Estimation of Cointegrated Multivariate Continuous-Time Autoregressive Moving Average Processes

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

an der Fakultät für Mathematik des
Karlsruher Instituts für Technologie (KIT)
genehmigte

DISSERTATION

von

Dipl.-Math. oec. Markus Scholz

aus Bad Mergentheim

Tag der mündlichen Prüfung: 22. Juni 2016

Referent: Prof. Dr. Vicky Fasen-Hartmann
Korreferent: Prof. Dr. Robert Stelzer

Abstract

We extend in this thesis the cointegrated discrete-time VAR model, which was introduced by Engle and Granger, to a continuous-time setting using cointegrated multivariate continuous-time autoregressive moving average (MCARMA) processes. The concept of cointegration describes the phenomenon, that two or more non-stationary processes, which are integrated, can have stationary linear combinations. Cointegration therefore models stochastic trends of some or all the variables. There is empirical evidence that cointegration arises e.g. in financial data.

We derive a canonical representation for cointegrated MCARMA processes and investigate its properties. Moreover, we derive similar results to the Johansen-Granger Representation Theorem in this thesis. A question that imposes itself in this framework is how to estimate the parameters of the cointegrated MCARMA model from discrete-time observations. Since the necessary uniform convergence results do not hold for the log-likelihood function, we use a stepwise approach. For this reason, we separate the parameter vector into two vectors, where the parameters in the first vector model the cointegration space and the parameters in the second vector model the stationary part. To this end, we show super-consistency for the estimator of the cointegration parameters. In the next step, we establish the consistency for the estimator of the stationary parameters. Moreover, we derive the limiting distributions of the estimators. Lastly, we present a simulation study in order to demonstrate the applicability of the estimation procedure.

Besides, we also consider a decomposition of stationary MCARMA processes into multivariate Ornstein-Uhlenbeck processes. With the help of this decomposition we derive a weak VARMA representation of the sampled process and the integrated sequence of MCARMA processes. Last but not least, we analyze the covariance structure of these representations.

Acknowledgements

First of all, special thanks go to my advisor Prof. Dr. Vicky Fasen for giving me the opportunity to write this thesis at the Institute for Stochastics. I want to express my sincere appreciation for all her helpful advice, encouragement and excellent guidance during the last three years of my doctoral research. Moreover, I am grateful to Prof. Dr. Robert Stelzer for serving as a second referee of this thesis.

I would like to thank all members of the Institute for Stochastics who made the last years an unforgettable part of my life. Especially, I would like to thank the colleagues who shared an office with me for countless great moments.

Financial support by the Deutsche Forschungsgemeinschaft through the DFG project „Statistik Lévy getriebener Modelle " is gratefully acknowledged by me.

Last but not least, I want to say a special thank you to my family and friends for their constant support and motivation. In particular, I want to express my special gratitude to Julia Mensdorf who is always there for me when I need her.

Contents

CHAPTER 1

Introduction

1.1. Historical Overview and Motivation

A typical problem in statistical applications is to analyze time series data and set up a model which fits the data sufficiently. For this purpose, the model should not only be quite general but also tractable. Furthermore, a model which reproduces the stylized facts, namely simplified presentation of empirical findings, of the considered time series is preferable. Examples for stylized facts are probabilistic properties but also sample path properties. If for example the data exhibits jumps, an appropriate model should incorporate jumps in the sample path as well. Besides, a model should not be to complex in order to have a good understanding of its properties, which, however, restricts the cases of possible applications. After having chosen a suitable model one typically wants to estimate the model parameters.

Since realizations in the future are often unpredictable, one generally assumes that an observation y_t is a realization of a random variable Y_t. Thus, it is common practice that data is modelled by a stochastic process $(Y_t)_{t \in I}$. Stochastic processes can be categorized in continuous-time and discrete-time processes depending on the nature of the index set I. In order to find a proper model for the data set, one must first choose one category. Even if observations are often made at discrete time points the underlying phenomena might be continuous in time. For example, in physics there are many continuous-time phenomena, like wind speed, water level, temperature and so forth, which are only observed at discrete time points. Apart from that, a continuous-time model might lead to beneficial properties as in the case of option pricing where the Black-Scholes model became standard. Other advantages of continuous-time models are that one can have irregularly spaced or high frequency data. A class of stochastic processes which is not only quite flexible but has also useful properties are continuous-time autoregressive moving average (CARMA) processes. We are going to use the multivariate versions, the so-called MCARMA processes, as the central process for our model in this thesis. A multivariate model has the advantage that it provides a joint model for several variables including the dependencies of these variables in comparison to individual models for each variable.

An important property of stochastic processes is stationarity. Loosely said, stationarity means that the statistical properties of the stochastic process do not change over time. Stationarity plays a major role in time series analysis due to its benefit for prediction. There is, however, empirical evidence that many time series in econometrics and finance are non-stationary, though their first differences are stationary. The first difference of a stochastic process $(Y_n)_{n\in\mathbb{Z}}$ is given by $\Delta Y_n := Y_n - Y_{n-1}$. The same is also true for many other fields of application. Since standard methods applied to non-stationary time series give spurious results, it is of great importance to have a class of models which captures the non-stationarity but also enables a broad statistical analysis. A particular subclass of non-stationary processes with useful properties for applications and estimation is the class of cointegrated time series. This is due to their close connection to stationary time series. Before we address the topic of cointegration, we first say a few words about multivariate continuous-time moving average (MCARMA) processes.

MCARMA Processes

CARMA processes date back to 1944 when Doob [30] introduced a univariate process which is nothing else but a Gaussian CARMA process. Brockwell [18] extended later in 2001 the definition of univariate CARMA processes by replacing the Brownian motion with a more general Lévy process. Lévy processes include for example Brownian motions, Poisson processes and α-stable processes. CARMA processes are the continuous-time analogue of the extensively studied autoregressive moving average (ARMA) processes. In the last years, there has been considerable interest in this class of processes as can be seen by various publications (some exemplary articles are Brockwell et. al. [19], [21], [22], [23], [24], Fasen et. al. [35], [36],[37], [38], Schlemm and Stelzer [90], [91], Todorov and Tauchen [97], . . .).

Marquardt and Stelzer [69] extended the univariate Lévy driven CARMA processes to the multivariate setting in 2007. The source of randomness in the model is a two-sided m-dimensional Lévy process $(L(t))_{t\in\mathbb{R}}$. A d-dimensional stationary MCARMA process $(Y(t))_{t\in\mathbb{R}}$ of order (p,q) for positive integers $p>q$ is formally defined as the solution to the stochastic differential equation

$$P(D)Y(t) = Q(D)DL(t), \quad D := \frac{\mathrm{d}}{\mathrm{d}t}, \quad t \in \mathbb{R}.$$

As for ARMA processes, we have an autoregressive polynomial $P(z) := I_m z^p + A_1 z^{p-1} + \ldots + A_p$ with matrix coefficients $A_1, \ldots, A_p \in \mathbb{R}^{d\times d}$ and a moving average polynomial $Q(z) := B_0 z^q + B_1 z^{q-1} + \ldots + B_q$ with matrix coefficients $B_0, B_1, \ldots, B_q \in \mathbb{R}^{d\times m}$. The parameters p and q determine the path properties of the MCARMA process, for example if the process has jumps or if it is smooth. The class of MCARMA processes allows for a rich theory in the sense of probabilistic and analytical properties. In particular, Lévy driven MCARMA processes allow for a broad class of marginal distributions.

The class of vector ARMA (VARMA) processes is equivalent to the class of discrete-time linear state space models. The same relation holds for stationary continuous-time linear stochastic state space models which are equivalent to causal MCARMA processes as was shown in [90]. Hence, another way to think of an MCARMA process is via the state space representation. A stationary continuous-time linear state space model is given by the state and observation equations

$$dX(t) = AX(t)\mathrm{d}t + B\mathrm{d}L(t) \quad \text{and} \quad Y(t) = CX(t), \quad t \in \mathbb{R},$$

where $A \in \mathbb{R}^{N\times N}$ has eigenvalues with strictly negative real parts, $B \in \mathbb{R}^{N\times m}$ and $C \in \mathbb{R}^{d\times N}$.

As already mentioned, there are often only observations at discrete time points available. Thus, discretizations of MCARMA processes are of particular interest, especially for estimating the model parameters of an MCARMA process using observations at discrete time points. The sampled process $(Y(nh))_{n\in\mathbb{Z}}$ is the MCARMA process $(Y(t))_{t\in\mathbb{R}}$ observed at equidistant time points. The sampled process satisfies a discrete-time linear state space model. For univariate CARMA processes Brockwell, Davis and Yang [21] showed that the sampled process satisfies a weak ARMA representation and for MCARMA processes Schlemm and Stelzer [90] derived a weak vector ARMA representation. However, the autoregressive polynomial has only complex coefficients instead of matrix coefficents. We derive in this thesis a weak vector ARMA representation for the sampled process $(Y(nh))_{n\in\mathbb{Z}}$ where the autoregressive polynomial is a matrix polynomial. For this purpose, we also derive a decomposition of an MCARMA process into the sum of MCAR processes. Such a decomposition was already used in the univariate case by Brockwell, Davis and Yang [21] and in a different form by Schlemm and Stelzer [90] in order to derive their (vector) ARMA representations.

Cointegration

As previously indicated, stationary processes do not seem to be sufficient to describe many time series data sets. In the year 1982, Nelson and Plosser [70] provided statistical evidence that many macroeconomic variables contain a stochastic trend. There are numerous works providing empirical evidence that many economic and financial data sets exhibit a non-stationary behavior but their first differences are stationary. A first solution to this problem was to work with first differences instead of levels to be able to apply standard methods for stationary processes. However, many relations between (economic) variables are stated in terms of the actual levels rather than differences. Hence, models on first differences are limited in containing such relations.

Granger and Newbold [43] observed that for multivariate time series following a common stochastic trend, statistical inference using the standard methods leads to spurious results. Granger discovered that non-stationary time series can fluctuate around a long-run equilibrium. These time series are non-stationary but behave stationary around the stochastic trend. Clive Granger coined the term cointegration for these multivariate time series. Cointegration means that the multivariate time series is non-stationary itself but their differences are stationary and there exist stationary linear combinations of the components. The seminal works by Granger [42] in 1981 and Engle and Granger [33] in 1987 introduced the concept of cointegration and thereby established a rich field of research. The concept of cointegration became quite popular in econometrics but is also applied in many other fields like physics, biology or social sciences. Cointegration relations have been found, among others, between spot and future (forward) prices, interest rates of different maturities or different countries, dividends and prices or stock prices in an industrial sector (c.f. Brenner and Kroner [17] and references therein). The importance of the concept of cointegration was substantiated by the Nobel price in economics for Clive Granger's discovery of cointegration in 2003. He won the Nobel price together with Robert Engle for their developments in time series analysis.

Cointegrated time series are a subclass of integrated processes, namely processes which have stationary first differences, and thus also of non-stationary processes. They are very closely related to stationary processes due to the defining property of having stationary first differences as well as stationary linear combinations. Hence, one can still use several properties of stationary processes while working with a cointegrated time series. Moreover, the property of having a long-run equilibrium implies that the components cannot drift too far away from the equilibrium but fluctuate around the equilibrium. Thus, in a cointegrated time series deviations from the equilibrium are stationary. This fact is used in the so-called error correction model where the short-run dynamic is considered. The error correction form was introduced in 1964 by Dennis Sargan [87] in a model of wages and prices for the United Kingdom. The classical error correction model for a cointegrated d-dimensional VARMA(p, p) process $Y_n = P(B)Y_{n-1} + Q(B)\epsilon_n$ for $n \in \mathbb{N}$ with noise $(\epsilon_n)_{n\in\mathbb{N}}$ is given by

$$\Delta Y_n = \alpha\beta^{\mathsf{T}} Y_{n-1} + \widetilde{P}(B)\Delta Y_{n-1} + Q(B)\epsilon_n,$$

where the polynomial $\widetilde{P}(z)$ is constructed from the matrix coefficients of the original polynomial $P(z)$. It has since become a useful tool in cointegration analysis for estimating the long-run as well as the short-run behavior. A cointegrated model can fit data with common stochastic trends better than a stationary model for the afore-stated reasons.

One of the most important results in cointegration analysis is the Granger representation theorem derived by Engle and Granger [33] in 1987. The Granger representation theorem connects the moving average, autoregressive and error correction representations for cointegrated time series. The Johansen-Granger representation theorem (c.f. Johansen [53] and [54]) characterizes cointegration precisely by making assumptions on the autoregressive polynomial. Johansen [52], [53] also presented a maximum-likelihood estimation method for cointegrated vector autoregressive processes using a reduced rank regression method. A reduced rank model is a multivariate regression model where the coefficient matrix has a reduced rank. Moreover, Johansen developed the Johansen cointegration rank test as a consequence of his reduced rank regression method. This a sequential test in order to determine the number of cointegration relations.

The most attention in the field of cointegration analysis was given to the discrete-time setting. However, in 1991 Phillips [77] considered cointegration in the continuous-time case. The natural analogue of first differences is differentiation in the continuous-time framework. Phillips considered differentiable stochastic processes in a special triangular model. Chambers [25] considered the connection between cointegrated models in continuous and discrete time. Other examples of cointegrated continuous-time models were considered for example by Kessler and Rahbek [57], who considered a cointegrated Gaussian CAR$(1,0)$ model, or by Fasen [34], [35], where special cases of MCARMA models were investigated.

In the work of Comte [27] a characterization of cointegration for CAR(p) processes was derived in combination with a moving average representation of the cointegrated CAR(p) process. Moreover, a more general definition for integration in continuous-time was given. Instead of defining integration as having a stationary derivative, a process is called integrated if it has stationary increments. This includes particulary Lévy processes and thus also Brownian motions. A multivariate process is then obviously called cointegrated if it has stationary increments and there exist stationary linear combinations. This definition has the advantage to include a much broader class of processes since for example Lévy processes are in general not differentiable. If the process is differentiable, these two definitions do coincide (c.f. Comte [27]).

In this thesis, we present a general cointegrated MCARMA model which includes the models of Comte [27] and Phillips [77]. This definition extends the aforementioned stationary MCARMA model to the non-stationary subclass of (co-)integrated processes. Furthermore, we derive the probabilistic properties of this model and an analogous result to the Granger-Representation theorem, i.e. a moving average representation and characterization of cointegration for MCARMA processes. This characterization is not only given with respect to the autoregressive polynomial but also with respect to the matrices of the state space form. Additionally, we investigate its sampled version and therefore the connection between the continuous-time cointegrated model and its discrete-time version.

Statistical estimation

A question which naturally arises in this context is how to estimate the model parameters of a cointegrated MCARMA process given discrete-time observations. Closely related to this is the problem of model identification. This problem occurs on the one hand due to the estimation of a continuous-time model which is only observed at discrete time points (aliasing effect) and on the other hand due to the multivariate setting, where different models can have the same output. This means models can include many redundancies and can become indistinguishable. It is essential to prevent this from happening by having a sufficient set of assumptions for deriving a consistent estimator. Statistical inference for a cointegrated MCARMA process was only considered for special cases so far. For example, Kessler and Rahbek [58] consider an ergodic Gaussian MCAR(1) process and present an estimation method for this process. They also present sufficient conditions to solve the identifiability problem and avoid the aliasing effect.

Since standard estimations methods for stationary processes are not suitable for non-stationary processes one has to use different methods in this setting. Saikkonen [85], [86] presented in the middle of the 1990s a step-wise estimation approach for cointegrated processes in the discrete-time framework. The stochastic equicontinuity condition and the concept of continuous weak convergence, which is closely related to uniform convergence, make it possible to deal with different rates of convergence. Since a cointegrated process is non-stationary it has different rates of convergence for different directions in the parameter space. However, Saikkonen derives a consistency result and the asymptotic distributions of the estimator using continuous weak convergence results and a suitable stochastic equicontinuity condition. He especially considered the maximum likelihood estimator as an example. A quasi-maximum likelihood method was considered by Schlemm and Stelzer [90] for stationary Lévy driven MCARMA processes observed equidistantly at discrete time points in order to estimate the model parameters. They showed the strong consistency of the quasi-maximum likelihood estimator (QMLE) and also the asymptotic normality. Furthermore, the identifiability problem is solved for this case.

We present in this thesis a step-wise quasi-maximum likelihood estimation method, which is based on the ideas of Saikkonen as well as the quasi-maximum likelihood approach for stationary MCARMA processes by Schlemm and Stelzer [90]. The QMLE for the cointegrated model is consistent, whereby different rates of consistency apply. The QMLE of the short-run parameters is consistent with the standard rate of $\sqrt{n}$, whereas the QMLE of the long-run parameters is super-consistent. Furthermore, we derive the asymptotic distributions of the long-run and short-run parameter estimators. The derived results are in line with the results for cointegrated models in the discrete-time setting. Furthermore, the model identifiability problem is solved for cointegrated MCARMA processes. The assumptions needed to derive these results are standard assumptions and similar to set of assumptions in Schlemm and Stelzer [90] and Saikkonen [85], [86]. The applicability of the estimation procedure is eventually tested in simulation studies.

1.2. Outline of the Thesis

This thesis is divided into six chapters. Chapter 2 summarizes known results about Lévy processes and MCARMA processes, which are the fundamental processes considered in this thesis. In Chapter 3 (weak) VARMA representations of MCARMA and integrated MCARMA processes observed at discrete time points are derived and their autocovariance structure is analyzed. Chapter 4, Chapter 5 and Chapter 6 consider cointegrated MCARMA processes, which comprise representations, characterization, statistical inference and simulation studies.

A list of notations and general abbreviations is given in the end of this thesis on page 228. In order to improve the readability we put some technical and auxiliary results in appendices in the end of each chapter. Furthermore, we state a list of the assumptions of each chapter in Appendix A. A collection of basic formulas and rules in the field of matrix theory can be found in Appendix B. We are going to employ these formulas and rules several times in this thesis and hence we summarized the most important ones in that section.

In the following we outline the thesis and give a short introduction to the results and content of each chapter. A more detailed description can be found in the beginning of each chapter.

Chapter 2: For the sake of understanding we present known results on multivariate Lévy processes in Section 2.2 and moreover, Section 2.3 gives an introduction to multivariate MCARMA processes. In Section 2.4 important properties of MCARMA processes are summarized, which we use on several occasions. Lastly, a definition of integrated CARMA processes and an extension to the multivariate case is given in Section 2.5.

Chapter 3: We consider in Chapter 3 stationary and integrated MCARMA processes and their observations in discrete time. In Section 3.2 we recall basic definitions in the field of matrix polynomials and recall some useful results. We need the theory on matrix polynomials to factorize the autoregressive polynomial $P(z)$ of a stationary MCARMA process Y. This factorization enables us to decompose an MCARMA(p,q) process Y into the sum of p-dependent multivariate Ornstein-Uhlenbeck processes Y_k in Section 3.3.

The result of Section 3.4 exploits the decomposition from the previous section yielding a (weak) VARMA$(p, p-1)$ of a stationary MCARMA(p,q) process Y observed at discrete time points. Next, Section 3.5 deals with a process derived from integrated MCARMA processes $\int_0^t Y(u)\,\mathrm{d}u$ observed at discrete time points, which is an MCARMA$(p+1,q)$ process itself. In particular, we consider the first difference of the sampled integrated process, namely the integrated sequence given by $I_n^{(h)} := \int_{(n-1)h}^{nh} Y(u)\,\mathrm{d}u$. Similar as in Section 3.4, we derive a (weak) VARMA representation of order (p,p) in this case. Besides, the autocovariance structure of both discrete-time processes is investigated.

Chapter 4: This chapter characterizes cointegrated Lévy driven MCARMA processes using two different perspectives. After a brief recapitulation of cointegration of VARMA processes in the discrete time framework we present the corresponding definition of cointegration for the continuous-time setting in Section 4.2. Since the source of randomness in our model are Lévy processes, the definition of integration as non-stationary processes having stationary increments is the feasible definition for the cointegrated model presented in this chapter. Additionally, we derive a continuous-time error correction form analogous to the discrete-time version. However, the main result of this section is the characterization of cointegration for MCARMA processes with respect to the autoregressive and moving average polynomials.

Section 4.3 considers cointegrated MCARMA models from a different perspective. The state space representation of MCARMA processes is used to derive decoupling of the linear system into a stationary and non-stationary part. Solving this state space equation yields that the d-dimensional cointegrated MCARMA process Y satisfies $Y(t) = C_1 B_1 L(t) + Y_2(t)$. This means Y is nothing else but the sum of a Lévy process and a stationary MCARMA process. This representation is the key to derive several probabilistic properties of the cointegrated process, which are also derived in this section.

Note that the characterization and the representation combined are a continuous-time version of the well-known Johansen-Granger Representation Theorem for cointegrated MCARMA models. Besides, we investigate which properties of the cointegrated MCARMA process, observed at discrete time points, are inherited from the continuous-time model. The rank of the matrix C_1 determines the number of common stochastic trends and the orthogonal complement of C_1 spans the cointegration space and consequently the cointegration rank can also be determined via the difference of the dimension of the process Y and the rank of the matrix C_1.

Lastly, we apply the Kalman filter to the sampled process in Section 4.4 in order to obtain the linear innovations $\varepsilon^{(h)}$ of the process. The linear innovations enable us to derive a transfer function error correction form $\varepsilon_n^{(h)} = \Pi Y_{n-1}^{(h)} + \overline{k}(B)\Delta Y_n^{(h)}$ of the sampled process with the linear innovations $\varepsilon^{(h)}$ as noise process. The difference to the classical error correction form is the linear filter $\overline{k}(z)$ of infinite order.

We see in this section that the matrix Π contains the cointegration information, that is the rank of Π, which is equal to the cointegration rank r and the matrix Π can be decomposed into the product $\Pi = \alpha\beta^T$ of full rank matrices with appropriate dimensions. Note that α is the adjustment matrix and the columns of β span the cointegration space. Moreover, the connection between β and the orthogonal complement of the matrix C_1 become self-evident. Eventually, the probabilistic properties of $\varepsilon^{(h)}$ are investigated. Lastly, the applicability of the Kalman filter in the non-stationary setting with unit roots is guaranteed due to the derivations in Appendix Section 4.6.

Chapter 5: Chapter 5 forms the main part of this thesis. It contains a step-wise estimation procedure for cointegrated Lévy driven MCARMA processes based on equidistant observations in discrete time. We use the results on the linear innovations from the previous chapter in order to calculate the pseudo-Gaussian log-likelihood function $\mathcal{L}_n^{(h)}(\vartheta)$ of observations $(y_1, \ldots, y_n)$, which we state in Section 5.2. Furthermore, we make a collection of assumptions on the driving Lévy process and the parametrization. Most of these assumptions are standard in the quasi-maximum likelihood approach for stationary processes.

Due to the non-stationarity we separate the parameter space Θ into short-run and long-run parameters in order to use a step-wise estimation approach. This basically means that we collect the parameters corresponding to the non-stationary behavior in a sub-vector $\vartheta_1 \in \Theta_1$ and the parameters corresponding to the stationary part in another sub-vector $\vartheta_2 \in \Theta_2$. Likewise, we separate the log-likelihood function into a sum $\mathcal{L}_n^{(h)}(\vartheta) = \mathcal{L}_{n,1}^{(h)}(\vartheta) + \mathcal{L}_{n,2}^{(h)}(\vartheta_2)$. The first summand $\mathcal{L}_{n,1}^{(h)}(\vartheta)$ depends on all parameters, whereas the second summand only depends on the stationary parameters $\mathcal{L}_{n,2}^{(h)}(\vartheta_2)$. We employ this partitioned form in the rest of this chapter.

Since stochastic equicontinuity and continuous weak convergence are essential to derive the asymptotic distribution, we recall these concepts in Section 5.4 and derive several continuous weak convergence for processes appearing in our model. Furthermore, we prove the stochastic equicontinuity condition for these processes in this section.

Section 5.3 deals with the identifiability problem and aliasing effect, which appear in the estimation of multivariate continuous-time systems using discrete time observations. The parametrization should be chosen such that different values of the parameter must generate different probability distributions of the observations, namely the model should be identifiable. We state sufficient conditions on the parametrization in order to have an identifiable model and derive a unique parametrization using the decoupled state space representation.

In Section 5.5 we show the super-consistency of the long-run quasi-maximum likelihood estimator. We do not merely derive a consistency result, but also determine the order of consistency in a second step. In the third step, we prove the consistency of the short-run quasi-maximum likelihood estimator with the knowledge of the order of consistency of the long-run parameter estimator. This step-wise proof is necessary as different rates of convergence apply to different directions of the parameter space.

Finally, we derive the asymptotic distributions of the long-run and short-run estimator in Section 5.6. For this purpose, we first prove the weak convergence of the score vector to a vector consisting of a stochastic integral and of a normal distributed random variable. Afterwards, we show the convergence of the Hessian matrix to a block-diagonal matrix, which is almost surely positive definite. Then by applying a mean value expansion of the score vector and using the results of this chapter, we obtain the asymptotic distributions of the estimators. The long-run estimator is mixed normal and the short-run estimator is asymptotically normal. Besides, these results yield also the asymptotic independence of the estimators.

Chapter 6: We apply in this section the step-wise quasi-maximum likelihood estimation procedure in simulation studies. For this purpose we present in Section 6.2 a unique parametrization for the long-run matrix C_1 satisfying the assumption of the previous chapter. The matrix C_1 contains the cointegration information. To be more precise, for an output process $Y = C_1 B_1 L(t) + Y_2(t)$ of dimension d and matrix $C_1 \in M_{d,c}(\mathbb{R})$ with full rank c we know that the process has $d - c$ stationary linear combinations and consequently c common stochastic trends. Moreover, the orthogonal complement of this matrix spans the cointegration space. The algorithm presented describes how to construct the matrix C_1 in a unique way from a given long-run parameter vector. Furthermore, references on appropriate parametrization for the remaining matrices are given.

In the end, we present the results of the simulation studies in Section 6.3. We consider a bivariate model with one common stochastic trend and a three-dimensional model with two common stochastic trends. Moreover, we compare the results for a Brownian motion and a normal-inverse Gaussian process respectively. We will see that the simulation studies verify the practical applicability of our step-wise quasi-maximum likelihood estimation method for cointegrated Lévy driven MCARMA processes.

Chapter 2

Preliminaries

2.1. Introduction

This preliminary chapter shall serve as an introduction to the model class we want to investigate in the following. Since all results in this chapter are already known, we briefly recall the most important results.

We introduce in this chapter some basic results on Lévy processes. Lévy processes will be the source of randomness in our model. Not only do Lévy processes include many widely used processes as for example Brownian motion, Poisson process and many others, they can also be considered as the continuous analogue of random walks. In the first subsection we give a short overview over Lévy processes.

In discrete time the class of multivariate autoregressive moving average (VARMA) processes are well studied. This model class consist of two parts. On the one hand, we have the autoregressive part modelling the linear dependence of the output process on its prior values. On the other hand, we have the moving average part, which is a function of its past innovations. Linear stochastic state space models have a close connection to ARMA models. These model classes are equivalent, see e.g. in Hannan and Deistler [46]. A comprehensive overview over discrete-time ARMA processes and state space models can be found for example in Brockwell and Davis [20].

A natural generalization of these models is a continuous-time model. Continuous-time autoregressive moving average processes (CARMA) date back to 1944 when Doob

[30] introduced a process which can be interpreted as an Gaussian CARMA process. Brockwell introduced later a univariate CARMA process driven by a Lévy process in [18]. Marquardt and Stelzer extended then CARMA processes to the multivariate case in [69] and their connection to continuous time linear stochastic state space models was shown in [90].

We recapitulate in the second subsection the definition of a multivariate CARMA (MCARMA) process. Moreover, we recall in the last subsection some of the basic properties of MCARMA processes. These definitions and results will build the basis for the integrated and cointegrated MCARMA models.

Integrated CARMA (ICARMA) processes have been introduced by Brockwell and Lindner [23] in 2013 in order to model spot volatility. Their definition of an integrated CARMA process extends naturally to the multivariate setting. We state this extended definition in Section 2.5. The property of integration is a necessary condition for a cointegrated process. However, this definition of an integrated MCARMA process has the disadvantage that it does not enable the process to be cointegrated. Such an integrated process will always be integrated but will have no stationary linear combination. Later on, we introduce a different definition of an integrated MCARMA process which is more flexible and hence we can have cointegration relations in this case.

2.2. Multivariate Lévy processes

We recall in this section the definition of a multivariate Lévy process and some of its properties. We use these processes throughout this thesis as the driving process of our models. For a profound treatment of Lévy processes see e.g. the textbooks by Applebaum [4], Bertoin [10] and Sato [88].

Definition 2.2.1 (Lévy process)
We say that a stochastic process $L = (L(t))_{t\geq 0}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a (m-dimensional) ***Lévy process*** *if it satisfies the following four properties:*

1. $L(0) = 0_m \quad \mathbb{P} - a.s.$
2. *L has independent increments, i.e. for each $n \in \mathbb{N}$ and each $0 \leq t_0 < t_1 < \ldots < t_n < \infty$ the random variables $L(t_0),\ L(t_1) - L(t_0),\ \ldots,\ L(t_n) - L(t_{n-1})$ are independent.*
3. *L has stationary increments, i.e. for all $t, s \geq 0$ we have $L(t+s) - L(s) \stackrel{d}{=} L(t)$.*

4. *L is stochastically continuous, i.e. for all $\varepsilon > 0$ and for all $s \geq 0$ we have* $\lim_{t\to s} \mathbb{P}(\|L(t) - L(s)\| > \varepsilon) = 0$.

A Lévy process is uniquely determined by its characteristic function in the Lévy-Khintchine form $\mathbb{E}\left[\mathrm{e}^{i\langle u, L(t)\rangle}\right] = \mathrm{e}^{t\Psi(u)}$, $u \in \mathbb{R}^m$, $t \geq 0$, where Ψ is given by

$$\Psi(u) = i\langle\gamma, u\rangle - \frac{1}{2}\langle u, \Sigma u\rangle + \int_{\mathbb{R}^m} \left(\mathrm{e}^{i\langle u,x\rangle} - 1 - i\langle u, x\rangle \mathbb{1}_{\{\|x\|\leq 1\}}\right) \nu(\mathrm{d}x).$$

The unique characteristic triplet (γ, Σ, ν) consists of the drift vector $\gamma \in \mathbb{R}^m$, the Gaussian covariance matrix $\Sigma \in \mathbb{R}^{m\times m}$, which is a positive semi-definite matrix, and the Lévy measure ν, which is a measure on $\mathbb{R}^m$ satisfying $\nu(0_m) = 0$ and $\int_{\mathbb{R}^m} \min(1, \|x\|^2)\nu(\mathrm{d}x) < \infty$. In the following, we consider the cádlág version of the Lévy process.

A two-sided Lévy process $L = (L(t))_{t\in\mathbb{R}}$ is defined by taking two independent copies $\{L_1(t)\}_{t\geq 0}$ and $\{L_2(t)\}_{t\geq 0}$ and set

$$L(t) := \begin{cases} L_1(t), & t \geq 0, \\ -L_2(-t-), & t < 0, \end{cases}$$

where $L(s-) := \lim_{t\nearrow s} L(t)$.

For a proper definition of stochastic integration with respect to Lévy processes see e.g. the books Applebaum [4] and Protter [82].

2.3. Lévy-driven Multivariate CARMA Processes

We recall in this section Lévy-driven multivariate CARMA processes, which were first defined by Marquardt and Stelzer (see [69]). This definition was slightly extended by Schlemm and Stelzer (see [90]), where the driving Lévy process was allowed to have a different dimension than the multivariate CARMA process. Subsequently, we summarize the main theory on MCARMA processes. We use the notation A^* to denote the complex conjugate matrix of A in the following.

Definition 2.3.1
Let $\mathcal{B}(\mathbb{R})$ denote the Borel-σ-algebra over $\mathbb{R}$. A family $\{\zeta(\Delta)\}_{\Delta\in\mathcal{B}(\mathbb{R})}$ of $\mathbb{C}^m$-valued random variables is called an m-dimensional ***random orthogonal measure*** *if*

(a) *$\zeta(\Delta) \in L^2$ for all bounded $\Delta \in \mathcal{B}(\mathbb{R})$,* *(b)* $\zeta(\emptyset) = 0$,

(c) $\zeta(\Delta_1 \cup \Delta_2) = \zeta(\Delta_1) + \zeta(\Delta_2)$ *a.s. if* $\Delta_1 \cap \Delta_2 = \emptyset$ *and*

(d) $F : \mathcal{B}(\mathbb{R}) \to M_m(\mathbb{C})$, $\Delta \mapsto \mathbb{E}[\zeta(\Delta)\zeta(\Delta)^*]$ *defines a* σ*-additive positive definite matrix measure (i.e. a* σ*-additive set function that assumes values in the positive semi-definite matrices) and it holds that*

$$\mathbb{E}[\zeta(\Delta_1)\zeta(\Delta_2)^*] = F(\zeta(\Delta_1) \cap \zeta(\Delta_2))$$

for all $\Delta_1, \Delta_2 \in \mathcal{B}(\mathbb{R})$.

F *is referred to as the spectral measure of* ζ.

The stochastic integrals $\int_\Delta f(t)\zeta(\mathrm{d}t)$ of deterministic Lebesgue-measurable functions $f : \mathbb{R} \to M_m(\mathbb{C})$ with respect to a random orthogonal measure ζ are defined in the usual L^2-sense. The integration is defined componentwise

$$\left(\int_\Delta f(t)\zeta(\mathrm{d}t)\right) := \begin{pmatrix} \left(\int_\Delta f(t)\zeta(\mathrm{d}t)\right)_1 \\ \vdots \\ \left(\int_\Delta f(t)\zeta(\mathrm{d}t)\right)_i \\ \vdots \\ \left(\int_\Delta f(t)\zeta(\mathrm{d}t)\right)_m \end{pmatrix} := \begin{pmatrix} \sum_{k=1}^m \int_\Delta f_{1k}(t)\zeta_k(\mathrm{d}t) \\ \vdots \\ \sum_{k=1}^m \int_\Delta f_{ik}(t)\zeta_k(\mathrm{d}t) \\ \vdots \\ \sum_{k=1}^m \int_\Delta f_{mk}(t)\zeta_k(\mathrm{d}t) \end{pmatrix}.$$

The integral is defined whenever

$$\int_\Delta f(t)F(\mathrm{d}t)f(t)^* := \left(\sum_{k,l=1}^m \int_\mathbb{R} f_{ik}(t)\bar{f}_{ik}(t)F_{kl}(\mathrm{d}t)\right)_{1\le i,j\le m} < \infty,$$

and f is said to be in $L^2(F)$. As we consider only the case, where we have spectral measures with constant density with respect to the Lebesgue measure λ on $\mathbb{R}$, i.e. $F(dt) = C\lambda(dt) := C\ dt$ holds for some positive definite $C \in M_m(\mathbb{C})$. Then it is sufficient for the existence of the integral that $\int_\Delta \|f(t)\|^2\,\mathrm{d}t < \infty$, for some norm $\|\cdot\|$ on $M_m(\mathbb{C})$. We denote by $M_m(\mathbb{C})$ the space of all m-dimensional complex valued matrices. Additionally, we denote the space of square integrable matrix-valued functions by

$$L^2(\mathbb{R}, M_m(\mathbb{C})) := L^2(M_m(\mathbb{C})) := \left\{ f : \mathbb{R} \to M_m(\mathbb{C}),\ \int_\mathbb{R} \|f(t)\|^2\,\mathrm{d}t < \infty \right\}.$$

Furthermore, we have for two functions $f, g \in L^2(F)$

$$\mathbb{E}\left[\int_\Delta f(t)\zeta(\mathrm{d}t)\left(\int_\Delta g(t)\zeta(\mathrm{d}t)\right)^*\right] = \int_\Delta f(t)Cg(t)^*\,\mathrm{d}t. \tag{2.1}$$

The next theorem shows the existence of a random orthogonal measure and a corresponding spectral measure.

Theorem 2.3.2 (Marquardt and Stelzer (2007), Theorem 3.5)
Let $L = (L(t))_{t\in\mathbb{R}}$ be an m-dimensional square integrable Lévy process with $\mathbb{E}[L(1)] = 0$ and $\mathbb{E}[L(1)L(1)^*] = \Sigma_L$. Then, there exists an m-dimensional random orthogonal measure Φ_L with spectral measure F_L such that $\mathbb{E}[\Phi_L(\Delta)] = 0$ for any bounded Borel set Δ,

$$F_L(dt) = \frac{\Sigma_L}{2\pi}dt$$

and

$$L(t) = \int_{-\infty}^{\infty} \frac{\mathrm{e}^{i\mu t} - 1}{i\mu}\,\Phi_L(\mathrm{d}\mu).$$

The random measure Φ_L is uniquely determined by

$$\Phi_L([a,b)) = \int_{-\infty}^{\infty} \frac{\mathrm{e}^{i\mu a} - \mathrm{e}^{i\mu b}}{2\pi i\mu}\,L(\mathrm{d}\mu)$$

for all $-\infty < a < b < \infty$.

For the polynomial

$$P : \mathbb{C} \to M_m(\mathbb{C}), \quad z \mapsto I_m z^p + A_1 z^{p-1} + A_2 z^{p-2} + \ldots + A_p,$$

with matrix coefficients $A_1, A_2, \ldots, A_p \in M_m(\mathbb{C})$ and $p \in \mathbb{N}$, we have the corresponding companion matrix given by

$$A = \begin{pmatrix} 0 & I_m & 0 & \ldots & 0 \\ 0 & 0 & I_m & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \ldots & \ldots & 0 & I_m \\ -A_p & -A_{p-1} & \ldots & \ldots & -A_1 \end{pmatrix} \in M_{mp}(\mathbb{C}). \tag{2.2}$$

The following result of Marquardt and Stelzer [69], is the main result in order to define multivariate CARMA processes. Since matrices do not commute in general, the definition is not as straightforward as in the one dimensional case. However, the following theorem by Marquardt and Stelzer enables a definition of MCARMA processes, which contains the autoregressive moving average structure.

Theorem 2.3.3 (Marquardt and Stelzer (2007), Theorem 3.12)
Let $L = (L(t))_{t\in\mathbb{R}}$ be an m-dimensional square integrable Lévy process with mean zero and corresponding m-dimensional random orthogonal measure Φ as in Theorem 2.3.2 and $p, q \in \mathbb{N}_0$, $q < p$. Let further $A_1, A_2, \ldots, A_p$, $B_0, B_1, \ldots, B_q \in M_m(\mathbb{R})$, where $B_0 \neq 0$ and define $\beta_1 = \beta_2 = \ldots = \beta_{p-q-1} = 0_{m\times m}$ (if $p > q+1$) and

$$\beta_{p-j} = -\sum_{i=1}^{p-j-1} A_i \beta_{p-j-i} + B_{q-j} \quad \text{for } j = 0, 1, 2, \ldots, q.$$

Assume that A is defined as in (2.2) satisfying $\sigma(A) \subset (-\infty, 0) + i\mathbb{R}$, which implies $A_p \in \mathcal{G}l_m(\mathbb{R})$. We denote by $G = (G_1^*(t), \ldots, G_p^*(t))^*$ an mp-dimensional process and set $\beta = (\beta_1^*, \ldots, \beta_p^*)^*$.

Then the stochastic differential equation

$$dG(t) = AG(t)dt + \beta dL(t) \tag{2.3}$$

is uniquely solved by the process G given by

$$G_j(t) = \int_{-\infty}^{\infty} \mathrm{e}^{i\lambda t} w_j(i\lambda)\, \Phi(\mathrm{d}\lambda), \quad j = 1, 2, \ldots, p, \quad t \in \mathbb{R},$$

where for $j = 1, 2, \ldots, p-1$

$$w_j(z) = \frac{1}{z}(w_{j+1}(z) + \beta_j),$$

and

$$w_p(z) = \frac{1}{z}\left(-\sum_{k=0}^{p-1} A_{p-k} w_{k+1}(z) + \beta_p\right).$$

The strictly stationary process G can also be represented as

$$G(t) = \int_{-\infty}^{t} \mathrm{e}^{A(t-s)} \beta\, L(\mathrm{d}s), \quad t \in \mathbb{R}. \tag{2.4}$$

Moreover, $G(0)$ and $\{L(t)\}_{t\geq 0}$ are independent, in particular, $\mathbb{E}[G_j(0)L(t)^*] = 0$ for *all $t \geq 0$, $j = 1, 2, \ldots, p$. Finally, it holds that*

$$w_p(z) = P(z)^{-1}\left(\beta_p z^{p-1} - \sum_{j=0}^{p-2}\sum_{k=0}^{j} A_{p-k}\beta_{p+k-j-1}z^j\right) \quad \text{and}$$
$$w_1(z) = P(z)^{-1}Q(z),$$

where the "autoregressive polynomial" is given by

$$P(z) = I_m z^p + A_1 z^{p-1} + \ldots + A_p, \text{ for } z \in \mathbb{C}$$

and the "moving average polynomial" by

$$Q(z) = B_0 z^q + B_1 z^{q-1} + \ldots + B_q \text{ for } z \in \mathbb{C} \tag{2.5}$$

and $\int_{-\infty}^{\infty} \|w_j(i\lambda)\|^2 \,\mathrm{d}\lambda < \infty$ for all $j \in \{1, 2, \ldots, p\}$.

Note that the process G in Theorem 2.3.2 is a multivariate Ornstein-Uhlenbeck process. From this fact one can derive many probabilistic properties of MCARMA processes.

Let us now give the definition of a multivariate continuous-time autoregressive moving average process as stated in Marquardt and Stelzer [69].

Definition 2.3.4 (MCARMA process)
Let $L = (L(t))_{t\in\mathbb{R}}$ be a two-sided square integrable m-dimensional Lévy process with $\mathbb{E}[L(1)] = 0$ and $\mathbb{E}[L(1)L(1)^*] = \Sigma_L$.

*An m-dimensional **Lévy-driven continuous-time autoregressive moving average process** $(Y(t))_{t\in\mathbb{R}}$ of order (p, q) for $p > q$ (MCARMA(p, q)) is defined as*

$$Y(t) = \int_{-\infty}^{\infty} \mathrm{e}^{i\lambda t} P(i\lambda)^{-1} Q(i\lambda)\, \Phi(\mathrm{d}\lambda), \quad t \in \mathbb{R}, \tag{2.6}$$

where the autoregressive and moving average polynomial are given by

$$P(z) := I_m z^p + A_1 z^{p-1} + \ldots + A_p,$$

and

$$Q(z) := B_0 z^q + B_1 z^{q-1} + \ldots + B_q$$

for $z \in \mathbb{C}$.

Furthermore, Φ *is the Lévy orthogonal random measure of Theorem 2.3.2 satisfying* $\mathbb{E}[\Phi(d\lambda)] = 0$ *as well as* $\mathbb{E}[\Phi(d\lambda)\Phi(d\lambda)^*] = \frac{d\lambda}{2\pi}\Sigma_L$*. Here* $A_j \in M_m(\mathbb{R})$*,* $j = 1, \ldots, p$ *and* $B_i \in M_m(\mathbb{R})$*,* $i = 1, 2, \ldots, q$ *are matrices satisfying* $B_q \neq 0$ *and*

$$\mathcal{N}(P) := \{z \in \mathbb{C} : \det(P(z)) = 0\} \subset \mathbb{R} \setminus \{0\} + i\mathbb{R}.$$

Normally, one is interested in a causal process. This means that the process is a function of past values and does not depend on future values, that is the process adapted to the natural filtration of the driving Lévy process. Hence, we give the definition of a causal MCARMA process.

Definition 2.3.5 (Causal MCARMA process)
Let $L = (L(t))_{t\in\mathbb{R}}$ *be a two-sided* m*-dimensional Lévy process satisfying*

$$\int_{\|x\|\geq 1} \log \|x\| \, \nu(\mathrm{d}x) < \infty. \tag{2.7}$$

Assume further that $p, q \in \mathbb{N}_0$ *with* $q < p$ *and* $A_1, A_2, \ldots, A_p, B_0, B_1, \ldots, B_q \in M_m(\mathbb{R})$*, where* $B_0 \neq 0$*. Define the matrices* A, β *and the polynomial* P *as in Theorem 2.3.3 and assume* $\sigma(A) = \mathcal{N}(P) \subset (-\infty, 0) + i\mathbb{R}$.

Then the m*-dimensional process*

$$Y(t) = (I_m, 0_m, \ldots, 0_m)G(t), \tag{2.8}$$

where G *is the unique stationary solution to*

$$dG(t) = AG(t)dt + \beta dL(t)$$

is called a ***causal MCARMA***(p, q) ***process****. Again, the process* G *is referred to as the state space representation.*

Additionally, the stationary MCARMA process can also be represented as a moving average process as can be seen by the next theorem.

Theorem 2.3.6 (Marquardt and Stelzer (2007), Theorem 3.22)
The MCARMA process (2.6) can be represented as a moving average process

$$Y(t) = \int_{-\infty}^{\infty} g(t-s) \, L(\mathrm{d}s), \quad t \in \mathbb{R}, \tag{2.9}$$

where the kernel matrix function $g : \mathbb{R} \to M_m(\mathbb{R})$ *is given by*

$$g(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\mu t} P(i\mu)^{-1} Q(i\mu) \, d\mu. \tag{2.10}$$

For causal MCARMA process defined in Definition 2.3.5 an analogous result holds if the kernel function g is replaced by

$$\tilde{g}(s) = (I_m, 0_m, \dots, 0_m) e^{As} \beta \mathbb{1}_{[0,\infty)}(s). \tag{2.11}$$

Remark 2.3.7
All the results in this and the next subsection remain valid if we allow the dimension of the Lévy process to be different than the dimension of the MCARMA process. We have in this case a two-sided Lévy process L with values in $\mathbb{R}^d$ and the m-dimensional Lévy-driven CARMA process, with polynomials $P(z) \in M_m(\mathbb{R}(z))$ and $Q(z) \in M_{m,d}(\mathbb{R}(z))$, is the solution of p^{th}-order linear differential equation

$$P(D)Y(t) = Q(D)DL(t),$$

which is interpreted as being equivalent to the state space representation $dG(t) = AG(t)dt + \beta dL(t)$, $Y(t) = (I_m, 0_m, \dots, 0_m)G(t)$, $t \in \mathbb{R}$, where we have now matrices with the following dimensions: $\beta \in M_{mp\times d}(\mathbb{R})$, and $(I_m, 0_m, \dots, 0_m) \in M_{m,pm}(\mathbb{R})$.

2.4. Properties of Multivariate CARMA Processes

Lastly, let us recall some useful properties of stationary MCARMA processes in this section.

Proposition 2.4.1 (Marquardt and Stelzer (2007), Proposition 3.26)
The processes defined in Definition 2.3.4 and Definition 2.3.5 are strictly stationary.

Proposition 2.4.2 (Marquardt and Stelzer (2007), Proposition 3.27)
If the driving Lévy process L *has the characteristic triplet* (γ, Σ, ν)*, the distribution of the MCARMA process* $Y(t)$ *is infinitely divisible for* $t \in \mathbb{R}$*. Moreover, the characteristic triplet of the stationary distribution is* $(\gamma_Y^\infty, \Sigma_Y^\infty, \nu_Y^\infty)$*, where*

$$\gamma_Y^\infty = \int_{\mathbb{R}} g(s)\gamma \, ds + \int_{\mathbb{R}} \int_{\mathbb{R}^m} g(s)x \left(\mathbb{1}_{\{\|g(s)x\| \le 1\}} - \mathbb{1}_{\{\|x\| \le 1\}} \right) \nu(dx) \, ds,$$
$$\Sigma_Y^\infty = \int_{\mathbb{R}} g(s)\Sigma g^*(s) \, ds$$

$$\text{and} \quad \nu_Y^{\infty}(B) = \int_{\mathbb{R}} \int_{\mathbb{R}^m} \mathbb{1}_B(g(s)x)\, \nu(\mathrm{d}x)\, \mathrm{d}s. \tag{2.12}$$

For a causal MCARMA process the same result holds with g replaced by $\tilde{g}$.

Proposition 2.4.3 (Marquardt and Stelzer (2007), Proposition 3.28)
Let $Y = (Y(t))_{t\in\mathbb{R}}$ be the MCARMA process defined by (2.3.4). Then, its matrix-valued autocovariance function is given by

$$\Gamma_y(h) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{i\lambda h} P(i\lambda)^{-1} Q(i\lambda) \Sigma_L Q(i\lambda)^* (P(i\lambda)^{-1})^*\, \mathrm{d}\lambda, \quad h \in \mathbb{R} \tag{2.13}$$

and the spectral density is given by

$$f_y(\lambda) = \frac{1}{2\pi} P(i\lambda)^{-1} Q(i\lambda) \Sigma_L Q(i\lambda)^* (P(i\lambda)^{-1})^*, \quad \lambda \in \mathbb{R}. \tag{2.14}$$

Definition 2.4.4
An $\mathbb{R}^d$-valued continuous-time ***linear state space model*** *$(\mathcal{A}, \mathcal{B}, \mathcal{C}, L)$ of dimension N is characterized by an $\mathbb{R}^m$-valued driving Lévy process L, a state transition matrix $\mathcal{A} \in M_N(\mathbb{R})$, an input matrix $\mathcal{B} \in M_{N,m}(\mathbb{R})$ and an observation matrix $\mathcal{C} \in M_{d,N}(\mathbb{R})$. It consists of a state equation of Ornstein-Uhlenbeck type*

$$\mathrm{d}X(t) = \mathcal{A}X(t)\mathrm{d}t + \mathcal{B}\mathrm{d}L(t) \tag{2.15}$$

and an observation equation

$$Y(t) = \mathcal{C}X(t). \tag{2.16}$$

The $\mathbb{R}^N$-valued process $X = (X(t))_{t\in\mathbb{R}}$ is the state vector process and the $\mathbb{R}^d$-valued process $Y = (Y(t))_{t\in\mathbb{R}}$ is the output process.

The next result characterizes the connection between continuous-time state-space models and MCARMA processes, which is very useful in the following.

Proposition 2.4.5 (Schlemm and Stelzer (2012), Corollary 3.4)
Assume that the driving Lévy process L satisfies $\mathbb{E}L(1) = 0$ and $\mathbb{E}\|L(1)\|^2 < \infty$. The classes of causal MCARMA and causal continuous-time state space models are equivalent.

Last but not least, we recall a result on the sampled process of a stationary MCARMA processes by Schlemm and Stelzer [90]. The corresponding sampled process $Y^{(h)} =$

$(Y_n^{(h)})_{n\in\mathbb{Z}}$ of a continuous-time stochastic process $Y = (Y(t))_{t\in\mathbb{R}}$ is defined by inserting nh into the original process i.e. $Y_n^{(h)} = Y(nh)$. The constant $h > 0$ determines the sampling distance. Consequently, we observe the process at equidistant time points. The sampled process is of particular interest in the case of discrete-time observation of the original continuous-time process.

Lemma 2.4.6 (Schlemm and Stelzer (2012), Lemma 5.2)
Assume that Y is an MCARMA process as in Definition 2.4.4. The sampled process $Y^{(h)}$ has the state space representation given by the state equation

$$X_n = \mathrm{e}^{Ah} X_{n-1} + R_n^{(h)}, \tag{2.17}$$

and observation equation

$$Y_n^{(h)} = CX_n, \tag{2.18}$$

with noise

$$R_n^{(h)} = \int_{(n-1)h}^{nh} \mathrm{e}^{A(nh-u)}\beta \, \mathrm{d}L(u). \tag{2.19}$$

The sequence $(R_n^{(h)})_{n\in\mathbb{Z}}$ is i.i.d. with mean zero and covariance matrix

$$\tilde{\Sigma} = \mathbb{E}R_n^{(h)} R_n^{(h)\mathsf{T}} = \int_0^h \mathrm{e}^{Au}\beta\Sigma_L\beta^*\mathrm{e}^{A^*u}\,\mathrm{d}u. \tag{2.20}$$

We have now recalled the most important definitions and results about stationary MCARMA processes. After this brief review of stationary MCARMA processes we are going to extend these processes to the non-stationary case in the following.

2.5. Integrated MCARMA Processes

We define in this section integrated MCARMA processes and derive some characterizations. Integrated CARMA processes were first defined in Brockwell and Lindner [23] in the univariate setting. An integrated CARMA process is given by the integral of a stationary CARMA process.

Definition 2.5.1 (Integrated CARMA process)
Let $Y = (Y(t))_{t\in\mathbb{R}}$ be a univariate stationary CARMA(p, q) process with integers $p > q$. The non-stationary d-times ***integrated CARMA****(p, d, q)* ***(ICARMA)***

process $I^{(d)}$*, for* $d \in \mathbb{N}_0$ *is defined as*

$$I^{(d)}(t) = \int_0^t \int_0^{u_{d-1}} \dots \int_0^{u_1} Y(u)\,\mathrm{d}u\,\mathrm{d}u_1 \dots \mathrm{d}u_{d-1}. \tag{2.21}$$

We denote a d-times continuous-time integrated process I by $I \sim \mathcal{I}(d)$. Note that the $ICARMA(p,d,q)$ process $I^{(d)}$ can be represented in the following way

$$I^{(d)}(t) = \int_0^t \frac{(t-u)^{d-1}}{(d-1)!} Y(u)\,\mathrm{d}u, \quad d \in \mathbb{N}_0, \tag{2.22}$$

which can be seen by induction. The representation is similar to the representation of a fractionally integrated CARMA process as in Marquardt [68].

We can extend the definition straightforwardly to the multivariate case as can be seen in the next definition.

Definition 2.5.2 (Multivariate Integrated CARMA process)
Let $Y = (Y(t))_{t\in\mathbb{R}} = (Y_1(t), Y_2(t), \dots, Y_m(t))^{\mathsf{T}}_{t\in\mathbb{R}}$ *be an* m*-dimensional stationary multivariate* $CARMA(p,q)$ *process with parameters* $p > q$*. The non-stationary* ***multivariate*** d***-times integrated CARMA***(p,d,q) ***(MICARMA) process*** $I^{(d)}$*,* $d \in \mathbb{N}_0$*, is defined as*

$$\begin{aligned} I^{(d)}(t) &:= \int_0^t \int_0^{u_{d-1}} \dots \int_0^{u_1} Y(u)\,\mathrm{d}u\,\mathrm{d}u_1 \dots \mathrm{d}u_{d-1} \\ &:= \begin{pmatrix} \int_0^t \int_0^{u_{d-1}} \dots \int_0^{u_1} Y_1(u)\,\mathrm{d}u\,\mathrm{d}u_1 \dots \mathrm{d}u_{d-1} \\ \vdots \\ \int_0^t \int_0^{u_{d-1}} \dots \int_0^{u_1} Y_m(u)\,\mathrm{d}u\,\mathrm{d}u_1 \dots \mathrm{d}u_{d-1} \end{pmatrix}. \end{aligned} \tag{2.23}$$

The definition of an integrated CARMA process is not restricted to Riemann integrals, e.g. we could also define it with respect to stochastic integrals, for example integrate with respect to a Lévy process.

Remark 2.5.3
If we have the differential equation $P(D)Y(t) = Q(D)DL(t)$ and define $P_d(z) := P(z) \cdot z^d$, we have

$$P_d(D)I^{(d)}(t) = P(D)D^d I^{(d)}(t) = P(D)Y(t) = Q(D)DL(t) \tag{2.24}$$

and thus $I^{(d)}$ is itself an MCARMA$(p+d,q)$ process.

There is an analogous representation of the MICARMA process similar to the representation (2.22) in the univariate case

$$I^{(d)}(t) = \int_0^t \frac{(t-u)^{d-1}}{(d-1)!} Y(u)\,\mathrm{d}u = \int_0^t \operatorname{diag}\left(\frac{(t-u)^{d-1}}{(d-1)!}\right) Y(u)\,\mathrm{d}u. \tag{2.25}$$

The integral in the last equation is again defined componentwise. Using the state space equation we can derive another representation.

Proposition 2.5.4
The MICARMA(p,d,q) process $I^{(d)}$ can be represented in the following way

$$\begin{aligned} I^{(d)}(t) &= \int_0^t \frac{(t-u)^{d-1}}{(d-1)!} Y(u)\,\mathrm{d}u \\ &= [I_m, 0_m, \dots, 0_m] A^{-d} \left[G(t) - \sum_{j=0}^{d-1} \frac{A^j}{j!} \left(t^j G(0) - \int_0^t (t-u)^j \beta\,\mathrm{d}L(u) \right) \right], \end{aligned} \tag{2.26}$$

where G and β are defined as in Theorem 2.3.3.

Proof. This can be seen directly by using the state space representation (2.3) recursively. □

CHAPTER 3

ARMA Representations of MCARMA and MICARMA Processes Observed at Discrete Time Points

3.1. Introduction

Although we consider in this thesis continuous-time models, the discrete-time models are of special interest for us. The reason for this is that despite having a continuous-time model, we observe only the process at discrete time points. Hence, we analyze the representation of the observed process and moreover its autocovariance structure. The aim of this chapter is to derive a vector autoregressive moving average (VARMA) representation for stationary and integrated multivariate autoregressive moving average (MICARMA) processes observed at discrete time points.

Stationary processes have practical properties, however, data often suggest a non-stationary behavior. Integrated ARMA (ARIMA) processes are a generalization of stationary ARMA processes allowing for a certain degree of non-stationarity. Taking the first difference of an ARIMA reduces the non-stationarity. Using this fact one can still use theory on stationary processes for integrated time series by taking the first differences. The textbooks of Box, Jenkins and Reinsel [15] and Brockwell and Davis [20] cover the basic theory on ARIMA processes.

The continuous-time integrated ARMA (ICARMA) processes, which was introduced by Brockwell and Lindner [23] in 2013 is the continuous-time analogue of an

ARIMA process. It is naturally defined by considering the integral of a stationary MCARMA process Y, namely $\int_0^t Y(t)\,\mathrm{d}t$. We consider the property of the integrated sequence, which is nothing else than the increments of the ICARMA process, that is $\int_{(n-1)h}^{nh} Y(t)\,\mathrm{d}t$.

In order to obtain the VARMA representations of either the stationary MCARMA process or the integrated sequence of an MCARMA process $\int_{(n-1)}^{n(h)} Y(t)\,\mathrm{d}t$, we need a decomposition of a stationary MCARMA process into MCAR processes. In particular we obtain the decomposition $Y(t) = \sum_{k=1}^{p} Y_k(t)$ of a stationary MCARMA(p, q) process, where $Y_k(t)$ are MCAR processes. Decompositions of CARMA processes into (multivariate) Ornstein-Uhlenbeck processes were considered in the univariate case by Brockwell, Davis and Yang [21] and by Schlemm and Stelzer [90] in the multivariate setting. We decompose an MCARMA process into the sum of MCAR processes. The difference between our decomposition and the one presented in Schlemm and Stelzer is the kind of process we decompose the original process into.

Necessary for this decomposition are results from the field of matrix polynomials. References about matrix analysis and matrix polynomials are for example the textbooks of Horn and Johnson [50] and Gohberg et al. [41]. For the sake of comprehensibility, we briefly recall the main definitions and the required results for this chapter in Section 3.2 and present an extension to repeated solvents in Appendix 3.7.

Finally, we present weak VARMA representations of stationary MCARMA and integrated MCARMA observed at discrete time points in Section 3.4. A related VARMA representation for stationary MCARMA processes can be found in Schlemm and Stelzer [90], where a different decomposition was used. Furthermore, an ARMA representation of a sampled univariate CARMA process was also considered in Brockwell and Lindner [22] and for a univariate integrated CARMA process the result was given in Brockwell and Lindner [23].

3.2. Theory on Matrix Polynomials

In this section we review main results about matrix polynomials. Of greater interest for the following considerations is some criterion if a matrix polynomial can be factorized into „linear factors“. This means we have matrix valued „roots“ for which we can factor the matrix polynomial as in the one-dimensional case. However, since we do not have the Fundamental Theorem of Algebra for matrix polynomials and the commutativity property does not hold, we need the following definitions and results to specify a similar result at least for special cases.

First, we define matrix polynomials. Since in general matrix multiplication is not commutative, we need to make a difference of the polynomial depending on the side we multiply the matrix.

Definition 3.2.1
*An n^{th}***-degree,** *m^{th}***-order monic λ-matrix** *$A : \mathbb{C} \to M_m(\mathbb{C})$ is given by*

$$A(\lambda) = A_0\lambda^n + A_1\lambda^{n-1} + \ldots + A_{n-1}\lambda + A_n, \tag{3.1}$$

where $A_k \in M_m(\mathbb{C}),\ k = 0, 1, \ldots, n,\ A_0 = I_m$ and $\lambda \in \mathbb{C}$.

Let X be an $m \times m$ dimensional complex matrix. The ***right matrix polynomial*** *$A_R : M_m(\mathbb{C}) \to M_m(\mathbb{C})$ is given by*

$$A_R(X) = A_0X^n + A_1X^{n-1} + \ldots + A_{n-1}X + A_n \tag{3.2}$$

and analogously the ***left matrix polynomial*** *A_L is given by*

$$A_L(X) = X^nA_0 + X^{n-1}A_1 + \ldots + XA_{n-1} + A_n.$$

By considering the determinant of a matrix polynomial, we obtain a one-dimensional polynomial with „classical“ roots. These roots of the determinant of the matrix polynomial and vectors related to these roots are important for the multivariate extension of univariate roots. Therefore, we give now a formal definition.

Definition 3.2.2
Let λ_i be a complex number such that

$$\det(A(\lambda_i)) = 0, \tag{3.3}$$

then λ_i is called ***latent root*** *of $A(\lambda)$. A complex $m \times 1$ vector p_i satisfying*

$$A(\lambda_i)p_i = 0_{m\times 1} \tag{3.4}$$

is called ***right latent vector*** *of $A(\lambda)$ associated to λ_i. Similarly, q_i is a* ***left latent vector*** *if*

$$q_i^\intercal A(\lambda_i) = 0_{1\times m}.$$

Finally, we are going extend the definition of a root to the matrix polynomial case. Since matrices do not commutate in general, we have to differ between left and right roots, depending on the side we are multiplying the matrix.

Definition 3.2.3

Let R be an $m \times m$ complex matrix such that

$$A_R(R) = A_0R^n + A_1R^{n-1} + \ldots + A_{n-1}R + A_n = 0_{m\times m}, \tag{3.5}$$

*then R is called a **right solvent** of the λ-matrix $A(\lambda)$ and an $m \times m$ complex matrix L is a **left solvent** if $A_L(L) = L^nA_0 + L^{n-1}A_1 + \ldots + LA_{n-1} + A_n = 0_{m\times m}$.*

The companion form of a matrix is a special form closely related to matrix polynomials. The coefficient matrices build the last row of the companion matrix. This special form plays a major role for (Vector-)ARMA and (Multivariate-)CARMA processes as one can use this form to represent these processes. Moreover, it can be easier to derive some properties of the processes using the companion form instead of the matrix polynomial form.

Definition 3.2.4

*Given a right matrix polynomial $M(X) = X^n + A_1X^{n-1} + \ldots + A_n$ the corresponding **block companion matrix** is given by*

$$A_C = \begin{pmatrix} 0 & I_m & 0 & \ldots & 0 \\ 0 & 0 & I_m & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \ldots & \ldots & 0 & I_m \\ -A_n & -A_{n-1} & \ldots & \ldots & -A_1 \end{pmatrix} \in M_{mn}(\mathbb{C}). \tag{3.6}$$

A block companion matrix A_C is invertible if A_n is invertible. If A_C is invertible, the inverse of the block companion matrix is given by

$$A_C^{-1} = \begin{pmatrix} -D_n & -D_{n-1} & \ldots & \ldots & -D_1 \\ I_m & 0 & \ldots & \ldots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & I_m & 0 \end{pmatrix} \in M_{mn}(\mathbb{C}), \tag{3.7}$$

where $D_i = A_n^{-1}A_{i-1}$, for $i = 2, \ldots n$ and $D_1 = A_n^{-1}$.

Note that the characteristic polynomial of the companion matrix is given by

$$\det(A_C - \lambda I_m) = (-1)^{mn} \det(I_m\lambda^n + A_1\lambda^{n-1} + \ldots + A^n). \tag{3.8}$$

This equation shows clearly the connection between eigenvalues of the companion matrix and the roots of the matrix polynomial. A profound theory on matrix polynomials and solvents can be found for example in Dennis et al. [29] and [28].

After we have listed the most important definitions we can now summarize some useful theorems. We start with a useful connection of latent roots of the matrix polynomial and eigenvalues of the corresponding companion matrix.

Theorem 3.2.5 (Dennis et al. (1976), Theorem 3.2)
If λ is latent root of $A(\lambda)$ and p is a right latent vectors, then λ is an eigenvalue of the matrix A_C and

$$v := \begin{pmatrix} p^{\mathsf{T}} & \lambda p^{\mathsf{T}} & \cdots & \lambda^{n-1} p^{\mathsf{T}} \end{pmatrix}^{\mathsf{T}} \in M_{mn\times 1}(\mathbb{C})$$

is a right eigenvector of A_C.

As earlier mentioned, we can characterize a solvent by its latent roots and latent vectors, which can be seen easily by the next result of Dennis [28].

Theorem 3.2.6 (Dennis et al. (1976), Lemma 4.1)
If $A(\lambda)$ has m linearly independent right latent vectors $p_1, \dots, p_m$ corresponding to the latent roots $\lambda_1, \dots, \lambda_m$, then

$$P\Lambda P^{-1}$$

is a right solvent, where $P := (p_1, \dots, p_m) \in M_m(\mathbb{C})$ and $\Lambda := diag(\lambda_1, \dots, \lambda_m)$.

In order to factorize an n^{th}-degree matrix polynomial into linear factors we need criteria for having n factors. There exist special cases, when one can completely factorize a matrix polynomial. In order to characterize these cases we need the following definition.

Definition 3.2.7
A set of right (left) solvents $R_k,\ k = 1, \dots, n$ ($L_k,\ k = 1, \dots, n$) is called a ***complete set of right (left) solvents*** *of $A(\lambda)$ if*

$$\sigma(A(\lambda)) = \bigcup_{k=1}^{n} \sigma(R_k), \quad \left(\sigma(A(\lambda)) = \bigcup_{k=1}^{n} \sigma(L_k)\right), \tag{3.9}$$

where $\sigma(A(\lambda))$ is the spectrum of $A(\lambda)$ and $\sigma(R_k)$ ($\sigma(L_k)$) is the spectrum of the right (left) solvent R_k (L_k).

As in Markus and Mereuca [66], a right solvent R of $A(\lambda)$ is called regular if

$$\sigma(R) \cap \sigma(A^{(1)}(\lambda)) = \emptyset, \tag{3.10}$$

where the monic λ-matrix $A^{(1)}(\lambda)$ of degree $n-1$ is given by

$$A(\lambda) = A^{(1)}(\lambda)(\lambda I_m - R). \tag{3.11}$$

Similarly, a left solvent of $A(\lambda)$ is called regular if

$$\sigma(L) \cap \sigma(\widehat{A}^{(1)}(\lambda)) = \emptyset,$$

where the monic λ-matrix $\widehat{A}^{(1)}(\lambda)$ of degree $n-1$ satisfies $A(\lambda) = (\lambda I_m - L)\widehat{A}^{(1)}(\lambda)$. Hence, a complete set of regular right (left) solvents R_k $(L_k), k = 1, \ldots, n$ of $A(\lambda)$ is given by

$$\sigma(R_k) \cap \sigma(R_j) = \emptyset, \ k \neq j, \ j, k = 1, \ldots, n$$

and

$$\sigma(A(\lambda)) = \bigcup_{k=1}^{n} \sigma(R_k), \tag{3.12}$$

or respectively

$$\sigma(L_k) \cap \sigma(L_j) = \emptyset, \ k \neq j, \ j, k = 1, \ldots, n \quad \text{and} \quad \sigma(A(\lambda)) = \bigcup_{k=1}^{n} \sigma(L_k).$$

The easiest case, where we have a complete set of eigenvalues is characterized in the next theorem of Dennis et al. [28].

Theorem 3.2.8 (Dennis et al. (1976), Theorem 4.1)
If the latent roots of $A(\lambda)$ are distinct, then $A(X)$ has a complete set of right solvents.

Next, we recall the definition of matrix residues and use Cauchy's integral theorem. Let $X \in M_m(\mathbb{C})$ with distinct eigenvalues. We denote by $\lambda_1, \ldots, \lambda_q$ for $q \leq m$ the distinct eigenvalues of the matrix X. Cauchy's integral theorem (see e.g. Lax [60], Theorem 17.5) states that for the matrix function $f(X) := \sum_{n=0}^{\infty} a_n X^n$ we have

$$f(X) = \frac{1}{2\pi i} \oint_{\Gamma} f(\lambda)(\lambda I_m - X)^{-1} \, \mathrm{d}\lambda = \frac{1}{2\pi i} \oint_{\Gamma} (\lambda - I_m X)^{-1} f(\lambda) \, \mathrm{d}\lambda, \tag{3.13}$$

where $\lambda_1, \ldots, \lambda_q$ lies in the interior of Γ. Note that Γ is some closed contour in $\mathbb{C}$ winding once around each eigenvalue of X.

Thus, we have for the right matrix polynomial $A_R : M_m(\mathbb{C}) \to M_m(\mathbb{C})$

$$\begin{aligned} A_R(X) &= \frac{1}{2\pi i} \oint_\Gamma A(\lambda)(\lambda I_m - X)^{-1}\,\mathrm{d}\lambda \\ &= \sum_{k=0}^{n} A_k \left[\frac{1}{2\pi i} \oint_\Gamma \lambda^{n-k}(\lambda - I_m X)^{-1}\,\mathrm{d}\lambda \right] \\ &= \sum_{k=0}^{n} A_k X^{n-k}. \end{aligned} \tag{3.14}$$

Consequently, for matrix exponentials we have by Lax [60], Theorem 17.5, a similar representation given by

$$\mathrm{e}^{Xt} = \frac{1}{2\pi i} \oint_\Gamma \mathrm{e}^{zt}(zI_m - X)^{-1}\,\mathrm{d}z. \tag{3.15}$$

For a more detailed theory on matrix residues of rational λ-matrices see e.g. Tsay and Shieh [98]. We state the definition of rational λ-matrices and consider their residues.

Definition 3.2.9
A ***strictly proper rational left λ-matrix*** *with n^{th}-degree m^{th}-order has the representation*

$$F(\lambda) = A_l(\lambda)^{-1} B_l(\lambda), \tag{3.16}$$

$A_l(\lambda) = \sum_{k=0}^{n} A_{lk}\lambda^{n-k}$ *with* $A_{l0} = I_m$, *and* $B_l(\lambda) = \sum_{k=0}^{n-1} B_{lk}\lambda^{n-1-k}$. *The rational λ-matrix $F(\lambda)$ is irreducible if $A_l(\lambda)$ and $B_l(\lambda)$ are left coprime.*

A rational right λ-matrix $F(\lambda) = B_r(\lambda)A_r(\lambda)^{-1}$ is defined analogously. We denote by $\mathrm{adj}(A)$ the adjugate of A. An alternative representation for a rational λ-matrix $F(\lambda)$ is given by

$$F(\lambda) = \frac{1}{\det(A_l(\lambda))}\,\mathrm{adj}(A_l(\lambda))B_l(\lambda).$$

For an irreducible $F(\lambda)$ the roots of $\det(A_l(\lambda))$ are referred to as the poles of $F(\lambda)$ and $\det(A_l(\lambda)) = \det(A_r(\lambda))$ holds. Furthermore, we have $\mathrm{adj}(A_l(\lambda))B_l(\lambda) = B_r(\lambda)\,\mathrm{adj}(A_r(\lambda))$. If $A_l(\lambda)$ has a complete set of regular right solvents R_i for $i = 1, \ldots, n$, then the right solvents R_i are called the regular left block poles of $F(\lambda)$.

Definition 3.2.10
Let R be a regular right solvent of $A(\lambda)$, then the ***residue of the rational λ-matrix*** *$F(\lambda)$ at R is defined by*

$$Res[F,R] := \frac{1}{2\pi i}\oint_{\Gamma_R} F(\lambda)\,d\lambda, \tag{3.17}$$

where Γ_R is a simple closed contour such that $\sigma(R)$ is contained in the interior of Γ_R and $\sigma(A(\lambda)) \setminus \sigma(R)$ is contained in the exterior of Γ_R.

The next theorem characterizes a matrix residual and the corresponding rational left matrix function to a solvent, which can be found in the paper of Tsay and Shieh [98].

Theorem 3.2.11 (Tsay and Shieh (1982), Theorem 3.4)
If $F(\lambda)$ is defined as above and $A(\lambda)$ has a set of regular right solvents $R_1, \dots R_q$ with $\sigma(R_k)$, $k = 1, \dots, q$ and $q \leq n$ lying in the interior of Γ. Then

$$\frac{1}{2\pi i}\oint_{\Gamma} F(\lambda)\,\mathrm{d}\lambda = \sum_{k=1}^{q} Res[F,R_k] = \sum_{k=1}^{q} F_L^{(k)}(R_k), \tag{3.18}$$

where $F_L^{(k)}(R)$ is the rational left matrix function of $F^{(k)}(\lambda) := (\lambda I_m - R_k)F(\lambda)$ evaluated at R.

Finally, we recall another theorem of Tsay and Shieh [98], which is going to be the key result for the proof of the decomposition presented in the next section. It enables us to separate a rational λ-matrix into a sum using the right solvents and the corresponding matrix residual.

Theorem 3.2.12 (Tsay and Shieh (1982), Theorem 4.1)
If $F(\lambda)$ is a strictly proper, irreducible rational λ-matrix, where $A(\lambda)$ has a complete set of regular right solvents $R_1, \dots R_n$, then

$$F(\lambda) = \sum_{k=1}^{n} (\lambda I_m - R_k)^{-1}\, Res[F,R_k]. \tag{3.19}$$

One can also consider repeated right solvents instead of non-recurring solvents. For the sake of completeness we have summarized the main results for this generalization in Appendix 3.7. The result in this chapter hold also for repeated solvents. However, the representations are more technical and thus we forego to present the more general result.

3.3. Decomposition of an MCARMA process

Before we present the key results, we first state some assumptions which we need repeatedly throughout the rest of this chapter. In order to have a properly defined MCARMA process we make a moment assumption on the Lévy process. Furthermore, we make two assumptions on the transition matrix $\mathcal{A}$. The first assumption on $\mathcal{A}$ guarantees that we have a stationary MCARMA process and the second one will not only simplify the notations but also be a sufficient condition for the corresponding matrix polynomial to have a complete set of regular right solvents.

Assumption A1
The Lévy process L satisfies $\mathbb{E}L(1) = 0$ and $\mathbb{E}\|L(1)\|^2 < \infty$.

Assumption A2
The eigenvalues of $\mathcal{A}$ in equation (2.15) and consequently of $A \in M_N(\mathbb{C})$ in (2.2), have strictly negative real parts, where $N = pm$.

Assumption A3
The eigenvalues $\lambda_1, \dots, \lambda_N$ of $\mathcal{A}$ in equation (2.15) and consequently of $A \in M_N(\mathbb{C})$ in (2.2), are distinct, where the dimensions satisfies $N = pm$.

The following theorem is the multivariate extension to the result of Brockwell, Davis and Yang [21], Proposition 2, where a univariate CARMA process was decomposed into CAR(1) processes. In the work of Schlemm and Stelzer [90], Proposition 5.1), another decomposition was considered. In the latter case, an MCARMA process was considered and decomposed into a sum of dependent Ornstein-Uhlenbeck processes. However, we have a matrix in the exponential in contrast to a scalar as in the results of Schlemm and Stelzer [90].

Theorem 3.3.1
Let Y be a MCARMA(p,q) process and assume that Assumption A1-Assumption A3 hold. Then there exist a complete regular set of right solvents R_k of the autoregressive matrix polynomial $P(\lambda)$, for $k = 1, \dots, p$, such that the process Y can be decomposed into a sum of dependent, complex-valued multivariate Ornstein-Uhlenbeck processes as

$$Y(t) = \sum_{k=1}^{p} Y_k(t), \tag{3.20a}$$

where

$$Y_k(t) = \mathrm{e}^{R_k(t-s)} Y_k(s) + \int_s^t \mathrm{e}^{R_k(t-u)}\, Res[F, R_k]\, \mathrm{d}L(u), \tag{3.20b}$$

for $s, t \in \mathbb{R}$ with $s < t$ and the matrix fraction description F is given by $F(\lambda) := P(\lambda)^{-1}Q(\lambda)$.

Proof. Due to Assumption A3 we can apply Theorem 3.2.8 and hence we have a complete set of regular right solvents of $P(\lambda)$. Further, let $F(\lambda) := P(\lambda)^{-1}Q(\lambda)$ be the rational matrix function, where $P(\lambda)$ and $Q(\lambda)$ are defined as in (2.5).

By Theorem 3.2.12 we have

$$F(\lambda) = P(\lambda)^{-1}Q(\lambda) = \sum_{k=1}^{p} (\lambda I_m - R_k)^{-1}\, \mathrm{Res}[F, R_k]. \tag{3.21}$$

Then, with Cauchy's integral formula (3.15) we obtain

$$\begin{aligned}
g(t) &= (I_m, 0_m, \ldots, 0_m) e^{At} \beta \mathbb{1}_{[0,\infty)}(t) \\
&= \frac{1}{2\pi i} \oint_\Gamma \mathrm{e}^{t\lambda} (I_m, 0_m, \ldots, 0_m)(\lambda I_{pm} - A)^{-1} \beta \, \mathrm{d}\lambda \\
&= \frac{1}{2\pi i} \oint_\Gamma \mathrm{e}^{t\lambda} P(\lambda)^{-1} Q(\lambda) \mathrm{d}\lambda \\
&= \frac{1}{2\pi i} \sum_{k=1}^{p} \oint_\Gamma \mathrm{e}^{t\lambda} (\lambda I_m - R_k)^{-1}\, \mathrm{Res}[F, R_k] \mathrm{d}\lambda \\
&= \frac{1}{2\pi i} \sum_{k=1}^{p} \oint_{\Gamma_k} \mathrm{e}^{t\lambda} (\lambda I_m - R_k)^{-1}\, \mathrm{Res}[F, R_k] \mathrm{d}\lambda \\
&= \sum_{k=1}^{p} \mathrm{e}^{tR_k}\, \mathrm{Res}[F, R_k],
\end{aligned} \tag{3.22}$$

where Γ_k is a simple closed contour such that $\sigma(R_k)$ lies in the interior of Γ_k and the residuary spectrum $\sigma(A(\lambda)) \setminus \sigma(R_k)$ lies in the exterior of Γ_k and $\Gamma := \bigcup_{k=1}^n \Gamma_k$.

Finally, we obtain by using (3.22) and Remark 3.23 in Marquardt and Stelzer [69] that

$$\begin{aligned}
Y(t) &= \int_{-\infty}^{t} g(t-s) \mathrm{d}L(s) \\
&= \sum_{k=1}^{p} \int_{-\infty}^{t} \mathrm{e}^{R_k(t-s)}\, \mathrm{Res}[F, R_k]\, \mathrm{d}L(s)
\end{aligned}$$

$$= \sum_{k=1}^{p} \mathrm{e}^{R_k(t-s)} Y_k(s) + \sum_{k=1}^{p} \int_s^t \mathrm{e}^{R_k(t-u)} \operatorname{Res}[F, R_k] \, \mathrm{d}L(u)$$

and thus we have completed the proof. □

Using this decomposition we can also calculate the covariance matrix of the decomposed process which depends now on the residuals of the rational matrix function $F(\lambda)$. A representation for the covariance matrix of the decomposed process is derived in the next proposition.

Proposition 3.3.2

Let Y be an MCARMA process, which satisfies Assumption A1, Assumption A2 and Assumption A3. Then we have the covariance matrix $\gamma_Y(l) = Cov(Y(t+l), Y(t))$ of the decomposed process given by

$$\gamma_Y(l) = \sum_{i=1}^{p} \sum_{j=1}^{p} \mathrm{e}^{|l| R_i} \int_0^{\infty} \mathrm{e}^{u R_i} \, Res[F, R_i] \Sigma_L \, Res[F, R_j]^{\mathsf{T}} \mathrm{e}^{u R_j^{\mathsf{T}}} \, \mathrm{d}u. \tag{3.23}$$

Proof. Due to the assumptions we can apply Theorem 3.3.1 and obtain

$$\begin{aligned}
\gamma_Y(l) &= \operatorname{Cov}(Y(t+l), Y(t)) \\
&= \operatorname{Cov}\left(\sum_{i=1}^{p} Y_i(t+l), \sum_{j=1}^{p} Y_j(t) \right) \\
&= \sum_{i,j=1}^{p} \operatorname{Cov}\left(\int_{-\infty}^{t+l} \mathrm{e}^{R_i(t+|l|-u)} \operatorname{Res}[F, R_i] \, \mathrm{d}L(u), \int_{-\infty}^{t} \mathrm{e}^{R_j(t-u)} \operatorname{Res}[F, R_j] \, \mathrm{d}L(u) \right) \\
&= \sum_{i,j=1}^{p} \mathrm{e}^{|l| R_i} \int_0^{\infty} \mathrm{e}^{u R_i} \operatorname{Res}[F, R_i] \Sigma_L \operatorname{Res}[F, R_j]^{\mathsf{T}} \mathrm{e}^{u R_j^{\mathsf{T}}} \, \mathrm{d}u.
\end{aligned}$$

Thus, we have shown the representation in the proposition. □

Last but not least, we comment briefly how the decomposition looks like in the case of repeated right solvents. In this case the notation gets more complicated due to the repeated solvents.

Remark 3.3.3

Assume that we have in Theorem 3.3.1 a complete set of repeated solvents R_k, for $k = 1, \ldots, \mu$, with multiplicities $\nu_1, \ldots, \nu_\mu$ respectively, instead of distinct eigenvalues. Then the stationary MCARMA process Y can be decomposed similar as in

Theorem 3.3.1 using Theorem 3.7.8 as

$$Y(t) = \sum_{i=1}^{\mu} \sum_{j=1}^{\nu_i} Y_{ij}(t), \tag{3.24}$$

with

$$Y_{ij}(t) = \mathrm{e}^{R_i(t-s)} Y_{ij}(s) + \int_s^t \mathrm{e}^{R_i(t-u)} F_{ij} \, \mathrm{d}L(u), \tag{3.25}$$

where F_{ij} is the corresponding matrix residue given by (3.54).

3.4. ARMA Representation and Autocovariance Structure of MCARMA Processes Observed in Discrete Time

We consider now MCARMA processes observed at discrete time points. The aim of this section is to derive VARMA representations for both cases in the subsequent section. Hereafter, we restrict ourselves to the case of distinct eigenvalues for the sake of simplicity in notations, namely Assumption A3 holds in the following. Besides, we assume throughout the rest of this chapter that Assumption A1 and Assumption A2 hold. Hence, we have a causal stationary MCARMA process. Furthermore, we use in the following the state space representation of the sampled process given as in Lemma 2.4.6.

Let us first prove an auxiliary lemma, c.f. Brockwell and Lindner [22] for the one-dimensional case. To distinguish the notation between the continuous-time process and the sampled discrete-time process, we write Y_n for $Y(n)$ in the following and accordingly $Y_{k,n}$ for $Y_k(n)$.

