Estimation of Continuous-Time ARMA Models and Random Matrices with Dependent Entries

Eckhard Schlemm

Dissertation

Fakultät für Mathematik

Technische Universität München

D-85748 Garching

Juni 2011

TECHNISCHE UNIVERSITÄT MÜNCHEN
Lehrstuhl für Mathematische Statistik

Estimation of Continuous-Time ARMA Models and Random Matrices with Dependent Entries

Eckhard Schlemm

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende:	Univ.-Prof. Dr. Nina Gantert
Prüfer der Dissertation:	1. Univ.-Prof. Dr. Robert Stelzer Universität Ulm
	2. Prof. Víctor Manuel Pérez Abreu Carrión, Ph. D. Universidad de Guanajuato & CIMAT, Mexico
	3. Univ.-Prof. Dr. Thomas Klein (nur mündliche Prüfung)

Die Dissertation wurde am 16.06.2011 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 19.09.2011 angenommen.

My surface is myself.
Under which
to witness, youth is
buried. Roots?
Everybody has roots.

William Carlos Williams, Paterson

Zusammenfassung

In dieser Arbeit werden statistische Fragen für lineare stochastische Prozesse untersucht. Konsistenz und asymptotische Normalität des Quasi-Maximum-Likelihood Schätzers für mehrdimensionale autoregressive Moving-Average (CARMA) Prozesse in stetiger Zeit werden bewiesen. Um Eigenschaften des zugrundeliegenden Lévy-Prozesses zu schätzen, wird die verallgemeinerte Momentenmethode erweitert und auf approximative, aus einem beobachteten CARMA-Prozess rekonstruierte Lévy-Zuwächse angewandt. Die Methode führt zu konsistenten und asymptotisch normal verteilten Schätzwerten, wenn hochfrequente Beobachtungen zur Verfügung stehen. Ein zweites Ziel der Arbeit ist es, die asymptotische Spektralverteilung empirischer Kovarianzmatrizen linearer Prozesse durch deren Spektraldichte zu charakterisieren. Schließlich werden ein zentraler Grenzwertsatz für ein Perkolationsproblem bewiesen und die Übergangskerne einer verwandten Markovkette explizit beschrieben.

Abstract

Several aspects of the statistical analysis of linear processes are investigated. For equidistantly observed multivariate Lévy-driven continuous-time autoregressive moving average (CARMA) processes we prove consistency and asymptotic normality of the quasi maximum likelihood estimator. To infer further characteristics of the driving Lévy process, we extend the classical generalized method of moments and apply it to approximate Lévy increments that are reconstructed from a sampled CARMA process. This approach results in consistent and asymptotically Gaussian estimates if high-frequency observations are available. Another objective of our work is to characterize the limiting spectral distribution of sample covariance matrices of linear processes through their second-order properties. Finally, we prove a Central Limit Theorem for a first-passage percolation problem, and describe the higher order transition kernels of a related Markov chain.

Contents

1. Introduction

1.1. Background and motivation

In the development of mathematical theories and models, one often faces the problem of reconciling the antagonistic notions of generality and tractability. Restricting the complexity of a model usually brings about the possibility of a deeper understanding of its properties, but also the danger of its range of potential applications being reduced. For decades – and still today – much of the research in the fields of mathematical statistics and probability theory has been devoted to the definition and analysis of stochastic processes which are, on the one hand, applicable to a wide range of real-world problems and, on the other hand, mathematically tractable. These two requirements are not mutually exclusive, indeed the possibility of rigorous statistical analysis of a stochastic model is a key prerequisite for its successful application in practice. In addition to being sufficiently simple, stochastic processes are considered practically useful only if they are able to reproduce the stylized facts of the observed time series which they are meant to model. These stylized facts are simplified generalizations of empirical findings and may include distributional properties such as skewness or heavy tails, path properties such as jumps or a certain degree of smoothness, volatility clustering and long-range dependence, see e. g. Cont (2001) for an empirical analysis of financial data.

Of particular importance is the question whether a time series should be modelled by a continuous-time or a discrete-time process, where the parameter of a stochastic process is referred to as time even though, in practice, it might just as well represent a space variable or some other physical quantity. The answer to this question might either be guided by physical considerations as in the case of modelling temperature changes, which is an inherently continuous-time phenomenon, or motivated by the availability of different mathematical techniques. For instance, the pricing of financial derivatives is conveniently done within the continuous-time framework, where Itô's formula or variants thereof can be used, even though the continuous nature of real asset prices is at least debatable due to the presence of microstructure noise (Aït-Sahalia and Yu, 2009; Amihud, Mendelson and Pedersen, 2006; Hansen and Lunde, 2006).

In practice, even phenomena for which a continuous-time model is appropriate are often observed and recorded at discrete points in time only, and it is a major challenge to develop a theory of how to recover characteristics of a suitable stochastic process from this partial information. This question becomes particularly relevant if the available observations are unequally spaced or are recorded at a high frequency, because in such situations discrete-time models with their preferred fixed time scale can usually not be employed successfully. Need for the ability to accommodate such irregular observations, the wish for mathematical elegance, and the belief that many physical and economic quantities change at least in an approximate way continually are the main reasons for the recently observed upsurge of continuous-time processes in stochastic modelling.

One particularly rich class of stochastic processes that are versatile and at the same time amenable to rigorous analysis are linear processes. The term linear process is not used consistently in the literature, and, in fact, there is a hierarchy of definitions that begins with finite-dimensional linear models and ends with a notion of linearity that contains almost every stationary stochastic process. In the following, we will use several of these different definitions and exercise the discretion to adopt that particular notion of linearity which is most useful for the respective purpose.

In order to be able to employ a stochastic model in practice, it is crucial that one understands in detail its statistical properties. This includes, but is not limited to, inference of model parameters and hypothesis testing, a prerequisite of which is a detailed understanding of the distribution of certain statistics derived from the model.

In this thesis, selected aspects of the statistical analysis of linear processes will be treated from a theoretical point of view and illustrated by enlightening simulation studies and examples. The results allow for the recently introduced promising continuous-time ARMA models to be used in applications, they extend existing results in the literature about sample covariance matrices of general infinite-order moving average processes in a mathematically appealing way, and they provide an explicit description of the finite-time behaviour of a first-passage percolation problem.

Continuous-time autoregressive moving average processes

The most restrictive definition of discrete-time linear stochastic processes are finite-order autoregressive moving average (ARMA) processes and finite-dimensional linear stochastic state space models, about which there exists an abundant literature (e. g. Brockwell and Davis, 1991; Hannan, 1970). It has been known for a long time that these two classes of stochastic processes are equivalent, and that they constitute, from a second-order point of view, the processes with rational spectral densities (Hannan and Deistler, 1988). Not surprisingly, the special explicit structure of these processes allows for an extremely rich theory about their probabilistic and analytical properties (e. g. Caines, 1988), and strong results with respect to statistical inference (e. g. Hannan, 1973; Hannan and Kavalieris, 1984a,b), but also restricts the areas of applicability of ARMA processes due to their inability to exhibit volatility clustering or long-range dependence. Since, as described above, continuous-time processes are in many aspects superior to discrete-time ones, continuous-time analogues of ARMA and linear state space models have been defined in order to make this convenient linear structure available for models defined in continuous time (Brockwell, 2001b; Doob, 1944; Marquardt and Stelzer, 2007). Formally, a continuous-time autoregressive moving average (CARMA) process $\boldsymbol{Y}$ with autoregressive polynomial $P(z) = z^p + A_1 z^{p-1} + \ldots + A_p$ and moving average polynomial $Q(z) = B_0 z^q + B_1 z^{q-1} + \ldots + B_q$ is defined as the solution of the differential equation

$$P(\mathrm{D})\boldsymbol{Y}(t) = Q(\mathrm{D})\mathrm{D}\boldsymbol{L}(t), \quad \mathrm{D} \coloneqq \frac{\mathrm{d}}{\mathrm{d}t}, \tag{1.1.1}$$

which resembles the difference equation defining an ARMA process. The randomness is introduced into the model by the driving Lévy process $\boldsymbol{L}$, which allows for a CARMA process to exhibit a wide variety of marginal distributions. The orders p and q determine the smoothness of the sample paths of a CARMA process, which may be discontinuous if $p - q$ equals one. This is an important feature because many economic time series are thought to exhibit jumps (see, e. g., Barndorff-Nielsen and Shephard, 2001a). By allowing the process $\boldsymbol{Y}$ to be multidimensional, it is possible to employ one joint model for several quantities and, thus, to capture their dependencies. With respect to volatility clustering and long-range dependence CARMA processes have the same shortcomings as their discrete-time counterparts; they serve, however, as building blocks for more complicated models possessing these features (e. g. Barndorff-Nielsen and Stelzer, 2011; Brockwell and Marquardt, 2005; Haug and Stelzer, 2011).

A substantial part of this thesis will deal with statistical inference, or more precisely parametric estimation, for multivariate Lévy-driven CARMA processes. Equation (1.1.1) entails that the definition of a particular CARMA process requires the specification of the integer-valued autoregressive and moving average orders p, q, of the coefficient matrices A_i, B_j, and of the driving Lévy process $\boldsymbol{L}$; it is thus a multi-step procedure, of which the second and third step are treated in this thesis. To make allowance for the dominant role of digital data processing and the fact that, for many phenomena of interest, continuous-time observations are not available, the estimation is based on observations of the CARMA process at discrete points in time.

To estimate the coefficient matrices A_i, B_j of a discretely observed CARMA process, we make extensive use of the linear structure of an equidistantly sampled continuous-time state space model. We derive different ARMA and state space representations for these sampled processes and investigate their probabilistic and analytical properties. Extending the work of Boubacar Mainassara and Francq (2011); Francq and Zakoïan (1998), we prove asymptotic properties for the quasi maximum likelihood estimator of a very general class of second-order linear state space models in discrete time; more precisely, we allow for both system and observation noise and impose only a strong-mixing assumption in the sense of Rosenblatt (1956). These new results, together with our detailed understanding of discretized CARMA processes, enable us to show that the same desirable asymptotic properties, strong consistency and asymptotic normality, also hold for the quasi maximum likelihood estimator of discretely observed CARMA processes.

Our approach to estimating the driving Lévy process is based on Brockwell, Davis and Yang (2011), who suggested the following method for the special case of a univariate CARMA process of order $(p, q) = (2, 1)$. If the CARMA process satisfies a minimum phase condition, it is possible to express the values of the driving Lévy process as a function of the values of the CARMA process observed continuously since the infinite past. Refining this observation, one can compute a set of approximate increments of the driving process from discrete observations on a finite time interval. Since the increments of a Lévy process are independent and identically distributed (i. i. d.), and their common distribution uniquely determines the whole process, these approximate increments can be used to estimate a parametric model of the driving Lévy process by, e. g., maximum likelihood. We extend this method in two ways: we show how to recover the driving process from a continuously observed multivariate CARMA process of any order $p > q$, and instead of restricting attention to maximum likelihood estimators, we allow the sample of approximate increments to be used with any suitable generalized method of moments estimator as defined by Hansen (1982). So far, asymptotic properties of this class of estimators were known when the sample is part of a sequence which is sufficiently independent for a Central Limit Theorem to hold. We relax this assumption and consider data that can be represented as a general additive perturbation of an i. i. d. sequence. Without imposing weak-dependence conditions on the perturbation, we prove that general method of moments estimators based on such a sample are consistent and asymptotically normally distributed if the length of the sample tends to infinity, and the noise goes to zero in a suitable way; these results are used to show that a parametric model of the driving Lévy process of multivariate CARMA process can be estimated if high-frequency observations are available, and to derive the asymptotic properties of the corresponding estimators.

Sample covariance matrices of linear processes

A very broad class of stochastic processes, which generalize causal finite-order ARMA processes, and which are also often referred to as linear processes, consists of infinite-order moving averages of the form $X_t = \sum_{j=0}^{\infty} c_j Z_{t-j}$, for some i. i. d. sequence Z. Such processes extend the modelling capabilities of finite-dimensional state space models by including long-range dependent processes such as fractionally integrated ARMA processes (Granger and Joyeux, 1980; Hosking, 1981). Much fewer and, in general, only weaker results can be proved for this more general class of linear processes than for ARMA processes.

A fundamental paradigm in statistics is that population quantities should be estimated by corresponding sample quantities, which, in the theory of stochastic processes or more general dynamical systems, is formalized by the concept of ergodicity (Krengel, 1985). In particular, the instantaneous covariance matrix of a p-dimensional stochastic process X, which is an important measure for dependencies between the components of X, can be estimated by the quantity $n^{-1}\boldsymbol{X}\boldsymbol{X}^T$, where n denotes the length of the sample and the columns of the matrix $\boldsymbol{X}$ are given by the individual observations; if the process X is ergodic, this estimate will converge to the true covariance matrix as the number n of observations tends to infinity. The prominent role played by the sample covariance matrix $n^{-1}\boldsymbol{X}\boldsymbol{X}^T$ in multivariate statistics is described in Anderson (2003); Muirhead (1982).

In practice, one is often confronted with the situation that the number of variables is of the same order of magnitude as the number of observations, and the basic assumption in the classical limit theorems that n tends to infinity while p remains fixed might then not be plausible. To overcome this and similar problems, a tremendous amount of research has been dedicated to the analysis of spectral properties of large random matrices (see, e. g., Bai and Silverstein, 2010, for an introduction). It is by now a classical result in random matrix theory that the distribution of the eigenvalues of the matrix $n^{-1}\boldsymbol{X}\boldsymbol{X}^T$ converge to a non-trivial limiting measure if the entries of the $p \times n$ matrix $\boldsymbol{X}$ are i. i. d. random variables with mean zero, and both p and n converge to infinity such that the ratio p/n tends to a positive finite limit (Marchenko and Pastur, 1967); this behaviour is qualitatively different from the situation where p is fixed, in which case each of these eigenvalues converges to the common variance of the entries of $\boldsymbol{X}$.

Building on recent results about the limiting spectral distribution of products of random matrices (Bai and Zhou, 2008; Pan, 2010), we investigate asymptotic spectral properties of sample covariance matrices of a special class of high-dimensional stochastic processes, the components of which are modelled by independent infinite-order moving average processes with identical second-order characteristics. Our main result is an explicit characterization of the Stieltjes transform of the limiting distribution of the eigenvalues of these sample covariance matrices in terms of the spectral densities of the underlying linear processes. Furthermore, we obtain the same result for a related random matrix model in which the assumption of independence between the rows is relaxed.

First-passage percolation

If one relaxes the definition of a linear process even further and defines it as an infinite-order moving average not of an i. i. d. sequence but rather of a merely uncorrelated random sequence, then every purely non-deterministic stationary stochastic process in discrete time with finite second moments is linear (Caines, 1988; Wold, 1954). Since this characterization of linearity applies to many stochastic processes which are thought of as being genuinely non-linear, such as GARCH or regime-dependent ARMA processes (Tong, 1990), it is usually not used in the literature. For the important special case of Markov chains, however, which satisfy an ergodicity condition, many interesting mathematical results, such as Central Limit Theorems, can be obtained without a strong notion of linearity (e. g., Chen, 1999).

In the last part of this thesis, we consider a first-passage percolation problem on a random graph as a model for a porous medium, going back to Hammersley and Welsh (1965). In this interpretation, the geometry of the medium is approximated by a discrete graph, and its heterogeneous permeability is represented by random weights on its edges, which determine the time it takes a fluid to travel from a vertex of the graph to one of its neighbours. Because of an Ergodic Theorem for subadditive stochastic processes (Kingman, 1968), it is known that, for many first-passage percolation models, a fluid asymptotically advances into the medium at a constant speed, often referred to as the inverse time constant, which is characteristic for the model under consideration.

Here, we investigate an essential one-dimensional model with independent edge weights, for which the time constant is known explicitly. By proving a Central Limit Theorem for an associated Markov chain and performing a quantitative analysis of its multi-step transition kernels, we provide a characterization of the finite-time behaviour of this particular first-passage percolation model. Moreover, we establish the necessary tools to understand the fluctuations of the asymptotic speed at which a fluid percolates through the medium in that model.

1.2. Outline of the thesis

In the following, we outline the structure and contents of this thesis. Each of the following six chapters is based on a paper and therefore essentially self-contained. Unified notation is used within each of the three parts of the thesis, with a summary of generally used symbols and abbreviations given after the bibliography on page 249.

The first part, consisting of Chapters 2 to 4, deals with the statistical analysis and estimation of Lévy-driven multivariate continuous-time autoregressive moving average (abbreviated MCARMA) processes which are observed on an equidistant time grid. It is based on the three papers Brockwell and Schlemm (2011); Schlemm and Stelzer (2011, 2012).

In Chapter 2 it is shown that, similar to the discrete-time theory, the class of multivariate CARMA processes is equivalent to the class of continuous-time linear state space models in the sense that the output process of any state space model possesses a CARMA representation and vice versa. The second topic of the chapter is to investigate the probabilistic properties of an equidistantly sampled CARMA process, which, in the univariate setting, is known to be a discrete-time ARMA process driven by a weak white noise sequence. We generalize this result to the multidimensional setting and show under a mild continuity assumption on the driving Lévy process that the noise sequence is not only uncorrelated, but exponentially completely regular (β-mixing) and, in particular, strongly mixing. It is verified that this continuity assumption is satisfied in most practically relevant situations, including the case where the driving Lévy process has a non-singular Gaussian component, is compound Poisson with an absolutely continuous jump size distribution, or has an infinite Lévy measure admitting a density around zero. We thus establish a strong notion of asymptotic independence for the linear innovations of a sampled CARMA process, which are useful in the development of an estimation theory for this class of stochastic processes.

Thereafter, in Chapter 3, we turn our attention to parametric inference and consider quasi maximum likelihood (QML) estimation for general non-Gaussian discrete-time linear state space models and equidistantly observed multivariate Lévy-driven continuous-time autoregressive moving average processes. In the discrete-time setting, we prove strong consistency and asymptotic normality of the QML estimator under standard moment assumptions and a strong-mixing condition on the output process of the state space model, but without the requirement that its linear innovations form a sequence of martingale differences. In the second part of the chapter, we further investigate probabilistic and analytical properties of sampled continuous-time state space models, and we apply our results from the discrete-time setting to derive the asymptotic properties of the QML estimator of discretely recorded MCARMA processes. Under natural identifiability conditions, the estimators are again consistent and asymptotically normally distributed for any sampling frequency. We also demonstrate the practical applicability of our method through a simulation study and a data example from econometrics.

Our discussion of statistical inference for multivariate continuous-time ARMA processes is concluded in Chapter 4, where we propose a procedure for estimating a parametric model of the driving Lévy process if high-frequency observations are available. Beginning with a new state space representation, we develop a method to recover the driving Lévy process exactly from a continuous record of the observed CARMA process. We use tools from numerical analysis and the theory of infinitely divisible distributions to extend this result to allow for the approximate recovery of unit increments of the driving Lévy process from discrete-time observations of the MCARMA process. These approximate increments can be represented as a perturbation of the true i. i. d. increments by a dependent noise sequence, which we analyse in detail as a function of the inverse sampling frequency h. We establish the asymptotic properties of generalized method of moments (GMM) estimators based on an additively disturbed i. i. d. sample, and we use this result to show that, if $h = h_N$ is chosen dependent on the length N of the observation period such that $h_N N$ converges to zero, then any suitable GMM estimator based on the reconstructed sample of unit increments of the driving process has the same asymptotic distribution as the one based on the true increments. In particular, these estimators are consistent and asymptotically normally distributed. We illustrate the theoretical results by a simulation study, in which we estimate the parameters of a discretely observed Gamma-driven CARMA process of order $(3, 1)$.

In the second part of the thesis, which incorporates material from the two papers Pfaffel and Schlemm (2011, 2012), certain statistics of linear processes are investigated within the framework of random matrix theory.

In Chapter 5 we derive the distribution of the eigenvalues of a large sample covariance matrix when the data is dependent in time. More precisely, the dependence for each variable $i = 1, \ldots, p$ is modelled as a linear process $(X_{i,t})_{t\in\mathbb{Z}} = (\sum_{j=0}^{\infty} c_j Z_{i,t-j})_{t\in\mathbb{Z}}$, where the random variables $\{Z_{i,t}\}$ are assumed to be independent with finite fourth moments and to satisfy a Lindeberg-type condition. A sample of n observations from such a p-dimensional data-generating process is represented by the matrix $\boldsymbol{X} = (X_{i,t})_{it} \in \mathbb{R}^{p\times n}$. If the sample size n and the number of variables $p = p_n$ both converge to infinity such that their asymptotic ratio $y = \lim_{n\to\infty} n/p_n$ is positive, then the empirical spectral distribution of the sample covariance matrix $p^{-1}\boldsymbol{X}\boldsymbol{X}^T$ converges, as n tends to infinity, to a non-random distribution, which only depends on y and the spectral density of the linear process $(X_{1,t})_{t\in\mathbb{Z}}$. Our results contain the classical Marchenko–Pastur law as a special case, but also apply to more general sample covariance matrices, in particular to those derived from (fractionally integrated) ARMA processes.

A more complicated random matrix model, which is also derived from a linear process, and in which the entries are dependent across both rows and columns, is studied in Chapter 6. More precisely, we investigate matrices of the form $\boldsymbol{X} = (X_{(i-1)n+t})_{it} \in \mathbb{R}^{p\times n}$ derived from a linear process $(X_t)_{t\in\mathbb{Z}} = (\sum_j c_j Z_{t-j})_{t\in\mathbb{Z}}$, where the $\{Z_t\}$ satisfy the same assumptions as before. Under the assumption that both p and n tend to infinity such that the ratio p/n converges to a finite positive limit y, we show that the empirical spectral distribution of the sample covariance matrix $p^{-1}\boldsymbol{X}\boldsymbol{X}^T$ converges almost surely to the same deterministic measure occurring in the apparently simpler model studied in Chapter 5.

The final part of the thesis consists of Chapter 7 and contains material from Schlemm (2011). As a model for a porous medium we consider the first-passage percolation problem on the random graph with vertex set $\mathbb{N} \times \{0, 1\}$, edges joining vertices at Euclidean distance equal to unity, and independent exponential edge weights. We provide a Central Limit Theorem for the first-passage times l_n between the vertices $(0, 0)$ and $(n, 0)$, which can be interpreted as the time it takes a fluid to percolate a distance n through the medium. This extends earlier results about the almost sure convergence of the average speed l_n/n, as n tends to infinity, given in Schlemm (2009). We use generating function techniques to compute the n-step transition kernels of a closely related Markov chain which can be used to calculate explicitly the asymptotic variance in the Central Limit Theorem and to quantify the fluctuations of the average speed at which the fluid percolates.

1.3. Open problems for future research

Finally, in order to embed the results in this thesis into the context of on-going research, I would like to mention some natural open questions arising from the present work.

1.3.1. Estimation of CARMA processes

As described above, the complete specification of a continuous-time autoregressive moving average model is an intricate multi-step process, which is hierarchical in the sense that later steps depend on the results of earlier ones. The first step, the selection of the autoregressive and moving average orders p and q, or of an alternative set of structure indices describing the algebraic structure of vector ARMA processes, like the Kronecker indices and the McMillan degree, is not considered in this work at all. In the discrete-time setting, order selection for ARMA and state space models is usually performed by the minimization of an information criterion that quantifies the trade-off between the complexity of the considered model on the one hand, and a goodness-of-fit measure on the other hand. The inclusion of a penalty term that depends on the number of parameters in the model prevents an over-parametrized model, which often describes the observed data better, from being selected and ensures that the order of the process can be estimated consistently. The most famous choices appear to be modified versions of the Akaike and Bayesian Information Criterion, AIC and BIC, respectively, as well as the minimum description length (MDL) approach. For more information on these criteria the reader is invited to consult Akaike (1977); Hannan (1980); Hannan and Rissanen (1982); Rissanen (1983, 1986); Schwarz (1978). The difficulty in transferring these discrete-time results to the continuous-time case comes from the fact that, while a linear state space structure is preserved under sampling, its ARMA orders and Kronecker indices are usually not (Åström, Hagander and Sternby, 1984; Bar-Ness and Langholz, 1975; Hagiwara and Araki, 1988; Larsson, 2005; Söderström, 1990). Moreover, the linear innovations of the sampled process are, in general, not a martingale difference sequence, which would be necessary for the proofs of the discrete-time results to carry over easily; however, as shown in Chapter 2, the linear innovations satisfy a strong-mixing condition, which is likely to imply a sufficient amount of asymptotic independence for comparable results to be established.

Once the model order is known or estimated, the second step in the specification of a CARMA model is the determination of the coefficient matrices in the autoregressive and moving average polynomials. Quasi maximum likelihood estimation, the asymptotic properties of which are the subject of Chapter 3, is a robust way to perform this step which does not require assumptions on the distribution of the driving Lévy process except the finiteness of certain moments. Its practical applicability would be enhanced significantly if preliminary estimates as starting values for the maximization of the quasi likelihood could be obtained without having to solve a computationally expensive high-dimensional optimization problem. A potential solution to this preliminary estimation problem could be the use of subspace identification techniques as described in Bauer (2009); Bauer, Deistler and Scherrer (1999); Chiuso (2006); Chiuso and Picci (2005); Mari, Stoica and McKelvey (2000); Peternell, Scherrer and Deistler (1996); van Overschee and De Moor (1996). If one is interested in estimating heavy-tailed CARMA processes which do not satisfy the assumption of having finite second moments, different techniques need to be developed (see, e. g., Klüppelberg and Mikosch, 1993; Mikosch, Gadrich, Klüppelberg and Adler, 1995).

The estimation of the driving Lévy process is developed in Chapter 4 for the case that high-frequency observations of the CARMA process are available. The total number of observations used in the estimation procedure is proportional to the length of the observation horizon and to the sampling frequency. In practice, resources are limited, and it is therefore of considerable importance to investigate how one should choose the length of the observation period and the sampling frequency such that the possibility of recording a given number of observations is used as efficiently as possible. Another important future research project is the development of estimation approaches that do not rely on high-frequency observations. One possible avenue in this direction is to observe that, by the results in Chapter 2, the linear innovations of an equidistantly sampled CARMA process are themselves a vector ARMA process driven by an infinitely divisible noise sequence whose characteristic triplet is related to the characteristic triplet of the driving Lévy process via the formulæ given in Rajput and Rosiński (1989). It seems therefore possible to estimate the linear innovations using a Kalman filter as in Chapter 3, to invert the vector ARMA equations, and to estimate the characteristics of the noise sequence. From these, one can extract information about the characteristics of the driving Lévy process.

In the treatment of the different aspects of the estimation of CARMA processes in this thesis, it is assumed that the results of the preceding steps are known exactly: quasi maximum likelihood estimation of the autoregressive and moving average polynomials requires knowledge of the McMillan degree of the true model, estimation of the driving Lévy process cannot be carried out if one is ignorant of the coefficient matrices in these polynomials. An important extension of the results in this thesis would be to remove this restriction, and to allow for estimated approximate values of the McMillan degree and the coefficient matrices to be used in subsequent steps. We expect that the asymptotic results remain essentially unaltered for this genuinely multi-step estimation scheme, but the proof of this conjecture appears to be non-trivial. Finally, in order to contribute to a more widespread use of multivariate continuous-time linear processes in applications, it would be helpful to have a ready-to-use software implementation of the results presented in this thesis and the extensions just mentioned.

1.3.2. Covariance matrices of linear processes

In this section I will comment on some extensions of the random matrix model described in Chapter 5; these extensions are motivated partly by applications, partly by their intrinsic mathematical appeal. The main feature of the random matrix model considered in Chapter 5 is that the rows are independent with identical second-order properties, and that the dependence within the rows is modelled by a linear process. The assumption that the rows are independent has been relaxed in Chapter 6 by introducing a dependence between the rows, which turned out to be weak enough not to change the limiting spectral distribution. For a realistic model, in which the rows of a random matrix can be interpreted as, for example, price data of individual assets recorded over time, it is necessary to allow for matrices whose rows are neither independent nor identically distributed. While dependence in this context is very difficult to model adequately, non-identically distributed rows can be easily accommodated in a random matrix model of the form

$$\mathbb{R}^{p\times n} \ni \boldsymbol{X} = (X_{i,t})_{it}, \quad X_{i,t} = \sum_{j=0}^{\infty} c_j^{(i)} Z_{i,t-j},$$

where the ith row is given by an infinite-order moving average process with coefficients $\underline{c}_i := (c_j^{(i)})_j$. A quantitative way to explicitly characterize the limiting spectral distribution of $p^{-1}\boldsymbol{X}\boldsymbol{X}^T$ in terms of the spectral densities associated with the possibly non-identical filters $\underline{c}_i$ would constitute an important extension of our results.

In what has been said so far, the dimensions of the occurring random matrices had a convenient interpretation as the number of variables and time, respectively. If one does not adopt this point of view, it is perhaps more natural to model the random matrix $\boldsymbol{X}$ as the output of a genuinely two-dimensional filter applied to an array $(Z_{i,t})_{it}$ of random variables, that is

$$\mathbb{R}^{p\times n} \ni \boldsymbol{X} = (X_{i,t})_{it}, \quad X_{i,t} = \sum_{j,k=-\infty}^{\infty} c_{j,k} Z_{i-j,t-k}.$$

The model considered in Chapter 5 is a special case of this generalization, corresponding to a filter satisfying $c_{j,k} = 0$, for $j \neq 0$. In the Gaussian setting, a characterization of the limiting spectral distribution for this random matrix model has been derived in Hachem, Loubaton and Najim (2005). In view of the paradigm that many properties of random covariance matrices depend on the distribution of their entries only through their first few moments (Tao and Vu, 2010), it is a natural conjecture that comparable results also hold for non-Gaussian random fields $(Z_{i,t})_{it}$.

Part I

Estimation of Multivariate Lévy-driven CARMA Processes

2. Multivariate CARMA Processes, Continuous-Time State Space Models and Complete Regularity of the Innovations of the Sampled Processes

2.1. Introduction

Continuous-time autoregressive moving average (CARMA) processes are the continuous-time analogues of the widely known discrete-time ARMA processes (see, e. g., Brockwell and Davis, 1991, for a comprehensive introduction); they were first defined in Doob (1944) in the univariate Gaussian setting and have stimulated a considerable amount of research in recent years (see, e. g., Brockwell, 2001a, and references therein). In particular, the restriction of the driving process to Brownian motion was relaxed, and Brockwell (2001b) allowed for Lévy processes with a finite logarithmic moment. Because of their applicability to irregularly spaced observations and high-frequency data, they have turned out to be a versatile and powerful tool in the modelling of phenomena from the natural sciences, engineering and finance. Recently, Marquardt and Stelzer (2007) extended the concept to multivariate CARMA (MCARMA) processes with the intention of being able to model the joint behaviour of several dependent time series. MCARMA processes are thus also the continuous-time analogues of discrete-time vector ARMA (VARMA) models (see, e. g., Lütkepohl, 2005).

The aim of this chapter is twofold: first, we establish the equivalence between multivariate CARMA processes and multivariate continuous-time state space models, a correspondence which is well known in the discrete-time setting (Hannan and Deistler, 1988); second, we investigate the probabilistic properties of the discrete-time process obtained by recording the values of an MCARMA process at discrete, equally spaced points in time. A detailed understanding of the innovations of the weak VARMA process which arises is a prerequisite for proving asymptotic properties of various statistics of a discretely observed MCARMA process. One notion of asymptotic independence which is very useful in this context is complete regularity in the sense of Volkonskiĭ and Rozanov (1959) (see Section 2.4 for a precise definition), and we show that the innovations of a discretized MCARMA process have this desirable property. Our results therefore not only provide important insight into the probabilistic structure of CARMA processes, but they are also fundamental to the development of an estimation theory for non-Gaussian continuous-time state space models based on equidistant observations.

In this chapter, we show that a sampled MCARMA process is a discrete-time VARMA process with dependent innovations. While the mixing behaviour of ARMA and more general linear processes is fairly well understood (see, e. g., Athreya and Pantula, 1986; Mokkadem, 1988; Pham and Tran, 1985), the mixing properties of the innovations of a sampled continuous-time process have received very little attention. From Brockwell and Lindner (2009), it is only known that the innovations of a discretized univariate Lévy-driven CARMA process are weak white noise, which, by itself, is typically of little help in applications. We show that the linear innovations of a sampled MCARMA process satisfy a set of VARMA equations, and we conclude that, under a mild continuity assumption on the driving Lévy process, they are geometrically completely regular and, in particular, geometrically strongly mixing. This continuity assumption is further shown to be satisfied for most of the practically relevant choices of the driving Lévy process, including processes with a non-singular Gaussian component as well as compound Poisson processes with an absolutely continuous jump size distribution, and infinite activity processes whose Lévy measures admit a density in a neighbourhood of zero.

Outline of the chapter The chapter is structured as follows. In Section 2.2 we review some well-known properties of Lévy processes, which we will use later on. The class of multivariate CARMA processes, in a slightly more general form than in the original definition of Marquardt and Stelzer (2007), is introduced and described in detail in Section 2.3; it is further shown to be equivalent to the class of continuous-time state space models. In Section 2.4 the main result about the mixing properties of the sampled processes is stated and demonstrated to be applicable in many practical situations. The proofs of the results are presented in Section 2.5.

Notation We use the following notation. The space of $m \times n$ matrices with entries in the ring $\mathbb{K}$ is denoted by $M_{m,n}(\mathbb{K})$ or $M_m(\mathbb{K})$ if $m = n$. A^T denotes the transpose of the matrix A, the matrices $\mathbf{1}_m$ and 0_m are the identity and the zero element of $M_m(\mathbb{K})$, respectively, and $A \otimes B$ stands for the Kronecker product of the matrices A and B. The zero vector in $\mathbb{R}^m$ is denoted by $\mathbf{0}_m$, and $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ represent the Euclidean norm and inner product, respectively. Finally, $\mathbb{K}[z]$ ($\mathbb{K}\{z\}$) is the ring of polynomial (rational) expressions in z over $\mathbb{K}$, and $I_B(\cdot)$ is the indicator function of the set B.

2.2. Multivariate Lévy processes

In this section we review the definition of a multivariate Lévy process and some elementary facts about these processes which we will use later. More details and proofs can be found in, for instance, Sato (1999).

Definition 2.1 (Lévy process) A (one-sided) $\mathbb{R}^m$-valued Lévy process $\boldsymbol{L} = (\boldsymbol{L}(t))_{t\in\mathbb{R}^+}$ is a stochastic process on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with the following properties:

i) the increments of $\boldsymbol{L}$ are stationary and independent, that is the distribution of $\boldsymbol{L}(t+s) - \boldsymbol{L}(t)$ does not depend on t, and for every $n \in \mathbb{N}$ and all $0 < t_0 < t_1 \ldots < t_n < \infty$, the random variables $\boldsymbol{L}(t_1) - \boldsymbol{L}(t_0), \boldsymbol{L}(t_2) - \boldsymbol{L}(t_1), \ldots, \boldsymbol{L}(t_n) - \boldsymbol{L}(t_{n-1})$ are independent,

ii) almost surely, $\boldsymbol{L}(0) = \boldsymbol{0}_m$,

iii) the process $\boldsymbol{L}$ is continuous in probability.

Every $\mathbb{R}^m$-valued Lévy process$(\boldsymbol{L}(t))_{t\geqslant 0}$ can without loss of generality be assumed to be càdlàg and is completely characterized by its characteristic function in the Lévy–Khintchine form $\mathbb{E}\mathrm{e}^{\mathrm{i}\langle \boldsymbol{u},\boldsymbol{L}(t)\rangle} = \exp\{t\psi^{\boldsymbol{L}}(\boldsymbol{u})\}$, $\boldsymbol{u} \in \mathbb{R}^m$, $t \geqslant 0$, where the characteristic exponent $\psi^{\boldsymbol{L}}$ has the special form

$$\psi^{\boldsymbol{L}}(\boldsymbol{u}) = \mathrm{i}\langle \gamma^{\boldsymbol{L}}, \boldsymbol{u}\rangle - \frac{1}{2}\langle \boldsymbol{u}, \Sigma^{\mathcal{G}}\boldsymbol{u}\rangle + \int_{\mathbb{R}^m} \left[\mathrm{e}^{\mathrm{i}\langle \boldsymbol{u},\boldsymbol{x}\rangle} - 1 - \mathrm{i}\langle \boldsymbol{u}, \boldsymbol{x}\rangle I_{\{\|x\|\leqslant 1\}}\right] \nu^{\boldsymbol{L}}(\mathrm{d}\boldsymbol{x}).$$

The vector $\gamma^{\boldsymbol{L}} \in \mathbb{R}^m$ is called the *drift*, the positive semidefinite, symmetric $m \times m$ matrix $\Sigma^{\mathcal{G}}$ is the *Gaussian covariance matrix* and $\nu^{\boldsymbol{L}}$ is a measure on $\mathbb{R}^m$, referred to as the *Lévy measure*, satisfying

$$\nu^{\boldsymbol{L}}(\{\boldsymbol{0}_m\}) = 0, \quad \int_{\mathbb{R}^m} \min(\|\boldsymbol{x}\|^2, 1)\nu^{\boldsymbol{L}}(\mathrm{d}\boldsymbol{x}) < \infty.$$

We will work with two-sided Lévy processes $\boldsymbol{L} = (\boldsymbol{L}(t))_{t\in\mathbb{R}}$. These are obtained from two independent copies $(\boldsymbol{L}_1(t))_{t\geqslant 0}$, $(\boldsymbol{L}_2(t))_{t\geqslant 0}$ of a one-sided Lévy process via the construction

$$\boldsymbol{L}(t) = \begin{cases} \boldsymbol{L}_1(t), & t \geqslant 0, \\ -\lim_{s\nearrow -t} \boldsymbol{L}_2(s), & t < 0. \end{cases}$$

Throughout, we restrict our attention to Lévy processes with zero means and finite second moments.

Assumption L1 The Lévy process $\boldsymbol{L}$ satisfies $\mathbb{E}\boldsymbol{L}(1) = \boldsymbol{0}_m$ and $\mathbb{E}\,\|\boldsymbol{L}(1)\|^2 < \infty$.

The assumption that $\mathbb{E}\boldsymbol{L}(1) = \boldsymbol{0}_m$ is made only for notational convenience and is not essential for our results. The premise that $\boldsymbol{L}$ has finite variance is, in contrast, a true restriction, which is often made in the analysis of ARMA processes in both discrete and continuous time. The treatment of the infinite variance case requires different techniques and often does not lead to comparable results (see, e. g., Klüppelberg and Mikosch, 1993; Mikosch et al., 1995). It is well known that $\boldsymbol{L}$ has finite second moments if and only if $\int_{\|\boldsymbol{x}\|\geqslant 1}\|\boldsymbol{x}\|^2\nu(\mathrm{d}\boldsymbol{x})$ is finite, and that $\Sigma^{\boldsymbol{L}} = \mathbb{E}\boldsymbol{L}(1)\boldsymbol{L}(1)^T$ is then given by $\int_{\|\boldsymbol{x}\|\geqslant 1}\boldsymbol{x}\boldsymbol{x}^T\nu^{\boldsymbol{L}}(\mathrm{d}\boldsymbol{x}) + \Sigma^{\mathcal{G}}$.

2.3. Lévy-driven multivariate CARMA processes and state space models

If $\boldsymbol{L}$ is a two-sided Lévy process with values in $\mathbb{R}^m$, and $p > q$ are positive integers, then the d-dimensional $\boldsymbol{L}$-driven autoregressive moving average (MCARMA) process with autoregressive polynomial

$$z \mapsto P(z) := \mathbf{1}_d z^p + A_1 z^{p-1} + \ldots + A_p \in M_d(\mathbb{R}[z]) \tag{2.3.1a}$$

and moving average polynomial

$$z \mapsto Q(z) := B_0 z^q + B_1 z^{q-1} + \ldots + B_q \in M_{d,m}(\mathbb{R}[z]) \tag{2.3.1b}$$

is heuristically thought of as the solution of the formally written pth-order linear differential equation

$$P(\mathrm{D})\boldsymbol{Y}(t) = Q(\mathrm{D})\mathrm{D}\boldsymbol{L}(t), \quad \mathrm{D} \equiv \frac{\mathrm{d}}{\mathrm{d}t}, \tag{2.3.2}$$

which is the continuous-time analogue of a set of discrete-time ARMA equations. We note that we allow for the driving Lévy process $\boldsymbol{L}$ and the $\boldsymbol{L}$-driven multivariate CARMA process $\boldsymbol{Y}$ to have different dimensions and thus slightly extend the original definition of Marquardt and Stelzer (2007). All the results we need from this paper are easily seen to continue to hold in this more general setting. Since, in general, Lévy processes are not differentiable, the differential equation (2.3.2) is purely formal and is, as usually, interpreted as being equivalent to the *state space representation*

$$\mathrm{d}\boldsymbol{G}(t) = \mathcal{A}\boldsymbol{G}(t)\mathrm{d}t + \mathcal{B}\mathrm{d}\boldsymbol{L}(t), \quad \boldsymbol{Y}(t) = \mathcal{C}\boldsymbol{G}(t), \quad t \in \mathbb{R}, \tag{2.3.3}$$

where the matrices $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are given by

$$\mathcal{A} = \begin{pmatrix} 0 & \mathbf{1}_d & 0 & \ldots & 0 \\ 0 & 0 & \mathbf{1}_d & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \ldots & \ldots & 0 & \mathbf{1}_d \\ -A_p & -A_{p-1} & \ldots & \ldots & -A_1 \end{pmatrix} \in M_{pd}(\mathbb{R}), \tag{2.3.4a}$$

$$\mathcal{B} = \begin{pmatrix} \beta_1^T & \cdots & \beta_p^T \end{pmatrix}^T \in M_{pd,m}(\mathbb{R}), \tag{2.3.4b}$$

where

$$\beta_{p-j} = - I_{\{0,\ldots,q\}}(j) \left[\sum_{i=1}^{p-j-1} A_i \beta_{p-j-i} - B_{q-j} \right],$$

and

$$\mathcal{C} = \begin{pmatrix} \mathbf{1}_d & 0_d & \ldots & 0_d \end{pmatrix} \in M_{d,pd}(\mathbb{R}). \tag{2.3.4c}$$

In view of representation (2.3.3), MCARMA processes are linear continuous-time state space models. We will consider this class of processes and see that it is in fact equivalent to the class of MCARMA models.

