x}_{t,u}\rangle\left\langle\hat{x}_{t,u}|_{\hat{x}_{t,\bullet}},\hat{x}_{t,u}|_{\hat{x}_{t,\bullet}}\right\rangle^{-1}\hat{x}_{t,u}|_{\hat{x}_{t,\bullet}}$, where $\hat{x}_{t,u}|_{\hat{x}_{t,\bullet}}$ are the residual of the regression of $\hat{x}_{t,u}$ on $\hat{x}_{t,\bullet}$. Consequently, we can transform

$$\begin{aligned} &T\cdot\mathrm{tr}\left(\langle\hat{\varepsilon}_t^c-\hat{\varepsilon}_t,\hat{\varepsilon}_t^c-\hat{\varepsilon}_t\rangle\langle\varepsilon_t^c,\varepsilon_t^c\rangle^{-1}\right) \\ &= T\cdot\mathrm{tr}\left(\langle\hat{\varepsilon}_t^c,\hat{x}_{t,u}\rangle\left\langle\hat{x}_{t,u}|_{\hat{x}_{t,\bullet}},\hat{x}_{t,u}|_{\hat{x}_{t,\bullet}}\right\rangle^{-1}\langle\hat{x}_{t,u},\hat{\varepsilon}_t^c\rangle\langle\hat{\varepsilon}_t^c,\hat{\varepsilon}_t^c\rangle^{-1}\right) \\ &= T\cdot\mathrm{tr}\left(\langle\hat{\varepsilon}_t^c,\hat{x}_{t,u}\rangle\langle\hat{x}_{t,u},\hat{x}_{t,u}\rangle^{-1}\langle\hat{x}_{t,u},\hat{\varepsilon}_t^c\rangle\langle\hat{\varepsilon}_t^c,\hat{\varepsilon}_t^c\rangle^{-1}\right) + o_p(1) \end{aligned}$$

since $\langle\hat{\varepsilon}_t^c,\hat{\varepsilon}_t^c\rangle$ and $\langle\hat{\varepsilon}_t,\hat{\varepsilon}_t\rangle$ converge to the true $\Sigma_\circ$ and the terms in $\left\langle\hat{x}_{t,u}|_{\hat{x}_{t,\bullet}},\hat{x}_{t,u}|_{\hat{x}_{t,\bullet}}\right\rangle$ are dominated by $\langle\hat{x}_{t,u},\hat{x}_{t,u}\rangle$. Consequently,

$$\begin{aligned} &T\cdot\mathrm{tr}\left(\langle\hat{\varepsilon}_t^c,\hat{x}_{t,u}\rangle\langle\hat{x}_{t,u},\hat{x}_{t,u}\rangle^{-1}\langle\hat{x}_{t,u},\hat{\varepsilon}_t^c\rangle\langle\hat{\varepsilon}_t^c,\hat{\varepsilon}_t^c\rangle^{-1}\right) \\ &= T\cdot\mathrm{tr}\left(\langle\hat{\varepsilon}_t^c,\hat{\varepsilon}_t^c\rangle^{-1/2}\langle\hat{\varepsilon}_t^c,\hat{x}_{t,u}\rangle\langle\hat{x}_{t,u},\hat{x}_{t,u}\rangle^{-1}\langle\hat{x}_{t,u},\hat{\varepsilon}_t^c\rangle\langle\hat{\varepsilon}_t^c,\hat{\varepsilon}_t^c\rangle^{-1/2}\right) \\ &= T\cdot\mathrm{tr}\left(\langle\hat{\varepsilon}_t^c,\hat{\varepsilon}_t^c\rangle^{-1/2}\langle\hat{\varepsilon}_t^c,\hat{x}_{t,e}|_{\hat{x}_{t,g}}\rangle\left\langle\hat{x}_{t,e}|_{\hat{x}_{t,g}},\hat{x}_{t,e}|_{\hat{x}_{t,g}}\right\rangle^{-1}\left\langle\hat{x}_{t,e}|_{\hat{x}_{t,g}},\hat{\varepsilon}_t^c\right\rangle\langle\hat{\varepsilon}_t^c,\hat{\varepsilon}_t^c\rangle^{-1/2}\right) \\ &\quad + T\cdot\mathrm{tr}\left(\langle\hat{\varepsilon}_t^c,\hat{\varepsilon}_t^c\rangle^{-1/2}\langle\hat{\varepsilon}_t^c,\hat{x}_{t,g}\rangle\langle\hat{x}_{t,g},\hat{x}_{t,g}\rangle^{-1}\langle\hat{x}_{t,g},\hat{\varepsilon}_t^c\rangle\langle\hat{\varepsilon}_t^c,\hat{\varepsilon}_t^c\rangle^{-1/2}\right) \\ &=: T\cdot\mathrm{tr}(M_1) + T\cdot\mathrm{tr}(M_2). \end{aligned}$$