Lemma 3.4.1
For each $l \in \mathbb{N}_0$ and all complex $m \times m$ matrices $C_1, \ldots, C_l$ it holds that

$$\begin{aligned} Y_{k,n}^{(h)} = & \sum_{r=1}^{l} C_r Y_{k,n-r}^{(h)} + \left[\mathrm{e}^{hlR_k} - \sum_{r=1}^{l} C_r \mathrm{e}^{h(l-r)R_k} \right] Y_{k,n-l}^{(h)} \\ & + \sum_{r=1}^{l-1} \left[\mathrm{e}^{hrR_k} - \sum_{j=1}^{r} C_j \mathrm{e}^{h(r-j)R_k} \right] N_{k,n-r}^{(h)} , \end{aligned} \tag{3.26}$$

where $N_{k,n}^{(h)} := \int_{(n-1)h}^{nh} \mathrm{e}^{R_k(nh-u)} \, Res[F, R_k] \, \mathrm{d}L(u)$, where the λ-matrix F and the right *solvents R_k are given as in Theorem 3.3.1.*

Proof. We rewrite (2.18) as stated in the lemma. We show the claim by induction. The assertion is clear for $l = 0$ since we always set the empty sum to zero. Assume the equation holds for some $l \in \mathbb{N}$. We can then rewrite $Y_{k,n}$ using (3.20b)

$$\begin{aligned} Y_{k,n}^{(h)} &= \sum_{r=1}^{l} C_r Y_{k,n-r}^{(h)} + \left[\mathrm{e}^{hlR_k} - \sum_{r=1}^{l} C_r \mathrm{e}^{h(l-r)R_k} \right] \left(\mathrm{e}^{hR_k} Y_{k,n-(l+1)}^{(h)} + N_{k,n-l}^{(h)} \right) \\ &\quad + \sum_{r=1}^{l-1} \left[\mathrm{e}^{hrR_k} - \sum_{j=1}^{r} C_j \mathrm{e}^{h(r-j)R_k} \right] N_{k,n-r}^{(h)} \\ &= \sum_{r=1}^{l+1} C_r Y_{k,n-r}^{(h)} + \left[\mathrm{e}^{h(l+1)R_k} - \sum_{r=1}^{l} C_r \mathrm{e}^{h(l+1-r)R_k} - C_{l+1} \right] Y_{k,n-(l+1)}^{(h)} \\ &\quad + \sum_{r=1}^{l} \left[\mathrm{e}^{hrR_k} - \sum_{j=1}^{r} C_j \mathrm{e}^{h(r-j)R_k} \right] N_{k,n-r}^{(h)}, \end{aligned}$$

which completes the induction step. □

A polynomial $P \in M_m(\mathbb{R}[z])$ is called monic if its leading coefficient is equal to I_m and Schur-stable if the zeros of $z \mapsto \det P(z)$ all lie in the complement of the closed unit disc. Eventually, we can obtain a VARMA representation for the sampled version of an MCARMA process.

Theorem 3.4.2

Assume that Y is an MCARMA process as in Definition 2.3.4 satisfying Assumption A1, Assumption A2 and Assumption A3 and $Y^{(h)}$ is its sampled version. Define the Schur-stable polynomial $\Phi \in M_m(\mathbb{R}[z])$ by

$$\Phi(z) = (I_m - \mathrm{e}^{hS_p} z) \cdots (I_m - \mathrm{e}^{hS_1} z) =: I_m - \Phi_1 z^1 - \ldots - \Phi_p z^p, \qquad (3.27)$$

such that e^{hR_k}, $k = 1, \ldots, p$ are right solvents of the polynomial $\Phi(z)$ and the matrices S_i are similar matrices to the right solvents R_i, for $i = 1, \ldots, p$, which depend on the order of the factorization.

Then there exists a monic Schur-stable polynomial $\Theta \in M_m(\mathbb{R}[z])$ of degree at most $p - 1$ such that

$$\Phi(B) Y_n^{(h)} = \Theta(B) \varepsilon_n^{(h)}, \quad n \in \mathbb{Z}, \qquad (3.28)$$

where B denotes the backshift operator, i.e. $B^j Y_n^{(h)} = Y_{n-j}^{(h)}$ for every non-negative *integer j and the white noise sequence $\varepsilon_n^{(h)}$ (not necessarily i.i.d.). Thus, $Y_n^{(h)}$ admits a weak $VARMA(p, p-1)$ representation.*

Proof. By setting $t = nh$ and $s = (n-1)h$ in (3.20b), we obtain that $Y_n^{(h)} = \sum_{k=1}^{p} Y_{k,n}^{(h)}$, where $Y_{k,n}^{(h)}$ are the sampled version of the component MCAR(1) process of Theorem 3.3.1. They satisfy

$$Y_{k,n}^{(h)} = \mathrm{e}^{hR_k} Y_{k,n-1}^{(h)} + N_{k,n}^{(h)}, \tag{3.29}$$

with

$$N_{k,n}^{(h)} = \int_{(n-1)h}^{nh} \mathrm{e}^{R_k(nh-u)} \operatorname{Res}[F, R_k] \, \mathrm{d}L(u) \tag{3.30}$$

and hence we obtain $(I_m - \mathrm{e}^{hR_k} B) Y_{k,n}^{(h)} = N_{k,n}^{(h)}$.

For each $l \in \mathbb{N}_0$ and all complex $dm \times m$ matrices $\Phi_1, \ldots, \Phi_l$ it holds by Lemma 3.4.1 that with

$$\begin{aligned} Y_{k,n}^{(h)} = & \sum_{r=1}^{p} \Phi_r Y_{k,n-r}^{(h)} + \left[\mathrm{e}^{hlR_k} - \sum_{r=1}^{p} \Phi_r \mathrm{e}^{h(p-r)R_k} \right] Y_{k,n-l}^{(h)} \\ & + \sum_{r=1}^{p-1} \left[\mathrm{e}^{hrR_k} - \sum_{j=1}^{r} \Phi_j \mathrm{e}^{h(r-j)R_k} \right] N_{k,n-r}^{(h)}. \end{aligned} \tag{3.31}$$

Note that by [67], Theorem 5.2., (c.f. Theorem 3.7.7) we only have similar matrices in the factorization instead of the right solvents itself. The fact that e^{hR_k} is a right solvent of $z \mapsto \Phi(z)$ implies that

$$e^{phR_k} - \Phi_1 e^{(p-1)hR_k} - \ldots - \Phi_p = 0.$$

Hence, we have with $l = p$ and $C_r = \Phi_r$ for (3.31) that

$$\Phi(B) Y_{k,n}^{(h)} = \sum_{r=0}^{p-1} \left[\mathrm{e}^{hrR_k} - \sum_{j=1}^{r} \Phi_j \mathrm{e}^{h(r-j)R_k} \right] N_{k,n-r}^{(h)}.$$

Summation over k and rearranging leads to

$$\Phi(B) Y_n^{(h)} = \sum_{k=1}^{p} W_{k,n-k+1}^{(h)} := U_n^{(h)}, \tag{3.32}$$

where the i.i.d. sequences $(W_{k,n}^{(h)})_{n \in \mathbb{Z}}$, $k \in \{1, \ldots, p\}$, are defined analogous as above by

$$W_{k,n}^{(h)} := \int_{(n-1)h}^{nh} \sum_{r=0}^{p-1} \left[\mathrm{e}^{hrR_k} - \sum_{j=1}^{r} \Phi_j \mathrm{e}^{h(r-j)R_k} \right] \mathrm{e}^{R_k(nh-u)} \operatorname{Res}[F, R_k] \mathrm{d}L(u). \quad (3.33)$$

By a multivariate generalization of Brockwell and Davis [20], Proposition 3.2.1, there exists a monic Schur-stable (due to Assumption A2) polynomial

$$\Theta(z) = I_m + \Theta_1 z + \ldots + \Theta_{p-1} z^{N-1}$$

and a multivariate white-noise process $(\varepsilon_n^{(h)}) \sim WN(0, \Sigma)$, $n \in \mathbb{Z}$ and $\Theta_i \in M_m(\mathbb{R})$, $i = 1, \ldots, p$, such that (3.28) holds. □

Next, we derive the covariance matrix of the series $(U_n^{(h)})_{n\in\mathbb{N}}$, which is defined in the proof of the last theorem.

Proposition 3.4.3
Let $U_n^{(h)}$ be the multivariate time series defined in (3.32). Then we obtain for the covariance matrix $\gamma_{U^{(h)}}$ at lag $l = -(p-1), \ldots, p-1$ that

$$\begin{aligned}\gamma_{U^{(h)}}(l) = \sum_{i=1}^{p-|l|} \Bigg[\sum_{\nu=1}^{p} \sum_{\mu=1}^{p} \left(\mathrm{e}^{h(i+m-1)R_\nu} - \sum_{j=1}^{i+|l|-1} \Phi_j \mathrm{e}^{h(i+|l|-j-1)R_\nu} \right) \\ \cdot \Sigma_{\nu,\mu}^{(h)} \left(\mathrm{e}^{h(i-1)R_\mu} - \sum_{j=1}^{i-1} \Phi_j \mathrm{e}^{h(i-j-1)R_\mu} \right)^{\mathsf{T}} \Bigg], \end{aligned} \quad (3.34)$$

and $\gamma_{U^{(h)}}(l) = 0$ for $|l| \geq p$, where

$$\Sigma_{\nu,\mu}^{(h)} = \int_0^h \mathrm{e}^{R_\nu(h-u)}\, Res[F, R_\nu]\ \Sigma_L\ \ Res[F, R_\mu]^{\mathsf{T}} \mathrm{e}^{R_\mu^{\mathsf{T}}(h-u)}\, \mathrm{d}u. \quad (3.35)$$

Proof. Let $l \in \{-(p-1), \ldots, p-1\}$, then

$$\begin{aligned}\gamma_{U^{(h)}}(l) &= \operatorname{Cov}\left(U_{n+|l|}^{(h)}, U_n^{(h)} \right) \\ &= \operatorname{Cov}\left(W_{1,n+|l|}^{(h)} + \ldots + W_{p,n+|l|-p+1}^{(h)}, W_{1,n}^{(h)} + \ldots + W_{p,n-p+1}^{(h)} \right) \\ &= \sum_{i=1}^{p-|l|} \operatorname{Cov}\left(W_{i+|l|,n-i+1}^{(h)}, W_{i,n-i+1}^{(h)} \right) \\ &\overset{(3.33)}{=} \sum_{i=1}^{p-|l|} \left[\sum_{\nu=1}^{p} \sum_{\mu=1}^{p} \left(\mathrm{e}^{h(i+|l|-1)R_\nu} - \sum_{j=1}^{i+|l|-1} \Phi_j \mathrm{e}^{h(i+|l|-j-1)R_\nu} \right) \right.\end{aligned}$$

$$\operatorname{Cov}\left(\int_{(n-i)h}^{(n-i+1)h} \mathrm{e}^{R_\nu[(n-i+1)h-u]}\operatorname{Res}[F,R_\nu]\,\mathrm{d}L(u),\right.$$
$$\left.\int_{(n-i)h}^{(n-i+1)h} \mathrm{e}^{R_\mu[(n-i+1)h-u]}\operatorname{Res}[F,R_\mu]\,\mathrm{d}L(u)\right)$$
$$\left.\left(\mathrm{e}^{h(i-1)R_\mu} - \sum_{j=1}^{i-1}\Phi_j \mathrm{e}^{h(i-j-1)R_\mu}\right)^{\mathsf{T}}\right],$$

where

$$\begin{aligned}
\Sigma_{\nu,\mu}^{(h)} &:= \operatorname{Cov}\left(N_{\nu,n-i+1}^{(h)}, N_{\mu,n-i+1}^{(h)}\right)\\
&= \operatorname{Cov}\left(\int_{(n-i)h}^{(n-i+1)h} \mathrm{e}^{R_\nu[(n-i+1)h-u]}\operatorname{Res}[F,R_\nu]\,\mathrm{d}L(u),\right.\\
&\qquad\left.\int_{(n-i)h}^{(n-i+1)h} \mathrm{e}^{R_\mu[(n-i+1)h-u]}\operatorname{Res}[F,R_\mu]\,\mathrm{d}L(u)\right)\\
&= \int_0^h \mathrm{e}^{R_\nu(h-u)}\operatorname{Res}[F,R_\nu]\;\Sigma_L\;\operatorname{Res}[F,R_\mu]^{\mathsf{T}}\mathrm{e}^{R_\mu^{\mathsf{T}}(h-u)}\,\mathrm{d}u
\end{aligned}$$

and finally the assertion follows. □

Proposition 3.4.4
Let $U_n^{(h)}$ be the multivariate time series defined in (3.32). An alternative representation for the covariance matrix $\gamma_{U^{(h)}}$ is given by

$$\gamma_{U^{(h)}}(l) = \sum_{k=0}^{p-|l|-1}\Theta_k\Sigma_\varepsilon\Theta_{k+|l|}^{\mathsf{T}}, \qquad \text{for} \quad |l| \le p-1 \tag{3.36a}$$

$$\text{and} \qquad \gamma_{U^{(h)}}(l) = 0, \qquad \text{for} \quad |l| > p-1, \tag{3.36b}$$

where the matrix coefficient $\Theta_0 = I_m$ and Σ_ε is the covariance matrix of the white noise sequence $\varepsilon_n^{(h)}$ from equation (3.28).

Proof. With the moving average representation found in Theorem 3.4.2, we obtain directly the alternative representation for the covariance matrix $\gamma_{U^{(h)}}$ at lag $l = -(p-1),\ldots,p-1$. □

Note that comparing the two representations of the autocovariance matrix yields

$$\sum_{k=0}^{p-|l|-1}\Theta_k\Sigma_\varepsilon\Theta_{k+|l|}^{\mathsf{T}} = \sum_{i=1}^{p-|l|}\left[\sum_{\nu=1}^{p}\sum_{\mu=1}^{p}\left(\mathrm{e}^{h(i+m-1)R_\nu} - \sum_{j=1}^{i+|l|-1}\Phi_j \mathrm{e}^{h(i+|l|-j-1)R_\nu}\right)\right.$$

$$\cdot\, \Sigma_{\nu,\mu}^{(h)} \left(\mathrm{e}^{h(i-1)R_\mu} - \sum_{j=1}^{i-1} \Phi_j \mathrm{e}^{h(i-j-1)R_\mu} \right)^{\mathsf{T}} \Bigg] .$$

3.5. ARMA Representation and Autocovariance Structure of the Integrated Sequence

Now that we have derived the (weak) VARMA representation of a sampled stationary MCARMA process let us turn to the integrated sequence. Hence, we obtain similar results for the integrated sequence using again the decomposition from Section 3.2.

Assume that we have in this section a multivariate integrated CARMA (MICARMA) process given as in Definition 2.5.2. In this context, we consider the integrated sequence, that is an MICARMA process sampled at discrete time points. Its definition is the same as in Brockwell and Lindner [23].

Definition 3.5.1 (Integrated Sequence)
The ***integrated sequence*** $I^{(h)} := (I_n^{(h)})_{n\in\mathbb{N}}$ *is defined as*

$$I_n^{(h)} := I^{(1)}(nh) - I^{(1)}((n-1)h) = \int_{(n-1)h}^{nh} Y(u)\,\mathrm{d}u, \quad n \in \mathbb{Z}, \tag{3.37}$$

where Y *is a stationary multivariate CARMA process satisfying Assumption A1, Assumption A2 and Assumption A3.*

The integrated sequence is thus simply an integrated MCARMA process sampled at integer time points. The covariance matrix of the integrated sequence is given in the following proposition.

Proposition 3.5.2
Let $I^{(h)}$ *be the integrated sequence as defined in Definition 3.5.1. The covariance matrix of* $(I_n^{(h)})_{n\in\mathbb{N}}$ *is given by*

$$\gamma_{I^{(h)}}(l) = \sum_{i=1}^{p}\sum_{j=1}^{p} h^2 \mathrm{e}^{lR_i} \int_0^\infty \mathrm{e}^{yR_i}\, Res[F,R_i]\Sigma_L\, Res[F,R_j]^{\mathsf{T}} \mathrm{e}^{yR_j^{\mathsf{T}}}\,\mathrm{d}y. \tag{3.38}$$

Proof. This follows immediately from Proposition 3.3.2 and Fubini's Theorem. We justify now the use of Fubini's Theorem. By the Hölder inequality we have

$$\begin{aligned}
\mathbb{E}\big[\|Y_i(t+l)Y_j(t)^\mathsf{T}\|\big] &\leq \big(\mathbb{E}\|Y_i(t+l)\|^2\big)^{\frac{1}{2}}\big(\mathbb{E}\|Y_j(t)\|^2\big)^{\frac{1}{2}} \\
&\leq \left(\int_0^\infty \operatorname{tr}\left(\mathrm{e}^{yR_i}\operatorname{Res}[F,R_i]\Sigma_L\operatorname{Res}[F,R_i]^\mathsf{T}\mathrm{e}^{yR_i^\mathsf{T}}\right)\,\mathrm{d}y\right)^{\frac{1}{2}} \\
&\quad\cdot\left(\int_0^\infty \operatorname{tr}\left(\mathrm{e}^{yR_j}\operatorname{Res}[F,R_j]\Sigma_L\operatorname{Res}[F,R_j]^\mathsf{T}\mathrm{e}^{yR_j^\mathsf{T}}\right)\,\mathrm{d}y\right)^{\frac{1}{2}} \\
&\leq C\cdot\left(\int_0^\infty \left\|\mathrm{e}^{yR_i}\right\|^2\,\mathrm{d}y\right)^{\frac{1}{2}} + C\cdot\left(\int_0^\infty \left\|\mathrm{e}^{yR_j}\right\|^2\,\mathrm{d}y\right)^{\frac{1}{2}} \\
&\leq K\cdot\int_0^\infty \mathrm{e}^{-2\cdot y\cdot k}\,\mathrm{d}y = K\cdot\frac{1}{2k}
\end{aligned}$$

for some constants $C, K, k > 0$. Note that R_i as well as R_j have eigenvalues with strictly negative real part and hence the last inequality holds. Lastly, we have

$$\int_{lh}^{(l+1)h}\int_0^h \frac{K}{2y}\mathrm{d}u\mathrm{d}v = \frac{K\cdot h^2}{2\cdot k} < \infty.$$

Now we use Proposition 3.3.2 and Fubini's Theorem and obtain

$$\begin{aligned}
\gamma_{I^{(h)}}(l) &= \int_{lh}^{(l+1)h}\int_0^h \operatorname{Cov}(Y(u),Y(v))\,\mathrm{d}u\,\mathrm{d}v \\
&= \int_{lh}^{(l+1)h}\int_0^h \operatorname{Cov}\left(\sum_{i=1}^p Y_i(t+l),\sum_{j=1}^p Y_j(t)\right)\,\mathrm{d}u\,\mathrm{d}v \\
&= \sum_{i=1}^p\sum_{j=1}^p\int_{lh}^{(l+1)h}\int_0^h \mathrm{e}^{lR_i}\int_0^\infty \mathrm{e}^{yR_i}\operatorname{Res}[F,R_i]\Sigma_L\operatorname{Res}[F,R_j]^\mathsf{T}\mathrm{e}^{yR_j^\mathsf{T}}\,\mathrm{d}y\,\mathrm{d}u\,\mathrm{d}v
\end{aligned}$$

which completes the proof. □

Analogously to Theorem 3.4.2 we proceed in the same manner with the integrated sequence. First, we derive that the integrated sequence is an autoregressive process with p-dependent noise.

Theorem 3.5.3

If Y is an MCARMA process satisfying Assumption A1, Assumption A2 and Assumption A3, the integrated sequence $(I_n^{(h)})_{n\in\mathbb{N}}$ defined in (3.37) is an autoregressive process driven by a p-dependent noise sequence $\widetilde{U}$. It satisfies the difference equation

$$\Phi(B)I_n^{(h)} = \widetilde{U}_n := \sum_{k=1}^p\int_{-h}^0 W_{k,n-k+1}^{(h)}(s)\,\mathrm{d}s, \quad n\in\mathbb{Z}, \tag{3.39}$$

where $\Phi(z)$ *is given by* (3.27) *and* $W_{k,n}^{(h)}(s)$ *given by*

$$W_{k,n}^{(h)}(s) := \sum_{r=0}^{p-1} \int_{(n-1)h+s}^{nh+s} \left(\mathrm{e}^{hrR_i} - \sum_{j=1}^{r} \Phi_j \mathrm{e}^{h(r-j)R_i} \right) \mathrm{e}^{R_i(nh+s-u)}\, Res[F, R_k]\, \mathrm{d}L(u). \tag{3.40}$$

Proof. It follows directly from (3.32) that every stationary MCARMA process Y satisfies the difference equation

$$\Phi(B)Y(nh+s) = U_n^{(h)}(s)$$

for any fixed $s \in [0, h]$ and for all $n \in \mathbb{Z}$, where

$$U_n^{(h)}(s) := \sum_{i=1}^{p} W_{k,n-k+1}^{(h)}(s).$$

Hence, it follows with $I_n^{(h)} = \int_{-h}^{0} Y(nh+s)\,\mathrm{d}s$ that the representation (3.39) is valid.

The summands on the right-hand side of (3.39) are not independent but they depend only on increments of the Lévy process L over the interval $[(n-p-1)h, nh]$ and hence the sequence $(\widetilde{U}_n)_{n\in\mathbb{Z}}$ is p-dependent. □

In order to obtain a (weak) VARMA representation it remains to show that the p-dependent noise sequence from the previous theorem can be represented as a moving average process.

Corollary 3.5.4

If Assumption A1 holds true, the sequence $(\widetilde{U}_n)_{n\in\mathbb{Z}}$ *is a* p*-dependent stationary sequence with mean zero and finite variance. It follows that* $\widetilde{U}_n$ *can be expressed as a moving average process*

$$\widetilde{U}_n = \mathbb{E}\widetilde{U}_0 + \varepsilon_n + \Theta_1 \varepsilon_{n-1} + \ldots + \Theta_p \varepsilon_{n-p}, \tag{3.41}$$

with coefficients $\Theta_i \in M_m(\mathbb{R}[z])$*. The process* $(\varepsilon_n)_{n\in\mathbb{Z}}$ *is a weak white noise sequence with zero mean.*

Moreover, the (moving average) polynomial $\Theta(z) := I_m + \Theta_1 z + \ldots + \Theta_p z^p$ *has no zeros in the interior of the unit disc. Thus,* $(I_n^{(h)})_{n\in\mathbb{Z}}$ *is a weak VARMA(p,q) process with* $q = p$ *or* $q < p$ *if* $\Theta_p = 0_m$*.*

Proof. Again by a multivariate generalization of Brockwell and Davis [20], Proposition 3.2.1, there exist coefficients $\Theta_i \in M_m(\mathbb{R})$, $i = 1, \dots, p$ and a multivariate white-noise sequence $(\varepsilon_n)_{n\in\mathbb{Z}}$, such that (3.41) holds. □

Note that if the driving Lévy process has mean $\mathbb{E}[L(1)] = \mu_L$, the mean of the MCARMA process Y is given by

$$\mathbb{E}[Y(t)] = -A_p^{-1} B_q \mu_L.$$

Hence, for the integrated process the mean is given by

$$\mathbb{E}[I_n^{(h)}] = \int_{(n-1)h}^{nh} \mathbb{E}[Y(t)]\,\mathrm{d}t = -hA_p^{-1} B_q \mu_L. \tag{3.42}$$

Once more, we take a closer look on the covariance matrices.

Proposition 3.5.5
Let $\widetilde{U}_n$ be the multivariate time series defined in (3.39). Then we obtain for the covariance matrix $\gamma_{\widetilde{U}}$ at lag $l = -(p-1), \dots, p-1$

$$\gamma_{\widetilde{U}}(l) = \sum_{i=1}^{p-|l|} \int_{-h}^{0}\int_{-h}^{0} \Bigg[\sum_{\nu=1}^{p}\sum_{\mu=1}^{p} \Bigg(\mathrm{e}^{h(i+|l|-1)R_\nu} - \sum_{j=1}^{i+|l|-1} \Phi_j \mathrm{e}^{h(i+|l|-j-1)R_\nu} \Bigg) \\ \Sigma_{N^{(h)}}(s) \Bigg(\mathrm{e}^{h(i-1)R_\mu} - \sum_{j=1}^{i-1} \Phi_j \mathrm{e}^{h(i-j-1)R_\mu} \Bigg)^{\mathsf{T}} \Bigg] \,\mathrm{d}s\,\mathrm{d}v, \tag{3.43}$$

where $\Sigma_{\mu,\nu}^{(h)}(s,r) := Cov\left(N_{\nu,n-i+1}^{(h)}(s), N_{\mu,n-i+1}^{(h)}(r)\right)$.

Proof. We obtain

$$\begin{aligned}
\gamma_{\widetilde{U}}(l) &= \mathrm{Cov}\left(\widetilde{U}_{n+|l|}, \widetilde{U}_n\right) \\
&= \mathrm{Cov}\Bigg(\int_{-h}^{0} W_{1,n+|l|}^{(h)}(s)\,\mathrm{d}s + \dots + \int_{-h}^{0} W_{p,n+|l|-p+1}^{(h)}(s)\,\mathrm{d}s, \\
&\qquad \int_{-h}^{0} W_{1,n}^{(h)}(r)\,\mathrm{d}r + \dots + \int_{-h}^{0} W_{p,n-p+1}^{(h)}(r)\,\mathrm{d}r \Bigg) \\
&= \sum_{i=1}^{p-|l|} \int_{-h}^{0}\int_{-h}^{0} \mathrm{Cov}\left(W_{i+|l|,n-i+1}^{(h)}(s), W_{i,n-i+1}^{(h)}(r) \right)\,\mathrm{d}s\,\mathrm{d}r \\
&\overset{(3.40)}{=} \sum_{i=1}^{p-|l|} \int_{-h}^{0}\int_{-h}^{0} \Bigg[\sum_{\mu,\nu=1}^{p} \Bigg(\mathrm{e}^{h(i+|l|-1)R_\nu} - \sum_{j=1}^{i+|l|-1} \Phi_j \mathrm{e}^{h(i+|l|-j-1)R_\nu} \Bigg)
\end{aligned}$$

$$\cdot \operatorname{Cov}\left(\int_{(n-i)h+s}^{(n-i+1)h+s} \mathrm{e}^{R_\nu[(n-i+1)h+s-u]} \operatorname{Res}[F, R_\nu] \, \mathrm{d}L(u), \right.$$
$$\left. \int_{(n-i)h+r}^{(n-i+1)h+r} \mathrm{e}^{R_\mu[(n-i+1)h+r-u]} \operatorname{Res}[F, R_\mu] \, \mathrm{d}L(u) \right)$$
$$\left. \cdot \left(\mathrm{e}^{h(i-1)R_\mu} - \sum_{j=1}^{i-1} \Phi_j \mathrm{e}^{h(i-j-1)R_\mu} \right)^{\mathsf{T}} \right] \mathrm{d}s \, \mathrm{d}r,$$

$$\text{where} \quad \Sigma_{\mu,\nu}^{(h)}(s,r) := \operatorname{Cov}\left(\int_{(n-i)h+s}^{(n-i+1)h+s} \mathrm{e}^{R_\nu[(n-i+1)h+s-u]} \operatorname{Res}[F, R_\nu] \, \mathrm{d}L(u), \right.$$
$$\left. \int_{(n-i)h+r}^{(n-i+1)h+r} \mathrm{e}^{R_\mu[(n-i+1)h+r-u]} \operatorname{Res}[F, R_\mu] \, \mathrm{d}L(u) \right).$$

The order of integration can be exchanged due to Fubini's theorem. □

Last but not least we consider the autocovariance structure of the p-dependent stationary sequence $\widetilde{U}$.

Proposition 3.5.6
Let $(\widetilde{U}_n)_{n\in\mathbb{N}}$ be the multivariate time series defined in (3.39) and the matrix coefficients Φ_i for $i = 0, \dots, p$ are defined as in (3.27). Then the autocovariance function of the process $\widetilde{U}$ is given by

$$\gamma_{\widetilde{U}}(l) = \begin{cases} \sum_{i=0}^{p} \sum_{j=0}^{p} \Phi_i \gamma_{I^{(h)}}(|l| - j + i) \Phi_j^{\mathsf{T}}, & \text{if } |l| \in \{0, \dots, p\}, \\ 0, & \text{otherwise}. \end{cases} \tag{3.44}$$

Proof. This follows immediately from (3.39) and Fubini's theorem. □

3.6. Conclusion

Often continuous-time models are only observed at discrete time points. Hence, discretised versions of the continuous time model are of great interest. We have derived in this chapter a decomposition of stationary Lévy driven MCARMA(p, q) processes into the sum of dependent MCAR processes, namely $Y(t) = \sum_{k=1}^{p} Y_k(t)$. This decomposition is also the key result, which enabled us to derive a weak VARMA representation of order $(p, p-1)$ of a sampled stationary MCARMA(p, q) process. We considered the stationary MCARMA process Y observed at equidistant time points.

We also derived a weak VARMA representation of order (p,p) for the integrated sequence $\int_{(n-1)h}^{nh} Y(t)\,\mathrm{d}t$, which is for example of interest for stochastic volatility modelling. Moreover, we investigated the autocovariance structure of these discretised models. These results could not only be useful in the estimation procedure, but also in finding a criterion for strict stationarity as in the univariate case.

3.7. Appendix: Partial Fraction Expansion for Repeated Solvents

We have recalled a partial fraction expansion without repeating right solvents. However, a right solvent could also appear more often than once. This chapter presents the extension to these cases. For this purpose, we use a partial fraction expansion for repeated right solvents. Results on factorization of matrix polynomials with repeated solvents can be found for example in Maroulas [67].

Definition 3.7.1
For a matrix polynomial $A(\lambda)$ of degree n and order m we define

$$A^{(k)}(\lambda) := \frac{d^k}{d\lambda^k} A(\lambda).$$

*A matrix R is defined to be a **right solvent with multiplicity** ν if and only if it is a common solvent of the equations*

$$A(R) = 0_m, \quad A^{(1)}(R) = 0_m, \quad \dots, \quad A^{(\nu-1)}(R) = 0_m \tag{3.45}$$

with $A^{(\nu)}(R) \neq 0_m$. Thus we can write for the matrix polynomial $A(\lambda)$

$$A(\lambda) = A_\nu(\lambda)(\lambda I_m - R)^\nu. \tag{3.46}$$

Another special matrix form is the so-called Vandermonde matrix, which is closely related to the companion matrix and as a consequence with matrix polynomials.

Definition 3.7.2
*Given $m \times m$ matrices $S_1, \dots, S_n$, the **block Vandermonde matrix** is defined by*

$$V(S_1, \dots, S_n) = \begin{pmatrix} I_m & I_m & \dots & I_m \\ S_1 & S_2 & \dots & S_n \\ \vdots & \vdots & & \vdots \\ S_1^{n-1} & S_2^{n-1} & \dots & S_n^{n-1} \end{pmatrix} \in M_{mn}(\mathbb{C}). \tag{3.47}$$

The Vandermonde matrix is extended in the next definition, which is necessary in order to characterize criteria for a factorization with repeated right solvents.

Definition 3.7.3
Suppose $R_1, \ldots, R_q$ are right solvents of the matrix polynomial $A(\lambda)$ with multiplicities $\nu_1, \ldots, \nu_q$ respectively. We define an $mn \times m(\nu_1 + \ldots + \nu_q)$-dimensional ***confluent Vandermonde matrix*** *$W(R_1, \ldots, R_q) = W$ by $W = [W_1, \ldots, W_q]$, where*

$$W_k = \begin{pmatrix} I_m & 0_m & 0_m & \ldots & 0_m \\ R_k & I_m & 0_m & & \cdot \\ R_k^2 & 2R_k & I_m & \ddots & \cdot \\ R_k^3 & 3R_k^2 & 3R_k & \ddots & 0_m \\ \vdots & \vdots & \vdots & & I_m \\ \vdots & \vdots & \vdots & & \vdots \\ R_k^{n-1} & (n-1)R_k^{n-2} & \binom{n-1}{2}R_k^{n-3} & \ldots & \binom{n-1}{\nu_k-1}R_k^{n-\nu_k} \end{pmatrix}. \tag{3.48}$$

Theorem 3.7.4 (Maroulas (1985), Theorem 2.1)
If the confluent Vandermonde matrix W as defined in (3.48) is left invertible, then there exists a matrix polynomial $A(\lambda)$ having roots $R_1, \ldots, R_q$ with multiplicities $\nu_1, \ldots, \nu_q$ respectively.

The invertibility of the confluent Vandermonde matrix can be guaranteed if the eigenvalues are located in a certain way. A sufficient criterion is given in the next theorem.

Theorem 3.7.5 (Maroulas (1985), Theorem 3.4)
Let $R_1, \ldots, R_q$ solvents of a matrix polynomial $A(\lambda)$, of multiplicities $\nu_1, \ldots, \nu_q$ respectively. The matrix W is invertible if and only if

$$\sigma(A) = \bigcup_{j=1}^{q} \sigma(R_j)$$

and

$$\sigma(R_j) \cap \sigma(R_i) = \emptyset$$

for $j, i = 1, \ldots, q,\ j \neq i$.

This theorem gives an intuitive interpretation, when we have a complete set of right solvents with multiplicities. The eigenvalues corresponding to a right solvent are not allowed to belong to another right solvent, i.e. the spectra of right solvents must have an empty intersection. Additionally, the union of the spectra of the right solvents must be the spectra of the original matrix A.

A necessary and sufficient condition to have a root R of multiplicity ν is given by the following theorem.

Theorem 3.7.6 (Maroulas (1985), Theorem 4.1)
The monic polynomial $A(\lambda)$ has a right divisor $(\lambda I_m - R)^\nu$ if and only if there exists an invariant subspace $\mathcal{M}$ of the companion matrix A_C of $A(\lambda)$ of the form

$$\mathcal{M} = im(\mathcal{M}_0) \oplus im(\mathcal{M}_1) \oplus \ldots \oplus im(\mathcal{M}_{\nu-1}), \tag{3.49}$$

where

$$\mathcal{M}_0 = \begin{pmatrix} 0_m & \cdots & 0_m & I_m & \binom{k+1}{k} R^\intercal & \cdots & \binom{n-1}{k} (R^{n-k-1})^\intercal \end{pmatrix}^\intercal .$$

The set of solvents $R_1, \ldots, R_q$ is called complete for $A(\lambda)$ if and only if $\nu_1 + \ldots + \nu_q = n$, where n is the degree of $A(\lambda)$. This means the sum of the multiplicities must add up to the degree of the matrix polynomial.

Theorem 3.7.7 (Maroulas (1985), Theorem 5.2.)
The set of solvents $R_1, \ldots, R_q$, $(1 \leq q < n)$ form a complete set of solvents of $A(\lambda)$ with $\nu_1 + \ldots + \nu_q = n$. If the matrices $W(R_1, \ldots, R_k)$, $k = 2, 3, \ldots q$ are invertible, then

$$A(\lambda) = T_q(\lambda) \cdots T_2(\lambda)(\lambda I_m - R_1)^\nu, \tag{3.50}$$

where

$$T_k(\lambda) = (\lambda I_m - Z_{k,\nu_k}) \cdots (\lambda I_m - Z_{k,2})(\lambda I_m - Z_{k,1}), \quad k = 2, \ldots, q \tag{3.51}$$

and the matrices $Z_{k,j}$ are similar to R_k for $j = 1, \ldots, \nu_k$.

The block partial fraction expansion with repeated block poles is given in Shieh, Chang and McInnis [93] or Levya-Ramos [61]. The next theorem is the analogue version of Theorem 3.2.12. It gives the partial fraction expansion with respect to repeated right solvents.

Theorem 3.7.8 (Shieh et. al. (1986), Theorem)
Let $R_1, \ldots, R_\mu$ be a complete set of right solvents of the n^{th} degree m^{th} order monic matrix polynomial $A(\lambda)$, where μ is the number of distinct solvents and ν_i is the multiplicity of R_i, with $n = \sum_{i=1}^{\mu} \nu_i$. The block partial fraction expansion of the

irreducible strictly proper rational left λ-matrix is given by

$$F(\lambda) = A(\lambda)^{-1}B(\lambda) = \sum_{i=1}^{\mu}\sum_{j=1}^{\nu_i}(\lambda I_m - R_i)^{-j}F_{i,j}, \tag{3.52}$$

where $F_{i,j} \in M_{m,m}(\mathbb{C})$ are the matrix residues associated with R_i.

A formula for the matrix residue is given in Section 6 of Levya-Ramos [61] (c.f. Equation (6.13)). For a complete set of distinct solvents we have

$$\begin{pmatrix} Res[F,R_1] \\ Res[F,R_2] \\ \vdots \\ Res[F,R_n] \end{pmatrix} = V(R_1,\dots,R_n)^{-1} \begin{pmatrix} I_m & 0_m & \dots & & 0_m \\ A_1 & I_m & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ A_{n-2} & & \ddots & I_m & 0_m \\ A_{n-1} & A_{n-2} & \dots & A_1 & I_m \end{pmatrix}^{-1} \begin{pmatrix} B_0 \\ B_2 \\ \vdots \\ B_{n-1} \end{pmatrix} \tag{3.53}$$

and for repeated solvents the formula changes to

$$\begin{pmatrix} F_{1,1} \\ \vdots \\ F_{1,\nu_1} \\ \vdots \\ F_{\mu,1} \\ \vdots \\ F_{\mu,\nu_\mu} \end{pmatrix} = W(R_1,\dots,R_\mu)^{-1} \begin{pmatrix} I_m & 0_m & \dots & & 0_m \\ A_1 & I_m & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ A_{n-2} & & \ddots & I_m & 0_m \\ A_{n-1} & A_{n-2} & \dots & A_1 & I_m \end{pmatrix}^{-1} \begin{pmatrix} B_0 \\ B_2 \\ \vdots \\ B_{n-1} \end{pmatrix}. \tag{3.54}$$

CHAPTER 4

Characterization of Cointegrated MCARMA Processes

4.1. Introduction

Many time series do not behave in a stationary way, they rather follow a stochastic trend. Such time series are obviously non-stationary and fluctuate around a long-run equilibrium. It was Clive Granger, who showed that statistical inference of such time series with the classical stationary methodology can lead to inadequate results. He coined the term cointegration for time series showing such a behavior (c.f. [42]).

The class of cointegrated time series is a subclass of non-stationary time series with the characterizing property that some linear combinations of a multivariate time series can be stationary. Examples of cointegrated time series include e.g. exchange rates or the connection between short and long-term interest rates. The seminal works by Granger in 1981 [42] and Engle and Granger in 1987 [33] lay the foundation for the field of cointegration analysis. Robert F. Engle and Clive Granger were awarded with the Nobel prize in 2003. Clive Granger was awarded for his discovery of the concept of cointegration. The official motivation of the committee was for his "methods of analyzing economic time series with common trends (cointegration)". Robert Engle was awarded for his developments in the field of autoregressive conditional heteroscedasticity models.

Also worthy of note is the work of Johansen [53] on discrete-time cointegrated vector autoregressive processes. Johansen presented in 1991 a likelihood approach to such cointegrated VAR models and developed a cointegration rank test in order to determine the number of cointegration relations in a model. His textbook [54] about cointegration became the standard literature in this field.

Cointegration in continuous time started being of interest in the early 1990s. In 1991 Phillips [77] considered spectral regression methods for error correction models and cointegrated systems in special triangular form in the continuous-time framework. In this work, Phillips regarded stochastic differential equations driven by a differentiable stationary process. The connection between cointegrated discrete-time models and continuous-time models were also analyzed by Chambers [25], where discrete-time representation of cointegrated continuous-time models were considered.

The special case of a p-dimensional Gaussian CAR$(1,0)$ process was considered in Kessler and Rahbek [57]. Comte [27] derived a characterization of integrated and cointegrated processes in continuous-time and particularly derived an error correction form and characterization of cointegration for CAR(p) processes. We extend this characterization of cointegration to general MCARMA(p,q) processes in Section 4.2. The cointegrated model of this chapter includes the models of Comte [27] and also the model of Phillips [77].

We briefly recall in Section 4.2 the concept of cointegration in discrete-time for vector autoregressive moving average processes. Then we utilize the definition of an integrated process of Comte [27], which does not necessarily need differentiability. In this definition, a continuous-time process is integrated if it has stationary increments but is itself non-stationary. Moreover, we characterize cointegration via its MCARMA form with the autoregressive polynomial $P(z)$ and the moving average polynomial $Q(z)$. We see that the property of cointegration is related to certain matrix coefficients of the autoregressive polynomial $P(z)$.

These results are very helpful for understanding the concept of cointegration for MCARMA processes. However, it is rather difficult to derive properties for general cointegrated MCARMA processes using the MCARMA representation. For this reason, we use continuous-time cointegrated state space models in Section 4.3. The continuous-time state space models have the advantage that we can derive an utterly helpful representation for the cointegrated model, which decouples the non-stationary and stationary part into subsystems. Thus, we can interpret the cointegrated process as a sum of a Lévy process and a stationary MCARMA process. We investigate the probabilistic properties of this model. Several of the properties can be derived due to the decoupled form.

After the continuous-time model, we consider the discrete-time sampled version of this process. The sampled process is observed at equidistant time points and has therefore an i.i.d. noise. The process has an analogue representation and consists thus of the sum of a random walk and a stationary process. We investigate the probabilistic properties of the sampled process as well.

In order to obtain an error correction form of the sampled process, we apply a linear filter to the model. We obtain with the help of the so-called Kalman filter the linear innovations, which is a white noise sequence. The name Kalman filter dates back to Rudolf E. Kalman, who developed a two-step algorithm, which produces estimates of unknown variables (see Kalman [56]). The Kalman filter is widely used in engineering sciences for state estimations of linear systems. In Section 4.4 the results of the Kalman filter are summarized and we obtain a error correction form with respect to the linear innovations.

It is not obvious how to obtain an error correction form for the sampled process. With the help of the linear innovations $\varepsilon^{(h)}$ we obtain the representation $\varepsilon_n^{(h)} = \Pi Y_{n-1}^{(h)} + \overline{k}(B)\Delta Y_n^{(h)}$, where $\overline{k}(z)$ is a linear filter. This error correction form has resemblance to the original error correction form presented by Engle and Granger [33] for VAR models. The difference to the classical result is that we have an infinite order linear filter $\overline{k}(z)$. We show that the cointegration information is contained in parts of the filter, which is helpful for estimation procedure later on and we are able to calculate the likelihood function with the help of the linear innovations. The applicability of the Kalman filter to state space models with unit roots is guaranteed by the considerations in Appendix 4.6.

4.2. Cointegrated Lévy Driven MCARMA processes

Before we start with cointegrated multivariate CARMA processes, we briefly consider their discrete time analogue, the vector ARMA (VARMA) processes. We recall the definition of cointegration for such processes. For more details on cointegrated processes in discrete-time see e.g. Johansen [54], Lütkepohl [62] or Reinsel [83].

The general form of an m-dimensional vector autoregressive moving average process $(Y_n)_{n\in\mathbb{N}}$ is given by the combination of a p^{th}-order vector autoregressive (VAR) process and a q^{th}-order moving average (MA) process

$$Y_n - \Phi_1 Y_{n-1} - \ldots - \Phi_p Y_{n-p} = \varepsilon_n - \Theta_1 \varepsilon_{n-1} - \ldots - \Theta_q \varepsilon_{n-q}, \quad n \in \mathbb{N}$$

or briefly $\Phi(B)Y_n = \Theta(B)\varepsilon_n$, for $n \in \mathbb{Z}$. The noise sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ is an m-dimensional zero mean white noise with nonsingular covariance matrix Σ_ε. The matrix coefficients are given by $\Phi_i, \Theta_j \in M_m(\mathbb{R})$, $i = 1, \dots, p$, $j = 0, \dots q$, where the autoregressive polynomial is given by $\Phi(B) := I_m - \Phi_1 B - \dots - \Phi_p B^p$ and the moving average polynomial by $\Theta(B) := I_m - \Theta_1 B - \dots - \Theta_q B^q$. Let B denote the backshift operator. The process Y is called a vector autoregressive moving average process of order (p, q) (VARMA(p, q)).

The VARMA(p, q) process Y is invertible if all roots of $\det(\Theta(z)) = 0$ are greater than one in absolute value, and stationary if all roots of $\det(\Phi(z)) = 0$ are greater than one in absolute value. In the case of stationarity, we have a moving average representation of Y given by $Y_n = \Psi(B)\varepsilon_n$, where $\Psi(z) = \Phi(z)^{-1}\Theta(z) = \sum_{i=0}^{\infty} \Psi_i z^i$. In this case, the covariance matrices are given by $\Gamma(l) = \sum_{i=0}^{\infty} \Psi_i \Sigma_\varepsilon \Psi_{i+l}^{\mathsf{T}}$.

In order to have integrated and cointegrated processes the roots of $\det(\Phi(z)) = 0$ must be greater or equal to one. Thus, we have no longer a stationary time series.

Definition 4.2.1
*If the process $(Y_n)_{n\in\mathbb{N}}$ is non-stationary but its first difference $((1 - B)Y_n)_{n\in\mathbb{N}}$ is stationary, we call $(Y_n)_{n\in\mathbb{N}}$ **integrated of order one**. If further there exists a vector β such that $\beta^{\mathsf{T}} Y$ is stationary, we say the process Y is **cointegrated** with cointegration vector β.*

Assume that the autoregressive polynomial of the process Y has $c < m$ unit roots, then the matrix $\Phi(1)$ has rank $r = m - c$ and r is the cointegration rank. This implies that we have r linearly independent vectors β_i such that $\beta^{\mathsf{T}} Y$ is stationary, where the matrix β consists of the vectors β_i for $i = 1, \dots, r$.

After this brief excursion to cointegration in discrete time, we now define cointegration in continuous-time. Before we are able to do this, we need first an integrated time series. Instead of the approach in Section 2.5, where the stationary process is „integrated“ (c.f. Definition 2.5.1), we approach cointegration differently by starting with a non-stationary process. An integrated d-dimensional MCARMA process in the sense of Definition 2.5.1 cannot be cointegrated since zero is an eigenvalue with multiplicity d and hence the cointegration rank is zero. However, the different definition we use in this chapter is flexible in the multiplicity of the eigenvalue zero and thus the cointegration rank can be modeled freely. The following alternative definition of integrated CARMA processes can be found in Comte [27].

Definition 4.2.2
A process $(Y(t))_{t\geq 0}$ with no deterministic component, which is non-stationary but

has stationary increments is said to be ***continuously integrated of order one****, denoted by* $Y \sim \mathcal{I}^c(1)$*. Further, a process* Y *is* ***integrated of order*** b*, denoted by* $Y \sim \mathcal{I}^c(b)$ *if its increments are continuously integrated of order* $b-1$*.*

Similarly as in the discrete-time setting, cointegration is defined by the existence of stationary linear combinations of the integrated components.

Definition 4.2.3
Let $(Y(t))_{t\geq 0}$ *be an integrated MCARMA process of order* 1*, i.e.* $Y(t) \sim \mathcal{I}^c(1)$*. We call* Y ***continuously cointegrated*** *with cointegration vector* $\beta \neq 0$*,* $\beta \in \mathbb{R}^d$*, if the linear combination* $(\beta^\mathsf{T} Y(t))_{t\geq 0}$ *is stationary.*

Denote by $Y \sim \mathcal{CI}(1,1)$ *that* Y *is cointegrated of order one. The* ***cointegration rank*** *is the number of linearly independent cointegrating relations and the space spanned by all linear independent cointegration vectors is the* ***cointegration space****.*

Higher orders of cointegration are then defined in the obvious way given in the next definition.

Definition 4.2.4
The $\mathcal{I}^c(b)$ *process* Y *is called* ***continuously cointegrated of order*** $(b,\bar{b})$ *(*$\mathcal{CI}(b,\bar{b})$*) for* $b \in \{1,\dots,d-1\}$*,* $\bar{b} \in \{1,\dots,b\}$ *with cointegration vectors* β *if there exists a non-zero vector* $\beta \in \mathbb{R}^d$ *such that* $\beta^\mathsf{T} Y \sim \mathcal{I}(b-\bar{b})$*, and there exists no non-zero vector* $\beta' \in \mathbb{R}^d$ *such that* $\beta'^\mathsf{T} Y \sim \mathcal{I}(b-b')$ *with* $b' > \bar{b}$*.*

Let us recall some basic results from Comte [27] on integrated and cointegrated processes, respectively. First, we see that the definition of cointegration using stationary increments includes the case of differentiation, which is the obvious analog of the first differences in discrete time.

Proposition 4.2.5 (Comte (1999), Proposition 1)
Let $(Y(t))_{t\geq 0}$ *be a process with finite second moments and its first-order mean square derivative* DY *exists. Then* $Y(t) \sim \mathcal{I}^c(1)$ *if and only if* Y *is non-stationary and* DY *is stationary.*

Furthermore, the integration and cointegration property of the continuous-time model directly transfers to its sampled version.

Proposition 4.2.6 (Comte (1999), Proposition 3)
Let the continuous-time process Y *be continuously integrated of order* b*,* $b \in \mathbb{N}$*, then the discrete-time process* $(Y_n^{(h)})_{n\in\mathbb{N}}$ *is integrated of order* b *(in the sense of Definition 4.2.3).*

Let $Y \sim \mathcal{CI}(b, \bar{b})$, for $b \geq \bar{b}$ and $b, \bar{b} \in \mathbb{N}$, then the discrete-time process $(Y_n^{(h)})_{n\in\mathbb{N}}$ is cointegrated of order $(b, \bar{b})$ (in the sense of Definition 4.2.3).

Remark 4.2.7
Theorem 3.4.2 can not be adapted to integrated processes in the sense of Definition 4.2.2, due to the non-stationarity of Y. Furthermore, an indirect approach is also not possible, because Y is a non-Markovian process. Hence, the information of the initial condition of Y might be lost, when we apply Theorem 3.4.2 to DY and then integrate DY again.

As from now we concentrate on the case $\mathcal{CI}(1, 1)$, i.e. on continuously cointegrated processes of order one. The differentiation operator should be understood as the mean square differential operator. Recall that an MCARMA(p,q) process is $p - q - 1$ times differentiable.

The following result is an extension of a result by Comte [27] for CAR(p) processes.

Proposition 4.2.8
Cointegration arises for a d-dimensional MCARMA(p, q) process Y if and only if $P(0) = A_p$ is singular, i.e. the corresponding companion matrix is not invertible. The error correction form in continuous time is given by

$$P^*(D)DY(t) = -A_pY(t) + Q(D)DL(t), \tag{4.1}$$

where the polynomial P^* is given by

$$P^*(z) := \frac{P(z) - A_p}{z}. \tag{4.2}$$

Proof. The crucial role of the matrix coefficient A_p is an immediate consequence of results known about the companion form, see Definition 3.2.4. The error correction form is then derived straightforwardly. □

The following result characterizes cointegration with respect to the matrix coefficients, where the coefficients A_p and A_{p-1} play a special role. The following result is an extension of Proposition 7 in Comte [27] to MCARMA processes.

However, we do not (yet) derive a moving average representation as in Comte [27]. Their idea of the proof cannot be extended to general MCARMA processes. On the one hand, we have the problem that the matrix multiplication is not commutative and hence we cannot use the same assumptions. On the other hand, the proof would end up with a process which is not an MCARMA process anymore.

Nonetheless, we get a characterization of cointegration. Denote by $A^{\perp} \in M_{d,r}(\mathbb{R})$ the full rank matrix satisfying $A^{\perp\mathsf{T}}A^{\perp} = I_r$ and $A^{\mathsf{T}}A^{\perp} = 0_{r\times(d-r)}$, this means $A^{\perp}$ is the orthogonal complement of A.

Theorem 4.2.9

Let Y be a solution of the differential equation

$$P(D)Y(t) = Q(D)DL(t), \quad t \geq 0, \tag{4.3}$$

with autoregressive polynomial

$$P(z) = I_m z^p + A_1 z^{p-1} + \ldots + A_p,$$

and moving average polynomial

$$Q(z) = B_0 z^q + B_1 z^{q-1} + \ldots + B_q,$$

with $Y(0) = 0$ as the initial condition. Let the following assumptions hold:

B1: If $\det P(z) = 0$ then either $\Re(z) < 0$ or $z = 0$.

B2: $rank(A_p) = rank(P(0)) = r < d$ and $A_p = \alpha\beta^{\mathsf{T}}$, where the adjustment matrix $\alpha \in M_{d,r}(\mathbb{R})$ and cointegration matrix $\beta \in M_{d,r}(\mathbb{R})$ have full rank r.

B3: $P'(0) = P^*(0) = A_{p-1}$ is such that the matrix $\alpha^{\perp\mathsf{T}}A_{p-1}\beta^{\perp}$ is of dimension $(d-r)\times(d-r)$ with full rank $(d-r)$.

Then we have that

i) the process DY is stationary,

ii) the process $\beta^{\mathsf{T}}Y$ is stationary,

and thus the Lévy driven MCARMA process Y is cointegrated of order one.

Proof. By multiplying (4.1) with α and $\alpha^{\perp\mathsf{T}}$ we obtain with $A_p = \alpha\beta^{\mathsf{T}}$ and $\alpha^{\perp\mathsf{T}}\alpha = 0_{(d-r)\times r}$ the following equations

$$\begin{aligned} \alpha^{\mathsf{T}}Q(D)DL(t) &= -\alpha^{\mathsf{T}}\alpha\beta^{\mathsf{T}}Y(t) + \alpha^{\mathsf{T}}P^*(D)DY(t), \\ \alpha^{\perp\mathsf{T}}Q(D)DL(t) &= \alpha^{\perp\mathsf{T}}P^*(D)DY(t). \end{aligned} \tag{4.4}$$

Since the system (4.4) is not invertible in Y and DY, we define new processes

$$Z(t) := (\beta^{\mathsf{T}}\beta)^{-1}\beta^{\mathsf{T}}Y(t) \quad \text{and} \quad V(t) := (\beta^{\perp\mathsf{T}}\beta^{\perp})^{-1}\beta^{\perp\mathsf{T}}DY(t) \quad \text{for } t \geq 0$$

and obtain thereby invertibility. The matrix $R := (\beta, \beta^\perp) \in M_{d,d}(\mathbb{R})$ of rank d satisfies

$$R(R^\mathsf{T} R)^{-1} R^\mathsf{T} = \beta(\beta^\mathsf{T}\beta)^{-1}\beta^\mathsf{T} + \beta^\perp(\beta^{\perp\mathsf{T}}\beta^\perp)^{-1}\beta^{\perp\mathsf{T}} = I_d \tag{4.5}$$

since it is the sum of the projection matrices on the range and the null space of β. Moreover, for $\bar{\beta} := \beta(\beta^\mathsf{T}\beta)^{-1} \in M_{d,r}(\mathbb{R})$ and $\bar{\beta}^\perp := \beta^\perp(\beta^{\perp\mathsf{T}}\beta^\perp)^{-1} \in M_{d,d-r}(\mathbb{R})$ we have due to (4.5) that $\beta\bar{\beta}^\mathsf{T} + \beta^\perp\bar{\beta}^{\perp\mathsf{T}} = I_d$ holds. Furthermore, we have

$$DY(t) = (\beta\bar{\beta}^\mathsf{T} + \beta^\perp\bar{\beta}^{\perp\mathsf{T}})DY(t) = \beta DZ(t) + \beta^\perp V(t). \tag{4.6}$$

Rewriting system (4.4) with the newly defined variables yields

$$\begin{aligned}\alpha^\mathsf{T} Q(D)DL(t) &= -\alpha^\mathsf{T}\alpha(\beta^\mathsf{T}\beta)Z(t) + \alpha^\mathsf{T} P^*(D)\beta DZ(t) + \alpha^\mathsf{T} P^*(D)\beta^\perp V(t),\\ \alpha^{\perp\mathsf{T}} Q(D)DL(t) &= \alpha^{\perp\mathsf{T}} P^*(D)\beta DZ(t) + \alpha^{\perp\mathsf{T}} P^*(D)\beta^\perp V(t).\end{aligned}$$

Rearranging the last expressions leads to

$$\widetilde{P}(D)(Z(t)^\mathsf{T}, V(t)^\mathsf{T})^\mathsf{T} = (\alpha, \alpha^\perp)^\mathsf{T} Q(D)DL(t), \tag{4.7}$$

where the matrix polynomial $\widetilde{P}$ is given by

$$\widetilde{P}(z) := \begin{pmatrix} \alpha^\mathsf{T}\alpha(\beta^\mathsf{T}\beta) + \alpha^\mathsf{T} P^*(z)\beta z & \alpha^\mathsf{T} P^*(z)\beta^\perp \\ \alpha^{\perp\mathsf{T}} P^*(z)\beta z & \alpha^{\perp\mathsf{T}} P^*(z)\beta^\perp \end{pmatrix}. \tag{4.8}$$

By assumption *B2* and *B3* we have

$$\begin{aligned}\det(\widetilde{P}(0)) &= \det\begin{pmatrix} \alpha^\mathsf{T}\alpha(\beta^\mathsf{T}\beta) & \alpha^\mathsf{T} P^*(0)\beta^\perp \\ 0_{(d-r)\times r} & \alpha^{\perp\mathsf{T}} P^*(0)\beta^\perp \end{pmatrix}\\ &= \det(\alpha^\mathsf{T}\alpha)\det(\beta^\mathsf{T}\beta)\det(\alpha^{\perp\mathsf{T}} P^*(0)\beta^\perp) \neq 0,\end{aligned}$$

where the matrices in the last line all have full rank and consequently a non-zero determinant. Then for $z \neq 0$, we can see due to (4.2) and (4.8) that $\widetilde{P}(z) = (\alpha, \alpha^\perp)^\mathsf{T} P(z)(\beta, \beta^\perp/z)$ and thus

$$\det(\widetilde{P}(z)) = \frac{1}{z^{d-r}}\det(\alpha, \alpha^\perp)^\mathsf{T}\det(P(z))\det(\beta, \beta^\perp) \neq 0.$$

Thus $\widetilde{P}$ has the same roots as P, except the null ones and the non-zero roots are assumed to have negative real part due to *B1*. Hence, the process (Z, V) is asymptotically (exponentially) stable and has a stationary solution. The process DY is also stationary, as a linear combination of stationary processes $DY(t) = \beta DZ(t) + \beta^{\perp} V(t)$, which finishes the proof of *i)*.

Besides $\beta^{\mathsf{T}} Y(t) = (\beta^{\mathsf{T}}\beta) Z(t)$ holds and therefore stationarity of $\beta^{\mathsf{T}} Y(t)$ follows. Consequently, we have shown *ii)*.

Thus, the process is continuously cointegrated and this completes the proof. □

We make now some remarks on the last result and its implications on cointegration for MCARMA models.

Remark 4.2.10
The assumption in the Theorem 4.2.9 have the following relevance:

- Assumption B1 guarantees that the process is non-stationary.
- Assumption B2 guarantees that there exist stationary linear combinations.
- Assumption B3 guarantees that the process is only integrated of order one and not of higher order.

Remark 4.2.11
It is not possible to adopt the proof of Comte [27], Proposition 7, to obtain a moving average representation for the cointegrated process of the form $Y(t) = C(\infty)L(t) + \int_{-\infty}^{t} \widetilde{C}(t-s)\,\mathrm{d}L(s)$, since we have now the matrix polynomial $\widetilde{P}(z)$, whose first matrix coefficient is not the identity matrix. Therefore the theory on MCARMA processes derived by Marquardt and Stelzer [69] cannot be applied.

However, we can make use of the state space representation and derive a moving average representation in Section 4.3. Additionally, we even get a more precise representation, because we are going to know the exact representation of $C(\infty)$ and $\widetilde{C}$. Furthermore, we can decouple the system into the non-stationary and stationary part which is nothing else but a stationary MCARMA process as defined in Marquardt and Stelzer [69].

Remark 4.2.12
If the cointegration rank is zero, i.e. $A_p = 0_{d\times d}$, we have no cointegration vector and thus the process is not cointegrated. However, the process is integrated of order one in the sense of Definition 2.5.2. On the other hand, if the rank of A_p is equal to d, i.e. A_p is of full rank, the process is stationary. This means that all eigenvalues have negative real part and *B1* is satisfied. Additionally, *B3* is automatically satisfied, whereas *B2* is violated. Therefore cointegration arises, when the rank of A_p satisfies

$0 < r < d$. Hence, it depends on the matrix A_p if we have a stationary, integrated or even cointegrated MCARMA process.

Remark 4.2.13
There are two natural ways to define an integrated MCARMA process. Both ways have the property that the process remains in the class of MCARMA processes. This can be seen by taking a closer look to the defining differential equations.

1. The first method starts with a stationary m-dimensional MCARMA(p,q) process Y, then an integral is taken to obtain the integrated process given by $I(t) = \int_0^t Y(s)\,\mathrm{d}s$. Assume that the process Y satisfies $P(D)Y(t) = Q(D)DL(t)$. Hence, the equation for the integrated process is

$$P^*(D)I(t) = P(D)DI(t) = PY(t) = Q(D)DL(t),$$

   where $DI(t) = Y(t)$ and $P^*(D) := zP(z)$. The order of $P^*(z)$ is $p^* := p + 1$. Obviously, I is also an MCARMA process with parameters (p^*, q). This, implies that $p^* > q$.

2. However, the second method takes as a starting point a non-stationary m-dimensional MCARMA(p, q) process X, where DX is stationary. Assume, that the process X satisfies $P(D)X(t) = Q(D)DL(t)$, $t \geq 0$. Therefore, we have

$$P(D)DX(t) = D[P(D)X(t)] = D[Q(D)DL(t)] = Q^*(D)DL(t),$$

   where $Q^*(z) := zQ(z)$. Clearly, $DX(t)$ is an MCARMA(p,q+1) process. Again, this implies that we need the assumption $p > q + 1$.

The different definitions of integrated processes are not equivalent. Both have in common that DY is stationary and DY is a MCARMA process, whereas in the first definition there exist no β, such that $\beta^\mathsf{T} Y$ is stationary. Due to the different definition of integration, A_p is not fixed to be zero, thus we allow the process to be cointegrated.

We use a different approach to derive useful properties of the cointegrated model in the next section. Instead of characterizing cointegration via the crucial matrix coefficients of the autoregressive polynomial, we use instead the matrices of the continuous-time state space representation of an MCARMA process. We can transform the state space representation and thus in particular the relevant matrices in an advantageous way, which was not accomplishable with the autoregressive moving average representation.

4.3. Representations of Cointegrated Lévy Driven MCARMA Processes

We work with the cointegrated Lévy driven MCARMA processes in state space form, i.e. we interpret it as the solution $(Y(t))_{t\geq t_0}$, for some $t_0 \in \mathbb{R}$, of a continuous-time state space model.

Definition 4.3.1
An $\mathbb{R}^d$-valued ***continuous-time linear state space model*** (A, B, C, L) *of dimension N is characterized by an $\mathbb{R}^m$-valued Lévy process, a transition matrix $A \in M_N(\mathbb{R})$, an input matrix $B \in M_{N,m}(\mathbb{R})$ and an observation matrix $C \in M_{d,N}(\mathbb{R})$. It consists of the state equation*

$$\mathrm{d}X(t) = AX(t)\mathrm{d}t + B\mathrm{d}L(t) \tag{4.9a}$$

and the observation equation

$$Y(t) = CX(t) \quad \text{for } t \geq t_0 \geq 0. \tag{4.9b}$$

The state vector process $(X(t))_{t\geq t_0}$ is an $\mathbb{R}^N$-valued process and the output process $(Y(t))_{t\geq t_0}$ is $\mathbb{R}^d$-valued.

Every solution of (4.9b) has the representation

$$Y(t) = C\exp(A(t-t_0))X(t_0) + C\int_{t_0}^{t} \exp(A(t-u))B\,\mathrm{d}L(u). \tag{4.10}$$

A solution Y is called causal, if for all $t \geq t_0$, $Y(t)$ is independent of the σ-algebra generated by $\{L(s) : s > t\}$. In the following let the next assumption always hold.

Assumption C1
The Lévy process L satisfies $\mathbb{E}L(1) = 0_d$ and $\mathbb{E}\|L(1)\|^2 < \infty$.

There exist state space models with different matrices (A, B, C), which generate the same output process. In order to describe this phenomena we introduce the notion of observationally equivalence. First, we give the definitions of some properties of linear state space systems, which we are going to need subsequently. These definitions enable us to imply restrictions on the state space model in order to achieve uniqueness in the state space model and output process relation. For this we define first the observability of a continuous-time linear state space model.

Definition 4.3.2

The continuous-time linear state space system (4.9) is ***observable*** *if the observability matrix*

$$\mathcal{O}_{CA} = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{N-1} \end{pmatrix} \in M_{dN,N}(\mathbb{R}) \tag{4.11}$$

has full rank, i.e. if rank$(\mathcal{O}_{CA}) = N$ *or equivalently*

$$\text{rank} \begin{pmatrix} \lambda I_N - A \\ C \end{pmatrix} = N \tag{4.12}$$

for all eigenvalues λ *of* A*.*

Another desired property for the state space model is to have a minimal dimension, i.e. that there is no state space system of smaller dimension producing the same output process.

Definition 4.3.3

The continuous-time linear state space system (4.9) is ***controllable*** *if the controllability matrix*

$$\mathcal{C}_{AB} = \begin{pmatrix} B & AB & \dots & A^{N-1}B \end{pmatrix} \in M_{N,mN}(\mathbb{R}) \tag{4.13}$$

has full rank, i.e. if rank$(\mathcal{C}_{AB}) = N$ *or equivalently*

$$\text{rank} \begin{pmatrix} \lambda I_N - A & B \end{pmatrix} = N \tag{4.14}$$

for all eigenvalues λ *of* A*.*

Another desired property for the state space model is to have a minimal dimension, i.e. that there is no state space system of smaller dimension producing the same output process.

Definition 4.3.4

The matrix triple (A, B, C) *is called an* ***algebraic realization*** *of a rational matrix function* $k \in M_{d,m}(\mathbb{R}\{z\})$ *of dimension* N *if* $k(z) = C(zI_N - A)^{-1}B$*, where* $A \in M_N(\mathbb{R})$*,* $B \in M_{N,m}(\mathbb{R})$ *and* $C \in M_{d,N}(\mathbb{R})$*. The matrix triple* (A, B, C) *is called* ***minimal*** *if there exists no other algebraic realization* $(\widetilde{A}, \widetilde{B}, \widetilde{C})$ *with dimension* *smaller than N. The dimension of a minimal realization of k is called the* ***McMillan degree*** *of k.*

Thus, non-minimality is a source of non-uniqueness of the state space model. Minimality guarantees that we consider only components of the state vector, which are relevant for the output. Therefore, this property implies a one-to-one correspondence of the unit root properties of the state process and the output process. If we would have a non-minimal system there could be unit roots having no effect on the output process. A minimal state space model is unique up to a change of basis of the state space.

Lemma 4.3.5 (Hannan and Deistler [46], Theorem 2.3.3.)
A realization (A, B, C) is minimal if and only if it is both controllable and observable.

Last but not least, we give the formal definition of observational equivalence of state space systems.

Definition 4.3.6
A rational matrix function $k: \; z \mapsto C(zI_n - A)^{-1}B$ is called ***transfer function*** *of the state space model (4.9). A minimal linear state space model (A, B, C) is called* ***observationally equivalent*** *to the minimal system $(\widetilde{A}, \widetilde{B}, \widetilde{C})$ if they give rise to the same transfer function.*

Hence, (A, B, C) and $(\widetilde{A}, \widetilde{B}, \widetilde{C})$ are observationally equivalent if and only if there exists a nonsingular transformation matrix $T \in GL_N(\mathbb{R})$ such that $A = T\widetilde{A}T^{-1}$, $B = T\widetilde{B}$ and $C = \widetilde{C}T^{-1}$. Such a transformation leads to a corresponding basis change of the state vector to $\widetilde{X}(t) = TX(t)$.

The aim is to define a cointegrated state-space model. For this purpose, we introduce a convenient canonical form of the state space model. This will be the analogous result to the canonical form in the discrete-time setting presented by Bauer and Wagner [8], Theorem 2 and Theorem 3. The advantage of this canonical form is that the non-stationary and stationary part are decoupled and can be transformed separately. Moreover, this enables us to use existing results on stationary state space models and Lévy processes in the following.

Theorem 4.3.7
Let (A, B, C, L) be a d-dimensional minimal state space model which satisfies $\sigma(A) \subset \{(-\infty, 0) + i\mathbb{R}\} \cup \{0\}$ and the algebraic and geometric multiplicity of the eigenvalue zero is equal to c, $0 \leq c \leq d$. Then there exists a unique observationally equivalent

minimal state space representation given by

$$\begin{pmatrix} \mathrm{d}X_1(t) \\ \mathrm{d}X_2(t) \end{pmatrix} = \begin{pmatrix} 0_c & 0_{c\times(N-c)} \\ 0_{(N-c)\times c} & A_2 \end{pmatrix} \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} + \begin{pmatrix} B_1 \mathrm{d}L(t) \\ B_2 \mathrm{d}L(t) \end{pmatrix}, \qquad (4.15)$$
$$Y(t) = C_1X_1(t) + C_2X_2(t), \quad t \geq t_0,$$

where

i) The non-stationary part of the transition matrix A is the zero matrix 0_c and for the stationary part A_2 it holds that $|\lambda_{max}(\exp(A_2))| < 1$.

ii) The matrix $C_1 \in M_{d,c}(\mathbb{R})$ satisfies $C_1^\mathsf{T} C_1 = I_c$ and C_1 is a positive lower triangular matrix (c.f. Definition 6.2.1).

iii) The block matrix $B_1 \in M_{c,d}(\mathbb{R})$ is not restricted.

iv) The stationary part of the transfer function, that is (A_2, B_2, C_2), is real-valued. It is given in a canonical form for stationary state space models, e.g. in echelon canonical form (c.f. Lütkepohl and Poskitt [64]).

The cointegrated process can be expressed as the solution of this canonical form. This solution consists of a sum of the initial value, a Lévy process and a stationary MCARMA process

$$Y(t) = C_1X_1(t_0) + C_1B_1L(t) + C_2 \int_{-\infty}^{t} \exp(A_2(t-u))B_2 \,\mathrm{d}L(u), \quad t \geq t_0 \qquad (4.16)$$

if we choose $X_2(t_0)$ appropriately.

Proof. Solving (4.15) leads directly to

$$Y(t) = C_1X_1(t_0) + C_2\exp(A_2(t-t_0))X_2(t_0) + C_1B_1L(t) + C_2 \int_{t_0}^{t} \exp(A_2(t-u))B_2 \,\mathrm{d}L(u),$$

which in the end gives (4.16).

We define

$$A^* := \begin{pmatrix} 0_{c\times c} & 0_{c\times(N-c)} \\ 0_{(N-c)\times c} & A_2 \end{pmatrix}, \quad B^* := \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} \quad \text{and} \quad C^* := \begin{pmatrix} C_1 & C_2 \end{pmatrix}. \qquad (4.17)$$

For the existence of the representation (4.15) we need to show that there exists a $T \in GL_N(\mathbb{R})$ which transforms the state space model (A, B, C) to the desirable form (A^*, B^*, C^*) satisfying all restrictions (i)-(iv). In the next step we have to show that this transformation matrix is unique which results in the uniqueness of this representation.

Existence: Due to the eigenvalue assumption on the matrix A the upper part of A^* is just the Jordan normal form corresponding to the eigenvalue zero. Thus there exists a transformation matrix T' such that

$$T'AT'^{-1} = \begin{pmatrix} 0_{c\times c} & 0_{c\times(N-c)} \\ 0_{(N-c)\times c} & A'_2 \end{pmatrix} =: A',$$

where the eigenvalues of A'_2 coincide with the non-zero eigenvalues of A which have by assumption strictly negative real parts. Otherwise A'_2 is not specified yet. Further, $B' := T'B = \begin{pmatrix} B_1'^{\mathsf{T}} B_2'^{\mathsf{T}} \end{pmatrix}^{\mathsf{T}}$ and $C' = CT'^{-1} = \begin{pmatrix} C'_1 & C'_2 \end{pmatrix}$. Since the block-diagonal structure of A' is preserved by block transformations, we consider in the following only block-diagonal transformation matrices $T'' = \text{Diag}(T''_1, T''_2)$ (see Gantmacher [40], p.231) resulting in

$$A'' := T''A'T''^{-1} = \begin{pmatrix} 0_{c\times c} & 0_{c\times(N-c)} \\ 0_{(N-c)\times c} & T''_2 A'_2 T_2''^{-1} \end{pmatrix}, \quad B'' := T''B' = \begin{pmatrix} T''_1 B'_1 \\ T''_2 B'_2 \end{pmatrix} \quad \text{and}$$
$$C'' := C'T''^{-1} = \begin{pmatrix} C'_1 T_1''^{-1} & C'_2 T_1''^{-1} \end{pmatrix}.$$

Obviously there exists a transformation matrix T''_1 such that $C_1 := C'_1 T_1''^{-1}$ satisfies (ii) and $B_1 := T''_1 B'_1$ Since the eigenvalues of A'_2 have strictly negative real parts, (A'_2, B'_2, C'_2) forms a stationary linear state space model. Hence, there exists a transformation matrix T''_2 such that

$$A_2 := T''_2 A'_2 T_2''^{-1}, \quad B_2 := T''_2 B'_2 \;\text{ and }\; C_2 := C'_2 T_2''^{-1}$$

satisfy (iv). Moreover, the eigenvalues of A'_2 and hence, A_2 have strictly negative real parts so that (i) is satisfied as well. Finally, $T = T''T'$ and $(A^*, B^*, C^*) = (A'', B'', C'')$.

Uniqueness: Assume there exists matrices

$$\widetilde{A} := \begin{pmatrix} 0_{c\times c} & 0_{c\times(N-c)} \\ 0_{(N-c)\times c} & \widetilde{A}_2 \end{pmatrix}, \quad \widetilde{B} := \begin{pmatrix} \widetilde{B}_1 \\ \widetilde{B}_2 \end{pmatrix} \quad \text{and} \quad \widetilde{C} := \begin{pmatrix} \widetilde{C}_1 & \widetilde{C}_2 \end{pmatrix},$$

so that the state space model $(\widetilde{A}, \widetilde{B}, \widetilde{C})$ satisfies the assumptions of this theorem as well. But then there exists a block diagonal transformation $T = \text{Diag}(T_1, T_2)$ with $(\widetilde{A}, \widetilde{B}, \widetilde{C}) = (TA^*T^{-1}, TB^*, C^*T^{-1})$. To be more precise $\widetilde{A}_2 = T_2 A_2 T_2^{-1}$, $\widetilde{B}_1 = T_1 B_1$, $\widetilde{B}_2 = T_2 B_2$, $\widetilde{C}_1 = C_1 T_1^{-1}$ and $\widetilde{C}_2 = C_2 T_2^{-1}$. Since (A_2, B_2, C_2) and $(\widetilde{A}_2, \widetilde{B}_2, \widetilde{C}_2)$, respectively are given in canonical form (cf. restriction (iv)), T_2 has to be the identity matrix. In order to prove uniqueness it remains to show that T_1 is the identity matrix as well. Due to $\widetilde{C}_1^\intercal \widetilde{C}_1 = I_c$ by (ii), we obtain

$$I_c = \widetilde{C}_1^\intercal \widetilde{C}_1 = (C_1 T_1)^\intercal C_1 T_1 = T_1^\intercal C_1^\intercal C_1 T_1 = T_1^\intercal T_1$$

and thus, T_1 is orthogonal. If we now exploit the fact that C_1 and $\widetilde{C}_1$ must both be lower triangular matrices, we further get that T_1 must be lower triangular itself since

$$C_1 T_1 = \begin{pmatrix} c_{11} & 0 & \cdots & \cdots & 0 \\ c_{21} & c_{22} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ c_{c1} & c_{c2} & \cdots & \cdots & c_{cc} \\ \vdots & \vdots & & & \vdots \\ c_{d1} & c_{d2} & \cdots & \cdots & c_{dc} \end{pmatrix} \begin{pmatrix} t_{11} & \cdots & t_{1c} \\ \vdots & \ddots & \vdots \\ t_{c1} & \cdots & t_{cc} \end{pmatrix}$$

$$= \begin{pmatrix} c_{11}t_{11} & c_{11}t_{12} & \cdots & c_{11}t_{1c} \\ * & \sum_{i=1}^2 c_{2i}t_{i2} & & \sum_{i=1}^2 c_{2i}t_{ic} \\ \vdots & & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ & & & \sum_{i=1}^c c_{ci}t_{ic} \\ * & \cdots & * & * \end{pmatrix} = \begin{pmatrix} \widetilde{c}_{11} & 0 & \cdots & \cdots & 0 \\ \widetilde{c}_{21} & \widetilde{c}_{22} & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & 0 \\ \widetilde{c}_{c1} & & \cdots & & \widetilde{c}_{cc} \\ \vdots & & & & \vdots \\ \widetilde{c}_{d1} & & \cdots & & \widetilde{c}_{dc} \end{pmatrix} = \widetilde{C}_1$$

holds and therefore the entries t_{ij}, for $i < j$, must be zero or otherwise the lower triangular structure would not be preserved. Real-valued orthogonal matrices have only eigenvalues equal to 1 or -1 and thus the diagonal of the transformation matrix T_1 consists only of the entries 1 or -1.

We utilize now the restriction that the first non-zero element $c_{j_i i}$ in each column of C_1 is positive. We start with the last column, where the first positive entry $c_{j_c c}$ is multiplied by t_{cc}. This product must be positive and hence t_{cc} must be positive. Since the diagonal entries of a triangular matrix are the eigenvalues itself, this implies that $t_{cc} = 1$. Thus, all columns of T_1 are orthonormal. The $(c-1)$-th column has only two entries which are non-zero and in order to be orthonormal to the c-th unit vector e_c it must be a unit vector itself, i.e. it must be e_{c-1}. Iterating this procedure leads to $T_1 = I_c$ and consequently we have due to the entire restrictions a unique form for the state space model. □

We abbreviate the stationary part with

$$Y_2(t) := C_2 \int_{-\infty}^{t} \exp(A_2(t-u))B_2 \, \mathrm{d}L(u), \quad t \geq t_0. \tag{4.18}$$

We show in the next lemma that we can either assume minimality of the original state space model (A, B, C, L) or of the stationary part of the state space system (A_2, B_2, C_2, L) combined with assumptions on the matrices B_1 and C_1. The assumptions on the decoupled system are often easier to verify.

Lemma 4.3.8
Each d-dimensional state space system (A, B, C, L), which satisfies $\sigma(A) \subset \{(-\infty, 0) + i\mathbb{R}\} \cup \{0\}$ with algebraic and geometric multiplicity of the eigenvalue zero is equal to $c \leq d$, is minimal if and only if B_1 has full row rank and C_1 has full column rank, i.e. rank $B_1 =$ rank $C_1 = c$, and the representation (A_2, B_2, C_2) of the stationary subsystem is minimal.

Proof. Minimality is equivalent to the conditions that the controllability and observability matrices $\mathcal{C}_{AB}$ (cf. (4.13)) and $\mathcal{O}_{AC}$ (cf. (4.11)), respectively have full rank. We prove an alternative criterion for observability (c.f. Bernstein [9], Proposition 12.3.13) . Therefore, we have to determine the rank for all eigenvalues λ of A of

$$\begin{pmatrix} \lambda I_c & 0_{c\times(N-c)} \\ 0_{(N-c)\times c} & \lambda I_{N-c} - A_2 \\ C_1 & C_2 \end{pmatrix}.$$

We consider two cases, beginning with the eigenvalue $\lambda = 0$, which simplifies the matrix to

$$\begin{aligned} \operatorname{rank} \begin{pmatrix} \lambda I_c & 0_{c\times(N-c)} \\ 0_{(N-c)\times c} & \lambda I_{N-c} - A_2 \\ C_1 & C_2 \end{pmatrix} &= \operatorname{rank} \begin{pmatrix} -A_2 & 0_{(N-c)\times c} \\ C_2 & C_1 \end{pmatrix} \\ = (N-c) + \operatorname{rank}(C_1 + C_2 A_2^{-1} 0_{(N-c)\times c}) &= (N-c) + c = N. \end{aligned}$$

The last equations follow by Bernstein [9], Proposition 2.8.3, the fact that C_1 has full rank c and the invertibility of A_2. In the case $\lambda \neq 0$ we get

$$\begin{aligned}
&\operatorname{rank}\begin{pmatrix} \lambda I_c & 0_{c\times(N-c)} \\ 0_{(N-c)\times c} & \lambda I_{N-c} - A_2 \\ C_1 & C_2 \end{pmatrix} \\
&= c + \operatorname{rank}\left(\begin{pmatrix} \lambda I_{N-c} - A_2 \\ C_2 \end{pmatrix} - \begin{pmatrix} 0_{(N-c)\times c} \\ C_1 \end{pmatrix} \cdot \lambda^{-1} I_c \cdot 0_{c\times(N-c)}\right) \\
&= c + \operatorname{rank}\begin{pmatrix} \lambda I_c - A_2 \\ C_2 \end{pmatrix} = N.
\end{aligned}$$

Again, we used Bernstein [9], Proposition 2.8.3, in combination with the fact that the stationary part is minimal by assumption and hence observable. As a consequence, we have shown that $\operatorname{rank}(\mathcal{O}_{CA}) = N$ and thus observability of the system. Analogously, we obtain $\operatorname{rank}(\mathcal{C}_{AB}) = N$ and together with the observability this gives the minimality of the system. □

Although there is written in Theorem 4.3.7 (iii) that we do not assume any restriction on B_1, the assumption on the minimality of the state space model implies that B_1 has full rank c.

Lemma 4.3.9
Let Y be given as in Theorem 4.3.7 with $0 < c < d$. Then Y is cointegrated with cointegration space spanned by $C_1^{\perp}$ and cointegration rank rank $C_1^{\perp} = d - c$, i.e. $C_1^{\perp} Y$ is stationary.

Proof. A conclusion of Lemma 4.3.8 is that $C_1 B_1 \neq 0_{d\times d}$ so that Y as given in (4.16) is indeed an integrated process since the Lévy process $(C_1 B_1 L(t))_{t\geq 0}$ is a non-stationary process with strictly stationary increments. Moreover, $C_1^{\perp} Y = C_1^{\perp} Y_2$ is a stationary process and hence, $C_1^{\perp}$ spans the cointegration space with $\operatorname{rank} C_1^{\perp} = d - c$. □

Note that the parameter c represents the number of common stochastic trends. From these considerations the next definition is well-defined.

Definition 4.3.10
*Let (A, B, C, L) be a d-dimensional minimal state space model of dimension N which satisfies $\sigma(A) \subset \{(-\infty, 0) + i\mathbb{R}\} \cup \{0\}$ and the algebraic and geometric multiplicity of the eigenvalue zero is equal to c with $0 < c < d$. Then the output process Y is called **cointegrated continuous-time linear state space model**.*

Note that we could define the cointegrated linear state space model in an alternative way: we define a sum of a stationary state space process and a Lévy process. This means we have $Y(t) = C_1 L_1(t) + Y_2(t)$, where L_1 is a LÃ©vy process and Y_2 is a stationary state space model. From Theorem 4.3.7 we know that both definitions are equivalent.

Corollary 4.3.11
The cointegrated state space model Y is causal.

Proof. This is obvious due to representation (4.16). □

The canonical form helps us later in the parameter estimation of the cointegrated MCARMA process. Given this representation, we can now deal with the integrated part and the stationary part completely separately. Because we do not observe the Lévy process, we have to do a stepwise procedure, where we first estimate the non-stationary parameters and after that the stationary parameters.

We know that the covariance of the cointegrated state space model can also be decomposed.