Definition 2.2 (State space model) A *continuous-time linear state space model* $(A, B, C, \boldsymbol{L})$ of dimension N with values in $\mathbb{R}^d$ is characterized by an $\mathbb{R}^m$-valued driving Lévy process $\boldsymbol{L}$, a state 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 a state equation of Ornstein–Uhlenbeck type

$$\mathrm{d}\boldsymbol{X}(t) = A\boldsymbol{X}(t)\mathrm{d}t + B\mathrm{d}\boldsymbol{L}(t), \quad t \in \mathbb{R}, \tag{2.3.5a}$$

and an observation equation

$$\boldsymbol{Y}(t) = C\boldsymbol{X}(t), \quad t \in \mathbb{R}. \tag{2.3.5b}$$

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

A solution $\boldsymbol{Y}$ of Eqs. (2.3.5) is called *causal* if, for all t, $\boldsymbol{Y}(t)$ is independent of the σ-algebra generated by the future values $\{\boldsymbol{L}(s) : s > t\}$. Every solution of the state equation (2.3.5a) satisfies

$$\boldsymbol{X}(t) = \mathrm{e}^{A(t-s)}\boldsymbol{X}(s) + \int_s^t \mathrm{e}^{A(t-u)} B\mathrm{d}\boldsymbol{L}(u), \quad s, t \in \mathbb{R}, \quad s < t. \tag{2.3.6}$$

The independent increment property of Lévy processes implies that $\boldsymbol{X}$ is a Markov process (see also Protter, 1990, for more information on stochastic integration). We always work under the following standard causal stationarity assumption.

Assumption E1 The eigenvalues of the matrix A have strictly negative real parts.

The following proposition is well known (Sato and Yamazato, 1984) and recalls conditions for the existence of a stationary causal solution of the state equation (2.3.5a) for easy reference.

Proposition 2.3 *If Assumptions L1 and E1 hold, then the state equation (2.3.5a) has a unique strictly stationary, causal solution $\boldsymbol{X}$ given by*

$$\boldsymbol{X}(t) = \int_{-\infty}^{t} \mathrm{e}^{A(t-u)} B \mathrm{d}\boldsymbol{L}(u), \quad t \in \mathbb{R}, \tag{2.3.7}$$

which has the same distribution as $\int_0^\infty \mathrm{e}^{Au} B \mathrm{d}\boldsymbol{L}(u)$. Moreover, $\boldsymbol{X}(t)$ has mean zero,

$$\begin{aligned} \mathbb{V}\mathrm{ar}(\boldsymbol{X}(t)) =& \mathbb{E}\boldsymbol{X}(t)\boldsymbol{X}(t)^T \\ =& \int_0^\infty \mathrm{e}^{Au} B \Sigma^{\boldsymbol{L}} B^T \mathrm{e}^{A^T u} \mathrm{d}u := \Gamma_0, \end{aligned} \tag{2.3.8a}$$

$$\begin{aligned} \mathbb{C}\mathrm{ov}\left(\boldsymbol{X}(t+h), \boldsymbol{X}(t)\right) =& \mathbb{E}\boldsymbol{X}(t+h)\boldsymbol{X}(t)^T \\ =& \mathrm{e}^{Ah}\Gamma_0, \quad h \geqslant 0, \end{aligned} \tag{2.3.8b}$$

and Γ_0 satisfies $A\Gamma_0 + \Gamma_0 A^T = -B\Sigma^{\boldsymbol{L}} B^T$.

It is an immediate consequence of Eqs. (2.3.5b) and (2.3.8b) that the output process $\boldsymbol{Y}$ has mean zero and autocovariance function $h \mapsto \gamma_{\boldsymbol{Y}}(h) = C\mathrm{e}^{Ah}\Gamma_0 C^T$, $h \geqslant 0$, and that $\boldsymbol{Y}$ can be written as a moving average of the driving Lévy process as

$$\boldsymbol{Y}(t) = \int_{-\infty}^{\infty} g(t-u) \mathrm{d}\boldsymbol{L}(u), \quad t \in \mathbb{R}; \quad g(t) = C\mathrm{e}^{At} B I_{[0,\infty)}(t). \tag{2.3.9}$$

These equations serve, with A, B and C defined as in Eqs. (2.3.4), as the definition of a multivariate CARMA process with autoregressive and moving average polynomials given by Eqs. (2.3.1). They also show that the behaviour of the process $\boldsymbol{Y}$ depends on the values of the individual matrices A, B and C only through the products $C\mathrm{e}^{At}B$, $t \in \mathbb{R}$. These products are, in turn, intimately related to the rational matrix function $H : z \mapsto C(z\mathbf{1}_N - A)^{-1}B$, which is called the *transfer function* of the state space model (2.3.5). A pair (P, Q), $P \in M_d(\mathbb{R}[z])$, $Q \in M_{d.m}(\mathbb{R}[z])$, of rational matrix functions is a *left matrix fraction description* for the rational matrix function $H \in M_{d,m}(\mathbb{R}\{z\})$ if $P(z)^{-1}Q(z) = H(z)$ for all $z \in \mathbb{C}$. The next theorem gives an answer to the question of what other state space representations besides Eq. (2.3.3) can be used to define an MCARMA process. The proof is given in Section 2.5.

Theorem 2.4 *If (P, Q) is a left matrix fraction description for the transfer function $z \mapsto C(z\mathbf{1}_N - A)^{-1}B$, then the stationary solution $\boldsymbol{Y}$ of the state space model $(A, B, C, \boldsymbol{L})$ defined by Eqs. (2.3.5) is an $\boldsymbol{L}$-driven MCARMA process with autoregressive polynomial P and moving average polynomial Q.*

Corollary 2.5 *The classes of MCARMA and causal continuous-time state space models are equivalent.*

Proof By definition, every MCARMA process is the output process of a state space model. Conversely, given any state space model $(A, B, C, \boldsymbol{L})$ with output process $\boldsymbol{Y}$, Caines (1988,

Appendix 2, Theorem 8) shows that the transfer function $H : z \mapsto C(zI_N - A)^{-1}B$ possesses a left matrix fraction description $H(z) = P(z)^{-1}Q(z)$. Hence, by Theorem 2.4, $\boldsymbol{Y}$ is an MCARMA process. □

2.4. Complete regularity of the innovations of sampled MCARMA processes

For a continuous-time stochastic process $\boldsymbol{Y} = (\boldsymbol{Y}(t))_{t\in\mathbb{R}}$ and a positive constant h, the corresponding sampled process $\boldsymbol{Y}^{(h)} = (\boldsymbol{Y}_n^{(h)})_{n\in\mathbb{Z}}$ is defined by $\boldsymbol{Y}_n^{(h)} = \boldsymbol{Y}(nh)$. A common problem in applications is the estimation of a set of model parameters based on observations of the values of a realization of a continuous-time process at equally spaced points in time. In order to make MCARMA processes $\boldsymbol{Y}$ amenable to parameter inference from equidistantly sampled observations, it is important to have a good understanding of the probabilistic properties of $\boldsymbol{Y}^{(h)}$. One such property which has turned out to be useful for the derivation of asymptotic properties of estimators is *mixing*, for which there are several different notions (see, e. g., Bradley, 2007, for a detailed exposition). Let I denote $\mathbb{Z}$ or $\mathbb{R}$. For a stationary process $\boldsymbol{X} = (X_n)_{n\in I}$ on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$, we write $\mathscr{F}_n^m = \sigma(X_j : j \in I, n < j < m)$, $-\infty \leqslant n < m \leqslant \infty$. The α-mixing coefficients $(\alpha(m))_{m\in I}$ were introduced in Rosenblatt (1956) and are defined by

$$\alpha(m) = \sup_{A\in\mathscr{F}_{-\infty}^0,\, B\in\mathscr{F}_m^\infty} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

If $\lim_{m\to\infty}\alpha(m) = 0$, then the process $\boldsymbol{X}$ is called *strongly mixing*, and if there exist constants $C > 0$ and $0 < \lambda < 1$ such that $\alpha_m < C\lambda^m$, $m \geqslant 1$, it is called *exponentially strongly mixing*. The β-mixing coefficients $(\beta(m))_{m\in I}$, introduced in Volkonskiĭ and Rozanov (1959), are similarly defined as

$$\beta(m) = \mathbb{E}\sup_{B\in\mathscr{F}_m^\infty} |\mathbb{P}(B|\mathscr{F}_{-\infty}^0) - \mathbb{P}(B)|.$$

If $\lim_{m\to\infty}\beta(m) = 0$, then the process $\boldsymbol{X}$ is called *completely regular* or *β-mixing*, and if there exist constants $C > 0$ and $0 < \lambda < 1$ such that $\beta_m < C\lambda^m$, $m \geqslant 1$, it is called *exponentially completely regular*. The equivalent definition

$$\beta(m) = \frac{1}{2} \sup_{\substack{I,J\geqslant 1 \\ (A_i)_{1\leqslant i\leqslant I}\in\left(\mathscr{F}_{-\infty}^0\right)^I,\, \bigcup_{i=1}^I A_i=\Omega \\ (B_j)_{1\leqslant j\leqslant J}\in\left(\mathscr{F}_m^\infty\right)^J,\, \bigcup_{j=1}^J B_j=\Omega}} \sum_{i=1}^{I}\sum_{j=1}^{J} \left|\mathbb{P}(A_i\cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)\right|,$$

given in Dedecker, Doukhan, Lang, León R., Louhichi and Prieur (2007, Eq. (1.2.4)), shows that complete regularity, just like strong mixing, is preserved under sampling, aggregation and linear transformations. It is clear from these definitions that $\alpha(m) \leqslant \beta(m)$ and that (exponential) complete regularity thus implies (exponential) strong mixing. It has been shown in Marquardt and Stelzer (2007, Proposition 3.34) that every causal multivariate CARMA process $\boldsymbol{Y}$ with a finite κth moment, $\kappa > 0$, is strongly mixing and this naturally carries over to the sampled process $\boldsymbol{Y}^{(h)}$ because the relevant σ-algebras, over which the supremum is taken, become smaller by sampling. We therefore do not investigate the mixing properties of the process $\boldsymbol{Y}^{(h)}$ itself, but rather of its linear innovations.

Definition 2.6 (Linear innovations) Let $(\boldsymbol{Y}_n)_{n\in\mathbb{Z}}$ be an $\mathbb{R}^d$-valued stationary stochastic process with finite second moments. The *linear innovations* $(\boldsymbol{\varepsilon}_n)_{n\in\mathbb{Z}}$ of $(\boldsymbol{Y}_n)_{n\in\mathbb{Z}}$ are then defined by

$$\boldsymbol{\varepsilon}_n = \boldsymbol{Y}_n - P_{n-1}\boldsymbol{Y}_n, \quad P_n = \text{orthogonal projection onto } \overline{\operatorname{span}}\left\{\boldsymbol{Y}_\nu : -\infty < \nu \leqslant n\right\}, \tag{2.4.1}$$

where the closure is taken in the Hilbert space of square-integrable random variables with inner product $(X, Y) \mapsto \mathbb{E}\langle X, Y\rangle$.

From now on, we work under an additional assumption, which is standard in the univariate case.

Assumption E2 The eigenvalues $\lambda_1, \ldots, \lambda_N$ of the state transition matrix A in Eq. (2.3.5a) are distinct.

A polynomial $p \in M_d(\mathbb{C}[z])$ is called *monic* if its leading coefficient is equal to $\mathbf{1}_d$ and *Schur-stable* if the zeros of $z \mapsto \det p(z)$ all lie in the complement of the closed unit disk. We first give a semi-explicit construction of a weak vector ARMA representation of $\boldsymbol{Y}^{(h)}$ with complex-valued coefficient matrices, a generalization of Brockwell et al. (2011, Proposition 3).

Theorem 2.7 (Weak ARMA representation) *Assume that $\boldsymbol{Y}$ is the output process of the state space system (2.3.5) satisfying Assumptions L1, E1 and E2, and that $\boldsymbol{Y}^{(h)}$ is its sampled version with linear innovations $\boldsymbol{\varepsilon}^{(h)}$. Define the Schur-stable polynomial $\varphi \in \mathbb{C}[z]$ by*

$$\varphi(z) = \prod_{\nu=1}^{N}\left(1 - \mathrm{e}^{h\lambda_\nu} z\right) =: \left(1 - \varphi_1 z - \ldots - \varphi_N z^N\right). \tag{2.4.2}$$

There then exists a monic Schur-stable polynomial $\Theta \in M_d(\mathbb{C}[z])$ of degree at most $N-1$, such that

$$\varphi(\mathrm{B})\boldsymbol{Y}_n^{(h)} = \Theta(\mathrm{B})\boldsymbol{\varepsilon}_n^{(h)}, \quad n \in \mathbb{Z}, \tag{2.4.3}$$

where B denotes the backshift operator, that is, $\mathrm{B}^j\boldsymbol{Y}_n^{(h)} = \boldsymbol{Y}_{n-j}^{(h)}$ for every non-negative integer j.

This result is very important for the proof of the mixing properties of the innovations sequence $\boldsymbol{\varepsilon}^{(h)}$ because it establishes an explicit linear relationship between $\boldsymbol{\varepsilon}^{(h)}$ and $\boldsymbol{Y}^{(h)}$. A good understanding of the mixing properties of $\varepsilon^{(h)}$ is not only theoretically interesting, but is also practically of considerable relevance for the purpose of statistical inference for multivariate CARMA processes. One estimation procedure in which the importance of the mixing properties of the innovations of the sampled process is clearly visible is quasi maximum likelihood (QML) estimation. Assume that $\Theta \subset \mathbb{R}^r$ is a compact parameter set and that a parametric family of MCARMA processes is given by the mapping $\Theta \ni \boldsymbol{\vartheta} \mapsto (A_{\boldsymbol{\vartheta}}, B_{\boldsymbol{\vartheta}}, C_{\boldsymbol{\vartheta}}, \boldsymbol{L}_{\boldsymbol{\vartheta}})$. It follows from Theorem 2.7 and Brockwell and Davis (1991, §11.5) that the Gaussian likelihood of observations $\boldsymbol{y}^L = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_L)$ under the model corresponding to a particular value $\boldsymbol{\vartheta}$ is given by

$$\mathscr{L}(\boldsymbol{\vartheta}, \boldsymbol{y}^L) = (2\pi)^{-Ld/2} \left(\prod_{n=1}^{L} \det V_{\boldsymbol{\vartheta},n} \right)^{-1/2} \exp\left\{ -\frac{1}{2} \sum_{n=1}^{L} \boldsymbol{e}_{\boldsymbol{\vartheta},n}^T V_{\boldsymbol{\vartheta},n}^{-1} \boldsymbol{e}_{\boldsymbol{\vartheta},n} \right\}, \tag{2.4.4}$$

where $\boldsymbol{e}_{\boldsymbol{\vartheta},n}$ is the residual of the minimum mean-squared error linear predictor of $\boldsymbol{Y}_n$ given the preceding observations, and $V_{\boldsymbol{\vartheta},n}$ is the corresponding covariance matrix. From a practical perspective, it is important to note that all quantities necessary to evaluate the Gaussian likelihood (2.4.4) can be conveniently computed by using the Kalman recursions (Brockwell and Davis, 1991, §12.2) and the state space representation given in Lemma 2.18. In case the observations $\boldsymbol{y}^L$ are (part of) a realization of the sampled MCARMA process $\boldsymbol{Y}_{\boldsymbol{\vartheta}_0}^{(h)}$ corresponding to the parameter value $\boldsymbol{\vartheta}_0$, the prediction error sequence $(\boldsymbol{e}_{\boldsymbol{\vartheta}_0,n})_{n\geqslant 1}$ is – up to an additive, exponentially decaying term which comes from the initialization of the Kalman filter – (part of) a realization of the innovations sequence $\varepsilon^{(h)}$ of $\boldsymbol{Y}_{\boldsymbol{\vartheta}_0}^{(h)}$. In order to be able to analyse the asymptotic behaviour of the natural QML estimator

$$\hat{\boldsymbol{\vartheta}}^L = \operatorname{argmax}_{\boldsymbol{\vartheta}\in\Theta} \mathscr{L}(\boldsymbol{\vartheta}, \boldsymbol{y}^L)$$

in the limit as $L \to \infty$, it is necessary to have a Central Limit Theorem for sums of the form

$$\frac{1}{\sqrt{L}} \sum_{n=1}^{L} \frac{\partial}{\partial \boldsymbol{\vartheta}} \left[\log\det V_{\boldsymbol{\vartheta},n} + \boldsymbol{e}_{\boldsymbol{\vartheta},n}^T V_{\boldsymbol{\vartheta},n}^{-1} \boldsymbol{e}_{\boldsymbol{\vartheta},n} \right] \Big|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0}. \tag{2.4.5}$$

Existing results in the literature (Bradley, 2007; Herrndorf, 1984; Ibragimov, 1962) ensure that various notions of weak dependence, and, in particular, strong mixing, are sufficient for a Central Limit Theorem for the expression (2.4.5) to hold, provided that the asymptotic independence is uniform in $\boldsymbol{\vartheta}$. Theorem 2.8 below is thus the necessary starting point for the development of an estimation theory for multivariate CARMA processes, which involves some additional issues like identifiability of parametrizations and is thus beyond the scope of this chapter. In Chapter 3 we will pursue this topic further and develop a quasi maximum likelihood estimation theory for state space models in both discrete and continuous time, which is, in particular, applicable to multivariate CARMA processes.

Before presenting a sufficient condition for the innovations $\boldsymbol{\varepsilon}^{(h)}$ to be completely regular, we first observe that the eigenvalues $\lambda_1, \ldots, \lambda_N$ of the matrix A are the roots of the characteristic polynomial $z \mapsto \det(z\mathbf{1}_N - A)$, which, by the Fundamental Theorem of Algebra, implies that they are either real or occur in complex conjugate pairs. We can therefore assume that they are ordered in such a way that, for some $r \in \{0, \ldots, N\}$,

$$\lambda_\nu \in \mathbb{R}, \quad 1 \leqslant \nu \leqslant r, \qquad\qquad \lambda_\nu = \overline{\lambda_{\nu+1}} \in \mathbb{C}\backslash\mathbb{R}, \quad \nu = r+1, r+3, \ldots, N-1.$$

By Lebesgue's decomposition theorem (Klenke, 2008, Theorem 7.33), every measure μ on $\mathbb{R}^d$ can be uniquely decomposed as $\mu = \mu_{\mathrm{c}} + \mu_{\mathrm{s}}$, where μ_{c} and μ_{s} are absolutely continuous and singular, respectively, with respect to the d-dimensional Lebesgue measure. If μ_{c} is not the zero measure, then we say that μ has a non-trivial absolutely continuous component.

Theorem 2.8 (Mixing Innovations) *Assume that $\boldsymbol{Y}$ is the output process of the linear continuous-time state space model $(A, B, C, \boldsymbol{L})$ satisfying Assumptions L1, E1 and E2. Denote by $\boldsymbol{\varepsilon}^{(h)}$ the linear innovations of the sampled process $\boldsymbol{Y}^{(h)}$, and further assume that the law of the $\mathbb{R}^{mN}$-valued random variable*

$$\mathscr{M}^{(h)} = \left[\begin{array}{ccccccc} \boldsymbol{M}_1^{(h)T} & \cdots & \boldsymbol{M}_r^{(h)T} & \underline{\boldsymbol{M}}_{r+1}^{(h)\;T} & \underline{\boldsymbol{M}}_{r+3}^{(h)\;T} & \cdots & \underline{\boldsymbol{M}}_{N-1}^{(h)\;T} \end{array}\right]^T, \tag{2.4.6}$$

where

$$\underline{\boldsymbol{M}}_\nu^{(h)} = \left[\begin{array}{cc} \operatorname{Re} \boldsymbol{M}_\nu^{(h)T} & \operatorname{Im} \boldsymbol{M}_\nu^{(h)T} \end{array}\right]^T, \quad \boldsymbol{M}_\nu^{(h)} = \int_0^h \mathrm{e}^{(h-u)\lambda_\nu} \mathrm{d}\boldsymbol{L}(u), \quad \nu = 1, \ldots, N, \tag{2.4.7}$$

has a non-trivial absolutely continuous component with respect to the mN-dimensional Lebesgue measure. Then, $\boldsymbol{\varepsilon}^{(h)}$ is exponentially completely regular.

The assumption on the distribution of $\mathscr{M}^{(h)}$ made in Theorem 2.8 is not very restrictive. Its verification is based on the following lemma, which allows us to derive sufficient conditions in terms of the Lévy process $\boldsymbol{L}$ which show that it is indeed satisfied in most practical situations.

Lemma 2.9 *There exist matrices $G \in M_{mN}(\mathbb{R})$ and $H \in M_{mN,m}(\mathbb{R})$ such that $\mathscr{M}^{(h)}$ equals $\mathscr{M}(h)$, where $(\mathscr{M}(t))_{t\geqslant 0}$ is the unique solution of the stochastic differential equation*

$$\mathrm{d}\mathscr{M}(t) = G\mathscr{M}(t)\mathrm{d}t + H\mathrm{d}\boldsymbol{L}(t), \quad \mathscr{M}(0) = \mathbf{0}_{mN}. \tag{2.4.8}$$

Moreover, rank $H = m$*, and the $mN \times mN$ matrix* $\left[\begin{array}{cccc} H & GH & \cdots & G^{N-1}H \end{array}\right]$ *is non-singular.*

The last part of the statement is referred to as *controllability* of the pair (G, H) and is essential in the proofs of the following explicit sufficient conditions for Theorem 2.8 to hold.

Proposition 2.10 *Assume that the Lévy process $\boldsymbol{L}$ has a non-singular Gaussian covariance matrix $\Sigma^{\mathcal{G}}$. Theorem 2.8 then holds.*

Proof By Sato (2006, Corollary 2.19), the law of $\mathscr{M}^{(h)}$ is infinitely divisible with Gaussian covariance matrix given by $\int_0^h \mathrm{e}^{Gu}H\Sigma^{\mathcal{G}}H^T\mathrm{e}^{G^Tu}\mathrm{d}u$. By the controllability of (G,H) and Bernstein (2005, Lemma 12.6.2) (see also Lemma 3.39 in the next chapter), this matrix is non-singular, and Sato (1999, Exercise 29.14) completes the proof. □

A simple Lévy process of practical importance which does not have a non-singular Gaussian covariance matrix is the *compound Poisson Process*, which is defined by $\boldsymbol{L}(t) = \sum_{n=1}^{N(t)} \boldsymbol{J}_n$, where $(N(t))_{t\in\mathbb{R}^+}$ is a Poisson process, and $(\boldsymbol{J}_n)_{n\in\mathbb{Z}}$ is an i. i. d. sequence independent of $(N(t))_{t\in\mathbb{R}^+}$; the law of $\boldsymbol{J}_n$ is called the *jump size distribution*. The proof of Priola and Zabczyk (2009, Theorem 1.1), in conjunction with Lemma 2.9, implies the following result.

Proposition 2.11 *Assume that $\boldsymbol{L}$ is a compound Poisson process with absolutely continuous jump size distribution. Theorem 2.8 then holds.*

Under a similar smoothness assumption, the conclusion of Theorem 2.8 also holds in the case of infinite activity Lévy processes. The statement follows from applying Priola and Zabczyk (2009, Theorem 1.1) to Eq. (2.4.8), see also Bodnarchuk and Kulik (2008); Picard (1996).

Proposition 2.12 *Assume that the Lévy measure $\nu^{\boldsymbol{L}}$ of $\boldsymbol{L}$ satisfies $\nu^{\boldsymbol{L}}(\mathbb{R}^m) = \infty$, and that there exists a positive constant ρ such that $\nu^{\boldsymbol{L}}$ restricted to the ball $\{\boldsymbol{x}\in\mathbb{R}^m : \|\boldsymbol{x}\| \leqslant \rho\}$ has a density with respect to the m-dimensional Lebesgue measure. Theorem 2.8 then holds.*

While the preceding three propositions already cover a wide range of Lévy processes encountered in practice, there are some relevant cases which are not yet taken care of, in particular the construction of the Lévy process as a vector of independent univariate Lévy processes (Corollary 2.16 below). To also cover this and related choices, we employ the polar decomposition for Lévy measures (Barndorff-Nielsen, Maejima and Sato, 2006, Lemma 2.1). By this result, for every Lévy measure $\nu^{\boldsymbol{L}}$, there exists a probability measure α on the $(m-1)$-sphere $S^{m-1} := \{\boldsymbol{x}\in\mathbb{R}^m : \|\boldsymbol{x}\| = 1\}$ and a family $\{\nu_{\boldsymbol{\xi}} : \boldsymbol{\xi}\in S^{m-1}\}$ of measures on $\mathbb{R}^+$, such that for each Borel set $B\in\mathcal{B}(\mathbb{R}^+)$, the function $\boldsymbol{\xi}\mapsto\nu_{\boldsymbol{\xi}}(B)$ is measurable and

$$\nu^{\boldsymbol{L}}(B) = \int_{S^{m-1}}\int_0^\infty I_B(\lambda\boldsymbol{\xi})\nu_{\boldsymbol{\xi}}(\mathrm{d}\lambda)\alpha(\mathrm{d}\boldsymbol{\xi}), \quad B\in\mathcal{B}(\mathbb{R}^m\backslash\{\boldsymbol{0}_m\}). \tag{2.4.9}$$

A linear subspace of a finite-dimensional vector space of codimension one is called a hyperplane.

Proposition 2.13 *If the Lévy measure $\nu^{\boldsymbol{L}}$ has a polar decomposition $(\alpha,\nu_{\boldsymbol{\xi}} : \boldsymbol{\xi}\in S^{m-1})$ such that for any hyperplane $\mathcal{H}\subset\mathbb{R}^m$, it holds that $\int_{S^{m-1}} I_{\mathbb{R}^m\backslash\mathcal{H}}(\boldsymbol{\xi})\int_0^\infty \nu_{\boldsymbol{\xi}}(\mathrm{d}\lambda)\alpha(\mathrm{d}\boldsymbol{\xi}) = \infty$, then Theorem 2.8 holds.*

Proof The proof rests on the main theorem of Simon (2010). We denote by $\operatorname{im} H$ the image of the linear operator associated with the matrix H. Since $\operatorname{rank} H = m$ and the pair (G, H) is controllable, we only have to show that $\nu^{\boldsymbol{L}}(\{\boldsymbol{x} \in \mathbb{R}^m : H\boldsymbol{x} \in \operatorname{im} H \backslash \mathcal{H}\}) = \infty$ for all hyperplanes $\mathcal{H} \subset \operatorname{im} H$, and since $\mathbb{R}^m \cong \operatorname{im} H$, the last condition is equivalent to $\nu^{\boldsymbol{L}}(\mathbb{R}^m \backslash \mathcal{H}) = \infty$ for all hyperplanes $\mathcal{H} \subset \mathbb{R}^m$. Using Eq. (2.4.9) and the fact that for every $\boldsymbol{\xi} \in S^{m-1}$ and every $\lambda \in \mathbb{R}^+$, the vector $\lambda\boldsymbol{\xi}$ is in $\mathcal{H}$ if and only if the vector $\boldsymbol{\xi}$ is, this is seen to be equivalent to the assumption of the proposition. □

Corollary 2.14 *If the Lévy measure $\nu^{\boldsymbol{L}}$ possesses a polar decomposition $(\alpha, \nu_{\boldsymbol{\xi}} : \boldsymbol{\xi} \in S^{m-1})$ such that $\alpha(S^{m-1} \backslash \mathcal{H})$ is positive for all hyperplanes $\mathcal{H} \in \mathbb{R}^m$, and $\nu_{\boldsymbol{\xi}}(\mathbb{R}^+) = \infty$ for α-almost every $\boldsymbol{\xi}$, then Theorem 2.8 holds.*

Corollary 2.15 *If the Lévy measure $\nu^{\boldsymbol{L}}$ has a polar decomposition $(\alpha, \nu_{\boldsymbol{\xi}} : \boldsymbol{\xi} \in S^{m-1})$ such that for some linearly independent vectors $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m \in S^{m-1}$, it holds that, for $k = 1, \ldots, m$, $\alpha(\boldsymbol{\xi}_k) > 0$ and $\nu_{\boldsymbol{\xi}_k}(\mathbb{R}^+) = \infty$, then Theorem 2.8 holds.*

Corollary 2.16 *Assume that $l \geqslant m$ is an integer and that the matrix $R \in M_{m,l}(\mathbb{R})$ has full rank m. If $\boldsymbol{L} = R\begin{pmatrix} L_1 & \cdots & L_l \end{pmatrix}^T$, where L_k, $k = 1, \ldots, l$, are independent univariate Lévy processes with Lévy measures ν_k^L satisfying $\nu_k^L(\mathbb{R}) = \infty$, then Theorem 2.8 holds.*

2.5. Proofs

2.5.1. Proofs for Section 2.3

Proof (of Theorem 2.4) The first step of the proof is to show that any pair (P, Q) of the form (2.3.1) is a left matrix fraction description of $\mathcal{C}(z\mathbf{1}_{pd} - \mathcal{A})^{-1}\mathcal{B}$, provided $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are defined as in Eqs. (2.3.4). We first show the relation

$$(z\mathbf{1}_{pd} - \mathcal{A})^{-1}\mathcal{B} = \begin{bmatrix} w_1(z)^T & \cdots & w_p^T(z) \end{bmatrix}^T, \tag{2.5.1}$$

where $w_j(z) \in M_{d,m}(\mathbb{R}\{z\})$, $j = 1, \ldots p$, are defined by the equations

$$w_j(z) = \frac{1}{z}(w_{j+1}(z) + \beta_j), \quad j = 1, \ldots, p-1, \tag{2.5.2a}$$

and

$$w_p(z) = \frac{1}{z}\left(-\sum_{k=0}^{p-1} A_{p-k} w_{k+1}(z) + \beta_p\right). \tag{2.5.2b}$$

Since it has been shown in Marquardt and Stelzer (2007, Theorem 3.12) that $w_1(z) = P(z)^{-1}Q(z)$, this will prove the assertion. Eq. (2.5.1) is clearly equivalent to $\mathcal{B} = (z\mathbf{1}_{pd} -$

$\mathcal{A})\begin{bmatrix} w_1(z)^T & \cdots & w_p^T(z) \end{bmatrix}^T$, which explicitly reads

$$\begin{aligned}\beta_j =& z w_j(z) - w_{j+1}(z), \quad j = 1, \ldots p-1, \\ \beta_p =& z w_p(z) + A_p w_1(z) + \ldots + A_1 w_p(z),\end{aligned}$$

and is thus equivalent to Eq. (2.5.2).

For the second step we consider a given state space model $(A, B, C, \boldsymbol{L})$. Using the spectral representation of the matrix exponential (Lax, 2002, Theorem 17.5),

$$e^{At} = \frac{1}{2\pi i}\int_\Gamma e^{zt}(z\mathbf{1}_N - A)^{-1}dz, \quad t \in \mathbb{R}, \tag{2.5.3}$$

where Γ is some closed contour in $\mathbb{C}$ winding around each eigenvalue of A exactly once, it follows that

$$\begin{aligned}\boldsymbol{Y}(t) =& \int_{-\infty}^t C e^{A(t-u)} B d\boldsymbol{L}(u) \\ =& \frac{1}{2\pi i}\int_{-\infty}^t \int_\Gamma e^{z(t-u)} C(z\mathbf{1}_N - A)^{-1} B dz d\boldsymbol{L}(u) \\ =& \frac{1}{2\pi i}\int_{-\infty}^t \int_\Gamma e^{z(t-u)} P(z)^{-1} Q(z) dz d\boldsymbol{L}(u) \\ =& \frac{1}{2\pi i}\int_{-\infty}^t \int_\Gamma e^{z(t-u)} \mathcal{C}(z\mathbf{1}_{pd} - \mathcal{A})^{-1} \mathcal{B} dz d\boldsymbol{L}(u) \\ =& \int_{-\infty}^t \mathcal{C} e^{\mathcal{A}(t-u)} \mathcal{B} d\boldsymbol{L}(u),\end{aligned}$$

where the matrices $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are defined in terms of (P, Q) by Eqs. (2.3.4). Thus, $\boldsymbol{Y}$ is a multivariate CARMA process with autoregressive polynomial P and moving average polynomial Q. □

2.5.2. Proofs for Section 2.4

In this section we present in detail the proofs of our main results, Theorems 2.7 and 2.8 and Lemma 2.9, as well as several auxiliary results. The first is a generalization of Brockwell et al. (2011, Proposition 2) expressing MCARMA processes as a sum of multivariate Ornstein–Uhlenbeck processes.

Proposition 2.17 *Let $\boldsymbol{Y}$ be the the output process of the state space system (2.3.5), and assume that Assumption E2 holds. Then, there exist vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N \in \mathbb{C}^d \backslash \{\mathbf{0}_d\}$ and $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N \in \mathbb{C}^m \backslash \{\mathbf{0}_m\}$, such that $\boldsymbol{Y}$ can be decomposed into a sum of dependent, complex-valued Ornstein–Uhlenbeck processes as $\boldsymbol{Y}(t) = \sum_{\nu=1}^N \boldsymbol{Y}_\nu(t)$, where*

$$\boldsymbol{Y}_\nu(t) = e^{\lambda_\nu(t-s)}\boldsymbol{Y}_\nu(s) + \boldsymbol{b}_\nu \int_s^t e^{\lambda_\nu(t-u)} d\langle \boldsymbol{s}_\nu, \boldsymbol{L}(u)\rangle, \quad s, t \in \mathbb{R}, \quad s < t. \tag{2.5.4}$$

Proof We first choose a left matrix fraction description (P, Q) of the transfer function $z \mapsto C(z\mathbf{1}_N - A)^{-1}B$ such that $z \mapsto \det P(z)$ and $z \mapsto \det Q(z)$ have no common zeros and $z \mapsto \det P(z)$ has no multiple zeros. This is always possible, by Assumption E2. Inserting the spectral representation (2.5.3) of e^{At} into the kernel $g(t)$ (Eq. (2.3.9)), we obtain $g(t) = \frac{1}{2\pi\mathrm{i}}\int_\Gamma \mathrm{e}^{zt}C(z\mathbf{1}_N - A)^{-1}B\mathrm{d}z I_{[0,\infty)}(t)$ and, by construction, the integrand equals $\mathrm{e}^{zt}P(z)^{-1}Q(z)I_{[0,\infty)}(t)$. After writing $P(z)^{-1} = \frac{1}{\det P(z)}\operatorname{adj} P(z)$, where adj denotes the adjugate of a matrix, an element-wise application of the Residue Theorem from complex analysis (Dieudonné, 1968, 9.16.1) shows that

$$g(t) = \sum_{\nu=1}^{N} \mathrm{e}^{\lambda_\nu t}\frac{1}{(\det P)'(\lambda_\nu)}\operatorname{adj} P(\lambda_\nu)Q(\lambda_\nu)I_{[0,\infty)}(t),$$

where $(\det P)'(\lambda_\nu) := \frac{d}{dz}\det P(z)|_{z=\lambda_\nu}$ is non-zero because $z \mapsto \det P(z)$ has only simple zeros. The same fact, in conjunction with the Smith decomposition of P (Bernstein, 2005, Theorem 4.7.5), also implies that $\operatorname{rank} P(\lambda_\nu) = d-1$, and thus $\operatorname{rank}\operatorname{adj} P(\lambda_\nu)$ equals one (Bernstein, 2005, Fact 2.14.7 *ii)*). Since $\det P$ and $\det Q$ have no common zeros, the matrix $[(\det P)'(\lambda_\nu)]^{-1}\operatorname{adj} P(\lambda_\nu)Q(\lambda_\nu)$ also has rank one and can thus be written as $\boldsymbol{b}_\nu \boldsymbol{s}_\nu^T$ for some non-zero $\boldsymbol{b}_\nu \in \mathbb{C}^d$ and $\boldsymbol{s}_\nu \in \mathbb{C}^m$ (Halmos, 1974, §51, Theorem 1). □

Lemma 2.18 *Assume that $\boldsymbol{Y}$ is the output process of the state space model (2.3.5). The sampled process $\boldsymbol{Y}^{(h)}$ then has the state space representation*

$$\boldsymbol{X}_n = \mathrm{e}^{Ah}\boldsymbol{X}_{n-1} + \boldsymbol{N}_n, \quad \boldsymbol{N}_n = \int_{(n-1)h}^{nh} \mathrm{e}^{A(nh-u)}B\mathrm{d}\boldsymbol{L}(u), \quad \boldsymbol{Y}_n^{(h)} = C\boldsymbol{X}_n^{(h)}. \tag{2.5.5}$$

The sequence $(\boldsymbol{N}_n)_{n\in\mathbb{Z}}$ is i. i. d. with mean zero and covariance matrix

$$\breve{\Sigma} = \mathbb{E}\boldsymbol{N}_n\boldsymbol{N}_n^T = \int_0^h \mathrm{e}^{Au}B\Sigma^{\boldsymbol{L}}B^T\mathrm{e}^{A^Tu}\mathrm{d}u. \tag{2.5.6}$$

Proof Eqs. (2.5.5) follow from setting $t = nh$, and $s = (n-1)h$ in the moving average representation (2.3.6). It is an immediate consequence of the Lévy process $\boldsymbol{L}$ having independent and stationary increments that the sequence $(\boldsymbol{N}_n)_{n\in\mathbb{Z}}$ is i. i. d., and that its covariance matrix $\breve{\Sigma}$ is given by Eq. (2.5.6). □

From this, we can now proceed to prove the weak vector ARMA representation of the process $\boldsymbol{Y}^{(h)}$.

Proof (of Theorem 2.7) It follows from setting $t = nh$, $s = (n-1)h$ in Eq. (2.5.4) that $\boldsymbol{Y}_n^{(h)}$ can be decomposed as $\boldsymbol{Y}_n^{(h)} = \sum_{\nu=1}^N \boldsymbol{Y}_{\nu,n}^{(h)}$, where $\boldsymbol{Y}_\nu^{(h)}$, satisfying

$$\boldsymbol{Y}_{\nu,n}^{(h)} = \mathrm{e}^{\lambda_\nu h}\boldsymbol{Y}_{\nu,n-1}^{(h)} + \boldsymbol{Z}_{\nu,n}^{(h)}, \quad \boldsymbol{Z}_{\nu,n}^{(h)} = \boldsymbol{b}_\nu \int_{(n-1)h}^{nh} \mathrm{e}^{\lambda_\nu(nh-u)}\mathrm{d}\langle \boldsymbol{s}_\nu, \boldsymbol{L}(u)\rangle,$$

are the sampled versions of the MCAR(1) processes from Proposition 2.17. Analogously to Brockwell and Lindner (2009, Lemma 2.1), we can show by induction that, for each $k \in \mathbb{N}_0$ and all complex $d \times d$ matrices $c_1, \ldots, c_k$, it holds that

$$\begin{aligned}\boldsymbol{Y}_{\nu,n}^{(h)} = \sum_{r=1}^{k} c_r \boldsymbol{Y}_{\nu,n-r}^{(h)} + \left[\mathrm{e}^{\lambda_\nu hk} - \sum_{r=1}^{k} c_r \mathrm{e}^{\lambda_\nu h(k-r)}\right] \boldsymbol{Y}_{\nu,n-k}^{(h)} \\ + \sum_{r=0}^{k-1} \left[\mathrm{e}^{\lambda_\nu hr} - \sum_{j=1}^{r} c_j \mathrm{e}^{\lambda_\nu h(r-j)}\right] \boldsymbol{Z}_{\nu,n-r}^{(h)}.\end{aligned} \tag{2.5.7}$$

If we then use the fact that $\mathrm{e}^{-h\lambda_\nu}$ is a root of $z \mapsto \varphi(z)$, which means that the expression $\mathrm{e}^{Nh\lambda_\nu} - \varphi_1 \mathrm{e}^{(N-1)h\lambda_n u} - \ldots - \varphi_N$ equals zero, and set $k = N$, $c_r = \mathbf{1}_d \varphi_r$, then Eq. (2.5.7) becomes

$$\varphi(\mathrm{B})\boldsymbol{Y}_{\nu,n}^{(h)} = \sum_{r=0}^{N-1} \left[\mathrm{e}^{rh\lambda_\nu} - \sum_{j=1}^{r} \varphi_j \mathrm{e}^{\lambda_\nu h(r-j)}\right] \boldsymbol{Z}_{\nu,n-r}^{(h)}.$$

Summing over ν and rearranging shows that this can be written as

$$\varphi(\mathrm{B})\boldsymbol{Y}_{n}^{(h)} = \sum_{\nu=1}^{N} \boldsymbol{V}_{\nu,n-\nu+1}^{(h)}, \tag{2.5.8}$$

where the i. i. d. sequences $\left(\boldsymbol{V}_{\nu,n}^{(h)}\right)_{n\in\mathbb{Z}}$, $\nu \in \{1, \ldots, N\}$, are defined by

$$\boldsymbol{V}_{\nu,n}^{(h)} = \int_{(n-1)h}^{nh} \sum_{\mu=1}^{N} \boldsymbol{b}_\mu \left[\mathrm{e}^{\lambda_\mu h(\nu-1)} - \sum_{\kappa=1}^{\nu-1} \varphi_\kappa \mathrm{e}^{\lambda_\mu h(\nu-\kappa-1)}\right] \mathrm{e}^{\lambda_\mu(nh-u)} \mathrm{d}\langle \boldsymbol{s}_\mu, \boldsymbol{L}(u)\rangle. \tag{2.5.9}$$

By a straightforward generalization of Brockwell and Davis (1991, Proposition 3.2.1), there exists a monic Schur-stable polynomial $\Theta(z) = \mathbf{1}_d + \Theta_1 z + \ldots + \Theta_{N-1} z^{N-1}$ with matrix-valued coefficients and a white noise sequence $\tilde{\boldsymbol{\varepsilon}}$ such that the $(N-1)$-dependent sequence $\varphi(\mathrm{B})\boldsymbol{Y}^{(h)}$ has the moving average representation $\varphi(\mathrm{B})\boldsymbol{Y}_n^{(h)} = \Theta(\mathrm{B})\tilde{\boldsymbol{\varepsilon}}_n$. Since both φ and Θ are monic, and φ is Schur stable (by Assumption E1), $\tilde{\boldsymbol{\varepsilon}}$ is the innovations process of $\boldsymbol{Y}^{(h)}$ and so it follows that $\tilde{\boldsymbol{\varepsilon}} = \boldsymbol{\varepsilon}^{(h)}$ because the innovations of a stochastic process are uniquely determined. □

As a corollary, we obtain that the innovations sequence $\boldsymbol{\varepsilon}^{(h)}$ itself satisfies a set of strong VARMA equations, the attribute *strong* referring to the fact that the noise sequence is i. i. d., not merely white noise.