Define $\tilde{v}_{\hat{\alpha}} := \hat{\alpha}'_{\perp} \langle \hat{\varepsilon}^c_t, \hat{\varepsilon}^c_t \rangle^{1/2}$ and $v_{\hat{\alpha}} := (\tilde{v}_{\hat{\alpha}}\tilde{v}'_{\hat{\alpha}})^{-1/2}\tilde{v}_{\hat{\alpha}}$, such that $v_{\hat{\alpha}}v'_{\hat{\alpha}} = I_c$. Then $\tilde{w}_{\hat{\alpha}} := \hat{\alpha}' \langle \hat{\varepsilon}^c_t, \hat{\varepsilon}^c_t \rangle^{-1/2}$ is orthogonal to $\tilde{v}_{\hat{\alpha}}$. Similarly we set $w_{\hat{\alpha}} := (\tilde{w}_{\hat{\alpha}}\tilde{w}'_{\hat{\alpha}})^{-1/2}\tilde{w}_{\hat{\alpha}}$. Then

$$\operatorname{tr}(I_s M_1) = \operatorname{tr}(v'_{\hat{\alpha}}v_{\hat{\alpha}}M_1) + \operatorname{tr}(w'_{\hat{\alpha}}w_{\hat{\alpha}}M_1) = \operatorname{tr}(v_{\hat{\alpha}}M_1v'_{\hat{\alpha}}) + \operatorname{tr}(w_{\hat{\alpha}}M_1w'_{\hat{\alpha}}).$$

The second term is zero since

$$\begin{aligned}
&\operatorname{tr}(w_{\hat{\alpha}}M_1w'_{\hat{\alpha}}) = \operatorname{tr}(\tilde{w}_{\hat{\alpha}}M_1\tilde{w}'_{\hat{\alpha}}(\tilde{w}'_{\hat{\alpha}}\tilde{w}_{\hat{\alpha}}))\\
&= \operatorname{tr}\left(\hat{\alpha}' \langle \hat{\varepsilon}^c_t, \hat{\varepsilon}^c_t \rangle^{-1} \langle \hat{\varepsilon}^c_t, \hat{x}_{t,e}|_{\hat{x}_{t,g}} \rangle \langle \hat{x}_{t,e}|_{\hat{x}_{t,g}}, \hat{x}_{t,e}|_{\hat{x}_{t,g}} \rangle^{-1} \langle \hat{x}_{t,e}|_{\hat{x}_{t,g}}, \hat{\varepsilon}^c_t \rangle \langle \hat{\varepsilon}^c_t, \hat{\varepsilon}^c_t \rangle^{-1} \hat{\alpha}\tilde{w}'_{\hat{\alpha}}(\tilde{w}'_{\hat{\alpha}}\tilde{w}_{\hat{\alpha}})^{-1}\right)
\end{aligned}$$