Proposition 4.3.12
Assume that Y is a cointegrated state space process as in (4.15). Then

$$\mathbb{E}[Y(t)] = C_1 \mathbb{E}[X_1(t_0)] \quad \text{for } t \geq t_0.$$

Suppose $X_1(t_0) = 0$. Then for $t \geq t_0$ and $s \geq 0$ we have

$$\begin{aligned} Cov(Y(t), Y(t+s)) &= C_2 \exp(A_2\, s)\Gamma_0 C_2^\mathsf{T} + \int_0^t C_2 \exp(A_2 u) B_2 \Sigma_L (C_1 B_1)^\mathsf{T} \, \mathrm{d}u \\ &\quad + \int_s^{t+s} C_1 B_1 \Sigma_L B_2^\mathsf{T} \exp(A_2^\mathsf{T} u) C_2^\mathsf{T} \, \mathrm{d}u + t \cdot C_1 B_1 \Sigma_L (C_1 B_1)^\mathsf{T}, \end{aligned}$$

where Γ_0 is the covariance matrix of the stationary process X_2 given by

$$\Gamma_0 := \int_0^\infty \exp(A_2 u) B_2 \Sigma_L \left(\exp(A_2 u) B_2\right)^\mathsf{T} \, \mathrm{d}u.$$

Proof. We obtain for the expectation evidently

$$\begin{aligned} \mathbb{E}[Y(t)] &= \mathbb{E}\left[C_1 X_1(t_0) + C_1 B_1 L(t) + C_2 \int_{-\infty}^t \exp(A_2(t-u)) B_2 \, \mathrm{d}L(u)\right] \\ &= C_1 \mathbb{E}[X_1(t_0)]. \end{aligned}$$

Setting $X_1(t_0) = 0$ leads then to

$$\begin{aligned}
&\mathrm{Cov}(Y(t), Y(t+s)) = \mathbb{E}\left[Y(t)Y(t+s)^{\mathsf{T}}\right] \\
&= \mathbb{E}\Big[C_1B_1L(t)\left(C_1B_1L(t+s)\right)^{\mathsf{T}} + C_1B_1L(t)Y_2(t+s)^{\mathsf{T}} \\
&\qquad + Y_2(t)\left(C_1B_1L(t+s)\right)^{\mathsf{T}} + Y_2(t)Y_2(t+s)^{\mathsf{T}}\Big] \\
&= C_1B_1\mathbb{E}\left[L(t)L(t+s)^{\mathsf{T}}\right](C_1B_1)^{\mathsf{T}} + \mathbb{E}[Y_2(t)Y_2(t+s)^{\mathsf{T}}] \\
&\qquad + C_1B_1\mathbb{E}\left[\int_0^t 1\,\mathrm{d}L(u)\left(\int_{-\infty}^{t+s}\exp(A_2(t+s-u))B_2\,\mathrm{d}L(u)\right)^{\mathsf{T}}\right]C_2^{\mathsf{T}} \\
&\qquad + C_2\mathbb{E}\left[\int_{-\infty}^{t}\exp(A_2(t-u))B_2\,\mathrm{d}L(u)\left(\int_0^{t+s}1\,\mathrm{d}L(u)\right)^{\mathsf{T}}\right](C_1B_1)^{\mathsf{T}} \\
&= C_1B_1\mathbb{E}\left[L(t)L(t+s)^{\mathsf{T}}\right](C_1B_1)^{\mathsf{T}} + \mathbb{E}[Y_2(t)Y_2(t+s)^{\mathsf{T}}] \\
&\qquad + C_1B_1\mathbb{E}\left[\int_0^t 1\,\mathrm{d}L(v)\left(\int_0^{t}\exp(A_2(t+s-u))B_2\,\mathrm{d}L(u)\right)^{\mathsf{T}}\right]C_2^{\mathsf{T}} \\
&\qquad + C_2\mathbb{E}\left[\int_0^{t}\exp(A_2(t-u))B_2\,\mathrm{d}L(u)\left(\int_0^{t}1\,\mathrm{d}L(v)\right)^{\mathsf{T}}\right](C_1B_1)^{\mathsf{T}}
\end{aligned}$$

and finally the claimed result follows by calculating all the remaining expectations using Equation (3.8) as well as Proposition 3.13 in Marquardt and Stelzer [69]. □

The time dependence of the covariance matrix is obvious in this representation and hence this process is indeed non-stationary. We assume for reasons of simplicity from now on that $t_0 = 0$. Note that the process Y_2 is a causal stationary MCARMA process in the sense of Marquardt and Stelzer [69], Definition 3.20. Hereby, the causality also applies for the cointegrated process.

Lemma 4.3.13
The cointegrated MCARMA process Y given as in Definition 4.3.10 is causal.

Proof. This is obvious due to representation (4.16) (c.f. Definition 3.20 in Marquardt and Stelzer [69]). □

Recall the results about the sampled process by Schlemm and Stelzer [90], Lemma 5.2, which leads to the following discrete-time representation for the sampled process. We derive the same decoupling for the sampled process as in the continuous-time case. Hence, we have also the separation of the stationary part and integrated part in the state space system. Moreover, we can clearly see the connection of the eigenvalue zero of the transition matrix A with unit roots in the discrete time case since e^A has eigenvalues equal to one if A has eigenvalues equal to zero.

Lemma 4.3.14

Assume that Y is an MCARMA process as in Definition 4.3.10. The sampled process $Y^{(h)}$ has the state space representation

$$\begin{pmatrix} X_{n,1}^{(h)} \\ X_{n,2}^{(h)} \end{pmatrix} = \begin{pmatrix} X_{n-1,1}^{(h)} \\ \mathrm{e}^{A_2 h} X_{n-1,2}^{(h)} \end{pmatrix} + \begin{pmatrix} R_{n,1}^{(h)} \\ R_{n,2}^{(h)} \end{pmatrix}$$

and observation equation

$$Y_n^{(h)} = C_1 X_{n,1}^{(h)} + C_2 X_{n,2}^{(h)}, \tag{4.19}$$

with noise term

$$R_n^{(h)} = \begin{pmatrix} R_{n,1}^{(h)} \\ R_{n,2}^{(h)} \end{pmatrix} = \begin{pmatrix} B_1 \left(L(nh) - L((n-1)h)\right) \\ \int_{(n-1)h}^{nh} \mathrm{e}^{A_2(nh-u)} B_2 \,\mathrm{d}L(u) \end{pmatrix}, \quad n \in \mathbb{N}. \tag{4.20}$$

The sequence $(R_n^{(h)})_{n\in\mathbb{N}}$ is an i.i.d. sequence with mean zero and covariance matrix

$$\widetilde{\Sigma}^{(h)} = \mathbb{E} R_n^{(h)} R_n^{(h)\mathsf{T}} = \int_0^h \begin{pmatrix} B_1 \Sigma_L B_1^\mathsf{T} & \mathrm{e}^{A_2 u} B_2 \Sigma_L B_1^\mathsf{T} \\ B_1 \Sigma_L B_2^\mathsf{T} \mathrm{e}^{A_2^\mathsf{T} u} & \mathrm{e}^{A_2 u} B_2 \Sigma_L B_2^\mathsf{T} \mathrm{e}^{A_2^\mathsf{T} u} \end{pmatrix} \mathrm{d}u. \tag{4.21}$$

We write for the different parts of the covariance matrix

$$\widetilde{\Sigma}^{(h)} = \begin{pmatrix} \widetilde{\Sigma}_{11}^{(h)} & \widetilde{\Sigma}_{12}^{(h)} \\ \widetilde{\Sigma}_{21}^{(h)} & \widetilde{\Sigma}_{22}^{(h)} \end{pmatrix}. \tag{4.22}$$

Furthermore, we have that $(R_n^{(h)})_{n\in\mathbb{Z}}$ has finite r^{th}-moments for some $r > 0$ if the driving Lévy process L has finite r^{th}-moments. This implies the existence of r^{th}-moments for the sampled process $Y_n^{(h)}$.

Proof. The state space representation follows at once by setting $t = nh$ in (4.15) and the same holds for the covariance matrix.

The existence of the r^{th}-moment follows immediately from Proposition 3.30 in Marquardt and Stelzer [69] and Lemma 3.15 in Schlemm and Stelzer [91]. □

Note that due to Theorem 1 in Bauer and Wagner [8] the class of cointegrated $I(1)$-processes with c common trends and the class of state-space models, satisfying the assumptions made in this section, are equivalent.

Lemma 4.3.15

The solution of the sampled process given in Lemma 4.3.14 is given by

$$Y_n^{(h)} = C_1 X_1^{(h)}(t_0) + C_1 B_1 L(nh) + Y_{n,2}^{(h)}, \quad n \in \mathbb{N}, \tag{4.23}$$

where the stationary part is given by

$$Y_{n,2}^{(h)} = C_2 \int_{-\infty}^{nh} \mathrm{e}^{A_2(nh-u)} B_2 \,\mathrm{d}L(u). \tag{4.24}$$

Proof. In the same manner this follows by inserting $t = nh$ into (4.16) and (4.18). □

The first two moments of the sampled process are derived in the next lemma.

Lemma 4.3.16

The expectation of the sampled cointegrated state space model (4.23) is given by

$$\mathbb{E}[Y_n^{(h)}] = C_1 \mathbb{E}[X_1^{(h)}(t_0)] \tag{4.25}$$

and for the covariance we get, if we assume $\mathbb{E}[X_1^{(h)}(t_0)] = 0$, that

$$\begin{aligned} Cov(Y_n^{(h)}, Y_{n+s}^{(h)}) = & C_2 \exp(A_2\, s) \widetilde{\Sigma}_{2,2}^{(h)} C_2^\mathsf{T} && (4.26) \\ & + \int_0^{nh} C_2 \exp(A_2 u) B_2 \Sigma_L (C_1 B_1)^\mathsf{T} \,\mathrm{d}u \\ & + \int_0^{(n+s)h} C_1 B_1 \Sigma_L B_2^\mathsf{T} \exp(A_2^\mathsf{T} u) C_2^\mathsf{T} \,\mathrm{d}u \\ & + nh \cdot C_1 \widetilde{\Sigma}_{1,1}^{(h)} C_1^\mathsf{T}. && (4.27) \end{aligned}$$

Proof. Setting $t = nh$ in Proposition 4.3.12 proves the claim. □

We need a certain degree of independence of the stationary part of the sampled process for the estimation procedure. Hence, we give the definition of strong mixing which gives us a sufficient degree of independence. For more details on mixing processes see e.g. Bradley [16] or Doukhan [31].

Definition 4.3.17

A continuous-time stationary stochastic process $X = (X_t)_{t\in\mathbb{R}}$ is called ***strongly mixing*** *if*

$$\alpha_l := \sup \left\{ |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{F}_{-\infty}^m, B \in \mathcal{F}_{m+l}^\infty \right\} \xrightarrow{l\to\infty} 0,$$

where $\mathcal{F}_{-\infty}^{m} := \sigma(X_t : t \leq n)$ *and* $\mathcal{F}_{m+l}^{\infty} := \sigma(X_t : t \geq m+l)$, $m \in \mathbb{N}$.

With regard to already known results, we see quite easily that the stationary part of the sampled process satisfies the following mixing property.

Lemma 4.3.18

The stationary part Y_2 *of the continuous cointegrated MCARMA process* Y *with finite second moments as given in Theorem 4.3.7 is exponentially strongly mixing and the same holds true for the stationary part* $Y_2^{(h)}$ *of the sampled process* $Y^{(h)}$. *There exists a constant* $\delta > 0$ *such that the mixing coefficients* $\alpha_{Y_{st}^{(h)}}$ *of the sampled process satisfy*

$$\sum_{l=0}^{\infty} \left[\alpha_{Y_2^{(h)}}(l) \right]^{\frac{\delta}{2+\delta}} < \infty.$$

The same holds true for $C_1^{\perp} Y = C_1^{\perp} Y_2$ *and* $C_1^{\perp} Y^{(h)} = C_1^{\perp} Y_2^{(h)}$.

Proof. Due to Proposition 3.34 in Marquardt and Stelzer[69], the assertion for the stationary process holds. This property transfers to the sampled process right away and we also have the condition on the mixing coefficients. The last claim follows directly by Bradley [16], Remark 1.8 b). □

In the further considerations, we need the first difference of the sampled cointegrated process and the stationary part of this process. We have already calculated the first two moments for the stationary part in a previous lemma and the mixing property of the stationary part of the sampled process. Hence, we turn our attention to the first difference and calculate a representation in the next lemma.

Lemma 4.3.19

The first difference of the sampled cointegrated process $Y^{(h)}$ *(c.f. Lemma 4.3.14) is given by*

$$\begin{aligned} \Delta Y_n^{(h)} &= C_1 R_{n,1}^{(h)} + \Delta Y_{n,2}^{(h)} \\ &= C_1 R_{n,1}^{(h)} + C_2 \int_{(n-1)h}^{nh} \mathrm{e}^{A_2(nh-u)} B_2 \, \mathrm{d}L(u) \\ &\quad + C_2 \left(\mathrm{e}^{A_2 h} - I_d \right) \int_{-\infty}^{(n-1)h} \mathrm{e}^{A_2((n-1)h-u)} B_2 \, \mathrm{d}L(u), \end{aligned} \tag{4.28}$$

where $R_{n,1}^{(h)}$ *is given as in (4.20).*

Proof. To show this we use equation (4.23) and (4.24). We obtain

$$\begin{aligned}
\Delta Y_n^{(h)} =& C_1 X_1^{(h)}(t_0) + C_1 B_1 L(nh) + Y_{n,2}^{(h)} - C_1 X_1^{(h)}(t_0) - C_1 B_1 L((n-1)h) + Y_{n-1,2}^{(h)} \\
=& C_1 R_{n,1}^{(h)} + C_2 \int_{-\infty}^{nh} \mathrm{e}^{A_2(nh-u)} B_2 \,\mathrm{d}L(u) - C_2 \int_{-\infty}^{(n-1)h} \mathrm{e}^{A_2((n-1)h-u)} B_2 \,\mathrm{d}L(u) \\
=& C_1 R_{n,1}^{(h)} + C_2 \int_{(n-1)h}^{nh} \mathrm{e}^{A_2(nh-u)} B_2 \,\mathrm{d}L(u) \\
&+ C_2 \left(\mathrm{e}^{A_2 h} - I_d\right) \int_{-\infty}^{(n-1)h} \mathrm{e}^{A_2((n-1)h-u)} B_2 \,\mathrm{d}L(u),
\end{aligned}$$

where the second and third summand together are nothing else than $\Delta Y_{n,2}^{(h)}$. □

Note that the first difference of the sampled cointegrated process $\Delta Y^{(h)}$ is obviously stationary and the r^{th}-moment exists if the r^{th}-moment of the Lévy process exists since due to Lemma 4.3.14 $Y^{(h)}$ has then a finite r^{th}-moment. Furthermore, note that $\Delta Y_2^{(h)}$ is also strongly mixing since its the difference and consequently a measurable function of the finite past values of a strongly mixing process (c.f. Remark 1.8 b) in Bradley [16]). Moreover, we know that $(C_1 R_{n,1}^{(h)})_{n\in\mathbb{N}}$ is obviously an i.i.d. sequence.

The last lemma showed that we can interpret the first difference of the sampled process $Y^{(h)}$ also as a sum of an i.i.d. sequence and the difference of the sampled process of the stationary part $Y_2^{(h)}$ of the cointegrated MCARMA process Y. This representation is obvious due to the decoupled state space system and thus the form of the solution in Lemma 4.3.15.

Last but not least, we present typical realizations of two bivariate cointegrated MCARMA processes in Figure 4.1. On the left side we have an MCARMA process driven by a Brownian motion and on the right side driven by a normal-inverse Gaussian process. Moreover, the figure shows the sampled versions the stationary linear combination and the common stochastic trend of each process.

4.4. Error Correction Form of Cointegrated State Space Model Sampled at a Discrete Time-grid

Assume that we have a cointegrated multivariate CARMA process driven by a Lévy process as given in Definition 4.3.10. The d-dimensional cointegrated process has the

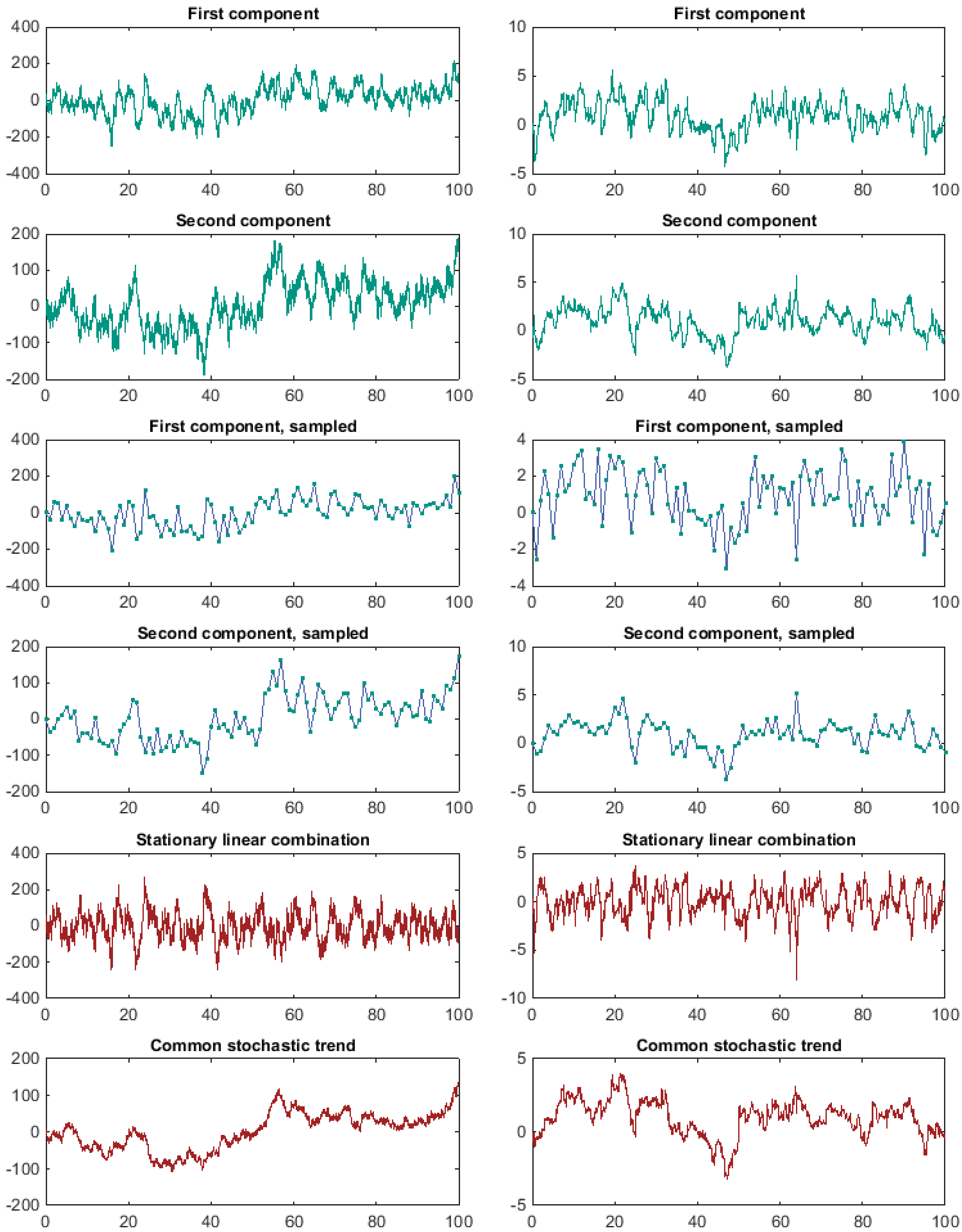

Figure 4.1.: Two Examples of Cointegrated Processes
Left: Brownian Motion - Right: Normal-inverse Gaussian

representation (c.f. (4.16))

$$Y(t) = C_1 X_1(t_0) + C_1 B_1 L(t) + C_2 \int_{-\infty}^{t} \exp(A_2(t-u)) B_2 \, \mathrm{d}L(u) \quad \text{for } t \geq t_0$$

and satisfies the assumptions of Theorem 4.3.7. Accordingly, it consist of a starting value, an integrated part and a stationary part, which coincides with a stationary MCARMA process. Furthermore, assume that $X_1(t_0)$ is independent of $(L(t))_{t\geq 0}$ and the covariance matrix of the Lévy process Σ_L is positive definite.

Since we want to estimate the model parameters from observations of the MCARMA process Y at discrete time points, we consider the sampled version of the cointegrated MCARMA process.

We briefly recall the state space representation with state equation

$$\begin{pmatrix} X_{n,1}^{(h)} \\ X_{n,2}^{(h)} \end{pmatrix} = \begin{pmatrix} X_{n-1,1}^{(h)} \\ \mathrm{e}^{A_2 h} X_{n-1,2}^{(h)} \end{pmatrix} + \begin{pmatrix} R_{n,1}^{(h)} \\ R_{n,2}^{(h)} \end{pmatrix},$$

noise

$$\begin{pmatrix} R_{n,1}^{(h)} \\ R_{n,2}^{(h)} \end{pmatrix} = \begin{pmatrix} B_1 \left(L(nh) - L((n-1)h) \right) \\ \int_{(n-1)h}^{nh} \mathrm{e}^{A_2(nh-u)} B_2 \, \mathrm{d}L(u) \end{pmatrix}$$

and the observation equation given by

$$Y_n^{(h)} = C_1 X_{n,1}^{(h)} + C_2 X_{n,2}^{(h)}. \tag{4.29}$$

Definition 4.4.1
Assume that the process L has finite second moments. The ***linear innovations*** $\varepsilon^{(h)} = (\varepsilon_n^{(h)})_{n\in\mathbb{Z}}$ *of* $Y^{(h)}$ *are defined by*

$$\varepsilon_n^{(h)} = Y_n^{(h)} - P_{n-1} Y_n^{(h)}, \tag{4.30}$$

where P_n is the orthogonal projection onto $\overline{span}\{Y_i^{(h)} : -\infty < i \leq n\}$ and the closure is taken in the Hilbert space of square-integrable random variables with inner product $(X,Y) \mapsto \mathbb{E}\langle X, Y\rangle$.

We want to obtain an error correction form and calculate the likelihood function corresponding to this state space model later on. However, the linear state space model is not in innovation form and thus we apply a linear filter to this model. We use the Kalman filter, which is a useful tool in estimation theory to calculate the linear innovations. Before we can apply the Kalman filter to the cointegrated model we have to check the applicability of the Kalman filter to the cointegrated model. The Kalman filter can be applied to the cointegrated model if Assumption K1-Assumption K3 in Section 4.6 are satisfied.

The discrete-time state space model (4.29) satisfies Assumption K1, Assumption K2 and Assumption K3. To see this note that $W_n = 0_d$ holds, which immediately implies $R \equiv 0_{d\times d}$. Assumption K1 is automatically satisfied due to the definition of the cointegrated MCARMA model. Furthermore, we have $\Gamma Z_n = R_n^{(h)}$ in our model, which has a positive definite covariance matrix. Hence, Assumption K2 is satisfied. In our model we have $C = H$ with $\operatorname{rank} C = d$. This implies $C\widetilde{\Sigma}^{(h)}C^T$ has full rank and thus Assumption K3 holds.

Therefore, by Appendix 4.6 we can apply the Kalman filter to the cointegrated model. Hence, we obtain the linear innovations of $Y^{(h)}$ and a new state space model in innovation form. In order to estimate the cointegrated process, we use the linear innovations. They describe the new information at time point n, which was not available one step before, i.e. at time point $n-1$.

Let us sum up the important results concerning the Kalman filter in the next proposition.

Proposition 4.4.2
The discrete-time algebraic Riccati equation

$$\Omega^{(h)} = \mathrm{e}^{Ah}\Omega^{(h)}\mathrm{e}^{A^{\mathsf{T}}h} - \mathrm{e}^{Ah}\Omega^{(h)}C^{\mathsf{T}}(C\Omega^{(h)}C^{\mathsf{T}})^{-1}C\Omega^{(h)}\mathrm{e}^{A^{\mathsf{T}}h} + \widetilde{\Sigma}^{(h)} \tag{4.31}$$

has a positive definite solution $\Omega^{(h)} \in \mathbb{S}_N^{++}(\mathbb{R})$ and the steady state Kalman gain matrix $K^{(h)}$ is given by

$$K^{(h)} := \mathrm{e}^{Ah}\Omega^{(h)}C^{\mathsf{T}}(C\Omega^{(h)}C^{\mathsf{T}})^{-1} \in M_{N,d}(\mathbb{R}). \tag{4.32}$$

The linear innovations $\varepsilon^{(h)}$ of $Y^{(h)}$ are the unique stationary solution of the linear state space model

$$\begin{aligned} \widehat{X}_n^{(h)} &= \left(\mathrm{e}^{Ah} - K^{(h)}C\right)\widehat{X}_{n-1}^{(h)} + K^{(h)}Y_{n-1}^{(h)}, \\ \varepsilon_n^{(h)} &= Y_n^{(h)} - C\widehat{X}_n^{(h)}, \qquad n \in \mathbb{N}. \end{aligned} \tag{4.33}$$

Then, we have a moving average representation for the linear innovations given by

$$\begin{aligned}\varepsilon_n^{(h)} &= \Big(I_d - C\big[I_N - \big(\mathrm{e}^{Ah} - K^{(h)}C\big)B\big]^{-1}K^{(h)}B\Big)\,Y_n^{(h)} \\ &= Y_n^{(h)} - C\sum_{i=1}^{\infty}\big(\mathrm{e}^{Ah} - K^{(h)}C\big)^{i-1}K^{(h)}Y_{n-i}^{(h)},\end{aligned} \tag{4.34}$$

where B denotes the backshift operator, which is defined by $BY_n^{(h)} = Y_{n-1}^{(h)}$. The covariance matrix of the innovations is given by

$$V^{(h)} = \mathbb{E}\varepsilon_n^{(h)}\varepsilon_n^{(h)\mathsf{T}} = C\Omega^{(h)}C^{\mathsf{T}} \in \mathbb{S}_d^{++}. \tag{4.35}$$

Thus, the process $Y^{(h)}$ is in innovations form given by

$$\begin{aligned}\widehat{X}_n^{(h)} &= \mathrm{e}^{Ah}\widehat{X}_{n-1}^{(h)} + K^{(h)}\varepsilon_{n-1}^{(h)}, \\ Y_n^{(h)} &= C\widehat{X}_n^{(h)} + \varepsilon_n^{(h)}, \qquad n \in \mathbb{N}.\end{aligned} \tag{4.36}$$

Define now the rational matrix-valued transfer function

$$\begin{aligned}k(z) :&= I_d - C\big[I_N - \big(\mathrm{e}^{Ah} - K^{(h)}C\big)z\big]^{-1}K^{(h)}z \\ &= I_d - C\sum_{i=1}^{\infty}\big(\mathrm{e}^{Ah} - K^{(h)}C\big)^{i-1}K^{(h)}z^i, \quad \text{for } z \in \mathbb{C}.\end{aligned} \tag{4.37}$$

Obviously, we have $k(0) = I_d$ and

$$k(1) = I_d - C\big[I_N - \big(\mathrm{e}^{Ah} - K^{(h)}C\big)\big]^{-1}K^{(h)} \in M_d(\mathbb{R}). \tag{4.38}$$

Due to Lemma 4.6.7 from Appendix 4.6 we know that $|\lambda_{max}(\mathrm{e}^{Ah} - K^{(h)}C)| < 1$, hence $I_N - \big(\mathrm{e}^{Ah} - K^{(h)}C\big)$ is invertible and $k(1)$ is well defined. Since the matrix e^{Ah} has eigenvalues equal to one, we have an integrated process.

Lemma 4.4.3
The eigenvalues of e^{Ah} are strictly within the unit circle or at one and the algebraic multiplicity of the eigenvalue one of e^{Ah} is equal to its geometric multiplicity c.

Proof. This assertion is automatically satisfied by the sampled process from (4.29), due to the eigenvalue condition on the cointegrated continuous-time model (see Theorem 4.3.7). Recall that the matrix exponential has the following property $\mathrm{e}^{Ah} = \exp(\mathrm{Diag}(0_c, A_2h)) = \mathrm{Diag}(I_c, \exp(A_2h))$. □

Hence, by Lemma 4.4.3 we have indeed a cointegrated state space model (4.36) of order one. We want to show that $k(1)$ maintains the requirements for cointegration. Hence, we show that $k(1)$ has rank $r = d - c$ and contains the information about the cointegration space. In order to analyze the matrix $k(1)$ we need first the following assumption guaranteeing that the filtered system is still minimal.

Assumption C2
The linear system given in (4.36) is controllable, i.e.

$$\mathcal{C}_{e^{Ah}K^{(h)}} := \begin{pmatrix} K^{(h)} & e^{Ah}K^{(h)} & \dots & \left(e^{Ah}\right)^{N-1} K^{(h)} \end{pmatrix} \in M_{N,pN}(\mathbb{R}) \tag{4.39}$$

has rank N.

This assumption can be formulated in a different way using an alternative criterion for controllability (c.f. [9, Proposition 12.6.13]). It holds by Fact 2.11.1 in Bernstein [9] that for all $\lambda \in \sigma(e^{Ah})$

$$\operatorname{rank} \mathcal{C}_{e^{Ah}K^{(h)}} = N \quad \Leftrightarrow \quad \ker\left(e^{Ah} - \lambda I_N\right)^{\mathsf{T}} \cap \ker K^{(h)\mathsf{T}} = \{0\}. \tag{$*$}$$

Obviously, $\ker\left(e^{Ah} - \lambda I_N\right)^{\mathsf{T}}$ is the eigenspace corresponding to the eigenvalue λ of $e^{A^{\mathsf{T}}h}$. Taking a closer look on $\ker K^{(h)\mathsf{T}}$ we see that the following equalities must hold

$$\begin{aligned} \ker K^{(h)\mathsf{T}} &= \ker\left(e^{Ah}\Omega^{(h)}C^{\mathsf{T}}(C\Omega^{(h)}C^{\mathsf{T}})^{-1}\right)^{\mathsf{T}} = \ker\left(\Omega^{(h)}C^{\mathsf{T}}\right)^{\mathsf{T}} \\ &= \ker\left(C\Omega^{(h)}\right) = \ker\left(C\right)\Omega^{(h)}. \end{aligned}$$

Therefore, the condition ($*$) is satisfied if $z' := \Omega^{(h)}z$ lies not in $\ker(C)$ for all eigenvectors z of $e^{A^{\mathsf{T}}h}$ corresponding to the eigenvalue λ. Besides, due to the observability condition (see Bernstein [9], Fact 2.11.3) we know that for all eigenvectors $\widetilde{z}$ corresponding to the eigenvalue λ of e^{Ah} that $\widetilde{z} \notin \ker C$.

Moreover, the linear system (4.36) is also observable due to the fact that the discrete-time and continuous-time observability matrix do coincide. Finally, minimality follows due to Assumption C2 and the observability of the system combined with Lemma 4.3.5.

The filtered minimal state space system (4.36) is obviously given in decoupled form

$$\begin{aligned} \begin{pmatrix} \widehat{X}^{(h)}_{n+1,1} \\ \widehat{X}^{(h)}_{n+1,2} \end{pmatrix} &= \begin{pmatrix} I_c & 0_{c\times(N-c)} \\ 0_{(N-c)\times c} & e^{A_2 h} \end{pmatrix} \begin{pmatrix} \widehat{X}^{(h)}_{n,1} \\ \widehat{X}^{(h)}_{n,2} \end{pmatrix} + \begin{pmatrix} K_1^{(h)} \\ K_2^{(h)} \end{pmatrix} \varepsilon_n^{(h)}, \\ Y_n^{(h)} &= \begin{pmatrix} C_1 & C_2 \end{pmatrix} \begin{pmatrix} \widehat{X}^{(h)}_{n,1} \\ \widehat{X}^{(h)}_{n,2} \end{pmatrix} + \varepsilon_n^{(h)}. \end{aligned} \tag{4.40}$$

Further, we can also make an assumption on the subsystems instead of the original system. The system $(\mathrm{e}^{Ah}, K^{(h)}, C)$ is minimal if and only if $K_1^{(h)}$ has full row rank and C_1 has full column rank, i.e. $\operatorname{rank} K_1^{(h)} = \operatorname{rank} C_1 = c$ and the representation $(\mathrm{e}^{A_2 h}, K_2^{(h)}, C_2)$ of the stationary subsystem is minimal due to Lemma 4.3.8. This is automatically satisfied due to the minimality of the continuous-time cointegrated model and Assumption C2.

We adapt now a result from Ribarits and Hanzon, [84] (Lemma 3.1) in order characterize the rank of $k(1)$. Furthermore, we see that the cointegration information is contained in the linear filter, to be precise in the matrix $k(1)$.

Lemma 4.4.4

Suppose that the linear state space model (4.36) satisfies Assumption C2 and has the representation (4.40). Further, let r denote the number of linearly independent cointegration relations of the corresponding process $Y^{(h)}$. Then we have

1. $r = d - c$,
2. *rank* $k(1) =$ *rank* $\Big(I_d - C\big[I_N - \big(\mathrm{e}^{Ah} - K^{(h)}C\big)\big]^{-1} K^{(h)}\Big) = r.$

Proof. Lemma 4.4.3 gives us the eigenvalue structure of the sampled process.

1. $\widehat{X}_{n,2}^{(h)}$ is stationary with the special choice of $\widehat{X}_{0,2}^{(h)} = \sum_{j=0}^{\infty} \mathrm{e}^{A_2 hj} K_2^{(h)} \varepsilon_{-j}^{(h)}$, thus we have $\widehat{X}_{n,2}^{(h)} = \sum_{j=1}^{\infty} \mathrm{e}^{A_2 h(j-1)} K_2^{(h)} \varepsilon_{n-j}^{(h)}$. For any arbitrary initial value $\widehat{X}_{0,1}^{(h)}$, which we assume to be independent of $(\varepsilon_n^{(h)})_{n\in\mathbb{N}}$, the solution for $n \geq 0$, is given by

$$\begin{aligned} Y_n^{(h)} &= C_1 \widehat{X}_{0,1}^{(h)} + C_1 K_1^{(h)} \sum_{j=1}^{n-1} \varepsilon_{n-j}^{(h)} + \left[\sum_{j=1}^{\infty} C_2 \mathrm{e}^{A_2 h(j-1)} K_2^{(h)} \varepsilon_{n-j}^{(h)} + \varepsilon_n^{(h)} \right] \qquad (4.41) \\ &= C_1 \widehat{X}_{0,1}^{(h)} + C_1 K_1^{(h)} \sum_{j=1}^{n-1} \varepsilon_j^{(h)} + Y_{n,2}^{(h)}, \end{aligned}$$

where the part in the brackets corresponds to a stationary discrete time moving average process.

Due to this representation and the minimality, the number of common trends is obviously equal to $\operatorname{rank} C_1 K_1^{(h)} = c$. By (4.41) we see that $(Y_n^{(h)})$ consists of a random constant $(C_1 \widehat{X}_{0,1}^{(h)})$, a stationary part $(Y_{n,2}^{(h)})$ and an integrated process $(C_1 K_1^{(h)} \sum_{j=1}^{n-1} \varepsilon_{n-j}^{(h)})$.

Denote by $C_1^{\perp} \in M_{d,(d-c)}(\mathbb{R})$ the full rank matrix satisfying $C_1^{\perp \mathsf{T}} C_1^{\perp} = I_{d-c}$ and $C_1^{\mathsf{T}} C_1{}^{\perp} = 0_{c\times(d-c)}$, this means $C_1{}^{\perp}$ is the orthogonal complement of C_1 and

$\operatorname{rank}(C_1^\perp) = d - c$. Thus, it follows that

$$(C_1^{\perp\mathsf{T}} Y_n^{(h)})_{n\in\mathbb{Z}} = (C_1^{\perp\mathsf{T}} Y_{n,2}^{(h)})_{n\in\mathbb{Z}},$$

where $Y_2^{(h)}$ is given above. Because r denotes the number of independent cointegration relations, we have $r \geq \operatorname{rank} C_1^\perp = d - c$. On the other hand, since $C_1^T C_1 = I_c$ the number of cointegration relations $r \leq d - c$. Finally, $r = d - c$ which concludes the first part. Hence, the column space of $C_1^\perp$ spans the cointegration space.

2. We obtain for $k(1)$ the following representation by applying the decoupling into subsystems to (4.38)

$$k(1) = I_d - \begin{pmatrix} C_1 & C_2 \end{pmatrix} \begin{pmatrix} K_1^{(h)} C_1 & K_1^{(h)} C_2 \\ K_2^{(h)} C_1 & K_2^{(h)} C_2 + I_{N-c} - \mathrm{e}^{A_2 h} \end{pmatrix}^{-1} \begin{pmatrix} K_1^{(h)} \\ K_2^{(h)} \end{pmatrix}. \tag{4.42}$$

Since $K_1^{(h)}$ and C_1 have full rank c, the $c \times c$ matrix $K_1^{(h)} C_1$ is regular and has also rank c. We set $\widetilde{N} := N - c$. Furthermore, the matrix $(K_2^{(h)} C_2 + I_{\widetilde{N}} - \mathrm{e}^{A_2 h})$ is also nonsingular due to the Kalman filter, which implies $|\lambda_{max}(\mathrm{e}^{A_2 h} - K_2^{(h)} C_2)| < 1$. Thus, we can apply the Matrix Inversion Lemma (see e.g. Bernstein [9], Proposition 2.8.7) and obtain

$$k(1) = I_d - \begin{pmatrix} C_1 & C_2 \end{pmatrix} \cdot M \cdot \begin{pmatrix} K_1^{(h)} \\ K_2^{(h)} \end{pmatrix}, \tag{4.43}$$

where the matrix M is defined by

$$M := \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \tag{4.44}$$

where

$$\begin{aligned}
M_{11} &:= (K_1^{(h)} C_1)^{-1} + (K_1^{(h)} C_1)^{-1} K_1^{(h)} C_2 Q^{-1} K_2^{(h)} C_1 (K_1^{(h)} C_1)^{-1}, \\
M_{12} &:= -(K_1^{(h)} C_1)^{-1} K_1^{(h)} C_2 Q^{-1}, \\
M_{21} &:= -Q^{-1} K_2^{(h)} C_1 (K_1^{(h)} C_1)^{-1}, \\
M_{22} &:= Q^{-1},
\end{aligned}$$

with

$$Q := I_{\widetilde{N}} - \mathrm{e}^{A_2 h} + K_2^{(h)} C_2 - K_2^{(h)} C_1 (K_1^{(h)} C_1)^{-1} K_1^{(h)} C_2.$$

Define $P := I_d - C_1(K_1^{(h)}C_1)^{-1}K_1^{(h)} \in M_d(\mathbb{R})$, which is obviously idempotent since $P^2 = P$ holds. Note that the matrix product $K_1^{(h)}C_1$ is a nonsingular $c \times c$ matrix and consequently $C_1(K_1^{(h)}C_1)^{-1}K_1^{(h)}$ has rank c. Then we can rewrite the matrix $k(1)$ once more using the representation (4.44) and obtain

$$\begin{aligned}
k(1) =& I_d - \begin{pmatrix} C_1 & C_2 \end{pmatrix} \cdot M \cdot \begin{pmatrix} K_1^{(h)} \\ K_2^{(h)} \end{pmatrix} \\
=& I_d - C_1 M_{11} K_1^{(h)} - C_2 M_{21} K_1^{(h)} - C_1 M_{12} K_2^{(h)} - C_2 M_{22} K_2^{(h)} \\
=& (I_d - C_1(K_1^{(h)}C_1)^{-1}K_1^{(h)}) - C_2 Q^{-1} K_2^{(h)} \\
&+ C_1(K_1^{(h)}C_1)^{-1}K_1^{(h)} C_2 Q^{-1} K_2^{(h)} \\
&+ C_2 Q^{-1} K_2^{(h)} C_1 (K_1^{(h)}C_1)^{-1} K_1^{(h)} \\
&- C_1(K_1^{(h)}C_1)^{-1}K_1^{(h)} C_2 Q^{-1} K_2^{(h)} C_1 (K_1^{(h)}C_1)^{-1} K_1^{(h)} \\
=& P - (I_d - C_1(K_1^{(h)}C_1)^{-1}K_1^{(h)}) C_2 Q^{-1} K_2^{(h)} (I_d - C_1(K_1^{(h)}C_1)^{-1}K_1^{(h)}) \\
=& P - P C_2 Q^{-1} K_2^{(h)} P.
\end{aligned}$$

Since the matrix P is idempotent and $I - P = C_1(K_1^{(h)}C_1)^{-1}K_1^{(h)}$ has obviously rank c, we have due to the rank equation for an idempotent matrix, i.e. $\operatorname{rank} P + \operatorname{rank}(I - P) = d$ (see e.g. Bernstein [9], Fact 3.12.9), that $\operatorname{rank} P = d - c$. As above, by the Matrix Inversion Lemma (see e.g. Bernstein [9], Corollary 2.8.8) we can rewrite the matrix Q^{-1} as

$$\begin{aligned}
Q^{-1} =& [I_{\widetilde{N}} - \mathrm{e}^{A_2 h} + K_2^{(h)} P^2 C_2]^{-1} \\
=& \big(I_{\widetilde{N}} - \mathrm{e}^{A_2 h}\big)^{-1} \\
&- \big(I_{\widetilde{N}} - \mathrm{e}^{A_2 h}\big)^{-1} K_2^{(h)} P [I_d + P C_2 \big(I_{\widetilde{N}} - \mathrm{e}^{A_2 h}\big)^{-1} K_2^{(h)} P]^{-1} P C_2 \big(I_{\widetilde{N}} - \mathrm{e}^{A_2 h}\big)^{-1}.
\end{aligned}$$

For the sake of brevity, we write $R := C_2\big(I_{\widetilde{N}} - \mathrm{e}^{A_2 h}\big)^{-1} K_2^{(h)}$. Substituting the previous result into the formula of $k(1)$ leads to

$$\begin{aligned}
k(1) &= P - PRP + PRP(I_d + PRP)^{-1} PRP \\
&= P - PRP + (PRP)^2 (I_d + PRP)^{-1} \\
&= [(P - PRP)(P + PRP) + (PRP)^2](I_d + PRP)^{-1} \\
&= P(I_d + PRP)^{-1},
\end{aligned}$$

where we used the fact that $(I_d + AB)^{-1}A = A(I_d + BA)^{-1}$ for matrices $A(:= PRP)$ and $B(:= I_d)$ such that $I_d + AB$ is nonsingular (see e.g. Bernstein [9], Fact 2.16.16) for the second equality. Since we have

$$\operatorname{rank} k(1) = \operatorname{rank} P(I_d + PRP)^{-1} = \operatorname{rank} P = d - c.$$

□

Since $\operatorname{rank} k(1) = d - c$ there exists $\alpha, \beta \in M_{d,r}(\mathbb{R})$ with full row rank such that $k(1) = -\alpha\beta^T$. Note that $k(z) - k(1)z = 0$ for $z = 1$ and $k(z) - k(1)z = I_d$ for $z = 0$. Hence, we can rewrite $k(z)$ as

$$\begin{aligned} k(z) &= k(1)z + [k(z) - k(1)z] = k(1)z + (1-z)[I_d - \widetilde{k}(z)] \\ &= k(1)z + I_d(1-z) - \widetilde{k}(z)(1-z) = -\alpha\beta^{\mathsf{T}} z + I_d(1-z) - \widetilde{k}(z)(1-z), \end{aligned}$$

where

$$\widetilde{k}(z) := I_d - \frac{k(z) - k(1)z}{1-z} \tag{4.45}$$

We can now state an error correction form, where we consider a linear state space model instead of a VAR process in the classical error correction form. The so-called transfer function error correction form, was presented by Ribarits and Hanzon [84] for discrete-time state space models. However, we have a continuous-time state space model observed at discrete time points. Therefore, our error correction form will have a similar form as the one of Ribarits and Hanzon but we will have different matrices.

Definition 4.4.5
*The **error correction form** is given by*

$$\Delta Y_n^{(h)} = \alpha\beta^{\mathsf{T}} Y_{n-1}^{(h)} + \widetilde{k}(B)\Delta Y_n^{(h)} + \varepsilon_n^{(h)}, \quad n \in \mathbb{N}. \tag{4.46}$$

For comparison see the classical error correction model for a cointegrated VARMA process for example in Lütkepohl [62], Section 14.2.

Lemma 4.4.6
$\beta^{\mathsf{T}} Y^{(h)}$ is stationary and the rows of β span the cointegration space.

Proof. By the Matrix Inversion Lemma (see e.g. Bernstein [9], Corollary 2.8.8) and

(4.43) we obtain

$$\begin{aligned} k(1) &= I_d - \begin{pmatrix} C_1 & C_2 \end{pmatrix} \cdot \begin{pmatrix} K_1^{(h)} C_1 & K_1^{(h)} C_2 \\ K_2^{(h)} C_1 & K_2^{(h)} C_2 + I_{N-c} - \mathrm{e}^{A_2 h} \end{pmatrix}^{-1} \cdot \begin{pmatrix} K_1^{(h)} \\ K_2^{(h)} \end{pmatrix} \\ &= [I_d - C_2(I_{\widetilde{N}} - \mathrm{e}^{A_2 h} + K_2^{(h)} C_2)^{-1} K_2^{(h)}] \\ &\quad -[I_d - C_2(I_{\widetilde{N}} - \mathrm{e}^{A_2 h} + K_2^{(h)} C_2)^{-1} K_2^{(h)}] \\ &\qquad \cdot C_1 (K_1^{(h)} [I_d - C_2(I_{\widetilde{N}} - \mathrm{e}^{A_2 h} + K_2^{(h)} C_2)^{-1} K_2^{(h)}] C_1)^{-1} \\ &\qquad \cdot K_1^{(h)} [I_d - C_2(I_{\widetilde{N}} - \mathrm{e}^{A_2 h} + K_2^{(h)} C_2)^{-1} K_2^{(h)}]. \end{aligned}$$

We receive

$$\begin{aligned} k(1) C_1 =& [I_d - C_2(I_{\widetilde{N}} - \mathrm{e}^{A_2 h} + K_2^{(h)} C_2)^{-1} K_2^{(h)}] C_1 \\ & - [I_d - C_2(I_{\widetilde{N}} - \mathrm{e}^{A_2 h} + K_2^{(h)} C_2)^{-1} K_2^{(h)}] C_1 \\ & \cdot (K_1^{(h)} [I_d - C_2(I_{\widetilde{N}} - \mathrm{e}^{A_2 h} + K_2^{(h)} C_2)^{-1} K_2^{(h)}] C_1)^{-1} \\ & \cdot K_1^{(h)} [I_d - C_2(I_{\widetilde{N}} - \mathrm{e}^{A_2 h} + K_2^{(h)} C_2)^{-1} K_2^{(h)}] C_1 \\ =& [I_d - C_2(I_{\widetilde{N}} - \mathrm{e}^{A_2 h} + K_2^{(h)} C_2)^{-1} K_2^{(h)}] C_1 \\ & - [I_d - C_2(I_{\widetilde{N}} - \mathrm{e}^{A_2 h} + K_2^{(h)} C_2)^{-1} K_2^{(h)}] C_1 \cdot I_c \\ =& 0_{d \times c}. \end{aligned}$$

This means $\alpha\beta^T C_1 = k(1)C_1 = 0_{d\times c}$. Since α and β have full rank $r = d - c$ and $\operatorname{rank} C_1^{\perp} = d - c$, $C_1^{\perp}$ and β span the same space. Due to Lemma 4.3.9 we can conclude the statement. □

In particular $\epsilon^{(h)}$ is then as a sum of stationary processes stationary itself. In the following we present some alternative representations for the innovation sequence which help to derive some further properties of the innovation sequence.

Lemma 4.4.7
Write $k(z) = \sum_{j=1}^{\infty} L_j z^j$ and $\widetilde{k}(z) = \sum_{j=1}^{\infty} \widetilde{K}_j z^j$, and define

$$\overline{k}(z) := I_d - \widetilde{k}(z). \tag{4.47}$$

Then the following alternative representations for the innovation sequence hold.

(a) $\varepsilon_n^{(h)} = k(B) Y_{n,2}^{(h)} + \overline{k}(B) C_1 R_{n,1}^{(h)}, \quad n \in \mathbb{N}.$

(b) $\varepsilon_n^{(h)} = \sum_{j=0}^{\infty} \left(\widetilde{K}_j C_1 B^j \quad \sum_{k=0}^{j} L_k C_2 \mathrm{e}^{A_2 h (j-k)} B^j \right) R_n^{(h)}, \quad n \in \mathbb{N}$, *where* $\widetilde{K}_0 = 0$ *and* $\widetilde{K}_j = C \sum_{i=j+1}^{\infty} \left(\mathrm{e}^{Ah} - K^{(h)} C \right)^{i-1} K^{(h)}, \quad$ *for* $j \geq 1$.

From (a) we see that the innovations consist of two summands containing on the one hand, the filtered version of the stationary part and on the other hand, the filtered version of the increments of the driving Lévy process.

Proof.
(a) Using (4.45) we obtain

$$k(z) - k(1)z = (1-z)[I_d - \widetilde{k}(z)] = (1-z)\overline{k}(z)$$

so that (4.46) and (4.28) result in

$$\begin{aligned}
\varepsilon_n^{(h)} &= \Delta Y_n^{(h)} + k(1)BY_n^{(h)} - \widetilde{k}(B)\Delta Y_n^{(h)} = k(1)BY_{n,2}^{(h)} + \overline{k}(B)\Delta Y_n^{(h)} \\
&= k(1)BY_{n,2}^{(h)} + \overline{k}(B)\left(C_1 R_{n,1}^{(h)} + (1-B)Y_{n,2}^{(h)}\right) \\
&= k(B)Y_{n,2}^{(h)} + \overline{k}(B)C_1 R_{n,1}^{(h)}, \quad n \in \mathbb{N}.
\end{aligned}$$

(b) With the moving average representation of $Y_2^{(h)}$ and the Cauchy product we receive

$$\begin{aligned}
\varepsilon_n^{(h)} &= \overline{k}(B)C_1 R_{n,1}^{(h)} + k(B)Y_{n,2}^{(h)} \\
&= \sum_{j=0}^{\infty} \widetilde{K}_j B^j C_1 R_{n-j,1}^{(h)} + \sum_{j=0}^{\infty} L_j B^j C_2 \sum_{i=0}^{\infty} \mathrm{e}^{A_2 h i} B^i R_{n,2}^{(h)} \\
&= \sum_{j=0}^{\infty} \widetilde{K}_j B^j C_1 R_{n,1}^{(h)} + \sum_{i=0}^{\infty} \sum_{j=0}^{k} L_j B^j C_2 \mathrm{e}^{A_2 h (i-j)} B^{i-j} R_{n,2}^{(h)} \\
&= \sum_{j=0}^{\infty} \widetilde{K}_j B^j C_1 R_{n,1}^{(h)} + \sum_{j=0}^{\infty} \sum_{k=0}^{j} L_k C_2 \mathrm{e}^{A_2 h (j-k)} B^j R_{n,2}^{(h)} \\
&= \sum_{j=0}^{\infty} \begin{pmatrix} \widetilde{K}_j C_1 B^j & \sum_{k=0}^{j} L_k C_2 \mathrm{e}^{A_2 h (j-k)} B^j \end{pmatrix} \begin{pmatrix} R_{n,1}^{(h)} \\ R_{n,2}^{(h)} \end{pmatrix} \\
&= \sum_{j=0}^{\infty} \begin{pmatrix} \widetilde{K}_j C_1 B^j & \sum_{k=0}^{j} L_k C_2 \mathrm{e}^{A_2 h (j-k)} B^j \end{pmatrix} R_n^{(h)}, \quad n \in \mathbb{N}.
\end{aligned}$$

We determine the matrix coefficients $\widetilde{K}_i$ in the subsequent considerations. It can easily be seen that $\widetilde{k}(0) = \widetilde{K}_0 = 0$. By rearranging (4.45) we obtain $k(z) - k(1)z = (1-z)[I_d - \widetilde{k}(z)]$, which is equivalent to

$$(1-z)\left[I_d - \sum_{i=1}^{\infty} \widetilde{K}_i z^i\right] = \left(I_d - C\sum_{i=1}^{\infty} (\mathrm{e}^{Ah} - K^{(h)}C)^{i-1} K^{(h)} z^i\right)$$

$$- \left(I_d - C \sum_{i=1}^{\infty} (\mathrm{e}^{Ah} - K^{(h)} C)^{i-1} K^{(h)} \right) z$$

and thus by comparison of the coefficients we obtain

$$\begin{aligned}
z^1 : \quad & -CK^{(h)} - I_d + C \sum_{i=1}^{\infty} \left(\mathrm{e}^{Ah} - K^{(h)} C\right)^{i-1} K^{(h)} = -I_d - \widetilde{K}_1 \\
& \Rightarrow \quad \widetilde{K}_1 = C \sum_{i=2}^{\infty} \left(\mathrm{e}^{Ah} - K^{(h)} C\right)^{i-1} K^{(h)}, \\
z^2 : \quad & -C\left(\mathrm{e}^{Ah} - K^{(h)} C\right) K^{(h)} = \widetilde{K}_2 - \widetilde{K}_1 \\
& \Rightarrow \quad \widetilde{K}_2 = C \sum_{i=3}^{\infty} \left(\mathrm{e}^{Ah} - K^{(h)} C\right)^{i-1} K^{(h)}, \\
& \qquad \vdots \\
z^j : \quad & -C\left(\mathrm{e}^{Ah} - K^{(h)} C\right)^{j-1} K^{(h)} = \widetilde{K}_j - \widetilde{K}_{j-1} \\
& \Rightarrow \quad \widetilde{K}_j = C \sum_{i=j+1}^{\infty} \left(\mathrm{e}^{Ah} - K^{(h)} C\right)^{i-1} K^{(h)}, \quad \text{for } j \geq 1.
\end{aligned}$$

This concludes the proof. □

We already have seen that the linear innovations are stationary. We investigate now the ergodicity of the innovations and sum up useful properties of the linear innovation sequence in the next proposition. These enable us in the next chapter to derive limit results necessary for the theoretical foundation of the quasi maximum likelihood estimation.

Proposition 4.4.8
The linear innovations $\varepsilon^{(h)}$ given by (4.33) are a stationary, ergodic and uncorrelated sequence. Furthermore, the linear innovations $\varepsilon^{(h)}$ have finite second moments.

Proof. As already mentioned $\varepsilon^{(h)}$ is stationary because all other terms appearing in (4.46) are stationary. For the ergodic property we first note $R_1^{(h)}$ is an i.i.d. sequence and $Y_2^{(h)}$ is ergodic, which was already shown in Schlemm and Stelzer [91]. Then, the vector process

$$(Z_n^{(h)})_{n\in\mathbb{N}} = \begin{pmatrix} Y_{n,2}^{(h)} \\ C_1 R_{n,1}^{(h)} \end{pmatrix}_{n\in\mathbb{N}}$$

is obviously stationary and ergodic. Since, we can define a measurable function f

such that

$$\varepsilon_n^{(h)} = f(Z_n^{(h)}), \quad n \in \mathbb{N}$$

we obtain that $\varepsilon^{(h)}$ is ergodic with Bradley [16], Proposition 2.10 (ii). Note that we could have proved stationarity of the linear innovations $\varepsilon^{(h)}$ using the same arguments as for the ergodicity.

Recall that $|\lambda_{max}(e^{Ah} - K^{(h)}C)| < 1$ holds and the form of the transfer function (4.37). Hence, the existence of second moments of the linear innovations $\varepsilon^{(h)}$ follows directly from the finite second moments of the driving Lévy process and the fact that the transfer function $k(z)$ has exponentially decaying coefficients.

All remaining assertions follow by Hannan and Deistler, [46], Chapter 4, and Brockwell and Davis, [20], Chapter 12. □

Note that if we have a Lévy process consisting only of a Brownian motion, the noise is Gaussian. This even implies that the linear innovations are a sequence of i.i.d. random variables.

Due to the representations and results on the linear innovations $\varepsilon^{(h)}$ and the sampled process $Y^{(h)}$ in this section we are able to derive several asymptotic results later on. Not only does a law of large numbers for the innovations hold, but we have in particular a functional central limit theorem and thus weak convergence to a stochastic integral due to the integrated part of $Y^{(h)}$. These weak convergence results are the key results enabling us to derive all necessary asymptotic results, in order to prove the consistency of the step-wise quasi maximum likelihood estimator and derive the asymptotic distribution in Chapter 5. For a profound derivation of the auxiliary asymptotic results we refer to Appendix 5.8.

4.5. Conclusion

Many time series do not behave in a stationary way, e.g. financial time series data. Hence, non-stationary models are of particular interest in order to model such behavior. One particular class of non-stationary processes are cointegrated models. We extended in this chapter existing continuous-time cointegrated models to the general Lévy driven MCARMA case.

Moreover, we have seen in this section a type of Johansen-Granger Representation Theorem characterizing cointegration. We considered the cointegrated Lévy driven MCARMA process from two different perspectives. On the one hand, we used the representation with the autoregressive and moving average polynomial and on the other hand, the state space representation. In the first case we characterized the cointegration property in terms of certain matrix coefficients. In contrast, a state space model is cointegrated if the system matrix A has several eigenvalues equal to zero. This leads into a decoupled state space system, where the integrated part is decoupled of the stationary part. The derived representation is very flexible since we can freely model these two parts. This has the advantage that we can make a restriction to these parts only and can find very easily a unique parametrization. Above all, we investigated the probabilistic properties of the cointegrated process.

Furthermore, the solution of the cointegrated state space model is a sum of a Lévy process and a stationary MCARMA process. Hence, this representation enables to use the properties of the stationary MCARMA process. This representation is a continuous-time analogue to the representation in the discrete-time case given in the Johansen-Granger Representation Theorem. Besides, the cointegration space can be easily recovered from this representation. We investigated the representation and probabilistic properties of the sampled process, which inherits many properties from the continuous-time cointegrated MCARMA process.

Last but not least, we showed the applicability of the Kalman filter in the non-stationary setting with unit roots. By means of the Kalman filter, we obtained the linear innovations $\varepsilon^{(h)}$ of the discrete observations $Y^{(h)}$ and derived a error correction representation for the linear innovations which resembles the classical error correction form of cointegrated processes, i.e. we have $\varepsilon_n^{(h)} = \Pi Y_{n-1}^{(h)} + \overline{k}(B)\Delta Y_n^{(h)}$. Furthermore, we showed that the cointegration information is preserved in the filtered model. To be precise the matrix Π is singular and contains the cointegration space and thus the cointegration rank.

4.6. Appendix: Derivation of the Kalman Filter for State Space Models with Unit Roots

We are going use the notation of Schlemm and Stelzer [91] in the succeeding part, where we derive the Kalman filter for a cointegrated state space model. The standard approach does not work here since the continuous cointegrated model is only semistable and not asymptotically stable. This property transfers to the discretised process. Hence, we have to check the applicability of the Kalman filter first. For an introduction to Kalman Filtering and Linear State Space Models see e.g. Anderson and Moore [3] or Hannan and Deistler [46].

Before we can use the properties of the Kalman filter, we have to check if the Kalman filter is applicable in this setting. Due to the unit roots we have to check under which additional assumptions the standard framework of the Kalman filter works. As we will see in this section, we have to initialize the Kalman filter with a positive definite matrix. Moreover, we have that under standard assumption also the convergence of the Kalman filter for the cointegrated model holds. However, we do not verify the results for the special form the cointegrated model takes. Instead, we consider a more general state space model which includes the model considered in Chapter 5.

Definition 4.6.1
An $\mathbb{R}^d$-valued ***discrete-time linear stochastic state space model*** (F, H, Z, W) *of dimension N is characterized by a strictly stationary $\mathbb{R}^{p+d}$-valued sequence $(Z^\intercal \; W^\intercal)^\intercal$ with zero mean and finite covariance matrix*

$$\mathbb{E}\left[\begin{pmatrix} Z_n \\ W_n \end{pmatrix} \begin{pmatrix} Z_n^\intercal & W_n^\intercal \end{pmatrix}\right] = \delta_{m,n} \begin{pmatrix} Q & 0 \\ 0 & R \end{pmatrix}, \quad n, m \in \mathbb{Z}, \tag{4.48}$$

for some matrices $Q \in \mathbb{S}_p^+(\mathbb{R})$, $R \in \mathbb{S}_d^{++}(\mathbb{R})$ or $R \equiv 0_{d\times d}$; a state transition matrix $F \in M_N(\mathbb{R})$; and an observation matrix $H \in M_{d,N}(\mathbb{R})$. It consist of a state equation

$$X_n = F X_{n-1} + \Gamma Z_{n-1}, \quad n \in \mathbb{Z} \tag{4.49a}$$

and an observation equation

$$Y_n = H X_n + W_n, \quad n \in \mathbb{Z}, \tag{4.49b}$$

where $\Gamma \in M_{N,p}$. The $\mathbb{R}^N$-valued autoregressive process $X = (X_n)_{n\in\mathbb{Z}}$ is called the state vector process, and $Y = (Y_n)_{n\in\mathbb{Z}}$ is the output process.

Note that the concept of observability and controllability are the same as in the continuous time case. The linear system (4.49) is observable if the observability matrix $\mathcal{O}_{HF}$ has full rank. Observability guarantees that we can reconstruct the state vector X_n given the observations $Y_n, \ldots, Y_{n+N-1}$.

The same holds for controllability, where the linear system (4.49) is controllable if the controllability matrix $\mathcal{C}_{F\Gamma}$ has full rank. Controllability guarantees that for a fixed value X_n we can reach any arbitrary specified value X^* at time point $n + N$ by designing a certain input sequence for which we obtain the required terminal state.

We give now the analogue definition of minimality for the discrete-time case (c.f. Definition 4.3.4).

Definition 4.6.2
Let $k \in M_{d,m}(\mathbb{R}\{z\})$ be a rational matrix function. A matrix triple (F, Γ, H) is called an ***algebraic realization*** *of k of dimension N if $k(z) = H(zI_N - F)^{-1}\Gamma$, where $F \in M_N(\mathbb{R})$, $\Gamma \in M_{N,p}(\mathbb{R})$ and $H \in M_{d,N}(\mathbb{R})$. The matrix triple (F, Γ, H) is called* ***minimal*** *if there exists no other algebraic realization $(\widetilde{F}, \widetilde{\Gamma}, \widetilde{H})$ with dimension smaller than N.*

This is the state space equivalent to left coprimeness for VARMA processes. Minimality can be characterized by observability and controllability, which is stated in the next theorem.

Theorem 4.6.3 (Hannan and Deistler (1988), Theorem 2.3.3.)
The linear system (4.49) is minimal if it is controllable and observable, i.e.

$$rank \mathcal{C}_{F\Gamma} = rank \mathcal{O}_{HF} = N. \tag{4.50}$$

In order to prove the results in this section we postulate a number of assumptions.

Assumption K1
The initial state X_0 is independent of Z and W, i.e. $\mathbb{E}(X_0 Z_n^\mathsf{T}) = 0$ and $\mathbb{E}(X_0 W_n^\mathsf{T}) = 0$ for all n.

The formal derivation of the Kalman filter, in the case of $R \in \mathbb{S}_d^{++}(\mathbb{R})$, can be found e.g. in Harvey [48] or Chui and Chen [26]. In the case Gaussian case we derive the optimal minimum mean-square error estimator, whereas in the non-Gaussian case the Kalman filter is derived as the best linear estimator.

Normally in the Kalman filter setting one has the following results (see e.g. Brockwell and Davis [20], Proposition 12.2.3, or Hamilton [45], Proposition 13.2). The optimal linear filter for the system (4.49) is given by

$$\widehat{X}_{n+1} =(F - K_n H)\widehat{X}_n + K_n Y_n, \quad \widehat{X}_0 = \overline{x}_0, \tag{4.51}$$

where the Kalman gain matrix K_n satisfies

$$K_n :=F\Omega_n H^\mathsf{T}[H\Omega_n H^\mathsf{T} + R]^{-1} \in M_{N,d}(\mathbb{R}) \tag{4.52}$$

and Ω_n satisfies the matrix Riccati difference equation

$$\Omega_{n+1} =F\Omega_n F^\mathsf{T} - F\Omega_n H^\mathsf{T}(H\Omega_n H^\mathsf{T} + R)^{-1} H\Omega_n F^\mathsf{T} + \Gamma Q \Gamma^\mathsf{T} \in \mathbb{S}_N^{+}(\mathbb{R}). \tag{4.53}$$

If Ω_n converges for $n \to \infty$, then the limiting solution Ω satisfies the following algebraic Riccati equation obtained from (4.53)

$$\Omega = F\Omega F^{\mathsf{T}} - F\Omega H^{\mathsf{T}}(H\Omega H^{\mathsf{T}} + R)^{-1} H\Omega F^{\mathsf{T}} + \Gamma Q \Gamma^{\mathsf{T}} \in \mathbb{S}_N^{++}(\mathbb{R}) \tag{4.54}$$

and the steady state Kalman gain matrix K is given by

$$K := F\Omega H^{\mathsf{T}}[H\Omega H^{\mathsf{T}} + R]^{-1} \in M_{N,d}(\mathbb{R}). \tag{4.55}$$

However, for $R \equiv 0_{d\times d}$ some inverses above might not be well defined. Since we have not found a derivation of the Kalman filter for the case of $R \equiv 0_{d\times d}$ in the literature, without the problem of singular matrices (compare Hamilton [45], Proposition 13.1 and 13.2), we present a derivation in the subsequent part in order to validate the results we just presented in the last few lines for the cointegrated setting.

We follow Chui and Chen [26] and adapt their proofs to our setting in order to be able to work with the derived state space form in the MCARMA setting (c.f. Proposition 4.4.2).

The Kalman filter minimizes the mean square error of the estimated parameter vectors. In particular, the Kalman filter is the best estimator in the Gaussian case among all filters. Nevertheless, if the noise is non-Gaussian, it is the best linear filter among all linear filters, i.e. a non-linear estimator might perform better (c.f. Anderson and Moore [3], Section 5.4).

Assumption K2
Let $R \equiv 0_{d\times d}$ and $\Gamma Q \Gamma^{\mathsf{T}} \in \mathbb{S}_N^{++}(\mathbb{R})$, i.e. $\Gamma Q \Gamma^{\mathsf{T}}$ is positive-definite.

The a priori and a posteriori estimator of X_n in the Gaussian (respectively non-Gaussian) setting are given by

$$\begin{aligned} \widehat{X}_{n|n-1} &= \mathbb{E}[X_n | Y_0, \ldots, Y_{n-1}] (= P_{n-1} X_n), \\ \widehat{X}_{n|n} &= \mathbb{E}[X_n | Y_0, \ldots, Y_n] (= P_n X_n) \end{aligned}$$

and the error of the estimation is given by the error covariance matrix

$$\Omega_{n+1,n} = \mathbb{E}\left[(X_{n+1} - \widehat{X}_{n+1})(X_{n+1} - \widehat{X}_{n+1})^{\mathsf{T}} | Y_0, \ldots, Y_n\right].$$

For reasons of brevity we write $\Omega_n := \Omega_{n,n-1}$ and $\widehat{X}_n := \widehat{X}_{n|n}$.

The Kalman filtering equations are given by (see e.g. Anderson and Moore [3])

$$\begin{cases} \Omega_{0,-1} &= \operatorname{Var}(X_0) \\ \Omega_{n+1,n} &= F[\Omega_{n,n-1} - \Omega_{n,n-1}H^\top (H\Omega_{n,n-1}H^\top)^{-1}H\Omega_{n,n-1}]F^\top + \Gamma Q \Gamma^\top \\ &= F\Omega_{n,n}F^\top + \Gamma Q\Gamma^\top \\ \Omega_{n,n} &= \Omega_{n,n-1} - \Omega_{n,n-1}H^\top (H\Omega_{n,n-1}H^\top)^{-1}H\Omega_{n,n-1} \\ \widehat{X}_{0|-1} &= \mathbb{E}[X_0] \\ \widehat{X}_{n+1|n} &= (F - K_nH)\widehat{X}_{n|n-1} + K_nY_n = F\widehat{X}_{n|n} \\ \widehat{X}_{n|n} &= \widehat{X}_{n|n-1} + \Omega_{n,n-1}H^\top (H\Omega_{n,n-1}H^\top)^{-1}(Y_n - H\widehat{X}_{n|n-1}) \\ K_n &= \Omega_{n,n-1}H^\top (H\Omega_{n,n-1}H^\top)^{-1}, \end{cases} \tag{4.56}$$

where K_n is the so-called Kalman gain matrix.

In order to guarantee well-defined Kalman filtering equations we need the following assumptions.

Assumption K3
Let H be of full rank and Ω_0 be positive definite.

Then $H\Omega_0H^\top$ is non-singular. Likewise, we show later that $H\Omega_nH^\top$ is non-singular for all $n \in \mathbb{N}$, hence the Kalman gain matrix and the Riccati equation are well-defined.

The recursion for the covariance Ω_n is given by using the Kalman filtering equations

$$\begin{aligned} \Omega_n &= F\Omega_{n-1,n-1}F^\top + \Gamma Q\Gamma^\top \\ &= (F - K_{n-1}H)\Omega_{n-1,n-2}F^\top + \Gamma Q\Gamma^\top \\ &= (F - F\Omega_{n-1,n-2}H^\top (H\Omega_{n-1,n-2}H^\top)^{-1}H)\Omega_{n-1,n-2}F^\top + \Gamma Q\Gamma^\top \\ &= (F - F\Omega_{n-1}H^\top (H\Omega_{n-1}H^\top)^{-1}H)\Omega_{n-1}F^\top + \Gamma Q\Gamma^\top, \end{aligned} \tag{4.57}$$

then we define analogous to Chui and Chen

$$\Psi(T) := (F - FTH^\top (HTH^\top)^{-1}H)TF^\top + \Gamma Q\Gamma^\top. \tag{4.58}$$

Hence, we have for Ω_n the discrete-time Riccati difference equation $\Omega_n = \Psi(\Omega_{n-1})$ and the discrete-time algebraic Riccati equation (DARE) $\Omega = \Psi(\Omega)$. We can rewrite the Riccati equation in the following form using (4.57)

$$\begin{aligned} \Omega_n =&(F - K_{n-1}H)\Omega_{n-1}F^\top + \Gamma Q\Gamma^\top \\ =&(F - K_{n-1}H)\Omega_{n-1}(F - K_{n-1}H)^\top + \Gamma Q\Gamma^\top \end{aligned}$$

$$
\begin{aligned}
&+ (F - K_{n-1}H)\Omega_{n-1}(K_{n-1}H)^{\mathsf{T}} \\
=&(F - K_{n-1}H)\Omega_{n-1}(F - K_{n-1}H)^{\mathsf{T}} + \Gamma Q \Gamma^{\mathsf{T}}
\end{aligned} \tag{4.59}
$$

The last equation holds because $(F - K_{n-1}H)\Omega_{n-1}(K_{n-1}H)^{\mathsf{T}}$ is zero, which can be seen by inserting the definition of K_n.

In order to show the convergence some auxiliary results are necessary, which are adapted versions of Lemmas from Chui and Chen [26]. The first Lemma provides an upper bound for the covariance Ω_n.

Lemma 4.6.4 (c.f. Chui and Chen (2009), Lemma 6.1)
Suppose that the linear system (4.49) is observable. Then there exists a non-negative definite symmetric constant matrix W independent of the positive-definite initial value Ω_0 such that for all $n \geq N + 1$

$$\Omega_n \leq W.$$

Proof. We define $\langle x, w \rangle := \operatorname{Cov}(x, w) = \mathbb{E}[(x - \mathbb{E}[x])(w - \mathbb{E}[w])^{\mathsf{T}}]$ and $\|w\|_N^2 := \langle w, w \rangle := \|w\|_N \|w\|_N^{\mathsf{T}}$. Let $\widehat{X}_{n-1|n-1}$ be the (linear) minimum variance estimate of X_{n-1} and

$$
\begin{aligned}
\Omega_n =& \|X_n - \widehat{X}_{n|n-1}\|_N^2 \\
=& \|F X_{n-1} + \Gamma Z_{n-1} - F \widehat{X}_{n-1|n-1}\|_N^2 \\
=& F \|X_{n-1} - \widehat{X}_{n-1|n-1}\|_N^2 F^{\mathsf{T}} + \Gamma Q \Gamma^{\mathsf{T}},
\end{aligned}
$$

which holds due to the independence of $X_{n-1} - \widehat{X}_{n-1|n-1}$ and ΓZ_{n-1}. Since $\widehat{X}_{n-1|n-1}$ is the (linear) minimum variance estimate, we have

$$\|X_n - \widehat{X}_{n-1|n-1}\|_N^2 \leq \|X_n - \tilde{X}_{n-1}\|_N^2$$

for any arbitrary unbiased estimate $\tilde{X}_{n-1}$ of X_{n-1}. The observation matrix has full rank, hence $\mathcal{O}_{HF}^{\mathsf{T}} \mathcal{O}_{HF}$ is a non-singular matrix given by

$$\mathcal{O}_{HF}^{\mathsf{T}} \mathcal{O}_{HF} = \sum_{i=0}^{N-1} (F^{\mathsf{T}})^i H^{\mathsf{T}} H F^i.$$

Choose

$$\tilde{X}_{n-1} := F^N (\mathcal{O}_{HF}^{\mathsf{T}} \mathcal{O}_{HF})^{-1} \sum_{i=0}^{N-1} (F^{\mathsf{T}})^i H^{\mathsf{T}} Y_{n-N-1+i}$$

$$
\begin{aligned}
&= F^N (\mathcal{O}_{HF}^\mathsf{T} \mathcal{O}_{HF})^{-1} \sum_{i=0}^{N-1} (F^\mathsf{T})^i H^\mathsf{T} \left(H F^i X_{n-N-1} + \sum_{j=0}^{i-1} H F^j \Gamma Z_{i-1-j} \right) \\
&= F^N X_{n-N-1} + F^N (\mathcal{O}_{HF}^\mathsf{T} \mathcal{O}_{HF})^{-1} \sum_{i=0}^{N-1} (F^\mathsf{T})^i H^\mathsf{T} \left(\sum_{j=0}^{i-1} H F^j \Gamma Z_{i-1-j} \right),
\end{aligned}
$$

for $n \geq N+1$. Then we obtain

$$
\begin{aligned}
&X_{n-1} - \tilde{X}_{n-1} \\
&= \sum_{i=0}^{N-1} F^i \Gamma Z_{N-1+i} - F^N (\mathcal{O}_{HF}^\mathsf{T} \mathcal{O}_{HF})^{-1} \sum_{i=0}^{N-1} (F^\mathsf{T})^i H^\mathsf{T} \left(\sum_{j=0}^{i-1} H F^j \Gamma Z_{i-1-j} \right).
\end{aligned}
$$

Obviously, we have that $\|X_n - \tilde{X}_{n-1}\|_N^2$ is independent of n for all $n \geq N+1$. Thus,

$$
\begin{aligned}
\Omega_n &= F \|X_{n-1} - \widehat{X}_{n-1|n-1}\|_N^2 F^\mathsf{T} + \Gamma Q \Gamma^\mathsf{T} \\
&\leq F \|X_{n-1} - \tilde{X}_{n-1|n-1}\|_N^2 F^\mathsf{T} + \Gamma Q \Gamma^\mathsf{T} \\
&\leq F \|X_N - \tilde{X}_{N|N}\|_N^2 F^\mathsf{T} + \Gamma Q \Gamma^\mathsf{T} =: W, \quad \text{for all } n \geq N+1.
\end{aligned}
$$

Additionally, W is independent of the initial condition $\Omega_0 = \Omega_{0,-1} = \|X_0 - \widehat{X}_{0|-1}\|_N^2$. Thus, we have shown all claims. □

The next Lemma states, that the Riccati equation satisfies an ordering property. The proof can be found in Anderson and Moore [3] or Chui and Chen [26].

Lemma 4.6.5 (Chui and Chen (2009), Lemma 6.2)
If A and B are both positive-definite and symmetric with $A \geq B$, then $\Psi(A) \geq \Psi(B)$ for $\Psi(\cdot)$ as in (4.58).

Due to the ordering property we can guarantee that the Riccati difference equation will be well-defined.

Lemma 4.6.6 (c.f. Chui and Chen (2009), Lemma 6.3)
Suppose that the linear system (4.49) is observable. Then with the initial condition $\Omega_0 = \Omega_{0,-1} = \Gamma Q \Gamma^\mathsf{T}$, the sequence $\{\Omega_n\}$ converges componentwise to some symmetric positive definite matrix $\Omega > 0$ as $n \to \infty$. Further, we have that Ω_n is positive definite for all $n \in \mathbb{N}$ and thus the Riccati difference equation is always well-defined.

Proof. Since by Assumption K2 $\Omega_0 = \Gamma Q \Gamma^\mathsf{T} > 0$ holds and by (4.59), we have

$$
\Omega_1 = (F - K_0 H) \Omega_0 (F - K_0 H)^\mathsf{T} + \Gamma Q \Gamma^\mathsf{T} > 0.
$$

Further, we have

$$\Omega_1 - \Omega_0 = (F - K_0 H)\Omega_0 (F - K_0 H)^{\mathsf{T}} + \Gamma Q \Gamma^{\mathsf{T}} - \Gamma Q \Gamma^{\mathsf{T}} \geq 0.$$

Since Ω_1 and Ω_0 are symmetric, we have by Lemma 4.6.5 that

$$\Omega_{n+1} \geq \Omega_n > 0 \quad \text{for all } n = 0, 1, \ldots.$$

Hence, Ω_n is monotonic nondecreasing and bounded by the matrix W due to Lemma 4.6.4. Therefore, we have for any $y \in \mathbb{R}^N$

$$0 < y^{\mathsf{T}} \Omega_0 y \leq y^{\mathsf{T}} \Omega_n y \leq y^{\mathsf{T}} W y,$$

thus the real-valued sequence $\{y^{\mathsf{T}} \Omega_n y\}$ is also bounded and monotonic nondecreasing and converges therefore to some positive constant. Choose $y = [0 \ldots 0\ 1\ 0 \ldots 0]^{\mathsf{T}}$, where the i^{th} component is non-zero. It follows that for $\Omega_n = [\Omega_{ij}^{(n)}]$

$$y^{\mathsf{T}} \Omega_n y = \Omega_{ii}^{(n)} \to p_{ii} \quad \text{as } n \to \infty,$$

where $p_{ii} > 0$. Now, choose $y = [0 \ldots 0\ 1\ 0 \ldots 0\ 1\ 0 \ldots 0]^{\mathsf{T}}$, where the i^{th} and j^{th} component is non-zero. Then we have

$$\begin{aligned} y^{\mathsf{T}} \Omega_n y &= \Omega_{ii}^{(n)} + \Omega_{ij}^{(n)} + \Omega_{ji}^{(n)} + \Omega_{jj}^{(n)} \\ &= \Omega_{ii}^{(n)} + 2\Omega_{ij}^{(n)} + \Omega_{jj}^{(n)} \to q \quad \text{as } n \to \infty, \end{aligned}$$

where $q > 0$. Thus

$$\Omega_{ij}^{(n)} \to \frac{1}{2} \left(q - \Omega_{ii} - \Omega_{jj}\right) \text{ as } n \to \infty,$$

i.e. $\Omega_n \to \Omega$. Since Ω_n is symmetric and positive definite, so is Ω. □

Now, we have the limit of the Kalman gain matrix given by Lemma 4.6.6

$$K = \lim_{n \to \infty} K_n = F \Omega H^{\mathsf{T}} (H \Omega H^{\mathsf{T}})^{-1}. \tag{4.60}$$

The analogue alternative representation for the (DARE) of an observable system is given by taking the limit in (4.59). We have

$$\Omega = (F - KH)\Omega(F - KH)^{\mathsf{T}} + \Gamma Q \Gamma^{\mathsf{T}}, \tag{4.61}$$

where K was defined in (4.60).

Lemma 4.6.7

Suppose that the linear system (4.49) is observable. Let the initial matrix be $\Omega_0 = \Gamma Q \Gamma^\mathsf{T}$ *and* Ω *be defined as in Lemma 4.6.6. Then all eigenvalues of* $(F - KH)$ *must be inside the unit circle and consequently the Kalman filter is asymptotically stable.*

Proof. Let $x \in \mathbb{C}^N$ denote an eigenvector of $(F - KH)^\mathsf{T}$ and λ its eigenvalue, then

$$(F - KH)^\mathsf{T} x = \lambda x. \tag{4.62}$$

Then

$$\begin{aligned} x^*(F - KH)\Omega(F - KH)^\mathsf{T} x &= [(F - KH)^\mathsf{T} x]^* \Omega [(F - KH)^\mathsf{T} x] \\ &= [\lambda x]^* \Omega [\lambda x] \\ &= |\lambda|^2 x^* \Omega x. \end{aligned} \tag{4.63}$$

Now, we multiply the (DARE) (4.61) from left with x^* and from right with x, we obtain

$$x^* \Omega x = |\lambda|^2 x^* \Omega x + x^* \Gamma Q \Gamma^\mathsf{T} x,$$

which is equivalent to

$$(1 - |\lambda|^2) x^* \Omega x = x^* \Gamma Q \Gamma^\mathsf{T} x. \tag{4.64}$$

Since $\Gamma Q \Gamma^\mathsf{T} > 0$ by Assumption K2, the right-hand side of (4.64) is positive. Further, according to Lemma 4.6.6 Ω is positive definite. In order to be well-defined $|\lambda| < 1$ must hold, i.e. all eigenvalues of $(F - KH)$ must be inside the unit circle. □

Lemma 4.6.8

Suppose that the linear system (4.49) is observable. For any positive definite symmetric initial matrix Ω_0*, we have* $\Omega_n > 0$ *and thereby* Ω_{n+1} *is well defined for all* $n \in \mathbb{N}$.

Proof. According to (4.59) we have by Bernstein [9], Proposition 8.1.2., that

$$\Omega_1 = (F - K_0 H)\Omega_0 (F - K_0 H)^\mathsf{T} + \Gamma Q \Gamma^\mathsf{T} > 0.$$

Again, since $\Gamma Q \Gamma^\mathsf{T} > 0$ and $(F - K_{n-1} H)\Omega_{n-1}(F - K_{n-1} H)^\mathsf{T} \geq 0$, we have by induction that $\Omega_n > 0$ and Ω_n is well-defined for all $n \in \mathbb{N}$ for an arbitrary positive symmetric initial matrix Ω_0. □

Lemma 4.6.9 (c.f. Chui and Chen (2009), Lemma 6.4)
Suppose that the linear system (4.49) is observable and Ω defined as in Lemma 4.6.6. Then the following relation

$$\Omega - \Omega_n = (F - KH)(\Omega - \Omega_n)(F - K_{n-1}H)^\intercal \tag{4.65}$$

holds for all $n = 1, 2, \ldots,$ and any positive-definite symmetric initial condition Ω_0.

Proof. By Lemma 4.6.8 all equations in the following proof are well defined. Let Ω_0 be an arbitrary symmetric positive definite initial matrix. Since $K_{n-1} = F\Omega_{n-1}H^\intercal(H\Omega_{n-1}H^\intercal)^{-1}$ and $\Omega_{n-1}^\intercal = \Omega_{n-1}$, the matrix $K_{n-1}H\Omega_{n-1}F^\intercal$ is non-negative definite and symmetric. Using (4.58), we have

$$\begin{aligned}\Omega - \Omega_n &= \Psi(\Omega) - \Psi(\Omega_{n-1})\\ &= (F\Omega F^\intercal - KH\Omega F^\intercal) - (F\Omega_{n-1}F^\intercal - K_{n-1}H\Omega_{n-1}F^\intercal)\\ &= F(\Omega - \Omega_{n-1})F^\intercal - KH\Omega F^\intercal + F\Omega_{n-1}H^\intercal K_{n-1}^\intercal. \end{aligned}\tag{4.66}$$

Now,

$$\begin{aligned}&(F - KH)(\Omega - \Omega_{n-1})(F - K_{n-1}H)^\intercal\\ &= F(\Omega - \Omega_{n-1})F^\intercal - KH\Omega F^\intercal + F\Omega_{n-1}H^\intercal K_{n-1}^\intercal + R_e, \end{aligned}\tag{4.67}$$

where

$$R_e = KH\Omega_{n-1}F^\intercal - F\Omega H^\intercal K_{n-1}^\intercal + KH\Omega H^\intercal K_{n-1}^\intercal - KH\Omega_{n-1}H^\intercal K_{n-1}^\intercal. \tag{4.68}$$

It remains to show that $R_e = 0$, then (4.66) and (4.67) are equal and the result follows. According to the definition of the Kalman gain matrix K_{n-1}, we have

$$K_{n-1}(H\Omega_{n-1}H^\intercal) = F\Omega_{n-1}H^\intercal \tag{4.69}$$

and hence, taking $n \to \infty$ with initial condition $\Omega_0 = \Gamma Q\Gamma^\intercal$ we gain this equation also for the limit

$$KH\Omega H^\intercal = F\Omega H^\intercal. \tag{4.70}$$

Lastly, if we insert (4.69) and (4.70) in (4.68), we have proven that $R_e = 0$. □

If we use (4.65) repeatedly, we obtain

$$\Omega - \Omega_n = (F - KH)^{n-N-1}(\Omega - \Omega_{N+1})B_n^{\mathsf{T}}, \tag{4.71}$$

where

$$B_n = (F - K_{n-1}H)\cdots(F - K_{N+1}H), \tag{4.72}$$

for $n = N+2, N+3, \ldots$ and $B_{n+1} = I_N$.

Lemma 4.6.10 (Chui and Chen (2009), Lemma 6.7)
Suppose that the linear system (4.49) is observable. Then

$$B_n B_n^{\mathsf{T}} \leq M, \quad n \geq N+1, \tag{4.73}$$

for some constant matrix M independent of Ω_0. Consequently, if $B_n = [b_{ij}^{(n)}]$, then it holds that $|b_{ij}^{(n)}| \leq m$, for some constant $m \in \mathbb{R}$ and for all i, j and n.