Corollary 2.19 *Assume that $\boldsymbol{Y}$ is the output process of the state space system (2.3.5) satisfying Assumptions L1, E1 and E2. Further assume that $\boldsymbol{\varepsilon}^{(h)}$ is the innovations sequence of the sampled process $\boldsymbol{Y}^{(h)}$. There then exists a monic, Schur-stable polynomial $\Theta \in M_d(\mathbb{C}[z])$ of degree at most $N-1$, a polynomial $\theta \in M_{d,dN}(\mathbb{R}[z])$ of degree $N-1$, and a C^{dN}-valued i. i. d. sequence*

$\boldsymbol{W}^{(h)} = (\boldsymbol{W}_n^{(h)})_{n\in\mathbb{Z}}$, *such that*

$$\Theta(\mathrm{B})\boldsymbol{\varepsilon}_n^{(h)} = \theta(\mathrm{B})\boldsymbol{W}_n^{(h)}, \quad n \in \mathbb{Z}. \tag{2.5.10}$$

Proof Combining Eqs. (2.4.3) and (2.5.8) gives

$$\boldsymbol{\varepsilon}_n^{(h)} + \Theta_1^{(h)}\boldsymbol{\varepsilon}_{n-1} + \ldots + \Theta_{N-1}^{(h)}\boldsymbol{\varepsilon}_{n-N+1} = \boldsymbol{V}_{1,n}^{(h)} + \boldsymbol{V}_{2,n-1}^{(h)} + \ldots + \boldsymbol{V}_{N,n-N+1}^{(h)}, \quad n \in \mathbb{Z}, \tag{2.5.11}$$

and with the definitions

$$\boldsymbol{W}_n^{(h)} = \left[\begin{array}{ccc} \boldsymbol{V}_{1,n}^{(h)T} & \ldots & \boldsymbol{V}_{N,n}^{(h)\,T} \end{array}\right]^T \in \mathbb{C}^{dN}, \quad n \in \mathbb{Z}, \tag{2.5.12a}$$

$$\theta(z) = \sum_{j=1}^{N} \theta_j z^{j-1}, \quad \theta_\nu = \left[\begin{array}{ccccccc} \underbrace{0_d \ \cdots \ 0_d}_{\nu-1 \text{ times}} & \mathbf{1}_d & \underbrace{0_d \ \cdots \ 0_d}_{N-\nu \text{ times}} \end{array}\right] \in M_{d,dN}(\mathbb{R}), \tag{2.5.12b}$$

Eq. (2.5.11) becomes $\Theta(\mathrm{B})\boldsymbol{\varepsilon}_n^{(h)} = \theta(\mathrm{B})\boldsymbol{W}_n^{(h)}$, showing that $\boldsymbol{\varepsilon}^{(h)}$ is indeed a vector ARMA process. □

Corollary 2.19 is the central step in establishing the complete regularity of the innovations process $\boldsymbol{\varepsilon}^{(h)}$.

Proof (of Theorem 2.8) We define the $\mathbb{R}^{mN}$-valued random variables

$$\mathscr{M}_n^{(h)} = \left[\begin{array}{cccccccc} \boldsymbol{M}_{n,1}^{(h)T} & \ldots & \boldsymbol{M}_{n,r}^{(h)T} & \underline{\boldsymbol{M}}_{n,r+1}^{(h)\,T} & \underline{\boldsymbol{M}}_{n,r+3}^{(h)\,T} & \ldots & \underline{\boldsymbol{M}}_{n,N-1}^{(h)\,T} \end{array}\right]^T, \quad n \in \mathbb{Z},$$

where, for $\nu = 1, \ldots, N$,

$$\underline{\boldsymbol{M}}_{n,\nu}^{(h)} = \left[\begin{array}{cc} \operatorname{Re} \boldsymbol{M}_{n,\nu}^{(h)T} & \operatorname{Im} \boldsymbol{M}_{n,\nu}^{(h)T} \end{array}\right]^T, \quad \boldsymbol{M}_{n,\nu}^{(h)} = \int_{(n-1)h}^{nh} \mathrm{e}^{\lambda_\nu(nh-u)} \mathrm{d}\boldsymbol{L}(u), \quad n \in \mathbb{Z}.$$

It is clear that the sequence $(\mathscr{M}_n^{(h)})_{n\in\mathbb{Z}}$ is i. i. d. and $\mathscr{M}^{(h)}$ is equal to $\mathscr{M}_1^{(h)}$. We shall now argue that the vector $\boldsymbol{W}_n^{(h)}$, as defined in Eq. (2.5.12a), is equal to a linear transformation of $\mathscr{M}_n^{(h)}$. By Eq. (2.5.9), we can write $\boldsymbol{W}_n^{(h)} = [\Gamma^T \otimes \mathbf{1}_d] \left[\begin{array}{ccc} (\boldsymbol{b}_1 \boldsymbol{s}_1^T \boldsymbol{M}_{n,1}^{(h)})^T & \cdots & (\boldsymbol{b}_N \boldsymbol{s}_N^T \boldsymbol{M}_{n,N}^{(h)})^T \end{array}\right]^T$, where $\Gamma = (\gamma_{\mu,\nu}) \in M_N(\mathbb{C})$ is given by $\gamma_{\mu,\nu} = \mathrm{e}^{\lambda_\mu h(\nu-1)} + \sum_{\kappa=1}^{\nu-1} \varphi_\kappa \mathrm{e}^{\lambda_\mu h(\nu-\kappa-1)}$. With the notation

$$B = \begin{pmatrix} \boldsymbol{b}_1 & \mathbf{0}_d & \ldots & \mathbf{0}_d \\ \mathbf{0}_d & \boldsymbol{b}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}_d \\ \mathbf{0}_d & \ldots & \mathbf{0}_d & \boldsymbol{b}_N \end{pmatrix} \in M_{dN,N}(\mathbb{C}),$$

and

$$S = \begin{pmatrix} \boldsymbol{s}_1^T & \mathbf{0}_d^T & \dots & \mathbf{0}_d^T \\ \mathbf{0}_d^T & \boldsymbol{s}_2^T & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}_d^T \\ \mathbf{0}_d^T & \dots & \mathbf{0}_d^T & \boldsymbol{s}_N^T \end{pmatrix} \in M_{N,mN}(\mathbb{C}),$$

we obtain $\begin{bmatrix} \left(\boldsymbol{b}_1\boldsymbol{s}_1^T\boldsymbol{M}_{n,1}^{(h)}\right)^T & \cdots & \left(\boldsymbol{b}_N\boldsymbol{s}_N^T\boldsymbol{M}_{n,N}^{(h)}\right)^T \end{bmatrix}^T = BS\begin{bmatrix} \boldsymbol{M}_{n,1}^{(h)T} & \cdots & \boldsymbol{M}_{n,N}^{(h)\,T} \end{bmatrix}^T$. We recall that, for $\nu = r+1, r+3, \dots, N-1$, the eigenvalues of A satisfy $\lambda_\nu = \overline{\lambda_{\nu+1}} \in \mathbb{C}\backslash\mathbb{R}$, which implies that

$$\boldsymbol{M}_{n,\nu}^{(h)} = \operatorname{Re}\boldsymbol{M}_{n,\nu}^{(h)} + \mathrm{i}\operatorname{Im}\boldsymbol{M}_{n,\nu}^{(h)} \quad \text{and} \quad \boldsymbol{M}_{n,\nu+1}^{(h)} = \overline{\boldsymbol{M}_{n,\nu}^{(h)}} = \operatorname{Re}\boldsymbol{M}_{n,\nu}^{(h)} - \mathrm{i}\operatorname{Im}\boldsymbol{M}_{n,\nu}^{(h)}.$$

Consequently, we obtain that $\begin{bmatrix} \boldsymbol{M}_{n,1}^{(h)T} & \cdots & \boldsymbol{M}_{n,N}^{(h)\,T} \end{bmatrix}^T = [K\otimes \mathbf{1}_m]\mathscr{M}_n^{(h)}$, where

$$K = \begin{pmatrix} \mathbf{1}_r & & & \\ & J & & \\ & & \ddots & \\ & & & J \end{pmatrix} \in M_N(\mathbb{C}), \quad J = \begin{pmatrix} 1 & \mathrm{i} \\ 1 & -\mathrm{i} \end{pmatrix},$$

so that, in total, $\boldsymbol{W}_n^{(h)} = F\mathscr{M}_n^{(h)}$ with $F = [\Gamma^T \otimes \mathbf{1}_d]BS[K\otimes\mathbf{1}_m] \in M_{dN,mN}(\mathbb{C})$. It follows that the VARMA equation (2.5.10) for $\boldsymbol{\varepsilon}^{(h)}$ becomes $\Theta(\mathrm{B})\boldsymbol{\varepsilon}_n^{(h)} = \tilde{\theta}(\mathrm{B})\mathscr{M}_n^{(h)}$, where $\tilde{\theta}(z) = \theta(z)F$. By the invertibility of Θ, the transfer function $k : z \mapsto \Theta(z)^{-1}\tilde{\theta}(z)$ is analytic in a disk containing the unit disk and permits a power series expansion $k(z) = \sum_{j=0}^{\infty}\Psi_j z^j$. We next argue that the impulse responses Ψ_j are necessarily *real* $d\times mN$ matrices. Since both $\boldsymbol{\varepsilon}_n^{(h)}$ and $\mathscr{M}_n^{(h)}$ are real-valued, it follows from taking the imaginary part of the equation $\boldsymbol{\varepsilon}_n^{(h)} = k(\mathrm{B})\mathscr{M}_n^{(h)}$ that $\mathbf{0}_d = \sum_{j=0}^{\infty}\operatorname{Im}\Psi_j\mathscr{M}_{n-j}^{(h)}$. Consequently, $0 = \mathbb{C}\mathrm{ov}(\mathbf{0}_d) = \sum_{j=0}^{\infty}\operatorname{Im}\Psi_j\,\mathbb{C}\mathrm{ov}(\mathscr{M}_{n-j}^{(h)})\operatorname{Im}\Psi_j^T$, and since each term in the sum is a positive semidefinite matrix, it follows that $\operatorname{Im}\Psi_j\,\mathbb{C}\mathrm{ov}(\mathscr{M}_{n-j}^{(h)})\operatorname{Im}\Psi_j^T = 0$ for every j. The existence of an absolutely continuous component of the law of $\mathscr{M}_{n-j}^{(h)}$ with respect to the mN-dimensional Lebesgue measure implies that $\mathbb{C}\mathrm{ov}(\mathscr{M}_{n-j}^{(h)})$ is non-singular, and it thus follows that $\operatorname{Im}\Psi_j = 0$ for every j. Hence, $k(z) \in M_{d,mN}(\mathbb{R})$ for all real z, and, consequently, $k \in M_{d,mN}(\mathbb{R}\{z\})$. Hannan and Deistler (1988, Theorem 1.2.1, *(iii)*) then implies that there exists a stable $(\mathscr{M}_n^{(h)})_{n\in\mathbb{N}}$-driven VARMA model for $\boldsymbol{\varepsilon}^{(h)}$ with real-valued coefficient matrices. It has been shown in Mokkadem (1988, Theorem 1) that a stable vector ARMA process is geometrically completely regular provided that the driving noise sequence is i. i. d. and absolutely continuous with respect to the Lebesgue measure. A careful analysis of the proof of this result shows that the existence of an absolutely continuous component of the law of the driving noise is already sufficient for the conclusion to hold. We briefly comment on the necessary modifications to the argument. We first note that under these weaker assumptions, the proof of Mokkadem (1988, Lemma 3) implies that the n-step transition probabilities $P^n(\boldsymbol{x},\cdot)$ of the Markov chain X associated with a vector ARMA model via its state space representation have an absolutely continuous component for all n greater than or equal to some n_0. This immediately implies aperiodicity and ϕ-irreducibility of X, where ϕ can be taken as the Lebesgue measure restricted to the support of the continuous component of $P^{n_0}(\boldsymbol{x},\cdot)$. The rest of the proof, in particular the verification of the Foster–Lyapunov drift condition for complete regularity, is unaltered. This shows that $\boldsymbol{\varepsilon}^{(h)}$ is geometrically completely regular and, in particular, strongly mixing with exponentially decaying mixing coefficients. □

Proof (of Lemma 2.9) By definition, $\boldsymbol{M}_\nu^{(h)} = \boldsymbol{M}_\nu(h)$, where $(\boldsymbol{M}_\nu(t))_{t\geqslant 0}$ is the solution to

$$\mathrm{d}\boldsymbol{M}_\nu(t) = \lambda_\nu \boldsymbol{M}_\nu(t)\mathrm{d}t + \mathrm{d}\boldsymbol{L}(t), \quad \boldsymbol{M}_\nu(0) = \mathbf{0}_m.$$

Taking the real and imaginary parts of this equation gives

$$\begin{aligned}
\mathrm{d}\,\mathrm{Re}\,\boldsymbol{M}_\nu(t) &= \mathrm{Re}\,\lambda_\nu \boldsymbol{M}_\nu(t)\mathrm{d}t + \mathrm{d}\boldsymbol{L}(t) = \left[\mathrm{Re}\,\lambda_\nu\,\mathrm{Re}\,\boldsymbol{M}_\nu(t) - \mathrm{Im}\,\lambda_\nu\,\mathrm{Im}\,\boldsymbol{M}_\nu(t)\right]\mathrm{d}t + \mathrm{d}\boldsymbol{L}(t),\\
\mathrm{d}\,\mathrm{Im}\,\boldsymbol{M}_\nu(t) &= \mathrm{Im}\,\lambda_\nu \boldsymbol{M}_\nu(t)\mathrm{d}t \qquad\qquad\;\; = \left[\mathrm{Re}\,\lambda_\nu\,\mathrm{Im}\,\boldsymbol{M}_\nu(t) + \mathrm{Im}\,\lambda_\nu\,\mathrm{Re}\,\boldsymbol{M}_\nu(t)\right]\mathrm{d}t,
\end{aligned}$$

and, consequently,

$$\mathrm{d}\begin{pmatrix} \mathrm{Re}\,\boldsymbol{M}_\nu(t) \\ \mathrm{Im}\,\boldsymbol{M}_\nu(t) \end{pmatrix} = [\Lambda_\nu \otimes \mathbf{1}_m] \begin{pmatrix} \mathrm{Re}\,\boldsymbol{M}_\nu(t) \\ \mathrm{Im}\,\boldsymbol{M}_\nu(t) \end{pmatrix}\mathrm{d}t + \begin{pmatrix} \mathbf{1}_m \\ 0_m \end{pmatrix}\mathrm{d}\boldsymbol{L}(t),$$

where

$$\Lambda_\nu = \begin{pmatrix} \mathrm{Re}\,\lambda_\nu & -\mathrm{Im}\,\lambda_\nu \\ \mathrm{Im}\,\lambda_\nu & \mathrm{Re}\,\lambda_\nu \end{pmatrix}$$

is the matrix representation of the complex number λ_ν. Using the fact that $\lambda_\nu \in \mathbb{R}$ for $\nu = 1,\ldots,r$, and $\lambda_\nu = \overline{\lambda_{\nu+1}} \in \mathbb{C}\backslash\mathbb{R}$ for $\nu = r+1, r+3, \ldots, N-1$, it follows that $\mathscr{M}^{(h)} = \mathscr{M}(h)$, where $(\mathscr{M}(t))_{t\geqslant 0}$ satisfies $\mathrm{d}\mathscr{M}(t) = G\mathscr{M}(t)\mathrm{d}t + H\mathrm{d}\boldsymbol{L}(t)$, and $G = \tilde{G}\otimes \mathbf{1}_m \in M_{mN}(\mathbb{R})$ and $H = \tilde{H}\otimes\mathbf{1}_m \in M_{mN,m}$ are given by

$$\begin{aligned}
\tilde{G} &= \mathrm{diag}(\lambda_1,\ldots,\lambda_r,\Lambda_{r+1},\Lambda_{r+3},\ldots,\Lambda_{N-1}),\\
\tilde{H} &= (\underbrace{1 \;\cdots\; 1}_{r\text{ times}} \;\; 1 \;\; 0 \;\; 1 \;\; 0 \;\; \cdots \;\; 1 \;\; 0\,)^T.
\end{aligned}$$

Since $\operatorname{rank} H = m$, the first claim of the lemma is proved. Next, we show that the controllability matrix $\mathscr{C} := \begin{bmatrix} H & GH & \cdots & G^{N-1}H \end{bmatrix} \in M_{mN}(\mathbb{R})$ is non-singular. With $\widetilde{\mathscr{C}} := \begin{bmatrix} \tilde{H} & \tilde{G}\tilde{H} & \cdots & \tilde{G}^{N-1}\tilde{H} \end{bmatrix}$ and by the properties of the Kronecker product, it follows that $\mathscr{C} = \widetilde{\mathscr{C}} \otimes \mathbf{1}_m$ and thus $\det \mathscr{C} = [\det \widetilde{\mathscr{C}}]^m$. The matrix $\widetilde{\mathscr{C}}$ is given explicitly by

$$\widetilde{\mathscr{C}} = \begin{pmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{N-1} \\ \vdots & & & & \vdots \\ 1 & \lambda_r & \lambda_r^2 & \cdots & \lambda_r^{N-1} \\ 1 & \operatorname{Re}\lambda_{r+1} & \operatorname{Re}\lambda_{r+1}^2 & \cdots & \operatorname{Re}\lambda_{r+1}^{N-1} \\ 0 & \operatorname{Im}\lambda_{r+1} & \operatorname{Im}\lambda_{r+1}^2 & \cdots & \operatorname{Im}\lambda_{r+1}^{N-1} \\ \vdots & & & & \vdots \\ 1 & \operatorname{Re}\lambda_{N-1} & \operatorname{Re}\lambda_{N-1}^2 & \cdots & \operatorname{Re}\lambda_{N-1}^{N-1} \\ 0 & \operatorname{Im}\lambda_{N-1} & \operatorname{Im}\lambda_{N-1}^2 & \cdots & \operatorname{Im}\lambda_{N-1}^{N-1} \end{pmatrix} = T \begin{pmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{N-1} \\ \vdots & & & & \vdots \\ 1 & \lambda_r & \lambda_r^2 & \cdots & \lambda_r^{N-1} \\ 1 & \lambda_{r+1} & \lambda_{r+1}^2 & \cdots & \lambda_{r+1}^{N-1} \\ i & i\overline{\lambda_{r+1}} & i\overline{\lambda_{r+1}^2} & \cdots & i\overline{\lambda_{r+1}^{N-1}} \\ \vdots & & & & \vdots \\ 1 & \lambda_{N-1} & \lambda_{N-1}^2 & \cdots & \lambda_{N-1}^{N-1} \\ i & i\overline{\lambda_{N-1}} & i\overline{\lambda_{N-1}^2} & \cdots & i\overline{\lambda_{N-1}^{N-1}} \end{pmatrix}$$

with $T \in M_N(\mathbb{R})$ given by $T = \operatorname{diag}(1, \ldots, 1, R, \ldots, R)$ and $R = \frac{1}{2}\begin{pmatrix} 1 & -\mathrm{i} \\ -\mathrm{i} & 1 \end{pmatrix}$. The formula for the determinant of a Vandermonde matrix (Bernstein, 2005, Fact 5.13.3) implies that

$$\det \mathscr{C} = \left[(-1)^{\frac{N-r}{2}} \prod_{1 \leqslant \mu < \nu \leqslant r} (\lambda_\mu - \lambda_\nu) \prod_{\substack{\mu,\nu \in I_{r,N} \\ \mu < \nu}} \operatorname{Im}\lambda_\mu |\lambda_\mu - \lambda_\nu|^2 |\overline{\lambda_\mu} - \lambda_\nu|^2 \prod_{\substack{1 \leqslant \mu \leqslant r \\ \nu \in I_{r,N}}} |\lambda_\mu - \lambda_\nu|^2 \right]^m,$$

where $I_{r,N} := \{r+1, r+3, \ldots, N-1\}$. Hence, $\det \mathscr{C}$ is not zero by Assumption E2, and the proof is complete. □

3. Quasi Maximum Likelihood Estimation for Strongly Mixing Linear State Space Models and Multivariate CARMA Processes

3.1. Introduction

Linear state space models have been used in time series analysis and stochastic modelling for many decades because of their wide applicability and analytical tractability (see, e. g., Brockwell and Davis, 1991; Hamilton, 1994, for a detailed account). In discrete time they are defined by the equations

$$\boldsymbol{X}_n = F\boldsymbol{X}_{n-1} + \boldsymbol{Z}_{n-1}, \quad \boldsymbol{Y}_n = H\boldsymbol{X}_n + \boldsymbol{W}_n, \quad n \in \mathbb{Z}, \tag{3.1.1}$$

where $\boldsymbol{X} = (\boldsymbol{X}_n)_{n\in\mathbb{Z}}$ is a latent state process, F, H are coefficient matrices, and $\boldsymbol{Z} = (\boldsymbol{Z}_n)_{n\in\mathbb{Z}}$, $\boldsymbol{W} = (\boldsymbol{W}_n)_{n\in\mathbb{Z}}$ are sequences of random variables, see Definition 3.1 for a precise formulation of this model. In this chapter we investigate the problem of estimating the coefficient matrices F, H as well as the covariances of $\boldsymbol{Z}$ and $\boldsymbol{W}$ from a sample of observed values of the output process $\boldsymbol{Y} = (\boldsymbol{Y}_n)_{n\in\mathbb{Z}}$, using a quasi maximum likelihood (QML) or generalized least squares approach. Given the importance of this problem in practice, it is surprising that a proper mathematical analysis of the quasi maximum likelihood estimation for the model (3.1.1) has only been performed in cases where the model is in the so-called innovations form

$$\boldsymbol{X}_n = F\boldsymbol{X}_{n-1} + K\boldsymbol{\varepsilon}_{n-1}, \quad \boldsymbol{Y}_n = H\boldsymbol{X}_n + \boldsymbol{\varepsilon}_n, \quad n \in \mathbb{Z}, \tag{3.1.2}$$

where the innovations $\boldsymbol{\varepsilon}$ form a martingale difference sequence (Hannan and Deistler, 1988, Chapter 4). This includes state space models in which the noise sequences $\boldsymbol{Z}, \boldsymbol{W}$ are Gaussian, because then the innovations, which are uncorrelated by definition, form an i. i. d. sequence. Restriction to these special cases excludes, however, the state space representations of aggregated linear processes, as well as of equidistantly observed continuous-time linear state space models.

In the first part of the present chapter we shall prove consistency and asymptotic normality of the quasi maximum likelihood estimator for the general linear state space model (3.1.1) under the assumptions that the noise sequences $\boldsymbol{Z}, \boldsymbol{W}$ are ergodic, and that the output process $\boldsymbol{Y}$ satisfies a strong-mixing condition in the sense of Rosenblatt (1956). This assumption is not very restrictive, and is, in particular, satisfied if the noise sequence $\boldsymbol{Z}$ is i. i. d. with an absolutely continuous component, and $\boldsymbol{W}$ is strongly mixing. Our results are a multivariate generalization of Francq and Zakoïan (1998), who considered the quasi maximum likelihood estimation for univariate strongly mixing ARMA processes. The very recent paper Boubacar Mainassara and Francq (2011), which deals with the structural estimation of weak vector ARMA processes, instead makes a mixing assumption about the innovations sequence $\boldsymbol{\varepsilon}$ of the process under consideration, which is very difficult to verify for state space models; their results can therefore not be used for the estimation of general discretely-observed linear continuous-time state space models. More importantly, their proof appears to be incomplete, because a crucial step in the proof of their Lemma 4 is claimed by the authors to be analogous to the corresponding step in the proof of Francq and Zakoïan (1998, Lemma 3). It is, however, not clear how the argument given there can be modified in order to be compatible with the assumption of strongly mixing innovations, which is weaker than the assumption of a strongly mixing output process as employed in Francq and Zakoïan (1998).

As alluded to above, one advantage of relaxing the assumption of i. i. d. innovations in a discrete-time state space model is the inclusion of sampled continuous-time state space models. These were introduced in the form continuous-time ARMA (CARMA) models in Doob (1944) as stochastic processes satisfying the formal analogue of the familiar autoregressive moving average equations of discrete-time ARMA processes, namely

$$a(\mathrm{D})Y(t) = b(\mathrm{D})\mathrm{D}W(t), \quad \mathrm{D} = \frac{\mathrm{d}}{\mathrm{d}t}, \tag{3.1.3}$$

where a and b are suitable polynomials, and W denotes a Brownian motion. In the recent past, a considerable body of research has been devoted to these processes (see, e. g., Brockwell, 2001a, and references therein). One particularly important extension of the model (3.1.3) was introduced in Brockwell (2001b), where the driving Brownian motion was replaced by a Lévy process with finite logarithmic moments. This allowed for a wide range of possibly heavy-tailed marginal distribution of the process Y as well as the occurrence of jumps in the sample paths, both characteristic features of many observed time series, e. g. in finance (Cont, 2001). Recently, Marquardt and Stelzer (2007) further generalized Eq. (3.1.3) to the multivariate setting, which gave researchers the possibility to model several dependent time series jointly by one linear continuous-time process. This extension is important, because many time series exhibit strong dependencies and can therefore not be modelled adequately on an individual basis. In that paper, the multivariate non-Gaussian equivalent of Eq. (3.1.3), namely $P(\mathrm{D})\boldsymbol{Y}(t) = Q(\mathrm{D})\mathrm{D}\boldsymbol{L}(t)$, for matrix-valued polynomials P and Q and a Lévy process $\boldsymbol{L}$, was interpreted by spectral techniques as a continuous-time state space model of the form

$$\mathrm{d}\boldsymbol{G}(t) = \mathcal{A}\boldsymbol{G}(t)\mathrm{d}t + \mathcal{B}\mathrm{d}\boldsymbol{L}(t), \quad \boldsymbol{Y}(t) = \mathcal{C}\boldsymbol{G}(t); \tag{3.1.4}$$

see Eq. (3.3.6) for an expression of the matrices $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$. The structural similarity between Eq. (3.1.1) and Eq. (3.1.4) is apparent, and it is essential for many of our arguments. Taking a different route, multivariate CARMA processes can be defined as the continuous-time analogue of discrete-time vector ARMA models, described in detail in Hannan and Deistler (1988); Lütkepohl (2005). As continuous-time processes, CARMA processes are suited particularly well to model irregularly spaced and high-frequency data, which makes them a flexible and efficient tool for building stochastic models of time series arising in the natural sciences, engineering and finance (e. g. Benth and Šaltytė Benth, 2009; Fan, Söderström, Mossberg, Carlsson and Zon, 1998; Na and Rhee, 2002; Todorov and Tauchen, 2006).

In the univariate Gaussian setting, several different approaches to the estimation problem of CARMA processes have been investigated (see, e. g., Larsson, Mossberg and Söderström, 2006; Nielsen, Madsen and Young, 2000, and references therein). Maximum likelihood estimation based on a continuous record was considered in Brown and Hewitt (1975); Feigin (1976); Pham (1977). Due to the fact that processes are typically not observed continuously and the limitations of digital computer processing, inference based on discrete observations has become more important in recent years; these approaches include variants of the Yule–Walker algorithm for time-continuous autoregressive processes (Hyndman, 1993), maximum likelihood methods (Brockwell et al., 2011; Duncan, Mandl and Pasik-Duncan, 1999), and randomized sampling (Leneman and Lewis, 1966; Rivoira, Moudden and Fleury, 2002) to overcome the aliasing problem. Alternative methods include discretization of the differential operator (Larsson and Söderström, 2002; Söderström, Fan, Carlsson and Mossberg, 1997), and spectral estimation (Gillberg and Ljung, 2009; Lahalle, Fleury and Rivoira, 2004; Lii and Masry, 1995; Masry, 1978). For the special case of Ornstein–Uhlenbeck processes, least squares and moment estimators have also been investigated without the assumptions of Gaussianity (Hu and Long, 2009; Spiliopoulos, 2009).

In the second part of this chapter we consider the estimation of general multivariate CARMA processes with finite second moments based on equally spaced discrete observations, exploiting the results about the quasi maximum likelihood estimation of general linear discrete-time state space models. Under natural identifiability assumptions we obtain strongly consistent and asymptotically normal estimators for the coefficient matrices of a second-order MCARMA process and the covariance matrix of the driving Lévy process, which together determine the second-order structure of the process. It is a natural restriction of the quasi maximum likelihood method that distributional properties of the driving Lévy process which are not determined by its covariance matrix cannot be estimated. However, once the autoregressive and moving average coefficients of a CARMA process are (approximately) known, and if high-frequency observations are available, a parametric model for the driving Lévy process can be estimated by the methods described in Chapter 4.

Outline of the chapter The organization of the chapter is as follows. In Section 3.2 we develop a quasi maximum likelihood estimation theory for general discrete-time linear stochastic state space models with finite second moments. In Section 3.2.1 we precisely define the class of linear stochastic state space models as well as the quasi maximum likelihood estimator. The following two sections 3.2.3 and 3.2.4 contain the proofs that, under a set of technical conditions, this estimator is strongly consistent and asymptotically normally distributed as the number of observations tends to infinity, see Theorems 3.7 and 3.8.

In Section 3.3 we use the results from Section 3.2 to establish asymptotic properties of a quasi maximum likelihood estimator for multivariate CARMA processes which are observed on a fixed equidistant time grid. As a first step, we review in Section 3.3.1 their definition as well as their relation to the class of continuous-time state space models. This is followed by an investigation of the probabilistic properties of a sampled MCARMA process in Section 3.3.2 and an analysis of the important issue of identifiability in Section 3.3.3. Finally, we are able to state and prove our main result, Theorem 3.50, about the strong consistency and asymptotic normality of the quasi maximum likelihood estimator for equidistantly sampled multivariate CARMA processes in Section 3.3.4.

In the final Section 3.4, we present canonical parametrizations, and we demonstrate the applicability of the quasi maximum likelihood estimation for continuous-time state space models with a simulation study and a data example from economics.

Notation We use the following notation: The space of $m \times n$ matrices with entries in the ring $\mathbb{K}$ is denoted by $M_{m,n}(\mathbb{K})$ or $M_m(\mathbb{K})$ if $m = n$. The set of symmetric matrices is denoted by $\mathbb{S}_m(\mathbb{K})$, and the symbols $\mathbb{S}_m^+(\mathbb{R})$ ($\mathbb{S}_m^{++}(\mathbb{R})$) stand for the subsets of positive semidefinite (positive definite) matrices, respectively. A^T denotes the transpose of the matrix A, $\operatorname{im} A$ its image, $\ker A$ its kernel, $\sigma(A)$ its spectrum, and $\mathbf{1}_m \in M_m(\mathbb{K})$ is the identity matrix. The vector space $\mathbb{R}^m$ is identified with $M_{m,1}(\mathbb{R})$ so that $\boldsymbol{u} = (u^1, \ldots, u^m)^T \in \mathbb{R}^m$ is a column vector. $\|\cdot\|$ represents the Euclidean norm, $\langle \cdot, \cdot \rangle$ the Euclidean inner product, and $\mathbf{0}_m \in \mathbb{R}^m$ the zero vector. $\mathbb{K}[X]$ ($\mathbb{K}\{X\}$) denotes the ring of polynomial (rational) expressions in X over $\mathbb{K}$, $I_B(\cdot)$ the indicator function of the set B, and $\delta_{n,m}$ the Kronecker symbol. The symbols $\mathbb{E}$, $\mathbb{V}$ar, and $\mathbb{C}$ov stand for the expectation, variance and covariance operators, respectively. Finally, we write ∂_m for the partial derivative operator with respect to the mth coordinate and $\nabla = \begin{pmatrix} \partial_1 & \cdots & \partial_r \end{pmatrix}$ for the gradient operator. When there is no ambiguity, we use $\partial_m f(\boldsymbol{\vartheta}_0)$ and $\nabla_{\boldsymbol{\vartheta}} f(\boldsymbol{\vartheta}_0)$ as shorthands for $\partial_m f(\boldsymbol{\vartheta})|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0}$ and $\nabla_{\boldsymbol{\vartheta}} f(\boldsymbol{\vartheta})|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0}$, respectively. A generic constant, the value of which may change from line to line, is denoted by C.

3.2. Quasi maximum likelihood estimation for discrete-time state space models

In this section we investigate quasi maximum likelihood (QML) estimation for general linear stochastic state space models in discrete time, and prove consistency and asymptotic normality under a set of technical assumptions described below. On the one hand, due to the wide applicability of state space systems in stochastic modelling and control, these results are interesting and useful in their own right. In the present chapter they will be applied in Section 3.3 to prove asymptotic properties of the QML estimator for discretely observed multivariate continuous-time ARMA processes. This is possible because every such process has a natural discrete-time state space structure, see Proposition 3.32.

Our theory extends existing results from the literature, in particular concerning the QML estimation of Gaussian state space models (Ansley and Kohn, 1985; Jones, 1980; Stoffer and Wall, 1991), of state space models whose innovations sequences are martingale differences (Hannan, 1969, 1975; Reinsel, 1997), and of weak univariate ARMA processes which satisfy a strong-mixing condition (Francq and Zakoïan, 1998). The techniques used in this section are based on Boubacar Mainassara and Francq (2011), who consider the estimation of structural discrete-time vector ARMA models with strongly mixing innovations.

3.2.1. Preliminaries and definition of the QML estimator

The general linear stochastic state space model for which we will develop a quasi maximum likelihood estimation theory is defined as follows.

Definition 3.1 (State space model) An $\mathbb{R}^d$-valued discrete-time linear stochastic state space model $(F, H, \boldsymbol{Z}, \boldsymbol{W})$ of dimension N is characterized by a strictly stationary $\mathbb{R}^{N+d}$-valued sequence $\begin{pmatrix} \boldsymbol{Z}^T & \boldsymbol{W}^T \end{pmatrix}^T$ with mean zero and finite covariance matrix

$$\mathbb{E}\left[\begin{pmatrix} \boldsymbol{Z}_n \\ \boldsymbol{W}_n \end{pmatrix}\begin{pmatrix} \boldsymbol{Z}_m^T & \boldsymbol{W}_m^T \end{pmatrix}\right] = \delta_{m,n}\begin{pmatrix} Q & R \\ R^T & S \end{pmatrix}, \quad n, m \in \mathbb{Z}, \tag{3.2.1}$$

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

$$\boldsymbol{X}_n = F\boldsymbol{X}_{n-1} + \boldsymbol{Z}_{n-1}, \quad n \in \mathbb{Z}, \tag{3.2.2a}$$

and an observation equation

$$\boldsymbol{Y}_n = H\boldsymbol{X}_n + \boldsymbol{W}_n, \quad n \in \mathbb{Z}. \tag{3.2.2b}$$

The $\mathbb{R}^N$-valued autoregressive process $\boldsymbol{X} = (\boldsymbol{X}_n)_{n\in\mathbb{Z}}$ is called the *state vector process*, and

$\boldsymbol{Y} = (\boldsymbol{Y}_n)_{n\in\mathbb{Z}}$ is called the *output process*.

The assumption that the processes $\boldsymbol{Z}$ and $\boldsymbol{W}$ are centred is not essential for our results, but simplifies the notation considerably. The following lemma collects important probabilistic properties of the output process $\boldsymbol{Y}$ of such a state space model. Its proof is standard (Brockwell and Davis, 1991, §12.1).

Lemma 3.2 *Assume that $(F, H, \boldsymbol{Z}, \boldsymbol{W})$ is a state space model according to Definition 3.1, and that the eigenvalues of F are less than unity in absolute value.*

i) There exists a unique stationary process $\boldsymbol{Y}$ solving Eqs. (3.2.2). This process has the moving average representation $\boldsymbol{Y}_n = \boldsymbol{W}_n + H\sum_{\nu=1}^{\infty} F^{\nu-1}\boldsymbol{Z}_{n-\nu}$.

ii) If both $\boldsymbol{Z}$ and $\boldsymbol{W}$ have a finite kth moments for some $k > 0$, then $\boldsymbol{Y}$ has finite kth moments as well.

iii) If the expected value of $\boldsymbol{Y}_n$ is finite, it is given by $\mathbb{E}\boldsymbol{Y}_n = \mathbb{E}\boldsymbol{W}_1 + H\sum_{\nu=1}^{\infty} F^{\nu-1}\mathbb{E}\boldsymbol{Z}_1$. In particular, if both $\boldsymbol{Z}$ and $\boldsymbol{W}$ have mean zero, then $\boldsymbol{Y}$ has mean zero as well.

Before we turn our attention to the estimation problem for this class of state space models, we review the necessary aspects of the theory of Kalman filtering, see Kalman (1960) for the original control-theoretic account and Brockwell and Davis (1991, §12.2) for a treatment in the context of time series analysis. The linear innovations of the output process $\boldsymbol{Y}$, which are introduced in the following definition, are of particular importance for the quasi maximum likelihood estimation of state space models. Intuitively, their role in quasi maximum likelihood estimation is similar to the one played by the residuals in least squares estimation.

Definition 3.3 (Linear innovations) Let $\boldsymbol{Y} = (\boldsymbol{Y}_n)_{n\in\mathbb{Z}}$ be an $\mathbb{R}^d$-valued stationary stochastic process with finite second moments. The *linear innovations* $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_n)_{n\in\mathbb{Z}}$ of $\boldsymbol{Y}$ are then defined by

$$\boldsymbol{\varepsilon}_n = \boldsymbol{Y}_n - P_{n-1}\boldsymbol{Y}_n, \quad P_n = \text{orthogonal projection onto } \overline{\operatorname{span}}\left\{\boldsymbol{Y}_\nu : -\infty < \nu \leqslant n\right\}, \tag{3.2.3}$$

where the closure is taken in the Hilbert space of square-integrable random variables with inner product $(X, Y) \mapsto \mathbb{E}\langle X, Y\rangle$.

This definition immediately implies that the innovations $\boldsymbol{\varepsilon}$ of a stationary stochastic process $\boldsymbol{Y}$ are stationary and uncorrelated. The following proposition is a combination of Brockwell and Davis (1991, Proposition 12.2.3) and Hamilton (1994, Proposition 13.2).