and we have $\hat{\alpha}' \langle \hat{\varepsilon}^c_t, \hat{\varepsilon}^c_t \rangle^{-1} \langle \hat{\varepsilon}^c_t, x_{t,e} \rangle = 0$ by Lemma 16 (ii). Thus, $\operatorname{tr}(w_{\hat{\alpha}}M_1w'_{\hat{\alpha}}) = 0$. Similarly

$$\operatorname{tr}\left(\tilde{w}_{\hat{\gamma}}M_2\tilde{w}'_{\hat{\gamma}}(\tilde{w}'_{\hat{\gamma}}\tilde{w}_{\hat{\gamma}})\right) = 0,$$

where $\tilde{w}_{\hat{\gamma}} = \hat{\gamma}' \langle \hat{\varepsilon}^c_t, \hat{\varepsilon}^c_t \rangle^{-1/2}$ by Lemma 16 (iii). Defining $\tilde{v}_{\hat{\gamma}} := \hat{\gamma}'_{\perp} \langle \hat{\varepsilon}^c_t, \hat{\varepsilon}^c_t \rangle^{1/2}$ and $v_{\hat{\gamma}} := (\tilde{v}_{\hat{\gamma}}\tilde{v}'_{\hat{\gamma}})^{-1/2}\tilde{v}_{\hat{\gamma}}$, it follows that

$$\begin{aligned}
-2\log Q(H(c_1,c_2)/H_\bullet) &= T\cdot\operatorname{tr}(v_{\hat{\alpha}}M_1v'_{\hat{\alpha}}) + T\cdot\operatorname{tr}\left(v_{\hat{\gamma}}M_2v'_{\hat{\gamma}}\right) + o_p(1)\\
&= T\cdot\operatorname{tr}\left(\langle \hat{\alpha}'_{\perp}\hat{\varepsilon}^c_t, \hat{\alpha}'_{\perp}\hat{\varepsilon}^c_t \rangle^{-1} \langle \hat{\alpha}'_{\perp}\hat{\varepsilon}^c_t, x_{t,e}|_{x_{t,g}} \rangle \langle x_{t,e}|_{x_{t,g}}, x_{t,e}|_{x_{t,g}} \rangle^{-1} \langle x_{t,e}|_{x_{t,g}}, \hat{\alpha}'_{\perp}\hat{\varepsilon}^c_t \rangle\right)\\
&\quad + T\cdot\operatorname{tr}\left(\langle \hat{\gamma}'_{\perp}\hat{\varepsilon}^c_t, \hat{\gamma}'_{\perp}\hat{\varepsilon}^c_t \rangle^{-1} \langle \hat{\gamma}'_{\perp}\hat{\varepsilon}^c_t, x_{t,g} \rangle \langle x_{t,g}, x_{t,g} \rangle^{-1} \langle x_{t,g}, \hat{\gamma}'_{\perp}\hat{\varepsilon}^c_t \rangle\right) + o_p(1)\\
&= T\cdot\operatorname{tr}\left(\left\langle \hat{\mathcal{B}}_E\hat{\varepsilon}^c_t, \hat{\mathcal{B}}_E\hat{\varepsilon}^c_t \right\rangle^{-1} \left\langle \hat{\mathcal{B}}_E\hat{\varepsilon}^c_t, x_{t,e}|_{x_{t,g}} \right\rangle \langle x_{t,e}|_{x_{t,g}}, x_{t,e}|_{x_{t,g}} \rangle^{-1} \left\langle x_{t,e}|_{x_{t,g}}, \hat{\mathcal{B}}_E\hat{\varepsilon}^c_t \right\rangle\right)\\
&\quad + T\cdot\operatorname{tr}\left(\left\langle \hat{\mathcal{B}}_2\hat{\varepsilon}^c_t, \hat{\mathcal{B}}_2\hat{\varepsilon}^c_t \right\rangle^{-1} \left\langle \hat{\mathcal{B}}_2\hat{\varepsilon}^c_t, x_{t,g} \right\rangle \langle x_{t,g}, x_{t,g} \rangle^{-1} \left\langle x_{t,g}, \hat{\mathcal{B}}_2\hat{\varepsilon}^c_t \right\rangle\right) + o_p(1).
\end{aligned}$$