Proof. According to Lemma 4.6.4, $\Omega_n \leq W$ for $n \geq N+1$. Using (4.59) we have

$$\begin{aligned} W \geq \Omega_n &= (F - K_{n-1}H)\Omega_{n-1}(F - K_{n-1}H)^{\mathsf{T}} + \Gamma Q \Gamma^{\mathsf{T}} \\ &\geq (F - K_{n-1}H)\Omega_{n-1}(F - K_{n-1}H)^{\mathsf{T}} \\ &\geq (F - K_{n-1}H)(F - K_{n-2}H)\Omega_{n-2}(F - K_{n-2}H)^{\mathsf{T}}(F - K_{n-1}H)^{\mathsf{T}} \\ &\geq \ldots \\ &\geq B_n \Omega_{N+1} B_n^{\mathsf{T}}. \end{aligned} \tag{4.74}$$

Since Ω_{N+1} is real, symmetric and by Lemma 4.6.8 positive definite, all its eigenvalues are positive real numbers. Let $\lambda_{min}^{(N+1)}$ be the smallest eigenvalue of Ω_{N+1}, hence according to Bernstein [9], Corollary 8.4.2, we have $\Omega_{N+1} \geq \lambda_{min}^{(N+1)} I_N$. All in all, we obtain

$$W \geq B_n \lambda_{min}^{(N+1)} I_N B_n^{\mathsf{T}} = \lambda_{min}^{(N+1)} B_n B_n^{\mathsf{T}}.$$

Choosing $M := \left(\lambda_{min}^{(N+1)}\right)^{-1} W$ gives us an upper bound. □

The results of Lemma 4.6.4, Lemma 4.6.5 and Lemma 4.6.6 guarantee the convergence of the Kalman filter for the special choice $\Omega_0 = \Gamma Q \Gamma^{\mathsf{T}}$. In addition, Lemma 4.6.7

gives us the asymptotic stability of the Kalman filter. In order to relax the restriction on the initial matrix we use Lemma 4.6.8 to ensure all equations are well-defined.

Eventually, Lemma 4.6.9 and Lemma 4.6.10 give us the convergence for arbitrary symmetric positive definite initial matrices. Finally, we have that B_n is bounded by Lemma 4.6.10 and Lemma 4.6.7 guarantees that the matrix $(F - KH)^{n-N-1}$ converges to zero as $n \to \infty$ and thus $\Omega_n \to \Omega$ as $n \to \infty$.

Theorem 4.6.11 (c.f. Chui and Chen (2009), Theorem 6.1)
Let the linear system (4.49) be observable and Assumption K1, Assumption K2 and Assumption K3 hold. Then, for any initial state X_0 such that $\Omega_0 = \Omega_{0,-1} = Var(X_0)$ is a positive definite and symmetric matrix, $\Omega_n = \Omega_{n,n-1} \to \Omega$ as $n \to \infty$.

Moreover, $\Omega > 0$ is symmetric and exists for any arbitrary initial value X_0. Furthermore, the order of convergence is geometric, that is,

$$tr(\Omega_n - \Omega)(\Omega_n - \Omega)^\mathsf{T} \leq Cr^n, \tag{4.75}$$

where $0 < r < 1$ and $C > 0$, independent of n. Consequently,

$$tr(K_n - K)(K_n - K)^\mathsf{T} \leq Cr^n. \tag{4.76}$$

Proof. In order to abbreviate the notations, we define $\widetilde{F} := (F - KH)$. Using (4.71) and Lemma 4.6.10 gives us

$$\begin{aligned}(\Omega_n - \Omega)(\Omega_n - \Omega)^\mathsf{T} &= \widetilde{F}^{n-N-1}(\Omega_{N+1} - \Omega)B_n B_n^\mathsf{T}(\Omega_{N+1} - \Omega)(\widetilde{F}^{n-N-1})^\mathsf{T} \\ &\leq \widetilde{F}^{n-N-1}\widetilde{M}(\widetilde{F}^{n-N-1})^\mathsf{T},\end{aligned}$$

for some positive definite symmetric constant matrix $\widetilde{M} = (\Omega_{N+1} - \Omega)M(\Omega_{N+1} - \Omega)$. Furthermore, we have $\widetilde{F}^n \to 0$ as $n \to \infty$ because Lemma 4.6.7 guarantees that all eigenvalues lie inside the unit disc. Next, Lemma 4.6.8 gives us that $\Omega > 0$, symmetric and exists for an arbitrary positive definite matrix Ω_0. By Bernstein [9], Fact 8.12.15., and Chui and Chen [26], Lemma 1.10, we have

$$\operatorname{tr}(\Omega_n - \Omega)(\Omega_n - \Omega)^\mathsf{T} \leq \operatorname{tr} \widetilde{F}^{n-N-1}(\widetilde{F}^{n-N-1})^\mathsf{T} \cdot \operatorname{tr} \widetilde{M} \leq Cr^n, \tag{4.77}$$

where $0 < r < 1$ and C is independent of n and depends only on $\widetilde{M}$. Further, we have

$$\begin{aligned}K_n - K =& F(\Omega_n - \Omega)H^\top (H\Omega_n H^\top)^{-1} \\ &+ F\Omega H^\top (H\Omega_n H^\top)^{-1}[H(\Omega - \Omega_n)H^\top](H\Omega H^\top)^{-1}.\end{aligned}$$

And therefore we have

$$\begin{aligned}(K_n - K)(K_n - K)^\top \leq& 2[F(\Omega_n - \Omega)H^\top (H\Omega_n H^\top)^{-1}][F(\Omega_n - \Omega)H^\top (H\Omega_n H^\top)^{-1}]^\top \\ &+ 2[F\Omega H^\top (H\Omega_n H^\top)^{-1}[H(\Omega - \Omega_n)H^\top](H\Omega H^\top)^{-1}] \\ &\cdot [F\Omega H^\top (H\Omega_n H^\top)^{-1}[H(\Omega - \Omega_n)H^\top](H\Omega H^\top)^{-1}]^\top.\end{aligned}$$

Additionally, we have

$$\Omega_1 = (F - K_0 H)\Omega_0 (F - K_0 H)^\top + \Gamma Q \Gamma^\top \geq \Gamma Q \Gamma^\top$$

and analogous $\Omega_n \geq \Gamma Q \Gamma^\top$ for $n \geq 1$. As a consequence, we have

$$H\Omega_n H^\top \geq H\Gamma Q \Gamma^\top H^\top$$

and

$$(H\Omega_n H^\top)^{-1} \leq (H\Gamma Q \Gamma^\top H^\top)^{-1}.$$

According to Chui and Chen [26], Lemma 1.9,

$$\operatorname{tr}\left((H\Omega_n H^\top)^{-1}(H\Omega_n H^\top)^{-1}\right) \leq \left(\operatorname{tr}(H\Gamma Q \Gamma^\top H^\top)^{-1}\right)^2$$

and again as in (4.77) we get

$$\begin{aligned}\operatorname{tr}(K_n - K)(K_n - K)^\top \leq& 2\operatorname{tr}(\Omega_n - \Omega)(\Omega_n - \Omega)^\top \operatorname{tr} FF^\top \operatorname{tr} H^\top H (\operatorname{tr}(H\Gamma Q \Gamma^\top H^\top)^{-1})^2 \\ &+ 2\operatorname{tr}\Omega\Omega^\top \operatorname{tr} FF^\top \operatorname{tr} H^\top H (\operatorname{tr}(H\Gamma Q\Gamma^\top H^\top)^{-1})^2 \operatorname{tr} HH^\top \\ &\cdot \operatorname{tr}(\Omega - \Omega_n)(\Omega - \Omega_n)^\top \operatorname{tr} H^\top H \operatorname{tr}(H\Omega H^\top)^{-1}(H\Omega H^\top)^{-1} \\ \leq& C_1 \operatorname{tr}(\Omega_n - \Omega)(\Omega_n - \Omega)^\top \\ \leq& Cr^n,\end{aligned}$$

where C_1 and C are constants, independent of n. □

We obtain the steady-state (limiting) Kalman filter by replacing the Kalman gain matrix K_n by its limit version K, therefore the prediction-correction equations are then given by

$$
\begin{cases}
\overline{X}_{0|-1} & = \mathbb{E}[X_0] \\
\overline{X}_{n+1|n} & = (F-KH)\overline{X}_{n|n-1} + KY_n = F\overline{X}_{n|n} \\
\overline{X}_{n|n} & = \overline{X}_{n|n-1} + \Omega H^\mathsf{T}(H\Omega H^\mathsf{T})^{-1}(Y_n - H\overline{X}_{n|n-1}) \\
K & = F\Omega H^\mathsf{T}(H\Omega H^\mathsf{T})^{-1}.
\end{cases} \tag{4.78}
$$

Theorem 4.6.12 (c.f. Chui and Chen (2009), Theorem 6.2)
Let the linear system (4.49) be observable and Assumption K1, Assumption K2 and Assumption K3 hold. Then

$$
\lim_{n\to\infty} \|X_n - \overline{X}_{n,n-1}\|_N^2 = \Omega = \lim_{n\to\infty} \|X_n - \widehat{X}_{n,n-1}\|_N^2 \tag{4.79a}
$$

and

$$
\lim_{n\to\infty} \|X_n - \overline{X}_{n,n}\|_N^2 = \Omega - \Omega H^\mathsf{T}(H\Omega H^\mathsf{T})^{-1}H\Omega = \lim_{n\to\infty} \|X_n - \widehat{X}_{n,n}\|_N^2. \tag{4.79b}
$$

Proof. The right hand side of (4.79a) is obvious and the right hand side of (4.79b) follows directly from the definition.

$$
\begin{aligned}
\lim_{n\to\infty} \|X_n - \widehat{X}_n\|_N^2 &= \lim_{n\to\infty} \Omega_{n,n} \\
&= \lim_{n\to\infty} \Omega_{n,n-1} - \Omega_{n,n-1}H^\mathsf{T}(H\Omega_{n,n-1}H^\mathsf{T})^{-1}H\Omega_{n,n-1} \\
&= \Omega - \Omega H^\mathsf{T}(H\Omega H^\mathsf{T})^{-1}H\Omega.
\end{aligned}
$$

For the left hand side of (4.79a) we need first some calculations. We have

$$
\begin{aligned}
X_n - \overline{X}_{n,n-1} &= (FX_{n-1} + \Gamma Z_{n-1}) - ((F-KH)\overline{X}_{n-1,n-2} + KY_{n-1}) \\
&= (F-KH)X_{n-1} + \Gamma Z_{n-1} - (F-KH)\overline{X}_{n-1,n-2} \\
&= (F-KH)(X_{n-1} - \overline{X}_{n-1,n-2}) + \Gamma Z_{n-1}.
\end{aligned}
$$

Hereby, we obtain with the independence

$$
\|X_n - \overline{X}_{n,n-1}\|_N^2 = (F-KH)\|X_{n-1} - \overline{X}_{n-1,n-2}\|_N^2(F-KH)^\mathsf{T} + \Gamma Q\Gamma^\mathsf{T}.
$$

Subtracting Ω from the previous result together with (4.61) gives us

$$
\|X_n - \overline{X}_{n,n-1}\|_N^2 - \Omega = (F-KH)\|X_{n-1} - \overline{X}_{n-1,n-2}\|_N^2(F-KH)^\mathsf{T} + \Gamma Q\Gamma^\mathsf{T} - \Omega
$$

$$=(F-KH)\left(\|X_{n-1}-\overline{X}_{n-1,n-2}\|_N^2-\Omega\right)(F-KH)^{\mathsf{T}}.$$

Iterating this procedure leads to

$$\|X_n-\overline{X}_{n,n-1}\|_N^2-\Omega=(F-KH)^n\left(\|X_0-\overline{X}_{0,-1}\|_N^2-\Omega\right)[(F-KH)^{\mathsf{T}}]^n$$

Since the eigenvalues of the matrix $(F-KH)$ lie in the unit circle, the matrix $(F-KH)^n$ tends to zero for $n\to\infty$ and finally we have

$$\lim_{n\to\infty}\|X_n-\overline{X}_{n,n-1}\|_N^2=\Omega. \tag{4.80}$$

For the left hand side (4.79b) we have by (4.78)

$$\begin{aligned}X_n-\overline{X}_{n,n}&=(X_n-\overline{X}_{n,n-1})-\Omega H^{\mathsf{T}}(H\Omega H^{\mathsf{T}})^{-1}(Y_n-H\overline{X}_{n|n-1})\\&=\left(I_N-\Omega H^{\mathsf{T}}(H\Omega H^{\mathsf{T}})^{-1}H\right)(X_n-\overline{X}_{n|n-1}).\end{aligned}$$

and consequently

$$\begin{aligned}\|X_n-\overline{X}_{n,n}\|_N^2=&\left(I_N-\Omega H^{\mathsf{T}}(H\Omega H^{\mathsf{T}})^{-1}H\right)\|X_n-\overline{X}_{n|n-1}\|_N^2\\&\cdot\left(I_N-\Omega H^{\mathsf{T}}(H\Omega H^{\mathsf{T}})^{-1}H\right)^{\mathsf{T}}\end{aligned}$$

and by taking the limit using the result (4.80).

$$\begin{aligned}&\lim_{n\to\infty}\|X_n-\overline{X}_{n,n}\|_N^2=\left(I_N-\Omega H^{\mathsf{T}}(H\Omega H^{\mathsf{T}})^{-1}H\right)\Omega\left(I_N-\Omega H^{\mathsf{T}}(H\Omega H^{\mathsf{T}})^{-1}H\right)^{\mathsf{T}}\\&=\Omega+\Omega H^{\mathsf{T}}(H\Omega H^{\mathsf{T}})^{-1}H\Omega H^{\mathsf{T}}(H\Omega H^{\mathsf{T}})^{-1}\Omega H-2\Omega H^{\mathsf{T}}(H\Omega H^{\mathsf{T}})^{-1}\Omega H\\&=\Omega-\Omega H^{\mathsf{T}}(H\Omega H^{\mathsf{T}})^{-1}\Omega H,\end{aligned}$$

which completes the proof of this theorem. □

Theorem 4.6.13 (c.f. Chui and Chen (2009), Theorem 6.3)
Let the linear system (4.49) be observable and Assumption K1, Assumption K2 and Assumption K3 hold.

Then there exist a real number r, with $0<r<1$ and a positive constant C, independent of n, such that

$$tr\|\widehat{X}_n-\overline{X}_n\|_N^2\le Cr^n. \tag{4.81}$$

Proof. The proof is an adapted version of Chui and Chen [26], where $R\equiv 0_{d\times d}$. Let G_n and G be defined by $G_n=\Omega_nH^{\mathsf{T}}(H\Omega_nH^{\mathsf{T}})^{-1}$ and $G:=\Omega H^{\mathsf{T}}(H\Omega H^{\mathsf{T}})^{-1}$ then we have

$$\begin{aligned}\widehat{X}_n - \overline{X}_n = &(I_N - GH)F(\widehat{X}_{n-1} - \overline{X}_{n-1})\\ &+ (G_n - G)(HF(X_{n-1} - \widehat{X}_{n-1}) + H\Gamma Z_{n-1}).\end{aligned}$$

Since $\langle Z_{n-1}, \widehat{X}_n - \overline{X}_n\rangle = 0$ and $\langle Z_{n-1}, X_n - \widehat{X}_n\rangle = 0$, we have by writing shortly $\tilde{A} := (I_N - GH)F$ that

$$\begin{aligned}\|\widehat{X}_n - \overline{X}_n\|_N^2 =&\tilde{A}\|\widehat{X}_{n-1} - \overline{X}_{n-1}\|_N^2\tilde{A}^\mathsf{T}\\ &+ (G_n - G)(HF\|X_{n-1} - \widehat{X}_{n-1}\|_N^2 F^\mathsf{T}H^\mathsf{T} + H\Gamma Q\Gamma^\mathsf{T}H^\mathsf{T})(G_n - G)^\mathsf{T}\\ &+ \tilde{A}\langle \widehat{X}_{n-1} - \overline{X}_{n-1}, X_{n-1} - \widehat{X}_{n-1}\rangle F^\mathsf{T}H^\mathsf{T}(G_n - G)^\mathsf{T}\\ &+ (G_n - G)HF\langle X_{n-1} - \widehat{X}_{n-1}, \widehat{X}_{n-1} - \overline{X}_{n-1}\rangle\tilde{A}^\mathsf{T}\end{aligned}$$

and repeating this $n-1$-times we obtain

$$\begin{aligned}&\|\widehat{X}_n - \overline{X}_n\|_N^2\\ &= \tilde{A}^n\|\widehat{X}_0 - \overline{X}_0\|_N^2(\tilde{A}^n)^\mathsf{T}\\ &\quad+ \sum_{i=0}^{n-1}\tilde{A}^i(G_{n-i} - G)HF\|X_{n-1-i} - \widehat{X}_{n-1-i}\|_N^2F^\mathsf{T}H^\mathsf{T}(G_{n-i} - G)^\mathsf{T}(\tilde{A}^i)^\mathsf{T}\\ &\quad+ \sum_{i=0}^{n-1}\tilde{A}^i(G_{n-i} - G)H\Gamma Q\Gamma^\mathsf{T}H^\mathsf{T}(G_{n-i} - G)^\mathsf{T}(\tilde{A}^i)^\mathsf{T}\\ &\quad+ \sum_{i=0}^{n-1}\tilde{A}^{i+1}\langle\widehat{X}_{n-1-i} - \overline{X}_{n-1-i}, X_{n-1-i} - \widehat{X}_{n-1-i}\rangle F^\mathsf{T}H^\mathsf{T}(G_{n-i} - G)^\mathsf{T}\tilde{A}^i\\ &\quad+ \sum_{i=0}^{n-1}\tilde{A}^i(G_{n-i} - G)HF\langle X_{n-1-i} - \widehat{X}_{n-1-i}, \widehat{X}_{n-1-i} - \overline{X}_{n-1-i}\rangle(\tilde{A}^{i+1})^\mathsf{T}.\end{aligned}$$

Furthermore, we have

$$\begin{aligned}0 &\le \langle\widehat{X}_j - \overline{X}_j - (X_j - \widehat{X}_j), \widehat{X}_j - \overline{X}_j - (X_j - \widehat{X}_j)\rangle\\ &= \langle\widehat{X}_j - \overline{X}_j, \widehat{X}_j - \overline{X}_j\rangle - \langle\widehat{X}_j - \overline{X}_j, X_j - \widehat{X}_j\rangle\\ &\quad - \langle X_j - \widehat{X}_j, \widehat{X}_j - \overline{X}_j\rangle + \langle X_j - \widehat{X}_j, X_j - \widehat{X}_j\rangle\end{aligned}$$

and thus with Theorem 4.6.12, we obtain

$$\begin{aligned}&\langle\widehat{X}_j - \overline{X}_j, X_j - \widehat{X}_j\rangle + \langle X_j - \widehat{X}_j, \widehat{X}_j - \overline{X}_j\rangle\\ &\le \langle\widehat{X}_j - \overline{X}_j, \widehat{X}_j - \overline{X}_j\rangle + \langle X_j - \widehat{X}_j, X_j - \widehat{X}_j\rangle\end{aligned}$$

$$
\begin{aligned}
&= 2\|X_j - \widehat{X}_j\|_N^2 + \|X_j - \overline{X}_j\|_N^2 + \langle X_j - \overline{X}_j, \widehat{X}_j - X_j\rangle + \langle \widehat{X}_j - X_j, X_j - \overline{X}_j\rangle \\
&\leq 3\|X_j - \widehat{X}_j\|_N^2 + 2\|X_j - \overline{X}_j\|_N^2 \xrightarrow{j\to\infty} 5(\Omega - \Omega H^\mathsf{T}(H\Omega H^\mathsf{T})^{-1}H\Omega).
\end{aligned}
$$

Therefore, it follows that the terms $\langle \widehat{X}_j - \overline{X}_j, X_j - \widehat{X}_j\rangle F^\mathsf{T} H^\mathsf{T}$ are componentwise uniformly bounded.

Thus, we obtain Chui and Chen [26] Lemma 1.6, 1.7 and 1.10 as well as Theorem 4.6.11

$$
\begin{aligned}
\operatorname{tr}\Big(&\tilde{A}\langle \widehat{X}_{n-1-i} - \overline{X}_{n-1-i}, X_{n-1-i} - \widehat{X}_{n-1-i}\rangle F^\mathsf{T} H^\mathsf{T}(G_{n-i} - G)^\mathsf{T} \\
&+(G_{n-i} - G)HF\langle X_{n-1-i} - \widehat{X}_{n-1-i}, \widehat{X}_{n-1-i} - \overline{X}_{n-1-i}\rangle \tilde{A}\Big) \leq C_1 r_1^{n-i+1},
\end{aligned}
$$

for some $r_1, 0 < r_1 < 1$ and some positive constant C_1, which is independent of n and i. In the same manner it holds that

$$
\begin{aligned}
\operatorname{tr}\|\widehat{X}_n - \overline{X}_n\|_N^2 \leq& \operatorname{tr}\|\widehat{X}_0 - \overline{X}_0\|_N^2 C_2 r_2^n + \sum_{i=0}^{n-1} C_3 r_3^i C_4 r_4^{n-i} + \sum_{i=0}^{n-1} C_5 r_5^i C_1 r1^{n-i+1} \\
\leq& p(n) r_6^n,
\end{aligned}
$$

where $0 < r_2, r_3, r_4, r_5 < 1$, $r_6 = \max(r_1, r_2, r_3, r_4, r_5) < 1$, C_2, C_3, C_4, C_5 are positive constants independent of i and n, and $p(n)$ is a polynomial of n. Hence, there exists a real number r, such that $r_6 < t < 1$ and a positive constant C independent of n satisfying $p(n)\left(\frac{r_6}{r}\right)^n \leq C$, where

$$
\operatorname{tr}\|\widehat{X}_n - \overline{X}_n\|_N^2 \leq Cr^n.
$$

This completes the proof. □

Finally, we have checked that the Kalman filter is applicable for the cointegrated model, i.e. a model with unit roots. Hence, we can make use of its equations and properties in Chapter 5.

CHAPTER 5

Asymptotic Inference of Cointegrated Lévy Driven MCARMA Models

5.1. Introduction

This chapter forms the main part of this thesis and deals with the statistical inference of the cointegrated Lévy driven MCARMA model. The step-wise quasi-maximum likelihood estimator estimates the model parameters and thus also the cointegration space using equidistant observations in discrete time. The estimation method works not only for Gaussian MCARMA models, but also for more general Lévy driven models.

A famous method for estimating the parameters of a cointegrated VAR(p) model is the approach presented by Johansen [53] and [54] using a method of reduced rank regression. This method estimates the cointegration vectors, however, the parametrization of the cointegration vectors is unrestricted. Furthermore, the Johansen test for the cointegration rank is a direct implication of this estimation procedure. However, the cointegration vectors are not unique and one gets nested models. Applying such a method to the sampled cointegrated MCARMA model seems not applicable since several assumptions of this procedure are not satisfied. For example we have an infinite order error correction form. Moreover, identifiability would be an issue. The estimation of cointegrated VARMA models in echelon form was presented by Lütkepohl and Poskitt [64], where the identifiability problem is solved due to the unique parametrization.

Special cases of cointegrated MCARMA models were considered by Fasen [34], where a multiple regression model is treated with multivariate Ornstein-Uhlenbeck processes as integrated processes, and in Fasen [35], where the noise term is an MCARMA process embedded in the cointegrated model observed at a high-frequency time grid. Besides, statistical inference and identification of an ergodic continuously observed Gaussian MCAR(1) process were considered in Kessler and Rahbek [57] and observations at discrete time points in Kessler and Rahbek [58]. Kessler and Rahbek solve for this simple cointegrated MCARMA model the identification and aliasing problem and derive the asymptotic distribution of the co-integration parameters.

There exists several results for quasi-maximum likelihood estimation of stationary processes. There is a connection of some of these results to our estimation method for cointegrated MCARMA processes. Quasi-maximum likelihood estimation for strongly mixing stationary ARMA processes is considered in Francq and Zakoïan [39], Boubacar and Maïnassara [13] present an estimation procedure for weak VARMA processes, however, they use the strong mixing assumption on the linear innovations. As Schlemm and Stelzer [90] showed, this assumption is very difficult to verify in the stationary MCARMA setting. However, in another work by Schlemm and Stelzer [91] they show a quasi-maximum likelihood estimation, where the strongly mixing assumption is made for the stationary process itself. This assumption is always satisfied for stationary MCARMA processes due to a result by Marquardt and Stelzer [69]. We capitalize on this result since the cointegrated MCARMA process is the sum of a stationary MCARMA process, for which the strong mixing assumptions hold, and a Lévy process. Since we use a step-wise estimation method, main ideas of the estimation procedure for stationary MCARMA processes in Schlemm and Stelzer [91] are used and adapted to our setting.

The idea of a step-wise estimation approach for integrated and cointegrated models dates back to Saikkonen [85] and [86]. We customize this idea to the cointegrated MCARMA model. The ideas of Saikkonen were also employed in the work of Bauer and Wagner [7], who consider a cointegrated state space model in the discrete-time framework. We state the log-likelihood function in Section 5.2 and separate the log-likelihood function in the same way as the parameter space. To be more precise, we split the parameter vector into two parts containing on the one hand the long-run parameter and on the other hand the short-run parameter. Accordingly, the log-likelihood function is separated in one part depending on all parameters and the other only on the short-run parameters. This is the key idea to the step-wise approach.

As we need some kind of uniform convergence, we derive in Section 5.4 such results using the closely related continuous weak convergence. The results of Section 5.4 are based on the asymptotic results derived in Appendix 5.8. However, we use the concepts of continuous convergence in combination with a stochastic equicontinuity condition in order to deal with the different rates of convergence occurring due to the non-stationary setting.

A major problem arising in the estimation of multivariate continuous-time models is the identifiability problem. The models can not only have many redundancies due to a high dimension, but also they may be indistinguishable in the sampling procedure. We overcome these problems in Section 5.3 and find sufficient conditions on the parametrization in order to have a unique parameterizations, which is identifiable from the discrete time observations. Furthermore, we immediately obtain a unique basis of the cointegration space as well.

We prove in Section 5.5 the consistency of the quasi-maximum likelihood estimators. Because we have different rates of convergence, we also have different orders of consistency. The long-run parameter estimator is super-consistent, i.e. the estimator converges in probability to the true value proportional to the inverse of the sample size. Hence, it converges at a faster rate than the classical $\sqrt{n}$ rate for stationary estimators. We cannot show consistency of the short-run quasi-maximum likelihood estimator without the knowledge of the consistency rate of the long-run quasi-maximum likelihood estimator. Hence, we show the consistency result in three steps:

1. Consistency of the long-run parameter estimator.
2. Order of consistency of the long-run parameter estimator.
3. Consistency of the short-run parameter estimator.

In the end, we derive the asymptotic distributions of the parameter estimators in Section 5.6. The short-run parameter estimator is asymptotically normal, whereas the long-run parameter estimator is asymptotically mixed normal. In order to prove these result, we use a classical Taylor series expansion of the score vector. Thus, we derive the asymptotic behavior of the score vector. Furthermore, we show that the Hessian matrix converges to a block diagonal matrix which is positive definite. Finally, we employ the continuous convergence and stochastic equicontinuity results from Section 5.4 for the Taylor series expansion and thus derive the asymptotic distribution of the estimators using the previous results of this chapters.

5.2. Quasi-Maximum Likelihood Estimation

We estimate the model parameters via an adapted quasi-maximum likelihood estimation method. In difference to the classical quasi-maximum likelihood approach, we use a step-wise version. We state in this section most of the assumptions needed for the estimation procedure and derive some properties of the sampled system which are inherited from the assumptions on the continuous-time model. In particular, we show a separation of the parameter space and accordingly for the log-likelihood function with respect to the long-run and short-run parameters. This separation is the key idea to prove later on the asymptotic normality and consistency of the estimators.

Before we start with the estimation procedure we briefly recall the definition and the representation of a cointegrated MCARMA process. We refer to Definition 4.3.10 for a profound definition of a cointegrated MCARMA process.

A cointegrated continuous-time linear state space model (A, B, C, L) driven by a Lévy process L is given by the state equation

$$\mathrm{d}X(t) = AX(t)\mathrm{d}t + B\mathrm{d}L(t)$$

and the observation equation

$$Y(t) = CX(t),$$

for $t \geq 0$, where $A \in \mathbb{R}^{N\times N}$, $B \in \mathbb{R}^{N\times m}$ and $C \in \mathbb{R}^{d\times N}$. The spectrum A satisfies $\sigma(A) \subset \{(-\infty, 0) + i\mathbb{R}\} \cup \{0\}$ and the algebraic and geometric multiplicity of the eigenvalue zero is $c \leq d$. Assume that the m-dimensional Lévy process L has mean zero, satisfies $\mathbb{E}\|L(1)\|^2 < \infty$ and has a non-singular covariance matrix $\Sigma^L = \mathbb{E}L(1)L(1)^\mathsf{T}$. The solution $(Y(t))_{t\geq 0}$ is then given by

$$Y(t) = C\exp(At)X(0) + C\int_0^t \exp(A(t-u))B\,\mathrm{d}L(u)$$

for $t \geq 0$. A minimal state space system (A, B, C, L) has due to Theorem 4.3.7 the representation

$$Y(t) = C_1X_1(0) + C_1B_1L(t) + C_2\int_{-\infty}^t \exp(A_2(t-u))B_2\,\mathrm{d}L(u).$$

The stationary part is abbreviated by

$$Y_2(t) := C_2\int_{-\infty}^t \exp(A_2(t-u))B_2\,\mathrm{d}L(u).$$

Since we want to estimate the model parameters of the continuous-time model from discrete time observations, we are interested in an representation of the process observed at equidistant time points. The sampled version of this process is given, as in Lemma 4.3.14, by the state space representation with state equation

$$\begin{pmatrix} X_{n,1}^{(h)} \\ X_{n,2}^{(h)} \end{pmatrix} = \begin{pmatrix} X_{n-1,1}^{(h)} \\ \mathrm{e}^{A_2 h} X_{n-1,2}^{(h)} \end{pmatrix} + \begin{pmatrix} R_{n,1}^{(h)} \\ R_{n,2}^{(h)} \end{pmatrix},$$

and the observation equation is given by

$$Y_n^{(h)} = C_1 X_{n,1}^{(h)} + C_2 X_{n,2}^{(h)}, \quad n \in \mathbb{N}. \tag{5.1}$$

The noise sequence $(R_n^{(h)})_{n\in\mathbb{N}}$ is an i.i.d. sequence given by

$$R_n^{(h)} = \begin{pmatrix} R_{n,1}^{(h)} \\ R_{n,2}^{(h)} \end{pmatrix} = \begin{pmatrix} B_1 \left(L(nh) - L((n-1)h)\right) \\ \int_{(n-1)h}^{nh} \mathrm{e}^{A_2(nh-u)} B_2 \, \mathrm{d}L(u) \end{pmatrix}, \quad n \in \mathbb{N}$$

with mean zero and covariance matrix

$$\widetilde{\Sigma}^{(h)} = \mathbb{E} R_n^{(h)} R_n^{(h)\mathsf{T}} = \int_0^h \begin{pmatrix} B_1 \Sigma_L B_1^{\mathsf{T}} & \mathrm{e}^{A_2 u} B_2 \Sigma_L B_1^{\mathsf{T}} \\ B_1 \Sigma_L B_2^{\mathsf{T}} \mathrm{e}^{A_2^{\mathsf{T}} u} & \mathrm{e}^{A_2 u} B_2 \Sigma_L B_2^{\mathsf{T}} \mathrm{e}^{A_2^{\mathsf{T}} u} \end{pmatrix} \mathrm{d}u. \tag{5.2}$$

For some parameter space $\Theta \subset \mathbb{R}^s$, $s \in \mathbb{N}$, we have for each $\vartheta \in \Theta$ matrices $A_\vartheta \in M_N(\mathbb{R})$, $B_\vartheta \in M_{N,m}(\mathbb{R})$, $C_\vartheta \in M_{d,N}(\mathbb{R})$ and a Lévy process L_ϑ. Further, we denote the true parameter vector with ϑ^0. Let us now state some standard assumptions.

Assumption M1

Assume that the cointegrated MCARMA process is driven by a Lévy process L_ϑ with mean zero and non-singular covariance matrix $\Sigma_\vartheta^L = \mathbb{E} L_\vartheta(1) L_\vartheta(1)^{\mathsf{T}}$. Assume further that there exists a $\delta > 0$ such that $\mathbb{E}\|L_\vartheta(1)\|^{4+\delta} < \infty$.

Assumption M2

Assume that the matrix A_ϑ has c eigenvalues equal to zero and the remaining eigenvalues have strictly negative real parts for all $\vartheta \in \Theta$. Moreover, the matrix C_ϑ has full rank for all $\vartheta \in \Theta$.

At this point the cointegration rank needs to be known somehow beforehand. With this knowledge we know the dimensions of the subsystems and are therefore able to estimate the model adequately. In reality, it is necessary to estimate first the cointegration rank r and then proceed as in the subsequent considerations.

Assumption M3

The triplet $(A_\vartheta, B_\vartheta, C_\vartheta)$ is minimal for all $\vartheta \in \Theta$ with McMillan degree N.

Under Assumption M1 - Assumption M3 we also have the canonical form given in Theorem 4.3.7. Hence, we can consider the decoupled matrices of the subsystems $(A_{2,\vartheta}, B_{1,\vartheta}, B_{2,\vartheta}, C_{1,\vartheta}, C_{2,\vartheta}, L_\vartheta)$ instead of $(A_\vartheta, B_\vartheta, C_\vartheta, L_\vartheta)$, where the matrices have the following dimensions $A_{2,\vartheta} \in M_{N-c}(\mathbb{R})$, $B_{1,\vartheta} \in M_{c,m}(\mathbb{R})$, $B_{2,\vartheta} \in M_{d-c,m}(\mathbb{R})$, $C_{1,\vartheta} \in M_{d,c}(\mathbb{R})$ and $C_{2,\vartheta} \in M_{d,N-c}(\mathbb{R})$ for all $\vartheta \in \Theta$. Additionally, we need to impose the following rank condition on the parametrization.

Assumption M4

The matrices $B_{1,\vartheta}$ and $C_{1,\vartheta}$ have full rank c for all $\vartheta \in \Theta$.

As in Proposition 4.4.2 we have due to the properties of the Kalman filter the following matrices depending now on the parameter vector ϑ:
The unique solution $\Omega_\vartheta^{(h)}$ of the discrete-time algebraic Riccati equation

$$\Omega_\vartheta^{(h)} = \mathrm{e}^{A_\vartheta h}\Omega_\vartheta^{(h)}\mathrm{e}^{A_\vartheta^\mathsf{T} h} - \mathrm{e}^{A_\vartheta h}\Omega_\vartheta^{(h)}C_\vartheta^\mathsf{T}\left(C_\vartheta\Omega_\vartheta^{(h)}C_\vartheta^\mathsf{T}\right)^{-1}C_\vartheta\Omega_\vartheta^{(h)}\mathrm{e}^{A_\vartheta^\mathsf{T} h} + \widetilde{\Sigma}_\vartheta^{(h)}$$

is used to calculate the steady-state Kalman gain matrix with respect to the parameter $\vartheta \in \Theta$ given by $K_\vartheta^{(h)} = \mathrm{e}^{A_\vartheta h}\Omega_\vartheta^{(h)}C_\vartheta^\mathsf{T}\left(C_\vartheta\Omega_\vartheta^{(h)}C_\vartheta^\mathsf{T}\right)^{-1}$ and the prediction covariance matrix of the Kalman filter $V_\vartheta^{(h)} = C_\vartheta\Omega_\vartheta^{(h)}C_\vartheta^\mathsf{T}$.

Recall the sampled version of a continuous cointegrated MCARMA process as in (4.29). The class of continuous-time cointegrated state space models $(A_\vartheta, B_\vartheta, C_\vartheta, L_\vartheta)$, for $\vartheta \in \Theta$, is mapped at sampling distance h to the discrete-time state space models $(\mathrm{e}^{A_\vartheta h}, C_\vartheta, R_\vartheta^{(h)})$ for $\vartheta \in \Theta$. The i.i.d. noise sequence $R_\vartheta^{(h)}$ is given by

$$R_{k,\vartheta}^{(h)} = \begin{pmatrix} B_{1,\vartheta}\left(L_\vartheta(kh) - L_\vartheta((k-1)h)\right) \\ \int_{(k-1)h}^{kh} \mathrm{e}^{A_{2,\vartheta}(nh-u)}B_{2,\vartheta}\,\mathrm{d}L_\vartheta(u) \end{pmatrix} = \begin{pmatrix} R_{k,1,\vartheta}^{(h)} \\ R_{k,2,\vartheta}^{(h)} \end{pmatrix}, \quad k \in \mathbb{N}. \tag{5.3}$$

This state space model is not in innovation form and hence we use the result from the previous subsection to calculate the pseudo-innovations of the observations $(Y_1^{(h)}, \ldots, Y_n^{(h)})$ of the output process $Y^{(h)}$. Hence, the pseudo-innovations are given by

$$\varepsilon_k^{(h)}(\vartheta) = \left(I_d - C_\vartheta[I_N - (\mathrm{e}^{A_\vartheta h} - K_\vartheta^{(h)}C_\vartheta)B]^{-1}K_\vartheta^{(h)}B\right)Y_k^{(h)}, \quad k \in \mathbb{N} \tag{5.4}$$

and they can also be represented in the error correction form

$$\varepsilon_k^{(h)}(\vartheta) = -\Pi(\vartheta)Y_{k-1}^{(h)} + \overline{k}(B,\vartheta)\Delta Y_k^{(h)}, \quad k \in \mathbb{N} \tag{5.5}$$

where $\Pi(\vartheta) := \alpha(\vartheta)\beta^{\mathsf{T}}(\vartheta)$ and the transfer function $\overline{k}(z,\vartheta) := I_d - \overline{k}(z,\vartheta)$ is given similarly as in (4.47). We omit the notation of the sampling distance h in the denotation of $\Pi(\vartheta)$ and $\overline{k}(z,\vartheta)$ to save notation. Recall that the matrix coefficients of $\widetilde{k}(z,\vartheta)$ are given as in Lemma 4.4.7.

Minus two over n times the logarithm of the pseudo-Gaussian likelihood function, denoted with $\mathcal{L}_n$, is given by

$$\begin{aligned} \mathcal{L}_n^{(h)}(\vartheta) &:= \frac{1}{n}\sum_{k=1}^{n} \ell_n^{(h)}(\vartheta) \\ &:= \frac{1}{n}\sum_{k=1}^{n} \left[d\log 2\pi + \log\det V_\vartheta^{(h)} + \varepsilon_k^{(h)}(\vartheta)^{\mathsf{T}} \left(V_\vartheta^{(h)}\right)^{-1} \varepsilon_k^{(h)}(\vartheta)\right]. \end{aligned} \tag{5.6}$$

The quasi-maximum likelihood estimator $\widehat{\vartheta}_n$ can be obtained by minimizing the Gaussian log-likelihood function $\mathcal{L}_n^{(h)}(\vartheta)$. Hence, the estimator based on the sample $(Y_1^{(h)},\ldots,Y_n^{(h)})$ is given by

$$\widehat{\vartheta}_n = \operatorname{argmin}_{\theta\in\Theta}\mathcal{L}_n^{(h)}(\vartheta). \tag{5.7}$$

An alternative estimation method is the least squares estimation, where we calculate the sum of squares function given by

$$\mathcal{Q}_n^{(h)}(\vartheta) = \frac{1}{n}\sum_{k=1}^{n} \varepsilon_k^{(h)}(\vartheta)^{\mathsf{T}}\varepsilon_k^{(h)}(\vartheta).$$

However, in this thesis we use the quasi-maximum likelihood approach and prove the results for this approach. All results immediately hold then for the least squares estimation as well.

We need some further assumption concerning the parametrization in order to be able to estimate the model parameters via quasi-maximum likelihood estimation.

Assumption M5
The parameter space Θ is a compact subset of $\mathbb{R}^s$.

Assumption M6
The mappings $\vartheta \mapsto A_{2,\vartheta}$, $\vartheta \mapsto B_{i,\vartheta}$, $\vartheta \mapsto C_{i,\vartheta}$ for $i \in \{1,2\}$ and $\vartheta \mapsto \Sigma_\vartheta^L$ are continuous.

As a consequence, all matrix functions arising in the estimation procedure are continuous, c.f. subsection 5.9.1. Additionally, we need a representation of the pseudo-innovations and the behavior of the related matrix coefficients. The next lemma shows that the coefficients decay exponentially fast independently of the parameter vector we use.

Lemma 5.2.1

Assume that Assumption M1-Assumption M6 hold. Then the pseudo-innovations sequence $\varepsilon^{(h)}(\vartheta)$ as given in equation (5.5) has the following property:
The pseudo-innovations $\varepsilon^{(h)}(\vartheta)$ are linear functions of $Y^{(h)}$, i.e. there exist matrix sequences $(\widetilde{K}_i(\vartheta))_{i\in\mathbb{N}}$, such that

$$\varepsilon_k^{(h)}(\vartheta) = -\Pi(\vartheta)Y_{k-1}^{(h)} + \left(\Delta Y_k^{(h)} - \sum_{i=1}^{\infty} \widetilde{K}_i(\vartheta)\Delta Y_{k-i}^{(h)}\right) \quad k \in \mathbb{N}. \tag{5.8}$$

The matrices $\widetilde{K}_i(\vartheta)$ are uniformly exponentially bounded, i.e. there exist a positive constants c and $\rho < 1$, such that $\sup_{\vartheta\in\Theta} \|\widetilde{K}_i(\vartheta)\| \leq c\rho^i$, $i \in \mathbb{N}$.

Proof. The proof follows in the same line as Lemma 2.6 in Schlemm and Stelzer [91], using equation (5.5) and Lemma 5.9.2 ii). □

We can split up the innovation sequence analogously to Saikkonen [85],[86]. For this we separate the s-dimensional parameter vector $\vartheta = (\vartheta_1^\mathsf{T}, \vartheta_2^\mathsf{T})^\mathsf{T}$, where ϑ_1 denotes the s_1-dimensional vector of long-run parameters, i.e. the parameters corresponding to the non-stationary part and ϑ_2 the s_2-dimensional vector of short-run parameters corresponding to the stationary part.

Therefore, we write the parameter space as the product space of the sub-spaces $\Theta = \Theta_1 \times \Theta_2$ with $\Theta_1 \subset \mathbb{R}^{s_1}$ and $\Theta_2 \subset \mathbb{R}^{s_2}$ and parameterize the matrices with the following sub-vectors

$$(A_{2,\vartheta_2}, B_{1,\vartheta_2}, B_{2,\vartheta_2}, C_{1,\vartheta_1}, C_{2,\vartheta_2}, L_{\vartheta_2}), \qquad \text{for } \vartheta_1 \in \Theta_1,\ \vartheta_2 \in \Theta_2. \tag{5.9}$$

We know from our previous considerations that the matrix $C_{1,\vartheta}$ is responsible for the cointegration property. Hence, we use only for this matrix the sub-vector ϑ_1 for the long-run parameters. For all the other matrices we use the sub-vector ϑ_2 for short-run parameters.

This partitioning transfers immediately to the matrix $\Pi(\vartheta)$ because we know the rank of this matrix is equal to r. Therefore, we can factorize the matrix into two $(d \times r)$-dimensional matrices $\alpha(\vartheta)$ and $\beta(\vartheta)$ with rank r satisfying $\Pi(\vartheta) = \alpha(\vartheta)\beta(\vartheta)^\mathsf{T}$. We choose the rank factorization in such a way that $\beta(\cdot)$ depends only on ϑ_1. This is possible due to our previous results, where we have seen that $\beta(\vartheta_1)$ and $C_{1,\vartheta_1}^{\perp}$ must span the same cointegration space. Hence, we interpret $\beta(\vartheta_1)$ always as the orthogonal complement of C_{1,ϑ_1}.

The special form of the adjustment matrix $\alpha(\vartheta)$ is not of importance, we only need the rank of this matrix, which is equal to r for all $\vartheta \in \Theta$. If necessary for the further considerations, we write $\alpha(\vartheta)\beta(\vartheta_1)^{\mathsf{T}}$ otherwise we remain with the shorter notation $\Pi(\vartheta)$. Sometimes it is important to know if the columns of $\beta(\cdot)$ lie in the cointegration space or not. In such cases the more detailed representation is used.

Assumption M7
We assume that the true parameter vector ϑ^0 lies in the interior of the parameter space Θ.

Now we have with the separation of the parameter space the following decomposition of the pseudo-innovations

$$\varepsilon_k^{(h)}(\vartheta) = \varepsilon_{k,1}^{(h)}(\vartheta) + \varepsilon_{k,2}^{(h)}(\vartheta), \tag{5.10a}$$

where

$$\begin{aligned}\varepsilon_{k,1}^{(h)}(\vartheta) := &- \left[\Pi(\vartheta_1,\vartheta_2) - \Pi(\vartheta_1^0,\vartheta_2)\right] Y_{k-1}^{(h)} \\ &+ \left[\overline{k}(B,\vartheta_1,\vartheta_2) - \overline{k}(B,\vartheta_1^0,\vartheta_2)\right] \Delta Y_k^{(h)}\end{aligned} \tag{5.10b}$$

and

$$\varepsilon_{k,2}^{(h)}(\vartheta) := \varepsilon_{k,2}^{(h)}(\vartheta_2) = \overline{k}(B,\vartheta_1^0,\vartheta_2)\Delta Y_k^{(h)} - \Pi(\vartheta_1^0,\vartheta_2) Y_{k-1}^{(h)}. \tag{5.10c}$$

Furthermore, it holds that $\varepsilon_{k,1}^{(h)}(\vartheta_1^0,\vartheta_2) = 0$ for all $\vartheta_2 \in \Theta_2$ and $k \in \mathbb{N}$. When we use the true long-run parameter vector we have $\Pi(\vartheta_1^0,\vartheta_2)Y^{(h)} = \alpha(\vartheta_1^0,\vartheta_2)\beta(\vartheta_1^0)^{\mathsf{T}}Y^{(h)} = \Pi(\vartheta_1^0,\vartheta_2)Y_2^{(h)}$ for all $\vartheta_2 \in \Theta_2$ and thus we have a stationary process. The process $\left(\varepsilon_{k,2}^{(h)}(\vartheta_2)\right)_{k\in\mathbb{N}}$ is obviously stationary as it consists only of stationary processes.

However, the process $\left(\varepsilon_{k,1}^{(h)}(\vartheta)\right)_{k\in\mathbb{N}}$ is non-stationary for $\vartheta_1 \neq \vartheta_1^0$ under the assumption that we have an identifiable model. We achieve this in Section 5.3, where we solve the identifiability problem.

Henceforth, we have the separated log-likelihood function $\mathcal{L}_n^{(h)}(\vartheta)$ given by

$$\mathcal{L}_n^{(h)}(\vartheta) = \mathcal{L}_{n,1}^{(h)}(\vartheta) + \mathcal{L}_{n,2}^{(h)}(\vartheta_2), \tag{5.11a}$$

where

$$\begin{aligned}\mathcal{L}_{n,1}^{(h)}(\vartheta) &:= \frac{1}{n}\sum_{k=1}^{n}\varepsilon_{k,1}^{(h)}(\vartheta)^{\mathsf{T}}\big(V_{\vartheta_1,\vartheta_2}^{(h)}\big)^{-1}\varepsilon_{k,1}^{(h)}(\vartheta)\\
&\quad+\frac{1}{n}\sum_{k=1}^{n}\Big[2\cdot\varepsilon_{k,1}^{(h)}(\vartheta)^{\mathsf{T}}\big(V_{\vartheta_1,\vartheta_2}^{(h)}\big)^{-1}\varepsilon_{k,2}^{(h)}(\vartheta_2)+\varepsilon_{k,2}^{(h)}(\vartheta_2)^{\mathsf{T}}\big(V_{\vartheta_1,\vartheta_2}^{(h)}\big)^{-1}\varepsilon_{k,2}^{(h)}(\vartheta_2)\Big]\\
&\quad+\frac{1}{n}\sum_{k=1}^{n}\Big[\log\det V_{\vartheta_1,\vartheta_2}^{(h)}-\log\det V_{\vartheta_1^0,\vartheta_2}^{(h)}-\varepsilon_{k,2}^{(h)}(\vartheta_2)^{\mathsf{T}}\big(V_{\vartheta_1^0,\vartheta_2}^{(h)}\big)^{-1}\varepsilon_{k,2}^{(h)}(\vartheta_2)\Big]\\
&:= \frac{1}{n}\sum_{k=1}^{n}\ell_{k,1}^{(h)}(\vartheta)\end{aligned}\tag{5.11b}$$

and

$$\begin{aligned}\mathcal{L}_{n,2}^{(h)}(\vartheta_2) &:= \frac{1}{n}\sum_{k=1}^{n}\Big[d\log 2\pi+\log\det V_{\vartheta_2}^{(h)}+\varepsilon_{k,2}^{(h)}(\vartheta_2)^{\mathsf{T}}\big(V_{\vartheta_2}^{(h)}\big)^{-1}\varepsilon_{k,2}^{(h)}(\vartheta_2)\Big]\\
&:= \frac{1}{n}\sum_{k=1}^{n}\ell_{k,2}^{(h)}(\vartheta_2),\end{aligned}\tag{5.11c}$$

where we have written shortly $V_{\vartheta_2}^{(h)}$ for $V_{\vartheta_1^0,\vartheta_2}^{(h)}$. Furthermore, for reasons of brevity we write shortly ε_k for $\varepsilon_k(\vartheta^0)$, whenever we insert the true parameter ϑ^0. The same also goes for all the other cases, where we can save notation, e.g., we write C_1 instead of C_1^0 and so on.

Obviously, $\mathcal{L}_{n,2}^{(h)}(\vartheta_2)$ depends only on the short-run parameters, whereas $\mathcal{L}_{n,1}^{(h)}(\vartheta)$ depends on all parameters. Furthermore, we have the following relations $\mathcal{L}_{n,1}^{(h)}(\vartheta_1^0,\vartheta_2)=0$ and $\mathcal{L}_{n}^{(h)}(\vartheta_1^0,\vartheta_2)=\mathcal{L}_{n,2}^{(h)}(\vartheta_2)$ for every $\vartheta_2\in\Theta_2$. This immediately implies $\mathcal{L}_{n}^{(h)}(\vartheta^0)=\mathcal{L}_{n,2}^{(h)}(\vartheta^0)$.

Note that the interesting part for the asymptotic behavior of $\mathcal{L}_{n,1}^{(h)}(\vartheta)$ and $\mathcal{L}_{n,2}^{(h)}(\vartheta_2)$ are only the parts, where the pseudo-innovations are included, since the constant $d\log 2\pi$ and the $\log\det$ terms do not depend on n. For a representation of these parts we refer to Appendix 5.9.2.

Accordingly, the quasi-maximum likelihood estimator (5.7) is divided into two parts $\widehat{\vartheta}_{n,1}$ and $\widehat{\vartheta}_{n,2}$. The estimator $\widehat{\vartheta}_{n,1}$ estimates the long-run parameters, i.e. the subvector of parameters corresponding to the non-stationary part, and the estimator $\widehat{\vartheta}_{n,2}$ estimates the short-run parameters, i.e. the subvector of parameters corresponding to the stationary part.

For reasons of brevity, we write for the partial derivatives $\partial_i^1 := \frac{\partial}{\partial\vartheta_{1i}}$ with respect to the i^{th}-component of the non-stationary parameter vector $\vartheta_1 \in \Theta_1$ for $i \in \{1, \ldots, s_1\}$ and similar with respect to the stationary part $\partial_j^{st} := \frac{\partial}{\partial \vartheta_{2j}}$ with respect to the j^{th}-component of the stationary parameter vector $\vartheta_2 \in \Theta_2$ for $j \in \{1, \ldots, s_2\}$.

Assumption M8

Let the functions $\vartheta \mapsto A_{2,\vartheta}$, $\vartheta \mapsto B_{i,\vartheta}$, $\vartheta \mapsto C_{i,\vartheta}$ for $i \in \{1, 2\}$ and $\vartheta \mapsto \Sigma_\vartheta^L$ be twice continuously differentiable.

Like the continuity, the differentiability property is inherited by the matrices built from the system matrices as can be seen by Lemma 5.9.3. Moreover, the continuous differentiability and the compact parameter space imply Lipschitz continuity. This property is crucial as it limits the behavior of the functions sufficiently if we change the parameter vector.

The Jacobian matrix of the $d \times d$ matrix function $\overline{k}(z, \cdot)$ with respect to the $s_i \times 1$ parameter vector ϑ_i in the interior of Θ_i is defined by

$$\nabla_{\vartheta_i} \overline{k}(z, \cdot) := \frac{\partial \operatorname{vec}(\overline{k}(z, \cdot))}{\partial {\vartheta_i}^\mathsf{T}}$$

and analogous for the $d \times d$ matrix function $\Pi(\cdot) = \alpha(\cdot)\beta(\cdot)^\mathsf{T}$ with respect to the $s_i \times 1$ parameter vector ϑ_i in the interior of Θ_i by

$$\nabla_{\vartheta_i} \Pi(\cdot) := \frac{\partial \operatorname{vec}(\Pi(\cdot))}{\partial {\vartheta_i}^\mathsf{T}} = (I_d \otimes \alpha(\cdot)) \frac{\partial \operatorname{vec}(\beta(\cdot)^\mathsf{T})}{\partial {\vartheta_i}^\mathsf{T}} + (\beta(\cdot) \otimes I_d) \frac{\partial \operatorname{vec}(\alpha(\cdot))}{\partial {\vartheta_i}^\mathsf{T}},$$

for $i = 1, 2$. Note that the Jacobian matrices $\nabla_{\vartheta_i} \overline{k}(z, \vartheta)$ and $\nabla_{\vartheta_i} \Pi(\vartheta)$ have dimension $d^2 \times s_i$. For differentiation rules and helpful formulas of matrix differential calculus we refer to Appendix B.4.

The derivatives with respect to the stationary and non-stationary parameters of the pseudo-innovations are given by

$$\partial_i^1 \varepsilon_k^{(h)}(\vartheta) = \partial_i^1 \overline{k}(B, \vartheta) \Delta Y_k^{(h)} - \partial_i^1 \Pi(\vartheta) Y_{k-1}^{(h)}, \quad \text{for } i \in \{1, \ldots, s_1\} \tag{5.12a}$$

and

$$\partial_i^{st} \varepsilon_k^{(h)}(\vartheta) = \partial_i^{st} \overline{k}(B, \vartheta) \Delta Y_k^{(h)} - \partial_i^{st} \Pi(\vartheta) Y_{k-1}^{(h)}, \quad \text{for } i \in \{1, \ldots, s_2\}. \tag{5.12b}$$

Note that $\partial_i^{st} \varepsilon_k^{(h)}(\vartheta^0)$ is stationary due to the fact that $\partial_i^{st} \Pi(\vartheta^0) = (\partial_i^{st} \alpha(\vartheta^0)) \, \beta(\vartheta_1^0)^\mathsf{T}$. Hence, we still multiply with the matrix $\beta(\vartheta_1^0)^\mathsf{T}$, which is in the cointegration space and thereby the non-stationary part is canceled out. However, $\partial_i^1 \varepsilon_k^{(h)}(\vartheta^0)$ is non-stationary since $\partial_i^1 \Pi(\vartheta^0) = (\partial_i^1 \alpha(\vartheta^0))\, \beta(\vartheta_1^0)^\mathsf{T} + \alpha(\vartheta^0) \left(\partial_i^1 \beta(\vartheta_1^0)^\mathsf{T}\right)$. The first summand still cancels out the non-stationarity in contrast to the second one, as $\partial_i^1 \beta(\vartheta_1^0)^\mathsf{T}$ does in general not lie in the cointegration space.

The last lemma in this section is the analogue version of Lemma 5.2.1, where we consider now partial derivatives of the pseudo-innovations.

Lemma 5.2.2

Assume that Assumption M1-Assumption M6 and Assumption M8 hold. The pseudo-innovation sequence $\varepsilon^{(h)}(\vartheta)$ as defined in equation (5.5) has the following properties.

i) For each $v \in \{1, \dots, s\}$, the random sequences $\left(\partial_v \varepsilon_k^{(h)}(\vartheta)\right)_{k\in\mathbb{N}}$ is a linear function of $Y^{(h)}$, i.e. there exist matrix sequences $(\widetilde{K}_i^{(v)}(\vartheta))_{i\in\mathbb{N}}$, such that

$$\partial_v \varepsilon_k^{(h)}(\vartheta) = -\partial_v \Pi(\vartheta)^\mathsf{T} Y_{k-1}^{(h)} - \left(\sum_{i=1}^{\infty} \widetilde{K}_i^{(v)}(\vartheta) \Delta Y_{k-i}^{(h)} \right). \tag{5.13}$$

The matrices $\widetilde{K}_i^{(v)}(\vartheta)$ are uniformly exponentially bounded, i.e. there exist a positive constants c and $\rho < 1$, such that $\sup_{\vartheta\in\Theta} \|\widetilde{K}_i^{(v)}(\vartheta)\| \le c\rho^i$, $i \in \mathbb{N}$.

ii) For each $u, v \in \{1, \dots, s\}$, the random sequences $\left(\partial_u^2 \partial_v^2 \varepsilon_k^{(h)}(\vartheta)\right)_{k\in\mathbb{N}}$ are linear functions of Y, i.e. there exist matrix sequences $(\widetilde{K}_i^{(u,v)}(\vartheta))_{i\in\mathbb{N}}$, such that

$$\partial_u^2 \partial_v^2 \varepsilon_k^{(h)}(\vartheta) = -\partial_u^2 \partial_v^2 \Pi(\vartheta)^\mathsf{T} Y_{k-1}^{(h)} - \left(\sum_{i=1}^{\infty} \widetilde{K}_i^{(u,v)}(\vartheta) \Delta Y_{k-i}^{(h)} \right). \tag{5.14}$$

The matrices $\widetilde{K}_i^{(u,v)}(\vartheta)$ are uniformly exponentially bounded, i.e. there exist a positive constants c and $\rho < 1$, such that $\sup_{\vartheta\in\Theta} \|\widetilde{K}_i^{(u,v)}(\vartheta)\| \le c\rho^i$, $i \in \mathbb{N}$.

Proof. For a proof see Schlemm and Stelzer [91], Lemma 2.11, and recall that the coefficients $\widetilde{K}_i(\vartheta)$ are uniformly exponentially bounded due to Lemma 5.2.1. □

5.3. Identifiability

In order to properly estimate our model, it is important that we can guarantee a unique minimum to our likelihood function. The models should not be nested, i.e. that a smaller model can be included in a larger one. We want to exclude this and thus present in the following sufficient identifiability conditions. In other words we want the state space system to be identifiable. Therefore, redundancies in the parametrization are averted. Most importantly this gives us a unique minimum of the log-likelihood function for our considered model.

We have also to deal with the problem of the aliasing effect, which only affects the stationary part of our estimation procedure. This means that the reconstruction of the continuous-time model from the discrete time observation could be different. The aliasing effect is a general problem, which appears in the identification of parameters from discrete time data. For more details see e.g. Phillips [74] or Hansen and Sargent [47].

The aim is to derive sufficient assumptions which guarantee that we have different processes for different parameters. We derive in the following these assumptions. For this purpose we consider three different cases. First, we consider the identifiability for ϑ_1, then for ϑ_2 and in the last step combine both results.

Case 1:
Let us first assume that $\vartheta_2^0 = \vartheta_2$ and $\vartheta_1^0 \neq \vartheta_1 \in \Theta_1$. We show that the resulting processes must then be different. The aliasing effect plays no role in this case since the matrix C_{1,ϑ_1} does not depend on the sampling distance. To be precise the continuous-time model and the sampled model contain the same matrix C_{1,ϑ_1}, so the cointegration property is the same for both processes.

Assumption M9
The matrices C_{1,ϑ_1} and $C_{1,\vartheta_1}^{\perp} = \beta(\vartheta_1)$ are positive lower triangular matrices for all $\vartheta_1 \in \Theta_1$ as in Theorem 4.3.7 satisfying $C_{1,\vartheta_1}^{\mathsf{T}} C_{1,\vartheta_1} = I_c$ and $C_{1,\vartheta_1}^{\perp\mathsf{T}} C_{1,\vartheta_1}^{\perp} = I_{d-c}$.

We need to clarify, what we understand by the distance between subspaces. For this we use the gap metric, i.e. for two subspaces $W_1, W_2 \in \mathbb{R}^{n\times m}$ the distance between W_1 and W_2 is given by $\operatorname{dist}(W_1, W_2) := \|P_{W_1} - P_{W_2}\| = \max\{\|P_{W_1} \cdot P_{W_2^{\perp}}\|, \|P_{W_2} \cdot P_{W_1^{\perp}}\|\}$, where P_{W_i} is the orthogonal projection onto W_i for $i = 1, 2$. For more details on the gap metric see e.g. Fact 10.9.18. in Bernstein [9], Chapter II.4 in Stewart and Sun [94] or Chapter S4.3. in Gohberg et. al. [41]. This means that two subspaces are different as long as there exists a positive angle between them in at least one direction. In the following result, we have the equality between the two expressions in the maximum and hence the Gap metric simplifies in this case.

Lemma 5.3.1
Assume that Assumption M2 - Assumption M4 and Assumption M9 hold, then we

have for all $\vartheta_1^0 \neq \vartheta_1 \in \Theta_1$ *that*

$$\beta(\vartheta_1)^\mathsf{T} = C_{1,\vartheta_1}^{\perp\mathsf{T}} \neq C_1^{\perp} = \beta(\vartheta_1^0)^\mathsf{T} \tag{5.15}$$

and for all $\vartheta^0 \neq \vartheta \in \Theta$ *that*

$$\|\Pi(\vartheta)C_1\| = \|\alpha(\vartheta)\beta(\vartheta_1)^\mathsf{T} C_1\| > 0. \tag{5.16}$$

Proof. Recall that $C_1^\mathsf{T} C_1 = I_c$, where $C_1 \in \mathbb{R}^{d\times c}$ with $c \leq d$ and this implies $(C_1 C_1^\mathsf{T})C_1 = C_1$. Hereby, $C_1 C_1^\mathsf{T}$ is a projection on the c-dimensional subspace spanned by the orthonormal columns of C_1. As a consequence, it has c eigenvalues equal to one and $d-c$ eigenvalues equal to zero. Accordingly, $I_d - C_1 C_1^\mathsf{T}$ is the projection on the $(d-c)$-dimensional orthogonal complement of C_1.

Likewise, we see that $C_{1,\vartheta_1}^{\perp} C_{1,\vartheta_1}^{\perp\mathsf{T}}$ is the unique orthogonal projection on the $(d-c)$-dimensional subspace spanned by the columns of C_{1,ϑ_1}. The projection matrix $C_{1,\vartheta_1}^{\perp} C_{1,\vartheta_1}^{\perp\mathsf{T}}$ is the zero matrix if and only if $d = c$ and therefore $C_1 = I_d$ since we project onto a zero dimensional space. In this case the true matrix C_1 is already known and given by $C_1 = I_d$ and we have an integrated process in this case which we will exclude in the following considerations, i.e. we have $c < d$.

Eventually, we have due to Assumption M9 and the defining property of projection matrices ($P = P^2$) that

$$\begin{aligned}
&\|C_{1,\vartheta_1}^{\perp} C_{1,\vartheta_1}^{\perp\mathsf{T}} C_1\| = \operatorname{tr}\left(C_{1,\vartheta_1}^{\perp} C_{1,\vartheta_1}^{\perp\mathsf{T}} C_1 C_1^\mathsf{T} C_{1,\vartheta_1}^{\perp} C_{1,\vartheta_1}^{\perp\mathsf{T}}\right) = \operatorname{tr}\left(C_{1,\vartheta_1}^{\perp} C_{1,\vartheta_1}^{\perp\mathsf{T}} C_1 C_1^\mathsf{T}\right) \\
&= \operatorname{tr}\left(C_1 C_1^\mathsf{T} C_{1,\vartheta_1}^{\perp} C_{1,\vartheta_1}^{\perp\mathsf{T}} C_1 C_1^\mathsf{T}\right) = \|C_1 C_1^\mathsf{T} C_{1,\vartheta_1}^{\perp}\| = \operatorname{tr}\left(C_1 C_1^\mathsf{T} C_{1,\vartheta_1}^{\perp} C_{1,\vartheta_1}^{\perp\mathsf{T}}\right) \\
&= \operatorname{tr}\left(\left(I_d - C_1^{\perp} C_1^{\perp\mathsf{T}}\right) \cdot \left(I_d - C_{1,\vartheta_1} C_{1,\vartheta_1}^\mathsf{T}\right)\right) \\
&= \operatorname{tr}(I_d) - \operatorname{tr}\left(C_1^{\perp} C_1^{\perp\mathsf{T}}\right) - \operatorname{tr}\left(C_{1,\vartheta_1} C_{1,\vartheta_1}^\mathsf{T}\right) + \operatorname{tr}\left(C_1^{\perp} C_1^{\perp\mathsf{T}} C_{1,\vartheta_1} C_{1,\vartheta_1}^\mathsf{T}\right) \\
&= d - (d-c) - c + \|C_1^{\perp} C_1^{\perp\mathsf{T}} C_{1,\vartheta_1}\| \\
&= \|C_1^{\perp} C_1^{\perp\mathsf{T}} C_{1,\vartheta_1}\| = \|C_{1,\vartheta_1} C_{1,\vartheta_1}^\mathsf{T} C_1^{\perp}\| > 0.
\end{aligned}$$

Therefore, $C_{1,\vartheta_1}^{\perp}$ does not span the entire cointegration space and at least at one angle of the subspaces do not coincide. Thus, in at least one direction we have that the spaces are not orthogonal and it holds that

$$C_{1,\vartheta_1}^{\perp\mathsf{T}} C_1 \neq 0_{(d-c)\times c} \quad \text{and} \quad 0_{(d-c)\times c} \neq C_{1,\vartheta_1}^\mathsf{T} C_1^{\perp}.$$

The fact that $\alpha(\vartheta)$ has rank r for all $\vartheta \in \Theta$ proves the assertion. □

In conclusion, the lemma states that only for the true parameter value we have that all column vectors of $\beta(\vartheta_1)$ lie in the cointegration space. Since the same unique representation applies to the original subspace spanned by C_1, we immediately get $C_{1,\vartheta_1} \neq C_1$ for $\vartheta_1^0 \neq \vartheta_1 \in \Theta_1$.

Theorem 5.3.2
Assume that Assumption M1 - Assumption M4 and Assumption M9 hold. Then we have for $\vartheta_1^0 \neq \vartheta_1 \in \Theta_1$ that the processes cannot coincide, in other words we have for all $n \in \mathbb{N}$ and all $h > 0$ that

$$Y_n^{(h)}(\vartheta_1^0, \vartheta_2^0) \neq Y_n^{(h)}(\vartheta_1, \vartheta_2^0) \tag{5.17}$$

with probability one.

Proof. Recall the representation (4.23) for the sampled process. From the previous considerations we have $C_1 \neq C_{1,\vartheta_1}$ and B_1 has full rank due to Assumption M4. This implies

$$Y_n^{(h)}(\vartheta_1^0, \vartheta_2^0) - Y_n^{(h)}(\vartheta_1, \vartheta_2^0) = (C_1 - C_{1,\vartheta_1})X_1(0) + (C_1 - C_{1,\vartheta_1})B_{1,\vartheta_2}L_{\vartheta_2}(nh) \neq 0_d$$

with probability one. □

In the end, we know now that only for the true long-run parameter the span of $\beta(\vartheta_1^0)$ matches the cointegration space and thus the process $\left(\varepsilon_{k,1}^{(h)}(\vartheta)\right)_{k\in\mathbb{N}}$ is indeed always non-stationary for all long-run parameters $\vartheta_1 \neq \vartheta_1^0$.

Case 2:
On the other hand, if we assume now $\vartheta_2^0 \neq \vartheta_2 \in \Theta_2$ and $\vartheta_1 = \vartheta_1^0$, we know with the previous consideration that we can reduce the problem of identifiability to the stationary part in this setting

$$\begin{aligned}Y_n^{(h)}(\vartheta_1^0, \vartheta_2^0) - Y_n^{(h)}(\vartheta_1^0, \vartheta_2) &= Y_{n,2}^{(h)}(\vartheta_2^0) - Y_{n,2}^{(h)}(\vartheta_2)\\ &= C_2 \int_{-\infty}^{nh} \mathrm{e}^{A_2(nh-u)} B_2 \,\mathrm{d}L(u) - C_{2,\vartheta_2} \int_{-\infty}^{nh} \mathrm{e}^{A_{2,\vartheta_2}(nh-u)} B_{2,\vartheta_2} \,\mathrm{d}L(u).\end{aligned}$$

Since we only have to consider a stationary MCARMA process, we can utilize the results of Schlemm and Stelzer [91] (see Chapter 3.3 and 3.4) on the identifiability of a stationary MCARMA process.

In addition, we have to deal with the aliasing effect arising in the estimation of continuous-time models as already mentioned. We follow the ideas presented in Schlemm and Stelzer [91] in order to guarantee identifiability. First, we give some definitions and make further necessary assumptions.

Definition 5.3.3
Two stochastic processes, irrespective of whether their index sets are continuous or discrete, are L^2***-observationally equivalent*** *if their spectral densities are the same.*

Definition 5.3.4
A family $(Y_2(\vartheta_2), \vartheta_2 \in \Theta_2)$ *of continuous-time stochastic processes is* ***identifiable from the spectral density*** *if for every* $\vartheta_2 \neq \vartheta_2' \in \Theta_2$ *the processes* $Y_2(\vartheta_2)$ *and* $Y_2(\vartheta_2')$ *are not* L^2*-observationally equivalent.*

It is h***-identifiable from the spectral density****, for some* $h > 0$*, if for every* $\vartheta_2 \neq \vartheta_2' \in \Theta_2$*, the two sampled processes* $Y_2^{(h)}(\vartheta_2)$ *and* $Y_2^{(h)}(\vartheta_2')$ *are not* L^2*-observationally equivalent.*

Note that the spectral density $f_{Y_2}^{(h)} : [-\pi, \pi] \mapsto \mathbb{S}_d^+(\mathbb{R}(\mathrm{e}^{i\omega}))$ of the stationary part $Y_2^{(h)}$ is given by (see Schlemm and Stelzer [91], Proposition 3.6)

$$f_{Y_2}^{(h)}(\omega) = C_2 \left(\mathrm{e}^{i\omega} I_N - \mathrm{e}^{A_2 h}\right)^{-1} \widetilde{\Sigma}_{22}^{(h)} \left(\mathrm{e}^{i\omega} I_N - \mathrm{e}^{A_2^\mathsf{T} h}\right)^{-1} C_2^\mathsf{T}, \tag{5.18}$$

where $\widetilde{\Sigma}_{22}^{(h)} = \int_0^h \mathrm{e}^{A_2 u} B_2 \Sigma_L B_2^\mathsf{T} \mathrm{e}^{A_2^\mathsf{T} u} \mathrm{d}u$.

Assumption M10
Assume that the collection of the stationary parts of the output processes, denoted by

$$K(\Theta_2) := (Y_2(\vartheta_2), \vartheta_2 \in \Theta_2),$$

corresponding to the linear state space model $(A_{2,\vartheta_2}, B_{2,\vartheta_2}, C_{2,\vartheta_2}, L_{\vartheta_2})$ is identifiable from the spectral density.

Thus, we need another assumption in order to guarantee h-identifiability and to overcome the aliasing effect.

Assumption M11
For all $\vartheta_2 \in \Theta_2$ the spectrum of A_{2,ϑ_2} is a subset of $\{z \in \mathbb{C} : |\Im z| \leq \frac{\pi}{h}\}$.

Theorem 5.3.5 (Theorem 3.13, Schlemm and Stelzer [91])
Assume that $\Theta_2 \ni \vartheta_2 \mapsto (A_{2,\vartheta_2}, B_{2,\vartheta_2}, C_{2,\vartheta_2}, L_{\vartheta_2})$ *is a parameterizations of the stationary part of the continuous-time cointegrated process satisfying Assumption M1 - Assumption M3, Assumption M10 and Assumption M11. Then the corresponding*

collection of output processes $K(\Theta_2)$ is h-identifiable from the spectral density.

Hence, by the results of Schlemm and Stelzer [91] we can also guarantee identifiability of the stationary part due to the decoupled system in our setting and in addition to that find a sufficient condition that avoids the aliasing effect.

Case 3:
Lastly, we consider the case, where $\vartheta_2^0 \neq \vartheta_2 \in \Theta_2$ and $\vartheta_1^0 \neq \vartheta_1 \in \Theta_1$ holds. Identifiability follows directly due to a combination of both previous cases if we compare the stationary and non-stationary parts of the processes separately.

Now with these results in hand, we want to show that the identifiability transfers to the transfer function corresponding to the moving average representation of the pseudo-innovations, i.e. $\overline{k}(z,\vartheta) \neq \overline{k}(z,\vartheta^0)$ for $\vartheta^0 \neq \vartheta \in \Theta$.

Lemma 5.3.6
Assume that Assumption M1 - Assumption M4, and Assumption M9-Assumption M11 hold. Then it follows that for $\vartheta_1^0 \neq \vartheta_1 \in \Theta_1$ and $\vartheta_2 \neq \vartheta_2^0 \in \Theta$ there exists a complex number $z \in \mathbb{C}$ such that

$$C_\vartheta \left[I_N - \left(\mathrm{e}^{A_\vartheta h} - K_\vartheta^{(h)} C_\vartheta\right) z\right]^{-1} K_\vartheta^{(h)} \neq C\left[I_N - \left(\mathrm{e}^{Ah} - K^{(h)} C\right) z\right]^{-1} K^{(h)} \tag{5.19a}$$

or

$$V_{\vartheta_1^0,\vartheta_2}^{(h)} \neq V^{(h)} \tag{5.19b}$$

or for all $n \in \mathbb{N}$ and all $h > 0$ that with probability one

$$Y_n^{(h)}(\vartheta_1^0, \vartheta_2^0) \neq Y_n^{(h)}(\vartheta_1, \vartheta_2^0). \tag{5.19c}$$

Proof. If $\vartheta_1 \neq \vartheta_1^0$ Theorem 5.3.2 proves the claim. Thus it suffices that the non-stationary part is unique. If this is not the case a combination of Lemma 2.3 in Schlemm and Stelzer [91] and Theorem 5.3.5 proves the claim. □

5.4. Stochastic Equicontinuity and Continuous Convergence

The aim of this section is to derive in some sense uniform weak convergence results. Following the ideas of Saikkonen, we use the concept of continuous weak convergence, which is similar to uniform convergence. Several continuous convergence results combined with a stochastic equicontinuity condition are sufficient for the derivation of the asymptotic distribution later on.

First, we formally define continuous weak convergence and stochastic equicontinuity. For further details on these concepts see e.g. Sweeting [96], Basawa and Scott [6] and the articles of Saikkonen [85] and [86].

Let S be a separable metric space with metric d. Furthermore, we have an S-valued random vector $\mathcal{X}_n(\vartheta)$ defined on the probability space $(\Omega_n, \mathcal{A}_n, \mathbb{P}_{n,\vartheta})$, where $\mathbb{P}_{n,\vartheta}$ is a sequence of probability measures defined on the Borel sets of S for each $\vartheta \in \Theta$.

Definition 5.4.1

Let $\mathcal{C}(S)$ be the set of all bounded, continuous, real valued functions on S. We say that the probability measures $\mathbb{P}_{n,\vartheta_n}$ ***converge weakly in the continuous sense*** *to $\mathbb{P}_\vartheta$ ($\mathbb{P}_{n,\vartheta} \xrightarrow[c]{w} \mathbb{P}_\vartheta$) if and only if for all $f \in \mathcal{C}(X)$, for all $\vartheta \in \Theta$ and all sequences $\vartheta_n \to \vartheta$*

$$\int f \mathrm{d}\mathbb{P}_{n,\vartheta_n} \to \int f \mathrm{d}\mathbb{P}_\vartheta \tag{5.20}$$

for $n \to \infty$ holds.

The concept of continuous convergence cannot only be defined for weak convergence, but also for convergence in probability.

Definition 5.4.2

Let ϑ_n and ϑ be as in Definition 5.4.1 and $c(\vartheta)$ be a deterministic function from Θ to S. We say that the sequence of random vectors $\mathcal{X}_n(\vartheta_n)$ ***converges continuously in probability*** *to $c(\vartheta)$, denoted with $\mathcal{X}_n(\vartheta) \xrightarrow[c]{p} c(\vartheta)$ if and only if for every $\varepsilon > 0$ we have*

$$\mathbb{P}_{n,\vartheta_n}\left(d(\mathcal{X}_n(\vartheta_n), c(\vartheta)) > \varepsilon\right) \to 0 \tag{5.21}$$

as $n \to \infty$.

We briefly recall the definition of stochastic equicontinuity as in the works of Saikkonen [85], [86], in 1993 and 1995. For this purpose, we define the closed ball with radius δ by

$$B(\vartheta_2, \delta) := \{\vartheta_2^* \in \Theta_2 : \|\vartheta_2^* - \vartheta_2\| \leq \delta\}$$

and the set

$$N_n(\vartheta_1, \delta) := \{\vartheta_1^* \in \Theta_1 : \|D_{1n}(\vartheta_1^* - \vartheta_1)\| \le \delta\}, \tag{5.22}$$

where $D_{1n} \in M_{r_1,r_1}(\mathbb{R})$ is a diagonal matrix, whose diagonal elements are positive and increasing functions of n. Thus, the set $N_n(\vartheta_1, \delta)$ is decreasing in n.

Definition 5.4.3 (**Condition SE**)
Denote with $\mathcal{C}$ the class of all sequences ϑ_n with $\vartheta_n \to \vartheta^0$ and ϑ^0 varies over all points in Θ. For every $\{\vartheta_n\} \in \mathcal{C}$, every $\varepsilon > 0$ and every $\eta > 0$, there exists an integer $n(\varepsilon, \eta)$ and a real number $\delta_2 > 0$ such that for all $\delta_1 > 0$,

$$P_{n,\vartheta_n}\left(\sup_{\vartheta \in N_n(\vartheta_{n,1},\delta_1)\times B(\vartheta_{n,2},\delta_2)} \|\mathcal{X}_n(\vartheta) - \mathcal{X}_n(\vartheta_n)\| > \varepsilon\right) \le \eta, \tag{SE}$$

for $n \ge n(\varepsilon, \eta)$. The set $N_n(\vartheta_1, \delta)$ is defined as in (5.22).

The probability measure $P_{n,\vartheta}$ describes the distribution of the n observations which are used to estimate the parameter ϑ.

Note that if the condition (SE) holds true for the case ϑ reduced to ϑ_2, we even have uniform convergence due to Theorem 2.1 in Newey [71]. The theorem states that if we have a compact parameter space, pointwise convergence and condition (SE) holds for an open subset, these three assertions together are equivalent to uniform convergence. Moreover, it is in general easier to show the stochastic equicontinuity condition in contrast to uniform convergence.

Note that $\overline{k}(z, \vartheta_1, \vartheta_2) - \overline{k}(z, \vartheta_1^0, \vartheta_2)$ is uniformly exponentially bounded by Lemma 5.2.1, i.e. there exist constants $c < \infty$ and $0 < \rho < 1$ such that

$$\begin{aligned} &\sup_{\vartheta\in\Theta} \|\widetilde{K}_j(\vartheta_1, \vartheta_2) - \widetilde{K}_j(\vartheta_1^0, \vartheta_2)\| \\ &\le \sup_{\vartheta\in\Theta} \|\widetilde{K}_j(\vartheta_1, \vartheta_2)\| + \sup_{\vartheta_2\in\Theta_2} \|\widetilde{K}_j(\vartheta_1^0, \vartheta_2)\| \\ &\le c\rho^j. \end{aligned} \tag{5.23}$$

The next result shows the stochastic equicontinuity condition for different combinations of filtered version of the stationary processes $\Delta Y^{(h)}$ and $Y_2^{(h)}$. It is an adaption of results shown by Saikkonen (c.f. [85], Corollary 4.1).

Theorem 5.4.4
Assume that Assumption M1-Assumption M8 hold. Let $\overline{L}(z, \vartheta) = \sum_{i=0}^{\infty} \overline{L}_i(\vartheta) z^i$ and

$\underline{L}(z,\vartheta) = \sum_{i=0}^{\infty} \underline{L}_j(\vartheta) z^j$, $\vartheta \in \Theta$ *be* $d \times d$ *uniformly exponentially bounded families of matrix polynomials. Furthermore, let* $\overline{\xi}$ *and* $\underline{\xi}$ *be placeholders either for the stationary process* $\Delta Y^{(h)}$ *or* $Y_2^{(h)}$*. Suppose the matrix* $D_{in} = n^{\gamma_i} I_{s_i \times s_i}$ *is given as in Definition 5.4.3 with* $\gamma_i > 0$*, for* $i = 1, 2$*. Denote by* $\Gamma_{\overline{\xi}\underline{\xi}}(h) := \mathbb{E}\overline{\xi}_k \underline{\xi}_{k+h}^{\mathsf{T}}$.

Then we have

$$\frac{1}{n}\sum_{k=1}^{n} \overline{L}(z,\vartheta)\overline{\xi}_k \underline{\xi}_{k+h}^{\mathsf{T}} \underline{L}(z,\vartheta)^{\mathsf{T}} \xrightarrow[c]{p} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \overline{L}_i(\vartheta)\Gamma_{\overline{\xi}\underline{\xi}}(h+i-j)\underline{L}_j(\vartheta)^{\mathsf{T}} \tag{5.24}$$

and the stochastic equicontinuity condition (SE) *holds with* $\mathcal{X}_n(\vartheta)$ *given by the left-hand side of* (5.24).

Proof. The proof follows directly by Corollary 4.1 of Saikkonen [85] if we can show that all assumptions of the Corollary are satisfied. Note that we have two different coefficient matrices, whereas the results in Saikkonen [85] are shown for the same coefficient matrix. However, this result also holds if the coefficient matrices are different as long as each sequence of matrix coefficients satisfies the necessary conditions as mentioned in the paper of Saikkonen [85] (p. 163).

Since we have uniformly stable families of matrix polynomials, Assumption 4.1 of Saikkonen [85] is obviously satisfied. Moreover, the stationary processes $\Delta Y^{(h)}$ and $Y_2^{(h)}$ do not depend on ϑ and have finite moments due to Assumption M1. The last remaining conditions are the necessary convergence results, which are satisfied due to Lemma 5.8.2.

The last claim of the theorem is a consequence of Lemma 5.2.1 and this completes the proof. □

We need for the following proofs a continuous version of Slutsky's Theorem, which can be derived from Proposition 2.4 and 2.5 in Saikkonen [85]. Proposition 2.4 is a version of the continuous mapping theorem for continuous weak convergence. We neglect the difference between the different versions and speak only of the continuous mapping theorem or Slutsky's Theorem. Whether we use the standard version or the version for continuous weak convergence will be clear from the context.

Theorem 5.4.5
Assume that Assumption M1-Assumption M8 hold. Suppose $D_{in} = n^{\gamma_i} I_{s_i \times s_i}$ *as in Definition 5.4.3 for* $\gamma_i > 0$, $i = 1, 2$. *We can conclude for any value* γ_2 *that*

i) the weak convergence result

$$n^{-2}\sum_{k=1}^{n}\Pi(\vartheta)Y_{k-1}^{(h)}Y_{k-1}^{(h)\mathsf{T}}\Pi(\vartheta)^{\mathsf{T}}$$
$$\xrightarrow[c]{w}\Pi(\vartheta)C_1B_1\int_0^1 W(r)W(r)^{\mathsf{T}}\,\mathrm{d}r\,(\Pi(\vartheta)C_1B_1)^{\mathsf{T}} \tag{5.25}$$

and condition (SE) holds with $\gamma_1 > \frac{1}{2}$ for

$$\mathcal{X}_n(\vartheta) = n^{-1}\sum_{k=1}^{n}\Pi(\vartheta)Y_{k-1}^{(h)}Y_{k-1}^{(h)\mathsf{T}}\Pi(\vartheta)^{\mathsf{T}};$$

ii) the weak convergence result

$$n^{-1}\sum_{k=1}^{n}\Pi(\vartheta)Y_{k-1}^{(h)}\Delta Y_k^{(h)\mathsf{T}}\overline{k}(B,\vartheta)^{\mathsf{T}}$$
$$\xrightarrow[c]{w}\Pi(\vartheta)\big(C_1B_1\big)\int_0^1 W(r)\mathrm{d}W(r)^{\mathsf{T}}\,\big(C_1B_1\big)^{\mathsf{T}}\overline{k}(1,\vartheta)^{\mathsf{T}}$$
$$+\Pi(\vartheta)\big(C_1B_1\big)\int_0^1 W(r)\mathrm{d}W(r)^{\mathsf{T}}\,\widetilde{\Psi}(1)^{\mathsf{T}}\overline{k}(1,\vartheta)^{\mathsf{T}}+\Sigma_1(\vartheta) \tag{5.26}$$

and condition (SE) holds for $\mathcal{X}_n(\vartheta)$ given by the left-hand side of (5.26) with $\gamma_1 > 0$, where $\widetilde{\Psi}(1)$ is defined in Appendix 5.8 and

$$\Sigma_1(\vartheta) := \Pi(\vartheta)\Gamma_{Y\Delta Y}\overline{k}(1,\vartheta)^{\mathsf{T}}+\Pi(\vartheta)C_1B_1\widetilde{\Sigma}^{(h)}B_2\sum_{j=1}^{\infty}\widetilde{\Psi}_j\sum_{i=j+1}^{\infty}\widetilde{K}_i(\vartheta)$$
$$+\sum_{j=0}^{\infty}\Pi(\vartheta)\big[\Gamma_{Y_2\Delta Y}(-j)-\Gamma_{Y_2\Delta Y}(-j+1)\big]\sum_{i=j+1}^{\infty}\widetilde{K}_i(\vartheta)$$

is a matrix valued function depending on the parameter vector $\vartheta \in \Theta$.