Proposition 3.4 *Assume that $\boldsymbol{Y}$ is the output process of the state space model (3.2.2), that at least one of the matrices Q and S is positive definite, and that the absolute values of the eigenvalues of F are less than unity. Then the following hold.*

i) The discrete-time algebraic Riccati equation

$$\Omega = F\Omega F^T + Q - \left[F\Omega H^T + R\right] \left[H\Omega H^T + S\right]^{-1} \left[F\Omega H^T + R\right]^T \tag{3.2.4}$$

has a unique positive semidefinite solution $\Omega \in \mathbb{S}_N^+(\mathbb{R})$.

ii) The absolute values of the eigenvalues of the matrix $F - KH \in M_N(\mathbb{R})$ are less than one, where

$$K = \left[F\Omega H^T + R\right] \left[H\Omega H^T + S\right]^{-1} \in M_{N,d}(\mathbb{R}) \tag{3.2.5}$$

is the steady-state Kalman gain matrix.

iii) The linear innovations $\boldsymbol{\varepsilon}$ of $\boldsymbol{Y}$ are the unique stationary solution to

$$\hat{\boldsymbol{X}}_n = (F - KH)\,\hat{\boldsymbol{X}}_{n-1} + K\boldsymbol{Y}_{n-1}, \quad \boldsymbol{\varepsilon}_n = \boldsymbol{Y}_n - H\hat{\boldsymbol{X}}_n, \quad n \in \mathbb{Z}. \tag{3.2.6a}$$

Using the backshift operator B, which is defined by $\mathrm{B}\,\boldsymbol{Y}_n = \boldsymbol{Y}_{n-1}$, this can be written equivalently as

$$\begin{aligned}\boldsymbol{\varepsilon}_n =& \left\{\mathbf{1}_d - H\left[\mathbf{1}_N - (F - KH)\,\mathrm{B}\right]^{-1} K\,\mathrm{B}\right\}\boldsymbol{Y}_n \\ =& \boldsymbol{Y}_n - H\sum_{\nu=1}^{\infty}(F - KH)^{\nu-1}K\boldsymbol{Y}_{n-\nu}, \qquad n \in \mathbb{Z}.\end{aligned} \tag{3.2.6b}$$

The covariance matrix $V = \mathbb{E}\boldsymbol{\varepsilon}_n\boldsymbol{\varepsilon}_n^T \in \mathbb{S}_d^+(\mathbb{R})$ of the innovations $\boldsymbol{\varepsilon}$ is given by

$$V = \mathbb{E}\boldsymbol{\varepsilon}_n\boldsymbol{\varepsilon}_n^T = H\Omega H^T + S. \tag{3.2.7}$$

iv) The process $\boldsymbol{Y}$ has the innovations representation

$$\hat{\boldsymbol{X}}_n = F\boldsymbol{X}_{n-1} + K\boldsymbol{\varepsilon}_{n-1}, \quad \boldsymbol{Y}_n = H\boldsymbol{X}_n + \boldsymbol{\varepsilon}_n, \quad n \in \mathbb{Z}, \tag{3.2.8a}$$

which, similar to Eqs. (3.2.6), allows for the moving average representation

$$\begin{aligned}\boldsymbol{Y}_n =& \left\{\mathbf{1}_d - H\left[\mathbf{1}_N - F\,\mathrm{B}\right]^{-1} K\,\mathrm{B}\right\}\boldsymbol{Y}_n \\ =& \boldsymbol{\varepsilon}_n + H\sum_{\nu=1}^{\infty}F^{\nu-1}K\boldsymbol{\varepsilon}_{n-\nu}, \qquad n \in \mathbb{Z}.\end{aligned} \tag{3.2.8b}$$

We now consider parametric families of state space models. For some parameter space $\Theta \subset \mathbb{R}^r$, $r \in \mathbb{N}$, the mappings

$$F_{(\cdot)} : \Theta \to M_N(\mathbb{R}), \qquad H_{(\cdot)} : \Theta \to M_{d,N}, \tag{3.2.9a}$$

together with a collection of strictly stationary stochastic processes $\boldsymbol{Z}_{\boldsymbol{\vartheta}}$, $\boldsymbol{W}_{\boldsymbol{\vartheta}}$, $\boldsymbol{\vartheta} \in \Theta$, with finite second moments determine a parametric family $(F_{\boldsymbol{\vartheta}}, H_{\boldsymbol{\vartheta}}, \boldsymbol{Z}_{\boldsymbol{\vartheta}}, \boldsymbol{W}_{\boldsymbol{\vartheta}})_{\boldsymbol{\vartheta}\in\Theta}$ of linear state space models according to Definition 3.1. For the variance and covariance matrices of the noise sequences $\boldsymbol{Z}, \boldsymbol{W}$ we use the notation (cf. Eq. (3.2.1)) $Q_{\boldsymbol{\vartheta}} = \mathbb{E}\boldsymbol{Z}_{\boldsymbol{\vartheta},n}\boldsymbol{Z}_{\boldsymbol{\vartheta},n}^T$, $S_{\boldsymbol{\vartheta}} = \mathbb{E}\boldsymbol{W}_{\boldsymbol{\vartheta},n}\boldsymbol{W}_{\boldsymbol{\vartheta},n}^T$, and $R_{\boldsymbol{\vartheta}} = \mathbb{E}\boldsymbol{Z}_{\boldsymbol{\vartheta},n}\boldsymbol{W}_{\boldsymbol{\vartheta},n}^T$, which defines the functions

$$Q_{(\cdot)} : \Theta \to \mathbb{S}_N^+(\mathbb{R}), \qquad S_{(\cdot)} : \Theta \to \mathbb{S}_d^+, \qquad R_{(\cdot)} : \Theta \to M_{N,d}(\mathbb{R}). \tag{3.2.9b}$$

It is well known (Brockwell and Davis, 1991, Eq. (11.5.4)) that for this model, minus twice the logarithm of the Gaussian likelihood of $\boldsymbol{\vartheta}$ based on a sample $\boldsymbol{y}^L = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_L)$ of observations can be written as

$$\mathscr{L}(\boldsymbol{\vartheta}, \boldsymbol{y}^L) = \sum_{n=1}^{L} l_{\boldsymbol{\vartheta},n} = \sum_{n=1}^{L} \left[d \log 2\pi + \log\det V_{\boldsymbol{\vartheta}} + \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}^T V_{\boldsymbol{\vartheta}}^{-1} \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n} \right], \tag{3.2.10}$$

where $\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}$ and $V_{\boldsymbol{\vartheta}}$ are given by analogues of Eqs. (3.2.6a) and (3.2.7), namely

$$\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n} = \left\{ \mathbf{1}_d - H_{\boldsymbol{\vartheta}} \left[\mathbf{1}_N - (F_{\boldsymbol{\vartheta}} - K_{\boldsymbol{\vartheta}} H_{\boldsymbol{\vartheta}})\, \mathrm{B} \right]^{-1} K_{\boldsymbol{\vartheta}}\, \mathrm{B} \right\} \boldsymbol{Y}_n, \quad n \in \mathbb{Z}, \tag{3.2.11a}$$

$$V_{\boldsymbol{\vartheta}} = H_{\boldsymbol{\vartheta}} \Omega_{\boldsymbol{\vartheta}} H_{\boldsymbol{\vartheta}}^T + S_{\boldsymbol{\vartheta}}, \tag{3.2.11b}$$

and $K_{\boldsymbol{\vartheta}}, \Omega_{\boldsymbol{\vartheta}}$ are defined in the same way as K, Ω in Eqs. (3.2.4) and (3.2.5). In the following we always assume that $\boldsymbol{y}^L = (\boldsymbol{Y}_{\boldsymbol{\vartheta}_0,1}, \ldots, \boldsymbol{Y}_{\boldsymbol{\vartheta}_0,L})$ is a sample from the output process of the state space model $(F_{\boldsymbol{\vartheta}_0}, H_{\boldsymbol{\vartheta}_0}, \boldsymbol{Z}_{\boldsymbol{\vartheta}_0}, \boldsymbol{W}_{\boldsymbol{\vartheta}_0})$ corresponding to the parameter value $\boldsymbol{\vartheta}_0$. We therefore call $\boldsymbol{\vartheta}_0$ the *true parameter value*. It is important to note that $\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0}$ are the true innovations of $\boldsymbol{Y}_{\boldsymbol{\vartheta}_0}$, and that therefore $\mathbb{E}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,n}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,n}^T = V_{\boldsymbol{\vartheta}_0}$, but that this relation fails to hold for other values of $\boldsymbol{\vartheta}$. This is due to the fact that $\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}}$ is not the true innovations sequence of the state space model corresponding to the parameter value $\boldsymbol{\vartheta}$, which could only be computed from knowledge of the fictitious output process $\boldsymbol{Y}_{\boldsymbol{\vartheta}}$, which is not observed. We therefore call the sequence $\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}}$ *pseudo-innovations*.

The goal of this section is to investigate how the value $\boldsymbol{\vartheta}_0$ can be estimated from $\boldsymbol{y}^L$ by maximizing Eq. (3.2.10). The first difficulty one is confronted with is that the pseudo-innovations $\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}}$ are defined in terms of the full history of the process $\boldsymbol{Y} = \boldsymbol{Y}_{\boldsymbol{\vartheta}_0}$, which is not observed. It is therefore necessary to use an approximation to these innovations which can be computed from the finite sample $\boldsymbol{y}^L$. One such approximation is obtained if, instead of using the steady-state Kalman filter described in Proposition 3.4, one initializes the filter at $n = 1$ with some prescribed values. More precisely, we define the approximate pseudo-innovations $\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{\vartheta}}$ via the recursion

$$\hat{\boldsymbol{X}}_{\boldsymbol{\vartheta},n} = (F_{\boldsymbol{\vartheta}} - K_{\boldsymbol{\vartheta}} H_{\boldsymbol{\vartheta}})\, \hat{\boldsymbol{X}}_{\boldsymbol{\vartheta},n-1} + K_{\boldsymbol{\vartheta}} \boldsymbol{Y}_{n-1}, \quad \hat{\boldsymbol{\varepsilon}}_{\boldsymbol{\vartheta},n} = \boldsymbol{Y}_n - H_{\boldsymbol{\vartheta}} \hat{\boldsymbol{X}}_{\boldsymbol{\vartheta},n}, \quad n \in \mathbb{N}, \tag{3.2.12}$$

and the prescription $\hat{X}_{\vartheta,1} = \hat{X}_{\vartheta,\text{initial}}$. The initial values $\hat{X}_{\vartheta,\text{initial}}$ are usually either sampled from the stationary distribution of X_ϑ, if that is possible, or set to some deterministic value. Alternatively, one can additionally define a positive semidefinite matrix $\Omega_{\vartheta,\text{initial}}$ and compute Kalman gain matrices $K_{\vartheta,n}$ recursively via Brockwell and Davis (1991, Eq. (12.2.6)). While this procedure might be advantageous for small sample sizes, the computational burden is significantly smaller when the steady-state Kalman gain is used. The asymptotic properties which we are dealing with in this chapter are expected to be the same for both choices because the Kalman gain matrices $K_{\vartheta,n}$ converge to their steady state values as n tends to infinity (Hamilton, 1994, Proposition 13.2).

The quasi maximum likelihood estimator $\hat{\vartheta}^L$ for the parameter ϑ based on the sample y^L is defined as

$$\hat{\vartheta}^L = \operatorname{argmin}_{\vartheta\in\Theta} \widehat{\mathscr{L}}(\vartheta, y^L), \tag{3.2.13}$$

where $\widehat{\mathscr{L}}(\vartheta, y^L)$ is obtained from $\mathscr{L}(\vartheta, y^L)$ by substituting $\hat{\varepsilon}_{\vartheta,n}$ from Eq. (3.2.12) for $\varepsilon_{\vartheta,n}$, that is

$$\begin{aligned}\widehat{\mathscr{L}}(\vartheta, y^L) &= \sum_{n=1}^{L} \hat{l}_{\vartheta,n} \\ &= \sum_{n=1}^{L} \left[d\log 2\pi + \log\det V_\vartheta + \hat{\varepsilon}_{\vartheta,n}^T V_\vartheta^{-1} \hat{\varepsilon}_{\vartheta,n} \right].\end{aligned} \tag{3.2.14}$$

3.2.2. Technical assumptions and main results

Our main results about the quasi maximum likelihood estimation for discrete-time state space models are Theorem 3.7, stating that the estimator $\hat{\vartheta}^L$ given by Eq. (3.2.13) is strongly consistent, which means that $\hat{\vartheta}^L$ converges to ϑ_0 almost surely, and Theorem 3.8, which asserts the asymptotic normality of $\hat{\vartheta}^L$ with the usual $L^{1/2}$ scaling. In order to prove these results, we need to impose the following conditions.

Assumption D1 The parameter space Θ is a compact subset of $\mathbb{R}^r$.

Assumption D2 The mappings $F_{(\cdot)}$, $H_{(\cdot)}$, $Q_{(\cdot)}$, $S_{(\cdot)}$, and $R_{(\cdot)}$ in Eqs. (3.2.9) are continuous.

The next condition guarantees that the models under consideration describe stationary processes.

Assumption D3 For every $\vartheta \in \Theta$, the following hold:

i) the eigenvalues of F_ϑ have absolute values less than unity,

ii) at least one of the two matrices Q_ϑ and S_ϑ is positive definite,

iii) the matrix V_ϑ is non-singular.

The next lemma shows that the assertions of Assumption D3 hold in fact uniformly in $\boldsymbol{\vartheta}$.

Lemma 3.5 *Suppose that Assumptions D1 to D3 are satisfied. Then the following hold.*

i) *There exists a positive number $\rho < 1$ such that, for all $\boldsymbol{\vartheta} \in \Theta$, it holds that*

$$\max\left\{|\lambda| : \lambda \in \sigma\left(F_{\boldsymbol{\vartheta}}\right)\right\} \leqslant \rho. \tag{3.2.15a}$$

ii) *There exists a positive number $\rho < 1$ such that, for all $\boldsymbol{\vartheta} \in \Theta$, it holds that*

$$\max\left\{|\lambda| : \lambda \in \sigma\left(F_{\boldsymbol{\vartheta}} - K_{\boldsymbol{\vartheta}} H_{\boldsymbol{\vartheta}}\right)\right\} \leqslant \rho, \tag{3.2.15b}$$

where $K_{\boldsymbol{\vartheta}}$ is defined by Eqs. (3.2.4) and (3.2.5).

iii) *There exists a positive number C such that $\left\|V_{\boldsymbol{\vartheta}}^{-1}\right\| \leqslant C$ for all $\boldsymbol{\vartheta}$.*

Proof Assertion i) is a direct consequence of Assumption D3, i), the assumed smoothness of $\boldsymbol{\vartheta} \mapsto F_{\boldsymbol{\vartheta}}$ (Assumption D2), the compactness of Θ (Assumption D1), and the fact (Bernstein, 2005, Fact 10.11.2) that the eigenvalues of a matrix are continuous functions of its entries. The claim ii) follows with the same argument from Proposition 3.4, ii) and the fact that the solution of a discrete-time algebraic Riccati equation is a continuous function of the coefficient matrices (Lancaster and Rodman, 1995, Chapter 14),(Sun, 1998). Moreover, by Eq. (3.2.7) and what was already said, the function $\boldsymbol{\vartheta} \mapsto V_{\boldsymbol{\vartheta}}$ is continuous, which shows that Assumption D3, iii) holds uniformly in $\boldsymbol{\vartheta}$ as well, and so iii) is proved. □

For the following assumption about the noise sequences $\boldsymbol{Z}$ and $\boldsymbol{W}$ we recall the notion of ergodicity from Durrett (2010, Chapter 6). Assume that $(X_n)_{n\in\mathbb{Z}}$ is a strictly stationary $\mathbb{R}^k$-valued process considered as a random variable on the canonical probability space $\left(\left(\mathbb{R}^k\right)^{\mathbb{Z}}, \mathscr{F}, \mathbb{P}\right)$, where $\mathscr{F} = \mathscr{B}\left(\left(\mathbb{R}^k\right)^{\mathbb{Z}}\right)$ is the Borel σ-algebra. An element $A \in \mathscr{F}$ is said to be shift-invariant if $\mathbb{P}\left(A \triangle \phi^{-1}(A)\right) = 0$, where $\triangle$ denotes the symmetric difference, and

$$\phi : \left(\mathbb{R}^k\right)^{\mathbb{Z}} \to \left(\mathbb{R}^k\right)^{\mathbb{Z}}, \quad (X_n)_{n\in\mathbb{Z}} \mapsto (X_{n+1})_{n\in\mathbb{Z}},$$

is the Bernoulli shift operator. The class of all shift-invariant sets $A \in \mathscr{F}$ is a σ-algebra, denoted by $\mathscr{I}$. The process X is said to be ergodic if $\mathscr{I}$ is the trivial σ-algebra, i. e.,$\mathbb{P}(A) \in \{0, 1\}$ for every $A \in \mathscr{I}$. The fundamental result about ergodic sequences is Birkhoff's Ergodic Theorem (Birkhoff, 1931), which allows for time averages to be approximated by population averages, or more precisely

$$\frac{1}{N}\sum_{n=1}^{N} X_n \xrightarrow[N\to\infty]{} \mathbb{E}\left[X_1 | \mathscr{I}\right],$$

both almost surely and in L^1. In particular, if X is ergodic, the right side equals $\mathbb{E}X_1$.

Assumption D4 The process $\begin{pmatrix} \boldsymbol{W}_{\vartheta_0}^T & \boldsymbol{Z}_{\vartheta_0}^T \end{pmatrix}^T$ is ergodic.

The assumption that the processes $\boldsymbol{Z}_{\vartheta_0}$ and $\boldsymbol{W}_{\vartheta_0}$ are ergodic implies via the moving average representation in Lemma 3.2, i) and Krengel (1985, Theorem 4.3) that the output process $\boldsymbol{Y} = \boldsymbol{Y}_{\vartheta_0}$ is ergodic. As a consequence, the pseudo-innovations $\boldsymbol{\varepsilon}_{\vartheta}$ defined in Eq. (3.2.11a) are ergodic for every $\vartheta \in \Theta$.

Our first identifiability assumption precludes redundancies in the parametrization of the state space models under consideration and is therefore necessary for the true parameter value ϑ_0 to be estimated consistently. It will be used in Lemma 3.13 to show that the quasi likelihood function given by Eq. (3.2.14) asymptotically has a unique global minimum at ϑ_0.

Assumption D5 For all $\vartheta_0 \neq \vartheta \in \Theta$, there exists a $z \in \mathbb{C}$ such that

$$H_{\vartheta}\left[\mathbf{1}_N - (F_{\vartheta} - K_{\vartheta}H_{\vartheta})\,z\right]^{-1} K_{\vartheta} \neq H_{\vartheta_0}\left[\mathbf{1}_N - (F_{\vartheta_0} - K_{\vartheta_0}H_{\vartheta_0})\,z\right]^{-1} K_{\vartheta_0}, \tag{3.2.16a}$$

or

$$V_{\vartheta} \neq V_{\vartheta_0}. \tag{3.2.16b}$$

Assumption D5 can be rephrased in terms of the spectral densities $f_{\boldsymbol{Y}_\vartheta}$ of the output processes $\boldsymbol{Y}_{\vartheta}$ of the state space models $(F_\vartheta, H_\vartheta, \boldsymbol{Z}_\vartheta, \boldsymbol{W}_\vartheta)$, which are defined as the inverse Fourier transforms of the autocovariance functions $\gamma_{\boldsymbol{Y}_\vartheta}$:

$$f_{\boldsymbol{Y}_\vartheta} : [-\pi,\pi] \to \mathbb{S}_d^+(\mathbb{R}), \quad \omega \mapsto \sum_{h\in\mathbb{Z}} \mathrm{e}^{-\mathrm{i}h\omega}\gamma_{\boldsymbol{Y}_\vartheta}(h); \quad \gamma_{\boldsymbol{Y}_\vartheta}(h) = \mathbb{E}\boldsymbol{Y}_{\vartheta,n}\boldsymbol{Y}_{\vartheta,n+h}^T.$$

This characterization will be very useful when we apply the estimation theory developed in this section to state space models that arise from sampling a continuous-time ARMA process.

Lemma 3.6 *If, for all $\vartheta_0 \neq \vartheta \in \Theta$, there exists an $\omega \in [-\pi,\pi]$ such that $f_{\boldsymbol{Y}_\vartheta}(\omega) \neq f_{\boldsymbol{Y}_{\vartheta_0}}(\omega)$, then Assumption D5 holds.*

Proof We recall from Hamilton (1994, Eq. (10.4.43)) that the spectral density $f_{\boldsymbol{Y}_\vartheta}$ of the output process $\boldsymbol{Y}_\vartheta$ of the state space model $(F_\vartheta, H_\vartheta, \boldsymbol{Z}_\vartheta, \boldsymbol{W}_\vartheta)$ is given by

$$f_{\boldsymbol{Y}_\vartheta}(\omega) = \frac{1}{2\pi}\mathscr{H}_\vartheta\left(\mathrm{e}^{\mathrm{i}\omega}\right) V_\vartheta \mathscr{H}_\vartheta\left(\mathrm{e}^{-\mathrm{i}\omega}\right)^T, \quad \omega \in [-\pi,\pi], \tag{3.2.17}$$

where $\mathscr{H}_\vartheta(z) := H_\vartheta\left[\mathbf{1}_N - (F_\vartheta - K_\vartheta H_\vartheta)\,z\right]^{-1} K_\vartheta + z$. If Assumption D5 does not hold, we have that both $\mathscr{H}_\vartheta(z) = \mathscr{H}_{\vartheta_0}(z)$ for all $z \in \mathbb{C}$, and $V_\vartheta = V_{\vartheta_0}$, and thus Eq. (3.2.17) implies that $f_{\boldsymbol{Y}_\vartheta}(\omega) = f_{\boldsymbol{Y}_{\vartheta_0}}(\omega)$, for all $\omega \in [-\pi,\pi]$, contradicting the assumption of the lemma. □

Under the assumptions described so far we obtain the following consistency result.

Theorem 3.7 (Consistency of $\hat{\boldsymbol{\vartheta}}^L$) *Assume that $(F_{\boldsymbol{\vartheta}}, H_{\boldsymbol{\vartheta}}, \boldsymbol{Z}_{\boldsymbol{\vartheta}}, \boldsymbol{W}_{\boldsymbol{\vartheta}})_{\boldsymbol{\vartheta}\in\Theta}$ is a parametric family of state space models according to Definition 3.1, and let $\boldsymbol{y}^L = (\boldsymbol{Y}_{\boldsymbol{\vartheta}_0,1}, \ldots, \boldsymbol{Y}_{\boldsymbol{\vartheta}_0,L})$ be a sample of length L from the output process of the model corresponding to $\boldsymbol{\vartheta}_0$. If Assumptions D1 to D5 hold, then the quasi maximum likelihood estimator $\hat{\boldsymbol{\vartheta}}^L = \operatorname{argmin}_{\boldsymbol{\vartheta}\in\Theta} \widehat{\mathscr{L}}(\boldsymbol{\vartheta}, \boldsymbol{y}^L)$ is strongly consistent, that is $\hat{\boldsymbol{\vartheta}}^L \to \boldsymbol{\vartheta}_0$ almost surely, as $L \to \infty$.*

We now describe the conditions which we need to impose in addition to Assumptions D1 to D5 for the asymptotic normality of the quasi maximum likelihood estimator to hold. The first one excludes the case that the true parameter value $\boldsymbol{\vartheta}_0$ lies on the boundary of the domain Θ.

Assumption D6 The true parameter value $\boldsymbol{\vartheta}_0$ is an element of the interior of Θ.

Next we need to impose a higher degree of smoothness than stated in Assumption D2 and a stronger moment condition than Assumption D4.

Assumption D7 The mappings $F_{(\cdot)}$, $H_{(\cdot)}$, $Q_{(\cdot)}$, $S_{(\cdot)}$, and $R_{(\cdot)}$ in Eqs. (3.2.9) are three times continuously differentiable.

By the results of the sensitivity analysis of the discrete-time algebraic Riccati equation in Sun (1998), the same degree of smoothness, namely C^3, also carries over to the mapping $\boldsymbol{\vartheta} \mapsto V_{\boldsymbol{\vartheta}}$.

Assumption D8 The process $\begin{pmatrix} \boldsymbol{W}_{\boldsymbol{\vartheta}_0}^T & \boldsymbol{Z}_{\boldsymbol{\vartheta}_0}^T \end{pmatrix}^T$ has finite $(4+\delta)$th moments for some $\delta > 0$, that is

$$\mathbb{E}\left\|\boldsymbol{W}_{\boldsymbol{\vartheta}_0,n}\right\|^{4+\delta} < \infty, \quad \mathbb{E}\left\|\boldsymbol{Z}_{\boldsymbol{\vartheta}_0,n}\right\|^{4+\delta} < \infty. \tag{3.2.18}$$

By Lemma 3.2, ii), Assumption D8 implies that the process $\boldsymbol{Y}$ has finite $(4+\delta)$th moments. In the definition of the general linear stochastic state space model and in Assumption D4, it was only assumed that the sequences $\boldsymbol{Z}$ and $\boldsymbol{W}$ are stationary and ergodic. This structure alone does not entail a sufficient amount of asymptotic independence for results like Theorem 3.8 to be established. Unless one is willing to restrict attention to Gaussian processes, not even imposing that $\boldsymbol{Z}$ and $\boldsymbol{W}$ are i. i. d. sequences leads a priori to a satisfactory asymptotic theory. It turns out that a certain degree of asymptotic independence stronger than ergodicity is required for the output process $\boldsymbol{Y}$. The notion of strong (or α-) mixing, which goes back to Rosenblatt (1956), is one possibility to define asymptotic independence quantitatively. It has turned out to be a very useful substitute for true independence allowing for many asymptotic results in the theory of inference for stochastic processes to be established. Given a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a stationary stochastic process $X = (X_t)_{t\in I}$ on that space, where I is either $\mathbb{R}$ or $\mathbb{Z}$, we introduce the σ-algebras $\mathscr{F}_n^m = \sigma(X_j : j \in I, n < j < m)$, $-\infty \leqslant n < m \leqslant \infty$. The strong-mixing coefficients $\alpha(m)$, $m \in I$, are defined by

$$\alpha(m) = \sup_{A\in\mathscr{F}_{-\infty}^0,\, B\in\mathscr{F}_m^\infty} \left|\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)\right|. \tag{3.2.19}$$

If $\lim_{m\to\infty}\alpha(m)=0$, the process X is called *strongly mixing*; it is called *exponentially strongly mixing* if $\alpha(m)=O(\lambda^m)$ for some $0<\lambda<1$. We assume that the process $\boldsymbol{Y}$ is strongly mixing and we impose a summability condition on the strong-mixing coefficients, which is known to be sufficient for a Central Limit Theorem for $\boldsymbol{Y}$ to hold (Bradley, 2007; Ibragimov, 1962).

Assumption D9 Denote by $\alpha_{\boldsymbol{Y}}$ the strong-mixing coefficients of the process $\boldsymbol{Y}=\boldsymbol{Y}_{\boldsymbol{\vartheta}_0}$. There exists a constant $\delta>0$ such that

$$\sum_{m=0}^{\infty}\left[\alpha_{\boldsymbol{Y}}(m)\right]^{\frac{\delta}{2+\delta}}<\infty. \tag{3.2.20}$$

In the case of exponential strong mixing, Assumption D9 is always satisfied, and it is no restriction to assume that the δ appearing in Assumptions D8 and D9 are the same. It follows from Mokkadem (1988) and the results in Chapter 2 that, because of the autoregressive structure of the state equation (3.2.2a), exponential strong mixing of the output process $\boldsymbol{Y}_{\boldsymbol{\vartheta}_0}$ can be assured by imposing the condition that the process $\boldsymbol{Z}_{\boldsymbol{\vartheta}_0}$ is an i. i. d. sequence whose marginal distributions possess a non-trivial absolutely continuous component in the sense of Lebesgue's decomposition theorem, see, e. g., Halmos (1950, §31, Theorem C) or the original account Lebesgue (1904).

Finally, we require another identifiability assumption, that will be used to ensure that the Fisher information matrix of the quasi maximum likelihood estimator is non-singular. This is necessary because the asymptotic covariance matrix in the asymptotic normality result for $\hat{\boldsymbol{\vartheta}}^L$ is directly related to the inverse of that matrix. Assumption D10 is formulated in terms of the first derivative of the parametrization of the model only, which makes it relatively easy to check in practice; the Fisher information matrix, in contrast, is related to the second derivative of the logarithmic Gaussian likelihood. For $j\in\mathbb{N}$ and $\boldsymbol{\vartheta}\in\Theta$, the vector $\psi_{\boldsymbol{\vartheta},j}\in\mathbb{R}^{(j+2)d^2}$ is defined as

$$\psi_{\boldsymbol{\vartheta},j}=\begin{pmatrix}\left[\mathbf{1}_{j+1}\otimes K_{\boldsymbol{\vartheta}}^T\otimes H_{\boldsymbol{\vartheta}}\right]\left[\begin{array}{cccc}(\operatorname{vec}\mathbf{1}_N)^T & (\operatorname{vec}F_{\boldsymbol{\vartheta}})^T & \cdots & \left(\operatorname{vec}F_{\boldsymbol{\vartheta}}^j\right)^T\end{array}\right]^T\\ \operatorname{vec}V_{\boldsymbol{\vartheta}}\end{pmatrix}, \tag{3.2.21}$$

where $\otimes$ denotes the Kronecker product of two matrices, and vec is the linear operator that transforms a matrix into a vector by stacking its columns on top of each other.

Assumption D10 There exists an integer $j_0\in\mathbb{N}$ such that the $[(j_0+2)d^2]\times r$ matrix $\nabla_{\boldsymbol{\vartheta}}\psi_{\boldsymbol{\vartheta}_0,j_0}$ has rank r.

Our main result about the asymptotic distribution of the quasi maximum likelihood estimator for discrete-time state space models is the following theorem. Equation (3.2.23) shows in particular that this asymptotic distribution is independent of the choice of the initial values $\hat{X}_{\vartheta,\text{initial}}$.

Theorem 3.8 (Asymptotic normality of $\hat{\vartheta}^L$) *Assume that $(F_{\vartheta}, H_{\vartheta}, Z_{\vartheta}, W_{\vartheta})_{\vartheta\in\Theta}$ is a parametric family of state space models according to Definition 3.1, and let $y^L = (Y_{\vartheta_0,1}, \ldots, Y_{\vartheta_0,L})$ be a sample of length L from the output process of the model corresponding to ϑ_0. If Assumptions D1 to D10 hold, then the maximum likelihood estimator $\hat{\vartheta}^L = \operatorname{argmin}_{\vartheta\in\Theta} \widehat{\mathscr{L}}(\vartheta, y^L)$ is asymptotically normally distributed with asymptotic covariance matrix $\Xi = J^{-1}IJ^{-1}$, that is*

$$\sqrt{L}\left(\hat{\vartheta}^L - \vartheta_0\right) \xrightarrow[L\to\infty]{d} \mathscr{N}(\mathbf{0}, \Xi), \tag{3.2.22}$$

where

$$I = \lim_{L\to\infty} L^{-1}\,\mathbb{V}\mathrm{ar}\left(\nabla_{\vartheta}\mathscr{L}\left(\vartheta_0, y^L\right)\right), \quad J = \lim_{L\to\infty} L^{-1}\nabla^2_{\vartheta}\mathscr{L}\left(\vartheta_0, y^L\right). \tag{3.2.23}$$

3.2.3. Proof of Theorem 3.7 – Strong consistency

In this section we prove the strong consistency of the quasi maximum likelihood estimator $\hat{\vartheta}^L$. As a first step we show that the stationary pseudo-innovations processes defined by the steady-state Kalman filter are uniformly approximated by their counterparts based on the finite sample y^L.

Lemma 3.9 *Under Assumptions D1 to D3, the pseudo-innovations sequences ε_{ϑ} and $\hat{\varepsilon}_{\vartheta}$ defined by the Kalman filter equations (3.2.6a) and (3.2.12) have the following properties.*

i) If the initial values $\hat{X}_{\vartheta,initial}$ are such that $\sup_{\vartheta\in\Theta}\left\|\hat{X}_{\vartheta,initial}\right\|$ is almost surely finite, then, with probability one, there exist a positive number C and a positive number $\rho < 1$, such that

$$\sup_{\vartheta\in\Theta}\left\|\varepsilon_{\vartheta,n} - \hat{\varepsilon}_{\vartheta,n}\right\| \leqslant C\rho^n, \quad n \in \mathbb{N}. \tag{3.2.24}$$

In particular, $\hat{\varepsilon}_{\vartheta_0,n}$ converges to the true innovations $\varepsilon_n = \varepsilon_{\vartheta_0,n}$ at an exponential rate.

ii) The sequences ε_{ϑ} are linear functions of Y, that is there exist matrix sequences $(c_{\vartheta,\nu})_{\nu\geqslant 1}$, such that

$$\varepsilon_{\vartheta,n} = Y_n + \sum_{\nu=1}^{\infty} c_{\vartheta,\nu} Y_{n-\nu}, \quad n \in \mathbb{Z}. \tag{3.2.25}$$

The matrices $c_{\vartheta,\nu}$ are uniformly exponentially bounded, that is there exist a positive constant C and a positive constant $\rho < 1$, such that

$$\sup_{\vartheta\in\Theta}\left\|c_{\vartheta,\nu}\right\| \leqslant C\rho^{\nu}, \quad \nu \in \mathbb{N}. \tag{3.2.26}$$

Proof We first prove part i) about the uniform exponential approximation of ε by $\hat{\varepsilon}$. Iterating the Kalman equations (3.2.6a) and (3.2.12), we find that, for $n \in \mathbb{N}$,

$$\varepsilon_{\vartheta,n} = \boldsymbol{Y}_n - H_\vartheta \left(F_\vartheta - K_\vartheta H_\vartheta\right)^{n-1} \hat{\boldsymbol{X}}_{\vartheta,1} - \sum_{\nu=1}^{n-1} H_\vartheta \left(F_\vartheta - K_\vartheta H_\vartheta\right)^{\nu-1} K_\vartheta \boldsymbol{Y}_{n-\nu}$$

and

$$\hat{\varepsilon}_{\vartheta,n} = \boldsymbol{Y}_n - H_\vartheta \left(F_\vartheta - K_\vartheta H_\vartheta\right)^{n-1} \hat{\boldsymbol{X}}_{\vartheta,\text{initial}} - \sum_{\nu=1}^{n-1} H_\vartheta \left(F_\vartheta - K_\vartheta H_\vartheta\right)^{\nu-1} K_\vartheta \boldsymbol{Y}_{n-\nu}.$$

Thus, using the fact that, by Lemma 3.5, the spectral radii of $F_\vartheta - K_\vartheta H_\vartheta$ are bounded by $\rho < 1$, it follows that

$$\begin{aligned}\sup_{\vartheta\in\Theta} \left\|\varepsilon_{\vartheta,n} - \hat{\varepsilon}_{\vartheta,n}\right\| &= \sup_{\vartheta\in\Theta} \left\| H_\vartheta \left(F_\vartheta - K_\vartheta H_\vartheta\right)^{n-1} \left(\boldsymbol{X}_{\vartheta,0} - \boldsymbol{X}_{\vartheta,\text{initial}}\right)\right\| \\ &\leqslant \|H\|_{L^\infty(\Theta)} \rho^{n-1} \sup_{\vartheta\in\Theta} \left\|\boldsymbol{X}_{\vartheta,0} - \boldsymbol{X}_{\vartheta,\text{initial}}\right\|,\end{aligned}$$

where $\|H\|_{L^\infty(\Theta)} := \sup_{\vartheta\in\Theta} \|H_\vartheta\|$ denotes the supremum norm of $H_{(\cdot)}$, which is finite by the Extreme Value Theorem. Since the last factor is almost surely finite by assumption, the claim (3.2.24) follows.

For part ii), we observe that Eq. (3.2.6a) and Lemma 3.5, ii) imply that ε_ϑ has the infinite-order moving average representation

$$\varepsilon_{\vartheta,n} = \boldsymbol{Y}_n - H_\vartheta \sum_{\nu=1}^{\infty} \left(F_\vartheta - K_\vartheta H_\vartheta\right)^{\nu-1} K_\vartheta \boldsymbol{Y}_{n-\nu}, \tag{3.2.27}$$

with uniformly exponentially bounded coefficients $c_{\vartheta,\nu} := -H_\vartheta \left(F_\vartheta - K_\vartheta H_\vartheta\right)^{\nu-1} K_\vartheta$. Explicitly, $\|c_{\vartheta.\nu}\| \leqslant \|H\|_{L^\infty(\Theta)} \|K\|_{L^\infty(\Theta)} \rho^{n-1}$. This shows Eqs. (3.2.25) and (3.2.26). □

Lemma 3.10 *Let $\mathscr{L}$ and $\widehat{\mathscr{L}}$ be given by Eqs.* (3.2.10) *and* (3.2.14)*. If Assumptions D1 to D3 are satisfied, then, almost surely,*

$$\frac{1}{L}\sup_{\vartheta\in\Theta}\left|\widehat{\mathscr{L}}(\vartheta, \boldsymbol{y}^L) - \mathscr{L}(\vartheta, \boldsymbol{y}^L)\right| \to 0, \quad \text{as } L \to \infty. \tag{3.2.28}$$

Proof We first observe that

$$\begin{aligned}\left|\widehat{\mathscr{L}}(\vartheta, \boldsymbol{y}^L) - \mathscr{L}(\vartheta, \boldsymbol{y}^L)\right| &= \sum_{n=1}^{L} \left[\hat{\varepsilon}_{\vartheta,n}^T V_\vartheta^{-1} \hat{\varepsilon}_{\vartheta,n} - \varepsilon_{\vartheta,n}^T V_\vartheta^{-1} \varepsilon_{\vartheta,n}\right] \\ &= \sum_{n=1}^{L} \left[\left(\hat{\varepsilon}_{\vartheta,n} - \varepsilon_{\vartheta,n}\right)^T V_\vartheta^{-1} \hat{\varepsilon}_{\vartheta,n} + \varepsilon_{\vartheta,n}^T V_\vartheta^{-1} \left(\hat{\varepsilon}_{\vartheta,n} - \varepsilon_{\vartheta,n}\right)\right].\end{aligned}$$

The fact that, by Lemma 3.5, iii), there exists a constant C such that $\left\|V_{\boldsymbol{\vartheta}}^{-1}\right\| \leqslant C$, for all $\boldsymbol{\vartheta} \in \Theta$, implies that

$$\frac{1}{L}\sup_{\boldsymbol{\vartheta}\in\Theta}\left|\widehat{\mathscr{L}}(\boldsymbol{\vartheta},\boldsymbol{y}^L)-\mathscr{L}(\boldsymbol{\vartheta},\boldsymbol{y}^L)\right| \leqslant \frac{C}{L}\sum_{n=1}^{L}\rho^n\left[\sup_{\boldsymbol{\vartheta}\in\Theta}\left\|\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{\vartheta},n}\right\|+\sup_{\boldsymbol{\vartheta}\in\Theta}\left\|\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}\right\|\right]. \tag{3.2.29}$$

Lemma 3.9, ii) and the assumption that $\boldsymbol{Y}$ has finite second moments imply that the expectation $\mathbb{E}\sup_{\boldsymbol{\vartheta}\in\Theta}\left\|\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}\right\|$ is finite. Applying Markov's inequality, one sees that, for every positive number ϵ,

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\rho^n\sup_{\boldsymbol{\vartheta}\in\Theta}\left\|\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}\right\|\geqslant\epsilon\right)\leqslant\mathbb{E}\sup_{\boldsymbol{\vartheta}\in\Theta}\left\|\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1}\right\|\sum_{n=1}^{\infty}\frac{\rho^n}{\epsilon}<\infty,$$

because $\rho < 1$. The Borel–Cantelli Lemma shows that $\rho^n\sup_{\boldsymbol{\vartheta}\in\Theta}\left\|\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}\right\|$ converges to zero almost surely, as $n\to\infty$. In an analogous way one can show that $\rho^n\sup_{\boldsymbol{\vartheta}\in\Theta}\left\|\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{\vartheta},n}\right\|$ converges to zero almost surely, and, consequently, so does the Cesàro mean in Eq. (3.2.29). The claim (3.2.28) thus follows. □

Lemma 3.11 *Assume that Assumptions D3 and D4 as well as the first part of Assumption D5, Eq. (3.2.16a), hold. If* $\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1}=\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,1}$ *almost surely, then* $\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0$.