Finally, Lemma 19 implies:

$$T\cdot(v_\alpha M_1 v'_\alpha) + T\cdot\operatorname{tr}\left(v_\gamma M_2 v'_\gamma\right) \to Q^\infty_r + Q^\infty_{r,s},$$

which proves the result in Theorem 12 (i) for $-2\log Q(H(c_1,c_2)/H_\bullet, \hat{\theta})$.
Next, the pseudo likelihood ratio rank test is investigated. Let $\hat{\theta}^{c_1,c_2}_n$ denote the PML estimator over $\mathbf{\Theta}^{c_1,c_2}_n$ and $\hat{\theta}_n$ denote the PML estimator over $\mathbf{\Theta}_n$. For $\mathcal{D}_t = 0$ we have $\theta_d = 0$, thus, $\hat{\varphi}^{c_1,c_2}_n = [\hat{\theta}^{c_1,c_2}_n, 0]$ and $\hat{\varphi}_n = [\hat{\theta}_n, 0]$. The following inequalities hold:

$$L^{\mathcal{C}}_T(\tilde{\mathcal{C}}_u(\hat{\theta}_n), \hat{\theta}_n) \le \mathcal{L}^{\varphi}_T(\hat{\varphi}^{c_1,c_2}_n) \le L^{ex}_T(\hat{C}^{\mathrm{OLS}}, \hat{\theta}^{c_1,c_2}_n) \le \mathcal{L}^{\varphi}_T(\hat{\varphi}_n),$$

where $\tilde{\mathcal{C}}_u(\hat{\theta}_n)$ is a (normalized) maximizer of the function $L^{\mathcal{C}}_T(\mathcal{C}_u, \hat{\theta}_n)$ over $\mathbb{U}(c_1, c_2, \hat{\theta}_n)$. The first inequality holds since $\hat{\varphi}^{c_1,c_2}_n$ maximizes the pseudo log-likelihood function over all systems with the specified structure. The second inequality holds because dropping the restrictions can only increase the pseudo log-likelihood function. The last inequality holds due to $\hat{\varphi}_n$ being the maximizer over all systems of order n. These inequalities imply

$$\begin{aligned}
-2\log Q(H(c_1,c_2)/H_\bullet, \hat{\theta}^{c_1,c_2}_n) &= -2\left(\mathcal{L}^{\varphi}_T(\hat{\varphi}^{c_1,c_2}_n) - L^{ex}_T(\hat{C}^{\mathrm{OLS}}, \hat{\theta}^{c_1,c_2}_n)\right)\\
&\le -2\left(\mathcal{L}^{\varphi}_T(\hat{\varphi}^{c_1,c_2}_n) - \mathcal{L}^{\varphi}_T(\hat{\varphi}_n)\right)\\
&\le -2\left(L^{\mathcal{C}}_T(\tilde{\mathcal{C}}_u(\hat{\theta}_n), \hat{\theta}_n) - \mathcal{L}^{\varphi}_T(\hat{\varphi}_n)\right) = -2\log Q(H(c_1,c_2)/H_\bullet, \hat{\theta}_n).
\end{aligned}$$