The stated weak convergence results also hold jointly as well as the stochastic equicontinuity condition (SE).

Proof. i) Due to Proposition 5.8.4 iii) and the continuous version of Slutzky's Theorem we have

$$\Pi(\vartheta)\left[n^{-2}\sum_{k=1}^{n}Y_{k-1}^{(h)}Y_{k-1}^{(h)\mathsf{T}}\right]\Pi(\vartheta)^{\mathsf{T}}$$
$$\xrightarrow[c]{w}\Pi(\vartheta)\int_0^1 C_1B_1W(r)W(r)^{\mathsf{T}}B_1^{\mathsf{T}}C_1^{\mathsf{T}}\,\mathrm{d}r\,\Pi(\vartheta)^{\mathsf{T}}. \tag{5.27}$$

In the next step we want to prove the stochastic equicontinuity condition. Note that $\Pi(\cdot)$ is Lipschitz continuous due to Lemma 5.9.3. Hence, we can find the following upper bound

$$\begin{aligned}
&\sup_{\vartheta^* \in N_n(\vartheta_{n,1},\delta_1)\times B(\vartheta_{n,2},\delta_2)} \|(\Pi(z,\vartheta_1^*,\vartheta_2^*) - \Pi(z,\vartheta_{n,1},\vartheta_2^*))\| \\
&\leq \sup_{\vartheta^* \in N_n(\vartheta_{n,1},\delta_1)\times B(\vartheta_{n,2},\delta_2)} c\cdot \|\vartheta_1^* - \vartheta_{n,1}\| \\
&\leq c\delta_1 n^{-\gamma_1}, \qquad (5.28)
\end{aligned}$$

for some constant $c > 0$. Note that we can rewrite

$$\begin{aligned}
&[\Pi(\vartheta^*) - \Pi(\vartheta_n)] Z_n [\Pi(\vartheta^*) - \Pi(\vartheta_n)]^\mathsf{T} \\
&\quad + \Pi(\vartheta_n) Z_n [\Pi(\vartheta^*) - \Pi(\vartheta_n)]^\mathsf{T} + [\Pi(\vartheta^*) - \Pi(\vartheta_n)] Z_n \Pi(\vartheta_n)^\mathsf{T} \\
&= \Pi(\vartheta^*) Z_n \Pi(\vartheta^*)^\mathsf{T} + \Pi(\vartheta_n) Z_n \Pi(\vartheta_n)^\mathsf{T} - \Pi(\vartheta_n) Z_n \Pi(\vartheta^*)^\mathsf{T} - \Pi(\vartheta^*) Z_n \Pi(\vartheta_n)^\mathsf{T} \\
&\quad + \Pi(\vartheta_n) Z_n \Pi(\vartheta^*)^\mathsf{T} - \Pi(\vartheta_n) Z_n \Pi(\vartheta_n)^\mathsf{T} + \Pi(\vartheta^*) Z_n \Pi(\vartheta_n)^\mathsf{T} - \Pi(\vartheta_n) Z_n \Pi(\vartheta_n)^\mathsf{T} \\
&= \Pi(\vartheta^*) Z_n \Pi(\vartheta^*)^\mathsf{T} - \Pi(\vartheta_n) Z_n \Pi(\vartheta_n)^\mathsf{T}, \qquad (5.29)
\end{aligned}$$

and

$$\begin{aligned}
&[\Pi(\vartheta^*) - \Pi(\vartheta_n)] Z_n [\Pi(\vartheta^*) - \Pi(\vartheta_n)]^\mathsf{T} \\
&= [\Pi(\vartheta_1^*,\vartheta_2^*) - \Pi(\vartheta_{n,1},\vartheta_2^*) + \Pi(\vartheta_{n,1},\vartheta_2^*) - \Pi(\vartheta_{n,1},\vartheta_{n,2})] Z_n \\
&\qquad \cdot [\Pi(\vartheta_1^*,\vartheta_2^*) - \Pi(\vartheta_{n,1},\vartheta_2^*) + \Pi(\vartheta_{n,1},\vartheta_2^*) - \Pi(\vartheta_{n,1},\vartheta_{n,2})]^\mathsf{T} \\
&= [\Pi(\vartheta_1^*,\vartheta_2^*) - \Pi(\vartheta_{n,1},\vartheta_2^*)] Z_n [\Pi(\vartheta_1^*,\vartheta_2^*) - \Pi(\vartheta_{n,1},\vartheta_2^*)]^\mathsf{T} \\
&\quad + [\Pi(\vartheta_{n,1},\vartheta_2^*) - \Pi(\vartheta_{n,1},\vartheta_{n,2})] Z_n [\Pi(\vartheta_1^*,\vartheta_2^*) - \Pi(\vartheta_{n,1},\vartheta_2^*)]^\mathsf{T} \\
&\quad + [\Pi(\vartheta_1^*,\vartheta_2^*) - \Pi(\vartheta_{n,1},\vartheta_2^*)] Z_n [\Pi(\vartheta_{n,1},\vartheta_2^*) - \Pi(\vartheta_{n,1},\vartheta_{n,2})]^\mathsf{T} \\
&\quad + [\Pi(\vartheta_{n,1},\vartheta_2^*) - \Pi(\vartheta_{n,1},\vartheta_{n,2})] Z_n \Pi(\vartheta_{n,1},\vartheta_2^*) - \Pi(\vartheta_{n,1},\vartheta_{n,2})]^\mathsf{T}, \qquad (5.30)
\end{aligned}$$

where Z_n is a placeholder for $n^{-1}\sum_{k=1}^n Y_{k-1}^{(h)} Y_{k-1}^{(h)\mathsf{T}}$. Thus, condition (SE) follows similar as in the proof of Corollary 2.2 in Newey [71] due to the submultiplicativity since

$$\begin{aligned}
&\sup_{\vartheta^* \in N_n(\vartheta_{n,1},\delta_1)\times B(\vartheta_{n,2},\delta_2)} \|\mathcal{X}_n(\vartheta^*) - \mathcal{X}_n(\vartheta_n)\| \\
&= \sup_{\vartheta^* \in N_n(\vartheta_{n,1},\delta_1)\times B(\vartheta_{n,2},\delta_2)} \Bigg\| \Pi(\vartheta^*) n^{-1} \sum_{k=1}^n Y_{k-1}^{(h)} Y_{k-1}^{(h)\mathsf{T}} \Pi(\vartheta^*)^\mathsf{T} \\
&\qquad\qquad - \Pi(\vartheta_n) n^{-1} \sum_{k=1}^n Y_{k-1}^{(h)} Y_{k-1}^{(h)\mathsf{T}} \Pi(\vartheta_n)]^\mathsf{T} \Bigg\|
\end{aligned}$$

$$
\begin{aligned}
&\overset{(5.29)}{\leq} \sup_{\vartheta^* \in N_n(\vartheta_{n,1},\delta_1)\times B(\vartheta_{n,2},\delta_2)} \left\| [\Pi(\vartheta^*) - \Pi(\vartheta_n)] n^{-1} \sum_{k=1}^{n} Y_{k-1}^{(h)} Y_{k-1}^{(h)\mathsf{T}} [\Pi(\vartheta^*) - \Pi(\vartheta_n)]^{\mathsf{T}} \right\| \\
&\quad + \sup_{\vartheta^* \in N_n(\vartheta_{n,1},\delta_1)\times B(\vartheta_{n,2},\delta_2)} 2 \cdot \left\| [\Pi(\vartheta^*) - \Pi(\vartheta_n)]\, n^{-1} \sum_{k=1}^{n} Y_{k-1}^{(h)} Y_{k-1}^{(h)\mathsf{T}} \Pi(\vartheta_n)^{\mathsf{T}} \right\| \\
&\overset{(5.30)}{\leq} \sup_{\vartheta^* \in N_n(\vartheta_{n,1},\delta_1)} \|\Pi(\vartheta_1^*, \vartheta_2^*) - \Pi(\vartheta_{n,1}, \vartheta_2^*)\|^2 \cdot \left\| n^{-1} \sum_{k=1}^{n} Y_{k-1}^{(h)} Y_{k-1}^{(h)\mathsf{T}} \right\| \\
&\quad + \sup_{\vartheta^* \in N_n(\vartheta_{n,1},\delta_1)\times B(\vartheta_{n,2},\delta_2)} 2 \cdot \|\Pi(\vartheta_1^*, \vartheta_2^*) - \Pi(\vartheta_{n,1}, \vartheta_2^*)\| \\
&\qquad\qquad \cdot \|\alpha(\vartheta_{n,1}, \vartheta_2^*) - \alpha(\vartheta_{n,1}, \vartheta_{n,2})\| \left\| n^{-1} \sum_{k=1}^{n} Y_{k-1}^{(h)} Y_{k-1}^{(h)\mathsf{T}} \beta(\vartheta_{n,1}) \right\| \\
&\quad + \sup_{\vartheta^* \in N_n(\vartheta_{n,1},\delta_1)} \|\alpha(\vartheta_{n,1}, \vartheta_2^*) - \alpha(\vartheta_{n,1}, \vartheta_{n,2})\|^2 \\
&\qquad\qquad \cdot \left\| n^{-1} \sum_{k=1}^{n} \beta(\vartheta_{n,1})^{\mathsf{T}} Y_{k-1}^{(h)} Y_{k-1}^{(h)\mathsf{T}} \beta(\vartheta_{n,1}) \right\| \\
&\quad + \sup_{\vartheta^* \in N_n(\vartheta_{n,1},\delta_1)\times B(\vartheta_{n,2},\delta_2)} 2 \cdot \|\Pi(\vartheta^*) - \Pi(\vartheta_n)\| \left\| n^{-1} \sum_{k=1}^{n} Y_{k-1}^{(h)} Y_{k-1}^{(h)\mathsf{T}} \Pi(\vartheta_n)^{\mathsf{T}} \right\| \\
&\overset{(5.28)}{\leq} c^2 \delta_1^2 \left\| n^{-1-2\gamma_1} \sum_{k=1}^{n} Y_{k-1}^{(h)} Y_{k-1}^{(h)\mathsf{T}} \right\| \\
&\quad + c^2 \delta_1 \delta_2 \left\| n^{-1-\gamma_1} \sum_{k=1}^{n} Y_{k-1}^{(h)} Y_{k-1,2}^{(h)\mathsf{T}} \beta(\vartheta_{n,1}) \right\| \\
&\quad + c\delta_2 \left\| n^{-1} \sum_{k=1}^{n} \beta(\vartheta_{n,1})^{\mathsf{T}} Y_{k-1,2}^{(h)} Y_{k-1,2}^{(h)\mathsf{T}} \beta(\vartheta_{n,1}) \right\| \\
&\quad + c\delta_1 \left\| n^{-1-\gamma_1} \sum_{k=1}^{n} Y_{k-1}^{(h)} Y_{k-1,2}^{(h)\mathsf{T}} \beta(\vartheta_{n,1}) \right\| \\
&\quad + c\delta_2 \left\| n^{-1} \sum_{k=1}^{n} Y_{k-1}^{(h)} Y_{k-1,2}^{(h)\mathsf{T}} \beta(\vartheta_{n,1}) \right\| .
\end{aligned}
$$

We are going to apply Proposition 5.8.4 to these result in order to determine the asymptotic behavior. The first term converges continuously in probability to zero due to Proposition 5.8.4 iii) for $\gamma_1 > \frac{1}{2}$.

The definition of continuous convergence combined with Lemma 5.8.2 for the third and Proposition 5.8.4 v) for the fifth term imply that they are of order $\mathcal{O}_{P_{n,\vartheta_n}}(1)$. Thus, there exists a $\delta_2 > 0$ that these terms are small enough, which yields the claim.

The remaining claims all hold for $\gamma_1 > 0$ and thus in particular for $\gamma_1 > \frac{1}{2}$. The second and fourth term converge as well continuously in probability to zero due to the fact that $\beta(\vartheta_{n,1})^{\mathsf{T}} Y_{k-1,2}$ is stationary under the probability measure P_{n,ϑ_n} (c.f. Saikkonen [85] p.170) and Proposition 5.8.4 v).

In summary, we have established condition (SE) for $\gamma_1 > \frac{1}{2}$.

ii) For the second part we obtain with the Beveridge-Nelson decomposition (c.f. Saikkonen [86], (9)) applied to $\overline{k}(B,\vartheta)\Delta Y_k^{(h)} = \overline{k}(1,\vartheta)\Delta Y_k^{(h)} + \eta_k(\vartheta) - \eta_{k-1}(\vartheta)$ with $\eta_k(\vartheta) := -\sum_{j=0}^{\infty}\sum_{i=j+1}^{\infty} \widetilde{K}_i(\vartheta)\Delta Y_{k-j}^{(h)}$ that

$$\begin{aligned}
& n^{-1}\sum_{k=1}^{n} \Pi(\vartheta) Y_{k-1}^{(h)} \Delta Y_k^{(h)\mathsf{T}} \overline{k}(B,\vartheta)^{\mathsf{T}} \\
&= n^{-1}\sum_{k=1}^{n} \Pi(\vartheta) Y_{k-1}^{(h)} \Delta Y_k^{(h)\mathsf{T}} \overline{k}(1,\vartheta)^{\mathsf{T}} + n^{-1}\sum_{k=1}^{n} \Pi(\vartheta) Y_{k-1}^{(h)} \left(\eta_k(\vartheta) - \eta_{k-1}(\vartheta)\right) \\
&= n^{-1}\sum_{k=1}^{n} \Pi(\vartheta) Y_{k-1}^{(h)} \Delta Y_k^{(h)\mathsf{T}} \overline{k}(1,\vartheta)^{\mathsf{T}} \\
&\quad + n^{-1}\sum_{k=1}^{n} \Pi(\vartheta) \sum_{i=1}^{k-1} C_1 R_{i,1}^{(h)} \left(\eta_k(\vartheta) - \eta_{k-1}(\vartheta)\right) \\
&\quad + n^{-1}\sum_{k=1}^{n} \Pi(\vartheta) Y_{k-1,2}^{(h)} \left(\eta_k(\vartheta) - \eta_{k-1}(\vartheta)\right) \\
&=: I_{n,1}(\vartheta) + I_{n,2}(\vartheta) + I_{n,3}(\vartheta).
\end{aligned}$$

We consider in the following the asymptotic behavior of $I_{n,i}(\vartheta)$ for $i = 1,2,3$.

Step 1:
The first term $I_{n,1}(\vartheta)$ converges due to the continuous mapping theorem and Proposition 5.8.4 iv).

Step 2:
One can easily see the convergence of the second term $I_{n,2}(\vartheta)$ by applying the summation by parts formula and the Beveridge-Nelson decomposition (c.f. Saikkonen [85], (A.8))

$$\begin{aligned}
I_{n,2} &= n^{-1}\sum_{k=1}^{n} \Pi(\vartheta) \sum_{i=1}^{k-1} C_1 R_{i,1}^{(h)} \left(\eta_k(\vartheta) - \eta_{k-1}(\vartheta)\right) \\
&= n^{-1}\Pi(\vartheta) C_1 R_{n,1}^{(h)} \eta_n(\vartheta) - n^{-1}\sum_{k=1}^{n} \Pi(\vartheta) C_1 R_{k,1}^{(h)} \eta_k(\vartheta) \\
&\xrightarrow[c]{p} \Pi(\vartheta) C_1 B_1 \widetilde{\Sigma}^{(h)} B_2 \sum_{j=1}^{\infty} \widetilde{\Psi}_j \sum_{i=j+1}^{\infty} \widetilde{K}_i(\vartheta).
\end{aligned}$$

Step 3:
The third term $I_{n,3}(\vartheta)$ converges due to Theorem 5.4.4 (c.f. Saikkonen [85], (A.9)) continuously in probability

$$
\begin{aligned}
I_{n,3}(\vartheta) &= n^{-1} \sum_{k=1}^{n} \Pi(\vartheta) Y^{(h)}_{k-1,2} \left(\eta_k(\vartheta) - \eta_{k-1}(\vartheta)\right) \\
&\xrightarrow[c]{p} \sum_{j=0}^{\infty} \Pi(\vartheta) \left(\Gamma_{Y_2 \Delta Y}(-j) - \Gamma_{Y_2 \Delta Y}(-j+1)\right) \sum_{i=j+1}^{\infty} \widetilde{K}_i(\vartheta).
\end{aligned}
$$

We set

$$
\begin{aligned}
\Sigma_1(\vartheta) :=& \Pi(\vartheta) \Gamma_{Y \Delta Y} \overline{k}(1,\vartheta)^{\mathsf{T}} + \Pi(\vartheta) C_1 B_1 \widetilde{\Sigma}^{(h)} B_2 \sum_{j=1}^{\infty} \widetilde{\Psi}_j \sum_{i=j+1}^{\infty} \widetilde{K}_i(\vartheta) \\
&+ \sum_{j=0}^{\infty} \Pi(\vartheta) \left[\Gamma_{Y_2 \Delta Y}(-j) - \Gamma_{Y_2 \Delta Y}(-j+1)\right] \sum_{i=j+1}^{\infty} \widetilde{K}_i(\vartheta)
\end{aligned}
$$

To sum it up we obtain with the previous considerations and Slutzky's lemma that the sum of the $I_{n,i}(\vartheta)$ converges continuously weakly to

$$
\begin{aligned}
&I_{n,1}(\vartheta) + I_{n,2}(\vartheta) + I_{n,3}(\vartheta) \\
&\xrightarrow[c]{w} \Pi(\vartheta_1) \big(C_1 B_1, 0_{d \times (N-c)}\big) \int_0^1 W(r) \, \mathrm{d}W(r)^{\mathsf{T}} \big(C_1 B_1, \widetilde{\Psi}(1)\big)^{\mathsf{T}} \overline{k}(1,\vartheta)^{\mathsf{T}} \\
&\quad + \Pi(\vartheta_1) \Gamma_{Y \Delta Y} \overline{k}(1,\vartheta)^{\mathsf{T}} + \Pi(\vartheta) C_1 B_1 \widetilde{\Sigma}^{(h)} B_2 \sum_{j=1}^{\infty} \widetilde{\Psi}_j \sum_{i=j+1}^{\infty} \widetilde{K}_i(\vartheta) \\
&\quad + \sum_{j=0}^{\infty} \Pi(\vartheta) \left(\Gamma_{Y_2 \Delta Y}(-j) - \Gamma_{Y_2 \Delta Y}(-j+1)\right) \sum_{i=j+1}^{\infty} \widetilde{K}_i(\vartheta). \qquad (5.31)
\end{aligned}
$$

Now, we have similar as in the proof of part (i)

$$
\begin{aligned}
&\mathcal{X}_n(\vartheta^*) - \mathcal{X}_n(\vartheta_n) \\
&= \frac{1}{n} \sum_{k=1}^{n} \left[\Pi(\vartheta^*) - \Pi(\vartheta_n)\right] Y^{(h)}_{k-1} \Delta Y^{(h)\mathsf{T}}_k \left[\overline{k}(B,\vartheta^*) - \overline{k}(B,\vartheta_n)\right]^{\mathsf{T}} \\
&\quad + \frac{1}{n} \sum_{k=1}^{n} \Pi(\vartheta_n) Y^{(h)}_{k-1} \Delta Y^{(h)\mathsf{T}}_k \left[\overline{k}(B,\vartheta^*_1) - \overline{k}(B,\vartheta_n)\right]^{\mathsf{T}} \\
&\quad + \frac{1}{n} \sum_{k=1}^{n} \left[\Pi(\vartheta^*) - \Pi(\vartheta_n)\right] Y^{(h)}_{k-1} \Delta Y^{(h)\mathsf{T}}_k \overline{k}(B,\vartheta_n)^{\mathsf{T}}. \qquad (5.32)
\end{aligned}
$$

We can also use the Lipschitz continuity of $\overline{k}(z,\cdot)$ due to Lemma 5.9.3. Note that we have a compact parameter space Θ and thus $\delta_1 > 0$ is bounded. By taking the supremum we have similar to (5.28)

$$\begin{aligned}
&\sup_{\vartheta^* \in N_n(\vartheta_{n,1},\delta_1)\times B(\vartheta_{n,2},\delta_2)} \left\| \left(\overline{k}(z,\vartheta_1^*,\vartheta_2^*) - \overline{k}(z,\vartheta_{n,1},\vartheta_{n,2}) \right) \right\| \\
&\leq \sup_{\vartheta^* \in N_n(\vartheta_{n,1},\delta_1)\times B(\vartheta_{n,2},\delta_2)} c \cdot \|\vartheta^* - \vartheta_n\| \\
&= \sup_{\vartheta^* \in N_n(\vartheta_{n,1},\delta_1)\times B(\vartheta_{n,2},\delta_2)} c \cdot (\|\vartheta_1^* - \vartheta_{n,1}\| + \|\vartheta_2^* - \vartheta_{n,2}\|)^{\frac{1}{2}} \\
&\leq \sup_{\vartheta^* \in N_n(\vartheta_{n,1},\delta_1)\times B(\vartheta_{n,2},\delta_2)} c \cdot (\|\vartheta_1^* - \vartheta_{n,1}\| + \|\vartheta_2^* - \vartheta_{n,2}\|) \\
&\leq c\delta_1 n^{-\gamma_1} + c\delta_2 \leq c_1(n^{-\gamma} + \delta_2),
\end{aligned} \tag{5.33}$$

for some constant $c, c_1 > 0$.

Finally, we make once more use of the submultiplicativity and this leads with a bounding argument to

$$\begin{aligned}
&\sup_{\vartheta^* \in N_n(\vartheta_{n,1},\delta_1)\times B(\vartheta_{n,2},\delta_2)} \left\| \mathcal{X}_n(\vartheta^*) - \mathcal{X}_n(\vartheta_n) \right\| \\
&\overset{(5.32)}{\leq} \sup_{\vartheta^* \in N_n(\vartheta_{n,1},\delta_1)\times B(\vartheta_{n,2},\delta_2)} \|\Pi(\vartheta^*) - \Pi(\vartheta_n)\| \cdot \left\| n^{-1} \sum_{k=1}^n Y_{k-1}^{(h)} \Delta Y_k^{(h)\mathsf{T}} \overline{k}(B,\vartheta_n)^\mathsf{T} \right\| \\
&\quad + \sup_{\vartheta^* \in N_n(\vartheta_{n,1},\delta_1)\times B(\vartheta_{n,2},\delta_2)} \left\| \left(\overline{k}(B,\vartheta^*) - \overline{k}(B,\vartheta_n) \right) \right\| \left\| n^{-1} \sum_{k=1}^n \Pi(\vartheta_n) Y_{k-1}^{(h)} \Delta Y_k^{(h)\mathsf{T}} \right\| \\
&\quad + \sup_{\vartheta^* \in N_n(\vartheta_{n,1},\delta_1)\times B(\vartheta_{n,2},\delta_2)} \left\| \left(\overline{k}(B,\vartheta^*) - \overline{k}(B,\vartheta_n) \right) \right\| \|\Pi(\vartheta^*) - \Pi(\vartheta_n)\| \\
&\qquad \cdot \left\| n^{-1} \sum_{k=1}^n Y_{k-1}^{(h)} \Delta Y_k^{(h)\mathsf{T}} \right\| \\
&\overset{(5.33)}{\leq} c_1 \left(n^{-\gamma_1} + \delta_2 \right) \left\| n^{-1} \sum_{k=1}^n Y_{k-1}^{(h)} \Delta Y_k^{(h)\mathsf{T}} \overline{k}(B,\vartheta_n)^\mathsf{T} \right\| \\
&\quad + c_1 \left(n^{-\gamma_1} + \delta_2 \right) \left\| n^{-1} \sum_{k=1}^n \Pi(\vartheta_n) Y_{k-1}^{(h)} \Delta Y_k^{(h)\mathsf{T}} \right\| \\
&\quad + c_1 \left(n^{-2\gamma_1} + \delta_2 \right) \left\| n^{-1} \sum_{k=1}^n Y_{k-1}^{(h)} \Delta Y_k^{(h)\mathsf{T}} \right\|.
\end{aligned}$$

The following assertions all hold for $\gamma_1 > 0$. To begin with, the first term in the norm converges continuously in probability due to (5.31) (c.f. Saikkonen [85], Theorem 4.2(iv) and 4.4). Next, the second term in the norm converges continuously in probability since it satisfies the conditions of Theorem 5.4.4 and

$\Pi(\vartheta_{n,1})Y_{k-1}^{(h)}$ is stationary under P_{n,ϑ_n} (c.f. Proof of Theorem 4.5, Saikkonen [85]). Finally, the last term in the norm also converges continuously on the basis of Proposition 5.8.4 iv). Thus, we have three terms which are all of order $\mathcal{O}_p(1)$ and prefactors which consists of a term converging to zero as $n \to \infty$ and $\delta_2 > 0$ can be chosen small enough such that the stochastic equicontinuity (SE) holds for $\gamma_1 > 0$.

In the end, it only remains to show joint convergence of the results. Note that we have shown all assumptions (c.f. Lemma 5.8.2, Lemma 5.8.3 and Proposition 5.8.4) which are necessary for the corresponding results in Saikkonen's articles [85] (Assumptions of Theorem 4.5) and [86] (Assumptions of Theorem 4.1). This being the case, one obtains joint convergence as a consequence of these results. □

5.5. Consistency of the Quasi-Maximum Likelihood Estimator

In order to show the consistency, we follow the ideas of Saikkonen [86]. Thus, we prove the consistency in three steps. In the first two steps, we prove the consistency of the long-run parameter estimator $\widehat{\vartheta}_{n,1}$ and determine the consistency rate. Lastly, we prove the consistency of the short-run parameter estimator $\widehat{\vartheta}_{n,2}$ by using the results from the first two steps.

Before we prove the consistency result we first show that the limiting function of the log-likelihood function of the stationary part has a unique minimum at ϑ_2^0. Let us derive first the limiting function.

Proposition 5.5.1

Assume that Assumption M1-Assumption M8 hold. Then we have that the sequence of random functions $\left(\mathcal{L}_{n,2}^{(h)}(\vartheta_2)\right)_{\vartheta_2\in\Theta_2}$ *converges continuously in probability for* $n \to \infty$ *to the limiting function* $\boldsymbol{\mathcal{L}}_2^{(h)} : \Theta_2 \to \mathbb{R}$ *given by*

$$\mathcal{L}_2^{(h)}(\vartheta_2) = d\log(2\pi) + \log\det V_{\vartheta_2}^{(h)} + \mathbb{E}\varepsilon_{1,2}^{(h)}(\vartheta_2)^\mathsf{T}\left(V_{\vartheta_2}^{(h)}\right)^{-1}\varepsilon_{1,2}^{(h)}(\vartheta_2).$$

Proof. We can prove the claim by using Theorem 5.4.4 and the continuous mapping theorem. Recall the representation of $\mathcal{L}_{n,2}^{(h)}(\vartheta_2)$

$$\mathcal{L}_{n,2}^{(h)}(\vartheta_2) = d\log 2\pi + \log\det V_{\vartheta_1^0,\vartheta_2}^{(h)} + \operatorname{tr}\left(\left(V_{\vartheta_1^0,\vartheta_2}^{(h)}\right)^{-1}\cdot\frac{1}{n}\sum_{k=1}^{n}\varepsilon_{k,2}^{(h)}(\vartheta_2)\varepsilon_{k,2}^{(h)}(\vartheta_2)^\mathsf{T}\right).$$

We realize that the only term of interest for the convergence is the last one. Note that we have by Lemma 5.9.2 iii) that $\|(V^{(h)}_{\vartheta_1^0,\vartheta_2})^{-1}\| \leq c$ for some constant $c > 0$. Besides, for all $\vartheta_2 \in \Theta_2$ we have the continuous convergence due to Theorem 5.4.4 and the form of $(\varepsilon^{(h)}_{k,2}(\vartheta_2))$ given in (5.10) with uniformly stable matrix coefficients due to Lemma 5.2.1. □

Lemma 5.5.2

Let Assumption M1-Assumption M8, Assumption M10 and Assumption M11 hold. The function $\mathcal{L}_2^{(h)} : \Theta_2 \to \mathbb{R}$ has a unique global minimum at ϑ_2^0.

Proof. The proof is analogous to Lemma 2.10 in Schlemm and Stelzer [91] since Assumption D5 is satisfied due to Lemma 5.3.6. □

Assume throughout the rest of this section that Assumption M1-Assumption M11 always hold.

We begin with the estimator $\widehat{\vartheta}_{n,1}$ of the parameter ϑ_1^0, which is connected to the non-stationary part. For this purpose, let us define the following set for $n \in \mathbb{N}$

$$N_{n,\gamma}(\vartheta_1^0, \delta) := \left\{\vartheta_1 \in \Theta_1 : \|\vartheta_1 - \vartheta_1^0\| \leq \delta n^{-\gamma}\right\}, \tag{5.34}$$

which is decreasing in n.

As in Saikkonen [86], we want to show that $\widehat{\vartheta}_{n,1} - \vartheta_1^0 = o_p(n^{-\gamma})$ holds for all $0 \leq \gamma < 1$ by proving a sufficient condition. The existence of the estimator is guaranteed due to the assumptions made. Among other things, we have a compact parameter space and a continuous log-likelihood function.

To show the consistency for the long run parameter, it suffices to show that for all $\delta > 0$ we have

$$\lim_{n\to\infty} P_{n,\vartheta^0}\left(\inf_{\vartheta \in \overline{N}_{n,\gamma}(\vartheta_1^0,\delta)\times\Theta_2} \mathcal{L}_n^{(h)}(\vartheta) - \mathcal{L}_n^{(h)}(\vartheta^0) > 0\right) = 1, \tag{5.35}$$

where the complement of $N_{n,\gamma}(\vartheta_1^0, \delta)$ is naturally given by

$$\overline{N}_{n,\gamma}(\vartheta_1^0, \delta) := \left\{\vartheta_1 \in \Theta_1 : \|\vartheta_1 - \vartheta_1^0\| \geq \delta n^{-\gamma}\right\}. \tag{5.36}$$

Let us now begin with the first step, where we show the consistency of the long-run parameter estimator.

Step 1:
First, we consider the special case, where $\gamma = 0$ in condition (5.35), i.e. for brevity we write for the set (5.36) shortly $\overline{N}_{n,0}(\vartheta_1^0, \delta) = \overline{B}(\vartheta_1^0, \delta)$. Using the fact that $\mathcal{L}_{n,1}^{(h)}(\vartheta_1^0, \vartheta_2) = 0$, we obtain

$$\begin{aligned}\inf_{\vartheta\in\overline{B}(\vartheta_1^0,\delta)\times\Theta_2} \mathcal{L}_n^{(h)}(\vartheta) - \mathcal{L}_n^{(h)}(\vartheta^0) &\geq \inf_{\vartheta\in\overline{B}(\vartheta_1^0,\delta)\times\Theta_2} \mathcal{L}_{n,1}^{(h)}(\vartheta) + \inf_{\vartheta\in\Theta_2} \left(\mathcal{L}_{n,2}^{(h)}(\vartheta_2) - \mathcal{L}_{n,2}^{(h)}(\vartheta_2^0)\right) \\ &= \inf_{\vartheta\in\overline{B}(\vartheta_1^0,\delta)\times\Theta_2} \mathcal{L}_{n,1}^{(h)}(\vartheta) + o_p(1). \qquad (5.37)\end{aligned}$$

Note that $\inf_{\vartheta\in\Theta_2}(\mathcal{L}_{n,2}^{(h)}(\vartheta_2) - \mathcal{L}_{n,2}^{(h)}(\vartheta_2^0)) = o_p(1)$ follows by the convergence result in Proposition 5.5.1 and the global minimum at ϑ_2^0 as shown in Lemma 5.5.2 (c.f. Saikkonen [86], Section 5.3).

We need to show that there exists for every $\delta > 0$ a constant $c > 0$ such that

$$\lim_{n\to\infty} P_{n,\vartheta_0}\left(\inf_{\vartheta\in\overline{B}(\vartheta_1^0,\delta)\times\Theta_2} \mathcal{L}_{n,1}^{(h)}(\vartheta) \geq c\right) = 1. \qquad (5.38)$$

Note that we can write the terms inside the norm as a trace and cyclical permutate the matrices. We will further use the submultiplicativity of the norm and Lemma 5.9.1 iii). We have for $\mathcal{L}_{n,1}^{(h)}(\vartheta)$, using the representation (5.11b), the following lower bound

$$\begin{aligned}\mathcal{L}_{n,1}^{(h)}(\vartheta) &\geq \left\|\frac{1}{n}\sum_{k=1}^n \left(V_\vartheta^{(h)}\right)^{-1}\varepsilon_{k,1}^{(h)}(\vartheta)\varepsilon_{k,1}^{(h)}(\vartheta)^\mathsf{T}\right\| - \left|\log\det V_\vartheta^{(h)}\right| - \left|\log\det V_{\vartheta_2}^{(h)}\right| \\ &\quad - \left\|\frac{2}{n}\sum_{k=1}^n \left(V_\vartheta^{(h)}\right)^{-1}\varepsilon_{k,1}^{(h)}(\vartheta)\varepsilon_{k,2}^{(h)}(\vartheta_2)^\mathsf{T}\right\| - \left\|\frac{1}{n}\sum_{k=1}^n \left(V_\vartheta^{(h)}\right)^{-1}\varepsilon_{k,2}^{(h)}(\vartheta_2)\varepsilon_{k,2}^{(h)}(\vartheta_2)^\mathsf{T}\right\| \\ &\quad - \left\|\frac{1}{n}\sum_{k=1}^n \left(V_{\vartheta_2}^{(h)}\right)^{-1}\varepsilon_{k,2}^{(h)}(\vartheta_2)\varepsilon_{k,2}^{(h)}(\vartheta_2)^\mathsf{T}\right\| \\ &\geq \sigma_{min}\left(\left(V_\vartheta^{(h)}\right)^{-1}\right)\left\|\frac{1}{n}\sum_{k=1}^n \varepsilon_{k,1}^{(h)}(\vartheta)\varepsilon_{k,1}^{(h)}(\vartheta)^\mathsf{T}\right\| - \left|\log\det V_\vartheta^{(h)}\right| - \left|\log\det V_{\vartheta_2}^{(h)}\right| \\ &\quad - \left\|\left(V_\vartheta^{(h)}\right)^{-1}\right\|\left(\left\|\frac{2}{n}\sum_{k=1}^n \varepsilon_{k,1}^{(h)}(\vartheta)\varepsilon_{k,2}^{(h)}(\vartheta_2)^\mathsf{T}\right\| + \left\|\frac{1}{n}\sum_{k=1}^n \varepsilon_{k,2}^{(h)}(\vartheta_2)\varepsilon_{k,2}^{(h)}(\vartheta_2)^\mathsf{T}\right\|\right) \\ &\quad - \left\|\left(V_{\vartheta_2}^{(h)}\right)^{-1}\right\|\left\|\frac{1}{n}\sum_{k=1}^n \varepsilon_{k,2}^{(h)}(\vartheta_2)\varepsilon_{k,2}^{(h)}(\vartheta_2)^\mathsf{T}\right\| \qquad (5.39a) \\ &\geq c\cdot\left\|\frac{1}{n}\sum_{k=1}^n \varepsilon_{k,1}^{(h)}(\vartheta)\varepsilon_{k,1}^{(h)}(\vartheta)^\mathsf{T}\right\| - c\cdot\left\|\frac{1}{n}\sum_{k=1}^n \varepsilon_{k,1}^{(h)}(\vartheta)\varepsilon_{k,2}^{(h)}(\vartheta_2)^\mathsf{T}\right\|\end{aligned}$$

$$- c \cdot \left\| \frac{1}{n} \sum_{k=1}^{n} \varepsilon_{k,2}^{(h)}(\vartheta_2) \varepsilon_{k,2}^{(h)}(\vartheta_2)^\mathsf{T} \right\| - c_1, \tag{5.39b}$$

for some constants $c, c_1 > 0$. The inequality including the minimal singular value σ_{min} of the non-singular matrix $\left(V_\vartheta^{(h)}\right)^{-1}$ follows by Bernstein [9], Corollary 9.6.7. Lastly, due to the compact parameter space, entailed by Assumption M5, and the continuity of $V_\vartheta^{(h)}$, we can also bound the log det terms and the minimal singular value. All in all, we see that the asymptotic behavior depends only on the parts, where combinations of $\varepsilon_2^{(h)}(\vartheta_2)$ and $\varepsilon_1^{(h)}(\vartheta)$ appear.

Let us consider all three terms in (5.39b) separately. We begin with the last one, which is stationary for all $\vartheta_2 \in \Theta_2$. We obtain with Theorem 5.4.4 and the form of $(\varepsilon_{k,2}^{(h)}(\vartheta_2))$ given in (5.10) the continuous convergence in probability, namely $\left\| \frac{1}{n} \sum_{k=1}^{n} \varepsilon_{k,2}^{(h)}(\vartheta_2) \varepsilon_{k,2}^{(h)}(\vartheta_2)^\mathsf{T} \right\| = \mathcal{O}_p(1)$.

The second term in (5.39b) is a combination of a stationary and a non-stationary process. Hence, we have to use additionally the second result of Theorem 5.4.5 to obtain the desired convergence result. Recall the representation of the pseudo-innovations $\varepsilon_1^{(h)}(\vartheta)$ and $\varepsilon_2^{(h)}(\vartheta_2)$ given in (5.10) and note that $\Pi(\vartheta_1^0, \vartheta_2) Y_k^{(h)} = \Pi(\vartheta_1^0, \vartheta_2) Y_{k,2}^{(h)}$. Therefore, we obtain the lower bound for the mixed term

$$\begin{aligned}
&\left\| \frac{1}{n} \sum_{k=1}^{n} \varepsilon_{k,1}^{(h)}(\vartheta) \varepsilon_{k,2}^{(h)}(\vartheta_2)^\mathsf{T} \right\| \\
&\geq \left\| \left[\Pi(\vartheta_1, \vartheta_2) - \Pi(\vartheta_1^0, \vartheta_2) \right] \frac{1}{n} \sum_{k=1}^{n} Y_{k-1}^{(h)} Y_{k-1,2}^{(h)\mathsf{T}} \Pi(\vartheta_1^0, \vartheta_2)^\mathsf{T} \right\| \\
&\quad - \left\| \left[\Pi(\vartheta_1, \vartheta_2) - \Pi(\vartheta_1^0, \vartheta_2) \right] \frac{1}{n} \sum_{k=1}^{n} Y_{k-1}^{(h)} \Delta Y_k^{(h)\mathsf{T}} \overline{k}(B, \vartheta_1^0, \vartheta_2)^\mathsf{T} \right\| \\
&\quad - \left\| \frac{1}{n} \sum_{k=1}^{n} \left(\overline{k}(B, \vartheta_1, \vartheta_2) - \overline{k}(B, \vartheta_1^0, \vartheta_2) \right) \Delta Y_k^{(h)} \Delta Y_k^{(h)\mathsf{T}} \overline{k}(B, \vartheta_1^0, \vartheta_2)^\mathsf{T} \right\| \\
&\quad - \left\| \frac{1}{n} \sum_{k=1}^{n} \left(\overline{k}(B, \vartheta_1, \vartheta_2) - \overline{k}(B, \vartheta_1^0, \vartheta_2) \right) \Delta Y_k^{(h)} Y_{k-1,2}^{(h)\mathsf{T}} \Pi(\vartheta_1^0, \vartheta_2)^\mathsf{T} \right\|.
\end{aligned} \tag{5.40}$$

The last two terms are stationary and hence we can use Theorem 5.4.4 and for the first two terms we use Theorem 5.4.5. Consequently, we have also for the mixed term $\left\| \frac{1}{n} \sum_{k=1}^{n} \varepsilon_{k,1}^{(h)}(\vartheta) \varepsilon_{k,2}^{(h)}(\vartheta_2)^\mathsf{T} \right\| = \mathcal{O}_p(1)$.

Ultimately, we have to deal with the first term in (5.39b) including a product of two non-stationary processes. As before, we derive a lower bound given by

$$\begin{aligned}
&\left\| \frac{1}{n}\sum_{k=1}^{n} \varepsilon_{k,1}^{(h)}(\vartheta)\varepsilon_{k,1}^{(h)}(\vartheta)^{\mathsf{T}} \right\| \\
\geq &\left\| \frac{1}{n}\sum_{k=1}^{n} \left[\Pi(\vartheta_1,\vartheta_2)-\Pi(\vartheta_1^0,\vartheta_2)\right] Y_{k-1}^{(h)}Y_{k-1}^{(h)\mathsf{T}} \left[\Pi(\vartheta_1,\vartheta_2)-\Pi(\vartheta_1^0,\vartheta_2)\right]^{\mathsf{T}} \right\| \\
&- \left\| \frac{1}{n}\sum_{k=1}^{n} (\overline{k}(B,\vartheta_1,\vartheta_2)-\overline{k}(B,\vartheta_1^0,\vartheta_2))\Delta Y_k^{(h)}\Delta Y_k^{(h)\mathsf{T}} \left(\overline{k}(B,\vartheta_1,\vartheta_2)-\overline{k}(B,\vartheta_1^0,\vartheta_2)\right)^{\mathsf{T}} \right\| \\
&- \left\| \frac{2}{n}\sum_{k=1}^{n} \left[\Pi(\vartheta_1,\vartheta_2)-\Pi(\vartheta_1^0,\vartheta_2)\right] Y_{k-1}^{(h)}\Delta Y_k^{(h)\mathsf{T}} (\overline{k}(B,\vartheta_1,\vartheta_2)-\overline{k}(B,\vartheta_1^0,\vartheta_2))^{\mathsf{T}} \right\|.
\end{aligned} \tag{5.41}$$

The first term is the dominant term, which we have to investigate in more detail. In contrast, the second and third terms are of order $\mathcal{O}_p(1)$ by Theorem 5.4.4 and Theorem 5.4.5 respectively.

Finally, let us now take a closer look on the dominant term in (5.41). Define $\lambda_{\min}(S(n))$ as the smallest eigenvalue of

$$S(n) := \frac{1}{n^2}\sum_{k=1}^{n}\sum_{i=1}^{k-1} R_{i,1}^{(h)} \sum_{i=1}^{k-1} R_{i,1}^{(h)\mathsf{T}},$$

where $R_{i,1}^{(h)\mathsf{T}}$ is as in (4.20). Note that $C_1 S(n) C_1^T$ is the dominant part of the term $\frac{1}{n^2}\sum_{k=1}^{n} Y_{k-1}^{(h)}Y_{k-1}^{(h)\mathsf{T}}$ due to the results in the proof of Proposition 5.8.4 iii). Furthermore, we have that $S(n)$ converges weakly to $B_1\int_0^1 W(r)W(r)^{\mathsf{T}}\,\mathrm{d}r B_1^{\mathsf{T}}$.

Hence, we have a lower bound, using again Bernstein [9], Corollary 9.6.7. and Proposition 5.8.4 iii), given by

$$\begin{aligned}
&\left\| \left[\Pi(\vartheta_1,\vartheta_2)-\Pi(\vartheta_1^0,\vartheta_2)\right] \frac{1}{n^2}\sum_{k=1}^{n} Y_{k-1}^{(h)}Y_{k-1}^{(h)\mathsf{T}} \left[\Pi(\vartheta_1,\vartheta_2)-\Pi(\vartheta_1^0,\vartheta_2)\right]^{\mathsf{T}} \right\| \\
&= \left\| \Pi(\vartheta_1,\vartheta_2)\frac{1}{n^2}\sum_{k=1}^{n} Y_{k-1}^{(h)}Y_{k-1}^{(h)\mathsf{T}}\Pi(\vartheta_1,\vartheta_2)^{\mathsf{T}} \right\| + \mathcal{O}_p\left(\frac{1}{n}\right) \\
&= \left\| C_1^{\mathsf{T}}\Pi(\vartheta)^{\mathsf{T}}\Pi(\vartheta)C_1 S(n) \right\| + \mathcal{O}_p\left(\frac{1}{n}\right) \\
&\geq \lambda_{\min}(S(n))\|\Pi(\vartheta)C_1\|^2 + \mathcal{O}_p\left(\frac{1}{n}\right),
\end{aligned} \tag{5.42}$$

where the sequence $\lambda_{\min}(S(n))$ converges weakly to the smallest eigenvalue of the almost surely positive-definite matrix $\int_0^1 B_1 W(r)W(r)^\mathsf{T} B_1^\mathsf{T}\,\mathrm{d}r$ (c.f. for example Johansen and Schaumburg [55]). Furthermore, the infimum of $\mathcal{L}_{n,1}^{(h)}(\vartheta)$ is taken over a set where $\vartheta_1 \in \overline{B}(\vartheta_1^0, \delta)$, thus by Lemma 5.3.1 we obtain $\|\Pi(\vartheta)C_1\|^2 \geq d_\delta$, for some constant $d_\delta > 0$. All in all, we obtain

$$\inf_{\vartheta \in \overline{B}(\vartheta_1^0,\delta)\times\Theta_2} \mathcal{L}_{n,1}^{(h)}(\vartheta) \geq n \cdot d_\delta \lambda_{min}(S(n)) - \mathcal{O}_p(1),$$

which tends to infinity at the rate $\mathcal{O}_p(n)$ and thus we have shown (5.38). In other words, we have shown that for a fixed lower bound for $\|\vartheta_1 - \vartheta_1^0\|$ and arbitrary ϑ_2 the infimum of the log-likelihood function tends to infinity. By doing this, we achieve consistency of the long-run parameter estimator.

Step 2:

Since the result from the first step is insufficient to show the consistency of the short-run parameter we determine now the consistency rate of the long-run parameter estimator. Thus, we use the complement of the set (5.34) without the restriction of $\gamma = 0$ in order to determine the consistency rate. In consequence, we consider the sufficient condition (5.35) with $0 < \gamma < 1$ and have

$$\begin{aligned}
&\inf_{\vartheta \in \overline{N}_{n,\gamma}(\vartheta_1^0,\delta)\times\Theta_2} n \cdot \left(\mathcal{L}_n^{(h)}(\vartheta) - \mathcal{L}_n^{(h)}(\vartheta^0)\right) \\
&\geq \inf_{\vartheta \in \overline{N}_{n,\gamma}(\vartheta_1^0,\delta)\times\Theta_2} n \cdot \mathcal{L}_{n,1}^{(h)}(\vartheta) + \inf_{\vartheta\in\Theta_2} n \cdot \left(\mathcal{L}_{n,2}^{(h)}(\vartheta_2) - \mathcal{L}_{n,2}^{(h)}(\vartheta_2^0)\right) \\
&= \inf_{\vartheta \in \overline{N}_{n,\gamma}(\vartheta_1^0,\delta)\times\Theta_2} n \cdot \mathcal{L}_{n,1}^{(h)}(\vartheta) + \mathcal{O}_p(1). \qquad (5.43)
\end{aligned}$$

In order to see the last equality we take an infeasible estimator $\widehat{\vartheta}_2^{st}$ for ϑ_2^0, which is the QML-estimator of Schlemm and Stelzer [91] minimizing $\mathcal{L}_{n,2}^{(h)}(\vartheta_2)$. This estimator is infeasible since it is only an estimator for a stationary MCARMA process Y_2. Note that $\mathcal{L}_{n,2}^{(h)}(\vartheta_2)$ depends only on the short-run parameter and the true long-run parameter value is already inserted in the log-likelihood function. For this reason, we can interpret this as a „classical" stationary estimation problem. However, this estimator is not applicable for this setting as a suitable estimator. We know that $\widehat{\vartheta}_{n,2}^{st} - \vartheta_2^0 = \mathcal{O}_p(n^{-\frac{1}{2}})$ and the score vector is asymptotically normal due to Theorem 3.16 in Schlemm and Stelzer [91]. Thus, applying a mean value expansion yields $n \cdot \left(\mathcal{L}_{n,2}^{(h)}(\widehat{\vartheta}_2^{st}) - \mathcal{L}_{n,2}^{(h)}(\vartheta_2^0)\right) = \left(\sqrt{n}\nabla\mathcal{L}_{n,2}^{(h)}(\underline{\vartheta}_{n,2})\right) \cdot \left(\sqrt{n}(\widehat{\vartheta}_{n,2}^{st} - \vartheta_2)\right) = \mathcal{O}_p(1)$ for an appropriate intermediate value $\underline{\vartheta}_2$ (c.f. also Saikkonen [86], p. 904).

The last line in (5.43) tends to infinity in probability if we can show for every $\delta > 0$

and every $d > 0$,

$$\lim_{n\to\infty} P_{n,\vartheta_0}\left(\inf_{\vartheta\in\overline{N}_{n,\gamma}(\vartheta_1^0,\delta)\times\Theta_2} n\cdot\mathcal{L}_{n,1}^{(h)}(\vartheta)\geq d\right) = 1. \tag{5.44}$$

To show this we first note that due to the results in the first step we only have to consider for n large enough the set

$$\overline{M}_{n,\gamma}(\vartheta_1^0,\delta_1) = \overline{N}_{n,\gamma}(\vartheta_1^0,\delta_1)\cap B(\vartheta_1^0,\delta_1)\subset B(\vartheta_1^0,\delta_1)$$

instead of the complete set $\overline{N}_{n,\gamma}(\vartheta_1^0,\delta_1)$.

We need another assumption, this time on the gradient of the matrix $\Pi(\cdot)$. A rank condition on the gradient is not only relevant for the proof of the consistency, but it is also central for the derivation of the asymptotic distribution in the next section.

Assumption M12
Assume that the $(d^2\times s_1)$- dimensional gradient matrix $\nabla_{\vartheta_1}\left(\Pi(\vartheta_1^0,\vartheta_2)^\mathsf{T}\right)$ has full column rank s_1 for all $\vartheta_2\in\Theta_2$.

We need the following result in order to know the exact speed of convergence of the subspace to the cointegration space. With the following lemma we are able to determine the consistency rate of $\widehat{\vartheta}_{n,1}$ afterwards.

Lemma 5.5.3
Assume that Assumption M2-Assumption M9 and Assumption M12 hold. For δ_1 small there exists a constant $c > 0$ such that

$$\inf_{(\vartheta_1,\vartheta_2)\in\overline{M}_{n,\gamma}(\vartheta_1^0,\delta_1)\times\Theta_2}\|\Pi(\vartheta)C_1\|\geq c\cdot\delta_1 n^{-\gamma}.$$

Proof. Applying a mean value expansion will prove the claim. First, we have due to Assumption M8 that $\Pi(\cdot)$ is continuously differentiable and Assumption M12 guarantees for δ_1 small enough that the gradient $\nabla_{\vartheta_1}\Pi(\vartheta)$ is of full column rank. Hence, the smallest eigenvalue of $\nabla_{\vartheta_1}\Pi(\vartheta)^\mathsf{T}\nabla_{\vartheta_1}\Pi(\vartheta)$ is bounded away from zero and accordingly the smallest singular value of $\nabla_{\vartheta_1}\Pi(\vartheta)$ for δ_1 small enough.

Recall the facts about the Frobenius norm, the vec operator and the Kronecker product in Appendix B. Hence, $\left(C_1^\mathsf{T}\otimes I_d\right)$ has dimension $dr\times d^2$ and rank dr. Then, the smallest singular value σ_{min} of $\left(C_1^\mathsf{T}\otimes I_d\right)$ is positive by Bernstein [9], Equation (5.6.4).

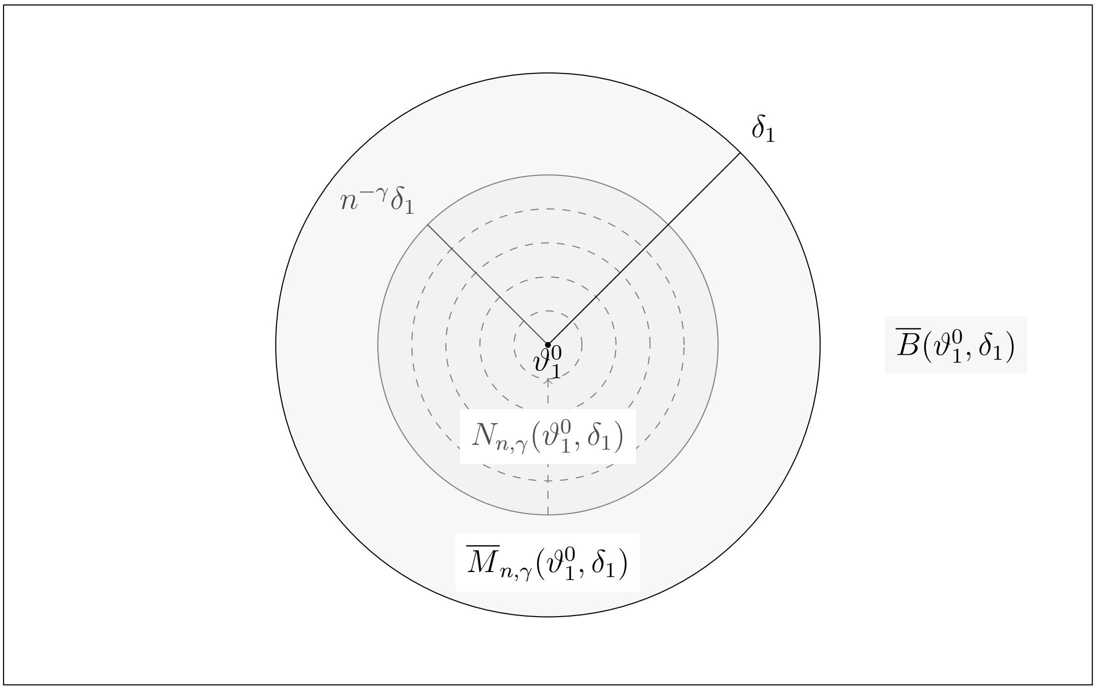

Figure 5.1.: Illustration of considered sets for the parameter ϑ_1.

Eventually, using a mean value expansion and Bernstein [9], Corollary 9.6.7 twice, leads to

$$\begin{aligned}
\left\|\Pi(\vartheta)C_1\right\| &= \left\|\mathrm{vec}(\Pi(\vartheta)C_1)\right\| \\
&= \left\| \underbrace{\mathrm{vec}\left(\Pi(\vartheta_1^0, \vartheta_2)C_1\right)}_{=0_{dr}} + \nabla_{\vartheta_1}\left(\Pi(\underline{\vartheta_1}, \vartheta_2)C_1\right)(\vartheta_1 - \vartheta_1^0) \right\| \\
&= \left\|\left(C_1^\mathsf{T} \otimes I_d\right)\nabla_{\vartheta_1}\left(\Pi(\underline{\vartheta_1}, \vartheta_2)\right)(\vartheta_1 - \vartheta_1^0)\right\| \\
&\geq \sigma_{min}\left(C_1^\mathsf{T} \otimes I_d\right)\sigma_{min}\left(\nabla_{\vartheta_1}\Pi(\underline{\vartheta_1}, \vartheta_2)\right)\|\vartheta_1 - \vartheta_1^0\| \\
&\geq c\|\vartheta_1 - \vartheta_1^0\| \geq c \cdot \delta n^{-\gamma},
\end{aligned} \tag{5.45}$$

for vectors $\underline{\vartheta}_{1,i} \in \Theta$ of the form $\underline{\vartheta}_{1,i} = \vartheta_1^0 + c_i(\widehat{\vartheta}_1 - \vartheta_1^0)$, $0 \leq c_i \leq 1$ such that $\nabla_{\vartheta_1}\Pi(\underline{\vartheta_1}, \vartheta_2)$ denotes the matrix, whose i^{th} row, for $i = 1, \ldots, s_1$, is equal to the i^{th} row of $\nabla_{\vartheta_1}\Pi(\underline{\vartheta}_{1,i}, \vartheta_2)$. □

Besides, due to the Lipschitz continuity of $\Pi(\cdot)$ we have in the set $\overline{M}_{n,\gamma}(\vartheta_1^0, \delta_1)$ the following upper bound

$$\left\|\Pi(\vartheta_1, \vartheta_2) - \Pi(\vartheta_1^0, \vartheta_2)\right\| \leq c_\Pi \left\|\vartheta_1 - \vartheta_1^0\right\| \leq c_\Pi \delta_1.$$

Knowing the speed of convergence of the subspace, we can relate this directly to the speed of convergence of $\vartheta_{n,1}$ to the true long-run parameter value. We have seen that it converges up to a constant with the same speed as $n \to \infty$ in the set $\overline{M}_{n,\gamma}(\vartheta_1^0, \delta_1)$.

Lemma 5.5.4

Assume that Assumption M2-Assumption M9 and Assumption M12 hold. Then we have as $n \to \infty$

$$\sup_{\vartheta_1 \in \overline{M}_{n,\gamma}(\vartheta_1^0,\delta)\times\Theta_2} \frac{\|\vartheta_1 - \vartheta_1^0\|}{n \cdot \|\Pi(\vartheta)C_1\|^2} = o(1). \tag{5.46}$$

Proof. First, we know that $\sup_{\vartheta_1 \in \overline{M}_{n,\gamma}(\vartheta_1^0,\delta)\times\Theta_2}(n\cdot\|\Pi(\vartheta)C_1\|)^{-1} = o(1)$ for all $0 \leq \gamma < 1$ by Lemma 5.5.3. Again, we have by using a intermediate step of Equation 5.45

$$\sup_{\vartheta_1 \in \overline{M}_{n,\gamma}(\vartheta_1^0,\delta)\times\Theta_2} \frac{\|\vartheta_1 - \vartheta_1^0\|}{\|\Pi(\vartheta)C_1\|} \leq \sup_{\vartheta_1 \in \overline{M}_{n,\gamma}(\vartheta_1^0,\delta)} c \cdot \frac{\|\vartheta_1 - \vartheta_1^0\|}{\|\vartheta_1 - \vartheta_1^0\|} < \infty,$$

for some positive constant c. □

As in Step 1 we have the lower bound of $\mathcal{L}_{n,1}^{(h)}(\vartheta)$ given by (5.39). We use similar derivations as in (5.40), (5.41) and (5.42) to further improve this lower bound. Moreover, we make use of the Lipschitz continuity of $\Pi(\cdot)$, $\overline{k}(z,\cdot)$ and $V_{(\cdot)}$, which finally leads to using Lemma 5.5.3

$$\begin{aligned}
n \cdot \mathcal{L}_{n,1}^{(h)}(\vartheta) &\overset{(5.39a)}{\geq} n \cdot c \cdot \left\| \frac{1}{n}\sum_{k=1}^{n} \varepsilon_{k,1}^{(h)}(\vartheta)\varepsilon_{k,1}^{(h)}(\vartheta)^{\mathsf{T}} \right\| - n \cdot c \cdot \left\| \frac{1}{n}\sum_{k=1}^{n} \varepsilon_{k,1}^{(h)}(\vartheta)\varepsilon_{k,2}^{(h)}(\vartheta_2)^{\mathsf{T}} \right\| \\
&\quad - n \cdot \left\| \left(V_{\vartheta}^{(h)}\right)^{-1} - \left(V_{\vartheta_2}^{(h)}\right)^{-1} \right\| \left\| \frac{1}{n}\sum_{k=1}^{n} \varepsilon_{k,2}^{(h)}(\vartheta_2)\varepsilon_{k,2}^{(h)}(\vartheta_2)^{\mathsf{T}} \right\| \\
&\quad - n \left| \log\det V_{\vartheta}^{(h)} - \log\det V_{\vartheta_2}^{(h)} \right| \\
&\geq n^2 \cdot c \Bigg[-\frac{\|\Pi(\vartheta_1,\vartheta_2) - \Pi(\vartheta_1^0,\vartheta_2)\|}{n} \underbrace{\left(\left\| \frac{1}{n}\sum_{k=1}^{n} Y_{k-1}^{(h)}\Delta Y_k^{(h)\mathsf{T}} \right\| + \left\| \frac{1}{n}\sum_{k=1}^{n} Y_{k-1}^{(h)} Y_{k-1,2}^{(h)\mathsf{T}} \right\| \right)}_{:=X_{n,1}} \\
&\qquad - \frac{\|\vartheta_1 - \vartheta_1^0\|}{n} \cdot \underbrace{\left(\left\| \frac{1}{n}\sum_{k=1}^{n} \Delta Y_k^{(h)}\Delta Y_k^{(h)\mathsf{T}} \right\| + \left\| \frac{1}{n}\sum_{k=1}^{n} Y_{k-1,2}^{(h)}\Delta Y_k^{(h)\mathsf{T}} \right\| \right)}_{:=X_{n,2}} \\
&\qquad - \frac{\|\vartheta_1 - \vartheta_1^0\|}{n} \underbrace{\left(\left\| \frac{1}{n}\sum_{k=1}^{n} \varepsilon_{k,2}^{(h)}(\vartheta_2)\varepsilon_{k,2}^{(h)}(\vartheta_2)^{\mathsf{T}} \right\| + 1 \right)}_{:=X_{n,3}} \\
&\qquad + \lambda_{\min}(S(n))\|\Pi(\vartheta)C_1\|^2 \Bigg]
\end{aligned}$$

$$\geq n^2 \cdot c\|\Pi(\vartheta)C_1\|^2 \cdot \underbrace{\left(\lambda_{\min}(S(n)) - \frac{c_1 \cdot \|\vartheta_1 - \vartheta_1^0\|}{n \cdot \|\Pi(\vartheta)C_1\|^2}(X_{n,1} + X_{n,2} + X_{n,3}) - o_p(1)\right)}_{:=X_n}$$

$$\geq n^{2-2\gamma} \cdot c \cdot \delta^2 \cdot X_n, \tag{5.47}$$

for some $c, c_1 > 0$ and $\delta > 0$.

Due to Lemma 5.5.4 the fraction in front of $X_{n1} + X_{n2} + X_{n3}$ is of order $o(1)$. Proposition 5.8.4 and Theorem 5.4.4 imply $X_{n1} + X_{n2} + X_{n3} = \mathcal{O}_p(1)$. Note that as previously $\lambda_{\min}(S(n))$ converges weakly to the smallest eigenvalue of the matrix $B_1 \int_0^1 W(r)W(r)^\mathsf{T}\,\mathrm{d}r\ B_1^\mathsf{T}$ due to Proposition 5.8.4 iii). Combining these results further implies that X_n converges to the smallest eigenvalue of the matrix $B_1 \int_0^1 W(r)W(r)^\mathsf{T}\,\mathrm{d}r\ B_1^\mathsf{T}$ in probability, which is almost surely positive definite. Thus, the smallest eigenvalue is almost surely positive. In conclusion, the product also diverges to infinity. Hence, we have shown (5.44), which finally leads to $\widehat{\vartheta}_{n,1} - \vartheta_1^0 = o_p(n^{-\gamma})$, for $0 \leq \gamma < 1$. In difference to the standard approach in the stationary setting, we need the order of consistency due to the different convergence rates depending if the parameters belong to the stationary or non-stationary part.

Step 3:
Next, we consider the consistency of the estimator $\widehat{\vartheta}_{n,2}$ of the short-run parameter with the help of the order of consistency we obtained in the previous step. We show the sufficient condition for consistency

$$\lim_{n\to\infty} P_{n,\vartheta_0}\left(\inf_{\vartheta\in\Theta_1\times\overline{B}(\vartheta_2^0,\delta)} n \cdot \left(\mathcal{L}_n^{(h)}(\vartheta) - \mathcal{L}_n^{(h)}(\vartheta^0)\right) > 0\right) = 1 \tag{5.48}$$

for $\delta > 0$. Let us assume for this step that $\frac{1}{2} < \gamma < 1$. Apparently, the parameter subspace Θ_1 is the union of the following sets $\Theta_1 = \overline{N}_{n,\gamma}(\vartheta_1^0, \delta_1) \cup N_{n,\gamma}(\vartheta_1^0, \delta_1)$ and thus we have already shown (5.48) for the set $\overline{N}_{n,\gamma}(\vartheta_1^0, \delta_1) \times \overline{B}(\vartheta_2^0, \delta)$ in Step 2. This insight enables us to use the convergence rate of the long-run estimator in the following. Hence, for arbitrary $\delta_1 > 0$ we obtain

$$\lim_{n\to\infty} P_{n,\vartheta_0}\left(\inf_{\vartheta\in N_{n,\gamma}(\vartheta_1^0,\delta_1)\times\overline{B}(\vartheta_2^0,\delta)} n \cdot \left(\mathcal{L}_n^{(h)}(\vartheta) - \mathcal{L}_n^{(h)}(\vartheta^0)\right) > 0\right)$$

$$\geq \lim_{n\to\infty} P_{n,\vartheta_0}\left(\inf_{\vartheta\in N_{n,\gamma}(\vartheta_1^0,\delta_1)\times\overline{B}(\vartheta_2^0,\delta)} n\mathcal{L}_{n,1}^{(h)}(\vartheta) + \inf_{\vartheta_2\in\overline{B}(\vartheta_2^0,\delta)} n\left(\mathcal{L}_{n,2}^{(h)}(\vartheta_2) - \mathcal{L}_{n,2}^{(h)}(\vartheta_2^0)\right) > 0\right)$$

$$\geq \lim_{n\to\infty} P_{n,\vartheta_0}\left(\inf_{\vartheta\in N_{n,\gamma}(\vartheta_1^0,\delta_1)\times\overline{B}(\vartheta_2^0,\delta)} |\mathcal{L}_{n,1}^{(h)}(\vartheta)| \leq \epsilon\,;\ \inf_{\vartheta_2\in\overline{B}(\vartheta_2^0,\delta)} \mathcal{L}_{n,2}^{(h)}(\vartheta_2) - \mathcal{L}_{n,2}^{(h)}(\vartheta_2^0) > \epsilon\right).$$

Consequently, we have to prove that the following two conditions hold in order to prove (5.48). First, we need to show that for all $\delta > 0$, every $\epsilon > 0$ and some $\delta_1 > 0$ we have

$$\lim_{n\to\infty} P_{n,\vartheta_0}\left(\sup_{\vartheta\in N_{n,\gamma}(\vartheta_1^0,\delta_1)\times\overline{B}(\vartheta_2^0,\delta)} |\mathcal{L}_{n,1}^{(h)}(\vartheta)| \leq \epsilon\right) = 1 \tag{5.49a}$$

and secondly that for all $\delta > 0$ and some $\epsilon > 0$

$$\lim_{n\to\infty} P_{n,\vartheta_0}\left(\inf_{\vartheta_2\in\overline{B}(\vartheta_2^0,\delta)} \mathcal{L}_{n,2}^{(h)}(\vartheta_2) - \mathcal{L}_{n,2}^{(h)}(\vartheta_2^0) > \epsilon\right) = 1. \tag{5.49b}$$

Eventually, to show (5.49a) we derive an upper bound in a similar way to the lower bound given in (5.39). Note that $\left\|\left(V_\vartheta^{(h)}\right)^{-1}\right\|$ is bounded due to Lemma 5.9.2 iii). As usual make use of the submultiplicativity of the norm and use (5.39a)

$$\begin{aligned}
|\mathcal{L}_{n,1}^{(h)}(\vartheta)| \leq{}& \left|\log\left(\frac{\det V_\vartheta^{(h)}}{\det V_{\vartheta_2}^{(h)}}\right)\right| + \left\|\left(V_\vartheta^{(h)}\right)^{-1}\right\| \left\|\frac{1}{n}\sum_{k=1}^n \varepsilon_{k,1}^{(h)}(\vartheta)\varepsilon_{k,1}^{(h)}(\vartheta)^\mathsf{T}\right\| \\
&+ 2\cdot\left\|\left(V_\vartheta^{(h)}\right)^{-1}\right\| \left\|\frac{1}{n}\sum_{k=1}^n \varepsilon_{k,1}^{(h)}(\vartheta)\varepsilon_{k,2}^{(h)}(\vartheta_2)^\mathsf{T}\right\| \\
&+ \left\|\frac{1}{n}\sum_{k=1}^n \left(\left(V_\vartheta^{(h)}\right)^{-1} - \left(V_{\vartheta_2}^{(h)}\right)^{-1}\right) \varepsilon_{k,2}^{(h)}(\vartheta_2)\varepsilon_{k,2}^{(h)}(\vartheta_2)^\mathsf{T}\right\| \\
\leq{}& \left|\log\left(\frac{\det V_\vartheta^{(h)}}{\det V_{\vartheta_2}^{(h)}}\right)\right| + c\left\|\frac{1}{n}\sum_{k=1}^n \varepsilon_{k,1}^{(h)}(\vartheta)^\mathsf{T}\varepsilon_{k,1}^{(h)}(\vartheta)\right\| \\
&+ \left\|\left(V_\vartheta^{(h)}\right)^{-1} - \left(V_{\vartheta_2}^{(h)}\right)^{-1}\right\| \left\|\frac{1}{n}\sum_{k=1}^n \varepsilon_{k,2}^{(h)}(\vartheta_2)^\mathsf{T}\varepsilon_{k,2}^{(h)}(\vartheta_2)\right\| \\
&+ c\left\|\frac{1}{n}\sum_{k=1}^n \varepsilon_{k,1}^{(h)}(\vartheta)^\mathsf{T}\varepsilon_{k,2}^{(h)}(\vartheta_2)\right\|,
\end{aligned} \tag{5.50}$$

for some constant $c > 0$.

Since $\log\det$ is a smooth function, we can use the Lipschitz continuity to obtain for $\vartheta_1 \in N_{n,\gamma}(\vartheta_1^0, \delta_1)$ (c.f. the definition of the set in (5.34))

$$\left|\log\left(\det V_\vartheta^{(h)}\right) - \log\left(\det V_{\vartheta_2}^{(h)}\right)\right| \leq c\|\vartheta_1 - \vartheta_1^0\| \leq c_1 n^{-\gamma},$$

for some constant $c, c_1 > 0$. Therefore, we have this upper bound depending on n. Alike, we use the Lipschitz continuity of the functions $\overline{k}(1,\cdot)$, $\Pi(\cdot)$ and $\left(V_{(\cdot)}^{(h)}\right)^{-1}$ and the submultiplicativity of the norm.

Finally, due to the representation given in (5.87) and the fact that $\Pi(\vartheta_1^0,\vartheta_2)Y_{k-1}^{(h)} = \Pi(\vartheta_1^0,\vartheta_2)Y_{k-1,2}^{(h)}$, we obtain the following upper bound

$$\begin{aligned}
&\sup_{\vartheta\in N_{n,\gamma}(\vartheta_1^0,\delta_1)\times\overline{B}(\vartheta_2^0,\delta)} |\mathcal{L}_{n,1}^{(h)}(\vartheta)| \\
&\overset{(5.50)}{\leq} \left|\log\left(\frac{\det V_\vartheta^{(h)}}{\det V_{\vartheta_2}^{(h)}}\right)\right| + c\left\|\frac{1}{n}\sum_{k=1}^n \varepsilon_{k,1}^{(h)}(\vartheta)^\mathsf{T}\varepsilon_{k,1}^{(h)}(\vartheta)\right\| + c\left\|\frac{1}{n}\sum_{k=1}^n \varepsilon_{k,1}^{(h)}(\vartheta)^\mathsf{T}\varepsilon_{k,2}^{(h)}(\vartheta_2)\right\| \\
&\quad + \left\|\left(V_\vartheta^{(h)}\right)^{-1} - \left(V_{\vartheta_2}^{(h)}\right)^{-1}\right\|\left\|\frac{1}{n}\sum_{k=1}^n \varepsilon_{k,2}^{(h)}(\vartheta_2)^\mathsf{T}\varepsilon_{k,2}^{(h)}(\vartheta_2)\right\| \\
&\leq c_1\cdot n^{-\gamma} + c_1\|\vartheta_1-\vartheta_1^0\|^2\left\|\frac{1}{n}\sum_{k=1}^n Y_{k-1}^{(h)}Y_{k-1}^{(h)\mathsf{T}}\right\| + c_1\|\vartheta_1-\vartheta_1^0\|^2\left\|\frac{1}{n}\sum_{k=1}^n Y_{k-1}^{(h)}\Delta Y_k^{(h)\mathsf{T}}\right\| \\
&\quad + c_1\|\vartheta_1-\vartheta_1^0\|^2\left\|\frac{1}{n}\sum_{k=1}^n \Delta Y_k^{(h)}\Delta Y_k^{(h)\mathsf{T}}\right\| + c_1\|\vartheta_1-\vartheta_1^0\|\left\|\frac{1}{n}\sum_{k=1}^n Y_{k-1}^{(h)}Y_{k-1,2}^{(h)\mathsf{T}}\right\| \\
&\quad + c_1\|\vartheta_1-\vartheta_1^0\|\left\|\frac{1}{n}\sum_{k=1}^n \Delta Y_k^{(h)}\Delta Y_k^{(h)\mathsf{T}}\right\| + c_1\|\vartheta_1-\vartheta_1^0\|\left\|\frac{1}{n}\sum_{k=1}^n Y_{k-1,2}^{(h)}\Delta Y_k^{(h)\mathsf{T}}\right\| \\
&\quad + c_1\|\vartheta_1-\vartheta_1^0\|\left\|\frac{1}{n}\sum_{k=1}^n Y_{k-1}^{(h)}\Delta Y_k^{(h)\mathsf{T}}\right\| + c_1\|\vartheta_1-\vartheta_1^0\|\left\|\frac{1}{n}\sum_{k=1}^n Y_{k-1,2}^{(h)}Y_{k-1,2}^{(h)\mathsf{T}}\right\| \\
&\leq c_1 n^{-\gamma} + c_1\left\|n^{-1-2\gamma}\sum_{k=1}^n Y_{k-1}^{(h)}Y_{k-1}^{(h)\mathsf{T}}\right\| + c_1\left\|n^{-1-2\gamma}\sum_{k=1}^n Y_{k-1}^{(h)}\Delta Y_k^{(h)\mathsf{T}}\right\| \\
&\quad + c_1\left\|n^{-1-2\gamma}\sum_{k=1}^n \Delta Y_k^{(h)}\Delta Y_k^{(h)\mathsf{T}}\right\| + c_1\left\|n^{-1-\gamma}\sum_{k=1}^n Y_{k-1}^{(h)}Y_{k-1,2}^{(h)\mathsf{T}}\right\| \\
&\quad + c_1\left\|n^{-1-\gamma}\sum_{k=1}^n \Delta Y_k^{(h)}\Delta Y_k^{(h)\mathsf{T}}\right\| + c_1\left\|n^{-1-\gamma}\sum_{k=1}^n Y_{k-1,2}^{(h)}\Delta Y_k^{(h)\mathsf{T}}\right\| \\
&\quad + c_1\left\|n^{-1-\gamma}\sum_{k=1}^n Y_{k-1}^{(h)}\Delta Y_k^{(h)\mathsf{T}}\right\| + c_1\left\|n^{-1-\gamma}\sum_{k=1}^n Y_{k-1,2}^{(h)}Y_{k-1,2}^{(h)\mathsf{T}}\right\|,
\end{aligned}$$

for some $c_1 > 0$. The different exponents arising depend on how often we used the Lipschitz continuity per summand. The second term of the right-hand side is of order $o_p(1)$ for $\gamma > \frac{1}{2}$. The remaining terms are all of $o_p(1)$ for every $\gamma > 0$ due to Proposition 5.8.4.

We have shown (5.49a) and only the second and final part (5.49b) is left to be considered. In order to show the condition (5.49b) note that Proposition 5.5.1 yields $\mathcal{L}_{n,2}^{(h)}(\vartheta_2) \xrightarrow[c]{p} \boldsymbol{\mathcal{L}}_2^{(h)}(\vartheta_2)$. Therefore, using the ideas of Poetscher and Prucha [80], Proof of Lemma 3.1, we obtain

$$\liminf_{n\to\infty}\inf_{\vartheta_2\in\overline{B}(\vartheta_2^0,\delta)}\left(\mathcal{L}_{n,2}^{(h)}(\vartheta_2) - \mathcal{L}_{n,2}^{(h)}(\vartheta_2^0)\right)$$

$$
\begin{aligned}
&= \liminf_{n\to\infty} \inf_{\vartheta_2 \in \overline{B}(\vartheta_2^0,\delta)} \Big(\mathcal{L}_{n,2}^{(h)}(\vartheta_2) - \boldsymbol{\mathcal{L}}_2^{(h)}(\vartheta_2) - \mathcal{L}_{n,2}^{(h)}(\vartheta_2^0) + \boldsymbol{\mathcal{L}}_2^{(h)}(\vartheta_2^0) + \boldsymbol{\mathcal{L}}_2^{(h)}(\vartheta_2) - \boldsymbol{\mathcal{L}}_2^{(h)}(\vartheta_2^0) \Big) \\
&\geq \liminf_{n\to\infty} \inf_{\vartheta_2 \in \overline{B}(\vartheta_2^0,\delta)} \Big(\mathcal{L}_{n,2}^{(h)}(\vartheta_2) - \boldsymbol{\mathcal{L}}_2^{(h)}(\vartheta_2) \Big) + \liminf_{n\to\infty} \Big(-\mathcal{L}_{n,2}^{(h)}(\vartheta_2^0) + \boldsymbol{\mathcal{L}}_2^{(h)}(\vartheta_2^0) \Big) \\
&\quad + \inf_{\vartheta_2 \in \overline{B}(\vartheta_2^0,\delta)} \Big(\boldsymbol{\mathcal{L}}_2^{(h)}(\vartheta_2) - \boldsymbol{\mathcal{L}}_2^{(h)}(\vartheta_2^0) \Big) \\
&> \epsilon, \qquad\qquad \mathbb{P} - \text{a.s.} \qquad (5.51)
\end{aligned}
$$

where the constant $\epsilon > 0$ is determined by $\inf_{\vartheta_2 \in \overline{B}(\vartheta_2^0,\delta)} \Big(\boldsymbol{\mathcal{L}}_2^{(h)}(\vartheta_2) - \boldsymbol{\mathcal{L}}_2^{(h)}(\vartheta_2^0) \Big)$ since $\boldsymbol{\mathcal{L}}_2^{(h)}(\vartheta_2)$ has its unique minimum at ϑ_2^0 due to Lemma 5.5.2.

In summary, we want to collect the obtained results of this section into one theorem, which is the main result of this section.

Theorem 5.5.5
Let Assumption M1-Assumption M12 hold. Then we obtain the super-consistency of the long-run parameter estimator $\widehat{\vartheta}_{n,1}$, i.e.

$$\widehat{\vartheta}_{n,1} - \vartheta_1^0 = o_p(n^{-\gamma}), \quad \text{for all } 0 \leq \gamma < 1, \tag{5.52a}$$

and the consistency of short-run parameter estimator $\widehat{\vartheta}_{n,2}$, i.e.

$$\widehat{\vartheta}_{n,2} - \vartheta_2^0 = o_p(1). \tag{5.52b}$$

5.6. Asymptotic Distributions

The aim of this section is to derive the asymptotic distributions of the short-run parameter estimator and the long-run parameter estimator. These two estimators have a different asymptotic behavior and different convergence rates. On the one hand, we prove the asymptotic normality of the estimator corresponding to the short-run parameters. However, on the other hand, we show that the estimator corresponding to the long-run parameters is mixed normal distributed.

Assume throughout this section that Assumption M1-Assumption M12 always hold.

5.6.1. Asymptotic Distribution of the Long-run Estimator

We derive in this section the asymptotic distribution of the long-run QMLE using a mean-value expansion. Thus, we consider first the asymptotic behaviour of the score vector and then of the Hessian matrix. First, we show the convergence to a stochastic integral of the gradient with respect to the non-stationary parameters. The partial derivatives with respect to i^{th}-component of the parameter vector ϑ, for $i = 1, \ldots, s$ of the log-likelihood function (5.6) are given by

$$\partial_i \mathcal{L}_n^{(h)}(\vartheta) = \frac{1}{n} \sum_{k=1}^{n} \partial_i \ell_{\vartheta,k}^{(h)}, \tag{5.53a}$$

where we obtain with the differentiation rules for matrix functions (c.f. Appendix B.4)

$$\begin{aligned}\partial_i \ell_{\vartheta,k}^{(h)} = \operatorname{tr}\left(\left(V_\vartheta^{(h)}\right)^{-1} \partial_i V_\vartheta^{(h)}\right) - \operatorname{tr}\left(\left(V_\vartheta^{(h)}\right)^{-1} \varepsilon_k^{(h)}(\vartheta) \varepsilon_k^{(h)}(\vartheta)^\mathsf{T} \left(V_\vartheta^{(h)}\right)^{-1} \partial_i V_\vartheta^{(h)}\right) \\ + 2 \cdot \left(\partial_i \varepsilon_k^{(h)}(\vartheta)^\mathsf{T}\right) \left(V_\vartheta^{(h)}\right)^{-1} \varepsilon_k^{(h)}(\vartheta). \end{aligned} \tag{5.53b}$$

Let us now derive the asymptotic behaviour of the gradient with respect to the long-run parameters.