Proof Assume, for the sake of contradiction, that $\boldsymbol{\vartheta}\neq\boldsymbol{\vartheta}_0$. By Assumption D5, there exist matrices $C_j\in M_d(\mathbb{R})$, $j\in\mathbb{N}_0$, such that, for $|z|\leqslant 1$,

$$H_{\boldsymbol{\vartheta}}\left[\mathbf{1}_N-(F_{\boldsymbol{\vartheta}}-K_{\boldsymbol{\vartheta}}H_{\boldsymbol{\vartheta}})z\right]^{-1}K_{\boldsymbol{\vartheta}}-H_{\boldsymbol{\vartheta}_0}\left[\mathbf{1}_N-(F_{\boldsymbol{\vartheta}_0}-K_{\boldsymbol{\vartheta}_0}H_{\boldsymbol{\vartheta}_0}z\right]^{-1}K_{\boldsymbol{\vartheta}_0}=\sum_{j=j_0}^{\infty}C_jz^j, \tag{3.2.30}$$

where $C_{j_0}\neq 0$, for some $j_0\geqslant 0$. Using Eq. (3.2.6b) and the assumed equality of $\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1}$ and $\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,1}$, this equation implies that $\mathbf{0}_d=\sum_{j=j_0}^{\infty}C_j\boldsymbol{Y}_{j_0-j}$ almost surely; in particular, the random variable $C_{j_0}\boldsymbol{Y}_0$ is almost surely equal to a linear combination of the components of $\boldsymbol{Y}_n$, $n<0$. It thus follows from the interpretation of the innovations sequence $\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0}$ as linear prediction errors for the process $\boldsymbol{Y}$ that $C_{j_0}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,0}$ is equal to zero, which implies that

$$\mathbb{E}C_{j_0}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,0}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,0}^TC_{j_0}^T=C_{j_0}V_{\boldsymbol{\vartheta}_0}C_{j_0}^T=0_d.$$

Since $V_{\boldsymbol{\vartheta}_0}$ is assumed to be non-singular, this implies that the matrix C_{j_0} is the null matrix, a contradiction to Eq. (3.2.30). □

Lemma 3.12 *If Assumptions D1 to D4 hold, then, with probability one, the sequence of random functions* $\boldsymbol{\vartheta}\mapsto L^{-1}\widehat{\mathscr{L}}(\boldsymbol{\vartheta},\boldsymbol{y}^L)$ *converges, as L tends to infinity, uniformly in* $\boldsymbol{\vartheta}$ *to the limiting function* $\mathscr{Q}:\Theta\to\mathbb{R}$ *defined by*

$$\mathscr{Q}(\boldsymbol{\vartheta})=d\log(2\pi)+\log\det V_{\boldsymbol{\vartheta}}+\mathbb{E}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1}^TV_{\boldsymbol{\vartheta}}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1}. \tag{3.2.31}$$

Proof In view of the approximation results in Lemma 3.10, it is enough to show that the sequence of random functions $\boldsymbol{\vartheta} \mapsto L^{-1}\mathscr{L}(\boldsymbol{\vartheta}, \boldsymbol{y}^L)$ converges uniformly to $\mathscr{Q}$. The proof of this assertion is based on the observation following Assumption D4 that for each $\boldsymbol{\vartheta} \in \Theta$ the sequence $\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}}$ is ergodic and its consequence that, by Birkhoff's Ergodic Theorem (Durrett, 2010, Theorem 6.2.1), the sequence $L^{-1}\mathscr{L}(\boldsymbol{\vartheta}, \boldsymbol{y}^L)$ converges to $\mathscr{Q}(\boldsymbol{\vartheta})$ point-wise. The stronger statement of uniform convergence follows from Assumption D1 that Θ is compact by an argument that is inspired by the proof of Ferguson (1996, Theorem 16): for $\delta > 0$, we write $B_\delta(\boldsymbol{\vartheta}) = \left\{\boldsymbol{\vartheta}' \in \Theta : \left\|\boldsymbol{\vartheta}' - \boldsymbol{\vartheta}\right\| < \delta\right\}$ for the open ball of radius δ around $\boldsymbol{\vartheta}$. The sequences $\underline{\sigma}^\delta_{\boldsymbol{\vartheta}} = \left(\underline{\sigma}^\delta_{\boldsymbol{\vartheta},n}\right)_{n\in\mathbb{Z}}$ and $\overline{\sigma}^\delta_{\boldsymbol{\vartheta}} = \left(\overline{\sigma}^\delta_{\boldsymbol{\vartheta},n}\right)_{n\in\mathbb{Z}}$, which are defined by

$$\underline{\sigma}^\delta_{\boldsymbol{\vartheta},n} = \inf_{\boldsymbol{\vartheta}' \in B_\delta(\boldsymbol{\vartheta})} \left[\boldsymbol{\varepsilon}^T_{\boldsymbol{\vartheta}',n} V^{-1}_{\boldsymbol{\vartheta}'} \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}',n} - \mathbb{E}\boldsymbol{\varepsilon}^T_{\boldsymbol{\vartheta}',1} V^{-1}_{\boldsymbol{\vartheta}'} \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}',1}\right]$$

and

$$\overline{\sigma}^\delta_{\boldsymbol{\vartheta},n} = \sup_{\boldsymbol{\vartheta}' \in B_\delta(\boldsymbol{\vartheta})} \left[\boldsymbol{\varepsilon}^T_{\boldsymbol{\vartheta}',n} V^{-1}_{\boldsymbol{\vartheta}'} \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}',n} - \mathbb{E}\boldsymbol{\varepsilon}^T_{\boldsymbol{\vartheta}',1} V^{-1}_{\boldsymbol{\vartheta}'} \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}',1}\right],$$

are strictly stationary, ergodic and monotone in δ. By Lemma 3.9, ii) there exists an integrable random variable Z such that $\underline{\sigma}^\delta_{\boldsymbol{\vartheta},1} < Z$ for all δ and all $\boldsymbol{\vartheta} \in \Theta$. Since, moreover, $\boldsymbol{\varepsilon}^T_{\boldsymbol{\vartheta},1} V^{-1}_{\boldsymbol{\vartheta}} \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1}$ is almost surely a continuous function of $\boldsymbol{\vartheta}$, and thus

$$\underline{\sigma}^\delta_{\boldsymbol{\vartheta},n} \xrightarrow[\delta\to 0]{\text{a.s.}} \boldsymbol{\varepsilon}^T_{\boldsymbol{\vartheta},n} V^{-1}_{\boldsymbol{\vartheta}} \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n} - \mathbb{E}\boldsymbol{\varepsilon}^T_{\boldsymbol{\vartheta},1} V^{-1}_{\boldsymbol{\vartheta}} \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1},$$

it follows from the Ergodic Theorem and Lebesgue's Dominated Convergence Theorem (Klenke, 2008, Corollary 6.26) that

$$\frac{1}{L}\sum_{n=1}^{L} \underline{\sigma}^\delta_{\boldsymbol{\vartheta},n} \xrightarrow[L\to\infty]{\text{a.s.}} \mathbb{E}\underline{\sigma}^\delta_{\boldsymbol{\vartheta},1} \xrightarrow[\delta\to 0]{} \mathbb{E}\left[\boldsymbol{\varepsilon}^T_{\boldsymbol{\vartheta},n} V^{-1}_{\boldsymbol{\vartheta}} \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n} - \mathbb{E}\boldsymbol{\varepsilon}^T_{\boldsymbol{\vartheta},1} V^{-1}_{\boldsymbol{\vartheta}} \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1}\right] = 0, \tag{3.2.32}$$

and similarly for $\overline{\sigma}^\delta_{\boldsymbol{\vartheta}}$. Since, for any $\boldsymbol{\vartheta}' \in B_\delta(\boldsymbol{\vartheta})$, it holds that

$$\frac{1}{L}\sum_{n=1}^{L} \underline{\sigma}^\delta_{\boldsymbol{\vartheta},n} \leqslant \frac{1}{L}\sum_{n=1}^{L} \boldsymbol{\varepsilon}^T_{\boldsymbol{\vartheta}',n} V^{-1}_{\boldsymbol{\vartheta}'} \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}',n} - \mathbb{E}\boldsymbol{\varepsilon}^T_{\boldsymbol{\vartheta}',1} V^{-1}_{\boldsymbol{\vartheta}'} \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}',1} \leqslant \frac{1}{L}\sum_{n=1}^{L} \overline{\sigma}^\delta_{\boldsymbol{\vartheta},n},$$

it follows that

$$\sup_{\boldsymbol{\vartheta}' \in B_\delta(\boldsymbol{\vartheta})} \left|\frac{1}{L}\sum_{n=1}^{L} \boldsymbol{\varepsilon}^T_{\boldsymbol{\vartheta}',n} V^{-1}_{\boldsymbol{\vartheta}'} \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}',n} - \mathbb{E}\boldsymbol{\varepsilon}^T_{\boldsymbol{\vartheta}',1} V^{-1}_{\boldsymbol{\vartheta}'} \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}',1}\right| \leqslant \left|\frac{1}{L}\sum_{n=1}^{L} \underline{\sigma}^\delta_{\boldsymbol{\vartheta},n}\right| + \left|\frac{1}{L}\sum_{n=1}^{L} \overline{\sigma}^\delta_{\boldsymbol{\vartheta},n}\right|.$$

Letting L tend to infinity on both sides of this inequality, we see that, almost surely,

$$\limsup_{L\to\infty}\sup_{\boldsymbol{\vartheta}'\in B_\delta(\boldsymbol{\vartheta})}\left|\frac{1}{L}\sum_{n=1}^{L}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}',n}^T V_{\boldsymbol{\vartheta}'}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}',n}-\mathbb{E}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}',1}^T V_{\boldsymbol{\vartheta}'}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}',1}\right|\leqslant\left|\mathbb{E}\underline{\sigma}_{\boldsymbol{\vartheta},1}^{\delta}\right|+\left|\mathbb{E}\overline{\sigma}_{\boldsymbol{\vartheta},1}^{\delta}\right|.$$

By Eq. (3.2.32) one finds, for every $\epsilon>0$ and every $\boldsymbol{\vartheta}\in\Theta$, a $\delta(\epsilon,\boldsymbol{\vartheta})>0$ such that $\left|\mathbb{E}\underline{\sigma}_{\boldsymbol{\vartheta},1}^{\delta}\right|+\left|\mathbb{E}\overline{\sigma}_{\boldsymbol{\vartheta},1}^{\delta}\right|<\epsilon$, for all $\delta<\delta(\epsilon,\boldsymbol{\vartheta})$. The collection of balls $\left\{B_{\delta(\epsilon,\boldsymbol{\vartheta})}(\boldsymbol{\vartheta})\right\}_{\boldsymbol{\vartheta}\in\Theta}$ covers Θ, and since the domain Θ is assumed to be compact, one can extract a finite subcover. This means that there exist a finite number of points $\boldsymbol{\vartheta}_1,\ldots,\boldsymbol{\vartheta}_k$ such that Θ is covered by the union of the $\delta(\epsilon,\boldsymbol{\vartheta}_i)$-balls centred at the $\boldsymbol{\vartheta}_i$, that is $\Theta\subset\bigcup_{i=1}^k B_{\delta(\epsilon,\boldsymbol{\vartheta}_i)}(\boldsymbol{\vartheta}_i)$. Defining $\delta(\epsilon)$ to be the minimum of the radii $\delta(\epsilon,\boldsymbol{\vartheta}_i)$, $i=1,\ldots,k$, it follows that, with probability one,

$$\begin{aligned}
&\limsup_{L\to\infty}\sup_{\boldsymbol{\vartheta}\in\Theta}\left|\frac{1}{L}\sum_{n=1}^{L}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}^T V_{\boldsymbol{\vartheta}}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}-\mathbb{E}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1}^T V_{\boldsymbol{\vartheta}'}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1}\right|\\
=&\limsup_{L\to\infty}\max_{i=1,\ldots,k}\sup_{\boldsymbol{\vartheta}\in B_{\delta(\epsilon,\boldsymbol{\vartheta}_i)}}\left|\frac{1}{L}\sum_{n=1}^{L}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}^T V_{\boldsymbol{\vartheta}}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}-\mathbb{E}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1}^T V_{\boldsymbol{\vartheta}}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1}\right|\\
\leqslant&\max_{i=1,\ldots,k}\left\{\left|\mathbb{E}\underline{\sigma}_{\boldsymbol{\vartheta}_i,1}^{\delta(\epsilon,\boldsymbol{\vartheta}_i)}\right|+\left|\mathbb{E}\overline{\sigma}_{\boldsymbol{\vartheta}_i,1}^{\delta(\epsilon,\boldsymbol{\vartheta}_i)}\right|\right\}\\
\leqslant&\epsilon,\quad\forall\delta\leqslant\delta(\epsilon).
\end{aligned}$$

Intersecting over a sequence ϵ_n which converges to zero proves the result. More precisely,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\boldsymbol{\vartheta}\in\Theta}\left|\frac{1}{L}\sum_{n=1}^{L}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}^T V_{\boldsymbol{\vartheta}}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}-\mathbb{E}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1}^T V_{\boldsymbol{\vartheta}'}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1}\right|\xrightarrow[L\to\infty]{}0\right)\\
\geqslant&\mathbb{P}\left(\limsup_{L\to\infty}\sup_{\boldsymbol{\vartheta}\in\Theta}\left|\frac{1}{L}\sum_{n=1}^{L}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}^T V_{\boldsymbol{\vartheta}}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}-\mathbb{E}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1}^T V_{\boldsymbol{\vartheta}'}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1}\right|\leqslant\epsilon_n\quad n=1,2\ldots\right)\\
=&\mathbb{P}\left(\bigcap_{n=1}^{\infty}\left\{\limsup_{L\to\infty}\sup_{\boldsymbol{\vartheta}\in\Theta}\left|\frac{1}{L}\sum_{n=1}^{L}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}^T V_{\boldsymbol{\vartheta}}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}-\mathbb{E}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1}^T V_{\boldsymbol{\vartheta}'}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1}\right|\leqslant\epsilon_n\right\}\right)=1,
\end{aligned}$$

because, by countable additivity, the probability of the intersection of a countably infinite number of events, each having full probability, is equal to unity. □

Lemma 3.13 *Under Assumptions D1 to D3 and D5, the function $\mathcal{Q}:\Theta\to\mathbb{R}$, as defined in Eq. (3.2.31), has a unique global minimum at $\boldsymbol{\vartheta}_0$.*

Proof We first observe that the difference $\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1}-\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,1}$ is an element of the Hilbert space spanned by the random variables $\{\boldsymbol{Y}_n,n\leqslant 0\}$, and that $\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,1}$ is, by definition, orthogonal to this space. This implies that the expectation $\mathbb{E}\left(\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1}-\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,1}\right)^T V_{\boldsymbol{\vartheta}}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,1}$ is equal to zero and, consequently, that $\mathscr{Q}(\boldsymbol{\vartheta})$ can be written as

$$\mathscr{Q}(\boldsymbol{\vartheta}) = d\log(2\pi) + \mathbb{E}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,1}^T V_{\boldsymbol{\vartheta}}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,1} + \mathbb{E}\left(\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1} - \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,1}\right)^T V_{\boldsymbol{\vartheta}}^{-1}\left(\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1} - \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,1}\right) + \log\det V_{\boldsymbol{\vartheta}}.$$

In particular, since $\mathbb{E}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,1}^T V_{\boldsymbol{\vartheta}_0}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,1} = \operatorname{tr}\left[V_{\boldsymbol{\vartheta}_0}^{-1}\mathbb{E}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,1}^T\right] = d$, it follows that

$$\mathscr{Q}(\boldsymbol{\vartheta}_0) = \log\det V_{\boldsymbol{\vartheta}_0} + d(1 + \log(2\pi)).$$

The elementary inequality $x - \log x \geqslant 1$, for $x > 0$, implies that

$$\operatorname{tr} M - \log\det M = \sum_{x\in\sigma(M)}(x - \log x) \geqslant d,$$

for all symmetric positive definite $d\times d$ matrices $M \in \mathbb{S}_d^{++}(\mathbb{R})$ with equality if and only if $M = \mathbf{1}_d$. Using this inequality for $M = V_{\boldsymbol{\vartheta}_0}^{-1}V_{\boldsymbol{\vartheta}}$, we thus obtain that, for all $\boldsymbol{\vartheta} \in \Theta$,

$$\begin{aligned}\mathscr{Q}(\boldsymbol{\vartheta}) - \mathscr{Q}(\boldsymbol{\vartheta}_0) =& d + \operatorname{tr}\left[V_{\boldsymbol{\vartheta}}^{-1}\mathbb{E}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,1}^T\right] - \log\det\left(V_{\boldsymbol{\vartheta}_0}^{-1}V_{\boldsymbol{\vartheta}}\right)\\ &+ \mathbb{E}\left(\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1} - \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,1}\right)^T V_{\boldsymbol{\vartheta}}^{-1}\left(\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1} - \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,1}\right) - \mathbb{E}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,1}^T V_{\boldsymbol{\vartheta}_0}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,1}\\ \geqslant& \mathbb{E}\left(\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1} - \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,1}\right)^T V_{\boldsymbol{\vartheta}}^{-1}\left(\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1} - \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,1}\right)\\ \geqslant& 0.\end{aligned}$$

It remains to argue that this chain of inequalities is in fact a strict inequality if $\boldsymbol{\vartheta} \neq \boldsymbol{\vartheta}_0$. If $V_{\boldsymbol{\vartheta}} \neq V_{\boldsymbol{\vartheta}_0}$, the first inequality is strict, and we are done. If $V_{\boldsymbol{\vartheta}} = V_{\boldsymbol{\vartheta}_0}$, the first part of Assumption D5, Eq. (3.2.16a), is satisfied. The second inequality is an equality if and only if $\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},1} = \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,1}$ almost surely, which, by Lemma 3.11, implies that $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$. Thus, the function $\mathscr{Q}$ has a unique global minimum at $\boldsymbol{\vartheta}_0$. □

Proof (of Theorem 3.7) We shall first show that the sequence $L^{-1}\widehat{\mathscr{L}}(\hat{\boldsymbol{\vartheta}}^L, \mathbf{y}^L)$ converges almost surely to $\mathscr{Q}(\boldsymbol{\vartheta}_0)$ as the sample size L tends to infinity. Assume that, for some positive number ϵ, it holds that $\sup_{\boldsymbol{\vartheta}\in\Theta}\left|L^{-1}\widehat{\mathscr{L}}(\boldsymbol{\vartheta}, \mathbf{y}^L) - \mathscr{Q}(\boldsymbol{\vartheta})\right| \leqslant \epsilon$. It then follows that

$$L^{-1}\widehat{\mathscr{L}}(\hat{\boldsymbol{\vartheta}}^L, \mathbf{y}^L) \leqslant L^{-1}\widehat{\mathscr{L}}(\boldsymbol{\vartheta}_0, \mathbf{y}^L) \leqslant \mathscr{Q}(\boldsymbol{\vartheta}_0) + \epsilon,$$

and

$$L^{-1}\widehat{\mathscr{L}}(\hat{\boldsymbol{\vartheta}}^L, \mathbf{y}^L) \geqslant \mathscr{Q}(\hat{\boldsymbol{\vartheta}}^L) - \epsilon \geqslant \mathscr{Q}(\boldsymbol{\vartheta}_0) - \epsilon,$$

where it was used that $\hat{\boldsymbol{\vartheta}}^L$ is defined to minimize $\widehat{\mathscr{L}}(\cdot, \mathbf{y}^L)$ and that, by Lemma 3.13, $\boldsymbol{\vartheta}_0$ minimizes $\mathscr{Q}(\cdot)$. In particular, it follows that $\left|L^{-1}\widehat{\mathscr{L}}(\hat{\boldsymbol{\vartheta}}^L, \mathbf{y}^L) - \mathscr{Q}(\boldsymbol{\vartheta}_0)\right| \leqslant \epsilon$. This observation and Lemma 3.12 immediately imply that

$$\mathbb{P}\left(\frac{1}{L}\widehat{\mathscr{L}}(\hat{\boldsymbol{\vartheta}}^L,\boldsymbol{y}^L)\xrightarrow[L\to\infty]{}\mathscr{Q}(\boldsymbol{\vartheta}_0)\right)\geqslant\mathbb{P}\left(\sup_{\boldsymbol{\vartheta}\in\Theta}\left|\frac{1}{L}\widehat{\mathscr{L}}(\boldsymbol{\vartheta},\boldsymbol{y}^L)-\mathscr{Q}(\boldsymbol{\vartheta})\right|\xrightarrow[L\to\infty]{}0\right)=1. \tag{3.2.33}$$

To complete the proof of the theorem, it suffices to show that, for every neighbourhood U of $\boldsymbol{\vartheta}_0$, with probability one, $\hat{\boldsymbol{\vartheta}}^L$ will eventually lie in U. For every such neighbourhood U of $\boldsymbol{\vartheta}_0$, we define the real number $\delta(U) := \inf_{\boldsymbol{\vartheta}\in\Theta\setminus U}\mathscr{Q}(\boldsymbol{\vartheta})-\mathscr{Q}(\boldsymbol{\vartheta}_0)$, which is strictly positive by Lemma 3.13. Then the following sequence of inequalities holds:

$$\begin{aligned}
\mathbb{P}\left(\hat{\boldsymbol{\vartheta}}^L\xrightarrow[L\to\infty]{}\boldsymbol{\vartheta}_0\right)=&\mathbb{P}\left(\forall U\,\exists L_0:\hat{\boldsymbol{\vartheta}}^L\in U\quad\forall L>L_0\right)\\
\geqslant&\mathbb{P}\left(\forall U\,\exists L_0:\mathscr{Q}(\hat{\boldsymbol{\vartheta}}^L)-\mathscr{Q}(\boldsymbol{\vartheta}_0)<\delta(U)\quad\forall L>L_0\right)\\
\geqslant&\mathbb{P}\left(\forall U\,\exists L_0:\left|\frac{1}{L}\widehat{\mathscr{L}}(\hat{\boldsymbol{\vartheta}}^L,\boldsymbol{y}^L)-\mathscr{Q}(\boldsymbol{\vartheta}_0)\right|<\frac{\delta(U)}{2}\right.\\
&\qquad\left.\text{and}\quad\left|\frac{1}{L}\widehat{\mathscr{L}}(\hat{\boldsymbol{\vartheta}}^L,\boldsymbol{y}^L)-\mathscr{Q}(\hat{\boldsymbol{\vartheta}}^L)\right|<\frac{\delta(U)}{2}\quad\forall L>L_0\right)\\
\geqslant&\mathbb{P}\left(\forall U\,\exists L_0:\left|\frac{1}{L}\widehat{\mathscr{L}}(\hat{\boldsymbol{\vartheta}}^L,\boldsymbol{y}^L)-\mathscr{Q}(\boldsymbol{\vartheta}_0)\right|<\frac{\delta(U)}{2}\quad\forall L>L_0\right)\\
&+\mathbb{P}\left(\forall U\,\exists L_0:\sup_{\boldsymbol{\vartheta}\in\Theta}\left|\frac{1}{L}\widehat{\mathscr{L}}(\boldsymbol{\vartheta},\boldsymbol{y}^L)-\mathscr{Q}(\boldsymbol{\vartheta})\right|<\frac{\delta(U)}{2}\quad\forall L>L_0\right)-1.
\end{aligned}$$

The first probability in the last line is equal to one by Eq. (3.2.33), the second because, by Lemma 3.12, the random functions $\boldsymbol{\vartheta}\mapsto L^{-1}\widehat{\mathscr{L}}(\boldsymbol{\vartheta},\boldsymbol{y}^L)$ converge almost surely uniformly to the function $\boldsymbol{\vartheta}\mapsto\mathscr{Q}(\boldsymbol{\vartheta})$. It thus follows that $\mathbb{P}\left(\hat{\boldsymbol{\vartheta}}^L\xrightarrow[L\to\infty]{}\boldsymbol{\vartheta}_0\right)=1$, which proves the theorem. □

3.2.4. Proof of Theorem 3.8 – Asymptotic normality

In this section we prove the assertion of Theorem 3.8, namely that the distribution of $L^{1/2}\left(\hat{\boldsymbol{\vartheta}}^L-\boldsymbol{\vartheta}_0\right)$ converges to a normal random variable with mean zero and covariance matrix $\Xi=J^{-1}IJ^{-1}$, an expression for which is given in Eq. (3.2.23). We first present a chain rule and some well-known explicit formulæ for the differentiation of common matrix-valued functions for easy reference; these can be found in Horn and Johnson (1994, Sections 6.5 and 6.6).

Proposition 3.14 *Assume that $g:\mathbb{R}\to M_{m,n}(\mathbb{R})$ and $f:M_{m,n}(\mathbb{R})\to\mathbb{R}$ are differentiable functions. The following chain rule holds.*

$$\frac{\partial}{\partial x}f(g(x))=\operatorname{tr}\left[\left(\left.\frac{\partial}{\partial M^T}f(M)\right|_{M=g(x)}\right)\left(\frac{\partial}{\partial x}g(x)\right)\right], \tag{3.2.34}$$

where, for $M = (m_{ij})_{ij} \in M_{m,n}(\mathbb{R})$ and $x \in \mathbb{R}$, we write

$$\frac{\partial}{\partial M^T} f(M) = \left(\frac{\partial}{\partial m_{ji}} f(M) \right)_{ij} \in M_{n,m}(\mathbb{R}), \tag{3.2.35a}$$

and

$$\frac{\partial}{\partial x} g(x) = \left(\frac{\partial}{\partial x} [g(x)]_{ij} \right)_{ij} \in M_{m,n}(\mathbb{R}). \tag{3.2.35b}$$

If $M \in M_m(\mathbb{R})$ is invertible, the following hold:

$$\frac{\partial}{\partial M^T} \log |\det M| = M^{-1}, \tag{3.2.36a}$$

$$\frac{\partial}{\partial M^T} \operatorname{tr}(AMB) = BA, \tag{3.2.36b}$$

$$\frac{\partial}{\partial M^T} \operatorname{tr}\left(AM^{-1}B\right) = -M^{-1}BAM^{-1}, \tag{3.2.36c}$$

where $A \in M_{k,m}(\mathbb{R})$ and $B \in M_{m,l}(\mathbb{R})$ are matrices of appropriate dimensions.

Next, we collect basic properties of $\partial_m \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}$ and $\partial_m \hat{\boldsymbol{\varepsilon}}_{\boldsymbol{\vartheta},n}$, where $\partial_m = \partial / \partial \vartheta^m$ denotes the partial derivative with respect to the mth component of $\boldsymbol{\vartheta}$; the following lemma mirrors Lemma 3.9.

Lemma 3.15 *Assume that Assumptions D1 to D3 and D7 hold. The pseudo-innovations sequences $\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}}$ and $\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{\vartheta}}$ defined by the Kalman filter equations (3.2.6a) and (3.2.12) have the following properties.*

i) If, for some $k \in \{1, \ldots, r\}$, the initial values $\hat{\mathbf{X}}_{\boldsymbol{\vartheta},initial}$ are such that both $\sup_{\boldsymbol{\vartheta}\in\Theta} \left\|\hat{\mathbf{X}}_{\boldsymbol{\vartheta},initial}\right\|$ and $\sup_{\boldsymbol{\vartheta}\in\Theta} \left\|\partial_k \hat{\mathbf{X}}_{\boldsymbol{\vartheta},initial}\right\|$ are almost surely finite, then, with probability one, there exist a positive number C and a positive number $\rho < 1$, such that

$$\sup_{\boldsymbol{\vartheta}\in\Theta} \left\|\partial_k \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n} - \partial_k \hat{\boldsymbol{\varepsilon}}_{\boldsymbol{\vartheta},n}\right\| \leqslant C\rho^n, \quad n \in \mathbb{N}. \tag{3.2.37}$$

ii) For each $k \in \{1, \ldots, r\}$, the random sequences $\partial_k \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}}$ are linear functions of $\mathbf{Y}$, that is there exist matrix sequences $\left(c_{\boldsymbol{\vartheta},\nu}^{(k)}\right)_{\nu\geqslant 1}$, such that

$$\partial_k \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n} = \sum_{\nu=1}^{\infty} c_{\boldsymbol{\vartheta},\nu}^{(k)} \mathbf{Y}_{n-\nu}, \quad n \in \mathbb{Z}. \tag{3.2.38}$$

The matrices $c_{\boldsymbol{\vartheta},\nu}^{(k)}$ are uniformly exponentially bounded, that is there exist a positive constant C and a positive constant $\rho < 1$, such that

$$\sup_{\boldsymbol{\vartheta}\in\Theta} \left\|c_{\boldsymbol{\vartheta},\nu}^{(k)}\right\| \leqslant C\rho^\nu, \quad \nu \in \mathbb{N}. \tag{3.2.39}$$

iii) *If, for some $k, l \in \{1, \ldots, r\}$, the initial values $\hat{\boldsymbol{X}}_{\boldsymbol{\vartheta},initial}$ are such that both $\sup_{\boldsymbol{\vartheta}\in\Theta} \left\|\hat{\boldsymbol{X}}_{\boldsymbol{\vartheta},initial}\right\|$, $\sup_{\boldsymbol{\vartheta}\in\Theta} \left\|\partial_i \hat{\boldsymbol{X}}_{\boldsymbol{\vartheta},initial}\right\|$, $i \in \{k, l\}$, and $\sup_{\boldsymbol{\vartheta}\in\Theta} \left\|\partial^2_{k,l} \hat{\boldsymbol{X}}_{\boldsymbol{\vartheta},initial}\right\|$ are almost surely finite, then, with probability one, there exist a positive number C and a positive number $\rho < 1$, such that*

$$\sup_{\boldsymbol{\vartheta}\in\Theta} \left\|\partial^2_{k,l} \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n} - \partial^2_{k,l} \hat{\boldsymbol{\varepsilon}}_{\boldsymbol{\vartheta},n}\right\| \leqslant C\rho^n, \quad n \in \mathbb{N}. \tag{3.2.40}$$

iv) *For each $k, l \in \{1, \ldots, r\}$, the random sequences $\partial^2_{k,l} \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}}$ are linear functions of $\boldsymbol{Y}$, that is there exist matrix sequences $\left(c^{(k,l)}_{\boldsymbol{\vartheta},\nu}\right)_{\nu \geqslant 1}$, such that*

$$\partial^2_{k,l} \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n} = \sum_{\nu=1}^{\infty} c^{(k,l)}_{\boldsymbol{\vartheta},\nu} \boldsymbol{Y}_{n-\nu}, \quad n \in \mathbb{Z}. \tag{3.2.41}$$

The matrices $c^{(k,l)}_{\boldsymbol{\vartheta},\nu}$ are uniformly exponentially bounded, that is there exist a positive constant C and a positive constant $\rho < 1$, such that

$$\sup_{\boldsymbol{\vartheta}\in\Theta} \left\|c^{(k,l)}_{\boldsymbol{\vartheta},\nu}\right\| \leqslant C\rho^\nu, \quad \nu \in \mathbb{N}. \tag{3.2.42}$$

Proof Analogous to the proof of Lemma 3.9, repeatedly interchanging differentiation and summation, and using the fact that, by Assumptions D1 to D3 and D7, both

$$\partial_k \left[H_{\boldsymbol{\vartheta}} \left(F_{\boldsymbol{\vartheta}} - K_{\boldsymbol{\vartheta}} H_{\boldsymbol{\vartheta}}\right)^{\nu-1} K_{\boldsymbol{\vartheta}}\right] \quad \text{and} \quad \partial^2_{k,l} \left[H_{\boldsymbol{\vartheta}} \left(F_{\boldsymbol{\vartheta}} - K_{\boldsymbol{\vartheta}} H_{\boldsymbol{\vartheta}}\right)^{\nu-1} K_{\boldsymbol{\vartheta}}\right]$$

are uniformly exponentially bounded in ν. □

Lemma 3.16 *For each $\boldsymbol{\vartheta} \in \Theta$ and every $m = 1, \ldots, r$, the random variable $\partial_m \mathscr{L}(\boldsymbol{\vartheta}, \boldsymbol{y}^L)$ has finite variance.*

Proof By Assumption D8 and the exponential decay of the coefficient matrices $c_{\boldsymbol{\vartheta},\nu}$ and $c^{(m)}_{\boldsymbol{\vartheta},\nu}$ proved in Lemma 3.9, ii) and Lemma 3.15, ii), it follows that

$$\mathbb{E} \left\|\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}\right\|^4 \leqslant \left[C \sum_{\nu=0}^{\infty} \rho^\nu\right]^4 \mathbb{E} \left\|\boldsymbol{Y}_1\right\|^4 < \infty,$$

and

$$\mathbb{E} \left\|\partial_m \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}\right\|^4 \leqslant \left[C \sum_{\nu=0}^{\infty} \rho^\nu\right]^4 \mathbb{E} \left\|\boldsymbol{Y}_1\right\|^4 < \infty.$$

Consequently, the derivative rules presented in Proposition 3.14 and the Cauchy–Schwarz inequality imply that, for some constant C,

$$\begin{aligned}\mathbb{E}\left|\partial_m\left(\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}^T V_{\boldsymbol{\vartheta}}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}\right)\right|^2 =&\mathbb{E}\left|-\operatorname{tr}\left[V_{\boldsymbol{\vartheta}}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}^T\left(\partial_m V_{\boldsymbol{\vartheta}}\right)\right]+2\left(\partial_m\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}^T\right)V_{\boldsymbol{\vartheta}}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}\right|^2\\ \leqslant& C\left\{\mathbb{E}\left\|\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}\right\|^4+\left(\mathbb{E}\left\|\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}\right|^4\mathbb{E}\left\|\partial_m\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}\right\|^4\right)^{1/2}\right\}\\ <&\infty,\end{aligned}$$

which proves that $\partial_m\mathscr{L}(\boldsymbol{\vartheta},\boldsymbol{y}^L)$ has finite second moments. □

We need the following multivariate covariance inequality which is a consequence of Davydov's inequality and the multidimensional generalization of an inequality used in the proof of Francq and Zakoïan (1998, Lemma 3). An overview of covariance inequalities for strongly mixing processes can be found in Bradley (2007); Doukhan (1994). For a positive real number α, we denote by $\lfloor\alpha\rfloor$ the greatest integer smaller than or equal to α.

Lemma 3.17 *Let $\boldsymbol{X}$ 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 C, such that for all $d\times d$ matrices A, B, every $n\in\mathbb{Z}$, $\Delta\in\mathbb{N}$, and time indices $\nu,\nu'\in\mathbb{N}_0$, $\mu,\mu'=0,1\ldots,\lfloor\Delta/2\rfloor$, it holds that*

$$\operatorname{Cov}\left(\boldsymbol{X}_{n-\nu}^T A\boldsymbol{X}_{n-\nu'};\boldsymbol{X}_{n+\Delta-\mu}^T B\boldsymbol{X}_{n+\Delta-\mu'}\right)\leqslant C\left\|A\right\|\left\|B\right\|\left[\alpha_{\boldsymbol{X}}\left(\left\lfloor\frac{\Delta}{2}\right\rfloor\right)\right]^{\delta/(\delta+2)},\tag{3.2.43}$$

where $\alpha_{\boldsymbol{X}}$ denotes the strong-mixing coefficients of the process $\boldsymbol{X}$, defined in Eq. (3.2.19).

Proof We first note that the bilinearity of $\operatorname{Cov}(\cdot;\cdot)$ and the elementary inequality $M_{ij}\leqslant\|M\|$, $M\in M_d(\mathbb{R})$, imply that

$$\begin{aligned}&\operatorname{Cov}\left(\boldsymbol{X}_{n-\nu}^T A\boldsymbol{X}_{n-\nu'};\boldsymbol{X}_{n+\Delta-\mu}^T B\boldsymbol{X}_{n+\Delta-\mu'}\right)\\ =&\sum_{i,j,s,t=1}^{d}A_{ij}B_{st}\operatorname{Cov}\left(X_{n-\nu}^i X_{n-\nu'}^j;X_{n+\Delta-\mu}^s X_{n+\Delta-\mu'}^t\right)\\ \leqslant& d^4\left\|A\right\|\left\|B\right\|\max_{i,j,s,t=1,\ldots,d}\operatorname{Cov}\left(X_{n-\nu}^i X_{n-\nu'}^j;X_{n+\Delta-\mu}^s X_{n+\Delta-\mu'}^t\right).\end{aligned}$$

Since the projection operator which maps a vector to one of its components is measurable, it follows that, for each i,j, the random variable $X_{n-\nu}^i X_{n-\nu'}^j$ is measurable with respect to $\mathscr{F}_{-\infty}^{n-\min\{\nu,\nu'\}}$, the σ-algebra generated by $\{\boldsymbol{X}_k:-\infty<k\leqslant n-\min\{\nu,\nu'\}\}$. Similarly, for each s,t, the random variable $X_{n+\Delta-\mu}^s X_{n+\Delta-\mu'}^t$ is measurable with respect to $\mathscr{F}_{n+\Delta-\max\{\mu,\mu'\}}^{\infty}$, the σ-algebra generated by $\{\boldsymbol{X}_k:n+\Delta-\max\{\mu,\mu'\}\leqslant k<\infty\}$. Davydov's inequality (Davydov, 1968, Lemma 2.1) implies that there exists a universal constant K such that

$$\begin{aligned}
&\mathbb{C}\mathrm{ov}\left(X^i_{n-\nu}X^j_{n-\nu'};X^s_{n+\Delta-\mu}X^t_{n+\Delta-\mu'}\right)\\
\leqslant &K\left(\mathbb{E}\left|X^i_{n-\nu}X^j_{n-\nu'}\right|^{2+\delta}\right)^{1/(2+\delta)}\left(\mathbb{E}\left|X^s_{n+\Delta-\mu}X^t_{n+\Delta-\mu'}\right|^{2+\delta}\right)^{1/(2+\delta)}\\
&\qquad\times\left[\alpha_X\left(\Delta-\max\left\{\mu,\mu'\right\}+\min\left\{\nu,\nu'\right\}\right)\right]^{\delta/(2+\delta)}\\
\leqslant &C\left[\alpha_X\left(\left\lfloor\frac{\Delta}{2}\right\rfloor\right)\right]^{\delta/(2+\delta)},
\end{aligned}$$

where it was used that $\Delta-\max\{\mu,\mu'\}+\min\{\nu,\nu'\}\geqslant\lfloor\Delta/2\rfloor$, and that strong-mixing coefficients are non-increasing. By the Cauchy–Schwarz inequality the constant C satisfies

$$C=K\left(\mathbb{E}\left|X^i_{n-\nu}X^j_{n-\nu'}\right|^{2+\delta}\right)^{1/(2+\delta)}\left(\mathbb{E}\left|X^s_{n+\Delta-\mu}X^t_{n+\Delta-\mu'}\right|^{2+\delta}\right)^{1/(2+\delta)}\leqslant K\left(\mathbb{E}\left\|\boldsymbol{X}_1\right\|^{4+2\delta}\right)^{\frac{2}{2+\delta}},$$

and thus does not depend on $n,\nu,\nu',\mu,\mu',\Delta$, nor on i,j,s,t. □

The next lemma is a multivariate generalization of Francq and Zakoïan (1998, Lemma 3). In the proof of Boubacar Mainassara and Francq (2011, Lemma 4) this generalization is used without providing details and, more importantly without imposing Assumption D9 about the strong mixing of $\boldsymbol{Y}$. In view of the derivative terms $\partial_m\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}$ in Eq. (3.2.45) it is not clear how the result of the lemma can be proved under the mere assumption of strong mixing of the innovations sequence $\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0}$. We therefore think that a detailed account, properly generalizing the arguments in the original paper (Francq and Zakoïan, 1998) to the multidimensional setting, is justified.

Lemma 3.18 *Suppose that Assumptions D1 to D3, D8 and D9 are satisfied. It then holds that, for every $\boldsymbol{\vartheta}\in\Theta$, the sequence $L^{-1}\,\mathbb{V}\mathrm{ar}\,\nabla_{\boldsymbol{\vartheta}}\mathscr{L}(\boldsymbol{\vartheta},\boldsymbol{y}^L)$ of deterministic matrices converges to a limit $I(\boldsymbol{\vartheta})$, as $L\to\infty$.*

Proof It is enough to show that, for each $\boldsymbol{\vartheta}\in\Theta$, and all $k,l=1,\ldots,r$, the sequence of real-valued random variables $I^{(k,l)}_{\boldsymbol{\vartheta},L}$, defined by

$$I^{(k,l)}_{\boldsymbol{\vartheta},L}=\frac{1}{L}\sum_{n=1}^{L}\sum_{t=1}^{L}\mathbb{C}\mathrm{ov}\left(\ell^{(k)}_{\boldsymbol{\vartheta},n},\ell^{(l)}_{\boldsymbol{\vartheta},t}\right),\tag{3.2.44}$$

converges to a limit as L tends to infinity, where $\ell^{(m)}_{\boldsymbol{\vartheta},n}=\partial_m l_{\boldsymbol{\vartheta},n}$ is the partial derivative of the nth term in expression (3.2.10) for $\mathscr{L}(\boldsymbol{\vartheta},\boldsymbol{y}^L)$. It follows from the derivative rules stated in Proposition 3.14 that

$$\ell^{(m)}_{\boldsymbol{\vartheta},n}=\mathrm{tr}\left[V_{\boldsymbol{\vartheta}}^{-1}\left(\mathbf{1}_d-\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}\boldsymbol{\varepsilon}^T_{\boldsymbol{\vartheta},n}V_{\boldsymbol{\vartheta}}^{-1}\right)\left(\partial_m V_{\boldsymbol{\vartheta}}\right)\right]+2\left(\partial_m\boldsymbol{\varepsilon}^T_{\boldsymbol{\vartheta},n}\right)V_{\boldsymbol{\vartheta}}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}.\tag{3.2.45}$$

By the assumed stationarity of the processes $\boldsymbol{\varepsilon}_{\vartheta}$, the covariances in the sum (3.2.44) depend only on the difference $n-t$. For the proof of the lemma it suffices to show that the sequence

$$\mathfrak{c}_{\vartheta,\Delta}^{(k,l)} = \mathbb{C}\mathrm{ov}\left(\ell_{\vartheta,n}^{(k)}, \ell_{n+\Delta,\vartheta}^{(l)}\right), \quad \Delta \in \mathbb{Z}, \tag{3.2.46}$$

is absolutely summable for all $k,l = 1,\ldots,r$, because then the Dominated Convergence Theorem implies that

$$I_{\vartheta,L}^{(k,l)} = \frac{1}{L}\sum_{\Delta=-L}^{L}(L-|\Delta|)\,\mathfrak{c}_{\vartheta,\Delta}^{(k,l)} \xrightarrow[L\to\infty]{} \sum_{\Delta\in\mathbb{Z}}\mathfrak{c}_{\vartheta,\Delta}^{(k,l)} < \infty. \tag{3.2.47}$$

In view of the of the symmetry $\mathfrak{c}_{\vartheta,\Delta}^{(k,l)} = \mathfrak{c}_{\vartheta,-\Delta}^{(k,l)}$, it is no restriction to assume that $\Delta \in \mathbb{N}$. In order to show that $\sum_{\Delta}\left|\mathfrak{c}_{\vartheta,\Delta}^{(k,l)}\right|$ is finite, we first use the bilinearity of $\mathbb{C}\mathrm{ov}(\cdot;\cdot)$ to estimate

$$\begin{aligned}
\left|\mathfrak{c}_{\vartheta,\Delta}^{(k,l)}\right| \leqslant& 4\left|\mathbb{C}\mathrm{ov}\left(\left(\partial_k\boldsymbol{\varepsilon}_{\vartheta,n}^T\right)V_{\vartheta}^{-1}\boldsymbol{\varepsilon}_{\vartheta,n};\left(\partial_l\boldsymbol{\varepsilon}_{\vartheta,n+\Delta}^T\right)V_{\vartheta}^{-1}\boldsymbol{\varepsilon}_{\vartheta,n+\Delta}\right)\right| \\
&+\left|\mathbb{C}\mathrm{ov}\left(\mathrm{tr}\left[V_{\vartheta}^{-1}\boldsymbol{\varepsilon}_{\vartheta,n}\boldsymbol{\varepsilon}_{\vartheta,n}^T V_{\vartheta}^{-1}\partial_k V_{\vartheta}\right];\mathrm{tr}\left[V_{\vartheta}^{-1}\boldsymbol{\varepsilon}_{\vartheta,n+\Delta}\boldsymbol{\varepsilon}_{\vartheta,n+\Delta}^T V_{\vartheta}^{-1}\partial_l V_{\vartheta}\right]\right)\right| + \\
&+2\left|\mathbb{C}\mathrm{ov}\left(\mathrm{tr}\left[V_{\vartheta}^{-1}\boldsymbol{\varepsilon}_{\vartheta,n}\boldsymbol{\varepsilon}_{\vartheta,n}^T V_{\vartheta}^{-1}\partial_k V_{\vartheta}\right];\left(\partial_l\boldsymbol{\varepsilon}_{\vartheta,n+\Delta}^T\right)V_{\vartheta}^{-1}\boldsymbol{\varepsilon}_{\vartheta,n+\Delta}\right)\right| + \\
&+2\left|\mathbb{C}\mathrm{ov}\left(\left(\partial_k\boldsymbol{\varepsilon}_{\vartheta,n}^T\right)V_{\vartheta}^{-1}\boldsymbol{\varepsilon}_{\vartheta,n};\mathrm{tr}\left[V_{\vartheta}^{-1}\boldsymbol{\varepsilon}_{\vartheta,n+\Delta}\boldsymbol{\varepsilon}_{\vartheta,n+\Delta}^T V_{\vartheta}^{-1}\partial_l V_{\vartheta}\right]\right)\right|.
\end{aligned}$$

Each of these four terms can be analysed separately. We give details only for the first one, the arguments for the other three terms being similar. Using the moving average representations (3.2.25) and (3.2.38) for $\boldsymbol{\varepsilon}_{\vartheta}$, $\partial_k\boldsymbol{\varepsilon}_{\vartheta}$ and $\partial_l\boldsymbol{\varepsilon}_{\vartheta}$, it follows that

$$\begin{aligned}
&\left|\mathbb{C}\mathrm{ov}\left(\left(\partial_k\boldsymbol{\varepsilon}_{\vartheta,n}^T\right)V_{\vartheta}^{-1}\boldsymbol{\varepsilon}_{\vartheta,n};\left(\partial_l\boldsymbol{\varepsilon}_{\vartheta,n+\Delta}^T\right)V_{\vartheta}^{-1}\boldsymbol{\varepsilon}_{\vartheta,n+\Delta}\right)\right| \\
=&\sum_{\nu,\nu',\mu,\mu'=0}^{\infty}\left|\mathbb{C}\mathrm{ov}\left(\boldsymbol{Y}_{n-\nu}^T c_{\vartheta,\nu}^{(k),T}V_{\vartheta}^{-1}c_{\vartheta,\nu'}\boldsymbol{Y}_{n-\nu'},\boldsymbol{Y}_{n+\Delta-\mu}^T c_{\vartheta,\mu}^{(l),T}V_{\vartheta}^{-1}c_{\vartheta,\mu'}\boldsymbol{Y}_{n+\Delta-\mu'}\right)\right|.
\end{aligned}$$

This sum can be split into one part I^+ in which at least one of the summation indices ν, ν', μ and μ' exceeds $\Delta/2$, and one part I^- in which all summation indices are less than or equal to $\Delta/2$. Using the fact that, by the Cauchy–Schwarz inequality,

$$\begin{aligned}
&\left|\mathbb{C}\mathrm{ov}\left(\boldsymbol{Y}_{n-\nu}^T c_{\vartheta,\nu}^{(k),T}V_{\vartheta}^{-1}c_{\vartheta,\nu'}\boldsymbol{Y}_{n-\nu'};\boldsymbol{Y}_{n+\Delta-\mu}^T c_{\vartheta,\mu}^{(l),T}V_{\vartheta}^{-1}c_{\vartheta,\mu'}\boldsymbol{Y}_{n+\Delta-\mu'}\right)\right| \\
\leqslant&\left\|V_{\vartheta}^{-1}\right\|^2\left\|c_{\vartheta,\nu}^{(k)}\right\|\left\|c_{\vartheta,\nu'}\right\|\left\|c_{\vartheta,\mu'}^{(l)}\right\|\left\|c_{\vartheta,\mu'}\right\|\mathbb{E}\left\|\boldsymbol{Y}_n\right\|^4,
\end{aligned}$$

it follows from Assumption D8 and the uniform exponential decay of $\|c_{\vartheta,\nu}\|$ and $\left\|c_{\vartheta,\nu}^{(m)}\right\|$

proved in Lemma 3.9, ii) and Lemma 3.15, ii) that there exist constants C and $\rho < 1$ such that

$$\begin{aligned} I^+ =& \sum_{\substack{\nu,\nu',\mu,\mu'=0\\ \max\{\nu,\nu',\mu,\mu'\}>\Delta/2}}^{\infty} \left|\operatorname{Cov}\left(\boldsymbol{Y}_{n-\nu}^T c_{\boldsymbol{\vartheta},\nu}^{(k),T} V_{\boldsymbol{\vartheta}}^{-1} c_{\boldsymbol{\vartheta},\nu'} \boldsymbol{Y}_{n-\nu'}, \boldsymbol{Y}_{n+\Delta-\mu}^T c_{\boldsymbol{\vartheta},\mu}^{(l),T} V_{\boldsymbol{\vartheta}}^{-1} c_{\boldsymbol{\vartheta},\mu'} \boldsymbol{Y}_{n+\Delta-\mu'}\right)\right| \\ \leqslant & C\rho^{\Delta/2}. \end{aligned} \tag{3.2.48}$$

For the contribution from all indices smaller than or equal to $\Delta/2$, Lemma 3.17 implies that

$$\begin{aligned} I^- =& \sum_{\nu,\nu',\mu,\mu'=0}^{\lfloor\Delta/2\rfloor} \left|\operatorname{Cov}\left(\boldsymbol{Y}_{n-\nu}^T c_{\boldsymbol{\vartheta},\nu}^{(k),T} V_{\boldsymbol{\vartheta}}^{-1} c_{\boldsymbol{\vartheta},\nu'} \boldsymbol{Y}_{n-\nu'}, \boldsymbol{Y}_{n+\Delta-\mu}^T c_{\boldsymbol{\vartheta},\mu}^{(l),T} V_{\boldsymbol{\vartheta}}^{-1} c_{\boldsymbol{\vartheta},\mu'} \boldsymbol{Y}_{n+\Delta-\mu'}\right)\right| \\ \leqslant & C\left[\alpha_{\boldsymbol{Y}}\left(\left\lfloor\frac{\Delta}{2}\right\rfloor\right)\right]^{\delta/(2+\delta)}. \end{aligned} \tag{3.2.49}$$

It thus follows from Assumption D9 that the sequences $\left|c_{\boldsymbol{\vartheta},\Delta}^{(k,l)}\right|$, $\Delta \in \mathbb{N}$, are summable, and Eq. (3.2.47) completes the proof of the lemma. □

We shall also need the following multivariate Chebyshev inequality.