The evaluations above show that the limit of the left hand side and the one of the right hand side coincide. Thus, the bounds imply that also the pseudo likelihood ratio in the middle converges to the same limit.
In order to derive the asymptotics of the test statistics in (ii) and (iii), note that in the case of a constant and no linear trend it holds that $\hat{\alpha}' \left\langle \hat{\varepsilon}^d_t, \hat{\varepsilon}^d_t \right\rangle^{-1} \left\langle \hat{\varepsilon}^d_t, 1 \right\rangle = 0$, where $\hat{\varepsilon}^d_t$ denotes the residual

in the corresponding restricted model with additional constant term. In the case of a constant and a linear trend $\hat{\alpha}' \left\langle \hat{\varepsilon}_t^{de}, \hat{\varepsilon}_t^{de} \right\rangle^{-1} \left\langle \hat{\varepsilon}_t^{de}, [1,t]' \right\rangle = 0$ and $\hat{\gamma}' \left\langle \hat{\varepsilon}_t^{de}, \hat{\varepsilon}_t^{de} \right\rangle^{-1} \left\langle \hat{\varepsilon}_t^{de}, 1 \right\rangle = 0$, where $\hat{\varepsilon}_t^{de}$ denotes the residual in the corresponding restricted model with additional constant and linear trend term. Both results are proven along the same lines as the results in Lemma 16. The derivation of the asymptotic distribution of the test statistics follows analogously, using the Taylor expansion of the test statistics, which leads to

$$\begin{aligned}
-2\log Q(H^d(c_1,c_2)/H_\bullet,\hat{\theta}) &= -T \log\det\left[\left\langle \hat{\varepsilon}_t^d, \hat{\varepsilon}_t^d \right\rangle \left\langle \hat{\varepsilon}_t, \hat{\varepsilon}_t \right\rangle^{-1}\right] \\
&= T \cdot \mathrm{tr}\left(\left\langle \hat{\varepsilon}_t^d, \hat{\varepsilon}_t^d \right\rangle^{-1} \left\langle \hat{\varepsilon}_t^d, \hat{x}_{t,ed}|_{\hat{x}_{t,g}} \right\rangle \left\langle \hat{x}_{t,ed}|_{\hat{x}_{t,g}}, \hat{x}_{t,ed}|_{\hat{x}_{t,g}} \right\rangle^{-1} \left\langle \hat{x}_{t,ed}|_{\hat{x}_{t,g}}, \hat{\varepsilon}_t^c \right\rangle\right) \\
&\quad + T \cdot \mathrm{tr}\left(\left\langle \hat{\varepsilon}_t^d, \hat{\varepsilon}_t^d \right\rangle^{-1} \left\langle \hat{\varepsilon}_t^d, \hat{x}_{t,g} \right\rangle \left\langle \hat{x}_{t,g}, \hat{x}_{t,g} \right\rangle^{-1} \left\langle \hat{x}_{t,g}, \hat{\varepsilon}_t^d \right\rangle\right) + o_p(1),
\end{aligned}$$

where $x_{t,ed} := [x_{t,e}', 1]'$ if only a constant is included and to

$$\begin{aligned}
-2\log Q(H^{de}(c_1,c_2)/H_\bullet,\hat{\theta}) &= -T \log\det\left[\left\langle \hat{\varepsilon}_t^{de}, \hat{\varepsilon}_t^{de} \right\rangle \left\langle \hat{\varepsilon}_t, \hat{\varepsilon}_t \right\rangle^{-1}\right] \\
&= T \cdot \mathrm{tr}\left(\left\langle \hat{\varepsilon}_t^{de}, \hat{\varepsilon}_t^{de} \right\rangle^{-1} \left\langle \hat{\varepsilon}_t^{de}, \hat{x}_{t,ee}|_{\hat{x}_{t,gd}} \right\rangle \left\langle \hat{x}_{t,ee}|_{\hat{x}_{t,gd}}, \hat{x}_{t,ee}|_{\hat{x}_{t,gd}} \right\rangle^{-1} \left\langle \hat{x}_{t,ee}|_{\hat{x}_{t,gd}}, \hat{\varepsilon}_t^{de} \right\rangle\right) \\
&\quad + T \cdot \mathrm{tr}\left(\left\langle \hat{\varepsilon}_t^{de}, \hat{\varepsilon}_t^{de} \right\rangle^{-1} \left\langle \hat{\varepsilon}_t^{de}, \hat{x}_{t,gd} \right\rangle \left\langle \hat{x}_{t,gd}, \hat{x}_{t,gd} \right\rangle^{-1} \left\langle \hat{x}_{t,gd}, \hat{\varepsilon}_t^{de} \right\rangle\right) + o_p(1),
\end{aligned}$$