Proposition 5.6.1

For the gradient with respect to the non-stationary parameters we have

$$\nabla_{\vartheta_1} \mathcal{L}_n^{(h)}(\vartheta^0) \xrightarrow{w} \mathcal{J}_1(\vartheta^0), \tag{5.54}$$

where the limit is given by $\mathcal{J}_1(\vartheta^0) := \begin{pmatrix} \mathcal{J}_1^{(1)}(\vartheta^0) & \cdots & \mathcal{J}_1^{(s_1)}(\vartheta^0) \end{pmatrix}^\mathsf{T}$ *with the following components*

$$\begin{aligned}\mathcal{J}_1^{(i)}(\vartheta^0) :=& 2\, tr\Bigg(\left(V_{\vartheta^0}^{(h)}\right)^{-1} \left(\partial_i^1 \Pi(\vartheta^0)\right) \left(C_1 B_1, 0_{d\times(N-c)}\right) \int_0^1 W(r)\, \mathrm{d}W(r)^\mathsf{T} \\ &\cdot \left(-\left(C_1 B_1, \widetilde{\Psi}(1)\right) \overline{k}(1, \vartheta^0) + \left(0_{d\times c}, \Psi(1)\right) \Pi(\vartheta^0)\right)^\mathsf{T}\Bigg) \\ &+ 2\, tr\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1} \Sigma_{\vartheta^0}^{(i)}\right).\end{aligned}$$

The matrix $\Sigma_{\vartheta^0}^{(i)}$ *is defined in the proof.*

Proof. We know from (5.53) that for $i = 1, \ldots, s_1$ we have

$$\begin{aligned}\partial_i^1 \mathcal{L}_n^{(h)}(\vartheta^0) = \operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1} \partial_i^1 V_{\vartheta^0}^{(h)}\right) \\ - \operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1} \left(\partial_i^1 V_{\vartheta^0}^{(h)}\right) \left(V_{\vartheta^0}^{(h)}\right)^{-1} \frac{1}{n} \sum_{k=1}^{n} \varepsilon_k^{(h)}(\vartheta^0) \varepsilon_k^{(h)}(\vartheta^0)^\mathsf{T}\right)\end{aligned}$$

$$+ 2 \cdot \operatorname{tr}\left(\frac{1}{n}\sum_{k=1}^{n}\left(V_{\vartheta^0}^{(h)}\right)^{-1}\left(\partial_i^1 \varepsilon_k^{(h)}(\vartheta^0)\right)\varepsilon_k^{(h)}(\vartheta^0)^{\mathsf{T}}\right).$$

Note that the second term converges due to Lemma 5.8.1 (Birkhoff's Ergodic Theorem), i.e. we obtain

$$\begin{aligned}
&\operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\left(\partial_i^1 V_{\vartheta^0}^{(h)}\right)\left(V_{\vartheta^0}^{(h)}\right)^{-1}\frac{1}{n}\sum_{k=1}^{n}\varepsilon_k^{(h)}(\vartheta^0)\varepsilon_k^{(h)}(\vartheta^0)^{\mathsf{T}}\right)\\
&\xrightarrow{a.s.} \operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\left(\partial_i^1 V_{\vartheta^0}^{(h)}\right)\left(V_{\vartheta^0}^{(h)}\right)^{-1}V_{\vartheta^0}^{(h)}\right) = \operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\partial_i^1 V_{\vartheta^0}^{(h)}\right).
\end{aligned}$$

Hence, this term and the first term cancel each other out. Thus, it only remains to show the convergence of the last term. Hence, we obtain with Proposition 5.8.4, Theorem 5.4.4, Theorem 5.4.5 and the continuous mapping theorem

$$\begin{aligned}
&2 \cdot \operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1} \cdot \frac{1}{n}\sum_{k=1}^{n}\left(\partial_i^1 \varepsilon_k^{(h)}(\vartheta^0)\right)\varepsilon_k^{(h)}(\vartheta^0)^{\mathsf{T}}\right)\\
&= -2\operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1} \cdot \frac{1}{n}\sum_{k=1}^{n}\partial_i^1\Pi(\vartheta^0)Y_{k-1}^{(h)}\Delta Y_k^{(h)\mathsf{T}}\overline{k}(B,\vartheta^0)^{\mathsf{T}}\right)\\
&\quad + 2\operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1} \cdot \frac{1}{n}\sum_{k=1}^{n}\partial_i^1\Pi(\vartheta^0)Y_{k-1}^{(h)}Y_{k-1,2}^{(h)\mathsf{T}}\Pi(\vartheta^0)^{\mathsf{T}}\right)\\
&\quad + 2\operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1} \cdot \frac{1}{n}\sum_{k=1}^{n}\partial_i^1\overline{k}(B,\vartheta^0)\Delta Y_k^{(h)}\Delta Y_k^{(h)\mathsf{T}}\overline{k}(B,\vartheta^0)^{\mathsf{T}}\right)\\
&\quad - 2\operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1} \cdot \frac{1}{n}\sum_{k=1}^{n}\partial_i^1\overline{k}(B,\vartheta^0)\Delta Y_k^{(h)}Y_{k-1,2}^{(h)\mathsf{T}}\Pi(\vartheta^0)^{\mathsf{T}}\right)\\
&\xrightarrow[c]{w} -2\operatorname{tr}\Bigg(\left(V_{\vartheta^0}^{(h)}\right)^{-1} \cdot \left(\partial_i^1\Pi(\vartheta^0)\right)\left(C_1B_1, 0_{d\times(N-c)}\right)\\
&\qquad\qquad \cdot \int_0^1 W(r)\mathrm{d}W(r)^{\mathsf{T}}\left(C_1B_1, \widetilde{\Psi}(1)\right)^{\mathsf{T}}\overline{k}(1,\vartheta^0)^{\mathsf{T}}\Bigg)\\
&\quad - 2\operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\left(\partial_i^1\Pi(\vartheta^0)\right)C_1B_1\widetilde{\Sigma}^{(h)}B_2\sum_{j=1}^{\infty}\widetilde{\Psi}_j\sum_{i=j+1}^{\infty}\widetilde{K}_i(\vartheta)\right)\\
&\quad - 2\operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1} \cdot \left(\partial_i^1\Pi(\vartheta^0)\right)\Gamma_{Y\Delta Y}\overline{k}(1,\vartheta)^{\mathsf{T}}\right)\\
&\quad - 2\operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1} \cdot \left(\partial_i^1\Pi(\vartheta^0)\right)\sum_{\mu=0}^{\infty}\left[\Gamma_{Y_2\Delta Y}(-\mu) - \Gamma_{Y_2\Delta Y}(-\mu+1)\right]\sum_{\nu=\mu+1}^{\infty}\widetilde{K}_\nu(\vartheta)\right)\\
&\quad + 2\operatorname{tr}\Bigg(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\left(\partial_i^1\Pi(\vartheta^0)\right)\left(C_1B_1, 0_{d\times(N-c)}\right)\\
&\qquad\qquad \cdot \int_0^1 W(r)\,\mathrm{d}W(r)^{\mathsf{T}}\left(0_{d\times c}, \Psi(1)\right)^{\mathsf{T}}\Pi(\vartheta^0)^{\mathsf{T}}\Bigg)
\end{aligned}$$

$$
\begin{aligned}
&+ 2\,\mathrm{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\left(\partial_i^1 \Pi(\vartheta^0)\right)\Gamma_{YY_2}\Pi(\vartheta^0)^\mathsf{T}\right)\\
&+ 2\,\mathrm{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\cdot \sum_{\nu=0}^{\infty}\sum_{\mu=0}^{\infty}\partial_i^1 \widetilde{K}_\nu(\vartheta^0)\Gamma_{\Delta Y\Delta Y}(\nu-\mu)\widetilde{K}_\mu(\vartheta^0)^\mathsf{T}\right)\\
&- 2\,\mathrm{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\cdot \sum_{\nu=0}^{\infty}\partial_i^1 \widetilde{K}_\nu(\vartheta^0)\Gamma_{\Delta YY_2}(\nu)\Pi(\vartheta^0)^\mathsf{T}\right)\\
= {}& 2\,\mathrm{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\left(\partial_i^1 \Pi(\vartheta^0)\right)\left(C_1B_1, 0_{d\times(N-c)}\right)\int_0^1 W(r)\,\mathrm{d}W(r)^\mathsf{T}\right.\\
&\qquad \left.\cdot\left(-\left(C_1B_1, \widetilde{\Psi}(1)\right)\overline{k}(1,\vartheta^0) + \left(0_{d\times c}, \Psi(1)\right)\Pi(\vartheta^0)\right)^\mathsf{T}\right) + 2\,\mathrm{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\Sigma_{\vartheta^0}^{(i)}\right)\\
=:{}& \mathcal{J}_1^{(i)}(\vartheta^0),
\end{aligned}
\tag{5.55}
$$

where $\Sigma_{\vartheta^0}^{(i)}$ is given by all the remaining parts of the limit. □

Next we continue with the Hessian matrix for this we need the second partial derivatives of the innovations which are given by

$$
\partial_{i,j}^2 \varepsilon_k^{(h)}(\vartheta) = -\partial_{i,j}^2 \Pi(\vartheta)Y_{k-1} + \partial_{i,j}^2 \widetilde{k}(B,\vartheta)\Delta Y_k, \qquad \text{for } i,j = 1,\dots,s \tag{5.56}
$$

and the second partial derivatives of the log-likelihood function (5.6) are given by

$$
\partial_{i,j}^2 \mathcal{L}_n^{(h)}(\vartheta) = \frac{1}{n}\sum_{k=1}^{n}\partial_{i,j}^2 \ell_{\vartheta,k}^{(h)}, \tag{5.57a}
$$

where we obtain with the differentiation rules for matrix functions (c.f. Appendix B.4) and equation (5.53b)

$$
\begin{aligned}
\partial_i\partial_j \ell_{\vartheta,k}^{(h)} = {}& \mathrm{tr}\left(\left(V_\vartheta^{(h)}\right)^{-1}\partial_{i,j}^2 V_\vartheta^{(h)} - \left(V_\vartheta^{(h)}\right)^{-1}\left(\partial_i V_\vartheta^{(h)}\right)\left(V_\vartheta^{(h)}\right)^{-1}\left(\partial_j V_\vartheta^{(h)}\right)\right)\\
&- \mathrm{tr}\left(\left(V_\vartheta^{(h)}\right)^{-1}\varepsilon_k^{(h)}(\vartheta)\varepsilon_k^{(h)}(\vartheta)^\mathsf{T}\left(V_\vartheta^{(h)}\right)^{-1}\partial_{i,j}^2 V_\vartheta^{(h)}\right)\\
&+ \mathrm{tr}\left(\left(V_\vartheta^{(h)}\right)^{-1}\left(\partial_j V_\vartheta^{(h)}\right)\left(V^{(h)}\right)^{-1}\varepsilon_k^{(h)}(\vartheta)\varepsilon_k^{(h)}(\vartheta)^\mathsf{T}\left(V_\vartheta^{(h)}\right)^{-1}\partial_i V_\vartheta^{(h)}\right)\\
&+ \mathrm{tr}\left(\left(V_\vartheta^{(h)}\right)^{-1}\varepsilon_k^{(h)}(\vartheta)\varepsilon_k^{(h)}(\vartheta)^\mathsf{T}\left(V_\vartheta^{(h)}\right)^{-1}\left(\partial_j V_\vartheta^{(h)}\right)\left(V_\vartheta^{(h)}\right)^{-1}\partial_i V_\vartheta^{(h)}\right)\\
&- \mathrm{tr}\left(\left(V_\vartheta^{(h)}\right)^{-1}\left(\partial_j \varepsilon_k^{(h)}(\vartheta)\varepsilon_k^{(h)}(\vartheta)^\mathsf{T}\right)\left(V_\vartheta^{(h)}\right)^{-1}\partial_i V_\vartheta^{(h)}\right)\\
&+ 2\cdot\left(\partial_{i,j}^2 \varepsilon_k^{(h)}(\vartheta)^\mathsf{T}\right)\left(V_\vartheta^{(h)}\right)^{-1}\varepsilon_k^{(h)}(\vartheta)\\
&+ 2\cdot\left(\partial_i \varepsilon_k^{(h)}(\vartheta)^\mathsf{T}\right)\left(V_\vartheta^{(h)}\right)^{-1}\left(\partial_j \varepsilon_k^{(h)}(\vartheta)\right)\\
&- 2\cdot\mathrm{tr}\left(\left(V_\vartheta^{(h)}\right)^{-1}\varepsilon_k^{(h)}(\vartheta)\left(\partial_i \varepsilon_k^{(h)}(\vartheta)^\mathsf{T}\right)\left(V_\vartheta^{(h)}\right)^{-1}\partial_j V_\vartheta^{(h)}\right).
\end{aligned}
\tag{5.57b}
$$

Now we can derive the asymptotic distribution of the Hessian matrix with respect to the long-run parameters.

Proposition 5.6.2
For each sequence $\overline{\vartheta}_n \to \vartheta^0$ *we have*

$$n^{-1}\partial_i^1\partial_j^1\mathcal{L}_n^{(h)}(\overline{\vartheta}_n) \xrightarrow{w} [Z_1(\vartheta^0)]_{i,j},$$

where the $(s_1 \times s_1)$*-dimensional matrix* $Z_1(\vartheta_0)$ *is given for* $i, j = 1, \dots, s_1$ *by*

$$[Z_1(\vartheta^0)]_{i,j} := 2 \cdot tr\left(\left(V^{(h)}\right)^{-1}\partial_i^1\Pi(\vartheta^0)C_1B_1\int_0^1 W(r)W(r)^\mathsf{T}\,\mathrm{d}r\ \left(\partial_j^1\Pi(\vartheta^0)C_1B_1\right)^\mathsf{T}\right). \tag{5.58}$$

Moreover, for some $\delta > 0$ *and* $0 \leq \gamma < 1$ *the condition* (SE) *holds for* $n^{-1}\partial_i^1\partial_j^1\mathcal{L}_n^{(h)}(\cdot)$ *and* $Z_1(\vartheta^0)$ *is a random matrix which is almost surely positive definite.*

Proof. First, we prove that the condition (SE) holds. Next, we derive the convergence result and lastly we show the positive definiteness of the limiting matrix.

Step 1: The stochastic equicontinuity condition for all parts is a direct consequence of Theorem 5.4.5 since the derivatives still satisfy the necessary assumptions. We could proceed similar as in the proof of these theorems. Under these circumstances, we would derive the results with the derivative with respect to ϑ of the transfer function instead of the original transfer function.

Step 2: The first term in (5.57) converges to zero due to the additional normalizing rate of n^{-1}. The remaining terms in (5.57) can all be dealt in a similar way except the last term. Thus, we consider exemplarily the fifth term

$$-\operatorname{tr}\left(\frac{1}{n^2}\sum_{k=1}^n \left(V_{\overline{\vartheta}_n}^{(h)}\right)^{-1}\left(\partial_j\varepsilon_k^{(h)}(\overline{\vartheta}_n)\varepsilon_k^{(h)}(\overline{\vartheta}_n)^\mathsf{T}\right)\left(V_{\overline{\vartheta}_n}^{(h)}\right)^{-1}\partial_i V_{\overline{\vartheta}_n}^{(h)}\right).$$

We can ignore the trace operator and the factors $V_\vartheta^{(h)}$ since the convergence follows then by the continuous mapping theorem. Hence, it suffices to consider

$$\frac{1}{n}\sum_{k=1}^n \partial_j\varepsilon_k^{(h)}(\overline{\vartheta}_n)\varepsilon_k^{(h)}(\overline{\vartheta}_n)^\mathsf{T}.$$

Due to the assumption $\overline{\vartheta}_n \xrightarrow{n\to\infty} \vartheta^0$, we can therefore apply a mean value expansion for a suitable intermediate value $\widetilde{\vartheta}_n$ and obtain

$$\begin{aligned}\frac{1}{n^2}\sum_{k=1}^{n}\partial_j\varepsilon_k^{(h)}(\overline{\vartheta}_n)\varepsilon_k^{(h)}(\overline{\vartheta}_n)^{\mathsf{T}} =&\frac{1}{n^2}\sum_{k=1}^{n}\partial_j\varepsilon_k^{(h)}(\overline{\vartheta}_n)\varepsilon_k^{(h)\mathsf{T}}\\ &+\sum_{l_1=1}^{s}\frac{1}{n^2}\sum_{k=1}^{n}\partial_j\varepsilon_k^{(h)}(\overline{\vartheta}_n)\partial_{l_1}\varepsilon_k^{(h)}(\widetilde{\vartheta}_n)^{\mathsf{T}}\left(\overline{\vartheta}_{n,l_1}-\vartheta^0_{l_1}\right).\end{aligned}$$

The first term converges to zero in probability as $n\to\infty$ due to the normalizing rate of n^{-2} since it is the product of a stationary and non-stationary process. The term $\frac{1}{n^2}\sum_{k=1}^{n}\partial_j\varepsilon_k^{(h)}(\overline{\vartheta}_n)\partial_{l_1}\varepsilon_k^{(h)}(\overline{\vartheta}_n)^{\mathsf{T}}$ is of order $\mathcal{O}_p(1)$ due to Theorem 5.4.5 and $\left(\overline{\vartheta}_{n,l_1}-\vartheta^0_{l_1}\right)$ converges to zero and thus the whole expression.

Finally, we have to deal with the last term of (5.57), i.e.

$$\frac{1}{n^2}\sum_{k=1}^{n}\left(\partial_j\varepsilon_k^{(h)}(\vartheta)\right)\left(\partial_i\varepsilon_k^{(h)}(\vartheta)^{\mathsf{T}}\right).$$

We obtain due Lemma 5.2.2 and Theorem 5.4.5 for $i,j=1,\ldots,s_1$ that

$$\begin{aligned}n^{-1}\partial_i^1\partial_j^1\mathcal{L}_n^{(h)}(\overline{\vartheta}_n) &= 2\cdot\operatorname{tr}\left(\left(V_{\overline{\vartheta}_n}^{(h)}\right)^{-1}\frac{1}{n^2}\sum_{k=1}^{n}\partial_i^1\varepsilon_k^{(h)}(\overline{\vartheta}_n)\partial_j^1\varepsilon_k^{(h)}(\overline{\vartheta}_n)^{\mathsf{T}}\right)+o_p(1)\\ &= 2\cdot\operatorname{tr}\left(\left(V_{\overline{\vartheta}_n}^{(h)}\right)^{-1}\frac{1}{n^2}\sum_{k=1}^{n}\partial_i^1\Pi(\overline{\vartheta}_n)Y_{k-1}^{(h)}Y_{k-1}^{(h)\mathsf{T}}\partial_j^1\Pi(\overline{\vartheta}_n)^{\mathsf{T}}\right)+o_p(1)\\ &\xrightarrow{w} 2\cdot\operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\partial_i^1\Pi(\vartheta^0)C_1B_1\int_0^1 W(r)W(r)^{\mathsf{T}}\,\mathrm{d}r\,\left(\partial_j^1\Pi(\vartheta^0)C_1B_1\right)^{\mathsf{T}}\right).\end{aligned}$$

Step 3: We ignore the prefactor 2 in (5.58) in the following and transform the matrix $Z_1(\vartheta_0)$. For this purpose define $M:=B_1\int_0^1 W(r)W(r)^{\mathsf{T}}\,\mathrm{d}rB_1^{\mathsf{T}}$, which is a $\mathbb{P}$-a.s. positive definite $c\times c$ matrix. Hence, we apply the Cholesky decomposition $M=LL^{\mathsf{T}}$. Then we have by using properties of the vec operator and the Kronecker product (see for example [9, Chapter 7.1])

$$\begin{aligned}&\operatorname{tr}\left(\left(V^{(h)}\right)^{-\frac{1}{2}}\partial_i^1\Pi(\vartheta^0)C_1M\left(\left(V^{(h)}\right)^{-\frac{1}{2}}\partial_i^1\Pi(\vartheta^0)C_1\right)^{\mathsf{T}}\right)\\ &=\operatorname{vec}\left(\left(V^{(h)}\right)^{-\frac{1}{2}}\partial_i^1\Pi(\vartheta^0)C_1L\right)^{\mathsf{T}}\operatorname{vec}\left(\left(V^{(h)}\right)^{-\frac{1}{2}}\partial_j^1\Pi(\vartheta^0)C_1L\right)\\ &=\operatorname{vec}\left(\partial_i^1\Pi(\vartheta^0)C_1\right)^{\mathsf{T}}\left(L\otimes\left(V^{(h)}\right)^{-\frac{1}{2}}\right)\left(L^{\mathsf{T}}\otimes\left(V^{(h)}\right)^{-\frac{1}{2}}\right)\operatorname{vec}\left(\partial_j^1\Pi(\vartheta^0)C_1\right)\\ &=\operatorname{vec}\left(\partial_i^1\Pi(\vartheta^0)C_1\right)^{\mathsf{T}}\left(M\otimes\left(V^{(h)}\right)^{-1}\right)\operatorname{vec}\left(\partial_j^1\Pi(\vartheta^0)C_1\right).\end{aligned}$$

Furthermore, due to Bernstein [9, Fact 7.4.23],

$$\operatorname{rank}\left(M \otimes \left(V^{(h)}\right)^{-1}\right) = \operatorname{rank}(M) \cdot \operatorname{rank}\left(\left(V^{(h)}\right)^{-1}\right)$$

holds and thus $M \otimes \left(V^{(h)}\right)^{-1}$ has full rank $c \cdot d$.

Now, if we consider the Hessian matrix $Z_1(\vartheta_0)$, we have with $v_i := \operatorname{vec}\left(\partial_i^1 \Pi(\vartheta_1^0)^\mathsf{T} C_1\right)$

$$\begin{aligned} Z_1(\vartheta_0) &= \begin{pmatrix} v_1^\mathsf{T}\left(M \otimes \left(V^{(h)}\right)^{-1}\right) v_1 & \cdots & v_1^\mathsf{T}\left(M \otimes \left(V^{(h)}\right)^{-1}\right) v_{s_1} \\ \vdots & & \vdots \\ v_{s_1}^\mathsf{T}\left(M \otimes \left(V^{(h)}\right)^{-1}\right) v_1 & \cdots & v_{s_1}^\mathsf{T}\left(M \otimes \left(V^{(h)}\right)^{-1}\right) v_{s_1} \end{pmatrix} \\ &= \begin{pmatrix} v_1 & \cdots & v_{s_1} \end{pmatrix}^\mathsf{T} M \begin{pmatrix} v_1 & \cdots & v_{s_1} \end{pmatrix} \\ &= \nabla_{\vartheta_1} \left(\Pi(\vartheta) C_1\right)_{\vartheta=\vartheta^0}^\mathsf{T} M \nabla_{\vartheta_1} \left(\Pi(\vartheta) C_1\right)_{\vartheta=\vartheta^0}. \end{aligned} \tag{5.59}$$

Thus, $Z_1(\vartheta_0)$ is obviously positive semi-definite. Due to Assumption M12 the $(dc \times s_1)$-dimensional matrix $\nabla_{\vartheta_1} \left(\Pi(\vartheta) C_1\right)_{\vartheta=\vartheta^0}$ is of full column rank and hence the product has full rank s_1. Therefore, we have a regular positive semi-definite matrix and as a consequence the positive definiteness $\mathbb{P}$-almost surely. □

After we have shown the asymptotic behaviour of the score vector and the Hessian matrix we are able to derive the mixed normality of the long-run parameter estimator.

Theorem 5.6.3

Assume that Assumption M1 - Assumption M12 hold then we have as $n \to \infty$

$$n(\widehat{\vartheta}_{n,1} - \vartheta_1^0) \xrightarrow{w} Z_1(\vartheta_0)^{-1} \cdot \mathcal{J}_1(\vartheta^0) \tag{5.60}$$

where $\mathcal{J}_1(\vartheta^0)$ is the weak limit of $\nabla_{\vartheta_1} \mathcal{L}_n^{(h)}(\vartheta^0)$ and $Z_1(\vartheta_0)$ is the weak limit of $n^{-1} \nabla_{\vartheta_1}^2 \mathcal{L}_n^{(h)}(\vartheta^0)$ as $n \to \infty$.

Proof. We have shown in Theorem 5.5.5 the consistency of $\widehat{\vartheta}_n$ where for the long-run QMLE $\widehat{\vartheta}_{n,1}$ holds $\widehat{\vartheta}_{n,1} - \vartheta_1^0 = o_p(n^{-\gamma})$, for $0 \leq \gamma < 1$. The true parameter $\vartheta^0 = ((\vartheta_1^0)^\mathsf{T}, (\vartheta_2^0)^\mathsf{T})^\mathsf{T}$ is an element of the interior of the compact parameter space $\Theta = \Theta_1 \times \Theta_2$ due to Assumption M5. Hence, the estimator $\widehat{\vartheta}_{n,1}$ is at some point also an element of the interior of Θ_1 with probability one.

Because the parametrization is assumed to be twice continuously differentiable, we can find the minimizing $\widehat{\vartheta}_{n,1}$ via the first order condition $\nabla_{\vartheta_1} \mathcal{L}_n^{(h)}(\vartheta_1, \vartheta_2) = 0_{s_1}$. We apply a Taylor expansion of the score vector around the point ϑ_1^0. Thus, we obtain the existence of parameter vectors $\underline{\vartheta}_{n,1,i} \in \Theta_1$ of the form $\underline{\vartheta}_{n,1,i} = \vartheta_1^0 + c_i(\widehat{\vartheta}_{n,1} - \vartheta_1^0)$, $0 \leq c_i \leq 1$ such that

$$0_{s_1} = \nabla_{\vartheta_1}\mathcal{L}_n^{(h)}(\vartheta_1^0, \widehat{\vartheta}_{n,2}) + n^{-1}\nabla^2_{\vartheta_1}\mathcal{L}_n^{(h)}(\underline{\vartheta}_{n,1}, \widehat{\vartheta}_{n,2})n(\widehat{\vartheta}_{n,1} - \vartheta_1^0), \tag{5.61}$$

where $\nabla^2_{\vartheta_1}\mathcal{L}_n^{(h)}(\underline{\vartheta}_{n,1}, \widehat{\vartheta}_{n,2})$ denotes the matrix whose i^{th} row, for $i = 1, \ldots, s_1$, is equal to the i^{th} row of $\nabla^2_{\vartheta_1}\mathcal{L}_n^{(h)}(\underline{\vartheta}_{n,1,i}, \widehat{\vartheta}_{n,2})$. We have already shown the asymptotic behavior of the first term in Proposition 5.6.1. Due to Proposition 5.6.2 we have that $n^{-1}\nabla^2_{\vartheta_1}\mathcal{L}_n^{(h)}(\underline{\vartheta}_{n,1}, \widehat{\vartheta}_{n,2})$ converges weakly to the random matrix $Z_1(\vartheta_0)$. Thus, (5.61) together with the almost sure positive definiteness of $Z_1(\vartheta_0)$, allows us to take the inverse and reorder the equation. Finally, we obtain

$$n(\widehat{\vartheta}_{n,1} - \vartheta^0) = -\left(n^{-1}\nabla^2_{\vartheta_1}\mathcal{L}_n^{(h)}(\underline{\vartheta}_{n,1}, \widehat{\vartheta}_{n,2})\right)^{-1}\nabla_{\vartheta_1}\mathcal{L}_n^{(h)}(\vartheta_1^0, \widehat{\vartheta}_{n,2}).$$

In particular, Theorem 5.4.5 also guarantees the joint convergence of the expressions. From the previous results and the continuous mapping theorem this converges weakly to the limit given in (5.60). □

5.6.2. ASYMPTOTIC DISTRIBUTION OF THE SHORT-RUN ESTIMATOR

Lastly, we derive the asymptotic normality of the short-run quasi-maximum likelihood estimator $\widehat{\vartheta}_{n,2}$ which we also prove by using a mean value expansion as for the long-run estimator. We prove in the next lemma that the partial derivatives have finite variance.

Lemma 5.6.4
For each $\vartheta_2 \in \Theta_2$ and every $i = 1, \ldots, s_2$, the random variable $\partial_i^{st}\big(\mathcal{L}_n^{(h)}(\vartheta_1^0, \vartheta_2)\big)$ has finite variance.

Proof. We have due to Lemma 5.9.5 and the Cauchy-Schwarz inequality

$$\begin{aligned}\mathbb{E}\Bigg| - \operatorname{tr}\Bigg(\big(V_{\vartheta_2}^{(h)}\big)^{-1}\varepsilon_{k,2}^{(h)}(\vartheta_2)\varepsilon_{k,2}^{(h)}(\vartheta_1^0, \vartheta_2)^{\mathsf{T}}\big(V_{\vartheta_2}^{(h)}\big)^{-1}\partial_i^{st}V_{\vartheta_2}^{(h)}\Bigg)& \\ + 2\cdot\left(\partial_i^{st}\varepsilon_{k,2}^{(h)}(\vartheta_2)^{\mathsf{T}}\right)\big(V_{\vartheta_2}^{(h)}\big)^{-1}\varepsilon_{k,2}^{(h)}(\vartheta_2)\Bigg|^2& \\ \leq C\cdot\mathbb{E}\|\varepsilon_{k,2}^{(h)}(\vartheta_2)\|^4 + C\cdot\left(\mathbb{E}\|\varepsilon_{k,2}^{(h)}(\vartheta_2)\|^4\mathbb{E}\|\partial_i^{st}\varepsilon_{k,2}^{(h)}(\vartheta_2)\|^4\right)^{\frac{1}{2}} < \infty&\end{aligned}$$

so that the statement follows with (5.53). □

Now we can prove the convergence of the covariance matrix of the score vector, i.e. the gradient of the log-likelihood function, where the true long-run parameter is inserted.

Lemma 5.6.5
We have for all $\vartheta_2 \in \Theta_2$ that

$$Var\left(\nabla_{\vartheta_2}\mathcal{L}_n^{(h)}(\vartheta_1^0,\vartheta_2)\right) \xrightarrow{n\to\infty} I(\vartheta_2), \tag{5.62}$$

where $I(\vartheta_2)$ is given for $i,j \in \{1,\ldots,s_2\}$ by

$$I(\vartheta_2) := \left[\sum_{l\in\mathbb{Z}} Cov\left(\partial_i^{st}\ell_{1,2}^{(h)}(\vartheta_2), \partial_j^{st}\ell_{1+l,2}^{(h)}(\vartheta_2)\right)\right]_{i,j} \tag{5.63}$$

and $\ell_{1,2}^{(h)}(\vartheta_2)$ are defined as in (5.11c).

Proof. We can derive the results in a similar way as in Lemma 2.14 in Schlemm and Stelzer [91]. Hence, we only sketch the proof to show the differences in the representations. We know by Lemma 4.3.18 that the strong-mixing coefficients $\alpha_{Y_2^{(h)}}$ of the stationary part of the sampled process satisfy $\sum_m[\alpha(m)]^{\frac{\delta}{2+\delta}} < \infty$ for some constant $\delta > 0$. Due to Remark 1.8b) in Bradley [16], we also know that $\Delta Y_2^{(h)}$ is strongly mixing.
Note that $\left(\varepsilon_{k,2}^{(h)}(\vartheta_2)\right)_{k\in\mathbb{N}}$ is a stationary sequence, compare with equation (5.10c). We can restrict ourselves to show that for all $\vartheta_2 \in \Theta_2$ and all $i,j = 1,\ldots,s_2$ the sequence $I_n^{(i,j)}(\vartheta)$ given by

$$I_n^{(i,j)}(\vartheta_2) := n^{-1}\sum_{k_1=1}^{n}\sum_{k_2=1}^{n} \text{Cov}\left(\partial_i^{st}\ell_{k_1,2}^{(h)}(\vartheta_2), \partial_j^{st}\ell_{k_2,2}^{(h)}(\vartheta_2)\right) \tag{5.64}$$

converges. Recall the representation of the partial derivatives in (5.53b). By stationarity, the covariance of (5.64) depends only on the difference l of k_1 and k_2. As in Schlemm and Stelzer [91], Lemma 2.14, it suffices to show that

$$\text{Cov}\left(\partial_i^{st}\ell_{1,2}^{(h)}(\vartheta_2), \partial_j^{st}\ell_{1+l,2}^{(h)}(\vartheta_2)\right), \quad l \in \mathbb{Z}, \tag{5.65}$$

is absolutely summable for all $i,j = 1,\ldots,s_2$. Then the Dominated Convergence Theorem implies that

$$I_n^{(i,j)}(\vartheta_2) = n^{-1}\sum_{l=-n}^{n}(n-|l|)\,\text{Cov}\left(\partial_i^{st}\ell_{1,2}^{(h)}(\vartheta_2), \partial_j^{st}\ell_{1+l,2}^{(h)}(\vartheta_2)\right)$$

$$\xrightarrow{n\to\infty} \sum_{l\in\mathbb{Z}} \operatorname{Cov}\left(\partial_i^{st}\ell_{1,2}^{(h)}(\vartheta_2), \partial_j^{st}\ell_{1+l,2}^{(h)}(\vartheta_2)\right) < \infty.$$

The absolute summability follows in a similar way as in the proof of Lemma 2.14 in Schlemm and Stelzer [91] using on the one hand the bilinearity of the covariance matrix, the Cauchy-Schwarz inequality and a covariance inequality which is a consequence of Davydov's inequality (c.f. Lemma 2.13, Schlemm and Stelzer [91]). To be more precise, we use the representation given in Lemma 5.2.1 and Lemma 5.2.2 and the same splitting as in the proof of Lemma 2.14 in Schlemm and Stelzer [91]. Furthermore, the exponential decay of the coefficients given in Lemma 5.2.1 and Lemma 5.2.2 is used. More details on the lengthy calculations are given in Appendix 5.9.3. □

Now we have all auxiliary results needed to consider the asymptotic behavior of the score vector. We will see that the gradient with respect to the short-run parameters is asymptotically normal with a truncation argument.

Proposition 5.6.6

For the gradient with respect to the stationary parameters we have the following asymptotic behavior

$$\sqrt{n} \cdot \nabla_{\vartheta_2}\mathcal{L}_n^{(h)}(\vartheta^0) \xrightarrow{w} \mathcal{N}(0, I(\vartheta_2^0)), \tag{5.66}$$

where $I(\vartheta_2^0)$ is the asymptotic covariance matrix given in (5.63).

Proof. We know that $\partial_i\varepsilon_k^{(h)}(\vartheta)$ is an element of the Hilbert space generated by $\left\{Y_l^{(h)}, l<k\right\}$. Furthermore, the fact that $\mathbb{E}\left[\varepsilon_k^{(h)}(\vartheta^0)\varepsilon_k^{(h)}(\vartheta^0)^\mathsf{T}\right] = V_{\vartheta^0}^{(h)}$ and the orthogonality of $\varepsilon_k^{(h)}(\vartheta^0)^\mathsf{T}$ to the Hilbert space generated by $\left\{Y_l^{(h)}, l<k\right\}$ shows that $\mathbb{E}\left[\nabla_{\vartheta_2}\mathcal{L}_n^{(h)}(\vartheta^0)\right] = 0_{s_2}$ (c.f. (5.53b)).

Recall that the derivatives of the pseudo-innovations with respect to the stationary parameters are stationary as well, since the non-stationary part is still canceled out by $\beta(\vartheta_1^0)^\mathsf{T} = C_1^\perp$.

Due to the representation (5.5) we can rewrite (5.53b) for $m \in \mathbb{N}$ as

$$\partial_i\mathcal{L}_n^{(h)}(\vartheta^0) = \frac{1}{n}\sum_{k=1}^{n}\left(Y_{m,k}^{(i)} - \mathbb{E}Y_{m,k}^{(i)}\right) + \frac{1}{n}\sum_{k=1}^{n}\left(Z_{m,k}^{(i)} - \mathbb{E}Z_{m,k}^{(i)}\right),$$

where

$$
\begin{aligned}
Y_{m,k}^{(i)} =& \operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\partial_i V_{\vartheta^0}^{(h)}\right)\\
&- \operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\Pi(\vartheta^0)Y_{k-1,2}^{(h)}Y_{k-1,2}^{(h)\mathsf{T}}\Pi(\vartheta^0)^{\mathsf{T}}\left(V_{\vartheta^0}^{(h)}\right)^{-1}\partial_i V_{\vartheta^0}^{(h)}\right)\\
&+ \sum_{\iota_1=0}^{m}\operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\widetilde{K}_{\iota_1}(\vartheta^0)\Delta Y_{k-\iota_1}^{(h)}Y_{k-1,2}^{(h)\mathsf{T}}\Pi(\vartheta^0)^{\mathsf{T}}\left(V_{\vartheta^0}^{(h)}\right)^{-1}\partial_i V_{\vartheta^0}^{(h)}\right)\\
&+ \sum_{\iota_2=0}^{m}\operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\Pi(\vartheta)Y_{k-1,2}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta^0)^{\mathsf{T}}\left(V_{\vartheta^0}^{(h)}\right)^{-1}\partial_i V_{\vartheta^0}^{(h)}\right)\\
&- \sum_{\iota_1,\iota_2=0}^{m}\operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\widetilde{K}_{\iota_1}(\vartheta^0)\Delta Y_{k-\iota_1}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta^0)^{\mathsf{T}}\left(V_{\vartheta^0}^{(h)}\right)^{-1}\partial_i V_{\vartheta^0}^{(h)}\right)\\
&+ 2\cdot\operatorname{tr}\left(\left(\partial_i\Pi(\vartheta^0)Y_{k-1,2}^{(h)}\right)\left(V_{\vartheta^0}^{(h)}\right)^{-1}Y_{k-1,2}^{(h)\mathsf{T}}\Pi(\vartheta^0)^{\mathsf{T}}\right)\\
&- 2\cdot\sum_{\iota_1=0}^{m}\operatorname{tr}\left(\left(\partial_i\widetilde{K}_{\iota_1}(\vartheta^0)\Delta Y_{k-\iota_1}^{(h)}\right)\left(V_{\vartheta^0}^{(h)}\right)^{-1}Y_{k-1,2}^{(h)\mathsf{T}}\Pi(\vartheta^0)^{\mathsf{T}}\right)\\
&- 2\cdot\sum_{\iota_2=0}^{m}\operatorname{tr}\left(\left(\partial_i\Pi(\vartheta^0)Y_{k-1,2}^{(h)}\right)\left(V_{\vartheta^0}^{(h)}\right)^{-1}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta^0)^{\mathsf{T}}\right)\\
&+ 2\cdot\sum_{\iota_1,\iota_2=0}^{m}\operatorname{tr}\left(\left(\partial_i\widetilde{K}_{\iota_1}(\vartheta^0)\Delta Y_{k-\iota_1}^{(h)}\right)\left(V_{\vartheta^0}^{(h)}\right)^{-1}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta^0)^{\mathsf{T}}\right) \qquad (5.67a)\\
Z_{m,k}^{(i)} =& U_{m,k}^{(i)} + V_{m,k}^{(i)} \qquad (5.67b)
\end{aligned}
$$

and

$$
\begin{aligned}
U_{m,k}^{(i)} =& \sum_{\iota_2=m+1}^{\infty}\operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\Pi(\vartheta)Y_{k-1,2}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta^0)^{\mathsf{T}}\left(V_{\vartheta^0}^{(h)}\right)^{-1}\partial_i V_{\vartheta^0}^{(h)}\right)\\
&- \sum_{\iota_1=0}^{\infty}\sum_{\iota_2=m+1}^{\infty}\operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\widetilde{K}_{\iota_1}(\vartheta^0)\Delta Y_{k-\iota_1}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta^0)^{\mathsf{T}}\left(V_{\vartheta^0}^{(h)}\right)^{-1}\partial_i V_{\vartheta^0}^{(h)}\right)\\
&- 2\cdot\sum_{\iota_2=m+1}^{m}\operatorname{tr}\left(\left(\partial_i\Pi(\vartheta^0)Y_{k-1,2}^{(h)}\right)\left(V_{\vartheta^0}^{(h)}\right)^{-1}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta^0)^{\mathsf{T}}\right)\\
&+ 2\cdot\sum_{\iota_1=0}^{\infty}\sum_{\iota_2=m+1}^{\infty}\operatorname{tr}\left(\left(\partial_i\widetilde{K}_{\iota_1}(\vartheta^0)\Delta Y_{k-\iota_1}^{(h)}\right)\left(V_{\vartheta^0}^{(h)}\right)^{-1}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta^0)^{\mathsf{T}}\right)\\
V_{m,k}^{(i)} =& \sum_{\iota_1=m+1}^{\infty}\operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\widetilde{K}_{\iota_1}(\vartheta^0)\Delta Y_{k-\iota_1}^{(h)}Y_{k-1,2}^{(h)\mathsf{T}}\Pi(\vartheta^0)^{\mathsf{T}}\left(V_{\vartheta^0}^{(h)}\right)^{-1}\partial_i V_{\vartheta^0}^{(h)}\right)\\
&- \sum_{\iota_1=m+1}^{\infty}\sum_{\iota_2=0}^{\infty}\operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\widetilde{K}_{\iota_1}(\vartheta^0)\Delta Y_{k-\iota_1}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta^0)^{\mathsf{T}}\left(V_{\vartheta^0}^{(h)}\right)^{-1}\partial_i V_{\vartheta^0}^{(h)}\right)
\end{aligned}
$$

$$- 2 \cdot \sum_{\iota_1=m+1}^{\infty} \operatorname{tr}\Big(\big(\partial_i \widetilde{K}_{\iota_1}(\vartheta^0)\Delta Y^{(h)}_{k-\iota_1}\big)\big(V^{(h)}_{\vartheta^0}\big)^{-1} Y^{(h)\mathsf{T}}_{k-1,2}\Pi(\vartheta^0)^{\mathsf{T}}\Big)$$
$$+ 2 \cdot \sum_{\iota_1=m+1}^{\infty}\sum_{\iota_2=0}^{\infty} \operatorname{tr}\Big(\big(\partial_i \widetilde{K}_{\iota_1}(\vartheta^0)\Delta Y^{(h)}_{k-\iota_1}\big)\big(V^{(h)}_{\vartheta^0}\big)^{-1} \Delta Y^{(h)\mathsf{T}}_{k-\iota_2}\widetilde{K}_{\iota_2}(\vartheta^0)^{\mathsf{T}}\Big).$$

We define

$$\mathcal{Y}_{m,k} = (Y^{(1)}_{m,k} \;\cdots\; Y^{(s_2)}_{m,k})^{\mathsf{T}} \quad \text{and} \quad \mathcal{Z}_{m,k} = (Z^{(1)}_{m,k} \;\cdots\; Z^{(s_2)}_{m,k})^{\mathsf{T}} \tag{5.68}$$

and use a truncation argument analogous to the proof of Lemma 2.16 in Schlemm and Stelzer [91]. We show the claim in three steps.

Step 1:
The process $\mathcal{Y}_{m,k}$ depends only on $m+1$ past values of $Y_2^{(h)}$. Hence, it inherits the strong mixing property of $Y_2^{(h)}$ and satisfies $\alpha_{\mathcal{Y}_m}(l) \leq \alpha_{Y_2^{(h)}}(\max\{0, l-m-1\})$. Thus by Lemma 4.3.18 we have $\sum_{l=1}^{\infty}(\alpha_{\mathcal{Y}_m}(l))^{\frac{\delta}{2+\delta}} < \infty$. Using the Cramér-Wold device and the univariate central limit theorem of Ibragimov [51], Theorem 1.7, for strongly mixing random variables we obtain

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(\mathcal{Y}_{m,k} - \mathbb{E}\mathcal{Y}_{m,k}) \xrightarrow{w} \mathcal{N}(0_{s_2}, \widehat{I}_m(\vartheta_2^0)) \tag{5.69}$$

as $n \to \infty$ and $\widehat{I}_m(\vartheta_2^0) = \sum_{l\in\mathbb{Z}} \operatorname{Cov}(\mathcal{Y}_{m,1}; \mathcal{Y}_{m,1+l})$.

Next we need to show that

$$\widehat{I}_m(\vartheta_2^0) \xrightarrow{m\to\infty} I(\vartheta_2^0). \tag{5.70}$$

Note that the bilinearity property of the covariance operator implies

$$\operatorname{Cov}\big(Y^{(i)}_{m,k}, Y^{(j)}_{m,k+l}\big) - \operatorname{Cov}\big(\partial_i \ell^{(h)}_{\vartheta^0,k}, \partial_j \ell^{(h)}_{\vartheta^0,k+l}\big)$$
$$= \operatorname{Cov}\big(Y^{(i)}_{m,k}, Y^{(j)}_{m,k+l} - \partial_j \ell^{(h)}_{\vartheta^0,k+l}\big) + \operatorname{Cov}\big(Y^{(i)}_{m,k} - \partial_i \ell^{(h)}_{\vartheta^0,k}, \partial_j \ell^{(h)}_{\vartheta^0,k+l}\big).$$

These terms can be treated in a similar way and hence we only consider the second one. Due to the definitions it yields that

$$Y^{(i)}_{m,k} - \partial_i \ell^{(h)}_{\vartheta^0,k} = -\sum_{\iota_1=m+1}^{\infty} \operatorname{tr}\Big(\big(V^{(h)}_{\vartheta^0}\big)^{-1}\widetilde{K}_{\iota_1}(\vartheta^0)\Delta Y^{(h)}_{k-\iota_1} Y^{(h)\mathsf{T}}_{k-1,2}\Pi(\vartheta^0)^{\mathsf{T}}\big(V^{(h)}_{\vartheta^0}\big)^{-1}\partial_i V^{(h)}_{\vartheta^0}\Big)$$
$$- \sum_{\iota_2=m+1}^{\infty} \operatorname{tr}\Big(\big(V^{(h)}_{\vartheta^0}\big)^{-1}\Pi(\vartheta)Y^{(h)}_{k-1,2}\Delta Y^{(h)\mathsf{T}}_{k-\iota_2}\widetilde{K}_{\iota_2}(\vartheta^0)^{\mathsf{T}}\big(V^{(h)}_{\vartheta^0}\big)^{-1}\partial_i V^{(h)}_{\vartheta^0}\Big)$$

$$
\begin{aligned}
&+ \sum_{\substack{\iota_1,\iota_2\\ \max\{\iota_1,\iota_2\}>m}} \operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\widetilde{K}_{\iota_1}(\vartheta^0)\Delta Y_{k-\iota_1}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta^0)^{\mathsf{T}}\left(V_{\vartheta^0}^{(h)}\right)^{-1}\partial_i V_{\vartheta^0}^{(h)}\right)\\
&+2\cdot\sum_{\iota_1=m+1}^{\infty}\operatorname{tr}\left(\left(\partial_i\widetilde{K}_{\iota_1}(\vartheta^0)\Delta Y_{k-\iota_1}^{(h)}\right)\left(V_{\vartheta^0}^{(h)}\right)^{-1}Y_{k-1,2}^{(h)\mathsf{T}}\Pi(\vartheta^0)^{\mathsf{T}}\right)\\
&+2\cdot\sum_{\iota_2=m+1}^{\infty}\operatorname{tr}\left(\left(\partial_i\Pi(\vartheta^0)Y_{k-1,2}^{(h)}\right)\left(V_{\vartheta^0}^{(h)}\right)^{-1}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta^0)^{\mathsf{T}}\right)\\
&-2\cdot\sum_{\substack{\iota_1,\iota_2\\ \max\{\iota_1,\iota_2\}>m}}\operatorname{tr}\left(\left(\partial_i\widetilde{K}_{\iota_1}(\vartheta^0)\Delta Y_{k-\iota_1}^{(h)}\right)\left(V_{\vartheta^0}^{(h)}\right)^{-1}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta^0)^{\mathsf{T}}\right).
\end{aligned}
$$

We obtain with the Cauchy-Schwarz inequality, the exponentially decreasing coefficients and the finite $(4+\delta)$-moments (c.f. Assumption M1) that $\operatorname{Var}(Y_{m,k}^{(i)}-\partial_i\ell_{\vartheta^0,k}^{(h)})\leq C\rho^m$, which does no longer depend on n. Thus, the L^2 continuity of the covariance operator implies that $\operatorname{Cov}\left(Y_{m,k}^{(i)}-\partial_i\ell_{\vartheta^0,k}^{(h)},\partial_j\ell_{\vartheta^0,k+l}^{(h)}\right)$ converges uniformly in l and at an exponential rate to zero as $m\to\infty$.

Hence, we have $\operatorname{Cov}\left(Y_{m,k}^{(i)},Y_{m,k+l}^{(j)}\right)\xrightarrow{m\to\infty}\operatorname{Cov}\left(\partial_i\ell_{\vartheta^0,k}^{(h)},\partial_j\ell_{\vartheta^0,k+l}^{(h)}\right)$ and the same arguments as in the proof of Lemma 5.6.5 guarantee that there exists a summable sequence, which dominates $|\operatorname{Cov}\left(Y_{m,k}^{(i)},Y_{m,k+l}^{(j)}\right)|$. Finally, these two results imply that the covariance matrix converges as in (5.70).

Step 2:
In this step, we show that $\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left(\mathcal{Z}_{m,k}-\mathbb{E}\mathcal{Z}_{m,k}\right)$ is asymptotically negligible. It holds that

$$
\begin{aligned}
&\operatorname{tr}\operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\mathcal{Z}_{m,k}\right)\\
&\leq 2\left(\operatorname{tr}\operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\mathcal{U}_{m,k}\right)+\operatorname{tr}\operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\mathcal{V}_{m,k}\right)\right),
\end{aligned}
\tag{5.71}
$$

where $\mathcal{U}_{m,k}$ and $\mathcal{V}_{m,k}$ are defined as $\mathcal{Z}_{m,k}$. Since both terms can be treated similarly we consider only the first one

$$
\begin{aligned}
\operatorname{tr}\operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\mathcal{U}_{m,k}\right)&=\frac{1}{n}\operatorname{tr}\sum_{k,k'=1}^{n}\operatorname{Cov}(\mathcal{U}_{m,k},\mathcal{U}_{m,k'})\\
&\leq\frac{1}{n}\sum_{i,j=1}^{s_2}\sum_{l=-n+1}^{n-1}(n-|l|)\mathfrak{u}_{m,l}^{(i,j)}\\
&\leq\sum_{i,j=1}^{s_2}\sum_{l\in\mathbb{Z}}|\mathfrak{u}_{m,l}^{(i,j)}|,
\end{aligned}
\tag{5.72}
$$

where

$$\mathfrak{u}_{m,l}^{(i,j)} = \mathrm{Cov}(U_{m,k}^{(i)}, U_{m,k+l}^{(j)}).$$

As in the proof of Lemma 2.16 in Schlemm and Stelzer [91], we find an upper bound for $|\mathfrak{u}_{m,l}^{(i,j)}|$ independent of i and j, using the Davydov inequality, Cauchy-Schwarz inequality and the same arguments as in the proof of Lemma 5.6.5. Namely, we obtain

$$|\mathfrak{u}_{m,l}^{(i,j)}| \leq C\rho^m \left(\left[\alpha_{Y_2^{(h)}} \left(\left\lfloor \frac{l}{2} \right\rfloor \right) \right]^{\frac{\delta}{\delta+2}} + \rho^{\frac{l}{2}} \right).$$

Moreover, we have by considering the sum over l the upper bound

$$\begin{aligned}\sum_{l=0}^{\infty} |\mathfrak{u}_{m,l}^{(i,j)}| &\leq \sum_{l=0}^{2m} |\mathfrak{u}_{m,l}^{(i,j)}| + \sum_{l=2m+1}^{\infty} |\mathfrak{u}_{m,l}^{(i,j)}| \\ &\leq C\rho^m \left(m + \sum_{l=0}^{\infty} \left[\alpha_{Y_2^{(h)}}(l) \right]^{\frac{\delta}{\delta+2}} \right),\end{aligned}$$

which implies $\mathrm{tr}\,\mathrm{Var}\left(\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \mathcal{U}_{m,k}\right) \leq C(m + \widetilde{C})\rho^m$ due to (5.72).

With the same steps one obtains an equivalent bound for $\mathrm{tr}\,\mathrm{Var}\left(\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \mathcal{V}_{m,k}\right)$ and thus we have with (5.71)

$$\mathrm{tr}\,\mathrm{Var}\left(\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \mathcal{Z}_{m,k}\right) \leq C(m + \widetilde{C})\rho^m. \tag{5.73}$$

Step 3:
With the multivariate Chebyshev inequality (see e.g. Schlemm [89], Lemma 3.19) we obtain for every $\epsilon > 0$ that

$$\begin{aligned}&\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\left\| \sqrt{n}\nabla_{\vartheta_2}\mathcal{L}_n^{(h)}(\vartheta^0) - \frac{1}{\sqrt{n}} \sum_{k=1}^{n} [\mathcal{Y}_{m,k} - \mathbb{E}\mathcal{Y}_{m,k}] \right\| > \epsilon \right) \\ &\leq \lim_{m\to\infty} \limsup_{n\to\infty} \frac{s_2}{\epsilon^2} \mathrm{tr}\,\mathrm{Var}\left(\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \mathcal{Z}_{m,k} \right) \\ &\leq \lim_{m\to\infty} \frac{s_2}{\epsilon^2} C(m + \widetilde{C})\rho^m = 0,\end{aligned}$$

where we used the result from Step 2. All in all, the previous results and Proposition 6.3.9 of [20] yield the asymptotic normality in (5.66). □

Next, we consider the asymptotic behaviour of the Hessian matrix with respect to the stationary parameters. For this we define the following matrix, which plays an important role in order to make sure that the Hessian matrix is positive definite. We denote shortly $F_\vartheta := \mathrm{e}^{A_\vartheta h} - K_\vartheta^{(h)} C_\vartheta$. The function is similar to the function in Schlemm and Stelzer [91], Assumption C11. However, we define F_ϑ slightly different since we do not have a moving average representation of $Y^{(h)}$ with respect to the innovations. Though, we have a moving average representation of $\varepsilon^{(h)}$ with respect to $Y^{(h)}$, c.f. (4.34). Hence, we have to adapt the criterion slightly and define the function

$$\psi_{\vartheta,j} := \begin{pmatrix} \left[1_{j+1} \otimes K_\vartheta^\mathsf{T} \otimes C_\vartheta\right] \left[(\mathrm{vec}\ 1_N)^\mathsf{T} \quad (\mathrm{vec}\ F_\vartheta)^\mathsf{T} \ \ldots \ (\mathrm{vec}\ F_\vartheta^j)^\mathsf{T}\right]^\mathsf{T} \\ \mathrm{vec}\ V_\vartheta^{(h)} \end{pmatrix}. \tag{5.74}$$

Assumption M13
Assume that there exists a positive index j_0 such that the $[(j_0+2)d^2 \times s_2]$ matrix $\nabla_{\vartheta_2} \psi_{\vartheta^0, j_0}$ has rank s_2.

Proposition 5.6.7
Assume that Assumption M13 additionally holds. Then, for each sequence $\overline{\vartheta}_n = (\overline{\vartheta}_{n,1}, \overline{\vartheta}_{n,2}) \to \vartheta^0$ with $\overline{\vartheta}_{n,1} \in N_{n,\gamma}(\vartheta_1^0, \delta)$, we have

$$\partial_i^{st} \partial_j^{st} \mathcal{L}_n^{(h)}(\overline{\vartheta}_n) \xrightarrow{p} [Z_{st}]_{i,j},$$

where the $(s_2 \times s_2)$-dimensional matrix Z_{st} is given for $i, j = 1, \ldots, s_2$ by

$$\begin{aligned} [Z_{st}]_{i,j} :=& 2\mathbb{E}\left(\partial_i \varepsilon_1^{(h)}(\vartheta^0)^\mathsf{T}\right)\left(V^{(h)}\right)^{-1}\left(\partial_j \varepsilon_1^{(h)}(\vartheta^0)\right) \\ &+ tr\left(\left(V^{(h)}\right)^{-\frac{1}{2}}\left(\partial_i V^{(h)}\right)\left(V^{(h)}\right)^{-1}\partial_j V^{(h)}\left(V^{(h)}\right)^{-\frac{1}{2}}\right). \end{aligned}$$

Moreover, for some $\delta > 0$ and $0 \leq \gamma < 1$ the condition (SE) holds for $\partial_i^{st}\partial_j^{st}\mathcal{L}_n^{(h)}(\overline{\vartheta}_n)$ and the limiting matrix Z_{st} is almost surely a non-singular constant matrix.

Proof. We proceed as in the proof of Proposition 5.6.2.

Step 1: The condition (SE) is again a direct consequence of Theorem 5.4.4 and Theorem 5.4.5.

Step 2: The first term in (5.57) is asymptotically a constant. Let us now consider all other terms in (5.57) except the last one. They follow all in a similar way, hence we consider again for example the relevant part of the fifth term. As in the proof of

Proposition 5.6.2 we use a mean value expansion

$$\frac{1}{n}\sum_{k=1}^{n}\partial_j\varepsilon_k^{(h)}(\overline{\vartheta}_n)\varepsilon_k^{(h)}(\overline{\vartheta}_n)^{\mathsf{T}} = \frac{1}{n}\sum_{k=1}^{n}\partial_j\varepsilon_k^{(h)}(\overline{\vartheta}_n)\varepsilon_k^{(h)\mathsf{T}} + \sum_{l_1=1}^{s}\frac{1}{n}\sum_{k=1}^{n}\partial_j\varepsilon_k^{(h)}(\overline{\vartheta}_n)\partial_{l_1}\varepsilon_k^{(h)}(\widetilde{\vartheta}_n)^{\mathsf{T}}\left(\overline{\vartheta}_{n,l_1}-\vartheta_{l_1}^0\right).$$

The first term converges due to Theorem 5.4.5 to $\mathbb{E}\left[\partial_j\varepsilon_1^{(h)}(\vartheta^0)\varepsilon_1^{(h)}(\vartheta^0)^{\mathsf{T}}\right]$. For the second term we have to differ between two cases. First, if ∂_{l_1} is taken with respect to a non-stationary parameter, we have that $\left(\overline{\vartheta}_{n,l_1}-\vartheta_{l_1}^0\right)$ is of order $o_p(n^{-\gamma})$ for $0\leq\gamma<1$ due to the assumption. The condition (SE) implies that for n large enough $\frac{1}{n}\sum_{k=1}^{n}\partial_j\varepsilon_k^{(h)}(\overline{\vartheta}_n)\partial_{l_1}\varepsilon_k^{(h)}(\widetilde{\vartheta}_n)^{\mathsf{T}}$ is stochastically bounded. Then as a consequence we have convergence to zero. Secondly, if ∂_{l_1} is taken with respect to a stationary parameter we have that $\left(\overline{\vartheta}_{n,l_1}-\vartheta_{l_1}^0\right)$ is of order $o_p(1)$ and the analogous argumentation leads to the convergence to zero.

After these reflections, it remains to consider the last term of (5.57) which has not yet been investigated, i.e. we have due to Theorem 5.4.4

$$2\operatorname{tr}\left(\left(V_{\overline{\vartheta}_n}^{(h)}\right)^{-1}\frac{1}{n}\sum_{k=1}^{n}\partial_i^{st}\varepsilon_k^{(h)}(\overline{\vartheta}_n)\partial_j^{st}\varepsilon_k^{(h)}(\overline{\vartheta}_n)^{\mathsf{T}}\right) \xrightarrow{p} 2\operatorname{tr}\left(\left(V^{(h)}\right)^{-1}\mathbb{E}\left[\partial_i^{st}\varepsilon_1^{(h)}(\vartheta^0)\partial_j^{st}\varepsilon_1^{(h)}(\vartheta^0)^{\mathsf{T}}\right]\right).$$

Note that $\varepsilon_k^{(h)}(\vartheta^0)$ is orthogonal to the Hilbert space spanned by $\{Y_i, i<k\}$. But $\partial_i\varepsilon_k^{(h)}(\vartheta^0)$ as well as $\partial_{i,j}^2\varepsilon_k^{(h)}(\vartheta^0)$ are an element of $\{Y_i, i<k\}$. Furthermore, we have $\mathbb{E}\left[\varepsilon_1^{(h)}(\vartheta^0)\varepsilon_1^{(h)}(\vartheta^0)^{\mathsf{T}}\right]=V^{(h)}$. Thus, combining the results for all terms leads finally to

$$\begin{aligned}
&\partial_i^{st}\partial_j^{st}\mathcal{L}_n^{(h)}(\overline{\vartheta}_n) \qquad (5.75)\\
&\xrightarrow{p}\operatorname{tr}\left(\left(V^{(h)}\right)^{-1}\partial_{i,j}^2V^{(h)}-\left(V^{(h)}\right)^{-1}\left(\partial_iV^{(h)}\right)\left(V^{(h)}\right)^{-1}\left(\partial_jV^{(h)}\right)\right)\\
&\quad-\operatorname{tr}\left(\left(V^{(h)}\right)^{-1}\mathbb{E}\left[\varepsilon_1^{(h)}(\vartheta^0)\varepsilon_1^{(h)}(\vartheta^0)^{\mathsf{T}}\right]\left(V^{(h)}\right)^{-1}\partial_{i,j}^2V^{(h)}\right)\\
&\quad+\operatorname{tr}\left(\left(V^{(h)}\right)^{-1}\left(\partial_jV^{(h)}\right)\left(V^{(h)}\right)^{-1}\mathbb{E}\left[\varepsilon_1^{(h)}(\vartheta^0)\varepsilon_1^{(h)}(\vartheta^0)^{\mathsf{T}}\right]\left(V^{(h)}\right)^{-1}\partial_iV^{(h)}\right)\\
&\quad+\operatorname{tr}\left(\left(V^{(h)}\right)^{-1}\mathbb{E}\left[\varepsilon_1^{(h)}(\vartheta^0)\varepsilon_1^{(h)}(\vartheta^0)^{\mathsf{T}}\right]\left(V^{(h)}\right)^{-1}\left(\partial_jV^{(h)}\right)\left(V^{(h)}\right)^{-1}\partial_iV^{(h)}\right)\\
&\quad-\operatorname{tr}\left(\left(V^{(h)}\right)^{-1}\mathbb{E}\left[\partial_j\varepsilon_1^{(h)}(\vartheta^0)\varepsilon_1^{(h)}(\vartheta^0)^{\mathsf{T}}\right]\left(V^{(h)}\right)^{-1}\partial_iV^{(h)}\right)\\
&\quad+2\cdot\operatorname{tr}\left(\left(V^{(h)}\right)^{-1}\mathbb{E}\left[\varepsilon_1^{(h)}(\vartheta^0)\left(\partial_{i,j}^2\varepsilon_1^{(h)}(\vartheta^0)^{\mathsf{T}}\right)\right]\right)
\end{aligned}$$

$$
\begin{aligned}
&- 2 \cdot \operatorname{tr}\left(\left(V^{(h)}\right)^{-1}\mathbb{E}\left[\varepsilon_1^{(h)}(\vartheta^0)\partial_i\varepsilon_1^{(h)}(\vartheta^0)^\mathsf{T}\right]\left(V^{(h)}\right)^{-1}\partial_j V^{(h)}\right)\\
&+ 2 \cdot \mathbb{E}\left[\left(\partial_i\varepsilon_1^{(h)}(\vartheta^0)^\mathsf{T}\right)\left(V^{(h)}\right)^{-1}\left(\partial_j\varepsilon_1^{(h)}(\vartheta^0)\right)\right]\\
= &\operatorname{tr}\left(\left(V^{(h)}\right)^{-1}\left(\partial_i V^{(h)}\right)\left(V^{(h)}\right)^{-1}\partial_j V^{(h)}\right)\\
&+ 2 \cdot \mathbb{E}\left[\left(\partial_i\varepsilon_1^{(h)}(\vartheta^0)^\mathsf{T}\right)\left(V^{(h)}\right)^{-1}\left(\partial_j\varepsilon_1^{(h)}(\vartheta^0)\right)\right].
\end{aligned}
\tag{5.76}
$$

Step 3: Next we check that Z_{st} is positive definite with probability one, which we show by contradiction. In conclusion, we have as $n \to \infty$ the following representations, which are similar to the ones in Schlemm and Stelzer [91], Lemma 3.22, or respectively Boubacar and Francq [14], Lemma 4. From Step 2 we know that

$$
\nabla^2_{\vartheta_2}\mathcal{L}_n^{(h)}(\vartheta^0) \xrightarrow{p} Z_{st} = Z_{st,1} + Z_{st,2}, \tag{5.77}
$$

where

$$
\begin{aligned}
&Z_{st,1} := 2 \cdot \left[\mathbb{E}\left(\partial_i\varepsilon_1^{(h)}(\vartheta^0)^\mathsf{T}\right)\left(V^{(h)}\right)^{-1}\left(\partial_j\varepsilon_1^{(h)}(\vartheta^0)\right)\right]_{i,j}\\
\text{and}\quad &Z_{st,2} := \left[\operatorname{tr}\left(\left(V^{(h)}\right)^{-\frac{1}{2}}\left(\partial_i V^{(h)}\right)\left(V^{(h)}\right)^{-1}\partial_j V^{(h)}\left(V^{(h)}\right)^{-\frac{1}{2}}\right)\right]_{i,j}.
\end{aligned}
$$

We can factorize $Z_{st,2}$ in the following way

$$
Z_{st,2} = \begin{pmatrix} a_1 & \dots & a_{s_2}\end{pmatrix}^\mathsf{T}\begin{pmatrix} a_1 & \dots & a_{s_2}\end{pmatrix}
$$

with

$$
a_i := \left(\left(V^{(h)}\right)^{-\frac{1}{2}} \otimes \left(V^{(h)}\right)^{-\frac{1}{2}}\right)\operatorname{vec}\left(\partial_i V^{(h)}\right)
$$

and thus $Z_{st,1}$ and $Z_{st,2}$ as given in (5.77) are obviously positive semi-definite. It remains to check that for any $c \in \mathbb{R}^{s_2} \setminus \{0_{s_2}\}$ we have $c^\mathsf{T} Z_{st,i} c > 0$ for at least one $i \in \{1, 2\}$. We assume for the sake of contradiction that there exists a vector c such that

$$
c^\mathsf{T} Z_{st,1} c + c^\mathsf{T} Z_{st,2} c = 0. \tag{$\Diamond$}
$$

The matrix $Z_{st,1}$ can be rewritten similar as in the proof of Proposition 5.6.2 as

$$
2\mathbb{E}\left(\partial_i\varepsilon_k^{(h)}(\vartheta^0)^\mathsf{T}\right)\left(V_{\vartheta^0}^{(h)}\right)^{-1}\left(\partial_j\varepsilon_k^{(h)}(\vartheta^0)\right)
$$

$$
\begin{aligned}
&= 2\,\mathrm{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\sum_{l_1=0}^{\infty}\sum_{l_2=0}^{\infty}\partial_i^{st}K_{l_1}(\vartheta^0)\Gamma_Y(l_1-l_2)\partial_j^{st}K_{l_2}(\vartheta^0)^{\mathsf{T}}\right)\\
&= 2\sum_{l_1=0}^{\infty}\sum_{l_2=0}^{\infty}\mathrm{vec}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-\frac{1}{2}}\partial_i^{st}K_{l_1}(\vartheta^0)\Gamma_Y(l_1-l_2)^{\frac{1}{2}}\right)^{\mathsf{T}}\\
&\qquad\qquad\cdot\mathrm{vec}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-\frac{1}{2}}\partial_j^{st}K_{l_2}(\vartheta^0)\Gamma_Y(l_1-l_2)^{\frac{1}{2}}\right)\\
&= 2\sum_{l_1=0}^{\infty}\sum_{l_2=0}^{\infty}\mathrm{vec}\left(\left(\partial_i^{st}K_{l_1}(\vartheta^0)\right)^{\mathsf{T}}\left(\Gamma_Y(l_1-l_2)\otimes V_{\vartheta^0}^{(h)}\right)^{-1}\right)\mathrm{vec}\left(\partial_j^{st}K_{l_2}(\vartheta^0)\right).
\end{aligned}
$$

Furthermore, it holds that $\left(\Gamma_Y(l_1-l_2)\otimes\left(V_{\vartheta^0}^{(h)}\right)^{-1}\right)$ has $\mathbb{P}$-a.s. full rank and this implies that under the assumption ($\Diamond$)

$$
\nabla_{\vartheta_2}\left[[K_{\vartheta^0}^{\mathsf{T}}\otimes C_{\vartheta^0}]\mathrm{vec}\left(\left(\mathrm{e}^{A_\vartheta h}-K_\vartheta^{(h)}C_\vartheta\right)^{l-1}\right)\right]c = 0_{d^2}
$$

must hold for all $l\in\mathbb{N}$. Analogous as in [91, Proof of Lemma 2.17.] we obtain that due to the assumption ($\Diamond$) the existence of a $c\neq 0_{s_2}$ such that $\nabla_{\vartheta_2}\mathrm{vec}\big(V_{\vartheta^0}^{(h)}\big)c=0$. The definition of $\psi_{\vartheta,j}$ in (5.74) implies that $\nabla_{\vartheta_2}\psi_{\vartheta^0,j_0}c=0_{(j+2)d^2}$ holds for all $j\in\mathbb{N}$, which is in contradiction with Assumption M13. Hence, Z_{st} is almost surely positive definite. □

Finally, we can state and prove the last main result of this chapter, namely, the asymptotic normality of the short-run estimator.

Theorem 5.6.8

Assume that Assumption M1 - Assumption M13 hold then we have as $n\to\infty$

$$
\sqrt{n}(\widehat{\vartheta}_{n,2}-\vartheta_2^0)\xrightarrow{w}\mathcal{N}(0,\Xi). \tag{5.78}
$$

The asymptotic covariance matrix of the stationary part is given by

$$
\Xi := Z_{st}^{-1}IZ_{st}^{-1}
$$

where

$$
I=\lim_{n\to\infty}Var\left(\nabla_{\vartheta_2}\mathcal{L}_n^{(h)}(\vartheta^0)\right)\quad\text{and}\quad Z_{st}=\lim_{n\to\infty}\nabla_{\vartheta_2}^2\mathcal{L}_n^{(h)}(\vartheta^0).
$$

Proof. We have shown in Theorem 5.5.5 the consistency of the estimator $\widehat{\vartheta}_n$. Due to Assumption M5 the estimator $\widehat{\vartheta}_{n,2}$ is at some point also an element of the interior of Θ_2 with probability one.

With the first order condition and a Taylor expansion of the score vector around the point ϑ_2^0 we obtain that there exist $\underline{\vartheta}_{n,i} \in \Theta$ of the form $\underline{\vartheta}_{n,i} = \vartheta^0 + c_i(\widehat{\vartheta}_n - \vartheta^0)$, $0 \leq c_i \leq 1$ such that

$$0_{s_2} = \sqrt{n}\nabla_{\vartheta}\mathcal{L}_n^{(h)}(\widehat{\vartheta}_{n,1}, \vartheta_2^0) + \nabla_{\vartheta}^2\mathcal{L}_n^{(h)}(\widehat{\vartheta}_{n,1}, \underline{\vartheta}_n)\sqrt{n}(\widehat{\vartheta}_{n,2} - \vartheta_2^0). \tag{5.79}$$

The first term is asymptotically normal due to Proposition 5.6.6. Note that due to Theorem 5.5.5 we have $\widehat{\vartheta}_{n,1} \in N_{n,\gamma}(\vartheta_1^0, \delta)$ for some $\delta > 0$. By Proposition 5.6.7 follows that $\nabla_{\vartheta}^2\mathcal{L}_n^{(h)}(\widehat{\vartheta}_{n,1}, \underline{\vartheta}_n)$ converges in probability in its continuous form to the matrix $Z_{st}(\vartheta^0)$, which is $\mathbb{P}$-a.s. a constant positive definite matrix. Hence, we obtain

$$\sqrt{n}(\widehat{\vartheta}_{n,2} - \vartheta_2^0) = -\left(\nabla_{\vartheta_2}^2\mathcal{L}_n^{(h)}(\widehat{\vartheta}_{n,1}, \underline{\vartheta}_n)\right)^{-1} \cdot \sqrt{n}\nabla_{\vartheta_2}\mathcal{L}_n^{(h)}(\vartheta^0).$$

Slutzky's theorem shows finally the claim. □

5.7. Conclusion

We developed in this paper a method to estimate the model parameters of a cointegrated Lévy driven MCARMA model based on equidistant observations in discrete time. For this purpose, we used a step-wise quasi-maximum likelihood approach. The estimation procedure works for quite general Lévy processes with finite $(4+\delta)$-th moments.

We separated the parameter space into a non-stationary and a stationary parameter space resulting in long-run and short-run parameter estimators. Using this splitting of the parameter vector, we showed the super-consistency of the long-run QMLE and the consistency of the short-run QMLE. Furthermore, we derived the asymptotic distributions of these estimators using the concept of weak continuous convergence, a stochastic equicontinuity condition and a Taylor series expansion. The long-run parameter estimator is mixed normal distributed, whereas the short-run parameter estimator is asymptotically normal as usually for a stationary QMLE. These results are in accordance with the results derived for discrete-time cointegrated models (c.f. for example Yap and Reinsel [99]). Besides, we showed the asymptotic independence of the estimators.

The identifiability problem is solved by a set of sufficient conditions on the parametrization, which guarantees that different values of the parameter generate different probability distributions of the observable variables. These identifiability conditions also avoid the aliasing effect. The assumptions we require are very similar to the stationary case as well as assumptions in the discrete-time case. Mainly, we need only some assumptions on the (co)integrated part, which are obviously not necessary in the stationary case.

5.8. Appendix: Asymptotic Results

We derive several necessary asymptotic results in the following. First, we state a law of large numbers for the innovations in the first auxiliary lemma. Then we derive limit results for parts of $\Delta Y_k^{(h)}$ and $Y_{k-1}^{(h)}$ in the next lemmas, which we use in the derivation of the results in Chapter 5.

Assume that throughout this chapter the assumptions of Section 4.4 are satisfied. We know from Proposition 4.4.8 that the innovation sequence ε is a stationary ergodic sequence in the case of a driving Lévy process. On the other side, in the case of a driving Brownian motion we even have that the linear innovations form an i.i.d. sequence. However, the derived properties for a driving Lévy process are sufficient for the asymptotic theory, which we derive in Chapter 5.

References to the asymptotic results needed for the subsequent proofs can be found e.g. in the papers of Phillips et al. [72], [78],[75], [76], [73] and [79]. A good summary of these results for an i.i.d. noise can be found in Lütkepohl [62], Appendix C, Hamilton [45], Chapter 18, or Johansen [54], Appendix B.7. Moreover, one can find references for asymptotic results also in Pötscher and Prucha [80] and [81]. For limit results and the general theory concerning ergodic processes, see e.g. Krengel [59], Bradley [16], Billingsley [12] or Doukhan [31].

The first result is a law of large numbers for the linear innovations, which follows due to the ergodicity and stationary of the linear innovations.

Lemma 5.8.1
The innovations ε_k sequence satisfies

$$n^{-1}\sum_{k=1}^{n-h}\varepsilon_k^{(h)}\varepsilon_{k+h}^{(h)\mathsf{T}} \xrightarrow{a.s.} \delta_{0,l}\mathbb{E}\left[\varepsilon_1^{(h)}\varepsilon_1^{(h)\mathsf{T}}\right] = \delta_{0,l}V^{(h)} = \delta_{0,l}C\Omega^{(h)}C^{\mathsf{T}}, \tag{5.80}$$

where $\delta_{0,l} = 1$ for $l = 0$ and zero else.

Proof. Note that $\left(\varepsilon_k^{(h)}\varepsilon_{k+h}^{(h)}\right)_{k\in\mathbb{N}}$ is stationary and ergodic due to the same argumentation as in the proof of Proposition 4.4.8 and $\mathbb{E}|\varepsilon_k^{(h)}\varepsilon_{k+h}^{(h)}| < \infty$. Thus, the proof is an immediate consequence of a law of large numbers for ergodic stationary processes

(Birkhoff's Ergodic Theorem) and equation (4.35). As a reference for the law of large numbers see e.g. Durrett [32], Theorem 7.2.1, or Bradley [16], 2.3 Ergodic Theorem. □

Recall the representations for $Y^{(h)}$ in Lemma 4.3.15, which gives us the following representations

$$Y_n^{(h)} = C_1 X_1^{(h)}(0) + C_1 B_1 L(nh) + Y_{n,2}^{(h)} \tag{5.81}$$

and $\Delta Y^{(h)}$ in (4.28)

$$\begin{aligned} \Delta Y_n^{(h)} &= C_1 R_{n,1}^{(h)} + \Delta Y_{n,2}^{(h)} \\ &= C_1 R_{n,1}^{(h)} + C_2 \sum_{j=0}^{\infty} \mathrm{e}^{A_2 h j} R_{n-j,2}^{(h)} - C_2 \sum_{j=0}^{\infty} \mathrm{e}^{A_2 h j} R_{n-1-j,2}^{(h)} \\ &= C_1 R_{n,1}^{(h)} + \sum_{j=0}^{\infty} \widetilde{\Psi}_j R_{n-j,2}^{(h)}, \end{aligned} \tag{5.82}$$

where the matrix polynomial $\Psi(z)$ is given by

$$\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j = I_d + \sum_{j=1}^{\infty} C_2 \mathrm{e}^{A_2 h j} z^j \tag{5.83}$$

and the new matrix polynomial is defined by $\widetilde{\Psi}(z) := (1-z)\Psi(z)$ with coefficient matrix $\widetilde{\Psi}(z)$, i.e. $\widetilde{\Psi}_i = \Psi_i - \Psi_{i-1}$ and $\widetilde{\Psi}_0 = \Psi_0 = I_d$ for $i \in \mathbb{N}$.

We consider now the properties of the moving average part of $\Delta Y^{(h)}$ and the integrated part of $Y^{(h)}$. Note that the first order difference of an infinite order moving average process is again a moving average process.

The covariance matrix of the difference process is denoted by

$$\Gamma_{\Delta Y}(l) := \mathbb{E}\left[\Delta Y_k^{(h)} \Delta Y_{k+l}^{(h)\mathsf{T}}\right], \quad \text{for } l \in \mathbb{N}_0, \tag{5.84}$$

and the covariance matrix for the stationary part $Y_2^{(h)}$ is analogously by

$$\Gamma_{Y_2}(l) := \mathbb{E}\left[Y_{k,2}^{(h)} Y_{k+l,2}^{(h)\mathsf{T}}\right], \quad \text{for } l \in \mathbb{N}_0. \tag{5.85}$$

The stationary part of $Y^{(h)}$ has a moving average representation with absolutely summable matrix coefficients Ψ_i and the same holds for the matrix coefficients $\widetilde{\Psi}_i$

the moving average representation of $\Delta Y^{(h)}$, i.e. we have

$$\sum_{s=0}^{\infty} s\|\Psi_s\| < \infty \qquad \text{and} \qquad \sum_{s=0}^{\infty} s\|\widetilde{\Psi}_s\| < \infty. \tag{5.86}$$

The absolute summability of $\|\Psi_s\|$ follows directly by

$$\sum_{s=0}^{\infty} \|\Psi_s\| \leq \|I_d\| + \sum_{s=1}^{\infty} \|C_2\| \left\|\mathrm{e}^{A_2 hs}\right\| < \infty,$$

due to the negative real part of the eigenvalues of A_2. Now we obtain with $\Psi_{-1} = 0_d$ by using the last result

$$\sum_{s=0}^{\infty} \|\widetilde{\Psi}_s\| = \sum_{s=0}^{\infty} \|\Psi_s - \Psi_{s-1}\| \leq \sum_{s=0}^{\infty} \|\Psi_s\| + \sum_{s=1}^{\infty} \|\Psi_{s-1}\| < \infty.$$

Lastly, $\sum_{s=1}^{\infty} s\|\Psi_s\| < \infty$ also holds true due to the exponentially decreasing matrix coefficients. Thus, with the previous inequality this holds also true for $\widetilde{\Psi}$. Moreover, since the linear innovations ε are a stationary ergodic process the same holds for $\Delta Y^{(h)}$ and $Y_2^{(h)}$, which can easily be seen as for ε itself.

Lemma 5.8.2
We have for $l \in \mathbb{N}_0$

i) $n^{-1} \sum_{k=1}^{n} Y_{k-1,2}^{(h)} \xrightarrow{a.s.} \mathbb{E}[Y_{k-1,2}^{(h)}] = 0$;

ii) $n^{-1} \sum_{k=1}^{n} Y_{k-1,2}^{(h)} Y_{k+l-1,2}^{(h)\mathsf{T}} \xrightarrow{a.s.} \mathbb{E}[Y_{k-1,2}^{(h)} Y_{k+l-1,2}^{(h)\mathsf{T}}] = \Gamma_{Y_2}(l)$;

iii) $n^{-1} \sum_{k=1}^{n} \Delta Y_k^{(h)} \Delta Y_{k+l}^{(h)\mathsf{T}} \xrightarrow{a.s.} \mathbb{E}[\Delta Y_k^{(h)} \Delta Y_{k+l}^{(h)\mathsf{T}}] = \Gamma_{\Delta Y}(l)$;

iv) $n^{-1} \sum_{k=1}^{n} Y_{k-1,2}^{(h)} \Delta Y_{k+l}^{(h)^{(h)\mathsf{T}}} \xrightarrow{a.s.} \mathbb{E}[Y_{k-1,2}^{(h)} \Delta Y_{k+l}^{(h)\mathsf{T}}] = \Gamma_{Y_2 \Delta Y}(l)$.

Proof. The moments exists for theses processes due to the results in Section 4.3. The results all follow by Birkhoff's Ergodic Theorem. This can easily be seen since ε is a stationary ergodic process and we consider in all cases measurable functions of ε, which inherit by Bradley [16], Proposition 2.10 (ii), the stationarity and ergodicity. □

We can restrict the temporal dependence of the first difference of the process.

Lemma 5.8.3
It holds that $\sum_{l=-\infty}^{\infty} \sup_n \|\mathbb{E}\Delta Y_n^{(h)} \Delta Y_{n+l}^{(h)\mathsf{T}}\| < \infty$.

Proof. The supremum can be neglected due to the stationarity of the process. Furthermore, it suffices to consider $l \in \mathbb{N}$ instead of $l \in \mathbb{Z}$ due to the symmetry. Note that ΔY is strongly mixing due Lemma 4.3.18 combined Remark 1.8 b) in Bradley [16] and thus we can apply the covariance inequality (Lemma 2.13, Schlemm and Stelzer). Recall that $(R_n^{(h)})_{n\in\mathbb{N}} = \begin{pmatrix} R_{n,1}^{(h)\mathsf{T}} & R_{n,2}^{(h)\mathsf{T}} \end{pmatrix}^{\mathsf{T}}_{n\in\mathbb{N}}$ is an i.i.d. sequence. Hence, we obtain by (5.82) that there exists constants $c_1, c_2 > 0$ such that

$$
\begin{aligned}
&\sum_{l=0}^{\infty} \|\mathbb{E}\Delta Y_n^{(h)} \Delta Y_{n+l}^{(h)\mathsf{T}}\| \\
&= \sum_{l=0}^{\infty} \left\| \mathbb{E}\left(C_1 R_{n,1}^{(h)} + \sum_{j=0}^{\infty} \widetilde{\Psi}_j R_{n-j,2}^{(h)} \right) \left(C_1 R_{n+l,1}^{(h)} + \sum_{j=0}^{\infty} \widetilde{\Psi}_j R_{n+l-j,2}^{(h)} \right)^{\mathsf{T}} \right\| \\
&\leq \left\| \mathbb{E} C_1 R_{1,1}^{(h)} R_{1,1}^{(h)\mathsf{T}} C_1^{\mathsf{T}} \right\| + \left\| \mathbb{E}\widetilde{\Psi}_0 R_{1,2}^{(h)} R_{1,1}^{(h)\mathsf{T}} C_1^{\mathsf{T}} \right\| + \sum_{l=0}^{\infty} \sum_{j=0}^{\infty} \left\| \mathbb{E} C_1 R_{n,1}^{(h)} R_{n+l-j,2}^{(h)\mathsf{T}} \widetilde{\Psi}_j^{\mathsf{T}} \right\| \\
&\quad + \sum_{l=0}^{\infty} \left\| \mathbb{E}\Delta Y_{n,2}^{(h)} \Delta Y_{n+l,2}^{(h)\mathsf{T}} \right\| \\
&\leq c_2 + \sum_{l=0}^{\infty} c_1 \cdot \|e^{A_2 hl}\| \left\| \int_0^h C_1 B_1 \Sigma_L^{(h)} B_2^{\mathsf{T}} e^{A_2^{\mathsf{T}} hu} \mathrm{d}u \right\| + c_1 \cdot \sum_{l=0}^{\infty} \left[\alpha_{Y_2^{(h)}}(l) \right]^{\frac{\delta}{2+\delta}} < \infty,
\end{aligned}
$$

due to the finite expectations, the eigenvalues with non-negative real part of A_2 and Lemma 4.3.18. □

Lastly we show some convergence results with respect to the cointegrated process.

Proposition 5.8.4

We have the following asymptotic results for $0 \leq r \leq 1$

i) $n^{-\frac{1}{2}} Y_{[nr]}^{(h)} \xrightarrow{w} C_1 B_1 W(r)$;

ii) $n^{-\frac{3}{2}} \sum_{k=1}^{n} Y_{k-1}^{(h)} \xrightarrow{w} \int_0^1 C_1 B_1 W(r)\, \mathrm{d}r$;

iii) $n^{-2} \sum_{k=1}^{n} Y_{k-1}^{(h)} Y_{k-1}^{(h)\mathsf{T}} \xrightarrow{w} \int_0^1 C_1 B_1 W(r) W(r)^{\mathsf{T}} B_1^{\mathsf{T}} C_1^{\mathsf{T}}\, \mathrm{d}r$;

iv) $$n^{-1} \sum_{k=1}^{n} Y_{k-1}^{(h)} \Delta Y_k^{(h)\mathsf{T}} \xrightarrow{w} \big(C_1 B_1, 0_{d\times(N-c)}\big) \int_0^1 W(r)\, \mathrm{d}W(r)^{\mathsf{T}} \big(C_1 B_1, \widetilde{\Psi}(1)\big)^{\mathsf{T}} + \Gamma_{Y\Delta Y},$$

v) $$n^{-1} \sum_{k=1}^{n} Y_{k-1}^{(h)} Y_{k-1,2}^{(h)\mathsf{T}} \xrightarrow{w} \big(C_1 B_1, 0_{d\times(N-c)}\big) \int_0^1 W(r)\, \mathrm{d}W(r)^{\mathsf{T}} \big(0_{d\times c}, \Psi(1)\big)^{\mathsf{T}} + \Gamma_{YY_2},$$

where W is a Brownian motion with covariance matrix $\Sigma_L^{(h)}$ and Γ_{YY_2} as well as $\Gamma_{Y\Delta Y}$ are covariance matrices defined in the proof.