Lemma 3.19 *Let $\boldsymbol{Z}$ be an $\mathbb{R}^d$ valued random variable with finite second moments. It holds that, for every $\epsilon > 0$,*

$$\mathbb{P}\left(\|\boldsymbol{Z} - \mathbb{E}\boldsymbol{Z}\| > \epsilon\right) \leqslant \frac{d}{\epsilon^2} \operatorname{tr} \mathbb{V}\mathrm{ar}\left(\boldsymbol{Z}\right). \tag{3.2.50}$$

Proof The claim is a consequence of the standard one-dimensional Chebyshev inequality. Using the subadditivity of the probability measure $\mathbb{P}$, we obtain that

$$\begin{aligned} \mathbb{P}\left(\|\boldsymbol{Z} - \mathbb{E}\boldsymbol{Z}\| > \epsilon\right) =& \mathbb{P}\left(\sum_{i=1}^d \left(Z^i - \mathbb{E}Z^i\right)^2 > \epsilon^2\right) \\ \leqslant & \mathbb{P}\left(\bigcup_{i=1}^d \left\{\left|Z^i - \mathbb{E}Z^i\right| > \frac{\epsilon}{\sqrt{d}}\right\}\right) \\ \leqslant & \sum_{i=1}^d \mathbb{P}\left(\left|Z^i - \mathbb{E}Z^i\right| > \frac{\epsilon}{\sqrt{d}}\right) \leqslant \frac{d}{\epsilon^2}\sum_{i=1}^d \mathbb{V}\mathrm{ar}\left(Z^i\right) = \frac{d}{\epsilon^2} \operatorname{tr} \mathbb{V}\mathrm{ar}\left(\boldsymbol{Z}\right). \end{aligned}$$ □

Lemma 3.20 *Let $\mathscr{L}$ and $\widehat{\mathscr{L}}$ be given by Eqs. (3.2.10) and (3.2.14). Assume that Assumptions D1 to D3 and D7 are satisfied. Then the following hold.*

i) For each $m = 1, \ldots, r$,

$$\frac{1}{\sqrt{L}} \sup_{\boldsymbol{\vartheta}\in\Theta} \left|\partial_m \widehat{\mathscr{L}}(\boldsymbol{\vartheta}, \boldsymbol{y}^L) - \partial_m \mathscr{L}(\boldsymbol{\vartheta}, \boldsymbol{y}^L)\right| \to 0, \quad \textit{in probability}, \tag{3.2.51}$$

as $L \to \infty$.

ii) For all $k, l = 1, \ldots, r$,

$$\frac{1}{L}\sup_{\boldsymbol{\vartheta}\in\Theta}\left|\partial^2_{k,l}\widehat{\mathscr{L}}(\boldsymbol{\vartheta},\boldsymbol{y}^L) - \partial^2_{k,l}\mathscr{L}(\boldsymbol{\vartheta},\boldsymbol{y}^L)\right| \to 0, \quad \textit{almost surely}, \tag{3.2.52}$$

as $L \to \infty$.

Proof Similar to the proof of Lemma 3.10. □

Lemma 3.21 *Under Assumptions D1, D3 and D7 to D9, the random variable $L^{-1/2}\nabla_{\boldsymbol{\vartheta}}\widehat{\mathscr{L}}(\boldsymbol{\vartheta}_0,\boldsymbol{y}^L)$ is asymptotically normally distributed with mean zero and covariance matrix $I(\boldsymbol{\vartheta}_0)$.*

Proof Because of Lemma 3.20, i) it is enough to show that $L^{-1/2}\nabla_{\boldsymbol{\vartheta}}\mathscr{L}\left(\boldsymbol{\vartheta}_0,\boldsymbol{y}^L\right)$ is asymptotically normally distributed with mean zero and covariance matrix $I(\boldsymbol{\vartheta}_0)$. We begin the proof by recalling the equation

$$\partial_i\mathscr{L}(\boldsymbol{\vartheta},\boldsymbol{y}^L) = \sum_{n=1}^{L}\left\{\operatorname{tr}\left[V_{\boldsymbol{\vartheta}}^{-1}\left(\mathbf{1}_d - \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}^T V_{\boldsymbol{\vartheta}}^{-1}\right)\partial_i V_{\boldsymbol{\vartheta}}\right] + 2\left(\partial_i\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}^T\right)V_{\boldsymbol{\vartheta}}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}\right\}, \tag{3.2.53}$$

which holds for every component $i = 1, \ldots, r$. The facts that $\mathbb{E}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,n}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,n}^T$ equals $V_{\boldsymbol{\vartheta}_0}$, and that $\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,n}$ is orthogonal to the Hilbert space generated by $\{\boldsymbol{Y}_t, t < n\}$, of which $\partial_i\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}^T$ is an element, show that $\mathbb{E}\partial_i\mathscr{L}\left(\boldsymbol{\vartheta}_0,\boldsymbol{y}^L\right) = 0$. Using Eq. (3.2.25), expression (3.2.53) can be rewritten as

$$\partial_i\mathscr{L}\left(\boldsymbol{\vartheta}_0,\boldsymbol{y}^L\right) = \sum_{n=1}^{L}\left[Y_{m,n}^{(i)} - \mathbb{E}Y_{m,n}^{(i)}\right] + \sum_{n=1}^{L}\left[Z_{m,n}^{(i)} - \mathbb{E}Z_{m,n}^{(i)}\right],$$

where, for every $m \in \mathbb{N}$, the processes $Y_m^{(i)}$ and $Z_m^{(i)}$ are defined by

$$\begin{aligned} Y_{m,n}^{(i)} =& \operatorname{tr}\left[V_{\boldsymbol{\vartheta}_0}^{-1}(\partial_i V_{\boldsymbol{\vartheta}_0})\right] + \sum_{\nu,\nu'=0}^{m}\left\{-\operatorname{tr}\left[V_{\boldsymbol{\vartheta}_0}^{-1}c_{\boldsymbol{\vartheta}_0,\nu}\boldsymbol{Y}_{n-\nu}\boldsymbol{Y}_{n-\nu'}^T c_{\boldsymbol{\vartheta},\nu'}^T V_{\boldsymbol{\vartheta}_0}^{-1}(\partial_i V_{\boldsymbol{\vartheta}_0})\right]\right. \\ &\left. +2\boldsymbol{Y}_{n-\nu}^T c_{\boldsymbol{\vartheta}_0,\nu}^{(i),T} V_{\boldsymbol{\vartheta}_0}^{-1} c_{\boldsymbol{\vartheta}_0,\nu'}\boldsymbol{Y}_{n-\nu'}\right\}, \end{aligned} \tag{3.2.54a}$$

$$Z_{m,n}^{(i)} = U_{m,n}^{(i)} + V_{m,n}^{(i)}, \tag{3.2.54b}$$

and

$$U_{m,n}^{(i)} = \sum_{\nu=0}^{\infty}\sum_{\nu'=m+1}^{\infty}\left\{-\operatorname{tr}\left[V_{\boldsymbol{\vartheta}_0}^{-1}c_{\boldsymbol{\vartheta}_0,\nu}\boldsymbol{Y}_{n-\nu}\boldsymbol{Y}_{n-\nu'}^T c_{\boldsymbol{\vartheta},\nu'}^T V_{\boldsymbol{\vartheta}_0}^{-1}(\partial_i V_{\boldsymbol{\vartheta}_0})\right] + 2\boldsymbol{Y}_{n-\nu}^T c_{\boldsymbol{\vartheta}_0,\nu}^{(i),T} V_{\boldsymbol{\vartheta}_0}^{-1} c_{\boldsymbol{\vartheta}_0,\nu'}\boldsymbol{Y}_{n-\nu'}\right\},$$

$$V_{m,n}^{(i)} = \sum_{\nu=m+1}^{\infty}\sum_{\nu'=0}^{m}\left\{-\operatorname{tr}\left[V_{\boldsymbol{\vartheta}_0}^{-1}c_{\boldsymbol{\vartheta}_0,\nu}\boldsymbol{Y}_{n-\nu}\boldsymbol{Y}_{n-\nu'}^T c_{\boldsymbol{\vartheta},\nu'}^T V_{\boldsymbol{\vartheta}_0}^{-1}(\partial_i V_{\boldsymbol{\vartheta}_0})\right] + 2\boldsymbol{Y}_{n-\nu}^T c_{\boldsymbol{\vartheta}_0,\nu}^{(i),T} V_{\boldsymbol{\vartheta}_0}^{-1} c_{\boldsymbol{\vartheta}_0,\nu'}\boldsymbol{Y}_{n-\nu'}\right\}.$$

It is convenient to also introduce the notations

$$\mathcal{Y}_{m,n} = \begin{pmatrix} Y_{m,n}^{(1)} & \cdots & Y_{m,n}^{(r)} \end{pmatrix}^T \quad \text{and} \quad \mathcal{Z}_{m,n} = \begin{pmatrix} Z_{m,n}^{(1)} & \cdots & Z_{m,n}^{(r)} \end{pmatrix}^T. \tag{3.2.55}$$

The rest of the proof proceeds in three steps: first we show that, for each $m \in \mathbb{N}$, the sequence $L^{-1/2}\sum_n [\mathcal{Y}_{m,n} - \mathbb{E}\mathcal{Y}_{m,n}]$ is asymptotically normally distributed with asymptotic covariance matrix I_m, and that I_m converges to $I(\boldsymbol{\vartheta}_0)$ as m tends to infinity. In the second step we prove that $L^{-1/2}\sum_n [\mathcal{Z}_{m,n} - \mathbb{E}\mathcal{Z}_{m,n}]$ goes to zero uniformly in L, as $m \to \infty$, and the last step is devoted to proving the asymptotic normality of $L^{-1/2}\nabla_{\boldsymbol{\vartheta}}\mathscr{L}\left(\boldsymbol{\vartheta}_0, \boldsymbol{y}^L\right)$.

Step 1 Since $\boldsymbol{Y}$ is stationary, it is clear that $\mathcal{Y}_m$ is a stationary process. Moreover, the strong-mixing coefficients $\alpha_{\mathcal{Y}_m}(k)$ of $\mathcal{Y}_m$ satisfy $\alpha_{\mathcal{Y}_m}(k) \leqslant \alpha_{\boldsymbol{Y}}(\max\{0, k-m\})$ because $\mathcal{Y}_{m,n}$ depends only on the finitely many values $\boldsymbol{Y}_{n-m}, \ldots, \boldsymbol{Y}_n$ of $\boldsymbol{Y}$ (see Bradley, 2007, Remark 1.8 b)). In particular, by Assumption D9, the strong-mixing coefficients of the processes $\mathcal{Y}_m$ satisfy the summability condition $\sum_k [\alpha_{\mathcal{Y}_m}(k)]^{\delta/(2+\delta)} < \infty$. Since, by the Cramér–Wold device, weak convergence of the sequence $L^{-1/2}\sum_{n=1}^L [\mathcal{Y}_{m,n} - \mathbb{E}\mathcal{Y}_{m,n}]$ to a multivariate normal distribution with mean zero and covariance matrix Σ is equivalent to the condition that, for every vector $\boldsymbol{u} \in \mathbb{R}^r$, the sequence $L^{-1/2}\boldsymbol{u}^T\sum_{n=1}^L [\mathcal{Y}_{m,n} - \mathbb{E}\mathcal{Y}_{m,n}]$ converges to a one-dimensional normal distribution with mean zero and variance $\boldsymbol{u}^T\Sigma\boldsymbol{u}$, we can apply the Central Limit Theorem for univariate strongly mixing processes (Herrndorf, 1984),(Ibragimov, 1962, Theorem 1.7) to obtain that

$$\frac{1}{\sqrt{L}}\sum_{n=1}^{L}[\mathcal{Y}_{m,n} - \mathbb{E}\mathcal{Y}_{m,n}] \xrightarrow[L\to\infty]{d} \mathscr{N}(\boldsymbol{0}_r, I_m), \quad \text{where} \quad I_m = \sum_{\Delta\in\mathbb{Z}} \mathbb{C}\mathrm{ov}\left(\mathcal{Y}_{m,n}; \mathcal{Y}_{m,n+\Delta}\right). \tag{3.2.56}$$

The claim that I_m converges to $I(\boldsymbol{\vartheta}_0)$ will follow if we can show that

$$\mathbb{C}\mathrm{ov}\left(Y_{m,n}^{(k)}; Y_{m,n+\Delta}^{(l)}\right) \xrightarrow[m\to\infty]{} \mathbb{C}\mathrm{ov}\left(\ell_{\boldsymbol{\vartheta}_0,n}^{(k)}; \ell_{\boldsymbol{\vartheta}_0,n+\Delta}^{(l)}\right), \quad \forall\Delta \in \mathbb{Z}, \tag{3.2.57}$$

and that $\left|\mathbb{C}\mathrm{ov}\left(Y_{m,n}^{(k)}; Y_{m,n+\Delta}^{(l)}\right)\right|$ is dominated by an absolutely summable sequence. For the first condition, we note that the bilinearity of $\mathbb{C}\mathrm{ov}(\cdot;\cdot)$ implies that

$$\begin{aligned}\mathbb{C}\mathrm{ov}\left(Y_{m,n}^{(k)}; Y_{m,n+\Delta}^{(l)}\right) - \mathbb{C}\mathrm{ov}\left(\ell_{\boldsymbol{\vartheta}_0,n}^{(k)}; \ell_{\boldsymbol{\vartheta}_0,n+\Delta}^{(l)}\right) =& \mathbb{C}\mathrm{ov}\left(Y_{m,n}^{(k)}; Y_{m,n+\Delta}^{(l)} - \ell_{\boldsymbol{\vartheta}_0,n+\Delta}^{(l)}\right)\\ &+ \mathbb{C}\mathrm{ov}\left(Y_{m,n}^{(k)} - \ell_{\boldsymbol{\vartheta}_0,n}^{(k)}; \ell_{\boldsymbol{\vartheta}_0,n+\Delta}^{(l)}\right).\end{aligned}$$

These two terms can be treated in a similar manner so we restrict our attention to the second one. The definitions of $Y_{m,n}^{(i)}$ (Eq. (3.2.54a)) and $\ell_{\boldsymbol{\vartheta},n}^{(i)}$ (Eq. (3.2.44)) allow us to compute

$$Y_{m,n}^{(k)} - \ell_{\boldsymbol{\vartheta}_0,n}^{(k)} = \sum_{\substack{\nu,\nu'\\ \max\{\nu,\nu'\}>m}} \left[\mathrm{tr}\left[V_{\boldsymbol{\vartheta}_0}^{-1} c_{\boldsymbol{\vartheta}_0,\nu} \boldsymbol{Y}_{n-\nu}\boldsymbol{Y}_{n-\nu'}^T c_{\boldsymbol{\vartheta},\nu'}^T V_{\boldsymbol{\vartheta}_0}^{-1}\partial_i V_{\boldsymbol{\vartheta}_0}\right] - 2\boldsymbol{Y}_{n-\nu}^T c_{\boldsymbol{\vartheta}_0,\nu}^{(i),T} V_{\boldsymbol{\vartheta}_0}^{-1} c_{\boldsymbol{\vartheta}_0,\nu'}\boldsymbol{Y}_{n-\nu'}\right].$$

As a consequence of the Cauchy–Schwarz inequality, Assumption D8 and the exponential bounds in Eq. (3.2.26), we therefore obtain that $\mathbb{V}\mathrm{ar}\left(Y_{m,n}^{(k)} - \ell_{\vartheta_0,n}^{(k)}\right) \leqslant C\rho^m$ independent of n. The L^2-continuity of $\mathbb{C}\mathrm{ov}(\cdot;\cdot)$ thus implies that the sequence $\mathbb{C}\mathrm{ov}\left(Y_{m,n}^{(k)} - \ell_{\vartheta_0,n}^{(k)}; \ell_{\vartheta_0,n+\Delta}^{(l)}\right)$ converges to zero as m tends to infinity at an exponential rate uniformly in Δ. The existence of a summable sequence dominating $\left|\mathbb{C}\mathrm{ov}\left(Y_{m,n}^{(k)}; Y_{m,n+\Delta}^{(l)}\right)\right|$ is ensured by the arguments given in the proof of Lemma 3.18, reasoning as in the derivation of Eqs. (3.2.48) and (3.2.49).

Step 2 In this step we shall show that there exist positive constants C and $\rho < 1$, independent of L, such that

$$\operatorname{tr} \mathbb{V}\mathrm{ar}\left(\frac{1}{\sqrt{L}} \sum_{n=1}^{L} \mathcal{Z}_{m,n}\right) \leqslant C\rho^m, \quad \mathcal{Z}_{m,n} \text{ given in Eq. (3.2.55).} \tag{3.2.58}$$

Since

$$\operatorname{tr} \mathbb{V}\mathrm{ar}\left(\frac{1}{\sqrt{L}} \sum_{n=1}^{L} \mathcal{Z}_{m,n}\right) \leqslant 2\left[\operatorname{tr} \mathbb{V}\mathrm{ar}\left(\frac{1}{\sqrt{L}} \sum_{n=1}^{L} \mathcal{U}_{m,n}\right) + \operatorname{tr} \mathbb{V}\mathrm{ar}\left(\frac{1}{\sqrt{L}} \sum_{n=1}^{L} \mathcal{V}_{m,n}\right)\right], \tag{3.2.59}$$

it suffices to consider the latter two terms. We first observe that

$$\begin{aligned}\operatorname{tr} \mathbb{V}\mathrm{ar}\left(\frac{1}{\sqrt{L}} \sum_{n=1}^{L} \mathcal{U}_{m,n}\right) &= \frac{1}{L} \operatorname{tr} \sum_{n,n'=1}^{L} \mathbb{C}\mathrm{ov}\left(\mathcal{U}_{m,n}; \mathcal{U}_{m,n'}\right) \\ &= \frac{1}{L} \sum_{k,l=1}^{r} \sum_{\Delta=-L+1}^{L-1} \left(L - |\Delta|\right) \mathfrak{u}_{m,\Delta}^{(k,l)} \leqslant \sum_{k,l=1}^{r} \sum_{\Delta\in\mathbb{Z}} \left|\mathfrak{u}_{m,\Delta}^{(k,l)}\right|,\end{aligned} \tag{3.2.60}$$

where

$$\begin{aligned}\mathfrak{u}_{m,\Delta}^{(k,l)} &= \mathbb{C}\mathrm{ov}\left(U_{m,n}^{(k)}; U_{m,n+\Delta}^{(l)}\right) \\ &= \sum_{\substack{\nu,\mu=0 \\ \nu',\mu'=m+1}}^{m} \mathbb{C}\mathrm{ov}\Big(-\operatorname{tr}\left[V_{\vartheta_0}^{-1} c_{\vartheta_0,\nu} \boldsymbol{Y}_{n-\nu} \boldsymbol{Y}_{n-\nu'}^T c_{\vartheta,\nu'}^T V_{\vartheta_0}^{-1} \partial_k V_{\vartheta_0}\right] + \boldsymbol{Y}_{n-\nu}^T c_{\vartheta_0,\nu}^{(k),T} V_{\vartheta_0}^{-1} c_{\vartheta_0,\nu'} \boldsymbol{Y}_{n-\nu'}\,; \\ &\qquad -\operatorname{tr}\left[V_{\vartheta_0}^{-1} c_{\vartheta_0,\mu} \boldsymbol{Y}_{n+\Delta-\mu} \boldsymbol{Y}_{n+\Delta-\mu'}^T c_{\vartheta,\mu'}^T V_{\vartheta_0}^{-1} \partial_l V_{\vartheta_0}\right] + \boldsymbol{Y}_{n+\Delta-\mu}^T c_{\vartheta_0,\mu}^{(l),T} V_{\vartheta_0}^{-1} c_{\vartheta_0,\mu'} \boldsymbol{Y}_{n+\Delta-\mu'}\Big).\end{aligned}$$

As before, under Assumption D8, the Cauchy–Schwarz inequality and the exponential bounds (3.2.26) and (3.2.39) for $\|c_{\vartheta_0,\nu}\|$ and $\left\|c_{\vartheta_0,\nu}^{(k)}\right\|$ imply that $\left|\mathfrak{u}_{m,\Delta}^{(k,l)}\right| < C\rho^m$. By arguments similar to the ones used in the proof of Lemma 3.17 it can be shown that Davydov's inequality implies that for $m < \lfloor \Delta/2 \rfloor$ it holds that

$$\left|\mathfrak{u}_{m,\Delta}^{(k,l)}\right| \leqslant C \sum_{\nu=0}^{\infty} \sum_{\nu'=m+1}^{\infty} \sum_{\mu,\mu'=0}^{\lfloor \Delta/2 \rfloor} \rho^{\nu+\nu'+\mu+\mu'} \left[\alpha_{\boldsymbol{Y}}\left(\left\lfloor \frac{\Delta}{2} \right\rfloor\right)\right]^{\delta/(2+\delta)} + C \sum_{\nu,\nu'=0}^{\infty} \sum_{\substack{\mu,\mu' \\ \max\{\mu,\mu'\} > \lfloor \Delta/2 \rfloor}} \rho^{\nu+\nu'+\mu+\mu'}$$

$$\leqslant C\rho^m \left\{ \left[\alpha_Y\left(\left\lfloor\frac{\Delta}{2}\right\rfloor\right)\right]^{\delta/(2+\delta)} + \rho^{\Delta/2} \right\}.$$

It thus follows that, independent of the value of k and l,

$$\sum_{\Delta=0}^{\infty} \left|\mathfrak{u}_{m,\Delta}^{(k,l)}\right| = \sum_{\Delta=0}^{2m} \left|\mathfrak{u}_{m,\Delta}^{(k,l)}\right| + \sum_{\Delta=2m+1}^{\infty} \left|\mathfrak{u}_{m,\Delta}^{(k,l)}\right| \leqslant C\rho^m \left\{ m + \sum_{\Delta=0}^{\infty} \left[\alpha_Y(\Delta)\right]^{\delta/(2+\delta)} \right\},$$

and therefore, by Eq. (3.2.60), that $\operatorname{tr}\mathbb{V}\mathrm{ar}\left(L^{-1/2}\sum_{n=1}^{L}\mathcal{U}_{m,n}\right) \leqslant C\rho^m$. In an analogous way one can show that $\operatorname{tr}\mathbb{V}\mathrm{ar}\left(L^{-1/2}\sum_{n=1}^{L}\mathcal{V}_{m,n}\right) \leqslant C\rho^m$, and thus the claim (3.2.58) follows with Eq. (3.2.59).

Step 3 In step 1 it has been shown that $L^{-1/2}\sum_n\left[\mathcal{Y}_{m,n} - \mathbb{E}\mathcal{Y}_{m,n}\right] \xrightarrow[L\to\infty]{d} \mathscr{N}(\mathbf{0}_r, I_m)$, and that $I_m \xrightarrow[m\to\infty]{} I(\boldsymbol{\vartheta}_0)$. In particular, the limiting normal random variables with covariances I_m converge weakly to a normal random variable with covariance matrix $I(\boldsymbol{\vartheta}_0)$. Step 2 together with the multivariate Chebyshev inequality (Lemma 3.19) implies that, for every $\epsilon > 0$,

$$\begin{aligned}
&\lim_{m\to\infty}\limsup_{L\to\infty}\mathbb{P}\left(\left\|\frac{1}{\sqrt{L}}\nabla_{\boldsymbol{\vartheta}}\mathscr{L}\left(\boldsymbol{\vartheta}_0, \boldsymbol{y}^L\right) - \frac{1}{\sqrt{L}}\sum_{n=1}^{L}\left[\mathcal{Y}_{m,n} - \mathbb{E}\mathcal{Y}_{m,n}\right]\right\| > \epsilon\right)\\
&= \lim_{m\to\infty}\limsup_{L\to\infty}\mathbb{P}\left(\left\|\frac{1}{\sqrt{L}}\sum_{n=1}^{L}\left[\mathcal{Z}_{m,n} - \mathbb{E}\mathcal{Z}_{m,n}\right]\right\| > \epsilon\right)\\
&\leqslant \lim_{m\to\infty}\limsup_{L\to\infty}\frac{r}{\epsilon^2}\operatorname{tr}\mathbb{V}\mathrm{ar}\left(\frac{1}{\sqrt{L}}\sum_{n=1}^{L}\mathcal{Z}_{m,n}\right)\\
&\leqslant \lim_{m\to\infty}\frac{Cr}{\epsilon^2}\rho^m = 0.
\end{aligned}$$

Brockwell and Davis (1991, Proposition 6.3.9) thus shows that

$$\frac{1}{\sqrt{L}}\nabla_{\boldsymbol{\vartheta}}\mathscr{L}\left(\boldsymbol{\vartheta}_0, \boldsymbol{y}^L\right) \xrightarrow[L\to\infty]{d} \mathscr{N}\left(\mathbf{0}_r, I(\boldsymbol{\vartheta}_0)\right),$$

which completes the proof. □

A very important step in the proof of asymptotic normality of quasi maximum likelihood estimators is to establish that the Fisher information matrix J, evaluated at the true parameter value, is non-singular. We shall now show that Assumption D10 is sufficient to ensure that J^{-1} exists for linear state space models. For vector ARMA processes, formulæ similar to Eqs. (3.2.62) below have been derived in the literature (see, e. g., Klein, Mélard and Saidi, 2008; Klein and Neudecker, 2000), but have not been used to derive criteria for J being non-singular. Our arguments are similar to Boubacar Mainassara and Francq (2011, Lemma 4).

Lemma 3.22 *Assume that Assumptions D1 to D4, D7 and D10 hold. With probability one, the matrix* $J = \lim_{L\to\infty} L^{-1}\nabla^2_{\boldsymbol{\vartheta}}\widehat{\mathscr{L}}(\boldsymbol{\vartheta}_0, \boldsymbol{y}^L)$ *exists and is non-singular.*

Proof We note that, by Lemma 3.20, ii), it is enough to show that $\lim_{L\to\infty} L^{-1}\nabla^2_{\boldsymbol{\vartheta}}\mathscr{L}(\boldsymbol{\vartheta}_0, \boldsymbol{y}^L)$ exists and is non-singular. As seen earlier, for every $i = 1, \ldots, r$,

$$\partial_i l_{\boldsymbol{\vartheta},n} = \operatorname{tr}\left[V_{\boldsymbol{\vartheta}}^{-1}\left(\mathbf{1}_d - \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}^T V_{\boldsymbol{\vartheta}}^{-1}\right)\partial_i V_{\boldsymbol{\vartheta}}\right] + 2\left(\partial_i \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}^T\right) V_{\boldsymbol{\vartheta}}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}. \tag{3.2.61}$$

Consequently, the second partial derivatives are given by

$$\begin{aligned}
\partial^2_{i,j} l_{\boldsymbol{\vartheta},n} = \operatorname{tr}\Big[&V_{\boldsymbol{\vartheta}}^{-1}\left(\partial^2_{i,j}V_{\boldsymbol{\vartheta}}\right) - V_{\boldsymbol{\vartheta}}^{-1}\left(\partial_i V_{\boldsymbol{\vartheta}}\right)V_{\boldsymbol{\vartheta}}^{-1}\left(\partial_j V_{\boldsymbol{\vartheta}}\right) - V_{\boldsymbol{\vartheta}}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}^T V_{\boldsymbol{\vartheta}}^{-1}\left(\partial^2_{i,j}V_{\boldsymbol{\vartheta}}\right)\\
&+V_{\boldsymbol{\vartheta}}^{-1}\left(\partial_i V_{\boldsymbol{\vartheta}}\right)V_{\boldsymbol{\vartheta}}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}^T V_{\boldsymbol{\vartheta}}^{-1}\left(\partial_j V_{\boldsymbol{\vartheta}}\right) + V_{\boldsymbol{\vartheta}}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}^T V_{\boldsymbol{\vartheta}}^{-1}\left(\partial_i V_{\boldsymbol{\vartheta}}\right)V_{\boldsymbol{\vartheta}}^{-1}\left(\partial_j V_{\boldsymbol{\vartheta}}\right)\\
&-V_{\boldsymbol{\vartheta}}^{-1}\left(\partial_i V_{\boldsymbol{\vartheta}}\right)V_{\boldsymbol{\vartheta}}^{-1}\left(\partial_i \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}^T\right)\Big] + 2\left(\partial^2_{i,j}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}^T\right)V_{\boldsymbol{\vartheta}}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n} + 2\left(\partial_i \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}^T\right)V_{\boldsymbol{\vartheta}}^{-1}\left(\partial_j \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}\right)\\
-2\operatorname{tr}\Big[&V_{\boldsymbol{\vartheta}}^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}\left(\partial_i \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}^T\right)V_{\boldsymbol{\vartheta}}^{-1}\left(\partial_j \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}\right)\Big].
\end{aligned}$$

By Lemma 3.2, iii), $\mathbb{E}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,n} = \mathbf{0}_d$, and by Eq. (3.2.7), $\mathbb{E}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,n}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,n}^T = V_{\boldsymbol{\vartheta}_0}$. The sequence $\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0}$ being the innovations of the process $\boldsymbol{Y}$ implies that $\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,n}$ is orthogonal to the Hilbert space spanned by $\{\boldsymbol{Y}_t, t < n\}$, of which, by Eq. (3.2.25), both $\partial_i\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,n}$ and $\partial^2_{i,j}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,n}$ are elements. It thus follows that

$$\mathbb{E}\left[\partial^2_{i,j} l_{\boldsymbol{\vartheta}_0,n}\right] = \operatorname{tr}\left[V_{\boldsymbol{\vartheta}_0}^{-1}\left(\partial_i V_{\boldsymbol{\vartheta}_0}\right)V_{\boldsymbol{\vartheta}_0}^{-1}\left(\partial_j V_{\boldsymbol{\vartheta}_0}\right)\right] + 2\mathbb{E}\left[\left(\partial_i \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}^T\right)V_{\boldsymbol{\vartheta}}^{-1}\left(\partial_j \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta},n}\right)\right].$$

Equations (3.2.25), (3.2.38) and (3.2.41), the ergodicity of $\boldsymbol{Y}$, and Krengel (1985, Theorem 4.3) imply that the sequence $\partial^2_{i,j} l_{\boldsymbol{\vartheta}_0}$ is ergodic, and Birkhoff's Ergodic Theorem shows that

$$\frac{1}{L}\nabla^2_{\boldsymbol{\vartheta}}\mathscr{L}\left(\boldsymbol{\vartheta}_0, \boldsymbol{y}^L\right) = \frac{1}{L}\sum_{n=1}^{L}\nabla^2_{\boldsymbol{\vartheta}} l_{\boldsymbol{\vartheta}_0,n} \xrightarrow[L\to\infty]{\text{a.s}} \mathbb{E}\left[\nabla^2_{\boldsymbol{\vartheta}} l_{\boldsymbol{\vartheta}_0,n}\right] =: J_1 + J_2,$$

where

$$J_1 = 2\mathbb{E}\left[\left(\nabla_{\boldsymbol{\vartheta}}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,1}\right)^T V_{\boldsymbol{\vartheta}_0}^{-1}\left(\nabla_{\boldsymbol{\vartheta}}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,1}\right)\right] \tag{3.2.62a}$$

and

$$J_2 = \left(\operatorname{tr}\left[V_{\boldsymbol{\vartheta}_0}^{-1/2}\left(\partial_i V_{\boldsymbol{\vartheta}_0}\right)V_{\boldsymbol{\vartheta}_0}^{-1}\left(\partial_j V_{\boldsymbol{\vartheta}_0}\right)V_{\boldsymbol{\vartheta}_0}^{-1/2}\right]\right)_{ij}. \tag{3.2.62b}$$

The matrix J_2 can be factorized as

$$J_2 = \begin{pmatrix}\boldsymbol{b}_1^T\\ \vdots\\ \boldsymbol{b}_r^T\end{pmatrix}\begin{pmatrix}\boldsymbol{b}_1 & \ldots & \boldsymbol{b}_r\end{pmatrix}, \quad \boldsymbol{b}_m = \left(V_{\boldsymbol{\vartheta}_0}^{-1/2}\otimes V_{\boldsymbol{\vartheta}_0}^{-1/2}\right)\operatorname{vec}\left(\partial_m V_{\boldsymbol{\vartheta}_0}\right), \tag{3.2.63}$$

and is thus positive semidefinite. Because J_1 is positive semidefinite as well, proving that J is non-singular is equivalent to proving that for any non-zero vector $\boldsymbol{c}\in\mathbb{R}^r$, the numbers $\boldsymbol{c}^T J_i \boldsymbol{c}$,

$i = 1,2$, are not both zero. Assume, for the sake of contradiction, that there exists such a vector $\boldsymbol{c} = (c_1, \ldots, c_r)^T$. The condition $\boldsymbol{c}^T J_1 \boldsymbol{c}$ implies that, almost surely, $\sum_{k=1}^r c_k \partial_k \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,n} = \mathbf{0}_d$, for all $n \in \mathbb{Z}$. It thus follows from the infinite-order moving average representation (3.2.8b) that

$$\sum_{\nu=1}^{\infty} \sum_{k=1}^{r} c_k \left(\partial_k \mathscr{M}_{\boldsymbol{\vartheta}_0,\nu}\right) \boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,-\nu} = \mathbf{0}_d, \tag{3.2.64}$$

where the Markov parameters $\mathscr{M}_{\boldsymbol{\vartheta},\nu}$ are given by $\mathscr{M}_{\boldsymbol{\vartheta},\nu} = -H_{\boldsymbol{\vartheta}} F_{\boldsymbol{\vartheta}}^{\nu-1} K_{\boldsymbol{\vartheta}}$, $\nu \geqslant 1$. Since the sequence $\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0}$ is uncorrelated with positive definite covariance matrix, Eq. (3.2.64) implies that

$$\sum_{k=1}^{r} c_k \left(\partial_k \mathscr{M}_{\boldsymbol{\vartheta}_0,\nu}\right) = \mathbf{0}_d, \quad \forall \nu \in \mathbb{N}.$$

Using the relation $\operatorname{vec}(ABC) = \left(C^T \otimes A\right) \operatorname{vec} B$ (Bernstein, 2005, Proposition 7.1.9), we see that the last display is equivalent to $\nabla_{\boldsymbol{\vartheta}} \left(\left[K_{\boldsymbol{\vartheta}_0}^T \otimes H_{\boldsymbol{\vartheta}_0}\right] \operatorname{vec} F_{\boldsymbol{\vartheta}_0}^{\nu-1}\right) \boldsymbol{c} = \mathbf{0}_{d^2}$ for every $\nu \in \mathbb{N}$. In view of Eq. (3.2.63), the condition $\boldsymbol{c}^T J_2 \boldsymbol{c} = 0$ implies that $\left(\nabla_{\boldsymbol{\vartheta}} \operatorname{vec} V_{\boldsymbol{\vartheta}_0}\right) \boldsymbol{c} = \mathbf{0}_{d^2}$. By the definition of $\psi_{\boldsymbol{\vartheta},j}$ in Eq. (3.2.21) it thus follows that $\nabla_{\boldsymbol{\vartheta}} \psi_{\boldsymbol{\vartheta}_0,j} \boldsymbol{c} = \mathbf{0}_{(j+2)d^2}$, for every $j \in \mathbb{N}$, which, by Assumption D10, is equivalent to the contradiction that $\boldsymbol{c} = \mathbf{0}_r$. □