where $x_{t,ee} := [x_{t,ee}', t]'$ and $x_{t,gd} := [x_{t,g}', 1]'$ if a constant and a linear trend are included. The rest follows as in Lemma 19 and the respective convergence results from Lemma 10. Convergence of the pseudo likelihood ratio then follows by the same arguments used in the case without deterministics.
■

Bibliography

Amann, H. and Escher, J. (2008). *Analysis III*, Birkhäuser Basel, Basel.

Aoki, M. (1990). *State Space Modeling of Time Series*, Springer, New York.

Bauer, D. (2001). Order Estimation for Subspace Procedures, *Automatica* **37**(10): 1561–1573.

Bauer, D. and Buschmeier, R. (2016). Asymptotic Properties of Subspace Methods for the Estimation of Seasonally Cointegrated Models.

Bauer, D., Matuschek, L., de Matos Ribeiro, P. and Wagner, M. (2020). A Parameterization of Models for Unit Root Processes: Structure Theory and Hypothesis Testing, *Econometrics* **8**(4): 42.

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. 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., Bauer, D., Matuschek, L. 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.

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.

Hylleberg, S., Engle, R. F., Granger, C. W. and Yoo, B. S. (1990). Seasonal Integration and Cointegration, *Journal of Econometrics* **44**(1–2): 215–238.

Hylleberg, S., Jorgensen, C. and Sorensen, N. (1993). Seasonality in Macroeconomic Time Series, *Empirical Economics* **18**(2): 321–335.

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. (1992). A Representation of Vector Autoregressive Processes Integrated of Order 2, *Econometric Theory 8* **8**(2): 188–202.

Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Auto-Regressive Models*, Oxford University Press, Oxford.

Johansen, S. (1996). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Advanced Texts in Econometrics, Oxford University Press, New York.

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.

Lewis, R. and Reinsel, G. C. (1985). Prediction of Multivariate Time Series by Autoregressive Model Fitting, *Journal of Multivariate Analysis* **16**(3): 393–411.

Lof, M. and Franses, P. (2001). On Forecasting Cointegrated Seasonal Time Series, *International Journal of Forecasting* **17**(4): 607–621.

Lütkepohl, H. and Claessen, H. (1997). Analysis of Cointegrated VARMA Processes, *Journal of Econometrics* **80**(2): 223–239.

Nielsen, H. B. and Rahbek, A. (2007). The Likelihood Ratio Test for Cointegration Ranks in the I(2) Model, *Econometric Theory* **23**(4): 615–637.

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. (2003). On the Specification of Cointegrated Autoregressive Moving-Average Forecasting Systems, *International Journal of Forecasting* **19**(3): 503–519.

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 Model, *Econometric Theory* **9**: 155–188.

Saikkonen, P. (1995). Problems With the Asymptotic Theory of Maximum Likelihood Estimation in Integrated and Cointegrated Systems, *Econometric Theory* **11**.

Saikkonen, P. and Luukkonen, R. (1997). Testing Cointegration in Infinite Order Vector Autoregressive Processes, *Journal of Econometrics* **81**(1): 93–126.

Sims, C., Stock, J. and Watson, M. (1990). Inference in Linear Time Series Models with Some Unit Roots, *Econometrica* **58**: 113–144.

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.

Zellner, A. and Palm, F. C. (1974). Time Series Analysis and Simultaneous Equation Econometric Models, *Journal of Econometrics* **2**(1): 17–54.