Proof. i) We have

$$n^{-\frac{1}{2}}Y_{\lfloor nr\rfloor}^{(h)} = n^{-\frac{1}{2}}Y_{\lfloor nr\rfloor,2}^{(h)} + n^{-\frac{1}{2}}C_1X_1^{(h)}(0) + n^{-\frac{1}{2}}\sum_{i=1}^{\lfloor nr\rfloor} C_1R_{i,1}^{(h)}.$$

Note that $C_1R_{i,1}^{(h)}$ is an i.i.d. sequence with mean zero and finite positive definite covariance matrix $\mathbb{E}R_{i,1}^{(h)}R_{i,1}^{(h)\mathsf{T}} = hB_1\Sigma_LB_1^{\mathsf{T}}$. The claim follows with a functional central limit theorem (see e.g. Johansen [54], Theorem B.12) combined with Slutzky's Theorem if the remaining part is $o_p(1)$. We use the Markov inequality and obtain for all $\epsilon > 0$

$$\begin{aligned}&\lim_{n\to\infty}\mathbb{P}\left(\left\|n^{-\frac{1}{2}}Y_{\lfloor nr\rfloor,2}^{(h)}\right\| \geq \epsilon\right) = \lim_{n\to\infty}\mathbb{P}\left(\left\|Y_{\lfloor nr\rfloor,2}^{(h)}\right\| \geq \sqrt{n}\epsilon\right)\\ &\leq \lim_{n\to\infty}\frac{1}{\sqrt{n}\epsilon}\mathbb{E}\left\|Y_{1,2}^{(h)}\right\| \longrightarrow 0.\end{aligned}$$

Hence, the same holds true for the starting value and we have shown the claim.

ii) We use the functional central limit theorem from i) and the continuous mapping theorem. Using equation (5.81), we have the following representation

$$\begin{aligned}n^{-\frac{3}{2}}\sum_{k=1}^{n}Y_{k-1}^{(h)} &= n^{-\frac{3}{2}}\sum_{k=1}^{n}Y_{k-1,2}^{(h)} + n^{-\frac{1}{2}}C_1X_1^{(h)}(0) + n^{-\frac{3}{2}}\sum_{k=1}^{n}\sum_{i=1}^{k-1}C_1R_{i,1}^{(h)}\\ &= \int_0^1\left(n^{-\frac{1}{2}}\sum_{i=1}^{[nr]}C_1R_{i,1}^{(h)}\right)\mathrm{d}r + o_p(1)\\ &\xrightarrow{w}\int_0^1 C_1B_1W(r)\,\mathrm{d}r.\end{aligned}$$

Recall that $Y_{k-1,2}^{(h)}$ is stationary and ergodic. Thus, the $o_p(1)$ term follows due Lemma 5.8.2 i) and for the starting value with the same argument as in the aforementioned part.

iii) Due to the representation given in equation (5.81), it holds by similar arguments as before and Lemma 5.8.2 that

$$n^{-2}\sum_{k=1}^{n}Y_{k-1}^{(h)}Y_{k-1}^{(h)\mathsf{T}} = n^{-2}\sum_{k=1}^{n}\sum_{i=1}^{k-1}C_1R_{i,1}^{(h)}R_{i,1}^{(h)\mathsf{T}}C_1^{\mathsf{T}} + o_p(1)$$

$$\xrightarrow{w} \int_0^1 C_1 B_1 W(r) W(r)^\mathsf{T} B_1^\mathsf{T} C_1^\mathsf{T} \,\mathrm{d}r.$$

iv) As already mentioned $R^{(h)}$ is an i.i.d. noise with mean zero and positive definite covariance matrix given in (5.2). We use Theorem B.13 of Johansen [54]. This convergence results allows different linear filters. The linear filter in our case are given by $(C_1, 0_{d\times(N-c)})$ and $(C_1, \widetilde{\Psi}(z))$. Due to the exponentially decreasing matrix coefficients of $\widetilde{\Psi}(z)$ the assumptions of the theorem are satisfied.

Thus, Theorem B.13 of Johansen combined with (5.82) and Lemma 5.8.2 iv) yields

$$\begin{aligned}
&n^{-1}\sum_{k=1}^{n} Y_{k-1}^{(h)} \Delta Y_k^{(h)\mathsf{T}} \\
&= n^{-1}\sum_{k=1}^{n}\sum_{i=1}^{k-1} C_1 R_{i,1}^{(h)} \Delta Y_k^{(h)\mathsf{T}} + n^{-1}\sum_{k=1}^{n} Y_{k-1,2}^{(h)} \Delta Y_k^{(h)\mathsf{T}} + o_p(1) \\
&\xrightarrow{w} \big(C_1 B_1, 0_{d\times(N-c)}\big) \int_0^1 W(r)\,\mathrm{d}W(r)^\mathsf{T} \big(C_1 B_1, \widetilde{\Psi}(1)\big)^\mathsf{T} \\
&\qquad + \sum_{l=1}^{\infty} (C_1, 0_{d\times(N-c)}) \widetilde{\Sigma}^{(h)} (C_1, \widetilde{\Psi}_l)^\mathsf{T} + \Gamma_{Y_2 \Delta Y}(0).
\end{aligned}$$

We set for reasons of brevity

$$\Gamma_{Y\Delta Y} := \sum_{l=1}^{\infty} (C_1, 0_{d\times(N-c)}) \widetilde{\Sigma}^{(h)} (C_1, \widetilde{\Psi}_l)^\mathsf{T} + \Gamma_{Y_2 \Delta Y}(0).$$

v) This claim follows similar to part iv). To sum it up, we can conclude

$$\begin{aligned}
&n^{-1}\sum_{k=1}^{n} Y_{k-2}^{(h)} Y_{k-1,2}^{(h)\mathsf{T}} \\
&= n^{-1}\sum_{k=1}^{n}\sum_{i=1}^{k-1} C_1 R_{i,1}^{(h)} Y_{k-1,2}^{(h)\mathsf{T}} + n^{-1}\sum_{k=1}^{n} Y_{k-1,2}^{(h)} Y_{k-1,2}^{(h)\mathsf{T}} + o_p(1) \\
&\xrightarrow{w} \big(C_1 B_1, 0_{d\times(N-c)}\big) \int_0^1 W(r)\,\mathrm{d}W(r)^\mathsf{T} \big(0_{d\times c}, \Psi(1)\big)^\mathsf{T} + \Gamma_{YY_2},
\end{aligned}$$

where Γ_{YY_2} is defined analogous to $\Gamma_{Y\Delta Y}$.

□

In the end, we have a foundation of asymptotic results on the processes appearing in the cointegrated model, which helps us in the proofs in Chapter 5. Furthermore, note that we have shown Assumption 4.3 of Saikkonen [85] with the results of this chapter.

5.9. Appendix: Auxiliary Results and Proofs of Chapter 5

5.9.1. Auxiliary Results of Chapter 5

We derive in this subsection several technical results which we need in the proofs of Chapter 5.

Lemma 5.9.1
Assumption M1-Assumption M6 imply that the following assertions hold for all $\vartheta \in \Theta$:

i) The functions $\vartheta \mapsto \mathrm{e}^{A_{2,\vartheta}h}$ and $\vartheta \mapsto \mathbb{E}R_{\vartheta,n}^{(h)}R_{\vartheta,n}^{(h)\mathsf{T}}$ are continuous.

ii) The covariance matrix $\widetilde{\Sigma}_{\vartheta}^{(h)} = \mathbb{E}R_{\vartheta,n}^{(h)}R_{\vartheta,n}^{(h)\mathsf{T}}$ is positive definite.

iii) The matrix $V_{\vartheta}^{(h)} = C_{\vartheta}\Omega_{\vartheta}^{(h)}C_{\vartheta}^{\mathsf{T}}$ is non-singular.

Proof. The proof is analogous to the proof of Lemma 3.14 in Schlemm and Stelzer [91].

i) The continuity is obvious since by Assumption M6 we have a composition of continuous functions.

ii) The positive-definiteness of $\widetilde{\Sigma}_{\vartheta}^{(h)}$ follows by Corollary 3.9 in [91]. The assumptions of Corollary 3.9 are satisfied due to Assumption M1 and Assumption M3.

iii) The non-singularity of $V_{\vartheta}^{(h)}$ follows by the full rank condition on C_{ϑ} in Assumption M2 and the non-singularity of $\Omega_{\vartheta}^{(h)}$, which follows by Proposition 4.4.2.

□

Note that $\Omega_{(\cdot)}^{(h)}$ is a continuous functions of the coefficient matrices (see Schlemm and Stelzer [91], Proof of Lemma 2.2 or Sun [95]). Hence, the steady-state Kalman gain matrix $K_{(\cdot)}^{(h)}$ is also a continuous function and this implies the continuity of $\Pi(\cdot)$ and $\overline{k}(z, \cdot)$. Due to the same reason $V_{(\cdot)}^{(h)}$ is as well a continuous function.

Some bounds on the matrix functions are given in the next lemma.

Lemma 5.9.2
Under Assumption M1-Assumption M6 the following results hold.

i) There exists a positive number $\rho \leq 1$ *such that for all* $\vartheta \in \Theta$ *it holds that*

$$max\{|\lambda| : \lambda \in \sigma(\mathrm{e}^{A_\vartheta h})\} \leq \rho.$$

Furthermore, there exists a positive number $\rho < 1$ *such that for all* $\vartheta \in \Theta$ *it holds that*

$$max\{|\lambda| : \lambda \in \sigma(\mathrm{e}^{A_{\vartheta,2}h})\} \leq \rho.$$

ii) There exists a positive number $\rho < 1$ *such that for all* $\vartheta \in \Theta$ *it holds that*

$$max\{|\lambda| : \lambda \in \sigma(\mathrm{e}^{A_\vartheta h} - K_\vartheta^{(h)} C_\vartheta)\} \leq \rho.$$

iii) There exists a positive number c *such that* $\left\| \left(V_\vartheta^{(h)}\right)^{-1} \right\| \leq c$ *for all* $\vartheta \in \Theta$.

Proof. The result of i) and iii) is a direct consequence of Lemma 2.2 in Schlemm and Stelzer [91]. Note that the second part of *i)* follows directly due to the decoupled state space form. Part iii) follows directly due to the derivation of the Kalman filter in Section 4.6. □

Next, we check that the differentiability of the parametrization transfers to the other matrices occurring in the following.

Lemma 5.9.3
Let Assumption M1-Assumption M6 and Assumption M8 hold. Then the following functions are all twice continuously differentiable as well:

1. $\vartheta \mapsto \exp(A_\vartheta)$;
2. $(A_\vartheta, B_\vartheta, \Sigma_\vartheta^L) \mapsto \int_0^h \mathrm{e}^{A_\vartheta u} B_\vartheta \Sigma_\vartheta^L B_\vartheta^\mathsf{T} \mathrm{e}^{A_\vartheta^\mathsf{T} u} \mathrm{d}u$;
3. $\vartheta \mapsto \Pi(\vartheta)$;
4. $\vartheta \mapsto \overline{k}(z, \vartheta)$.

Moreover, the functions $\Pi(\cdot)$, $\overline{k}(z,\cdot)$ *and* $\left(V_{(\cdot)}^{(h)}\right)^{-1}$ *are Lipschitz continuous, i.e. we have for some constants* $0 < c_V, c_\Pi, c_{\overline{k}} < \infty$ *that* $\|\Pi(\vartheta) - \Pi(\vartheta')\| \leq c_\Pi \|\vartheta - \vartheta'\|$, $\|\overline{k}(z,\vartheta) - \overline{k}(z,\vartheta')\| \leq c_{\overline{k}} \|\vartheta - \vartheta'\|$ *and* $\left\| \left(V_\vartheta^{(h)}\right)^{-1} - \left(V_{\vartheta'}^{(h)}\right)^{-1} \right\| \leq c_V \|\vartheta - \vartheta'\|$, *whenever* ϑ *and* ϑ' *are in* Θ.

Proof. The first two functions are twice continuously differentiable (c.f. Schlemm and Stelzer [91], Proof of Lemma 3.15). Note that due to the results on the algebraic Riccati equation in Sun [95] the matrix function $V_{(\cdot)}^{(h)}$ is twice continuous differentiable.

We prove the twice continuous differentiability of $\Pi(\cdot)$ and $\overline{k}(z,\cdot)$ together by showing that $k(z,\cdot)$ is twice continuous differentiable. Recall that the transfer function has the form $k(z,\vartheta) = I_d - C_\vartheta \sum_{l=1}^{\infty}(\mathrm{e}^{A_\vartheta h} - K_\vartheta^{(h)} C_\vartheta)^{l-1} K_\vartheta^{(h)} z^l$. Due to the uniform exponential bound of the matrix coefficients and the fact that the partial derivatives $\partial_i k(z,\vartheta)$ as well as $\partial_{i,j}^2 k(z,\vartheta)$ are also uniformly exponentially bounded due to Lemma 5.2.1 and Lemma 5.2.2, we can exchange summation and differentiation. Since we have only sums and products of at least twice continuous differentiable functions, $k(z,\cdot)$ is twice continuous differentiable itself. The same argumentation holds for $\Pi(\cdot)$ and $\overline{k}(z,\cdot)$ holds and the assertion follows.

Due to the continuous differentiability and the compact parameter space we receive that $\Pi(\cdot)$, $\overline{k}(z,\cdot)$ and $\left(V_{(\cdot)}^{(h)}\right)^{-1}$ are Lipschitz continuous. □

As already mentioned, we need the existence of higher moments, finite second moments are not sufficient to derive the aforementioned results. The assumption of finite moments of the Lévy process transfers to the other processes as can be seen in the following results.

The finite moments of the Lévy process due to Assumption M1 imply finite moments of the i.i.d. noise of the sampled process.

Lemma 5.9.4
The noise $R_\vartheta^{(h)}$ is given as in (4.20). It follows that $\mathbb{E}\|R_\vartheta^{(h)}\|^{4+\delta} < \infty$.

Proof. This follows similar to Lemma 3.15 in Schlemm and Stelzer [91]. □

Moreover, the pseudo-innovations and its partial derivatives have also finite fourth moment.

Lemma 5.9.5
It holds that $\mathbb{E}\|\varepsilon_k^{(h)}(\vartheta)\|^4 < \infty$ and $\mathbb{E}\|\partial_i \varepsilon_k^{(h)}(\vartheta)\|^4 < \infty$, for $i = 1, \ldots, s$.

Proof. Note that the following inequality $(a+b)^4 \leq 2^3(a^4+b^4)$ holds due to the concavity. We use (5.10) and obtain

$$\begin{aligned}
\mathbb{E}\|\varepsilon_k^{(h)}(\vartheta)\|^4 &\leq \mathbb{E}\left\|\Pi(\vartheta)^\mathsf{T} Y_{k-1}^{(h)} + \sum_{i=0}^{\infty} \widetilde{K}_i(\vartheta)\Delta Y_{k-i}^{(h)}\right\|^4 \\
&\leq C\cdot\mathbb{E}\left\|\Pi(\vartheta)^\mathsf{T} Y_{k-1}^{(h)}\right\|^4 + C\cdot\mathbb{E}\left\|\sum_{i=0}^{\infty} \widetilde{K}_i(\vartheta)\Delta Y_{k-i}^{(h)}\right\|^4 .
\end{aligned}$$

The matrix coefficients $\widetilde{K}_i(\vartheta)$ decay exponentially fast due to Lemma 5.2.1. Hence, we have using Assumption M1 and the finite fourth moments of $Y^{(h)}$ in the same manner as in the proof of Lemma 3.16 in Schlemm [89] that $\mathbb{E}\|\varepsilon_k^{(h)}(\vartheta)\|^4 < \infty$. Similarly we can derive the second statement with Lemma 5.2.2 i). □

5.9.2. Representation of Asymptotically Dominant Part

We see that the interesting parts for the asymptotic behavior are for $\mathcal{L}_{n,1}^{(h)}(\vartheta)$ given by

$$\begin{aligned}\mathcal{Q}_{n,1}^{(h)}(\vartheta) := \operatorname{tr}&\left(\frac{1}{n}\sum_{k=1}^{n}\left(V_{\vartheta}^{(h)}\right)^{-1}\varepsilon_{k,1}^{(h)}(\vartheta)\varepsilon_{k,1}^{(h)}(\vartheta)^{\mathsf{T}}\right)\\
&+2\cdot\operatorname{tr}\left(\frac{1}{n}\sum_{k=1}^{n}\left(V_{\vartheta}^{(h)}\right)^{-1}\varepsilon_{k,2}^{(h)}(\vartheta_2)\varepsilon_{k,1}^{(h)}(\vartheta)^{\mathsf{T}}\right)\\
&+\operatorname{tr}\left(\frac{1}{n}\sum_{k=1}^{n}\left[\left(V_{\vartheta}^{(h)}\right)^{-1}-\left(V_{\vartheta_2}^{(h)}\right)^{-1}\right]\varepsilon_{k,2}^{(h)}(\vartheta_2)\varepsilon_{k,2}^{(h)}(\vartheta_2)^{\mathsf{T}}\right)\end{aligned}$$

and for $\mathcal{L}_{n,2}^{(h)}(\vartheta_2)$ by

$$\mathcal{Q}_{n,2}^{(h)}(\vartheta_2) := \frac{1}{n}\sum_{k=1}^{n}\operatorname{tr}\left(\left(V_{\vartheta_2}^{(h)}\right)^{-1}\varepsilon_{k,2}^{(h)}(\vartheta_2)\varepsilon_{k,2}^{(h)}(\vartheta_2)^{\mathsf{T}}\right).$$

Using the representation (5.5) of the pseudo-innovations we can rewrite these expressions. Doing this for $\mathcal{Q}_{n,1}^{(h)}(\vartheta)$ gives us

$$\begin{aligned}\mathcal{Q}_{n,1}^{(h)}(\vartheta) &= \frac{1}{n}\sum_{k=1}^{n}\operatorname{tr}\left[\left(V_{\vartheta}^{(h)}\right)^{-1}\left[\Pi(\vartheta_1,\vartheta_2)-\Pi(\vartheta_1^0,\vartheta_2)\right]Y_{k-1}^{(h)}Y_{k-1}^{(h)\mathsf{T}}\left[\Pi(\vartheta_1,\vartheta_2)-\Pi(\vartheta_1^0,\vartheta_2)\right]^{\mathsf{T}}\right]\\
&+\frac{1}{n}\sum_{k=1}^{n}\operatorname{tr}\Bigg(\left(V_{\vartheta}^{(h)}\right)^{-1}\\
&\cdot\Big(-2\cdot\left[\Pi(\vartheta_1,\vartheta_2)-\Pi(\vartheta_1^0,\vartheta_2)\right]Y_{k-1}^{(h)}\Delta Y_k^{(h)\mathsf{T}}\left[\overline{k}(B,\vartheta_1,\vartheta_2)-\overline{k}(B,\vartheta_1^0,\vartheta_2)\right]^{\mathsf{T}}\\
&\quad+\left[\overline{k}(B,\vartheta_1,\vartheta_2)-\overline{k}(B,\vartheta_1^0,\vartheta_2)\right]\Delta Y_k^{(h)}\cdot\Delta Y_k^{(h)\mathsf{T}}\left[\overline{k}(B,\vartheta_1,\vartheta_2)-\overline{k}(B,\vartheta_1^0,\vartheta_2)\right]^{\mathsf{T}}\Big)\Bigg)\\
&+2\cdot\frac{1}{n}\sum_{k=1}^{n}\operatorname{tr}\Bigg(\left(V_{\vartheta_1,\vartheta_2}^{(h)}\right)^{-1}\Big(-\left[\Pi(\vartheta_1,\vartheta_2)-\Pi(\vartheta_1^0,\vartheta_2)\right]Y_{k-1}^{(h)}\Delta Y_k^{(h)\mathsf{T}}\overline{k}(B,\vartheta_1^0,\vartheta_2)^{\mathsf{T}}\\
&+\left[\overline{k}(B,\vartheta_1,\vartheta_2)-\overline{k}(B,\vartheta_1^0,\vartheta_2)\right]\Delta Y_k^{(h)}\Delta Y_k^{(h)\mathsf{T}}\overline{k}(B,\vartheta_1^0,\vartheta_2)^{\mathsf{T}}\\
&+\left[\Pi(\vartheta_1,\vartheta_2)-\Pi(\vartheta_1^0,\vartheta_2)\right]Y_{k-1}^{(h)}Y_{k-1,2}^{(h)\mathsf{T}}\Pi(\vartheta_1^0,\vartheta_2)^{\mathsf{T}}\\
&-\left[\overline{k}(B,\vartheta_1,\vartheta_2)-\overline{k}(B,\vartheta_1^0,\vartheta_2)\right]\Delta Y_k^{(h)}Y_{k-1,2}^{(h)\mathsf{T}}\Pi(\vartheta_1^0,\vartheta_2)^{\mathsf{T}}\Big)\Bigg)\end{aligned}$$

$$+\frac{1}{n}\sum_{k=1}^{n}\operatorname{tr}\Bigg(\Big[(V^{(h)}_{\vartheta_1,\vartheta_2})^{-1}-(V^{(h)}_{\vartheta_1^0,\vartheta_2})^{-1}\Big]\cdot\Big[\overline{k}(B,\vartheta_1^0,\vartheta_2)\Delta Y_k^{(h)}\Delta Y_k^{(h)\mathsf{T}}\overline{k}(B,\vartheta_1^0,\vartheta_2)^{\mathsf{T}}$$
$$+\Pi(\vartheta_1^0,\vartheta_2)Y^{(h)}_{k-1,2}Y^{(h)\mathsf{T}}_{k-1,2}\Pi(\vartheta_1^0,\vartheta_2)^{\mathsf{T}}-2\cdot\overline{k}(B,\vartheta_1^0,\vartheta_2)\Delta Y_k^{(h)}Y^{(h)\mathsf{T}}_{k-1,2}\Pi(\vartheta_1^0,\vartheta_2)^{\mathsf{T}}\Big]\Bigg) \quad (5.87)$$

and $\mathcal{Q}_n^{(h)}(\vartheta_2)$ as

$$\mathcal{Q}^{(h)}_{n,2}(\vartheta_2)=\operatorname{tr}\Bigg(\frac{1}{n}\sum_{k=1}^{n}(V^{(h)}_{\vartheta_1^0,\vartheta_2})^{-1}\Big[\overline{k}(B,\vartheta_1^0,\vartheta_2)\Delta Y_k^{(h)}\Delta Y_k^{(h)\mathsf{T}}\overline{k}(B,\vartheta_1^0,\vartheta_2)^{\mathsf{T}}$$
$$+\Pi(\vartheta_1^0,\vartheta_2)Y^{(h)}_{k-1,2}Y^{(h)\mathsf{T}}_{k-1,2}\Pi(\vartheta_1^0,\vartheta_2)^{\mathsf{T}}$$
$$-2\cdot\overline{k}(B,\vartheta_1^0,\vartheta_2)\Delta Y_k^{(h)}Y^{(h)\mathsf{T}}_{k-1,2}\Pi(\vartheta_1^0,\vartheta_2)^{\mathsf{T}}\Big]\Bigg). \quad (5.88)$$

In this chapter we are going to use an extension of Davydov's inequality, which was shown by Schlemm and Stelzer [91]. We briefly recall this result.

Lemma 5.9.6 (Lemma 2.13, Schlemm and Stelzer [91])
Let X be a strictly stationary, strongly mixing d-dimensional stochastic process with finite $(4+\delta)$-th moments for some $\delta>0$. Then there exists a constant κ, such that for all $d\times d$ matrices A,B, every $n\in\mathbb{Z}$, $l\in\mathbb{N}$, and time indices $\nu,\nu'\in\mathbb{N}_0$, $\mu,\mu'=0,1,\ldots,\lfloor\frac{l}{2}\rfloor$, it holds that

$$Cov(X^{\mathsf{T}}_{n-\nu}AX_{n-\nu'}\ ,\ X^{\mathsf{T}}_{n+l-\mu}BX_{n+l-\mu'})\leq\kappa\|A\|\|B\|\left[\alpha_X\left(\left\lfloor\frac{l}{2}\right\rfloor\right)\right]^{\frac{\delta}{\delta+2}}, \quad (5.89)$$

where α_X denote the strong mixing coefficients of the process X.

5.9.3. Parts of the Proof of Lemma 5.6.5

We give here intermediate steps of the proof Lemma 5.6.5 in order to make the upper bound clearer which guarantees the necessary absolute summability.

Instead of considering every term in (5.65), which occurs by using the representation (5.53b), we consider exemplarily only

$$\left|\operatorname{Cov}\left(\partial_i^{st}\varepsilon^{(h)}_{k,2}(\vartheta_2)\big(V^{(h)}_{\vartheta}\big)^{-1}\varepsilon^{(h)}_{k,2}(\vartheta_2)^{\mathsf{T}}\ ,\ \partial_j^{st}\varepsilon^{(h)}_{k+l,2}(\vartheta_2)\big(V^{(h)}_{\vartheta}\big)^{-1}\varepsilon^{(h)}_{k+l,2}(\vartheta_2)^{\mathsf{T}}\right)\right|.$$

The covariance matrix can be bounded from above using the representation given in Lemma 5.2.2 by

$$
\begin{aligned}
&\left|\mathrm{Cov}\left(\partial_i^{st}\varepsilon_{k,2}^{(h)}(\vartheta_2)\big(V_\vartheta^{(h)}\big)^{-1}\varepsilon_{k,2}^{(h)}(\vartheta_2)^\mathsf{T}\,,\,\partial_j^{st}\varepsilon_{k+l,2}^{(h)}(\vartheta_2)\big(V_\vartheta^{(h)}\big)^{-1}\varepsilon_{k+l,2}^{(h)}(\vartheta_2)^\mathsf{T}\right)\right|\\
&\leq 4\cdot c_1\cdot\Bigg(\left|\mathrm{Cov}\left(Y_{k-1,2}^{(h)}Y_{k-1,2}^{(h)\mathsf{T}}\,,\,Y_{k-1+l,2}^{(h)}Y_{k-1+l,2}^{(h)\mathsf{T}}\right)\right|\\
&\quad+\sum_{\iota_4=0}^{\infty}\left|\mathrm{Cov}\left(Y_{k-1,2}^{(h)}Y_{k-1,2}^{(h)\mathsf{T}}\,,\,Y_{k-1+l,2}^{(h)}\Delta Y_{k+l-\iota_4}^{(h)\mathsf{T}}\widetilde{K}_{\iota_4}(\vartheta_2)^\mathsf{T}\right)\right|\\
&\quad+\sum_{\iota_3=0}^{\infty}\left|\mathrm{Cov}\left(Y_{k-1,2}^{(h)}Y_{k-1,2}^{(h)\mathsf{T}}\,,\,\partial_j^{st}\widetilde{K}_{\iota_3}(\vartheta_2)\Delta Y_{k+l-\iota_3}^{(h)}Y_{k-1+l,2}^{(h)\mathsf{T}}\right)\right|\\
&\quad+\sum_{\iota_3,\iota_4=0}^{\infty}\left|\mathrm{Cov}\left(Y_{k-1,2}^{(h)}Y_{k-1,2}^{(h)\mathsf{T}}\,,\,\partial_j^{st}\widetilde{K}_{\iota_3}(\vartheta_2)\Delta Y_{k+l-\iota_3}^{(h)}\Delta Y_{k+l-\iota_4}^{(h)\mathsf{T}}\widetilde{K}_{\iota_4}(\vartheta_2)^\mathsf{T}\right)\right|\\
&\quad+\sum_{\iota_2=0}^{\infty}\left|\mathrm{Cov}\left(Y_{k-1,2}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta_2)^\mathsf{T}\,,\,Y_{k+l-1,2}^{(h)}Y_{k+l-1,2}^{(h)\mathsf{T}}\right)\right|\\
&\quad+\sum_{\iota_2,\iota_4=0}^{\infty}\left|\mathrm{Cov}\left(Y_{k-1,2}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta_2)^\mathsf{T}\,,\,Y_{k+l-1,2}^{(h)}\Delta Y_{k+l-\iota_4}^{(h)\mathsf{T}}\widetilde{K}_{\iota_4}(\vartheta_2)^\mathsf{T}\right)\right|\\
&\quad+\sum_{\iota_2,\iota_3=0}^{\infty}\left|\mathrm{Cov}\left(Y_{k-1,2}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta_2)^\mathsf{T}\,,\,\partial_j^{st}\widetilde{K}_{\iota_3}(\vartheta_2)\Delta Y_{k+l-\iota_3}^{(h)}Y_{k+l-1,2}^{(h)\mathsf{T}}\right)\right|\\
&\quad+\sum_{\iota_2,\iota_3,\iota_4=0}^{\infty}\left|\mathrm{Cov}\left(Y_{k-1,2}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta_2)^\mathsf{T}\,,\,\partial_j^{st}\widetilde{K}_{\iota_3}(\vartheta_2)\Delta Y_{k+l-\iota_3}^{(h)}\Delta Y_{k+l-\iota_4}^{(h)\mathsf{T}}\widetilde{K}_{\iota_4}(\vartheta_2)^\mathsf{T}\right)\right|\\
&\quad+\sum_{\iota_1=0}^{\infty}\left|\mathrm{Cov}\left(\partial_i^{st}\widetilde{K}_{\iota_1}(\vartheta_2)\Delta Y_{k-\iota_1}^{(h)}Y_{k-1,2}^{(h)\mathsf{T}}\,,\,Y_{k-1+l,2}^{(h)}Y_{k-1+l,2}^{(h)\mathsf{T}}\right)\right|\\
&\quad+\sum_{\iota_1,\iota_4=0}^{\infty}\left|\mathrm{Cov}\left(\partial_i^{st}\widetilde{K}_{\iota_1}(\vartheta_2)\Delta Y_{k-\iota_1}^{(h)}Y_{k-1,2}^{(h)\mathsf{T}}\,,\,Y_{k-1+l,2}^{(h)}\Delta Y_{k+l-\iota_4}^{(h)\mathsf{T}}\widetilde{K}_{\iota_4}(\vartheta_2)^\mathsf{T}\right)\right|\\
&\quad+\sum_{\iota_1,\iota_3=0}^{\infty}\left|\mathrm{Cov}\left(\partial_i^{st}\widetilde{K}_{\iota_1}(\vartheta_2)\Delta Y_{k-\iota_1}^{(h)}Y_{k-1,2}^{(h)\mathsf{T}}\,,\,\partial_j^{st}\widetilde{K}_{\iota_3}(\vartheta_2)\Delta Y_{k+l-\iota_3}^{(h)}Y_{k-1+l,2}^{(h)\mathsf{T}}\right)\right|\\
&\quad+\sum_{\iota_1,\iota_3,\iota_4=0}^{\infty}\left|\mathrm{Cov}\left(\partial_i^{st}\widetilde{K}_{\iota_1}(\vartheta_2)\Delta Y_{k-\iota_1}^{(h)}Y_{k-1,2}^{(h)\mathsf{T}}\,,\,\partial_j^{st}\widetilde{K}_{\iota_3}(\vartheta_2)\Delta Y_{k+l-\iota_3}^{(h)}\Delta Y_{k+l-\iota_4}^{(h)\mathsf{T}}\widetilde{K}_{\iota_4}(\vartheta_2)^\mathsf{T}\right)\right|\\
&\quad+\sum_{\iota_1,\iota_2=0}^{\infty}\left|\mathrm{Cov}\left(\partial_i^{st}\widetilde{K}_{\iota_1}(\vartheta_2)\Delta Y_{k-\iota_1}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta_2)^\mathsf{T}\,,\,Y_{k-1+l,2}^{(h)}Y_{k-1+l,2}^{(h)\mathsf{T}}\right)\right|\\
&\quad+\sum_{\iota_1,\iota_2,\iota_4=0}^{\infty}\left|\mathrm{Cov}\left(\partial_i^{st}\widetilde{K}_{\iota_1}(\vartheta_2)\Delta Y_{k-\iota_1}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta_2)^\mathsf{T}\,,\,Y_{k-1+l,2}^{(h)}\Delta Y_{k+l-\iota_4}^{(h)\mathsf{T}}\widetilde{K}_{\iota_4}(\vartheta_2)^\mathsf{T}\right)\right|\\
&\quad+\sum_{\iota_1,\iota_2,\iota_3=0}^{\infty}\left|\mathrm{Cov}\left(\partial_i^{st}\widetilde{K}_{\iota_1}(\vartheta_2)\Delta Y_{k-\iota_1}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta_2)^\mathsf{T}\,,\,\partial_j^{st}\widetilde{K}_{\iota_3}(\vartheta_2)\Delta Y_{k+l-\iota_3}^{(h)}Y_{k-1+l,2}^{(h)\mathsf{T}}\right)\right|\\
&\quad+\sum_{\iota_1,\iota_2,\iota_3,\iota_4=0}^{\infty}\Bigg|\mathrm{Cov}\Bigg(\partial_i^{st}\widetilde{K}_{\iota_1}(\vartheta_2)\Delta Y_{k-\iota_1}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta_2)^\mathsf{T}\,,\\
&\qquad\qquad\partial_j^{st}\widetilde{K}_{\iota_3}(\vartheta_2)\Delta Y_{k+l-\iota_3}^{(h)}\Delta Y_{k+l-\iota_4}^{(h)\mathsf{T}}\widetilde{K}_{\iota_4}(\vartheta_2)^\mathsf{T}\Bigg)\Bigg|\Bigg).
\end{aligned}\tag{$\bigstar$}
$$

We consider one example for each class of the different combinations appearing in the equation above. First, we have the covariance matrix, where only the stationary part appears. In this case we have with Lemma 2.13 in Schlemm and Stelzer [91] and Lemma 4.3.18 that

$$\left|\operatorname{Cov}\left(Y_{k-1,2}^{(h)}Y_{k-1,2}^{(h)\mathsf{T}},\, Y_{k-1+l,2}^{(h)}Y_{k-1+l,2}^{(h)\mathsf{T}}\right)\right| \leq C\cdot\left[\alpha_{Y_2^{(h)}}(l)\right]^{\frac{\delta}{2+\delta}}. \tag{$\clubsuit$}$$

Next, we consider the summand in the equation ($\bigstar$) above, where only the process $\Delta Y^{(h)}$ appears. It holds that

$$\begin{aligned}
&\sum_{\iota_1,\iota_2,\iota_3,\iota_4=0}^{\infty}\Big|\operatorname{Cov}\Big(\partial_i^{st}\widetilde{K}_{\iota_1}(\vartheta_2)\Delta Y_{k-\iota_1}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta_2)^{\mathsf{T}},\\
&\qquad\qquad \partial_j^{st}\widetilde{K}_{\iota_3}(\vartheta_2)\Delta Y_{k+l-\iota_3}^{(h)}\Delta Y_{k+l-\iota_4}^{(h)\mathsf{T}}\widetilde{K}_{\iota_4}(\vartheta_2)^{\mathsf{T}}\Big)\Big|\\
&=\sum_{\substack{\iota_1,\iota_2,\iota_3,\iota_4=0\\ \max\{\iota_1,\iota_2,\iota_3,\iota_4\}>\frac{l}{2}}}^{\infty}\Big|\operatorname{Cov}\Big(\partial_i^{st}\widetilde{K}_{\iota_1}(\vartheta_2)\Delta Y_{k-\iota_1}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta_2)^{\mathsf{T}},\\
&\qquad\qquad \partial_j^{st}\widetilde{K}_{\iota_3}(\vartheta_2)\Delta Y_{k+l-\iota_3}^{(h)}\Delta Y_{k+l-\iota_4}^{(h)\mathsf{T}}\widetilde{K}_{\iota_4}(\vartheta_2)^{\mathsf{T}}\Big)\Big|\\
&\quad+\sum_{\iota_1,\iota_2,\iota_3,\iota_4=0}^{\lfloor\frac{l}{2}\rfloor}\Big|\operatorname{Cov}\Big(\partial_i^{st}\widetilde{K}_{\iota_1}(\vartheta_2)\Delta Y_{k-\iota_1}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta_2)^{\mathsf{T}},\\
&\qquad\qquad \partial_j^{st}\widetilde{K}_{\iota_3}(\vartheta_2)\Delta Y_{k+l-\iota_3}^{(h)}\Delta Y_{k+l-\iota_4}^{(h)\mathsf{T}}\widetilde{K}_{\iota_4}(\vartheta_2)^{\mathsf{T}}\Big)\Big|\\
&=:S^+ + S^-.
\end{aligned}$$

We can apply the Cauchy-Schwarz inequality to S^+. Further, we use the finite fourth moment of $\Delta Y^{(h)}$ and the exponential decay of the coefficients to obtain that there exists constants $C>0$ and $\rho<1$ such that

$$\begin{aligned}
S^+ \leq &\sum_{\substack{\iota_1,\iota_2,\iota_3,\iota_4=0\\ \max\{\iota_1,\iota_2,\iota_3,\iota_4\}>\frac{l}{2}}}^{\infty}\left\|\partial_i^{st}\widetilde{K}_{\iota_1}(\vartheta_2)\right\|\left\|\widetilde{K}_{\iota_2}(\vartheta_2)\right\|\left\|\partial_j^{st}\widetilde{K}_{\iota_3}(\vartheta_2)\right\|\left\|\widetilde{K}_{\iota_4}(\vartheta_2)\right\|\mathbb{E}\left\|\Delta Y_k^{(h)}\right\|^4\\
&\leq C\cdot\rho^{\frac{l}{2}}.
\end{aligned}$$

Consequently S^+ is absolutely summable.

Next, we show that the same holds for S^-. In order to show this there are some tedious calculations necessary. However, we see immediately that we can write $\Delta Y_k^{(h)} = \Delta L(kh) + \Delta Y_{k,2}^{(h)}$. We obtain by using this decomposition on the covariance in S^- that

$$\begin{aligned}
&\left|\operatorname{Cov}\left(\partial_i^{st}\widetilde{K}_{\iota_1}(\vartheta_2)\Delta Y_{k-\iota_1}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta_2)^{\mathsf{T}}, \partial_j^{st}\widetilde{K}_{\iota_3}(\vartheta_2)\Delta Y_{k+l-\iota_3}^{(h)}\Delta Y_{k+l-\iota_4}^{(h)\mathsf{T}}\widetilde{K}_{\iota_4}(\vartheta_2)^{\mathsf{T}}\right)\right| \\
&\leq C\cdot\left|\operatorname{Cov}\left(\widetilde{K}_{\iota_1}(\vartheta_2)\Delta Y_{k-\iota_1}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta_2)^{\mathsf{T}},\right.\right. \\
&\qquad\left.\left.\partial_j^{st}\widetilde{K}_{\iota_3}(\vartheta_2)\Delta L((k+l-\iota_3)h)\Delta L((k+l-\iota_4)h)^{\mathsf{T}}\widetilde{K}_{\iota_4}(\vartheta_2)^{\mathsf{T}}\right)\right| \\
&\quad+\left|\operatorname{Cov}\left(\partial_i^{st}\widetilde{K}_{\iota_1}(\vartheta_2)\Delta Y_{k-\iota_1}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta_2)^{\mathsf{T}},\right.\right. \\
&\qquad\left.\left.\partial_j^{st}\widetilde{K}_{\iota_3}(\vartheta_2)\Delta L((k+l-\iota_3)h)\Delta Y_{k+l-\iota_4,2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_4}(\vartheta_2)^{\mathsf{T}}\right)\right| \\
&\quad+\left|\operatorname{Cov}\left(\partial_i^{st}\widetilde{K}_{\iota_1}(\vartheta_2)\Delta Y_{k-\iota_1}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta_2)^{\mathsf{T}},\right.\right. \\
&\qquad\left.\left.\partial_j^{st}\widetilde{K}_{\iota_3}(\vartheta_2)\Delta Y_{k+l-\iota_3,2}^{(h)}\Delta L((k+l-\iota_4)h)^{\mathsf{T}}\widetilde{K}_{\iota_4}(\vartheta_2)^{\mathsf{T}}\right)\right| \\
&\quad+\left|\operatorname{Cov}\left(\partial_i^{st}\widetilde{K}_{\iota_1}(\vartheta_2)\Delta Y_{k-\iota_1}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta_2)^{\mathsf{T}},\right.\right. \\
&\qquad\left.\left.\partial_j^{st}\widetilde{K}_{\iota_3}(\vartheta_2)\Delta Y_{k+l-\iota_3,2}^{(h)}\Delta Y_{k+l-\iota_4,2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_4}(\vartheta_2)^{\mathsf{T}}\right)\right|.
\end{aligned}\tag{5.90}$$

The sum in S^- goes from $\iota_3,\iota_4=0$ to $\lfloor\frac{l}{2}\rfloor$, which implies that $k+l-\iota_3>k$ as well as $k+l-\iota_4>k$. Thus, the first summand of (5.90) is equal to zero since $\Delta Y_{k-\iota_1}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}$ is independent of $\Delta L((k+l-\iota_3)h)\Delta L((k+l-\iota_4)h)^{\mathsf{T}}$ for all $\iota_1,\iota_2,\iota_3,\iota_4\in\left\{0,\ldots,\lfloor\frac{l}{2}\rfloor\right\}$. The second and third summand of (5.90) can be dealt with in a similar fashion. Hence, we only consider the second term. Therefore, we obtain for the second one

$$\begin{aligned}
&\left|\operatorname{Cov}\left(\partial_i^{st}\widetilde{K}_{\iota_1}(\vartheta_2)\Delta Y_{k-\iota_1}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta_2)^{\mathsf{T}},\right.\right. \\
&\qquad\left.\left.\partial_j^{st}\widetilde{K}_{\iota_3}(\vartheta_2)\Delta L((k+l-\iota_3)h)\Delta Y_{k+l-\iota_4,2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_4}(\vartheta_2)^{\mathsf{T}}\right)\right| \qquad (\diamond)\\
&\leq\left|\operatorname{Cov}\left(\partial_i^{st}\widetilde{K}_{\iota_1}(\vartheta_2)\Delta Y_{k-\iota_1}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta_2)^{\mathsf{T}},\right.\right. \\
&\qquad\left.\left.\partial_j^{st}\widetilde{K}_{\iota_3}(\vartheta_2)\Delta L((k+l-\iota_3)h)Y_{k+l-\iota_4,2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_4}(\vartheta_2)^{\mathsf{T}}\right)\right| \\
&\quad+\left|\operatorname{Cov}\left(\partial_i^{st}\widetilde{K}_{\iota_1}(\vartheta_2)\Delta Y_{k-\iota_1}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta_2)^{\mathsf{T}},\right.\right. \\
&\qquad\left.\left.\partial_j^{st}\widetilde{K}_{\iota_3}(\vartheta_2)\Delta L((k+l-\iota_3)h)Y_{k+l-\iota_4-1,2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_4}(\vartheta_2)^{\mathsf{T}}\right)\right| \\
&=0.
\end{aligned}$$

To see this we only consider the first one, because once more the two summands can be dealt with analogously. We know for $\iota_3<\iota_4$ that due to the independence the covariance is in this case equal to zero. For $0\leq\iota_4<\iota_3\leq\lfloor\frac{l}{2}\rfloor$ we have

$$\begin{aligned}
&\left|\operatorname{Cov}\left(\partial_i^{st}\widetilde{K}_{\iota_1}(\vartheta_2)\Delta Y_{k-\iota_1}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_2}(\vartheta_2)^{\mathsf{T}},\right.\right. \\
&\qquad\left.\left.\partial_j^{st}\widetilde{K}_{\iota_3}(\vartheta_2)\Delta L((k+l-\iota_3)h)Y_{k+l-\iota_4,2}^{(h)\mathsf{T}}\widetilde{K}_{\iota_4}(\vartheta_2)^{\mathsf{T}}\right)\right|
\end{aligned}$$

$$
\begin{aligned}
&\leq \Bigg| \operatorname{Cov}\Bigg(\partial_i^{st} \widetilde{K}_{\iota_1}(\vartheta_2) \Delta Y_{k-\iota_1}^{(h)} \Delta Y_{k-\iota_2}^{(h)\mathsf{T}} \widetilde{K}_{\iota_2}(\vartheta_2)^{\mathsf{T}}, \\
&\qquad \partial_j^{st} \widetilde{K}_{\iota_3}(\vartheta_2) \Delta L((k+l-\iota_3)h) \Big(C_2 \int_{M_1} \mathrm{e}^{A_2((k+l-\iota_4)h-u)} B_2 \, \mathrm{d}L(u) \Big)^{\mathsf{T}} \widetilde{K}_{\iota_4}(\vartheta_2)^{\mathsf{T}} \Bigg) \Bigg| \\
&\quad + \Bigg| \operatorname{Cov}\Bigg(\partial_i^{st} \widetilde{K}_{\iota_1}(\vartheta_2) \Delta Y_{k-\iota_1}^{(h)} \Delta Y_{k-\iota_2}^{(h)\mathsf{T}} \widetilde{K}_{\iota_2}(\vartheta_2)^{\mathsf{T}}, \\
&\qquad \partial_j^{st} \widetilde{K}_{\iota_3}(\vartheta_2) \Delta L((k+l-\iota_3)h) \Big(C_2 \int_{M_2} \mathrm{e}^{A_2((k+l-\iota_4)h-u)} B_2 \, \mathrm{d}L(u) \Big)^{\mathsf{T}} \widetilde{K}_{\iota_4}(\vartheta_2)^{\mathsf{T}} \Bigg) \Bigg| \\
&= 0, \qquad\qquad (\spadesuit)
\end{aligned}
$$

whereupon M_1 is the interval, where the processes of the second part in the covariance overlap, i.e. $M_1 := (-\infty, (k+l-\iota_4)h) \cap ((k+l-\iota_3-1)h, (k+l-\iota_3)h)$ and M_2 is the interval given by $M_2 := (-\infty, (k+l-\iota_4)h) \setminus M_1$. Hence, the second summand has to be zero, which follows once more by the independence of the $\Delta L((k+l-\iota_3)h)$ with the integral term. On the other side, the process $\Delta L((k+l-\iota_3)h)$ is not independent of the integral term over M_1, but the product on the right-hand side is independent of the product term on the left-hand side in the covariance. Thus, this covariance is also equal to zero.

Lastly, let us consider the fourth summand of (5.90). Analogously, it suffices to consider $Y_{k,2}^{(h)}$ instead of the $\Delta Y_{k,2}^{(h)}$ with the same arguments as in ($\diamond$). Thus, we exemplarily consider

$$
\begin{aligned}
&\Bigg| \operatorname{Cov}\Bigg(\partial_i^{st} \widetilde{K}_{\iota_1}(\vartheta_2) \big(Y_{k-\iota_1,2}^{(h)} + \Delta L(k-\iota_1) \big) \big(Y_{k-\iota_2,2}^{(h)} + \Delta L(k-\iota_2) \big)^{\mathsf{T}} \widetilde{K}_{\iota_2}(\vartheta_2)^{\mathsf{T}}, \\
&\qquad \partial_j^{st} \widetilde{K}_{\iota_3}(\vartheta_2) Y_{k+l-\iota_3,2}^{(h)} Y_{k+l-\iota_4,2}^{(h)\mathsf{T}} \widetilde{K}_{\iota_4}(\vartheta_2)^{\mathsf{T}} \Bigg) \Bigg| \\
&\leq \Bigg| \operatorname{Cov}\Bigg(\partial_i^{st} \widetilde{K}_{\iota_1}(\vartheta_2) Y_{k-\iota_1,2}^{(h)} Y_{k-\iota_2,2}^{(h)\mathsf{T}} \widetilde{K}_{\iota_2}(\vartheta_2)^{\mathsf{T}}, \\
&\qquad \partial_j^{st} \widetilde{K}_{\iota_3}(\vartheta_2) Y_{k+l-\iota_3,2}^{(h)} Y_{k+l-\iota_4,2}^{(h)\mathsf{T}} \widetilde{K}_{\iota_4}(\vartheta_2)^{\mathsf{T}} \Bigg) \Bigg| \\
&\quad + \Bigg| \operatorname{Cov}\Bigg(\partial_i^{st} \widetilde{K}_{\iota_1}(\vartheta_2) \Delta L\big((k-\iota_1)h\big) Y_{k-\iota_2,2}^{(h)\mathsf{T}} \widetilde{K}_{\iota_2}(\vartheta_2)^{\mathsf{T}}, \\
&\qquad \partial_j^{st} \widetilde{K}_{\iota_3}(\vartheta_2) Y_{k+l-\iota_3,2}^{(h)} Y_{k+l-\iota_4,2}^{(h)\mathsf{T}} \widetilde{K}_{\iota_4}(\vartheta_2)^{\mathsf{T}} \Bigg) \Bigg| \\
&\quad + \Bigg| \operatorname{Cov}\Bigg(\partial_i^{st} \widetilde{K}_{\iota_1}(\vartheta_2) Y_{k-\iota_1,2}^{(h)} \Delta L\big((k-\iota_2)h\big)^{\mathsf{T}} \widetilde{K}_{\iota_2}(\vartheta_2)^{\mathsf{T}}, \\
&\qquad \partial_j^{st} \widetilde{K}_{\iota_3}(\vartheta_2) Y_{k+l-\iota_3,2}^{(h)} Y_{k+l-\iota_4,2}^{(h)\mathsf{T}} \widetilde{K}_{\iota_4}(\vartheta_2)^{\mathsf{T}} \Bigg) \Bigg| \\
&\quad + \Bigg| \operatorname{Cov}\Bigg(\partial_i^{st} \widetilde{K}_{\iota_1}(\vartheta_2) \Delta L\big((k-\iota_1)h\big) \Delta L\big((k-\iota_2)h\big)^{\mathsf{T}} \widetilde{K}_{\iota_2}(\vartheta_2)^{\mathsf{T}}, \\
&\qquad \partial_j^{st} \widetilde{K}_{\iota_3}(\vartheta_2) Y_{k+l-\iota_3,2}^{(h)} Y_{k+l-\iota_4,2}^{(h)\mathsf{T}} \widetilde{K}_{\iota_4}(\vartheta_2)^{\mathsf{T}} \Bigg) \Bigg| \\
&\leq C \cdot \Big[\alpha_{\mathrm{Y}_2^{(h)}}(l) \Big]^{\frac{\delta}{2+\delta}} + C \cdot \rho^{\frac{l}{2}},
\end{aligned}
$$

for some constant $C > 0$, where we used similar arguments as in (♠). We justify the upper bound with the subsequent estimations.

The first summand is bounded from above by the mixing coefficients just as in (♣) due to Lemma 2.13 in Schlemm and Stelzer [91]. We have due to the definition of S^- that $\iota_3, \iota_4 < \frac{l}{2}$ holds and thus we can apply this lemma. As a consequence, there are three terms left. The second and third one are yet again be dealt with in the same manner. Thus, we have, with the same methods as in (♠) and the fact that $\iota_3, \iota_4 < \frac{l}{2}$, for the second term

$$
\begin{aligned}
&\Big| \operatorname{Cov}\Big(\partial_i^{st} \widetilde{K}_{\iota_1}(\vartheta_2) \Delta L\big((k-\iota_1)h\big) Y_{k-\iota_2,2}^{(h)\mathsf{T}} \widetilde{K}_{\iota_2}(\vartheta_2)^{\mathsf{T}} \;, \\
&\qquad\qquad \partial_j^{st} \widetilde{K}_{\iota_3}(\vartheta_2) Y_{k+l-\iota_3,2}^{(h)} Y_{k+l-\iota_4,2}^{(h)\mathsf{T}} \widetilde{K}_{\iota_4}(\vartheta_2)^{\mathsf{T}}\Big) \Big| \\
&\leq \Big| \operatorname{Cov}\Big(\partial_i^{st} \widetilde{K}_{\iota_1}(\vartheta_2) \Delta L\big((k-\iota_1)h\big) \Big(C_2 \int_N \mathrm{e}^{A_2((k-\iota_2)h-u)} B_2 \,\mathrm{d}L(u)\Big)^{\mathsf{T}} \widetilde{K}_{\iota_2}(\vartheta_2)^{\mathsf{T}}, \\
&\qquad \partial_j^{st} \widetilde{K}_{\iota_3}(\vartheta_2) \Big(C_2 \int_N \mathrm{e}^{A_2((k+l-\iota_3)h-u)} B_2 \,\mathrm{d}L(u)\Big) \\
&\qquad\quad \cdot \Big(C_2 \int_N \mathrm{e}^{A_2((k+l-\iota_4)h-u)} B_2 \,\mathrm{d}L(u)\Big) \widetilde{K}_{\iota_4}(\vartheta_2)^{\mathsf{T}}\Big) \Big|, \\
&\leq C \cdot \left(\mathbb{E}\left\| \Delta L\big((k-\iota_1)h\big) \Big(C_2 \mathrm{e}^{A_2(\iota_1-\iota_2)h} \int_N \mathrm{e}^{A_2((k-\iota_1)h-u)} B_2 \,\mathrm{d}L(u)\Big)^{\mathsf{T}} \right\|^2\right)^{\frac{1}{2}} \\
&\quad \cdot \left(\mathbb{E}\left\| \Big(C_2 \mathrm{e}^{A_2 \frac{l}{2} h} \mathrm{e}^{A_2(\frac{l}{2}-\iota_3)h} \int_N \mathrm{e}^{A_2(kh-u)} B_2 \,\mathrm{d}L(u)\Big) \right.\right. \\
&\qquad \left.\left. \cdot \Big(C_2 \mathrm{e}^{A_2 \frac{l}{2} h} \mathrm{e}^{A_2(\frac{l}{2}-\iota_4)h} \int_N \mathrm{e}^{A_2(kh-u)} B_2 \,\mathrm{d}L(u)\Big)^{\mathsf{T}} \right\|^2\right)^{\frac{1}{2}} \\
&\leq C \cdot \rho^{\frac{l}{2}},
\end{aligned}
$$

for some constants $c, C > 0$, where the set N is given by the intersection of the following sets

$$
\begin{aligned}
N := &\big((k-\iota_1-1)h, (k-\iota_1)h\big) \cap \big(-\infty, (k-\iota_2)h\big) \\
&\cap \big(-\infty, (k+l-\iota_3)h\big) \cap \big(-\infty, (k+l-\iota_4)h\big).
\end{aligned}
$$

Note that the set N is either the empty set if $\iota_2 - 1 > \iota_1$ or otherwise $N = \big((k-\iota_1-1)h, (k-\iota_1)h\big)$. The remaining parts of the integral can be ignored as before due to the independence. Hence, we have also shown for these two the absolute summability.

The last summand follows in the same way. The difference is the intersection, which is in this case given by

$$N' := \big((k-\iota_1-1)h, (k-\iota_1)h\big) \cap \big((k-\iota_2-1)h, (k-\iota_2)h\big) \\ \cap \big(-\infty, (k+l-\iota_3)h\big) \cap \big(-\infty, (k+l-\iota_4)h\big).$$

In conclusion, we have shown that $S^- \leq C \cdot \left[\alpha_{Y_2^{(h)}}(l)\right]^{\frac{\delta}{2+\delta}} + C \cdot \mathrm{e}^{-c \cdot lh}$ and thus absolutely summable.

If we proceed with all remaining terms of ($\bigstar$) and then for all remaining terms in (5.65) analogously, we are able to show that the original covariance matrix is absolute summable.

CHAPTER 6

Simulation Study

6.1. Introduction

The practical applicability of the quasi-maximum likelihood estimation procedure for cointegrated MCARMA processes in Chapter 5 is considered in this chapter. We test the theoretical results shown for the step-wise estimation procedure in simulation studies. The consistency of the long-run and short-run estimators, which we have shown in Section 5.5, suggest that the estimated results should be quite close in the mean to the true value for a large sample.

However, in order to apply the estimation method, we need first a parametrization of the model matrices satisfying the assumptions of Chapter 5. For this purpose, we present a suitable parametrization of the matrix C_1, i.e. the matrix corresponding to the long-run behavior of the time series. We describe an algorithm in Section 6.2 how to construct the matrix C_1 from a given parameter vector. This algorithm is based on the ideas of Bauer and Wagner [7], who parameterized a complex valued matrix satisfying similar constraints. The parametrization of the stationary part is standard as we use the echelon canonical form. The echelon canonical form is widely used in the VARMA context, see e.g. Lütkepohl and Poskitt [64] and the textbooks of Lütkepohl [62], or Hannan and Deistler [46]. In the context of linear state space models canonical representations can also be found, see for example in Guidorzi [44]. Another approach is to combine the error correction form with the echelon form, which was done in Lütkepohl and Claasen [63] for cointegrated VARMA models.

However, this approach has the disadvantage that it relies on the finite order form of the error correction form, whereas we have an infinite order term in the error correction form.

In Section 6.3 we present the results of the simulation studies. We consider two different cases, which differ in the dimension of the model. On the one hand we simulate a two-dimensional cointegrated MCARMA process and on the other hand a three-dimensional model. As a driving Lévy process we use a normal-inverse Gaussian process and a Brownian motion respectively. The two dimensional model is chosen in such a way that the stationary part corresponds one-to-one with the simulation study in Schlemm and Stelzer [91] for a stationary bivariate MCARMA process. We added one common stochastic trend to this model in order to make it cointegrated. The three-dimensional model is considered in order to have more complexity in the non-stationary part, where we added two common stochastic trends.

6.2. Parametrization

We give now an example for the parametrization of $C_{1,\vartheta_1} \in M_{d,c}(\mathbb{R})$, which is based on the parameterizations presented in Bauer and Wagner [7]. They presented a complex version of this parametrization. There is a major difference between the complex version and the real-valued version. This is due to the fact that the orthogonal constraints on the matrix have a greater effect in the real-valued case. If we consider a two dimensional subspace and choose a unique basis, it is enough to know the first vector. The second vector is then uniquely determined by the first one. This is not the case in the complex-valued case because $\mathbb{C}$ is isomorphic to $\mathbb{R}^2$. Hence, this gives more "degrees of freedom" for being orthogonal to a vector. Consequently, in the real-valued case more vectors are predetermined due to the orthogonality constraints. One can think of the parametrization as choosing a special, unique orthonormal basis of the space spanned by C_{1,ϑ_1}.

The parametrization of C_{1,ϑ_1} should satisfy Assumption M4, Assumption M6 and Assumption M8. Furthermore, C_{1,ϑ_1} must be of the form given in Theorem 4.3.7. Hence, we want the matrix C_{1,ϑ_1} to be a positive lower triangular matrix (c.f. Definition 6.2.1) with rank c satisfying $C_{1,\vartheta_1}^\mathsf{T} C_{1,\vartheta_1} = I_c$. We assume to know the indices $1 \leq j_1 < j_2 < \ldots < j_c \leq d$, which denotes the first positive entry in each column of C_{1,ϑ_1}. Furthermore, we denote the j^{th}-column of C_{1,ϑ_1} with c_j such that $C_{1,\vartheta_1} = [c_1, c_2, \ldots, c_c]$ because we parameterize the columns step by step.

Let us now formally define what we understand by a positive lower triangular matrix.

Definition 6.2.1

A matrix $C_1 = [c_{i,j}]_{i=1,\dots d, j=1,\dots,c} \in M_{d,c}(\mathbb{R})$ *is* ***positive lower triangular*** *if there exist indices* $1 \le j_1 < j_2 < \ldots < j_c \le d$*, such that* $c_{i,j} = 0, j_i < j$*,* $0 < c_{j_i,i}$*, i.e.* C_1 *is of the form*

$$\begin{pmatrix}
0 & 0 & 0 & \cdots & 0 \\
\vdots & & & & \vdots \\
0 & & & & 0 \\
c_{j_1,1} & \vdots & & \cdots & \\
* & & & & \vdots \\
 & & \vdots & & \\
 & 0 & & \cdots & \\
\vdots & c_{j_2,2} & & & \vdots \\
 & * & & & \\
 & & & & \\
 & \vdots & 0 & \cdots & \vdots \\
\vdots & & c_{j_3,3} & & \\
 & & * & & 0 \\
 & \vdots & & \cdots & c_{j_c,c} \\
\vdots & & \vdots & & * \\
 & & & & \vdots \\
* & * & * & \cdots & *
\end{pmatrix}, \tag{6.1}$$

where $*$ *denote arbitrary entries.*

We denote in the following with $f(\cdot)$ a function which maps a vector onto the unit-sphere, i.e. $f : \mathbb{R}^n \to S^n \subset \mathbb{R}^{n+1}$. Consequently, $f(x) \in \mathbb{R}^{n+1}$ is a real unit norm vector parameterized by the vector $x \in \mathbb{R}^n$. Furthermore, we require an additional constraint on f, namely that $e_1^\top f(x) > 0$ holds, where e_1 is the first unit vector of the appropriate dimension. One could use any mapping which satisfies the first part. One example satisfying these assumptions is the inverse stereographic projection. However, one has to be careful that the additional constraint is satisfied. For stereographic projections this can be done quite easily by restricting the parameter space.

We use in the following the inverse stereographic projection. Let us recall briefly the inverse stereographic projection and some of its properties.

Remark 6.2.2
The inverse stereographic projection $f : \mathbb{R}^n \to S^n \subset \mathbb{R}^{n+1}$, $\mathbb{R}^n \ni x \mapsto f(x) \in S^n$ is given by

$$f(x) = \frac{1}{\|x\|^2 + 1} \begin{pmatrix} \|x\|^2 - 1 \\ 2x \end{pmatrix}. \tag{6.2}$$

This function has a positive first entry for all $x \in \mathbb{R}^n$ satisfying $\|x\| > 1$. The inverse stereographic projection is a smooth and bijective function with Jacobian matrix

$$J_f(x) = \frac{2}{(\|x\|^2 + 1)^2} \begin{pmatrix} 2x^T \\ (\|x\|^2 - 1)\, I_n - 2xx^T \end{pmatrix}. \tag{6.3}$$

Although, we could use a general function f satisfying all restrictions we constrain ourselves hereafter on the inverse stereographic projection.

We need therefore at most

$$s_1 = \left(\sum_{i=1}^{c} d - j_i \right) - \binom{c}{2} \tag{6.4}$$

parameters for a given multi-index $\mathbf{j} = (j_1, \ldots, j_c)$, with $1 \leq j_1 < \ldots < j_c \leq d$. The subtraction of j_i is due to the normalization of each column to unity and the positive triangular form. The positiveness of the first entry in each column is needed since otherwise we would have two vectors satisfying the constraints. Thus we would not have uniqueness. Due to the orthogonality we can reduce the amount of parameters needed even further. Since all columns have to be pairwise orthogonal, the orthogonality implies $\binom{c}{2}$ additional equations, which is the reason for the last minus term.

Depending on the values of c, the dimension d and the multi-index $\mathbf{j}$ we might know a priori the last few columns due to the unique representation. These last columns consist in this case only of unit vectors, for example for the last column we have $c_c = e_d$. If $j_c = d$ holds, the column must be the unit vector to satisfy the unique representation. However, due to the orthogonality this implies that the last row must be equal to e_c^{T}. If we have also $j_{c-1} = d - 1$, this implies $c_{c-1} = e_{d-1}$. This is due to the fact that the last row of C_{1,ϑ_1} is equal to e_c^{T} and the orthogonality. We proceed in the same manner until the deviation of adjacent multi-indices is greater than one. Hence, we introduce the value $b_{d,c,\mathbf{j}}$, which denotes the number of the first column of C_{1,ϑ_1} from where on only unit vectors follow, i.e. $c_{b_{d,c,\mathbf{j}}} = e_{j_{b_{d,c,\mathbf{j}}}}, \ldots, c_c = e_d$. The value $b_{d,c,\mathbf{j}}$ is given by

$$b_{d,c,\mathbf{j}} := \inf\{1 \le i \le c : j_i - i = d - c\} \wedge c + 1, \tag{6.5}$$

where we set $\inf\{\emptyset\} = \infty$. We omit the dependence on the parameters and simply write b instead of $b_{d,c,\mathbf{j}}$ in the following.

As a consequence, all entries in the range j_b to d of the first $b-1$ columns of C_{1,ϑ_1} are equal to zero due to the orthogonality condition. For example, if $b \le c$ the matrix C_{1,ϑ_1} then has the following block form

$$C_{1,\vartheta_1} = \begin{pmatrix} \widetilde{C}_{1,\vartheta_1} & 0_{(d-c+b-1)\times(c-b+1)} \\ 0_{(d-c+b-1)\times(c-b+1)} & I_{(c-b+1)\times(c-b+1)} \end{pmatrix}, \tag{6.6}$$

where the $(d-c+b)\times(c-b)$ matrix block $\widetilde{C}_{1,\vartheta_1}$ satisfies the same constraints as the original matrix C_{1,ϑ_1}. Hence, the construction of the matrix C_{1,ϑ_1} reduces to constructing the upper left block $\widetilde{C}_{1,\vartheta_1}$. Note that we interpret 0_{-a}, for $a \ge 0$ as an empty vector and the same applies for matrices.

Before we start, let us introduce the auxiliary variable

$$n_i := (d - c + b - j_i) \vee 0 \tag{6.7}$$

for $1 \le i \le b-1$. The value of n_i can be interpreted as the number of components of the i^{th} column which depend somehow on the parameter vector. They are not as in the previous considerations inevitable equal to zero or one. We need at most $(n_i - i) \vee 0$ parameters for the i^{th} column. Quite often no additional parameters are needed to construct a column due to all the constraints (c.f. Table 6.1).

Before we present the parametrization we give one last definition. Since we want all matrices in the following to be of a special form, we introduce an abbreviation for this special form.

Definition 6.2.3
*We say an $m \times n$ matrix A, for $n \le m$, is in N_{plt}-**form** if it satisfies $A^T A = I_n$ and A is of positive lower triangular form.*

In the next lemma we present the algorithm, which constructs the matrix C_{1,ϑ_1} uniquely for a given multi-index $\mathbf{j}$. More importantly, we see that all the assumptions needed of the parametrization are indeed satisfied. Note that each column depends on the parameters of the previous parameterized columns due to the orthogonality.

Lemma 6.2.4

Assume that we have the multi-index $\boldsymbol{j}$ given as in Definition 6.2.1 with $1 \leq j_1 < j_2 < \ldots < j_c \leq d$. Furthermore, we assume that we have a real parameter vector $\vartheta_1 \in \mathbb{R}^{s_1}$ of dimension s_1, where s_1 is determined via (6.4). Then we can construct the matrix C_{1,ϑ_1} in the following manner.

First, we calculate $b_{d,c,\boldsymbol{j}}$ as defined in (6.5). Next, we know that the matrix has to have the form given in (6.6). Therefore, we can consider the reduced problem of constructing $\widetilde{C}_{1,\vartheta_1}$. In the following, we write without loss of generality for the columns of $\widetilde{C}_{1,\vartheta_1}$ as well $c_1, \ldots, c_{b-1}$.

The parameter vector is divided into

$$\vartheta_1 = [\vartheta_{1,1}^\mathsf{T}, \vartheta_{1,2}^\mathsf{T}, \ldots, \vartheta_{1,l_{d,c,\boldsymbol{j}}}^\mathsf{T}]^\mathsf{T},$$

where the number $l_{d,c,\boldsymbol{j}}$ of sub-vectors depends on the parameters d, c and $\boldsymbol{j}$. We do not specify the exact number here. However, the algorithm presented below will make clear how to divide the parameter vector in sub-vectors of appropriate size.

We denote with f_i always an inverse stereographic projection of a dimension depending on the size of the parameter sub-vector.

If $b = 1$ holds, the complete matrix is already given by $C_1 = \begin{pmatrix} 0_{(d-c)\times c} \\ I_{c\times c} \end{pmatrix}$. Hence, we assume from now on $b > 1$. It remains to determine the columns $c_1, \ldots, c_{b-1}$. We construct one column after the other by the following algorithm.

- *First Vector:*

  In this reduced block matrix, we have for the first vector no orthogonality constraint. The only constraint is the normalization and hence it is given right away by the inverse stereographic projection and the first index j_1. Insofar, we have the first column given by

$$c_1 = \begin{pmatrix} 0_{j_1-1,1} \\ f_1(\vartheta_{1,1}) \end{pmatrix}, \tag{6.8}$$

  where the vector $\vartheta_{1,1} \in \mathbb{R}^{n_1-1}$ is the vector containing the first $n_1 - 1$ entries of ϑ_1.

If $b = 2$ holds, we have determined all columns. Otherwise, it remains to calculate the columns $c_2, \ldots, c_{b-1}$.

- <u>*Second Vector:*</u>
We have to distinguish between two cases. In order to determine the case we are in, we consider the auxiliary variable n_{b-1} and check if it is equal to two. If $n_{b-1} = 2$, we have two entries of the $(b-1)^{th}$ column, which can depend on new parameters. However, due to the normalization and orthogonality to c_1 these entries are already uniquely determined. Thus, we need no additional parameters. In the other case, we cannot compute any column directly. Again, we have to take the normalization into account. This is done by the stereographic projection $f_2(\cdot)$. And the orthogonality to c_1 is guaranteed by the multiplication with a matrix Q_2, which is going to be specified further below. Obviously, the matrix Q_2 depends on the parameter sub-vector $\vartheta_{1,1}$ and can be interpreted as the mapping, which maps the normalized vector $f_2(\cdot)$ into the orthogonal complement of c_1.

  1. <u>*Case 1: $n_{b-1} \neq 2$:*</u>
  In this case the column vector c_2 is given by

$$c_2 = \begin{pmatrix} 0_{j_2-1,1} \\ Q_2 f_2(\vartheta_{1,2}) \end{pmatrix}, \tag{6.9a}$$

  where Q_2 denotes the unique orthogonal complement in N_{plt}-form of the matrix product

$$[0_{n_2,j_2-1}, I_{n_2}][c_1].$$

  Note that Q_2 has dimension $n_2 \times (n_2 - 1)$ and $f_1(\vartheta_{1,2})$ is a unit vector of dimension $n_2 - 1$. Hence, we need $n_2 - 2$ parameters in this case to parameterize the second column.

  2. <u>*Case 2: $n_{b-1} = 2$:*</u>
  In this case column vector c_{b-1} is already determined via the unique orthonormal complement of

$$[0_{2,j_{b-1}-1}, I_2][c_1]$$

  in N_{plt}-form. We denote this unique orthonormal complement in N_{plt}-form by Q_2^*, which is a two-dimensional vector. Consequently, c_{b-1} is given by

$$c_{b-1} = \begin{pmatrix} 0_{j_{b-1}-1,1} \\ Q_2^* \end{pmatrix}. \tag{6.9b}$$

  We did not need any parameters in this case since c_{b-1} has only two non-

zero entries. They are uniquely determined due to the normalization, the orthogonal constraint and the restriction of a positive first entry.

Denote by $C(2)$ *the matrix containing all columns of* $\widetilde{C}_{1,\vartheta_1}$ *which are already determined. Depending on the considered case, it holds that either* $C(2) = [c_1, c_2]$ *or* $C(2) = [c_1, c_{b-1}]$.

We proceed iteratively until we have determined the $b-1$ *columns. However, we state only one more step. The rest will follow in the same way.*

- *Third Vector:*
  We have already determined the two columns stacked in the matrix $C(2)$.

  a) *Case 1:* $n_{b-2} \neq 3$*:*
  In this case we determine the vector c_i. *Depending on the considered case, this is either going to be* c_2 *or* c_3, *i.e.* $i \in \{2, 3\}$. *Hence we have*

$$c_i = \begin{pmatrix} 0_{j_i-1,1} \\ Q_3 f_i(\vartheta_{1,i}) \end{pmatrix}, \tag{6.10a}$$

  where Q_3 *denotes the unique orthogonal complement in* N_{plt}*-form of*

$$[0_{n_i,j_i-1}, I_{n_i}]C(2).$$

  Note that Q_3 *has dimension* $n_i \times (n_i - 2)$ *and* $f(\vartheta_{1,i})$ *is a normalized vector of dimension* $n_i - 2$. *Hence, we need* $n_i - 3$ *parameters in this case to parameterize either the second or the third column.*

  b) *Case 2:* $n_{b-2} = 3$*:*
  In this case c_{b-2} *is already determined via the unique orthonormal complement of*

$$[0_{3,j_{b-1}-1}, I_3]C(2)$$

  in N_{plt}*-form. We denote this unique orthonormal complement in* N_{plt}*-form by* Q_3^*, *which is a three-dimensional vector. Consequently,* c_{b-2} *is given by*

$$c_{b-2} = \begin{pmatrix} 0_{j_{b-2}-1,1} \\ Q_3^* \end{pmatrix}. \tag{6.10b}$$

  Denote by $C(3)$ *the matrix containing all columns of* $\widetilde{C}_{1,\vartheta_1}$ *which are already determined. Depending on the considered case, the matrix* $C(3)$ *is given by one of the following possibilities* $C(3) = [c_1, c_2, c_3]$, $C(3) = [c_1, c_2, c_{b-1}]$ *or* $C(3) = [c_1, c_{b-2}, c_{b-1}]$.

The matrix $\widetilde{C}_{1,\vartheta_1}$ constructed by the above algorithm satisfies Assumption M4, Assumption M6 and Assumption M8 and is of the form given in Theorem 4.3.7, i.e. it satisfies Assumption M9.

Proof. The matrix C_{1,ϑ_1} is obviously in positive lower triangular form due to the construction. Note that Q_j satisfies $Q_j^{\mathsf{T}} Q_j = I_{n_i - i}$ by definition. Then, one can easily see that each column has norm one since

$$\|c_j\|^2 = \|Q_j f(\vartheta_{1,j})\|^2 = f(\vartheta_{1,j})^T Q_j^T Q_j f(\vartheta_{1,j}) = f(\vartheta_{1,j})^T f(\vartheta_{1,j}) = 1$$

and the same holds obviously true for Q_j^*.

Assume without loss of generality that $j < i$. The orthogonality of the columns is given due to

$$c_j^T c_i = c_j^{\mathsf{T}} Q_i f_i(\vartheta_{1,i}) = 0$$

since Q_i is the orthogonal complement of a matrix which contains c_j as a column. The rank condition in Assumption M4 is obviously satisfied as well. Hence, C_{1,ϑ_1} is given in the canonical form of Theorem 4.3.7.

To see the continuity (Assumption M6) and smoothness (Assumption M8) of $C_{1,\cdot}$ note that each column is the product of two smooth functions. After all the stereographic projection is a smooth function and Q_j^* as well as Q_i are function compositions of smooth functions.

In conclusion, the parametrization presented satisfies all desired assumptions. □

Furthermore, we know that the partial derivative of the inverse stereographic projections is orthogonal to the original inverse stereographic projection f, that is

$$\left(\frac{\partial}{\partial \vartheta_j} f(\vartheta_{1,j})^{\mathsf{T}}\right) f(\vartheta_{1,j}) = 0,$$

hence if follows that

$$\frac{\partial}{\partial \vartheta_j} C_{1,\vartheta_1} \notin \text{span } C_{1,\vartheta_1},$$

for $j = 1, \ldots, s_1$. Besides, the set of all partial derivatives is linearly independent.

We present now an example in order to clarify, how the parametrization algorithm works. In this example we have the following parameters given:

$$d = 6, \quad c = 4, \quad r = 2$$

$$i_1 = 1, \ i_2 = 2, \ i_3 = 3, \ i_4 = 5.$$

With this we can calculate b, which is equal to 5, furthermore we have the auxiliary variables given by $n_2 = 5$, $n_3 = 4$, $n_4 = 2$. The first column c_1 is parameterized by the parameter vector $\vartheta_{1,1} = (\vartheta_1, \vartheta_2, \vartheta_3, \vartheta_4, \vartheta_5)^\mathsf{T} \in \mathbb{R}^5$ and hence given as

$$c_1 = \begin{pmatrix} \frac{\|\vartheta_{1,1}\|^2-1}{\|\vartheta_{1,1}\|^2+1} \\ \frac{2\vartheta_1}{\|\vartheta_{1,1}\|^2+1} \\ \frac{2\vartheta_2}{\|\vartheta_{1,1}\|^2+1} \\ \frac{2\vartheta_3}{\|\vartheta_{1,1}\|^2+1} \\ \frac{2\vartheta_4}{\|\vartheta_{1,1}\|^2+1} \\ \frac{2\vartheta_5}{\|\vartheta_{1,1}\|^2+1} \end{pmatrix},$$

where $\|\vartheta_{1,1}\|^2 - 1$ must be a positive real number. Since $n_4 = 2$, we know directly due to the orthogonality constraints that

$$c_4 = \begin{pmatrix} 0 & 0 & 0 & 0 & \frac{\vartheta_5}{\sqrt{\vartheta_4^2+\vartheta_5^2}} & -\frac{\vartheta_4}{\sqrt{\vartheta_4^2+\vartheta_5^2}} \end{pmatrix}^\mathsf{T},$$

where ϑ_5 must be a positive real number. We see, that there is no free parameter needed for c_4, whereas we only get an additional constraint on the parameter space. Next, we compute c_2 with the knowledge of c_1 and c_4. First, we compute the unique positive orthogonal complement of

$$\begin{pmatrix} \frac{2\vartheta_1}{\|\vartheta_{1,1}\|^2+1} & 0 \\ \frac{2\vartheta_2}{\|\vartheta_{1,1}\|^2+1} & 0 \\ \frac{2\vartheta_3}{\|\vartheta_{1,1}\|^2+1} & 0 \\ \frac{2\vartheta_4}{\|\vartheta_{1,1}\|^2+1} & \frac{\vartheta_5}{\sqrt{\vartheta_4^2+\vartheta_5^2}} \\ \frac{2\vartheta_5}{\|\vartheta_{1,1}\|^2+1} & -\frac{\vartheta_4}{\sqrt{\vartheta_4^2+\vartheta_5^2}} \end{pmatrix}.$$

The orthogonal complement space must be three dimensional. As a consequence, we obtain

$$Q_2 = \begin{pmatrix} q_1 & 0 & 0 \\ q_1 \cdot \vartheta_1\vartheta_2 & q_2 & 0 \\ q_1 \cdot \vartheta_1\vartheta_3 & q_2 \cdot \vartheta_2\vartheta_3 & q_3 \\ q_1 \cdot \vartheta_1\vartheta_4 & q_2 \cdot \vartheta_2\vartheta_4 & q_3 \cdot \vartheta_3\vartheta_4 \\ q_1 \cdot \vartheta_1\vartheta_5 & q_2 \cdot \vartheta_2\vartheta_5 & q_3 \cdot \vartheta_3\vartheta_5 \end{pmatrix},$$

where $q_1 := \frac{\sqrt{\vartheta_2^2+\vartheta_3^2+\vartheta_4^2+\vartheta_5^2}}{\sqrt{\vartheta_1^2+\vartheta_2^2+\vartheta_3^2+\vartheta_4^2+\vartheta_5^2}}$, $q_2 := \frac{\sqrt{\vartheta_3^2+\vartheta_4^2+\vartheta_5^2}}{\sqrt{\vartheta_2^2+\vartheta_3^2+\vartheta_4^2+\vartheta_5^2}}$ and $q_3 := \frac{\sqrt{\vartheta_4^2+\vartheta_5^2}}{\sqrt{\vartheta_3^2+\vartheta_4^2+\vartheta_5^2}}$. In order to obtain c_2 we have to multiply Q_2 with $f(\vartheta_{1,2})$ using the parameter vector $\vartheta_{1,2} =$

$(\vartheta_6, \vartheta_7)^\mathsf{T} \in \mathbb{R}^2$. Calculating this product gives us

$$c_2 = Q_2 \cdot \begin{pmatrix} 0 \\ q_1 \cdot \frac{\|\vartheta_{1,2}\|^2-1}{\|\vartheta_{1,2}\|^2+1} \\ -q_1 \cdot \vartheta_1\vartheta_2 \cdot \frac{\|\vartheta_{1,2}\|^2-1}{\|\vartheta_{1,2}\|^2+1} + q_2 \cdot \frac{2\vartheta_6}{\|\vartheta_{1,2}\|^2+1} \\ -q_1 \cdot \vartheta_1\vartheta_3 \cdot \frac{\|\vartheta_{1,2}\|^2-1}{\|\vartheta_{1,2}\|^2+1} - q_2 \cdot \vartheta_2\vartheta_3 \cdot \frac{2\vartheta_6}{\|\vartheta_{1,2}\|^2+1} + q_3 \cdot \frac{2\vartheta_7}{\|\vartheta_{1,2}\|^2+1} \\ -q_1 \cdot \vartheta_1\vartheta_4 \cdot \frac{\|\vartheta_{1,2}\|^2-1}{\|\vartheta_{1,2}\|^2+1} - q_2 \cdot \vartheta_2\vartheta_4 \cdot \frac{2\vartheta_6}{\|\vartheta_{1,2}\|^2+1} - q_3 \cdot \vartheta_3\vartheta_4 \cdot \frac{2\vartheta_7}{\|\vartheta_{1,2}\|^2+1} \\ -q_1 \cdot \vartheta_1\vartheta_5 \cdot \frac{\|\vartheta_{1,2}\|^2-1}{\|\vartheta_{1,2}\|^2+1} - q_2 \cdot \vartheta_2\vartheta_5 \cdot \frac{2\vartheta_6}{\|\vartheta_{1,2}\|^2+1} - q_3 \cdot \vartheta_3\vartheta_5 \cdot \frac{2\vartheta_7}{\|\vartheta_{1,2}\|^2+1} \end{pmatrix},$$

where $\|\vartheta_{1,2}\|^2 - 1 > 0$ must hold. As of now, we also know c_3 due to the orthogonality constraints, which is given by the orthogonal complement and thus we get

$$c_3 := \begin{pmatrix} 0 \\ 0 \\ \frac{\vartheta_7\sqrt{(\vartheta_3^2+\vartheta_4^2+\vartheta_5^2)}}{\|\vartheta_{1,2}\|\sqrt{\vartheta_2^2+\vartheta_3^2+\vartheta_4^2+\vartheta_5^2}} \\ -\frac{\vartheta_2\vartheta_3\vartheta_7}{\|\vartheta_{1,2}\|\sqrt{\vartheta_2^2+\vartheta_3^2+\vartheta_4^2+\vartheta_5^2}\sqrt{\vartheta_3^2+\vartheta_4^2+\vartheta_5^2}} - \frac{\vartheta_6\sqrt{\vartheta_4^2+\vartheta_5^2}}{\|\vartheta_{1,2}\|\sqrt{\vartheta_3^2+\vartheta_4^2+\vartheta_5^2}} \\ -\frac{\vartheta_2\vartheta_4\cdot\vartheta_7}{\|\vartheta_{1,2}\|\sqrt{\vartheta_2^2+\vartheta_3^2+\vartheta_4^2+\vartheta_5^2}\sqrt{\vartheta_3^2+\vartheta_4^2+\vartheta_5^2}} - \frac{\vartheta_3\vartheta_4\cdot\vartheta_6}{\|\vartheta_{1,2}\|\sqrt{\vartheta_4^2+\vartheta_5^2}\sqrt{\vartheta_3^2+\vartheta_4^2+\vartheta_5^2}} \\ -\frac{\vartheta_2\vartheta_5\cdot\vartheta_7}{\|\vartheta_{1,2}\|\sqrt{\vartheta_2^2+\vartheta_3^2+\vartheta_4^2+\vartheta_5^2}\sqrt{\vartheta_3^2+\vartheta_4^2+\vartheta_5^2}} - \frac{\vartheta_3\vartheta_5\cdot\vartheta_6}{\|\vartheta_{1,2}\|\sqrt{\vartheta_4^2+\vartheta_5^2}\sqrt{\vartheta_3^2+\vartheta_4^2+\vartheta_5^2}} \end{pmatrix}.$$

Hence, we could calculate now $C_{1,\vartheta_1}^{\perp}$ by taking the unique orthogonal complement of C_{1,ϑ_1} in N_{plt}-form.

In Table 6.1 we show the number of parameters needed for different combinations of the number of common stochastic trends c, the dimension of the output process d and the multi-index j. We give the number of parameters per column vector, i.e. for example $(3, 2, 0)$ means that the first column vector is parameterized with three parameters the second one with two and the last needs no additional parameter. We omit the trivial case of $c = d$, because the process is then integrated but not cointegrated. Moreover, it follows that $C_1 = I_d$ and no estimation of long-run parameters has to be done. The table shows that in most cases we do not need many parameters due to all the constraints. Note further that for $c = d - 1$ we only have to parameterize the first vector. All other vectors are then given due to the orthogonality constraints. So the more complicated cases of parametrization occur only for $1 < c < d - 1$. For one common stochastic trend ($c = 1$) we always need $d - i_1$ parameters. Moreover, we excluded the multi-indices in the table, which belong to the class, where it suffices to consider the reduced problem. In the subsequent overview, we present the parameters needed up to dimension $d = 6$.