Proof (of Theorem 3.8) Since $\hat{\boldsymbol{\vartheta}}^L$ converges almost surely to $\boldsymbol{\vartheta}_0$ by the consistency result proved in Theorem 3.7, and $\boldsymbol{\vartheta}_0$ is an element of the interior of Θ by Assumption D6, the estimate $\hat{\boldsymbol{\vartheta}}^L$ is an element of the interior of Θ eventually almost surely. The assumed smoothness of the parametrization (Assumption D7) implies that the extremal property of $\hat{\boldsymbol{\vartheta}}^L$ can be expressed as the first order condition $\nabla_{\boldsymbol{\vartheta}} \widehat{\mathscr{L}}(\hat{\boldsymbol{\vartheta}}^L, \boldsymbol{y}^L) = \mathbf{0}_r$. A Taylor expansion of $\boldsymbol{\vartheta} \mapsto \nabla_{\boldsymbol{\vartheta}} \widehat{\mathscr{L}}(\boldsymbol{\vartheta}, \boldsymbol{y}^L)$ around the point $\boldsymbol{\vartheta}_0$ shows that there exist parameter vectors $\boldsymbol{\vartheta}_i \in \Theta$ of the form $\boldsymbol{\vartheta}_i = \boldsymbol{\vartheta}_0 + c_i(\hat{\boldsymbol{\vartheta}}^L - \boldsymbol{\vartheta}_0)$, $0 \leqslant c_i \leqslant 1$, such that

$$\mathbf{0}_r = L^{-1/2} \nabla_{\boldsymbol{\vartheta}} \widehat{\mathscr{L}}(\boldsymbol{\vartheta}_0, \boldsymbol{y}^L) + \frac{1}{L} \nabla_{\boldsymbol{\vartheta}}^2 \widehat{\mathscr{L}}(\underline{\boldsymbol{\vartheta}}^L, \boldsymbol{y}^L) L^{1/2} \left(\hat{\boldsymbol{\vartheta}}^L - \boldsymbol{\vartheta}_0\right), \tag{3.2.65}$$

where $\nabla_{\boldsymbol{\vartheta}}^2 \widehat{\mathscr{L}}(\underline{\boldsymbol{\vartheta}}^L, \boldsymbol{y}^L)$ denotes the matrix whose ith row, $i = 1, \ldots, r$, is equal to the ith row of $\nabla_{\boldsymbol{\vartheta}}^2 \widehat{\mathscr{L}}(\boldsymbol{\vartheta}_i, \boldsymbol{y}^L)$. By Lemma 3.21 the first term on the right hand side converges weakly to a multivariate normal random variable with mean zero and covariance matrix $I = I(\boldsymbol{\vartheta}_0)$. As in Lemma 3.12 one can show that the sequence

$$\left(\boldsymbol{\vartheta} \mapsto L^{-1} \nabla_{\boldsymbol{\vartheta}}^3 \widehat{\mathscr{L}}(\boldsymbol{\vartheta}, \boldsymbol{y}^L)\right)_{L \in \mathbb{N}} \tag{3.2.66}$$

of random functions converges almost surely uniformly to the continuous function $\boldsymbol{\vartheta} \mapsto \nabla_{\boldsymbol{\vartheta}}^3 \mathscr{Q}(\boldsymbol{\vartheta})$ taking values in the space $\mathbb{R}^{r \times r \times r}$. Since on the compact space Θ this function is bounded in the operator norm obtained from identifying $\mathbb{R}^{r \times r \times r}$ with the space of linear functions from $\mathbb{R}^r$ to $M_r(\mathbb{R})$, the sequence (3.2.66) is almost surely uniformly bounded, and we obtain that

$$\left\| \frac{1}{L}\nabla^2_{\boldsymbol{\vartheta}}\widehat{\mathscr{L}}(\underline{\boldsymbol{\vartheta}}^L, \boldsymbol{y}^L) - \frac{1}{L}\nabla^2_{\boldsymbol{\vartheta}}\widehat{\mathscr{L}}(\boldsymbol{\vartheta}_0, \boldsymbol{y}^L) \right\| \leqslant \sup_{\boldsymbol{\vartheta}\in\Theta} \left\| \frac{1}{L}\nabla^3_{\boldsymbol{\vartheta}}\widehat{\mathscr{L}}(\boldsymbol{\vartheta}, \boldsymbol{y}^L) \right\| \left\| \underline{\boldsymbol{\vartheta}}^L - \boldsymbol{\vartheta}_0 \right\| \xrightarrow[L\to\infty]{\text{a.s.}} 0,$$

because, by Theorem 3.7, the second factor almost surely converges to zero as L tends to infinity. It follows from Lemma 3.22 that $L^{-1}\nabla^2_{\boldsymbol{\vartheta}}\widehat{\mathscr{L}}(\underline{\boldsymbol{\vartheta}}^L, \boldsymbol{y}^L)$ converges to the matrix J almost surely, and thus from Eq. (3.2.65) that

$$L^{1/2}\left(\hat{\boldsymbol{\vartheta}}^L - \boldsymbol{\vartheta}_0\right) \xrightarrow{d} \mathscr{N}\left(\mathbf{0}_r, J^{-1}IJ^{-1}\right),$$

as $L \to \infty$. This shows Eq. (3.2.22) and completes the proof. □

In practice, one is interested in also estimating the asymptotic covariance matrix Ξ, which is useful in constructing confidence regions for the estimated parameters or in performing statistical tests. This problem has been considered in the framework of estimating weak VARMA processes in Boubacar Mainassara and Francq (2011) where the following procedure has been suggested, which is also applicable in our set-up. First, $J(\boldsymbol{\vartheta}_0)$ is estimated consistently by $\hat{J}^L = L^{-1}\nabla^2\widehat{\mathscr{L}_{\boldsymbol{\vartheta}}}\left(\hat{\boldsymbol{\vartheta}}^L, \boldsymbol{y}^L\right)$. For the computation of $\hat{J}^L$ we rely on the fact that the Kalman filter can not only be used to evaluate the Gaussian log-likelihood of a state space model but also its gradient and Hessian. The most straightforward, but computationally burdensome way of achieving this is by direct differentiation of the Kalman filter equations, which results in increasing the number of passes through the filter to $r+1$ and $r(r+3)/2$ for the gradient and the Hessian, respectively. More sophisticated algorithms, including the Kalman smoother and/or the backward filter have been devised and can be found in Kulikova and Semoushin (2006); Segal and Weinstein (1989). The construction of a consistent estimator of $I = I(\boldsymbol{\vartheta}_0)$ is based on the observation that $I = \sum_{\Delta\in\mathbb{Z}}\operatorname{Cov}(\ell_{\boldsymbol{\vartheta}_0,n}, \ell_{\boldsymbol{\vartheta}_0,n+\Delta})$, where $\ell_{\boldsymbol{\vartheta}_0,n} = \nabla_{\boldsymbol{\vartheta}}\left[\log\det V_{\boldsymbol{\vartheta}_0} + \boldsymbol{\varepsilon}^T_{\boldsymbol{\vartheta}_0,n}V^{-1}_{\boldsymbol{\vartheta}_0}\boldsymbol{\varepsilon}_{\boldsymbol{\vartheta}_0,n}\right]$. Assuming that $(\ell_{\boldsymbol{\vartheta}_0,n})_{n\in\mathbb{N}^+}$ admits an infinite-order AR representation $\Phi(\mathrm{B})\ell_{\boldsymbol{\vartheta}_0,n} = \boldsymbol{U}_n$, where $\Phi(z) = \mathbf{1}_r + \sum_{i=1}^\infty \Phi_i z^i$ and $(\boldsymbol{U}_n)_{n\in\mathbb{N}^+}$ is a weak white noise with covariance matrix $\Sigma_{\boldsymbol{U}}$, it follows from the interpretation of $I/(2\pi)$ as the value of the spectral density of $(\ell_{\boldsymbol{\vartheta}_0,n})_{n\in\mathbb{N}^+}$ at frequency zero that I can also be written as $I = \Phi^{-1}(1)\Sigma_{\boldsymbol{U}}\Phi(1)^{-1}$. The idea is to fit a long autoregression to $(\ell_{\hat{\boldsymbol{\vartheta}}^L,n})_{n=1,\ldots L}$, the empirical counterparts of $(\ell_{\boldsymbol{\vartheta}_0,n})_{n\in\mathbb{N}^+}$ which are defined by replacing $\boldsymbol{\vartheta}_0$ with the estimate $\hat{\boldsymbol{\vartheta}}^L$ in the definition of $\ell_{\boldsymbol{\vartheta}_0,n}$. This is done by choosing an integer $s > 0$, and performing a least-squares regression of $\ell_{\hat{\boldsymbol{\vartheta}}^L,n}$ on $\ell_{\hat{\boldsymbol{\vartheta}}^L,n-1}, \ldots, \ell_{\hat{\boldsymbol{\vartheta}}^L,n-s}$, $s+1 \leqslant n \leqslant L$. Denoting by $\hat{\Phi}^L_s(z) = \mathbf{1}_r + \sum_{i=1}^s \hat{\Phi}^L_{i,s} z^i$ the obtained empirical autoregressive polynomial and by $\hat{\Sigma}^L_s$ the empirical covariance matrix of the residuals of the regression, it was claimed in Boubacar Mainassara and Francq (2011, Theorem 4) that under the additional assumption $\mathbb{E}\left[\|\boldsymbol{\varepsilon}_n\|^{8+\delta}\right] < \infty$ the spectral estimator $\hat{I}^L_s = \left(\hat{\Phi}^L_s(1)\right)^{-1}\hat{\Sigma}^L_s\left(\hat{\Phi}^L_s(1)\right)^{T,-1}$ converges to I in probability as $L, s \to \infty$ if $s^3/L \to 0$. The covariance matrix of $\hat{\boldsymbol{\vartheta}}^L$ is then estimated consistently as

$$\widehat{\Xi}_s^L = \frac{1}{L}\left(\hat{J}^L\right)^{-1}\hat{I}_s^L\left(\hat{J}^L\right)^{-1}. \tag{3.2.67}$$

In the simulation study performed in Section 3.4.2, this estimator for Ξ performs convincingly.

3.3. Quasi maximum likelihood estimation for Lévy-driven multivariate CARMA processes

In this section we pursue the second main topic of the present chapter, a detailed investigation of the asymptotic properties of the quasi maximum likelihood estimator of discretely observed multivariate continuous-time autoregressive moving average processes. We will make use of the equivalence between MCARMA and continuous-time linear state space models, as well as of the important observation that the state space structure of a continuous-time process is preserved under equidistant sampling, which allows for the results of the previous section to be applied. The conditions we need to impose on the parametrization of the models under consideration are therefore closely related to the assumptions made in the discrete-time case, except that the mixing and ergodicity assumptions D4 and D9 are automatically satisfied (Marquardt and Stelzer, 2007, Proposition 3.34).

We start the section with a short recapitulation of the definition and basic properties of Lévy-driven continuous-time ARMA processes; this is followed by a discussion of the second-order properties of discretely observed CARMA process, leading to a set of accessible identifiability conditions. Section 3.3.4 contains our main result about the consistency and asymptotic normality of the quasi maximum likelihood estimator for equidistantly sampled MCARMA processes.

3.3.1. Lévy-driven multivariate CARMA processes and continuous-time state space models

A natural source of randomness in the specification of continuous-time stochastic processes are Lévy processes. For a thorough discussion of these processes we refer the reader to the monographs Applebaum (2004); Bertoin (1996); Sato (1999).

Definition 3.23 (Lévy process) A two-sided $\mathbb{R}^m$-valued *Lévy process* $(\boldsymbol{L}(t))_{t\geqslant 0}$ is a stochastic process, defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, with stationary, independent increments, continuous in probability, and satisfying $\boldsymbol{L}(0) = \mathbf{0}_m$ almost surely.

The class of Lévy processes includes many important processes such as Brownian motions, stable processes, and compound Poisson processes as special cases, which makes them very useful in stochastic modelling. Another advantage is that the property of having stationary independent increments implies that Lévy process have a rather particular structure which makes many problems analytically tractable. More precisely, the Lévy–Itô decomposition theorem asserts that every Lévy process can be additively decomposed into a Brownian motion, a compound Poisson process, and a square-integrable pure-jump martingale, where the three terms are independent. This is equivalent to the statement that the characteristic function of a Lévy process $\boldsymbol{L}$ has the special form

$$\mathbb{E}\mathrm{e}^{\mathrm{i}\langle \boldsymbol{u},\boldsymbol{L}(t)\rangle} = \exp\{t\psi^{\boldsymbol{L}}(\boldsymbol{u})\}, \quad \boldsymbol{u}\in\mathbb{R}^m, \quad t\in\mathbb{R}^+, \tag{3.3.1}$$

where the characteristic exponent $\psi^{\boldsymbol{L}}$ has the form

$$\psi^{\boldsymbol{L}}(\boldsymbol{u}) = \mathrm{i}\langle\gamma^{\boldsymbol{L}},\boldsymbol{u}\rangle - \frac{1}{2}\langle\boldsymbol{u},\Sigma^{\mathcal{G}}\boldsymbol{u}\rangle + \int_{\mathbb{R}^m}\left[\mathrm{e}^{\mathrm{i}\langle\boldsymbol{u},\boldsymbol{x}\rangle} - 1 - \mathrm{i}\langle\boldsymbol{u},\boldsymbol{x}\rangle I_{\{\|x\|\leqslant 1\}}\right]\nu^{\boldsymbol{L}}(\mathrm{d}\boldsymbol{x}). \tag{3.3.2}$$

$\gamma^{\boldsymbol{L}}\in\mathbb{R}^m$ is called the *drift vector*, $\Sigma^{\mathcal{G}}$ is a non-negative definite, symmetric $m\times m$ matrix called the *Gaussian covariance matrix*, and the *Lévy measure* $\nu^{\boldsymbol{L}}$ satisfies

$$\nu^{\boldsymbol{L}}(\{\boldsymbol{0}_m\}) = 0, \quad \int_{\mathbb{R}^m}\min(\|\boldsymbol{x}\|^2,1)\nu^{\boldsymbol{L}}(\mathrm{d}\boldsymbol{x}) < \infty.$$

For the present purpose it is enough to know that a Lévy process $\boldsymbol{L}$ has finite kth absolute moments, $k>0$, that is $\mathbb{E}\,\|\boldsymbol{L}(t)\|^k<\infty$, if and only if $\int_{\|\boldsymbol{x}\|\geqslant 1}\|\boldsymbol{x}\|^k\,\nu^{\boldsymbol{L}}(\mathrm{d}\boldsymbol{x})<\infty$ (Sato, 1999, Corollary 25.8), and that the covariance matrix of $\boldsymbol{L}(1)$, if it exists, is given by Sato (1999, Example 25.11)

$$\Sigma^{\boldsymbol{L}} := \mathbb{E}\left(\boldsymbol{L}(1)-\mathbb{E}\boldsymbol{L}(1)\right)\left(\boldsymbol{L}(1)-\mathbb{E}\boldsymbol{L}(1)\right)^T = \Sigma^{\mathcal{G}} + \int_{\|\boldsymbol{x}\|\geqslant 1}\boldsymbol{x}\boldsymbol{x}^T\nu^{\boldsymbol{L}}(\mathrm{d}\boldsymbol{x}).$$

Assumption L2 The Lévy process $\boldsymbol{L}$ has mean zero and finite second moments, which means in terms of the characteristic triplet of $\boldsymbol{L}$ that $\gamma^{\boldsymbol{L}}+\int_{\|\boldsymbol{x}\|\geqslant 1}\boldsymbol{x}\nu^{\boldsymbol{L}}(\mathrm{d}\boldsymbol{x})$ is zero, and that the integral $\int_{\|\boldsymbol{x}\|\geqslant 1}\|\boldsymbol{x}\|^2\,\nu^{\boldsymbol{L}}(\mathrm{d}\boldsymbol{x})$ is finite.

Just like i. i. d. sequences are used in time series analysis to define ARMA processes, Lévy processes can be used to construct (multivariate) continuous-time autoregressive moving average processes, called (M)CARMA processes. If $\boldsymbol{L}$ is a two-sided Lévy process with values in $\mathbb{R}^m$, and $p>q$ are integers, the d-dimensional $\boldsymbol{L}$-driven MCARMA(p,q) process with autoregressive polynomial

$$z\mapsto P(z) := \mathbf{1}_d z^p + A_1 z^{p-1} + \ldots + A_p \in M_d(\mathbb{R}[z]) \tag{3.3.3a}$$

and moving average polynomial

$$z \mapsto Q(z) := B_0 z^q + B_1 z^{q-1} + \ldots + B_q \in M_{d,m}(\mathbb{R}[z]) \tag{3.3.3b}$$

is defined as the solution to the formal differential equation

$$P(\mathrm{D})\boldsymbol{Y}(t) = Q(\mathrm{D})\mathrm{D}\boldsymbol{L}(t), \quad D \equiv \frac{\mathrm{d}}{\mathrm{d}t}. \tag{3.3.4}$$

It is often useful to allow for the dimensions of the driving Lévy process $\boldsymbol{L}$ and the $\boldsymbol{L}$-driven MCARMA process to be different, which is a slight extension of the original definition of Marquardt and Stelzer (2007). The results obtained in that paper remain true if our definition is used. In general, the paths of a Lévy process are not differentiable, so in order to make sense of Eq. (3.3.4), we interpret it as being equivalent to the *state space representation*

$$\mathrm{d}\boldsymbol{G}(t) = \mathcal{A}\boldsymbol{G}(t)\mathrm{d}t + \mathcal{B}\mathrm{d}\boldsymbol{L}(t), \quad \boldsymbol{Y}(t) = \mathcal{C}\boldsymbol{G}(t), \quad t \in \mathbb{R}, \tag{3.3.5}$$

where the matrices $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are given by

$$\mathcal{A} = \begin{pmatrix} 0 & \mathbf{1}_d & 0 & \ldots & 0 \\ 0 & 0 & \mathbf{1}_d & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \ldots & \ldots & 0 & \mathbf{1}_d \\ -A_p & -A_{p-1} & \ldots & \ldots & -A_1 \end{pmatrix} \in M_{pd}(\mathbb{R}), \tag{3.3.6a}$$

$$\mathcal{B} = \begin{pmatrix} \beta_1^T & \cdots & \beta_p^T \end{pmatrix}^T \in M_{pd,m}(\mathbb{R}), \tag{3.3.6b}$$

where

$$\beta_{p-j} = -I_{\{0,\ldots,q\}}(j)\left[\sum_{i=1}^{p-j-1} A_i \beta_{p-j-i} - B_{q-j}\right],$$

and

$$\mathcal{C} = \begin{pmatrix} \mathbf{1}_d & 0_d & \ldots & 0_d \end{pmatrix} \in M_{d,pd}(\mathbb{R}). \tag{3.3.6c}$$

It follows from representation (3.3.5) that MCARMA processes are special cases of linear multivariate continuous-time state space models, and in fact, the class of linear state space models is equivalent to the class of MCARMA models (Corollary 2.5). By considering the class of linear state space models, one can define representations of MCARMA processes which are different from Eq. (3.3.5) and better suited for the purpose of estimation.

Definition 3.24 (State Space Model) An $\mathbb{R}^d$-valued continuous-time linear state space model $(A, B, C, \boldsymbol{L})$ of dimension N is characterized by an $\mathbb{R}^m$-valued driving Lévy process $\boldsymbol{L}$, a state 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 a state equation of Ornstein–Uhlenbeck type

$$\mathrm{d}\boldsymbol{X}(t) = A\boldsymbol{X}(t)\mathrm{d}t + B\mathrm{d}\boldsymbol{L}(t), \quad t \in \mathbb{R}, \tag{3.3.7a}$$

and an observation equation

$$\boldsymbol{Y}(t) = C\boldsymbol{X}(t), \quad t \in \mathbb{R}. \tag{3.3.7b}$$

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

A solution $\boldsymbol{Y}$ to Eq. (3.3.7) is called *causal* if, for all t, $\boldsymbol{Y}(t)$ is independent of the σ-algebra generated by $\{\boldsymbol{L}(s) : s > t\}$. Every solution to Eq. (3.3.7a) satisfies

$$\boldsymbol{X}(t) = \mathrm{e}^{A(t-s)}\boldsymbol{X}(s) + \int_s^t \mathrm{e}^{A(t-u)}B\mathrm{d}\boldsymbol{L}(u), \quad \forall s, t \in \mathbb{R}, \quad s < t, \tag{3.3.8}$$

where the stochastic integral with respect to $\boldsymbol{L}$ is well-defined by Protter (1990, Theorem 3.9). The independent-increment property of Lévy processes implies that $\boldsymbol{X}$ is a Markov process. The following can be seen as the multivariate extension of Brockwell et al. (2011, Proposition 1) and recalls conditions for the existence of a stationary causal solution of the state equation (3.3.7a) for easy reference. We always work under the following assumption.

Assumption E1 The eigenvalues of A have strictly negative real parts.

Proposition 3.25 (Sato and Yamazato (1983, Theorem 5.1)) *If Assumptions E1 and L2 hold, then Eq. (3.3.7a) has a unique strictly stationary, causal solution $\boldsymbol{X}$ given by*

$$\boldsymbol{X}(t) = \int_{-\infty}^t \mathrm{e}^{A(t-u)}B\mathrm{d}\boldsymbol{L}(u), \quad t \in \mathbb{R}, \tag{3.3.9}$$

which has the same distribution as $\int_0^\infty \mathrm{e}^{Au}B\mathrm{d}\boldsymbol{L}(u)$. Moreover, $\boldsymbol{X}(t)$ has mean zero and second-order structure given by

$$\mathbb{V}\mathrm{ar}(\boldsymbol{X}(t)) =: \Gamma_0 = \int_0^\infty \mathrm{e}^{Au}B\Sigma^{\boldsymbol{L}}B^T\mathrm{e}^{A^Tu}\mathrm{d}u, \tag{3.3.10a}$$

$$\mathbb{C}\mathrm{ov}\left(\boldsymbol{X}(t+h), \boldsymbol{X}(t)\right) =: \gamma_{\boldsymbol{Y}}(h) = \mathrm{e}^{Ah}\Gamma_0, \quad h \geqslant 0, \tag{3.3.10b}$$

where the variance Γ_0 satisfies

$$A\Gamma_0 + \Gamma_0 A^T = -B\Sigma^{\boldsymbol{L}}B^T. \tag{3.3.10c}$$

It is an immediate consequence that the output process $\boldsymbol{Y}$ has mean zero and autocovariance function $\mathbb{R} \ni h \mapsto \gamma_{\boldsymbol{Y}}(h)$ given by $\gamma_{\boldsymbol{Y}}(h) = C\mathrm{e}^{Ah}\Gamma_0 C^T$, $h \geqslant 0$, and that $\boldsymbol{Y}$ itself can be written succinctly as a moving average of the driving Lévy process as

$$\boldsymbol{Y}(t) = \int_{-\infty}^{\infty} g(t-u)\mathrm{d}\boldsymbol{L}(u), \quad t \in \mathbb{R}; \quad g(t) = C\mathrm{e}^{At}BI_{[0,\infty)}(t). \tag{3.3.11}$$

As in Marquardt and Stelzer (2007, Proposition 3.30) one shows the following result about the existence of moments.

Proposition 3.26 *Let $\boldsymbol{Y}$ be the output process of the state space model* (3.3.7) *driven by the Lévy process $\boldsymbol{L}$. If $\boldsymbol{L}(1)$ is in $L^r(\Omega,\mathbb{P})$ for some $r>0$, then so are $\boldsymbol{Y}(t)$ and the state vector $\boldsymbol{X}(t)$, $t \in \mathbb{R}$.*

Equation (3.3.11) which, in conjunction with Eq. (3.3.6), serves as the definition of a multivariate CARMA process with autoregressive and moving average polynomials given by Eq. (3.3.3), shows that the behaviour of the process $\boldsymbol{Y}$ depends on the values of the individual matrices A, B, and C only through the products $C\mathrm{e}^{At}B$, $t \in \mathbb{R}$. The following lemma relates this analytical statement to an algebraic one about rational matrices, allowing us to draw a connection to the identifiability theory of discrete-time state space models.

Lemma 3.27 *Two matrix triplets (A,B,C), $(\tilde{A},\tilde{B},\tilde{C})$ of appropriate dimensions satisfy $C\mathrm{e}^{At}B = \tilde{C}\mathrm{e}^{\tilde{A}t}\tilde{B}$ for all $t \in \mathbb{R}$ if and only if $C(z\mathbf{1}-A)^{-1}B = \tilde{C}(z\mathbf{1}-\tilde{A})^{-1}\tilde{B}$ for all $z \in \mathbb{C}$.*

Proof If we start at the first equality and replace the matrix exponentials by their spectral representations (see Lax, 2002, Theorem 17.5), we obtain

$$\int_\gamma \mathrm{e}^{zt}C(z\mathbf{1}-A)^{-1}B\mathrm{d}z = \int_{\tilde{\gamma}} \mathrm{e}^{zt}\tilde{C}(z\mathbf{1}-\tilde{A})^{-1}\tilde{B}\mathrm{d}z, \quad \forall t \in \mathbb{R}, \tag{3.3.12}$$

where γ is a closed contour in $\mathbb{C}$ winding around each eigenvalue of A exactly once, and likewise for $\tilde{\gamma}$. Since we can always assume that $\gamma = \tilde{\gamma}$ by taking γ to be R times the unit circle, $R > \max\{|\lambda| : \lambda \in \sigma_A \cup \sigma_{\tilde{A}}\}$, we can write Eq. (3.3.12) as

$$\int_\gamma \mathrm{e}^{zt}\left[C(z\mathbf{1}-A)^{-1}B - \tilde{C}(z\mathbf{1}-\tilde{A})^{-1}\tilde{B}\right]\mathrm{d}z = 0, \quad \forall t \in \mathbb{R}. \tag{3.3.13}$$

Since the rational matrix function $\Delta(z) = C(z\mathbf{1}-A)^{-1}B - \tilde{C}(z\mathbf{1}-\tilde{A})^{-1}\tilde{B}$ has only poles with modulus less than R, it has an expansion around infinity, $\Delta(z) = \sum_{n=0}^\infty A_n z^{-n}$, $A_n \in M_d(\mathbb{C})$, which converges in a region $\{z \in \mathbb{C} : |z| > r\}$ containing γ. Using the fact that this series converges uniformly on the compact set γ and applying the Residue Theorem from complex analysis (Dieudonné, 1968, 9.16.1), which implies $\int_\gamma \mathrm{e}^{zt}z^{-n}\mathrm{d}z = t^n/n!$, Eq. (3.3.13) becomes $\sum_{n=0}^\infty \frac{t^n}{n!}A_{n+1} \equiv 0_N$. Consequently, by the Identity Theorem (Dieudonné, 1968, Theorem 9.4.3), A_n is the zero matrix for all $n > 1$, and since $\Delta(z) \to 0$ as $z \to \infty$, it follows that $\Delta(z) \equiv 0_{d,m}$. □

Because of its importance for the following discussion, the rational matrix function $H : z \mapsto C(z\mathbf{1}_N - A)^{-1}B$ is given a special name: it is called the *transfer function* of the state space model (3.3.7) and is intimately related to the spectral density $f_{\boldsymbol{Y}}$ of the output process $\boldsymbol{Y}$, which is defined as $f_{\boldsymbol{Y}}(\omega)=\int_{\mathbb{R}}\mathrm{e}^{-\mathrm{i}\omega h}\gamma_{\boldsymbol{Y}}(h)\mathrm{d}h$ – the Fourier transform of $\gamma_{\boldsymbol{Y}}$. Before we make this relation explicit, we prove the following lemma.

Lemma 3.28 *For any real number v, and matrices $A, B, \Sigma^{\boldsymbol{L}}, \Gamma_0$ as in Eq. (3.3.10a), it holds that*

$$\int_{-v}^{\infty}\mathrm{e}^{Au}B\Sigma^{\boldsymbol{L}}B^T\mathrm{e}^{A^Tu}\mathrm{d}u=\mathrm{e}^{-Av}\Gamma_0\mathrm{e}^{-A^Tv}. \tag{3.3.14}$$

Proof We define the functions $l,r:\mathbb{R}\to M_N(\mathbb{R})$ by

$$\begin{aligned}l(v)=&\int_{-v}^{\infty}\mathrm{e}^{Au}B\Sigma^{\boldsymbol{L}}B^T\mathrm{e}^{A^Tu}\mathrm{d}u,\\ r(v)=&\mathrm{e}^{-Av}\Gamma_0\mathrm{e}^{-A^Tv}.\end{aligned}$$

Clearly, both $l:v\mapsto l(v)$ and $r:v\mapsto r(v)$ are differentiable functions of v; taking the derivatives yields

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}v}l(v)=&\mathrm{e}^{-Av}B\Sigma^{\boldsymbol{L}}B^T\mathrm{e}^{-A^Tv},\\ \frac{\mathrm{d}}{\mathrm{d}v}r(v)=&-A\mathrm{e}^{-Av}\Gamma_0\mathrm{e}^{-A^Tv}-\mathrm{e}^{-Av}\Gamma_0A^T\mathrm{e}^{-A^Tv}.\end{aligned}$$

Using the identity $A\Gamma_0+\Gamma_0A^T=-B\Sigma^{\boldsymbol{L}}B^T$, Eq. (3.3.10c), one sees immediately that

$$\frac{\mathrm{d}}{\mathrm{d}v}l(v)=\frac{\mathrm{d}}{\mathrm{d}v}r(v),\quad\forall v\in\mathbb{R}.$$

Hence, l and r differ only by an additive constant. Since $l(0)$ equals $r(0)$ by the definition of Γ_0, the constant is zero, and $l(v)=r(v)$ for all real numbers v. □

Proposition 3.29 *Let $\boldsymbol{Y}$ be the output process of the state space model (3.3.7), and denote by $H: z\mapsto C(z\mathbf{1}_N-A)^{-1}B$ its transfer function. Then the relation*

$$f_{\boldsymbol{Y}}(\omega)=\frac{1}{2\pi}H(\mathrm{i}\omega)\Sigma^{\boldsymbol{L}}H(-\mathrm{i}\omega)^T \tag{3.3.15}$$

holds for all real ω; in particular, $\omega\mapsto f_{\boldsymbol{Y}}(\omega)$ is a rational matrix function.

Proof First, we recall (Bernstein, 2005, Proposition 11.2.2) that the Laplace transform of any matrix A is given by its resolvent, that is, for any complex number z,

$$(zI-A)^{-1}=\int_0^{\infty}\mathrm{e}^{-zu}\mathrm{e}^{Au}\mathrm{d}u. \tag{3.3.16}$$

We are now ready to compute

$$\begin{aligned}\frac{1}{2\pi}H(\mathrm{i}\omega)\Sigma^{\boldsymbol{L}}H(-\mathrm{i}\omega)^T =&\frac{1}{2\pi}C(\mathrm{i}\omega\mathbf{1}_N-A)^{-1}B\Sigma^{\boldsymbol{L}}B^T\left(-\mathrm{i}\omega\mathbf{1}_N-A^T\right)^{-1}C^T\\ =&\frac{1}{2\pi}C\left[\int_0^\infty \mathrm{e}^{-\mathrm{i}\omega u}\mathrm{e}^{Au}\mathrm{d}uB\Sigma^{\boldsymbol{L}}B^T\int_0^\infty \mathrm{e}^{\mathrm{i}\omega v}\mathrm{e}^{A^Tv}\mathrm{d}v\right]\mathrm{d}hC^T,\end{aligned}$$

where the last line follows from Eq. (3.3.16). Introducing the new variable $h=u-v$, and using Lemma 3.28, this becomes

$$\begin{aligned}&\frac{1}{2\pi}C\left[\int_0^\infty\int_{-v}^\infty \mathrm{e}^{-\mathrm{i}\omega h}\mathrm{e}^{Ah}\mathrm{e}^{Av}B\Sigma^{\boldsymbol{L}}B^T\mathrm{e}^{A^Tv}\mathrm{d}h\mathrm{d}v\right]C^T\\ =&\frac{1}{2\pi}C\left[\int_0^\infty\int_0^\infty \mathrm{e}^{-\mathrm{i}\omega h}\mathrm{e}^{Ah}\mathrm{e}^{Av}B\Sigma^{\boldsymbol{L}}B^T\mathrm{e}^{A^Tv}\mathrm{d}h\mathrm{d}v+\int_0^\infty\int_{-v}^0 \mathrm{e}^{-\mathrm{i}\omega h}\mathrm{e}^{Ah}\mathrm{e}^{Av}B\Sigma^{\boldsymbol{L}}B^T\mathrm{e}^{A^Tv}\mathrm{d}h\mathrm{d}v\right]C^T\\ =&\frac{1}{2\pi}C\left[\int_0^\infty \mathrm{e}^{-\mathrm{i}\omega h}\mathrm{e}^{Ah}\Gamma_0\mathrm{d}h+\int_{-\infty}^0 \mathrm{e}^{-\mathrm{i}\omega h}\Gamma_0\mathrm{e}^{-A^Th}\mathrm{d}h\right]C^T.\end{aligned}$$

By Eq. (3.3.10b) and the fact that the spectral density and the autocovariance function of a stochastic process are Fourier duals of each other, the last expression is equal to

$$\frac{1}{2\pi}\int_{-\infty}^\infty \mathrm{e}^{-\mathrm{i}\omega h}\gamma_{\boldsymbol{Y}}(h)\mathrm{d}h=f_{\boldsymbol{Y}}(\omega),$$

which completes the proof. □

One can also express the spectral density of a multivariate CARMA process in terms of its autoregressive and moving average polynomials.

Proposition 3.30 (Marquardt and Stelzer (2007, Proposition 3.28)) *The spectral density matrix function $f_{\boldsymbol{Y}}$ of an MCARMA process $\boldsymbol{Y}$ with autoregressive polynomial P and moving average polynomial Q is given by*

$$f_{\boldsymbol{Y}}(\omega)=\frac{1}{2\pi}P(\mathrm{i}\omega)^{-1}Q(\mathrm{i}\omega)\Sigma^{\boldsymbol{L}}Q(-\mathrm{i}\omega)^T\left(P(-\mathrm{i}\omega)^{-1}\right)^T. \tag{3.3.17}$$

A converse of Proposition 3.29, which will be useful in our later discussion of identifiability, is the Spectral Factorization Theorem. Its proof can be found in Rozanov (1967, Theorem 1.10.1) and also in Caines (1988, Theorem 4.1.4).

Theorem 3.31 *Every positive definite rational matrix function $f\in\mathbb{S}_d^+(\mathbb{C}\{\omega\})$ of full rank can be factorized as $f(\omega)=(2\pi)^{-1}W(\mathrm{i}\omega)W(-\mathrm{i}\omega)^T$, where the rational matrix function $z\mapsto W(z)\in M_{d,N}(\mathbb{R}\{z\})$, called a spectral factor, has full rank. For fixed N, the spectral factor W is uniquely determined up to an orthogonal transformation, i. e.*

$$W(z)\mapsto W(z)O, \tag{3.3.18}$$

for some orthogonal $N \times N$ matrix O.

3.3.2. Equidistant observations

We now turn to properties of the sampled process $\boldsymbol{Y}^{(h)} = (\boldsymbol{Y}_n^{(h)})_{n\in\mathbb{Z}}$ which is defined by $\boldsymbol{Y}_n^{(h)} = \boldsymbol{Y}(nh)$ and represents observations of the process $\boldsymbol{Y}$ at equally spaced points in time. A very fundamental observation is that the linear state space structure of the continuous-time process is preserved under sampling, as detailed in the following proposition. Of particular importance is the explicit formula (3.3.21) for the spectral density of the sampled process $\boldsymbol{Y}^{(h)}$.

Proposition 3.32 *Assume that $\boldsymbol{Y}$ is the output process of the state space model (3.3.7). Then the sampled process $\boldsymbol{Y}^{(h)}$ has the state space representation*

$$\boldsymbol{X}_n = \mathrm{e}^{Ah}\boldsymbol{X}_{n-1} + \boldsymbol{N}_n^{(h)}, \quad \boldsymbol{N}_n^{(h)} = \int_{(n-1)h}^{nh} \mathrm{e}^{A(nh-u)} B\mathrm{d}\boldsymbol{L}(u), \quad \boldsymbol{Y}_n^{(h)} = C\boldsymbol{X}_n^{(h)}. \tag{3.3.19}$$

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

$$\Sigma^{(h)} = \int_0^h \mathrm{e}^{Au} B\Sigma^{\boldsymbol{L}} B^T \mathrm{e}^{A^T u}\mathrm{d}u. \tag{3.3.20}$$

Moreover, the spectral density of $\boldsymbol{Y}^{(h)}$, denoted by $f_{\boldsymbol{Y}}^{(h)}$, is given by

$$f_{\boldsymbol{Y}}^{(h)}(\omega) = C\left(\mathrm{e}^{\mathrm{i}\omega}\mathbf{1}_N - \mathrm{e}^{Ah}\right)^{-1}\Sigma^{(h)}\left(\mathrm{e}^{-\mathrm{i}\omega}\mathbf{1}_N - \mathrm{e}^{A^T h}\right)^{-1} C^T; \tag{3.3.21}$$

in particular, $f_{\boldsymbol{Y}}^{(h)} : [-\pi,\pi] \to \mathbb{S}_d^+\left(\mathbb{R}\left\{\mathrm{e}^{\mathrm{i}\omega}\right\}\right)$ is a rational matrix function.

Proof Eqs. (3.3.19) follow from setting $t = nh$, $s = (n-1)h$ in Eq. (3.3.8). That the sequence $(\boldsymbol{Z}_n)_{n\in\mathbb{Z}}$ is i. i. d. as well as expression (3.3.20) for $\Sigma^{(h)}$ are immediate consequences of the Lévy process $\boldsymbol{L}$ having independent, homogeneous increments. Expression (3.3.21) follows from calculations analogous to the ones used for the continuous-time case (Proposition 3.29): using a Neumann series representation for the inverse of the matrix $\mathrm{e}^{\mathrm{i}\omega}\mathbf{1}_N - \mathrm{e}^{Ah}$ it follows that

$$\begin{aligned}
&C(\mathrm{e}^{\mathrm{i}\omega}\mathbf{1}_N - \mathrm{e}^{Ah})^{-1}\Sigma^{(h)}(\mathrm{e}^{-\mathrm{i}\omega}\mathbf{1}_N - \mathrm{e}^{A^T h})^{-1}C^T\\
=&C\left[\sum_{l,m\in\mathbb{N}_0} \mathrm{e}^{\mathrm{i}(m-l)\omega}\mathrm{e}^{Alh}\Sigma^{(h)}\mathrm{e}^{A^T mh}\right]C^T\\
=&C\left[\sum_{\substack{k\in-\mathbb{N}\\ l\in\mathbb{N}_0}} \mathrm{e}^{\mathrm{i}k\omega}\mathrm{e}^{A(l-k)h}\Sigma^{(h)}\mathrm{e}^{A^T lh} + \sum_{l\in\mathbb{N}_0}\mathrm{e}^{Alh}\Sigma^{(h)}\mathrm{e}^{A^T lh} + \sum_{\substack{k\in\mathbb{N}\\ l\in\mathbb{N}_0}}\mathrm{e}^{\mathrm{i}k\omega}\mathrm{e}^{Alh}\Sigma^{(h)}\mathrm{e}^{A^T(k+l)h}\right]C^T.
\end{aligned}$$

The observations that

$$\sum_{l\in\mathbb{N}_0} \mathrm{e}^{Alh}\Sigma^{(h)}\mathrm{e}^{A^Tlh} = \sum_{l\in\mathbb{N}_0} \mathrm{e}^{Alh}\left[\int_0^h \mathrm{e}^{Au}B\Sigma^{\boldsymbol{L}}B^T\mathrm{e}^{A^Tu}\mathrm{d}u\right]\mathrm{e}^{Alh} = \Gamma_0,$$

and that the autocovariance function of $\boldsymbol{Y}^{(h)}$ is given by $\gamma_{\boldsymbol{Y}^{(h)}}(k) = \gamma_{\boldsymbol{Y}}(kh)$, imply that the last expression can be further simplified to

$$C\left[\sum_{k\in\mathbb{N}} \mathrm{e}^{-\mathrm{i}k\omega}\mathrm{e}^{-Akh}\Gamma_0 + \Gamma_0 + \sum_{k\in-\mathbb{N}} \mathrm{e}^{-\mathrm{i}k\omega}\Gamma_0\mathrm{e}^{-A^Tkh}\right]C^T = \sum_{k\in\mathbb{Z}} \mathrm{e}^{-\mathrm{i}k\omega}\gamma_{\boldsymbol{Y}^{(h)}}(k) = f_{\boldsymbol{Y}}^{(h)}(\omega),$$

which proves the claim. □

In the following we analyse further the sampled state space model (3.3.19), in particular we will derive conditions for it to be minimal in the sense that the process $\boldsymbol{Y}^{(h)}$ is not the output process of any state space model of dimension less than N, and for the noise covariance matrix $\Sigma^{(h)}$ given in Eq. (3.3.20) to be non-singular. We begin by recalling some well-known notions from discrete-time realization and control theory. For a detailed account we refer to Åström (1970); Caines (1988); Sontag (1998), which also explain the origin of the terminology.

Definition 3.33 (Algebraic realization) Let $H \in M_{d,m}(\mathbb{R}\{z\})$ be a rational matrix function. A matrix triple (A, B, C) is called an *algebraic realization of H of dimension N* if $H(z) = C(z\mathbf{1}_N - A)^{-1}B$, where $A \in M_N(\mathbb{R})$, $B \in M_{N,m}(\mathbb{R})$, and $C \in M_{d,N}(\mathbb{R})$.

Every rational matrix function has many algebraic realizations of various dimensions. A particularly convenient class are the ones of minimal dimension, which have a number of useful properties.

Definition 3.34 (Minimality) Let $H \in M_{d,m}(\mathbb{R}\{z\})$ be a rational matrix function. A *minimal realization* of H is an algebraic realization of H of dimension smaller than or equal to the dimension of every other algebraic realization of H. The dimension of a minimal realization of H is the *McMillan degree* of H.

Two other important properties of algebraic realizations, which are intimately related to the notion of minimality and play a key role in the study of identifiability, are introduced in the following definitions.

Definition 3.35 (Controllability) An algebraic realization (A, B, C) of dimension N is *controllable* if the *controllability* matrix $\mathscr{C} = \begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix} \in M_{m,mN}(\mathbb{R})$ has full rank, i. e., if $\operatorname{rank}\mathscr{C} = N$.

Definition 3.36 (Observability) An algebraic realization (A, B, C) of dimension N is *observable* if the *observability* matrix $\mathscr{O} = \begin{bmatrix} C^T & (CA)^T & \cdots & (CA^{n-1})^T \end{bmatrix}^T \in M_{dN,N}(\mathbb{R})$ has full rank, i. e., if $\operatorname{rank}\mathscr{O} = N$.

Remark 3.37 We will often say that a state space system (3.3.7) is minimal, controllable or observable if the corresponding transfer function has this property.

The next theorem characterizes minimality in a useful way in terms of controllability and observability.