Erklärung zu den Beiträgen der einzelnen Autoren zu den Artikeln

Die vorliegende Dissertation besteht aus den drei Artikeln,

- Bauer, D., de Matos Ribeiro, P., Matuschek, L. and Wagner, M. (2020). A Parameterization of Models for Unit Root Processes: Structure Theory and Hypothesis Testing, *Econometrics* **8**(4): 42.
- Matuschek, L., Bauer, D., de Matos Ribeiro, P. and Wagner, M. (2021). Inference on Cointegrating Ranks and Spaces of Multiple Frequency I(1) Processes: A State Space Approach, *Mimeo. (unveröffentlicht)*
- Matuschek, L., Bauer, D., de Matos Ribeiro, P. and Wagner, M. (2021). Pseudo Maximum Likelihood Estimation and Inference for I(2) Processes: A State Space Approach, *Mimeo. (unveröffentlicht)*

die jeweils von Professor Dietmar Bauer, Professor Martin Wagner, Patrick de Matos Ribeiro und mir verfasst wurden. Der erstgenannte Artikel ist gleichzeitig auch Bestandteil der Dissertation von Patrick de Matos Ribeiro, die im Rahmen des gemeinsamen Projektes ebenfalls an der Fakultät Statistik der TU Dortmund eingereicht wurde.

Die Resultate des ersten Artikels "A Parameterization of Models for Unit Root Processes: Structure Theory and Hypothesis Testing" stammen von Patrick de Matos Ribeiro und mir. Ausgehend davon verfasste Dietmar Bauer einen Entwurf, bei dem die Gliederung des Artikels entstand. Bei der anschließenden Überarbeitung im Rahmen zweier Revisionen schrieb Martin Wagner die Einleitung. Die Revisionen der übrigen Abschnitte wurden in enger Abstimmung und Diskussion mit Martin Wagner durchgeführt, Abschnitte zwei und drei primär von Patrick de Matos Ribeiro, Abschnitte vier und fünf primär von mir. (geschätzte Anteile: MW 20%, DB 20%, PR 30%, LM 30%)

Die erste Version des zweiten Artikels "Inference on Cointegrating Ranks and Spaces of Multiple Frequency I(1) Processes: A State Space Approach" wurde wiederum von Patrick de Matos Ribeiro und mir in Zusammenarbeit mit Martin Wagner erstellt, wobei die Resultate und Beweise vorwiegend von mir stammen, während die Teile zur Simulationsstudie von Patrick de Matos Ribeiro erstellt wurden. Nach einer grundlegenden Überarbeitung von Dietmar Bauer, wurden die Resultate und entsprechenden Herleitungen von mir schließlich in einem dritten Schritt in Notation neu gefasst, präzisiert und um fehlende Bestandteile erweitert. In weiteren Iterationen zu bestimmten Beweisen und Textpassagen mit Dietmar Bauer, Patrick de Matos Ribeiro und schließlich Martin Wagner ist dann die endgültige Version entstanden. Der für die Tests verwendete Code wurde von mir unter Mithilfe von Patrick de Matos Ribeiro erstellt und verwendet auch Code von Dietmar Bauer zur Schätzung des Startwertes in der Optimierung. (geschätzte Anteile: MW 10%, DB 20%, PR 10%, LM 60%)

Der dritte Artikel schließlich, "Pseudo Maximum Likelihood Estimation and Inference for I(2) Processes: A State Space Approach", wurde auf Basis der erarbeiteten Resultate zu I(1) Prozessen von mir verfasst, einschließlich der Resultate und Beweise, des notwendigen Codes und der Simulationsstudie. Anschließend hat Dietmar Bauer hilfreiche Anmerkung und Korrekturvorschläge eingebracht. (geschätzte Anteile: DB 10%, LM 90%)

Dortmund, den 29.09.2020,
Lukas Matuschek

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Dissertation ist bisher keiner anderen Fakultät vorgelegt worden. Ich erkläre, dass ich bisher kein Promotionsverfahren erfolglos beendet habe und dass keine Aberkennung eines bereits erworbenen Doktorgrades vorliegt.

Lukas Matuschek