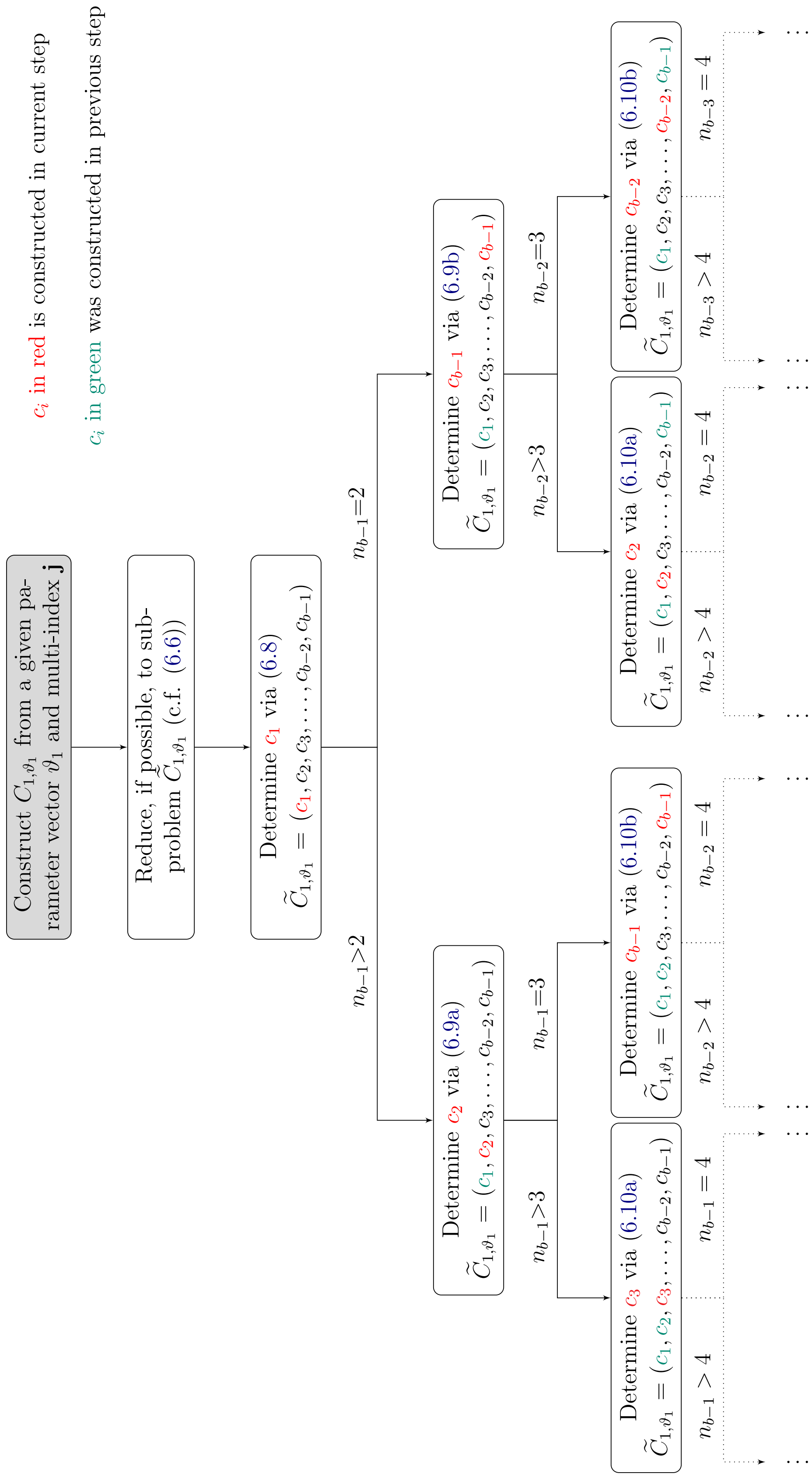

Figure 6.1.: Illustration of the algorithm for constructing C_{1,ϑ_1}

Dimension $d = 4$		
c	**j**	# Parameters
2	(1,2)	(3,1)
	(1,3)	(3,0)
	(2,3)	(2,0)

Dimension $d = 5$		
c	**j**	# Parameters
2	(1,2)	(4,2)
	(1,3)	(4,1)
	(2,3)	(3,1)
	(1,4)	(4,0)
	(2,4)	(3,0)
	(3,4)	(2,0)
3	(1,2,3)	(4,2,0)
	(1,2,4)	(4,1,0)
	(1,3,4)	(4,0,0)
	(2,3,4)	(3,0,0)

Dimension $d = 6$		
c	**j**	# Parameters
2	(1,2)	(5,3)
	(1,3)	(5,2)
	(1,4)	(5,1)
	(1,5)	(5,0)
	(2,3)	(4,2)
	(2,4)	(4,1)
	(2,5)	(4,0)
	(3,4)	(3,1)
	(3,5)	(3,0)
	(4,5)	(2,0)
3	(1,2,3)	(5,3,1)
	(1,2,4)	(5,3,0)
	(1,2,5)	(5,2,0)
	(1,3,4)	(5,2,0)
	(1,3,5)	(5,1,0)
	(1,4,5)	(5,0,0)
	(2,3,4)	(4,2,0)
	(2,3,5)	(4,1,0)
	(2,4,5)	(4,0,0)
	(3,4,5)	(3,0,0)
4	(1,2,3,4)	(5,3,0,0)
	(1,2,3,5)	(5,2,0,0)
	(1,2,4,5)	(5,1,0,0)
	(1,3,4,5)	(5,0,0,0)
	(2,3,4,5)	(4,0,0,0)

Table 6.1.: Number of Parameters needed for the Parametrization

We can freely select the parametrization for the stationary part due to the decoupled system derived in Theorem 4.3.7. Therefore, we choose the canonical echelon form for this part. We omit a detailed description of this parametrization and only mention that it satisfies all necessary assumptions which we made in Chapter 5 in order to gain consistency and the asymptotic distributions. A list of all assumptions can be found in Appendix A. For a reference on this canonical parametrization we refer to Schlemm and Stelzer [91], Section 4.1, where the necessary facts and results are summarized. Other references would be for example the article of Lütkepohl and Poskitt [64], and the textbooks of Lütkepohl [62], or Hannan and Deistler [46].

In order to use the echelon form for the stationary subsystem, namely the matrices $(A_{2,\vartheta_2}, B_{2,\vartheta_2}, C_{2,\vartheta_2})$, we need to know the Kronecker index. We assume that we can somehow estimate the Kronecker index of the model with an information criterion.

Well-known information criteria are for example the Akaike Information Criterion (AIC), which was introduced by Akaike [2] in 1973, or the Bayesian Information Criterion (BIC) dates back to Schwarz [92] in 1978. The profound treatment of the order selection problem for discrete-time weak VARMA processes can be found e.g. in Boubacar [13]. A recent work on information criteria for stationary MCARMA processes is Fasen and Kimmig [38]. However, to treat information criteria for the cointegrated model is beyond the scope of this thesis.

Lastly, we consider the parametrization of the remaining system matrices. The matrix B_{1,ϑ_2} is parameterized using the vec operator, i.e. each entry of B_{1,ϑ_2} consists of a an entry of a $(d \cdot c)$-dimensional sub-vector of ϑ_2. Moreover, the covariance matrix of the Lévy process $\Sigma^L_{\vartheta_2}$ is parameterized via the vech operator. We make use of the symmetric form and hence only need to parameterize the lower or upper triangular matrix respectively. As for B_{1,ϑ_2} each entry consists of a an entry of a $\left(\frac{m(m+1)}{2}\right)$-dimensional sub-vector of the short-run parameter vector ϑ_2.

6.3. Simulation Results

At this point, we would like to thank Eckhard Schlemm and Robert Stelzer, who kindly provided the MATLAB code for the simulation and parameter estimation of the stationary MCARMA process. This MATLAB code was the foundation for the simulation studies of this thesis. We adapted the code in order to include not only stationary MCARMA processes, but also the cointegrated MCARMA model. In this respect, we have used the same methods as described in Section 4.2 of Schlemm and Stelzer [91].

In the course of the extension we have heavily used the decoupling of the cointegrated model given by Theorem 4.3.7. Moreover, we have implemented the step-wise quasi-maximum likelihood estimation approach as described in Chapter 5. Evidently, we use the parametrization described in Section 6.2 for the matrix C_{1,ϑ_1} along with the other parametrization methods mentioned for the matrices corresponding to the short-run parameters. The exact parametrization for both model scenarios considered are presented in the following.

6.3.1. Bivariate Examples

We test the performance of the estimation method presented in Chapter 5 with the example used in Schlemm and Stelzer [91], Section 4.2. Hence, we simulate a bivariate

CARMA process with Kronecker indices $(1,2)$ and cointegration multi-index $i_1 = 1$. This implies that the stationary MCARMA process has order $(p,q) = (2,1)$ and we have one cointegration relation.

We consider as a driving Lévy process a normal-inverse Gaussian (NIG) process $(L(t))_{t\geq 0}$ with mean zero. Additionally, we simulate the case, where we have a Brownian motion as a driving Lévy process $(L(t))_{t\geq 0}$. We adjust the covariance matrices such that they coincide in both scenarios. The covariance matrix is thus given as in Schlemm and Stelzer by

$$\Sigma_L = \frac{1}{31^{\frac{3}{2}}} \begin{pmatrix} 82 & -28 \\ -28 & 64 \end{pmatrix} \approx \begin{pmatrix} 0.4751 & -0.1622 \\ -0.1622 & 0.3708 \end{pmatrix}. \tag{6.11}$$

We used the following set of parameters for the simulation of the bivariate CARMA process

$$\vartheta^0 = \begin{pmatrix} -1 & -2 & 1 & -2 & -3 & 1 & 2 & 1 & 1 & 0.4751 & -0.1622 & 0.3708 & 3 \end{pmatrix}^\mathsf{T}. \tag{6.12}$$

For this setting, the dimension of the matrices are as follows

$$A \in M_{4,4}(\mathbb{R}), \quad B \in M_{4,2}(\mathbb{R}) \text{ and } C \in M_{2,4}(\mathbb{R}).$$

The canonical parametrization of the model has the following state space form

$$\mathrm{d}X(t) = \begin{pmatrix} \vartheta_1 & \vartheta_2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \vartheta_3 & \vartheta_4 & \vartheta_5 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} X(t)\mathrm{d}t + \begin{pmatrix} \vartheta_1 & \vartheta_2 \\ \vartheta_6 & \vartheta_7 \\ \vartheta_3 + \vartheta_5\vartheta_6 & \vartheta_4 + \vartheta_5\vartheta_7 \\ \vartheta_8 & \vartheta_9 \end{pmatrix} \mathrm{d}L(t),$$

and

$$Y(t) = \begin{pmatrix} 1 & 0 & 0 & \frac{\vartheta_{13}^2 - 1}{\vartheta_{13}^2 + 1} \\ 0 & 1 & 0 & \frac{2\cdot\vartheta_{13}}{\vartheta_{13}^2 + 1} \end{pmatrix} X(t). \tag{6.13}$$

Furthermore, we set the initial value $X(0) = 0_3$ and the covariance matrix was parameterized by $(\vartheta_{10}, \vartheta_{11}, \vartheta_{12}) = \mathrm{vech}\Sigma_L$.

The cointegration space is directly found by taking the span of the orthogonal complement of $C_1 = \begin{pmatrix} 0.8 \\ 0.6 \end{pmatrix}$, i.e. we have $C_1^\perp = \begin{pmatrix} -0.6 \\ 0.8 \end{pmatrix}$.

We simulate on a grid $0, 0.01, 0.02, \ldots, 2000$ using an Euler scheme to the stochastic differential equation (6.13). The sampling distance was set to $h = 1$.

Case 1: Normal-Inverse Gaussian

Since the parameters were chosen accordingly to the simulation study of Schlemm and Stelzer, we also briefly recall the properties of a multivariate NIG-Lévyprocess (for more details see e.g. Barndorff-Nielsen [5]). Moreover, we remain closely to the notation of Schlemm and Stelzer [91] in this section.

The increment of a two-dimensional NIG-Lévy process $L(t) - L(t-1)$ has the density

$$f_{NIG}(x; \mu, \alpha, \beta, \delta, \Delta) = \frac{\delta \mathrm{e}^{\delta\kappa}}{2\pi} \cdot \frac{\mathrm{e}^{\langle \beta, x\rangle}(1 + \alpha g(x))}{\mathrm{e}^{\alpha g(x)} g(x)^3}, \tag{6.14}$$
$$\text{where} \quad g(x) = \sqrt{\delta^2 + \langle x - \mu, \Delta(x - \mu)\rangle}$$
$$\text{and} \quad \kappa^2 = \alpha^2 - \langle \beta, \Delta\beta\rangle > 0.$$

The covariance of the process is in this case given by

$$\Sigma = \delta(\alpha - \beta^\mathsf{T}\Delta\beta)^{-\frac{1}{2}} \left(\Delta + (\alpha^2 - \beta^\mathsf{T}\Delta\beta)^{-1}\Delta\beta\beta^T\Delta\right) \tag{6.15}$$

and in order to obtain the covariance matrix in (6.11) and mean zero, we have to set the parameters of the NIG-process to

$$\delta = 1, \quad \alpha = 3, \quad \beta = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \Delta = \begin{pmatrix} 1.2 & -0.5 \\ -0.5 & 1 \end{pmatrix} \quad \text{and} \quad \mu = -\frac{1}{2\sqrt{31}} \begin{pmatrix} 3 \\ 2 \end{pmatrix}.$$

The results for a sample of 350 replicates of the bivariate NIG-driven MCARMA process are summarized in Table 6.2. One realization of the cointegrated MCARMA process driven by a normal-inverse Gaussian process is given in Figure 6.2 and the corresponding stationary linear combination in Figure 6.3.

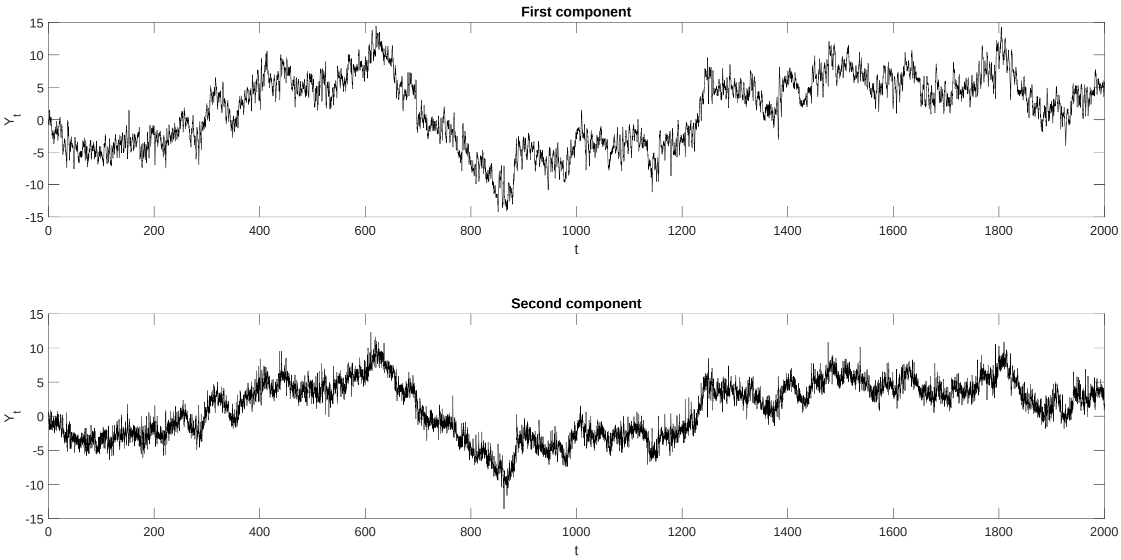

Figure 6.2.: Typical realization of bivariate NIG-driven CARMA process

	Bivariate NIG-driven MCARMA process			
parameter	true param.	sample mean	bias	sample std. dev.
ϑ_1	-1	-0.9857	-0.0143	0.0515
ϑ_2	-2	-2.0025	0.0025	0.0573
ϑ_3	1	0.9919	0.0081	0.0749
ϑ_4	-2	-1.9758	-0.0242	0.1126
ϑ_5	-3	-2.9774	-0.0226	0.0497
ϑ_6	1	1.0129	-0.0129	0.1071
ϑ_7	2	2.0005	-0.0005	0.0690
ϑ_8	1	1.0078	-0.0078	0.0684
ϑ_9	1	0.9872	0.0128	0.0761
ϑ_{10}	0.4751	0.4715	0.0036	0.0678
ϑ_{11}	-0.1622	-0.1572	-0.0050	0.0381
ϑ_{12}	0.3708	0.3698	0.0010	0.0314
ϑ_{13}	3	2.9999	0.0001	0.0075

Table 6.2.: Estimates for the parameters of a bivariate NIG-driven CARMA process

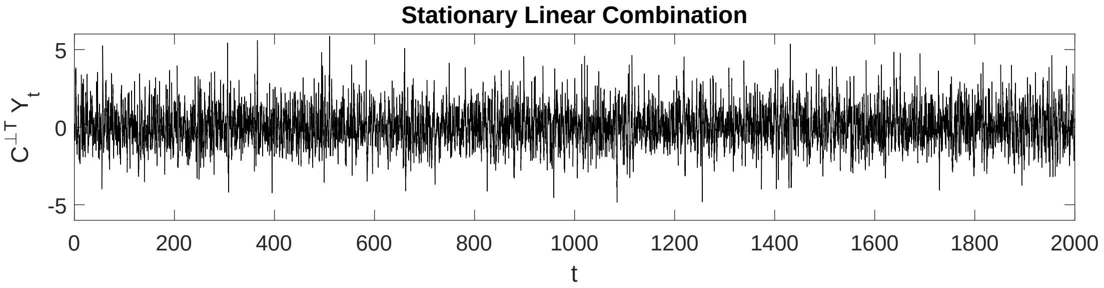

Figure 6.3.: Typical realization the stationary linear combination

---

Case 2: Brownian Motion

Now we consider the case we the Lévy process is a Brownian motion. The only difference to the previous case is the driving process. In this case the parameters $(\vartheta_{10}, \vartheta_{11}, \vartheta_{12})$ directly give the covariance matrix Σ_L. The results for a sample of 350 replicates of the bivariate BM-driven cointegrated MCARMA process are summarized in Table 6.3 and one realization is given in Figure 6.4.

	Bivariate BM-driven MCARMA process			
parameter	true param.	sample mean	bias	sample std. dev.
ϑ_1	-1	-0.9895	-0.0105	0.0425
ϑ_2	-2	-1.9934	-0.0066	0.0459
ϑ_3	1	0.9898	0.0102	0.0570
ϑ_4	-2	-1.9701	-0.0299	0.0872
ϑ_5	-3	-2.9898	-0.0102	0.0324
ϑ_6	1	1.0155	-0.0155	0.0789
ϑ_7	2	2.0068	-0.0068	0.0441
ϑ_8	1	1.0096	-0.0096	0.0482
ϑ_9	1	0.9777	0.0223	0.0599
ϑ_{10}	0.4751	0.5200	-0.0449	0.0518
ϑ_{11}	-0.1622	-0.1283	-0.0339	0.0266
ϑ_{12}	0.3708	0.3195	0.0513	0.0213
ϑ_{13}	3	2.9981	0.0019	0.0068

Table 6.3.: Estimates for the parameters of a bivariate BM-driven CARMA process

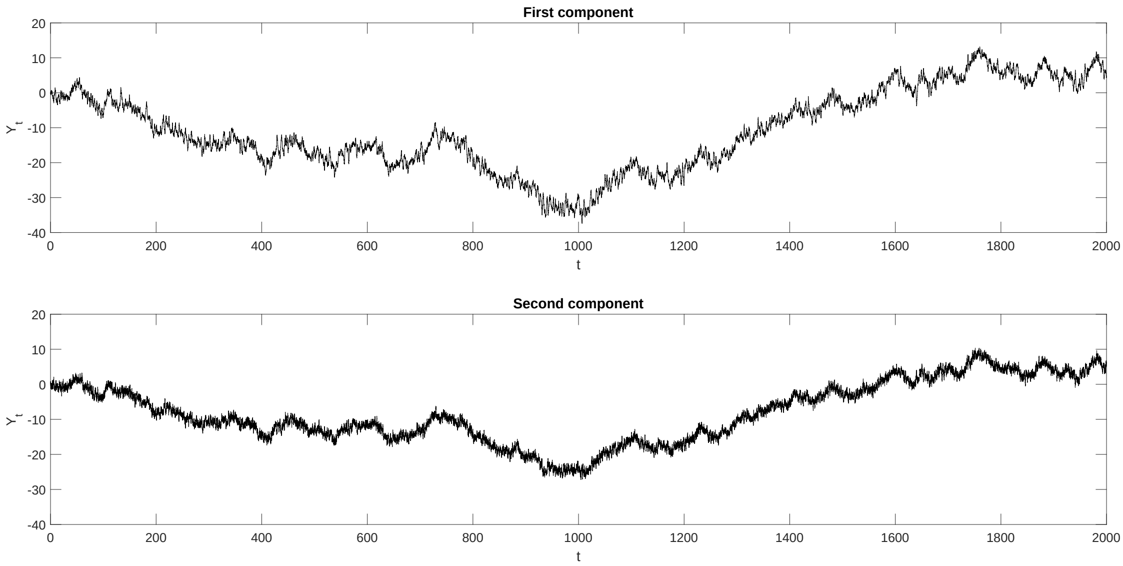

Figure 6.4.: Typical realization of bivariate BM-driven CARMA process

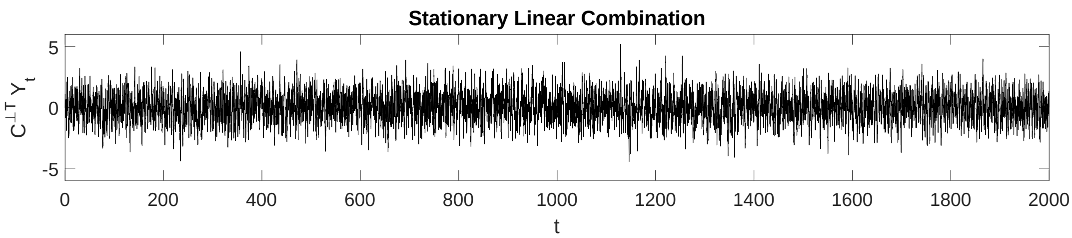

Figure 6.5.: Typical realization the stationary linear combination

---

In summary, we conclude that the theoretical results proven in Chapter 5 have given us an estimation procedure, which performs quite well in the two simulation studies for the two-dimensional case. The results of the NIG-case are comparable to the results of the simulation study in Section 4.2. of Schlemm and Stelzer [91]. This simulation study was done for a stationary MCARMA process with the same parameters. Hence, for reasons of comparability we have chosen the true parameters as in (6.12). Compare for this purpose Table 6.2 with Table 3 in Schlemm and Stelzer [91].

In Figure 6.6 we compare the bias and sample standard deviation of the two simulation studies. We have chosen two completely different Lévy processes with respect to the continuity of the sample paths. The Brownian motion has continuous sample paths, whereas the normal-inverse Gaussian process is a pure jump process. The Kalman filter as well as the quasi-maximum likelihood function are constructed marginally for

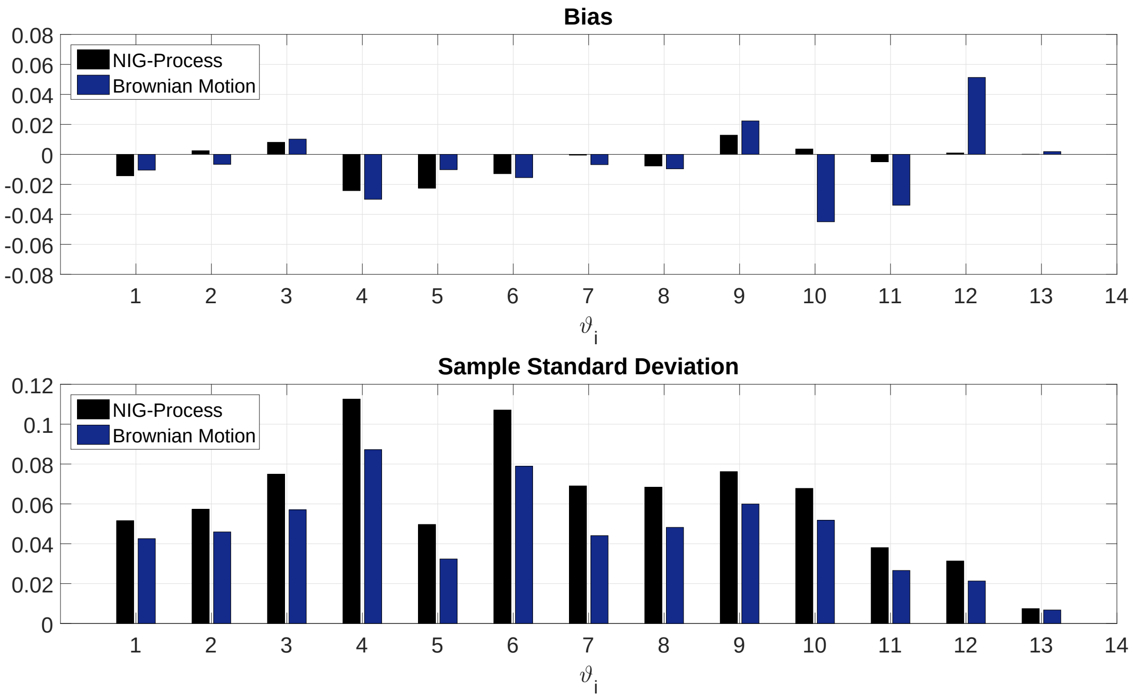

Figure 6.6.: Bias and Sample Standard Deviation of both Simulation Studies

the Gaussian case. The figure shows that the sample standard deviation is smaller in the case of a driving Brownian motion is to be expected. It is very apparent that the cointegration parameter is estimated extremely well. Accordingly, the cointegration parameter ϑ_{13} has in both cases the smallest bias and standard deviation of all parameters. This is based on the fact that in this simple two-dimensional model there is only one long-run parameter. Hence, the complexity of this estimation is quite low and consequently the optimization performs quite narrowly, which is in accordance with the consistency results shown in Section 5.5.

Lastly, we compare for the Gaussian MCARMA case the minimal likelihood values if we do not use the right model. To be precise, we estimated the Brownian motion-driven MCARMA model of this subsection using all possible models with respect to the cointegration rank of appropriate dimension. There are four cases which are given by

1. A stationary MCARMA model (A_2, B_2, C_2, L);
2. A cointegrated MCARMA model with multi-index (2);
3. The true cointegrated MCARMA model with multi-index (1);
4. An integrated MCARMA model.

Minimal value of the log-likelihood function							
True Model (1)		Stationary Model		Coint. Model (2)		Integrated model	
5.24	5.28	14.36	14.97	27.63	22.27	5.16	5.25
5.23	5.19	14.75	19.13	9.62	8.15	5.21	5.44
5.23	5.26	13.40	11.51	7.31	19.06	5.30	5.24
5.30	5.21	29.55	18.12	18.15	10.73	5.35	5.23
5.22	5.21	50.86	46.63	16.41	13.25	5.18	5.33
5.24	5.24	33.29	21.84	7.39	11.59	5.59	5.31
5.27	5.24	20.43	32.78	10.91	8.01	5.24	5.32
5.25	5.16	14.85	47.20	18.84	11.62	5.29	5.28
5.27	5.24	18.15	54.99	9.87	11.14	5.33	5.20
5.20	5.29	17.21	41.14	43.01	9.89	5.29	5.39
5.25	5.18	13.03	22.90	10.31	10.98	5.42	5.21
5.21	5.24	10.90	14.02	13.28	10.41	5.24	5.26
5.25	5.31	16.18	17.50	32.19	13.46	5.31	5.25
5.18	5.16	12.53	13.18	8.07	30.89	5.40	5.16
5.21	5.26	43.47	14.26	24.32	12.34	5.29	5.23
5.29	5.25	85.46	14.55	16.55	24.56	5.31	5.24
5.25	5.27	10.53	16.38	13.34	8.90	5.54	5.19
5.25	5.21	42.85	15.45	14.40	22.44	5.31	5.22
5.30	5.30	27.48	25.55	6.48	8.92	5.33	5.32
5.19	5.25	23.54	11.66	10.82	10.37	5.24	5.23
5.31	5.23	19.08	16.28	10.55	16.11	5.35	5.23
5.34	5.21	23.16	18.01	12.22	12.24	5.21	5.44
5.24	5.18	30.15	22.79	23.87	9.54	5.30	5.28
5.22	5.21	16.05	12.93	12.56	8.80	5.22	5.21
5.23	5.23	14.31	51.27	10.69	9.61	5.25	5.24
5.31	5.20	12.70	11.20	14.80	11.13	5.44	5.25
5.23	5.28	11.21	44.18	9.79	52.56	5.25	5.23
5.27	5.23	34.06	52.53	10.95	22.10	5.23	5.17
5.29	5.19	16.34	17.44	45.28	79.37	5.40	5.45
5.13	5.24	9.62	22.56	12.92	23.48	5.22	5.21
5.26	5.29	15.28	16.69	19.68	31.53	5.29	5.34
5.21	5.13	11.57	19.32	8.78	9.93	5.28	5.26
5.33	5.27	9.45	11.96	10.19	56.78	5.25	5.32
5.16	5.22	19.84	14.10	22.24	6.21	5.14	5.25
5.17	5.19	14.60	34.31	30.04	9.42	5.41	5.24
5.17	5.20	11.07	11.20	24.11	21.86	5.27	5.43
5.28	5.24	39.36	13.03	20.62	9.15	5.29	5.32
5.26	5.24	52.76	17.96	6.40	10.85	5.21	5.18
5.21	5.20	20.09	41.32	10.43	23.59	5.75	5.36
5.12	5.22	21.04	17.98	11.53	8.14	5.17	5.21
5.28	5.21	13.65	31.78	20.07	8.80	5.20	5.44
5.19	5.18	14.74	23.95	22.55	7.69	5.24	5.30
5.17	5.17	13.23	31.73	9.18	22.29	5.34	5.16
5.22	5.24	11.49	15.52	8.98	21.41	5.19	5.29
5.22	5.22	14.53	20.71	12.30	16.51	5.23	5.17
5.21	5.20	17.37	30.97	8.45	11.63	5.25	5.34
5.17	5.25	44.94	13.59	18.10	10.65	5.31	5.36
5.25	5.20	12.77	14.15	16.82	12.52	5.25	5.28
5.22	5.28	19.21	82.80	6.05	10.94	5.21	5.20
5.16	5.29	88.27	13.95	13.01	11.25	5.29	5.31

Table 6.4.: Minimum of the likelihood function for the four different models and 100 simulations each

We compare for each model the minimal value of the likelihood function, namely $\mathcal{L}_n^{(h)}(\widehat{\vartheta}_n)$, for 100 simulations in Table 6.4. We see that the true model and the integrated model behave similar in the sense of the value of the likelihood function.

The mean, standard deviation, minimal and maximal value of the minimum of likelihood function for the 100 simulations are summarized in Table 6.5.

	True Coint. Model (1)	Stationary Model	Cointegrated Model (2)	Integrated model
mean	5.2303	23.8473	16.2713	5.2851
st. dev.	0.0449	16.0159	11.3465	0.0956
min	5.1226	9.4492	6.0526	5.1367
max	5.3356	88.2747	79.3741	5.7509

Table 6.5.: Minimum of the Likelihood function for the four different models

The results indicate that the likelihood function is not converging for the cases where the stationary model and the wrong cointegrated model are chosen as would be expected by the theoretical results in Chapter 5. The reason for this is that there is no chance to estimate the true cointegration space due to the unique parametrization. Hence, the non-stationary subsystem is estimated with a stationary model in both cases. The likelihood function for the integrated model seems to converge which is probably the case due to the fact that the integrated model estimates the non-stationary part with a non-stationary model. However, it „overfits“ the model in a sense. In other words, the integrated model includes another dimension of non-stationary which is not present in the true model.

6.3.2. Three-dimensional Examples

We also simulate a three-dimensional CARMA process with Kronecker indices $(1, 2, 1)$ and cointegration multi-index $(i_1, i_2) = (1, 2)$. This implies that the MCARMA process has order $(p, q) = (2, 1)$ and two common stochastic trends. The cointegration space is accordingly a one-dimensional subspace of $\mathbb{R}^3$. We need 28 parameters in total to parameterize all matrices of this model.

As before we simulate the cointegrated process using either a normal-inverse Gaussian (NIG) process with mean zero or a Brownian motion. We set the covariance matrix for the three-dimensional process in both cases to

$$\Sigma_L \approx \begin{pmatrix} 0.5310 & -0.1934 & 0.1678 \\ -0.1934 & 0.3784 & -0.2227 \\ 0.1678 & -0.2227 & 0.5632 \end{pmatrix}. \tag{6.16}$$

The true parameter values for the simulation of the three-dimensional MCARMA process are given in Table 6.6.

ϑ_1	ϑ_2	ϑ_3	ϑ_4	ϑ_5	ϑ_6	ϑ_7	ϑ_8	ϑ_9	ϑ_{10}
-2	-3	-3	1	1	-1	2	-1	-3	-3

ϑ_{11}	ϑ_{12}	ϑ_{13}	ϑ_{14}	ϑ_{15}	ϑ_{16}	ϑ_{17}	ϑ_{18}	ϑ_{19}	ϑ_{20}
-1	-1	2	1	1	0	1	1	-2	0

ϑ_{21}	ϑ_{22}	ϑ_{23}	ϑ_{24}	ϑ_{25}	ϑ_{26}	ϑ_{27}	ϑ_{28}
0.5310	-0.1934	0.1678	0.3784	-0.2227	0.5632	1	2

Table 6.6.: Parameters for the simulation of the three-dimensional MCARMA process

The dimensions of the model matrices are as follows

$$A \in M_{6,6}(\mathbb{R}), \quad B \in M_{6,3}(\mathbb{R}) \text{ and } C \in M_{3,6}(\mathbb{R}).$$

The canonical parametrization of the model has the following state space form

$$\mathrm{d}X(t) = \begin{pmatrix} A_2 & 0_{4\times 2} \\ 0_{2\times 4} & 0_{2\times 2} \end{pmatrix} X(t)\mathrm{d}t + \begin{pmatrix} B_2 \\ B_1 \end{pmatrix} \mathrm{d}L(t)$$

and

$$Y(t) = \begin{pmatrix} C_2 & C_1 \end{pmatrix} X(t), \tag{6.17}$$

where the matrices are given by

$$A_2 = \begin{pmatrix} \vartheta_1 & \vartheta_2 & 0 & \vartheta_3 \\ 0 & 0 & 1 & 0 \\ \vartheta_4 & \vartheta_5 & \vartheta_6 & \vartheta_7 \\ \vartheta_8 & \vartheta_9 & \vartheta_{10} & \vartheta_{11} \end{pmatrix},$$

$$C_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad C_1 = \begin{pmatrix} \frac{\vartheta_{27}^2+\vartheta_{28}^2-1}{\vartheta_{27}^2+\vartheta_{28}^2+1} & 0 \\ \frac{2\cdot\vartheta_{27}}{\vartheta_{27}^2+\vartheta_{28}^2+1} & \frac{\vartheta_{28}}{\sqrt{\vartheta_{27}^2+\vartheta_{28}^2}} \\ \frac{2\cdot\vartheta_{28}}{\vartheta_{27}^2+\vartheta_{28}^2+1} & -\frac{\vartheta_{27}}{\sqrt{\vartheta_{27}^2+\vartheta_{28}^2}} \end{pmatrix},$$

$$B_2 = \begin{pmatrix} \vartheta_1 & \vartheta_2 & \vartheta_3 \\ \vartheta_{12} & \vartheta_{13} & \vartheta_{14} \\ \vartheta_4+\vartheta_6\vartheta_{12} & \vartheta_5+\vartheta_6\vartheta_{13} & \vartheta_7+\vartheta_6\vartheta_{14} \\ \vartheta_8+\vartheta_{10}\vartheta_{12} & \vartheta_9+\vartheta_{10}\vartheta_{13} & \vartheta_{11}+\vartheta_{10}\vartheta_{14} \end{pmatrix} \quad \text{and} \quad B_1 = \begin{pmatrix} \vartheta_{15} & \vartheta_{16} & \vartheta_{17} \\ \vartheta_{18} & \vartheta_{19} & \vartheta_{20} \end{pmatrix}.$$

Furthermore, we set the initial value $X(0) = 0_6$ and the covariance matrix is once again parameterized by the vector $(\vartheta_{21}, \vartheta_{22}, \vartheta_{23}, \vartheta_{24}, \vartheta_{25}, \vartheta_{26}) = \text{vech}\Sigma_L$.

In this model we have is now

$$C_1 = \begin{pmatrix} 0.6667 & 0 \\ 0.3333 & 0.8944 \\ 0.6667 & -0.4472 \end{pmatrix}$$

and the orthogonal complement is given by

$$C_1^{\perp} = \begin{pmatrix} 0.7454 \\ -0.2981 \\ -0.5963 \end{pmatrix}.$$

Once more, we simulate on a grid $0, 0.01, 0.02, \ldots, 2000$ using an Euler scheme to the stochastic differential equation (6.17) and use a sampling distance of $h = 1$.

Case 1: Normal-Inverse Gaussian

The NIG-process has covariance matrix (6.11) and mean zero if we set the parameters of the normal-inverse Gaussian distribution to

$$\delta = 1, \quad \alpha = 3, \quad \beta = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \Delta = \begin{pmatrix} 1.25 & -0.5 & \frac{1}{6}\sqrt{3} \\ -0.5 & 1 & -\frac{1}{3}\sqrt{3} \\ \frac{1}{6}\sqrt{3} & -\frac{1}{3}\sqrt{3} & \frac{4}{3} \end{pmatrix}$$

$$\text{and} \quad \mu = -\frac{1}{2\sqrt{31}} \begin{pmatrix} 3 \\ 2 \end{pmatrix}.$$

This implies that the covariance matrix of the NIG-process is indeed

$$\Sigma_L \approx \begin{pmatrix} 0.5310 & -0.1934 & 0.1678 \\ -0.1934 & 0.3784 & -0.2227 \\ 0.1678 & -0.2227 & 0.5632 \end{pmatrix}$$

as in (6.16).

We proceed as in the two dimensional case. The results for a sample of 350 replicates of the 3-dimensional NIG-driven MCARMA process are summarized in Table 6.7 and a realization can be found in Figure 6.7.

	3-dim. NIG-driven MCARMA process			
parameter	true param.	sample mean	bias	sample std. dev.
ϑ_1	-2	-1.9910	-0.0090	0.0583
ϑ_2	-3	-3.0042	0.0042	0.0407
ϑ_3	-3	-3.0194	0.0194	0.0456
ϑ_4	1	0.9887	0.0113	0.0440
ϑ_5	1	0.9977	0.0023	0.0351
ϑ_6	-1	-0.9861	-0.0139	0.0544
ϑ_7	2	2.0122	-0.0122	0.0396
ϑ_8	-1	-1.0039	0.0039	0.0442
ϑ_9	-3	-2.9937	-0.0063	0.0342
ϑ_{10}	-3	-2.9904	-0.0096	0.0490
ϑ_{11}	-1	-1.0055	0.0055	0.0449
ϑ_{12}	-1	-1.0023	0.0023	0.0386
ϑ_{13}	2	1.9984	0.0016	0.0363
ϑ_{14}	1	1.0034	-0.0034	0.0353
ϑ_{15}	1	0.9984	0.0016	0.0351
ϑ_{16}	0	-0.0345	0.0345	0.0644
ϑ_{17}	1	0.9840	0.0160	0.0521
ϑ_{18}	1	1.0010	-0.0010	0.0314
ϑ_{19}	-2	-1.9841	-0.0159	0.0388
ϑ_{20}	0	0.0111	-0.0111	0.0347
ϑ_{21}	0.5310	0.5279	0.0031	0.0605
ϑ_{22}	-0.1934	-0.1870	-0.0064	0.0385
ϑ_{23}	0.1678	0.1678	0.0000	0.0467
ϑ_{24}	0.3784	0.3816	-0.0032	0.0293
ϑ_{25}	-0.2227	-0.2127	0.0100	0.0334
ϑ_{26}	0.5632	0.5585	0.0047	0.0476
ϑ_{27}	1	1.0002	0.0002	0.0030
ϑ_{28}	2	2.0000	0.0000	0.0079

Table 6.7.: Estimates for the parameters of a three-dimensional NIG-driven CARMA process

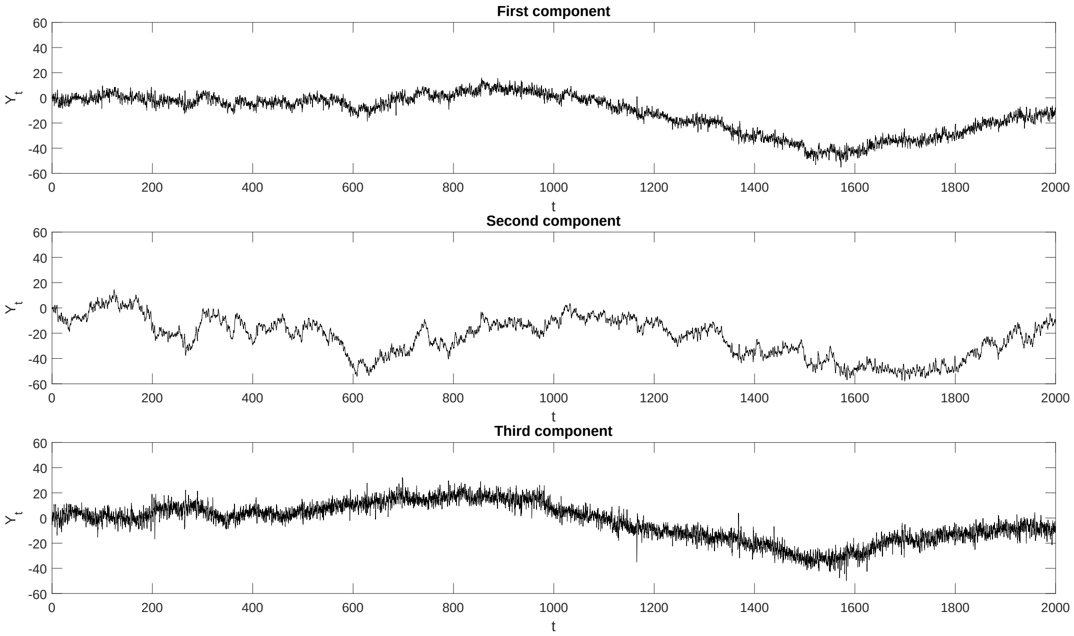

Figure 6.7.: Typical realization of 3-dimensional NIG-driven CARMA process

---

Case 2: Brownian Motion

Lastly, we have the Brownian Motion case for the three-dimensional model. The

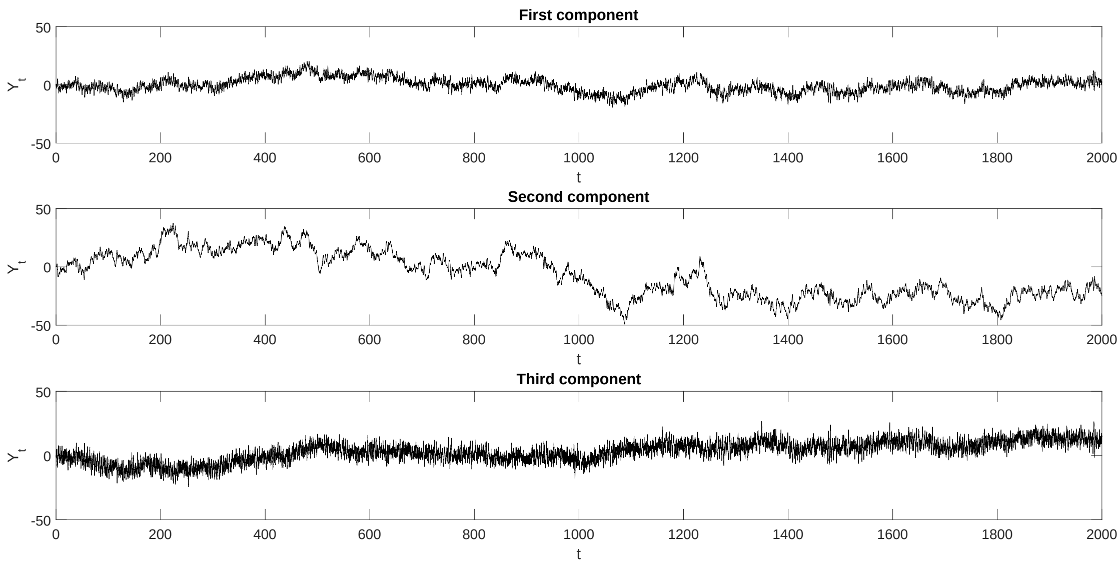

Figure 6.8.: Typical realization of 3-dimensional Brownian motion-driven CARMA process

covariance matrix is given as in (6.16) and thus coincides with the covariance matrix in the previous case.

	3-dim. BM-driven MCARMA process			
parameter	true param.	sample mean	bias	sample std. dev.
ϑ_1	-2	-1.9958	-0.0042	0.0475
ϑ_2	-3	-3.0005	0.0005	0.0339
ϑ_3	-3	-3.0309	0.0309	0.0401
ϑ_4	1	0.9987	0.0013	0.0381
ϑ_5	1	0.9895	0.0105	0.0316
ϑ_6	-1	-0.9763	-0.0237	0.0431
ϑ_7	2	2.0113	-0.0113	0.0342
ϑ_8	-1	-1.0075	0.0075	0.0399
ϑ_9	-3	-2.9896	-0.0104	0.0348
ϑ_{10}	-3	-2.9892	-0.0108	0.0444
ϑ_{11}	-1	-1.0097	0.0097	0.0461
ϑ_{12}	-1	-1.0242	0.0242	0.0367
ϑ_{13}	2	2.0077	-0.0077	0.0295
ϑ_{14}	1	0.9740	0.0260	0.0353
ϑ_{15}	1	1.0175	-0.0175	0.0284
ϑ_{16}	0	-0.0361	0.0361	0.0513
ϑ_{17}	1	0.9623	0.0377	0.0417
ϑ_{18}	1	0.9877	0.0123	0.0303
ϑ_{19}	-2	-1.9868	-0.0132	0.0306
ϑ_{20}	0	-0.0090	0.0090	0.0362
ϑ_{21}	0.5310	0.5849	-0.0539	0.0478
ϑ_{22}	-0.1934	-0.2037	0.0103	0.0328
ϑ_{23}	0.1678	0.1513	0.0165	0.0396
ϑ_{24}	0.3784	0.4209	-0.0425	0.0259
ϑ_{25}	-0.2227	-0.2209	-0.0018	0.0300
ϑ_{26}	0.5632	0.4814	0.0818	0.0356
ϑ_{27}	1	0.9995	0.0005	0.0033
ϑ_{28}	2	2.0004	-0.0004	0.0091

Table 6.8.: Estimates for the parameters of a three-dimensional MCARMA process driven by Brownian motion

The results for a sample of 350 replicates of the three-dimensional cointegrated MCARMA process driven by a Brownian motion are summarized in Table 6.8 and a realization can be found in Figure 6.8.

In conclusion, we see that in this case we have an excellent performance of the estimation procedure for both cases as well. We have comparable accuracy of the estimation results in terms of the bias and standard deviation of the sample. In this model we had two common stochastic trends. The bias and standard deviation of the long-run parameters ($\vartheta_{27}, \vartheta_{28}$) are also in this framework the lowest. In this respect,

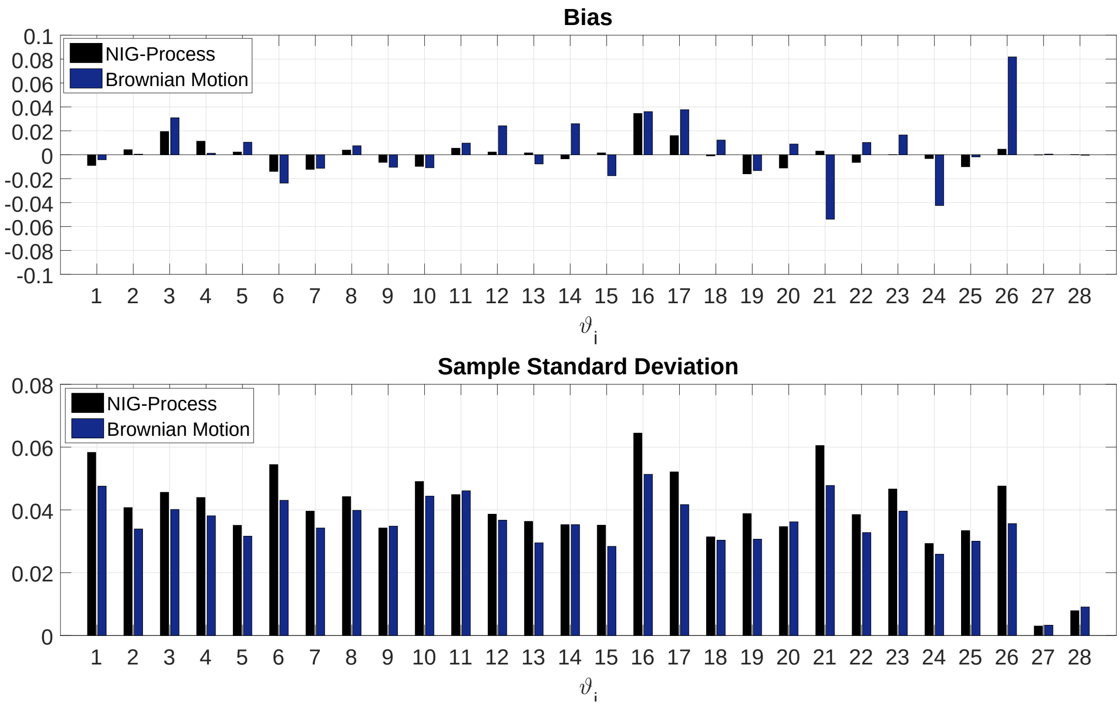

Figure 6.9.: Bias and Sample Standard Deviation of both Simulation Studies

there are no noteworthy differences in the simulation results between the four cases. The bias and standard deviation is again visualized in Figure 6.9.

Appendix

A. Summary of Assumptions

In this part of the appendix we sum up all the assumptions of each chapter in order to give an overview.

A.1. Assumptions in Chapter 3

Assumption A1
The Lévy process L satisfies $\mathbb{E}L(1) = 0$ and $\mathbb{E}\|L(1)\|^2 < \infty$.

Assumption A2
The eigenvalues of $\mathcal{A}$ in equation (2.15) and consequently of $A \in M_N(\mathbb{C})$ in (2.2), have strictly negative real parts, where $N = pm$.

Assumption A3
The eigenvalues $\lambda_1, \ldots, \lambda_N$ of $\mathcal{A}$ in equation (2.15) and consequently of $A \in M_N(\mathbb{C})$ in (2.2), are distinct, where the dimensions satisfies $N = pm$.

A.2. Assumptions in Chapter 4

Assumption C1
The Lévy process L satisfies $\mathbb{E}L(1) = 0_d$ and $\mathbb{E}\|L(1)\|^2 < \infty$.

Assumption C2
The linear system given in (4.36) is controllable, i.e.

$$\mathcal{C}_{\mathrm{e}^{Ah}K^{(h)}} := \left(K^{(h)} \quad \mathrm{e}^{Ah}K^{(h)} \quad \ldots \quad \left(\mathrm{e}^{Ah}\right)^{N-1}K^{(h)}\right) \in M_{N,pN}(\mathbb{R}) \tag{4.39}$$

has rank N.

Assumption K1

The initial state X_0 is independent of Z and W, i.e. $\mathbb{E}(X_0 Z_n^\mathsf{T}) = 0$ and $\mathbb{E}(X_0 W_n^\mathsf{T}) = 0$ for all n.

Assumption K2

Let $R \equiv 0_{d\times d}$ and $\Gamma Q \Gamma^\mathsf{T} \in \mathbb{S}_N^{++}(\mathbb{R})$, i.e. $\Gamma Q \Gamma^\mathsf{T}$ is positive-definite.

Assumption K3

Let H be of full rank and Ω_0 be positive definite.

A.3. Assumptions in Chapter 5

Assumption M1

Assume that the cointegrated MCARMA process is driven by a Lévy process L_ϑ with mean zero and non-singular covariance matrix $\Sigma_\vartheta^L = \mathbb{E}L_\vartheta(1)L_\vartheta(1)^\mathsf{T}$. Assume further that there exists a $\delta > 0$ such that $\mathbb{E}\|L_\vartheta(1)\|^{4+\delta} < \infty$.

Assumption M2

Assume that the matrix A_ϑ has c eigenvalues equal to zero and the remaining eigenvalues have strictly negative real parts for all $\vartheta \in \Theta$. Moreover, the matrix C_ϑ has full rank for all $\vartheta \in \Theta$.

Assumption M3

The triplet $(A_\vartheta, B_\vartheta, C_\vartheta)$ is minimal for all $\vartheta \in \Theta$ with McMillan degree N.

Assumption M4

The matrices $B_{1,\vartheta}$ and $C_{1,\vartheta}$ have full rank c for all $\vartheta \in \Theta$.

Assumption M5

The parameter space Θ is a compact subset of $\mathbb{R}^s$.

Assumption M6

The mappings $\vartheta \mapsto A_{2,\vartheta}$, $\vartheta \mapsto B_{i,\vartheta}$, $\vartheta \mapsto C_{i,\vartheta}$ for $i \in \{1,2\}$ and $\vartheta \mapsto \Sigma_\vartheta^L$ are continuous.

Assumption M7

We assume that the true parameter vector ϑ^0 lies in the interior of the parameter space Θ.

Assumption M8

Let the functions $\vartheta \mapsto A_{2,\vartheta}$, $\vartheta \mapsto B_{i,\vartheta}$, $\vartheta \mapsto C_{i,\vartheta}$ for $i \in \{1,2\}$ and $\vartheta \mapsto \Sigma_\vartheta^L$ be twice continuously differentiable.

Assumption M9
The matrices C_{1,ϑ_1} and $C^{\perp}_{1,\vartheta_1} = \beta(\vartheta_1)$ are positive lower triangular matrices for all $\vartheta_1 \in \Theta_1$ as in Theorem 4.3.7 satisfying $C^{\mathsf{T}}_{1,\vartheta_1} C_{1,\vartheta_1} = I_c$ and $C^{\perp\mathsf{T}}_{1,\vartheta_1} C^{\perp}_{1,\vartheta_1} = I_{d-c}$.

Assumption M10
Assume that the collection of the stationary parts of the output processes, denoted by

$$K(\Theta_2) := (Y_2(\vartheta_2), \vartheta_2 \in \Theta_2),$$

corresponding to the linear state space model $(A_{2,\vartheta_2}, B_{2,\vartheta_2}, C_{2,\vartheta_2}, L_{\vartheta_2})$ is identifiable from the spectral density.

Assumption M11
For all $\vartheta_2 \in \Theta_2$ the spectrum of A_{2,ϑ_2} is a subset of $\{z \in \mathbb{C} : |\Im z| \leq \frac{\pi}{h}\}$.

Assumption M12
Assume that the $(d^2 \times s_1)$- dimensional gradient matrix $\nabla_{\vartheta_1} \left(\Pi(\vartheta_1^0, \vartheta_2)^{\mathsf{T}}\right)$ has full column rank s_1 for all $\vartheta_2 \in \Theta_2$.

Assumption M13
Assume that there exists a positive index j_0 such that the $[(j_0 + 2)d^2 \times s_2]$ matrix $\nabla_{\vartheta_2} \psi_{\vartheta^0, j_0}$ has rank s_2.

B. Collection of Matrix Formulas

In this part of the appendix we sum up several basic formulas, facts and calculation rules in the field of matrix theory. This section ought to give an overview of formulas which are used several times throughout this thesis.

B.1. Facts on the Trace Operator and Frobenius Norm

For a reference on the properties of trace and Frobenius norm we refer to Bernstein [9]. Let $a, b \in \mathbb{R}$, $A \in M_{m,n}(\mathbb{R})$, $B \in M_{n,m}(\mathbb{R})$ and $C \in M_{m,k}$. Then the following formulas hold:

(i) $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ ((2.2.25), [9]);

(ii) $\operatorname{tr}(aA + bB) = a\operatorname{tr}(A) + b\operatorname{tr}(B)$ ((2.2.29), [9]);

(iii) Let A and B be symmetric then we have $\operatorname{tr}(AB) \leq |\operatorname{tr}(AB)| \leq \frac{1}{2}\operatorname{tr}(A^2 + B^2)$ (Fact 8.12.8, [9]);

(iv) $\|A\|_F = \|\mathrm{vec}A\|_F$ ((9.2.6), [9]);

(v) $\|AB\|_F \leq \begin{Bmatrix} \sigma_{max}(A)\|B\|_F \\ \sigma_{max}(B)\|A\|_F \end{Bmatrix} \leq \|A\|_F\|B\|_F$ (Corollary 9.3.7, [9]);

(vi) If $m \leq n$ then $\sigma_{min}(A)\|C\|_F \leq \|AC\|_F$ and if $m \leq k$ then $\sigma_{min}(C)\|A\|_F \leq \|AC\|_F$ (Corollary 9.6.7, [9]);

(vii) $\sigma_{max}(A) \leq \|A\|_F$ (Fact 9.8.12, [9]);

(viii) $\|A \otimes B\|_F = \|A\|_F\|B\|_F$ (Fact 9.14.37, [9]);

(ix) $\sigma_{max} = \sigma_1 \geq \cdots \geq \sigma_{min}(A) = \sigma_{\min\{n,m\}} > 0$ ((5.6.4), [9]).

B.2. Facts on Partitioned Matrices

For a reference on the properties of partitioned matrices we refer to Bernstein [9].

Proposition B.1 (Proposition 2.8.3., Bernstein [9])
Let $A \in M_{n,n}(\mathbb{R})$, $B \in M_{n,m}(\mathbb{R})$, $C \in M_{l,n}$, $D \in M_{l,m}$ and A is invertible. Then we have

$$rank \begin{pmatrix} A & B \\ C & D \end{pmatrix} = n + rank(D - CA^{-1}B) \tag{B.1}$$

and

$$\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(A)\det(D - CA^{-1}B). \tag{B.2}$$

Proposition B.2 (Proposition 2.8.7., Bernstein [9])
Let $A \in M_{n,n}(\mathbb{R})$, $B \in M_{n,m}(\mathbb{R})$, $C \in M_{m,n}$, $D \in M_{m,m}$. If A and $M := D - CA^{-1}B$ are invertible then

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1}BM^{-1}CA^{-1} & -A^{-1}BM^{-1} \\ -M^{-1}CA^{-1} & M^{-1} \end{pmatrix}. \tag{B.3}$$

Proposition B.3 (Corollary 2.8.8., Bernstein [9])
Let $A \in M_{n,n}(\mathbb{R})$, $B \in M_{n,m}(\mathbb{R})$, $C \in M_{m,n}$, $D \in M_{m,m}$. If A, $D - CA^{-1}B$ and D are invertible then

$$(A - BD^{-1}C)^{-1} = A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1}. \tag{B.4}$$

B.3. Vec Operator and Kronecker Product

For a reference on the properties of the vec operator and the Kronecker product see for example Chapter 7.1 in Bernstein [9] or Lütkepohl [62], Appendix A.11 and A.12.

Rules for the Kronecker product and Vectorization operator:
Let the matrices A, B, C, D have appropriate dimensions.

(i) $A \otimes (B + C) = A \otimes B + A \otimes C$;

(ii) $(A \otimes B)(C \otimes D) = AC \otimes BD$;

(iii) $(A \otimes B)^\mathsf{T} = A^\mathsf{T} \otimes B^\mathsf{T}$;

(iv) If A and B are invertible then $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$;

(v) If A and B are square matrices then $\operatorname{tr}(A \otimes B) = \operatorname{tr}(A)\operatorname{tr}(B)$;

(vi) $\operatorname{vec}(A + B) = \operatorname{vec}(A) + \operatorname{vec}(B)$;

(vii) $\operatorname{vec}(ABC) = (C^\mathsf{T} \otimes A)\operatorname{vec}(B) = (I \otimes AB)\operatorname{vec}(C) = (C^\mathsf{T}B^\mathsf{T} \otimes I)\operatorname{vec}(A)$;

(viii) $\operatorname{vec}(A + B) = \operatorname{vec}(A) + \operatorname{vec}(B)$;

(ix) $\operatorname{vec}(B^\mathsf{T})^\mathsf{T}\operatorname{vec}(A) = \operatorname{tr}(AB) = tr(BA) = \operatorname{vec}(A^\mathsf{T})^\mathsf{T}\operatorname{vec}(B)$.

B.4. Matrix Differential Calculus

For differentiation rules and helpful formulas of matrix differential calculus see e.g. the books by Horn and Johnson [49], Section 6.5 and 6.6, Magnus and Neudecker [65] or Abadir and Magnus [1].

Proposition B.4 (Chain Rule (Horn and Johnson, Corollary 6.6.19))
Assume that $g : \mathbb{R} \to M_{m,n}(\mathbb{R})$ and $h : M_{m,n}(\mathbb{R}) \to \mathbb{R}$ are differentiable functions. Then we have

$$\frac{\partial}{\partial x} h(g(x)) = tr\left(\left[\frac{\partial}{\partial M^{\mathsf{T}}} h(M)\Big|_{M=g(x)}\right]\left(\frac{\partial}{\partial x} g(x)\right)\right), \tag{B.5}$$

where $M = [m_{i,j}]_{i,j} \in M_{m,n}(\mathbb{R})$ and $x \in \mathbb{R}$. We denote

$$\frac{\partial}{\partial M^{\mathsf{T}}} h(M) = \left[\frac{\partial}{\partial m_{i,j}} h(M)\right]_{i,j} \in M_{n,m}(\mathbb{R}) \tag{B.6}$$

and

$$\frac{\partial}{\partial x} g(x) = \left[\frac{\partial}{\partial x} \left[g(x)\right]_{i,j}\right]_{i,j} \in M_{n,m}(\mathbb{R}). \tag{B.7}$$

Proposition B.5 (Differentiation Rules)
Let $M \in M_m(\mathbb{R})$ be invertible and $A \in M_{n,m}(\mathbb{R})$ and $B \in M_{m,k}(\mathbb{R})$. Furthermore, let $v \in \mathbb{R}^d$, $C(v) \in M_m(\mathbb{R})$ and $D(v) \in M_{k,r}(\mathbb{R})$ be differentiable functions, then we have the following formulas:

(i) $\frac{\partial}{\partial M^{\mathsf{T}}} tr(AMB) = BA$;

(ii) $\frac{\partial}{\partial M^{\mathsf{T}}} \log|\det M| = M^{-1}$;

(iii) $\frac{\partial}{\partial M^{\mathsf{T}}} tr(AM^{-1}B) = -M^{-1}BAM^{-1}$;

(iv) $\frac{\partial}{\partial v^{\mathsf{T}}} vec(AC(v)B) = (B^{\mathsf{T}} \otimes A)\frac{\partial vec(C(v))}{\partial v^{\mathsf{T}}}$;

(v) $\frac{\partial}{\partial v^{\mathsf{T}}} vec(C(v)BD(v)) = (I_r \otimes C(v)B)\frac{\partial vec(D(v))}{\partial v^{\mathsf{T}}} + (D(v)^{\mathsf{T}}B^{\mathsf{T}} \otimes I_m)\frac{\partial vec(C(v))}{\partial v^{\mathsf{T}}}$.

For proofs see e.g. Lütkepohl [62], Appendix A.13 (6), (7), (13), (15) and (17) or Horn and Johnson [49], Section 6.5.

Bibliography

[1] Abadir, K. M. & Magnus, J. R. (2005) *Matrix algebra*, vol. 1. Cambridge Univ. Press, Cambridge.

[2] Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. *Second International Symposium on Information Theory* 267–281.

[3] Anderson, B. D. O. & Moore, J. B. (1979) *Optimal filtering*. Prentice-Hall information and system sciences series, Prentice-Hall, Englewood Cliffs.

[4] Applebaum, D. (2004) *Lévy processes and stochastic calculus*. 1. publ., Cambridge Stud. Adv. Math. 93, Cambridge Univ. Press, Cambridge.

[5] Barndorff-Nielsen, O. E. (1997) Normal inverse Gaussian distributions and stochastic volatility modelling. *Scand. J. Stat.* **24** (1), 1–13.

[6] Basawa, I. V. & Scott, D. J. (1983) *Asymptotic optimal inference for non-ergodic models*. Lecture Notes in Statist. 17, Springer, New York.

[7] Bauer, D. & Wagner, M. (2002) Asymptotic properties of pseudo maximum likelihood estimates for multiple frequency I(1) processes. Diskussionsschriften, Universitaet Bern, Departement Volkswirtschaft.

[8] Bauer, D. & Wagner, M. (2012) A state space canonical form for unit root processes. *Econometric Theory* **28** (6), 1313–1349.

[9] Bernstein, D. S. (2009) *Matrix mathematics: theory, facts, and formulas*. 2nd ed., Princeton Univ. Press, Princeton.

[10] Bertoin, J. (2005) *Lévy processes*. 1st ed., Cambridge Tracts in Math. 121, Cambridge Univ. Press, Cambridge.

[11] Billingsley, P. (1995) *Probability and measure*. 3rd ed., Wiley Ser. Probab. Stat., Wiley, New York.

[12] Billingsley, P. (1999) *Convergence of probability measures*. 2nd ed., Wiley Ser. Probab. Stat., Wiley, New York.

[13] Boubacar Maïnassara, Y. (2012) Selection of weak VARMA models by modified akaike's information criteria. *J. Time Ser. Anal.* **33** (1), 121–130.

[14] Boubacar Maïnassara, Y. & Francq, C. (2011) Estimating structural VARMA models with uncorrelated but non-independent error terms. *J. Multivar. Anal.* **102** (3), 496–505.

[15] Box, G. E. P., Jenkins, G. M. & Reinsel, G. C. (1994) *Time series analysis: forecasting and control*. 3rd ed., Prentice Hall, Englewood Cliffs.

[16] Bradley, R. C. (2007) *Introduction to strong mixing conditions*, vol. 1. Kendrick, Heber City.

[17] Brenner, R. J. & Kroner, K. F. (1995) Arbitrage, cointegration, and testing the unbiasedness hypothesis in financial markets. *J. Financ. Quant. Anal.* **30** (01), 23–42.

[18] Brockwell, P. J. (2001) Lévy-driven CARMA processes. *Ann. Inst. Statist. Math.* **53** (1), 113–124.

[19] Brockwell, P. J. (2014) Recent results in the theory and applications of CARMA processes. *Ann. Inst. Statist. Math.* **66** (4), 647–685.

[20] Brockwell, P. J. & Davis, R. A. (1998) *Time Series: Theory and Methods*. 2nd ed., Springer Ser. Statist., Springer, New York.

[21] Brockwell, P. J., Davis, R. A. & Yang, Y. (2011) Estimation for non-negative Lévy-driven CARMA processes. *J. Bus. Econom. Statist.* **29** (2), 250–259.

[22] Brockwell, P. J. & Lindner, A. (2009) Existence and uniqueness of stationary Lévy-driven CARMA processes. *Stochastic Process. Appl.* **119** (8), 2660–2681.

[23] Brockwell, P. J. & Lindner, A. (2013) Integration of CARMA processes and spot volatility modelling. *J. Time Ser. Anal.* **34** (2), 156–167.

[24] Brockwell, P. J. & Schlemm, E. (2013) Parametric estimation of the driving Lévy process of multivariate CARMA processes from discrete observations. *J. Multivar. Anal.* **115**, 217–251.

[25] Chambers, M. J. (1999) Discrete time representation of stationary and non-stationary continuous time systems. *J. Econ. Dyn. Control* **23** (4), 619–639.

[26] Chui, C. K. & Chen, G. (2009) *Kalman filtering: with real-time applications*. 4th ed., Springer, Berlin.

[27] Comte, F. (1999) Discrete and continuous time cointegration. *J. Econometrics* **88** (2), 207–226.

[28] Dennis, J., Traub, J. & Weber, R. (1976) The Algebraic Theory of Matrix Polynomials. *SIAM J. Numer. Anal.* **13** (6), 831–845.

[29] Dennis, J., Jr., Traub, J. & Weber, R. (1971) On the matrix polynomial, lambda-matrix and block eigenvalue problems. *Computer Science Department tech. rep., Cornell University, Ithaca and Carnegie-Mellon University, Pittsburgh* .

[30] Doob, J. L. (1944) The elementary Gaussian processes. *Ann. Math. Stat.* **15** (3), 229–282.

[31] Doukhan, P. (1994) *Mixing: properties and examples*. Lecture Notes in Statist. 85, Springer, New York.

[32] Durrett, R. (2010) *Probability: theory and examples*. 4th ed., Camb. Ser. Stat. Probab. Math. 31, Cambridge Univ. Press, Cambridge.

[33] Engle, R. F. & Granger, C. W. J. (1987) Co-integration and error correction: Representation, estimation, and testing. *Econometrica* **55** (2), 251–276.

[34] Fasen, V. (2013) Time series regression on integrated continuous-time processes with heavy and light tails. *Econometric Theory* **29**, 28–67.

[35] Fasen, V. (2014) Limit theory for high frequency sampled MCARMA models. *Adv. Appl. Probab.* **46**, 846–877.

[36] Fasen, V. & Fuchs, F. (2013) On the limit behavior of the periodogram of high-frequency sampled stable CARMA processes. *Stochastic Process. Appl.* **123** (1), 229–273.

[37] Fasen, V. & Fuchs, F. (2013) Spectral estimates for high-frequency sampled continuous-time autoregressive moving average processes. *J. Time Ser. Anal.* **34** (5), 532–551.

[38] Fasen, V. & Kimmig, S. (2016+) Information criteria for multivariate CARMA processes. *Accepted for publication in Bernoulli* .

[39] Francq, C. & Zakoïan, J.-M. (1998) Estimating linear representations of nonlinear processes. *J. Statist. Plann. Inference* **68** (1), 145–165.

[40] Gantmacher, F. R. (1986) *Matrizentheorie*. Springer, Berlin , Heidelberg [u.a.].

[41] Gohberg, Y. Z., Lancaster, P. & Rodman, L. (1982) *Matrix polynomials*. Computer science and applied mathematics, Acad. Pr., New York.

[42] Granger, C. (1981) Some properties of time series data and their use in econometric model specification. *J. Econometrics* **16** (1), 121–130.

[43] Granger, C. & Newbold, P. (1974) Spurious regressions in econometrics. *J. Econometrics* **2** (2), 111 – 120.

[44] Guidorzi, R. (1975) Canonical structures in the identification of multivariable systems. *Automatica* **11** (4), 361–374.

[45] Hamilton, J. D. (1994) *Time series analysis*. Princeton Univ. Press, Princeton.

[46] Hannan, E. & Deistler, M. (2012) *The Statistical Theory of Linear Systems*. Classics Appl.Math., Society for Industrial and Applied Mathematics.

[47] Hansen, L. P. & Sargent, T. J. (1983) The dimensionality of the aliasing problem in models with rational spectral densities. *Econometrica* **51** (2), 377–387.

[48] Harvey, A. C. (1989) *Forecasting, structural time series models and the Kalman Filter*. Cambridge Univ. Press, Cambridge.

[49] Horn, R. A. & Johnson, C. R. (1994) *Topics in Matrix Analysis*. Cambridge Univ. Press, Cambridge.

[50] Horn, R. A. & Johnson, C. R. (2013) *Matrix analysis*. 2nd ed., Cambridge Univ. Press, Cambridge.

[51] Ibragimov, I. A. (1962) Some limit theorems for stationary processes. *Theory Probab. Appl.* **7** (4), 349–382.

[52] Johansen, S. (1988) Statistical analysis of cointegration vectors. *J. Econ. Dyn. Control* **12** (2), 231–254.

[53] Johansen, S. (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* **59** (6), 1551–1580.

[54] Johansen, S. (2009) *Likelihood-based inference in cointegrated vector autoregressive models*. Advanced texts in econometrics, Oxford Univ. Press, Oxford.

[55] Johansen, S. & Schaumburg, E. (1999) Likelihood analysis of seasonal cointegration. *J. Econometrics* **88** (2), 301–339.

[56] Kalman, R. E. (1960) A new approach to linear filtering and prediction problems. *J. Basic Eng.* (82), 35–45.

[57] Kessler, M. & Rahbek, A. (2001) Asymptotic likelihood based inference for cointegrated homogenous Gaussian diffusions. *Scand. J. Statist.* **28** (3), 455–470.

[58] Kessler, M. & Rahbek, A. (2004) Identification and inference for multivariate cointegrated and ergodic Gaussian diffusions. *Stat. Inference Stoch. Process.* **7** (2), 137–151.

[59] Krengel, U. (1985) *Ergodic theorems: with a supplement*. de Gruyter Stud. Math. 6, De Gruyter, Berlin.

[60] Lax, P. (2002) *Functional analysis*. Pure and applied mathematics, Wiley, New York.

[61] Levya-Ramos, J. (1991) Partial-fraction expansion in system analysis. *Internat. J. Control* **53** (3), 619–639.

[62] Lütkepohl, H. (2005) *New introduction to multiple time series analysis*. Springer, Berlin.

[63] Lütkepohl, H. & Claessen, H. (1997) Analysis of cointegrated VARMA processes. *J. Econometrics* **80** (2), 223–239.

[64] Lütkepohl, H. & Poskitt, D. S. (1996) Specification of Echelon-form VARMA models. *J. Bus. Econom. Statist.* **14** (1), 69–79.

[65] Magnus, J. R. & Neudecker, H. (2002) *Matrix differential calculus with applications in statistics and econometrics*. Wiley Ser. Probab. Stat., John Wiley, New York.

[66] Markus, A. S. & Mereuca, I. V. (1973) On the Complete n-Tuple of Roots of the Operator Equation Corresponding to a Polynomial Operator Bundle. *Mathematics of the USSR-Izvestiya* **7** (5), 1108–1131.

[67] Maroulas, J. (1985) Factorization of matrix polynomials with multiple roots. *Linear Algebra Appl.* **69**, 9–32.

[68] Marquardt, T. (2007) Multivariate fractionally integrated CARMA processes. *J. Multivar. Anal.* **98** (9), 1705–1725.

[69] Marquardt, T. & Stelzer, R. (2007) Multivariate CARMA processes. *Stochastic Process. Appl.* **117** (1), 96–120.

[70] Nelson, C. R. & Plosser, C. R. (1982) Trends and random walks in macroeconmic time series: some evidence and implications. *Journal of monetary economics* **10** (2), 139–162.

[71] Newey, W. K. (1991) Uniform convergence in probability and stochastic equicontinuity. *Econometrica* **59** (4), 1161–1167.

[72] Park, J. Y. & Phillips, P. C. B. (1988) Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory* **4** (3), 468–497.

[73] Park, J. Y. & Phillips, P. C. B. (1989) Statistical inference in regressions with integrated processes: Part 2. *Econometric Theory* **5** (1), 95–131.

[74] Phillips, P. C. B. (1973) The problem of identification in finite parameter continuous time models. *J. Econometrics* **1** (4), 351–362.

[75] Phillips, P. C. B. (1987) Weak convergence of sample covariance matrices to stochastic integrals via martingale approximations. Cowles Foundation Discussion Papers 846, Cowles Foundation for Research in Economics, Yale University.

[76] Phillips, P. C. B. (1988) Weak convergence to the matrix stochastic integral $\int_0^1 B\,dB'$. *J. Multivar. Anal.* **24** (2), 252–264.

[77] Phillips, P. C. B. (1991) Error correction and long-run equilibrium in continuous time. *Econometrica* **59** (4), 967–980.

[78] Phillips, P. C. B. & Durlauf, S. N. (1986) Multiple time series regression with integrated processes. *Rev. Econ. Stud.* **53** (4), 473–495.

[79] Phillips, P. C. B. & Solo, V. (1992) Asymptotics for linear processes. *Ann. Statist.* **20** (2), 971–1001.

[80] Pötscher, B. M. & Prucha, I. R. (1991) Basic structure of the asymptotic theory in dynamic nonlinear econometric models, Part 1: consistency and approximation concepts. *Econometric Reviews* **10** (2), 125–216.

[81] Pötscher, B. M. & Prucha, I. R. (1997) Basic structure of the asymptotic theory in dynamic nonlinear econometric models, Part 2: Asymptotic normality. *Nonlinear Models* **1**, 425–497.

[82] Protter, P. E. (1990) *Stochastic Integration and Differential Equations: A new Approach*. Appl. Math. 21, Springer, New York.

[83] Reinsel, G. (2003) *Elements of Multivariate Time Series Analysis*. Springer Ser. Statist., Springer, New York.

[84] Ribarits, T. & Hanzon, B. (2011) The state-space error correction model: Definiton, estiamtion and model selection. *unpublished manuscript received from the authors upon request* .

[85] 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.

[86] Saikkonen, P. (1995) Problems with the asymptotic theory of maximum likelihood estimation in integrated and cointegrated systems. *Econometric Theory* **11**, 888–911.

[87] Sargan, J. D. (1964) Wages and prices in the united kingdom: a study in econometric methodology. *Econometric analysis for national economic planning* **16**, 25–54.

[88] Sato, K.-I. (1999) *Lévy processes and infinitely divisible distributions*. Cambridge Stud. Adv. Math. 68, Cambridge Univ. Press, Cambridge.

[89] Schlemm, E. (2011) *Estimation of Continuous-Time ARMA Models and Random Matrices with Dependent Entries*. Dissertation, Technische Universität München, München.

[90] Schlemm, E. & Stelzer, R. (2012) Multivariate CARMA processes, continuous-time state space models and complete regularity of the innovations of the sampled processes. *Bernoulli* **18** (1), 46–63.

[91] Schlemm, E. & Stelzer, R. (2012) Quasi maximum likelihood estimation for strongly mixing state space models and multivariate Lévy-driven CARMA processes. *Electron. J. Stat.* **6**, 2185–2234.

[92] Schwarz, G. (1978) Estimating the dimension of a model. *Ann. Statist.* **6** (2), 461–464.

[93] Shieh, L. S., Chang, F. & Mcinnis, B. C. (1986) The block partial fraction expansion of a matrix fraction description with repeated block poles. *IEEE Trans. Automat. Control* **31** (3), 236–239.

[94] Stewart, G. W. & Sun, J. (1990) *Matrix perturbation theory*. Computer science and scientific computing, Acad. Pr., Boston, Mass.

[95] Sun, J.-G. (1998) Sensitivity analysis of the discrete-time algebraic riccati equation. *Linear Alg. Appl.* **275/276**, 595 – 615.

[96] Sweeting, T. J. (1980) Uniform asymptotic normality of the maximum likelihood estimator. *Ann. Statist.* **8** (6), 1375–1381.

[97] Todorov, V. & Tauchen, G. (2006) Simulation methods for Lévy-driven continuous-time autoregressive moving average (CARMA) stochastic volatility models. *J. Bus. Econom. Statist* **24** (4), 455–469.

[98] Tsay, Y. T. & Shieh, L. S. (1982) Some applications of rational matrices to problems in systems theory. *Internat. J. Systems Sci.* **13** (12), 1319–1337.

[99] Yap, S. F. & Reinsel, G. C. (1995) Estimation and testing for unit roots in a partially nonstationary vector autoregressive moving average model. *J. Amer. Statist. Assoc.* **90** (429), 253–267.

Index

List of Abbreviations

$A := B$	A is defined by B
$[a, b], (a, b), (a, b], [a, b)$	closed, open, half-open interval from a to b
$\mathbb{N}, \mathbb{N}_0$	$\{1, 2, \ldots\}, \{0, 1, 2, \ldots\}$
$\mathbb{Z}$	$\{\ldots, -2, -1, 0, 1, 2, \ldots\}$
$\mathbb{R}, \mathbb{R}_+, \mathbb{R}_-, \bar{\mathbb{R}}$	$(-\infty, \infty), [0, \infty), (-\infty, 0], [-\infty, \infty]$
$\mathbb{C}$	complex numbers
$\Re(z), \Im(z)$	real and imaginary part of $z \in \mathbb{C}$
S^n	unit n-sphere in $\mathbb{R}^{n+1}$
I_m	$m \times m$ identity matrix
0_m	$m \times m$ zero matrix
e_i	i-th unit vector
$\lambda_{\min}(A)$	smallest eigenvalue of the matrix A
$\lambda_{\max}(A)$	largest eigenvalue of the matrix A
$\sigma_{\min}(A)$	smallest singular value of the matrix A
$\sigma_{\max}(A)$	largest singular value of the matrix A
$M_{m,n}(\mathbb{R})$	space of all $m \times n$ real-valued matrices
$M_{m,n}(\mathbb{C})$	space of all $m \times n$ complex-valued matrices
M_m	space of all $m \times m$ matrices
$\mathbb{S}_N$	set of N-dimensional symmetric matrices
$\mathbb{S}_N^+$	set of N-dimensional symmetric positive-semidefinite matrices
$\mathbb{S}_N^{++}$	set of N-dimensional symmetric positive-definite matrices
$GL_m(\cdot)$	invertible $m \times m$ matrices

$A \geq 0$	A positive-semidefinite
$A > 0$	A positive-definite
$A \geq B$	$A - B$ positive-semidefinite
$A > B$	$A - B$ positive-definite
A^{-1}	inverse of the matrix A
A^{T}	transposed of the matrix A
A^*	complex conjugate of the matrix A
$\mathrm{adj}(A)$	adjugate of the matrix A
$\det(A)$	determinant of the matrix A
$\mathrm{tr}(A)$	trace of the matrix A
$\sigma(A)$	spectrum of the matrix A
$A \subset B$	A is contained in B or $A = B$
$\ker A$	kernel of A
$\mathrm{im}\ A$	image of A
$A^{\perp}$	If A is a $n \times p$ matrix with rank q then $A^{\perp}$ is $n \times (n-q)$ with rank $n-q$, lies in the null space of A and $A^{\perp T} A = 0_{(n-q)\times q}$ and $A^T A^{\perp} = 0_{q\times(n-q)}$
P_A	projection matrix onto the subspace spanned by the columns of the matrix A
$\mathrm{Diag}(i_1, \ldots, i_m)$	diagonal matrix with entries $i_1, \ldots, i_,$
M/N	is the set of all members of M which are not members of N
$\mathrm{vec}(A)$	is an operator which converts the matrix A into a column vector
$\mathrm{vech}(A)$	is an operator which converts the symmetric matrix A into a column vector by vectorizing only the lower triangular part of A
$A \otimes B$	is the Kronecker product of A and B
$a \vee b, a \wedge b$	maximum, minimum of a and b
a^+	$a^+ := 0 \vee a$
a^-	$a^- := 0 \wedge a$
$\log, \exp$	natural logarithm, exponential function
$\log^+(a)$	$\log^+(a) := \log(a \vee 1)$

$\lfloor x \rfloor$	is the largest integer less than or equal to x
$\lceil x \rceil$	is the largest integer greater than or equal to x
$\delta_{i,j}$	$\delta_{i,j} = 1$ for $i = j$ and 0 else
$\mathbb{1}_A$	indicator function of the set A
$\emptyset$	empty set
$\langle x, w \rangle$	inner product $x, w \in \mathbb{R}^m$
$\|x\|$	Euclidean norm of $x \in \mathbb{R}^m$
$\|A\|$	Frobenius norm of $A \in M_m$, i.e. $\|A\| = \|A\|_F = \sqrt{\operatorname{tr} A^\intercal A}$
$\langle A, B \rangle$	Frobenius inner product, i.e. $\langle A, B \rangle = \operatorname{tr}(A^\intercal B)$
$\angle(u, v)$	angle between the vectors u and v
$\mathcal{B}(\mathbb{R})$	Borel-σ-algebra over $\mathbb{R}$
L^p	Lebesgue spaces
$\mathbb{P}, \mathbb{E}, \mathrm{Var}, \mathrm{Cov}$	probability measure, expected value, variance and covariance
$X \overset{d}{=} Y$	the distribution of X coincides with the distribution of Y
$\xrightarrow{p}$	convergence in probability
$\xrightarrow{a.s.}$	almost sure convergence
$\xrightarrow{w}$	weak convergence
$\xrightarrow{L^2}$	convergence in L^2
$\xrightarrow[c]{w}$	continuous weak convergence
$\xrightarrow[c]{p}$	continuous convergence in probability
$\mathcal{N}(0, 1)$	standard normal distribution
$\mathcal{N}(\mu, \Sigma)$	normal distribution with mean μ and variance Σ
$B = \{B(t)\}_{t \geq 0}$	Brownian motion
D	Differential operator
$L = \{L(t)\}_{t \geq 0}$	Lévyprocess
ARMA	Autoregressive Moving Average Process
VARMA	Vector Autoregressive Moving Average Process
MA	Moving Average Process
VAR	Vector Autoregressive Process
ARIMA	Integrated Autoregressive Moving Average Process

CARMA	Continuous-Time Autoregressive Moving Average Process
ICARMA	Integrated Continuous-Time Autoregressive Moving Average Process
MCARMA	Multivariate Continuous-Time Autoregressive Moving Average Process
MICARMA	Multivariate Integrated Continuous-Time Autoregressive Moving Average Process
DARE	Discrete Algebraic Riccati Equation
QMLE	Quasi-Maximum Likelihood Filter
SE	Stochastic Equicontinuity Condition
TFECM	Transfer Function Error Correction Model