Theorem 3.38 (Hannan and Deistler (1988, Theorem 2.3.3)) *A realization (A, B, C) is minimal if and only if it is both controllable and observable.*

Lemma 3.39 *For all matrices $A \in M_N(\mathbb{R})$, $B \in M_{N,m}(\mathbb{R})$, $\Sigma \in \mathbb{S}_m^{++}(\mathbb{R})$, and every real number $t > 0$, the following linear subspaces of $\mathbb{R}^N$ are equal:*

$$i)\ \mathrm{im}\left[B, AB, \ldots, A^{N-1}B\right], \qquad ii)\ \mathrm{im}\int_0^t \mathrm{e}^{Au}B\Sigma B^T \mathrm{e}^{A^T u}\mathrm{d}u. \tag{3.3.22}$$

Proof The assertion is a generalization of Bernstein (2005, Lemma 12.6.2). The proof relies on two simple facts from linear algebra: first, that for any two subspaces V, W of $\mathbb{R}^N$, V is a subspace of W if and only if $W^\perp$ is a subset of $V^\perp$, where $^\perp$ denotes the orthogonal complement; second that for any matrix M the orthogonal complement of the range of M is equal to the kernel of M^T. In order to show that *i)* $\subset$ *ii)*, we pick $v \in \ker\left[\int_0^t \mathrm{e}^{Au}B\Sigma B^T \mathrm{e}^{A^T u}\mathrm{d}u\right]$. Then, clearly,

$$v^T\left[\int_0^t \mathrm{e}^{Au}B\Sigma B^T \mathrm{e}^{A^T u}\mathrm{d}u\right]v = 0;$$

this, together with the fact that for any $u \in [0, t]$, $\mathrm{e}^{Au}B\Sigma B^T \mathrm{e}^{A^T u}$ is positive semidefinite and the assumption that Σ is positive definite, implies that $v^T \mathrm{e}^{Au}B = \mathbf{0}_m^T$ for any $0 \leqslant u \leqslant t$. Differentiating with respect to u and evaluating at $u = 0$ yields $v^T A^i B = \mathbf{0}_m^T$ for all $i = 0, 1, \ldots, N-1$, which shows that v is an element of $\mathrm{im}\begin{bmatrix} B & AB & \cdots & A^{N-1}B \end{bmatrix}^\perp$. In order to prove the converse, that *ii)* $\subset$ *i)*, we choose $v \in \mathrm{im}\left[B, AB, \ldots, A^{N-1}B\right]^\perp$. Then, v is an element of $\ker\left[B, AB, \ldots, A^{N-1}B\right]^T$, and therefore $v^T A^i B = \mathbf{0}_m^T$ for all $i = 0, 1, \ldots, N-1$. Since, by the Cayley–Hamilton theorem, the matrix e^{Au} can be expanded as $\mathrm{e}^{Au} = \sum_{i=0}^{N-1}\psi_i(u)A^i$, it follows that

$$\left[\int_0^t \mathrm{e}^{Au}B\Sigma B^T \mathrm{e}^{A^T u}\mathrm{d}u\right]v = \int_0^t \mathrm{e}^{Au}B\Sigma\sum_{i=0}^{n-1}\psi_i(u)\left(v^T A^i B\right)^T \mathrm{d}u = \mathbf{0}_N,$$

which means that $v \in \ker\left[\int_0^t \mathrm{e}^{Au}B\Sigma B^T \mathrm{e}^{A^T u}\mathrm{d}u\right]$ and thus completes the proof. □

Corollary 3.40 *If the triple (A, B, C) is minimal of dimension N, and Σ is positive definite, then the $N \times N$ matrix $\check{\Sigma} = \int_0^h \mathrm{e}^{Au}B\Sigma B^T \mathrm{e}^{A^T u}\mathrm{d}u$ has full rank N.*

Proof By Theorem 3.38, minimality of (A, B, C) implies controllability, i. e. full rank of the controllability matrix $\left[\begin{array}{cccc} B & AB & \cdots & A^{N-1}B \end{array}\right]$. By Lemma 3.39, this is equivalent to Σ having full rank. $\square$

Proposition 3.41 *Assume that $\mathbf{Y}$ is the d-dimensional output process of the state space model (3.3.7) with (A, B, C) being a minimal realization of McMillan degree N. Then a sufficient condition for the sampled process $\mathbf{Y}^{(h)}$ to have the same McMillan degree, is the Kalman–Bertram criterion*

$$\lambda - \lambda' \neq \frac{2\pi i k}{h}, \qquad \forall (\lambda, \lambda') \in \sigma(A) \times \sigma(A), \qquad \forall k \in \mathbb{Z} \backslash \{0\}. \tag{3.3.23}$$

Proof We will prove the assertion by showing that the N-dimensional state space representation (3.3.19) is both controllable and observable, and thus, by Theorem 3.38, minimal. Observability has been shown in Sontag (1998, Proposition 5.2.11) using the Hautus criterion (Hautus, 1969). The key ingredient in the proof of controllability is Corollary 3.40, where we showed that the autocovariance matrix $\Sigma^{(h)}$ of $\mathbf{N}_n^{(h)}$, given by Eq. (3.3.20), has full rank; this shows that the representation (3.3.19) is indeed minimal and completes the proof. $\square$

Remark 3.42 Since, by Hannan and Deistler (1988, Theorem 2.3.4), minimal realizations are unique up to a change of basis $(A, B, C) \mapsto (TAT^{-1}, TB, CT^{-1})$, for some non-singular $N \times N$ matrix T, and such a transformation does not change the eigenvalues of A, the criterion (3.3.23) does not depend on what particular triple (A, B, C) one chooses.

Uniqueness of the principal logarithm (Higham, 2008, Theorem 1.31) implies the following.

Lemma 3.43 *Assume that the matrices $A, B \in M_N(\mathbb{R})$ satisfy $\mathrm{e}^{hA} = \mathrm{e}^{hB}$ for some $h > 0$. If the spectra σ_A, σ_B of A, B satisfy $|\operatorname{Im} \lambda| < \pi/h$ for all $\lambda \in \sigma_A \cup \sigma_B$, then $A = B$.*

Lemma 3.44 *Assume that $A \in M_N(\mathbb{R})$ satisfies Assumption E1. For every $h > 0$, the linear map*

$$\mathscr{M} : M_N(\mathbb{R}) \to M_N(\mathbb{R}), \quad M \mapsto \int_0^h \mathrm{e}^{Au} M \mathrm{e}^{A^T u} \mathrm{d}u, \tag{3.3.24}$$

is injective.

Proof If we apply the vectorization operator $\mathrm{vec} : M_N(\mathbb{R}) \to \mathbb{R}^{N^2}$ and use the well-known identity (Bernstein, 2005, Proposition 7.1.9) $\mathrm{vec}(UVW) = (W^T \otimes U)\,\mathrm{vec}(V)$ for matrices U, V and W of appropriate dimensions, we obtain the induced linear operator

$$\mathrm{vec} \circ \mathscr{M} \circ \mathrm{vec}^{-1} : \mathbb{R}^{N^2} \to \mathbb{R}^{N^2}, \quad \mathrm{vec}\, M \mapsto \int_0^h \mathrm{e}^{Au} \otimes \mathrm{e}^{Au} \mathrm{d}u\, \mathrm{vec}\, M.$$

To prove the claim that $\mathscr{M}$ is injective, it is thus sufficient to show that the matrix $\mathscr{A} := \int_0^h \mathrm{e}^{Au} \otimes \mathrm{e}^{Au} \mathrm{d}u \in M_{N^2}(\mathbb{R})$ is non-singular. We write $A \oplus A := A \otimes \mathbf{1}_N + \mathbf{1}_N \otimes A$. By

Bernstein (2005, Fact 11.14.37), $\mathscr{A} = \int_0^h \mathrm{e}^{(A\oplus A)u}\mathrm{d}u$ and since $\sigma(A \oplus A) = \{\lambda + \mu : \lambda, \mu \in \sigma(A)\}$ (Bernstein, 2005, Proposition 7.2.3), Assumption E1 implies that all eigenvalues of the matrix $A \oplus A$ have strictly negative real parts; in particular, $A \oplus A$ is invertible. Consequently, it follows from Bernstein (2005, Fact 11.13.14) that $\mathscr{A} = (A \oplus A)^{-1}\left[\mathrm{e}^{(A\oplus A)h} - \mathbf{1}_{N^2}\right]$. Since, for any matrix M, it holds that $\sigma(\mathrm{e}^M) = \{\mathrm{e}^\lambda, \lambda \in \sigma(M)\}$ (Bernstein, 2005, Proposition 11.2.3), the spectrum of $\mathrm{e}^{(A\oplus A)h}$ is a subset of the open unit disk. Using the fact that $\|M^n\|^{1/n}$ converges to the spectral radius of the matrix M as n tends to infinity, one sees that there exists a natural number n such that $\left\|\mathrm{e}^{n(A\oplus A)h}\right\|$ is strictly smaller than one. This implies that $\mathrm{e}^{(A\oplus A)h} - \mathbf{1}_{N^2}$ is non-singular, because

$$\left(\mathrm{e}^{(A\oplus A)h} - \mathbf{1}_{N^2}\right)^{-1} = \left(\mathbf{1}_{N^2} + \mathrm{e}^{(A\oplus A)h} + \ldots + \mathrm{e}^{(n-1)(A\oplus A)h}\right)\left(\mathrm{e}^{n(A\oplus A)h} - \mathbf{1}_{N^2}\right)^{-1}.$$

The existence of the last factor follows from its convergent Neumann series representation, which completes the proof. □

3.3.3. Overcoming the aliasing effect

One goal in this chapter is the estimation of multivariate CARMA processes or, equivalently, continuous-time state space models, based on discrete observations. We are now in the position to begin formulating precisely what assumptions we need to impose on the parametrizations of these models in order to ensure consistency and asymptotic normality of the quasi maximum likelihood estimator. In this brief section we concentrate on the issue of identifiability, and we derive sufficient conditions that prevent redundancies from being introduced into an otherwise properly specified model by the process of sampling, an effect known as aliasing (Hansen and Sargent, 1983; McCrorie, 2003).

For ease of notation we choose to parametrize the state matrix, the input matrix, and the observation matrix of the state space model (3.3.7), as well as the driving Lévy process $\boldsymbol{L}$; from these one can always obtain an autoregressive and a moving average polynomial which describe the same process by applying a left matrix fraction decomposition to the corresponding transfer function, see Patel (1981) and the upcoming Theorems 3.52 and 3.53. We hence assume that there is some compact parameter set $\Theta \subset \mathbb{R}^r$, and that, for each $\boldsymbol{\vartheta} \in \Theta$, one is given matrices $A_{\boldsymbol{\vartheta}}$, $B_{\boldsymbol{\vartheta}}$ and $C_{\boldsymbol{\vartheta}}$ of matching dimensions, as well as a Lévy process $\boldsymbol{L}_{\boldsymbol{\vartheta}}$. A basic assumption is that we always work with second order processes (cf. Assumption L2).

Assumption C1 For each $\boldsymbol{\vartheta} \in \Theta$, it holds that $\mathbb{E}\boldsymbol{L}_{\boldsymbol{\vartheta}} = \mathbf{0}_m$, that $\mathbb{E}\left\|\boldsymbol{L}_{\boldsymbol{\vartheta}}(1)\right\|^2$ is finite, and that the covariance matrix $\Sigma_{\boldsymbol{\vartheta}}^{\boldsymbol{L}} = \mathbb{E}\boldsymbol{L}_{\boldsymbol{\vartheta}}(1)\boldsymbol{L}_{\boldsymbol{\vartheta}}(1)^T$ is non-singular.

To ensure that the model corresponding to $\boldsymbol{\vartheta}$ describes a stationary output process we impose the analogue of Assumption E1.

Assumption C2 For each $\boldsymbol{\vartheta} \in \Theta$, the eigenvalues of $A_{\boldsymbol{\vartheta}}$ have strictly negative real parts.

Next, we restrict the model class so as to only contain minimal algebraic realizations of a fixed McMillan degree.

Assumption C3 For all $\boldsymbol{\vartheta} \in \Theta$, the triple $(A_{\boldsymbol{\vartheta}}, B_{\boldsymbol{\vartheta}}, C_{\boldsymbol{\vartheta}})$ is minimal with McMillan degree N.

Since we shall base the inference on a quasi maximum likelihood approach and thus on second-order properties of the observed process, we require the model class to be identifiable from these available information according to the following definitions.

Definition 3.45 (L^2-equivalence) 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 3.46 (Identifiability) A family $(\boldsymbol{Y}_{\boldsymbol{\vartheta}}, \boldsymbol{\vartheta} \in \Theta)$ of continuous-time stochastic processes is *identifiable from the spectral density* if, for every $\boldsymbol{\vartheta}_1 \neq \boldsymbol{\vartheta}_2$, the two processes $\boldsymbol{Y}_{\boldsymbol{\vartheta}_1}$ and $\boldsymbol{Y}_{\boldsymbol{\vartheta}_2}$ are not L^2-observationally equivalent. It is *h-identifiable from the spectral density*, $h > 0$, if, for every $\boldsymbol{\vartheta}_1 \neq \boldsymbol{\vartheta}_2$, the two sampled processes $\boldsymbol{Y}^{(h)}_{\boldsymbol{\vartheta}_1}$ and $\boldsymbol{Y}^{(h)}_{\boldsymbol{\vartheta}_2}$ are not L^2-observationally equivalent.

Assumption C4 The collection of output processes $K(\Theta) := (\boldsymbol{Y}_{\boldsymbol{\vartheta}}, \boldsymbol{\vartheta} \in \Theta)$ corresponding to the state space models $(A_{\boldsymbol{\vartheta}}, B_{\boldsymbol{\vartheta}}, C_{\boldsymbol{\vartheta}}, \boldsymbol{L}_{\boldsymbol{\vartheta}})$ is identifiable from the spectral density.

Since we shall use only observations of $\boldsymbol{Y}$ at discrete points in time separated by a sampling interval h, it would seem more natural to impose the stronger requirement that $K(\Theta)$ be h-identifiable. We will see, however, that this is implied by the previous assumptions if we additionally assume that the following holds.

Assumption C5 For all $\boldsymbol{\vartheta} \in \Theta$, the spectrum of $A_{\boldsymbol{\vartheta}}$ is a subset of

$$\left\{z \in \mathbb{C} : -\frac{\pi}{h} < \operatorname{Im} z < \frac{\pi}{h}\right\}.$$

Theorem 3.47 (Identifiability) *Assume that $\Theta \supset \boldsymbol{\vartheta} \mapsto \left(A_{\boldsymbol{\vartheta}}, B_{\boldsymbol{\vartheta}}, C_{\boldsymbol{\vartheta}}, \Sigma^{\boldsymbol{L}}_{\boldsymbol{\vartheta}}\right)$ is a parametrization of continuous-time state space models satisfying Assumptions C1 to C5. Then the corresponding collection of output processes $K(\Theta)$ is h-identifiable from the spectral density.*

Proof We will show that for every $\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2 \in \Theta$, $\boldsymbol{\vartheta}_1 \neq \boldsymbol{\vartheta}_2$, the sampled output processes $\boldsymbol{Y}^{(h)}_{\boldsymbol{\vartheta}_1}$ and $\boldsymbol{Y}^{(h)}_{\boldsymbol{\vartheta}_2}(h)$ are not L^2-observationally equivalent. Suppose, for the sake of contradiction, that the spectral densities of the sampled output processes were the same. Then the Spectral

Factorization Theorem (Theorem 3.31) would imply that there exists an orthogonal $N \times N$ matrix O such that

$$C_{\vartheta_1}(\mathrm{e}^{\mathrm{i}\omega}\mathbf{1}_N - \mathrm{e}^{A_{\vartheta_1}h})\Sigma_{\vartheta_1}^{(h),1/2}O = C_{\vartheta_2}(\mathrm{e}^{\mathrm{i}\omega}\mathbf{1}_N - \mathrm{e}^{A_{\vartheta_2}h})\Sigma_{\vartheta_2}^{(h),1/2}, \quad -\pi \leqslant \omega \leqslant \pi,$$

where $\Sigma_{\vartheta_i}^{(h),1/2}$, $i = 1,2$, are the unique positive definite matrix square roots of the matrices $\int_0^h \mathrm{e}^{A_{\vartheta_i}u}B_{\vartheta_i}\Sigma_{\vartheta_i}^{\boldsymbol{L}}B_{\vartheta_i}^T\mathrm{e}^{A_{\vartheta_i}^T u}\mathrm{d}u$, defined by spectral calculus. This means that the two triples

$$\left(\mathrm{e}^{A_{\vartheta_1}h}, \Sigma_{\vartheta_1}^{(h),1/2}O, C_{\vartheta_1}\right) \quad \text{and} \quad \left(\mathrm{e}^{A_{\vartheta_2}h}, \Sigma_{\vartheta_2}^{(h),1/2}, C_{\vartheta_2}\right)$$

are algebraic realizations of the same rational matrix function. Since Assumption C5 clearly implies the Kalman–Bertram criterion (3.3.23), it follows from Proposition 3.41 in conjunction with Assumption C3 that these realizations are minimal, and hence from Hannan and Deistler (1988, Theorem 2.3.4) that there exists an invertible matrix $T \in M_N(\mathbb{R})$ satisfying

$$\mathrm{e}^{A_{\vartheta_1}h} = T^{-1}\mathrm{e}^{A_{\vartheta_2}h}T, \qquad \Sigma_{\vartheta_1}^{(h),1/2}O = T^{-1}\Sigma_{\vartheta_2}^{(h),1/2}, \qquad C_{\vartheta_1} = C_{\vartheta_2}T. \tag{3.3.25}$$

It follows from the power series representation of the matrix exponential that $T^{-1}\mathrm{e}^{A_{\vartheta_2}h}T$ equals $\mathrm{e}^{T^{-1}A_{\vartheta_2}Th}$. Under Assumption C5, the first equation in conjunction with Lemma 3.43 therefore implies that $A_{\vartheta_1} = T^{-1}A_{\vartheta_2}T$. Using this, the second of the three equations (3.3.25) gives

$$\Sigma_{\vartheta_1}^{(h)} = \int_0^h \mathrm{e}^{A_{\vartheta_1}u}\left(T^{-1}B_{\vartheta_2}\right)\Sigma_{\vartheta_2}^{\boldsymbol{L}}\left(T^{-1}B_{\vartheta_2}\right)^T\mathrm{e}^{A_{\vartheta_1}^T u}\mathrm{d}u,$$

which, by Lemma 3.44, implies that $(T^{-1}B_{\vartheta_2})\Sigma_{\vartheta_2}^{\boldsymbol{L}}(T^{-1}B_{\vartheta_2})^T = B_{\vartheta_1}\Sigma_{\vartheta_1}^{\boldsymbol{L}}B_{\vartheta_1}^T$. Together with the last of the equations (3.3.25) and Proposition 3.32 it follows that, for every $\omega \in [-\pi, \pi]$,

$$\begin{aligned} f_{\vartheta_1}(\omega) =& C_{\vartheta_1}(\mathrm{i}\omega\mathbf{1}_N - A_{\vartheta_1})^{-1}B_{\vartheta_1}\Sigma_{\vartheta_1}^{\boldsymbol{L}}B_{\vartheta_1}^T(-\mathrm{i}\omega\mathbf{1}_N - A_{\vartheta_1}^T)^{-1}C_{\vartheta_1}^T \\ =& C_{\vartheta_2}(\mathrm{i}\omega\mathbf{1}_N - A_{\vartheta_2})^{-1}B_{\vartheta_2}\Sigma_{\vartheta_2}^{\boldsymbol{L}}B_{\vartheta_2}^T(-\mathrm{i}\omega\mathbf{1}_N - A_{\vartheta_2}^T)^{-1}C_{\vartheta_2}^T = f_{\vartheta_2}(\omega); \end{aligned}$$

this contradicts Assumption C4 that $\boldsymbol{Y}_{\vartheta_1}$ and $\boldsymbol{Y}_{\vartheta_2}$ are not L^2-observationally equivalent. □

3.3.4. Asymptotic properties of the QML estimator

In this section we apply the theory that we developed in Section 3.2 for the quasi maximum likelihood estimation of general discrete-time linear state space models to the estimation of continuous-time linear state space models or, equivalently, multivariate CARMA processes. We have already seen that a discretely observed MCARMA process can be represented by a discrete-time state space model and that, thus, a parametric family of MCARMA processes induces a parametric family of discrete-time state space models. More precisely, Eqs. (3.3.19) show that the process of sampling with spacing h maps the continuous-time state space models $(A_{\boldsymbol{\vartheta}}, B_{\boldsymbol{\vartheta}}, C_{\boldsymbol{\vartheta}}, \boldsymbol{L}_{\boldsymbol{\vartheta}})_{\boldsymbol{\vartheta}\in\Theta}$ to the discrete-time state space models

$$\left(\mathrm{e}^{A_{\boldsymbol{\vartheta}}h}, C_{\boldsymbol{\vartheta}}, \boldsymbol{N}_{\boldsymbol{\vartheta}}^{(h)}, \boldsymbol{0}\right)_{\boldsymbol{\vartheta}\in\Theta}, \quad \boldsymbol{N}_{\boldsymbol{\vartheta},n}^{(h)} = \int_{(n-1)h}^{nh} \mathrm{e}^{A_{\boldsymbol{\vartheta}}u} B_{\boldsymbol{\vartheta}} \mathrm{d}\boldsymbol{L}_{\boldsymbol{\vartheta}}(u). \tag{3.3.26}$$

which are not in the innovations form (3.1.2). The quasi maximum likelihood estimator $\hat{\boldsymbol{\vartheta}}^{L,(h)}$ is defined by Eq. (3.2.14), applied to the state space model (3.3.26), that is

$$\hat{\boldsymbol{\vartheta}}^{L,(h)} = \operatorname{argmin}_{\boldsymbol{\vartheta}\in\Theta} \widehat{\mathscr{L}}^{(h)}(\boldsymbol{\vartheta}, \boldsymbol{y}^{L,(h)}), \tag{3.3.27a}$$

$$\widehat{\mathscr{L}}^{(h)}(\boldsymbol{\vartheta}, \boldsymbol{y}^{L,(h)}) = \sum_{n=1}^{L} \left[d \log 2\pi + \log\det V_{\boldsymbol{\vartheta}}^{(h)} + \hat{\boldsymbol{\varepsilon}}_{\boldsymbol{\vartheta},n}^{(h),T} V_{\boldsymbol{\vartheta}}^{(h),-1} \hat{\boldsymbol{\varepsilon}}_{\boldsymbol{\vartheta},n}^{(h)}\right], \tag{3.3.27b}$$

where $\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{\vartheta}}^{(h)}$ are the pseudo-innovations of the observed process $\boldsymbol{Y}^{(h)} = \boldsymbol{Y}_{\boldsymbol{\vartheta}_0}^{(h)}$, which are computed from the sample $\boldsymbol{y}^{L,(h)} = (\boldsymbol{Y}_1^{(h)}, \ldots, \boldsymbol{Y}_L^{(h)})$ via the recursion

$$\hat{\boldsymbol{X}}_{\boldsymbol{\vartheta},n} = \left(\mathrm{e}^{A_{\boldsymbol{\vartheta}}h} - K_{\boldsymbol{\vartheta}}^{(h)} C_{\boldsymbol{\vartheta}}\right) \hat{\boldsymbol{X}}_{\boldsymbol{\vartheta},n-1} + K_{\boldsymbol{\vartheta}}^{(h)} \boldsymbol{Y}_{n-1}^{(h)}, \quad \hat{\boldsymbol{\varepsilon}}_{\boldsymbol{\vartheta},n}^{(h)} = \boldsymbol{Y}_n^{(h)} - C_{\boldsymbol{\vartheta}} \hat{\boldsymbol{X}}_{\boldsymbol{\vartheta},n}, \quad n \in \mathbb{N}.$$

The initial value $\hat{\boldsymbol{X}}_{\boldsymbol{\vartheta},1}$ may be chosen in the same ways as in the discrete-time case. The steady-state Kalman gain matrices $K_{\boldsymbol{\vartheta}}^{(h)}$ and pseudo-covariances $V_{\boldsymbol{\vartheta}}^{(h)}$ are computed as functions of the unique positive definite solution $\Omega_{\boldsymbol{\vartheta}}^{(h)}$ to the discrete-time algebraic Riccati equation

$$\Omega_{\boldsymbol{\vartheta}}^{(h)} = \mathrm{e}^{A_{\boldsymbol{\vartheta}}h} \Omega_{\boldsymbol{\vartheta}}^{(h)} \mathrm{e}^{A_{\boldsymbol{\vartheta}}^T h} + \Sigma_{\boldsymbol{\vartheta}}^{(h)} - \left[\mathrm{e}^{A_{\boldsymbol{\vartheta}}h} \Omega_{\boldsymbol{\vartheta}}^{(h)} C_{\boldsymbol{\vartheta}}^T\right] \left[C_{\boldsymbol{\vartheta}} \Omega_{\boldsymbol{\vartheta}}^{(h)} C_{\boldsymbol{\vartheta}}^T\right]^{-1} \left[\mathrm{e}^{A_{\boldsymbol{\vartheta}}h} \Omega_{\boldsymbol{\vartheta}}^{(h)} C_{\boldsymbol{\vartheta}}^T\right]^T,$$

namely

$$K_{\boldsymbol{\vartheta}}^{(h)} = \left[\mathrm{e}^{A_{\boldsymbol{\vartheta}}h} \Omega_{\boldsymbol{\vartheta}}^{(h)} C_{\boldsymbol{\vartheta}}^T\right] \left[C_{\boldsymbol{\vartheta}} \Omega_{\boldsymbol{\vartheta}}^{(h)} C_{\boldsymbol{\vartheta}}^T\right]^{-1}, \quad V_{\boldsymbol{\vartheta}}^{(h)} = C_{\boldsymbol{\vartheta}} \Omega_{\boldsymbol{\vartheta}}^{(h)} C_{\boldsymbol{\vartheta}}^T.$$

In order to obtain the asymptotic normality of the quasi maximum likelihood estimator for multivariate CARMA processes, it is therefore only necessary to make sure that Assumptions D1 to D10 hold for the model (3.3.26). The discussion of identifiability in the previous section allows us to specify accessible conditions on the parametrization of the continuous-time model under which the quasi maximum likelihood estimator is strongly consistent. In addition to the identifiability assumptions C3 to C5, we impose the following conditions.

Assumption C6 The parameter space Θ is a compact subset of $\mathbb{R}^r$.

Assumption C7 The functions $\boldsymbol{\vartheta} \mapsto A_{\boldsymbol{\vartheta}}$, $\boldsymbol{\vartheta} \mapsto B_{\boldsymbol{\vartheta}}$, $\boldsymbol{\vartheta} \mapsto C_{\boldsymbol{\vartheta}}$, and $\boldsymbol{\vartheta} \mapsto \Sigma_{\boldsymbol{\vartheta}}^{\boldsymbol{L}}$ are continuous. Moreover, for each $\boldsymbol{\vartheta} \in \Theta$, the matrix $C_{\boldsymbol{\vartheta}}$ has full rank.

Lemma 3.48 *Assumptions C1 to C3, C6 and C7 imply that the family* $\left(\mathrm{e}^{A_{\boldsymbol{\vartheta}}h}, C_{\boldsymbol{\vartheta}}, \boldsymbol{N}_{\boldsymbol{\vartheta}}^{(h)}, \boldsymbol{0}\right)_{\boldsymbol{\vartheta}\in\Theta}$ *of discrete-time state space models satisfies Assumptions D1 to D4.*

Proof Assumption D1 is clear. Assumption D2 follows from the observation that the functions $A \mapsto \mathrm{e}^{A}$ and $(A, B, \Sigma) \mapsto \int_0^h \mathrm{e}^{Au} B\Sigma B^T \mathrm{e}^{A^T u}\mathrm{d}u$ are continuous. By Assumptions C2, C6 and C7, and the fact that the eigenvalues of a matrix are continuous functions of its entries, it follows that there exists a positive real number ϵ such that, for each $\boldsymbol{\vartheta} \in \Theta$, the eigenvalues of $A_{\boldsymbol{\vartheta}}$ have real parts less than or equal to $-\epsilon$. The observation that the eigenvalues of e^{A} are given by the exponentials of the eigenvalues of A thus shows that Assumption D3, i) holds with $\rho := \mathrm{e}^{-\epsilon h} < 1$. Assumption C1 that the matrices $\Sigma_{\boldsymbol{\vartheta}}^{\boldsymbol{L}}$ are non-singular and the minimality assumption C3 imply by Corollary 3.40 that the noise covariance matrices $\Sigma_{\boldsymbol{\vartheta}}^{(h)} = \mathbb{E}\boldsymbol{N}_{\boldsymbol{\vartheta},n}^{(h)}\boldsymbol{N}_{\boldsymbol{\vartheta},n}^{(h),T}$ are non-singular, and thus Assumption D3, ii) holds. Further, by Proposition 3.4, the matrices $\Omega_{\boldsymbol{\vartheta}}$ are non-singular, and so are, because the matrices $C_{\boldsymbol{\vartheta}}$ are assumed to be of full rank, the matrices $V_{\boldsymbol{\vartheta}}$; this means that Assumption D3, iii) is satisfied. Assumption D4 is a consequence of Proposition 3.32, which states that the noise sequences $\boldsymbol{N}_{\boldsymbol{\vartheta}}$ are i. i. d. and, in particular, ergodic; their second moments are finite because of Assumption C1. □

In order to be able to show that the quasi maximum likelihood estimator $\hat{\boldsymbol{\vartheta}}^{L,(h)}$ is asymptotically normally distributed, we impose the following conditions in addition to the ones described so far.

Assumption C8 The true parameter value $\boldsymbol{\vartheta}_0$ is an element of the interior of Θ.

Assumption C9 The functions $\boldsymbol{\vartheta} \mapsto A_{\boldsymbol{\vartheta}}$, $\boldsymbol{\vartheta} \mapsto B_{\boldsymbol{\vartheta}}$, $\boldsymbol{\vartheta} \mapsto C_{\boldsymbol{\vartheta}}$, and $\boldsymbol{\vartheta} \mapsto \Sigma_{\boldsymbol{\vartheta}}^{\boldsymbol{L}}$ are three times continuously differentiable.

Assumption C10 There exists a positive number δ such that $\mathbb{E}\left\|\boldsymbol{L}_{\boldsymbol{\vartheta}_0}(1)\right\|^{4+\delta} < \infty$.

Lemma 3.49 *Assumptions C8 to C10 imply that Assumptions D6 to D8 hold for the model* (3.3.26)*.*

Proof Assumption D6 is clear. Assumption D7 follows from the fact that the functions $A \mapsto \mathrm{e}^{A}$ and $(A, B, \Sigma) \mapsto \int_0^h \mathrm{e}^{Au} B\Sigma B^T \mathrm{e}^{A^T u}\mathrm{d}u$ are not only continuous, but infinitely often differentiable. For Assumption D8 we need to show that the random variables $\boldsymbol{N} := \boldsymbol{N}_{\boldsymbol{\vartheta}_0,1}$ have bounded $(4+\delta)$th absolute moments. It follows from Rajput and Rosiński (1989, Theorem 2.7) that $\boldsymbol{N}$ is infinitely divisible with characteristic triplet (γ, Σ, ν) given by

$$\gamma = \int_0^h \mathrm{e}^{A_{\boldsymbol{\vartheta}_0}(h-s)} B_{\boldsymbol{\vartheta}_0}\left[\gamma_{\boldsymbol{\vartheta}_0}^{\boldsymbol{L}} + \int_{\mathbb{R}^d} \boldsymbol{x}\left(I_{[0,1]}\left(\left\|\mathrm{e}^{A_{\boldsymbol{\vartheta}_0}(h-s)} B_{\boldsymbol{\vartheta}_0}\boldsymbol{x}\right\|\right) - I_{[0,1]}(\|\boldsymbol{x}\|)\right)\nu^{\boldsymbol{L}_{\boldsymbol{\vartheta}_0}}(\mathrm{d}\boldsymbol{x})\right]\mathrm{d}s,$$

$$\Sigma = \int_0^h \mathrm{e}^{A_{\boldsymbol{\vartheta}_0}(h-s)} B_{\boldsymbol{\vartheta}_0}\Sigma^{\mathcal{G}} B_{\boldsymbol{\vartheta}_0}^T \mathrm{e}^{A_{\boldsymbol{\vartheta}_0}^T(h-s)}\mathrm{d}s,$$

$$\nu(B) = \int_0^h \int_{\mathbb{R}^m} I_B\left(\mathrm{e}^{A_{\boldsymbol{\vartheta}_0}(h-s)} B_{\boldsymbol{\vartheta}_0}\boldsymbol{x}\right)\nu^{\boldsymbol{L}_{\boldsymbol{\vartheta}_0}}(\mathrm{d}\boldsymbol{x})\mathrm{d}s, \quad B \in \mathscr{B}(\mathbb{R}^N \backslash \{\boldsymbol{0}_N\}),$$

where $\mathscr{B}(\cdot)$ denotes the Borel σ-algebra. These formulæ imply that

$$\int_{\|x\|\geqslant 1}\|x\|^{4+\delta}\nu(\mathrm{d}x)\leqslant\int_0^1\left\|\mathrm{e}^{A_{\vartheta_0}(h-s)}B_{\vartheta}\right\|^{4+\delta}\mathrm{d}s\int_{\|x\|\geqslant 1}\|x\|^{4+\delta}\nu^{L_{\vartheta_0}}\vartheta(\mathrm{d}x).$$

The first factor on the right side is finite by Assumptions C6 and C9, the second by Assumption C10 and the well known equivalence of finiteness of the αth absolute moment of an infinitely divisible distribution and finiteness of the αth absolute moments of the corresponding Lévy measure restricted to the exterior of the unit ball (Sato, 1999, Corollary 25.8). The same corollary shows that $\mathbb{E}\,\|N\|^{4+\delta}<\infty$ and thus Assumption D8. □

Our final assumption is the analogue of Assumption D10. It will ensure that the Fisher information matrix of the quasi maximum likelihood estimator $\hat{\vartheta}^{L,(h)}$ is non-singular by imposing a non-degeneracy condition on the parametrization of the model.

Assumption C11 There exists a positive index j_0 such that the $\left[(j_0+2)d^2\right]\times r$ matrix

$$\nabla_{\vartheta}\left(\begin{array}{c}\left[\mathbf{1}_{j_0+1}\otimes K_{\vartheta}^{(h),T}\otimes C_{\vartheta}\right]\left[\begin{array}{cccc}\left(\operatorname{vec}\mathrm{e}^{\mathbf{1}_N h}\right)^T & \left(\operatorname{vec}\mathrm{e}^{A_{\vartheta}h}\right)^T & \cdots & \left(\operatorname{vec}\mathrm{e}^{A_{\vartheta}^{j_0}h}\right)^T\end{array}\right]^T\\ \operatorname{vec}V_{\vartheta}\end{array}\right)_{\vartheta=\vartheta_0}$$

has rank r.

Theorem 3.50 (Asymptotic normality of $\hat{\vartheta}^{L,(h)}$) *Assume that $(A_{\vartheta},B_{\vartheta},C_{\vartheta},L_{\vartheta})_{\vartheta\in\Theta}$ is a parametric family of continuous-time state space models, and denote by $y^{L,(h)}=(Y_{\vartheta_0,1}^{(h)},\ldots,Y_{\vartheta_0,L}^{(h)})$ a sample of length L from the discretely observed output process corresponding to the parameter value $\vartheta_0\in\Theta$. Under Assumptions C1 to C7 the quasi maximum likelihood estimator $\hat{\vartheta}^{L,(h)}=$* $\operatorname{argmin}_{\vartheta\in\Theta}\widehat{\mathscr{L}}(\vartheta,y^{L,(h)})$ *is strongly consistent, that is*

$$\hat{\vartheta}^{L,(h)}\xrightarrow[L\to\infty]{a.s.}\vartheta_0.\tag{3.3.29}$$

If, moreover, Assumptions C8 to C11 hold, then $\hat{\vartheta}^{L,(h)}$ is asymptotically normally distributed, that is

$$\sqrt{L}\left(\hat{\vartheta}^{L,(h)}-\vartheta_0\right)\xrightarrow[L\to\infty]{d}\mathscr{N}(\mathbf{0},\Xi),\tag{3.3.30}$$

where the asymptotic covariance matrix $\Xi=J^{-1}IJ^{-1}$ is given by

$$I=\lim_{L\to\infty}L^{-1}\operatorname{Var}\left(\nabla_{\vartheta}\mathscr{L}\left(\vartheta_0,y^L\right)\right),\quad J=\lim_{L\to\infty}L^{-1}\nabla_{\vartheta}^2\mathscr{L}\left(\vartheta_0,y^L\right).\tag{3.3.31}$$

Proof Strong consistency is a consequence of Theorem 3.7 if we can show that the parametric family $\left(\mathrm{e}^{A_{\vartheta}h},C_{\vartheta},N_{\vartheta},\mathbf{0}\right)_{\vartheta\in\Theta}$ of discrete-time state space models satisfies Assumptions D1 to D5. The first four of these are shown to hold in Lemma 3.48. For the last one, we observe that, by Lemma 3.6, Assumption D5 is equivalent to the family of state space models (3.3.26) being identifiable from the spectral density. Under Assumptions C3 to C5 this is guaranteed by Theorem 3.47.

In order to prove Eq. (3.3.30), we shall apply Theorem 3.8 and therefore need to verify Assumptions D6 to D10 for the state space models $\left(e^{A_\vartheta h}, C_\vartheta, N_\vartheta, \mathbf{0}\right)_{\vartheta\in\Theta}$. The first three hold by Lemma 3.49, the last one as a reformulation of Assumption C11. Assumption D9, that the strong-mixing coefficients α of a sampled multivariate CARMA process satisfy $\sum_m [\alpha(m)]^{\delta/(2+\delta)} < \infty$, follows from Assumption C1 and Marquardt and Stelzer (2007, Proposition 3.34), where it was shown that MCARMA processes with a finite logarithmic moment are exponentially strongly mixing. □

3.4. Practical applicability

In this section we complement the theoretical results from Sections 3.2 and 3.3 by commenting on their applicability in practical situations. Canonical parametrizations are a classical subject of research about discrete-time dynamical systems, and most of the results carry over to the continuous-time case; without going into great detail we present the basic notions and results about these parametrizations. The assertions of Theorem 3.50 are confirmed by means of a simulation study for a bivariate non-Gaussian CARMA process. Finally, we estimate the parameters of a CARMA model for a bivariate time series from economics using our quasi maximum likelihood approach.

3.4.1. Canonical parametrizations

We present parametrizations of multivariate CARMA processes that satisfy the identifiability conditions C3 and C4, as well as the smoothness conditions C7 and C9; if, in addition, the parameter space Θ is restricted so that Assumptions C2, C5, C6 and C8 hold, and the driving Lévy process satisfies Assumption C1, the canonically parametrized MCARMA model can be estimated consistently. In order for this estimate to be asymptotically normally distributed, one must additionally impose Assumption C10 on the Lévy process and check that Assumption C11 holds – a condition which we are unable to verify analytically for the general model; for explicit parametrizations, however, it can be checked numerically with moderate computational effort. The parametrizations we are to present are well-known from the discrete-time setting; detailed descriptions with proofs can be found in Deistler (1983); Hannan (1971, 1976, 1979); Hannan and Deistler (1988); Luenberger (1967); Lütkepohl and Poskitt (1996); Poskitt (1992); Reinsel (1997); Rosenbrock (1970) or, from a slightly different perspective, in the control theory literature Gevers (1986); Gevers and Wertz (1983, 1984); Guidorzi (1975, 1981). We begin with a canonical decomposition for rational matrix functions.

Theorem 3.51 (Bernstein (2005, Theorem 4.7.5)) *Let $H \in M_{d,m}(\mathbb{R}\{z\})$ be a rational matrix function of rank r. There exist matrices $S_1 \in M_d(\mathbb{R}[z])$ and $S_2 \in M_m(\mathbb{R}[z])$ with constant determinant, such that $H = S_1 M S_2$, where*

$$M = \begin{bmatrix} \operatorname{diag}\{\epsilon_i/\psi_i\}_{i=1}^{r} & 0_{r,m-r} \\ 0_{d-r,r} & 0_{d-r,m-r} \end{bmatrix} \in M_{d,m}(\mathbb{R}\{z\}), \tag{3.4.1}$$

and $\epsilon_1, \ldots \epsilon_r, \psi_1, \ldots, \psi_r \in \mathbb{R}[z]$ are polynomials with leading coefficient one, uniquely determined by H satisfying the following conditions:

i) for each $i = 1, \ldots, r$, the polynomials ϵ_i and ψ_i have no common roots,

ii) for each $i = 1, \ldots, r-1$, the polynomial ϵ_i divides the polynomial ϵ_{i+1}, and

iii) for each $i = 1, \ldots, r-1$, the polynomial ψ_{i+1} divides the polynomial ψ_i.

The triple (S_1, M, S_2) is called the Smith–McMillan decomposition of H.

The degrees ν_i of the denominator polynomials ψ_i in the Smith–McMillan decomposition of a rational matrix function H are called the Kronecker indices of H, and they define the vector $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_d) \in \mathbb{N}^d$, where we set $\nu_k = 0$ for $k = r+1, \ldots, d$. They satisfy the important relation $\sum_{i=1}^{d} \nu_i = \delta_M(H)$, where $\delta_M(H)$ denotes the McMillan degree of H, that is the smallest possible dimension of an algebraic realization of H, see Definition 3.34. For $1 \leqslant i, j \leqslant d$, we also define the integers $\nu_{ij} = \min\{\nu_i + I_{\{i>j\}}, \nu_j\}$, and if the Kronecker indices of the transfer function of an MCARMA process $\boldsymbol{Y}$ are $\boldsymbol{\nu}$, we call $\boldsymbol{Y}$ an MCARMA$_{\boldsymbol{\nu}}$ process.

Theorem 3.52 (Echelon state space realization, Guidorzi (1975, Section 3)) *For positive integers d and m, let $H \in M_{d,m}(\mathbb{R}\{z\})$ be a rational matrix function with Kronecker indices $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_d)$. Then a unique minimal algebraic realization (A, B, C) of H of dimension $N = \delta_M(H)$ is given by the following structure.*

(i) The matrix $A = (A_{ij})_{i,j=1,\ldots,d} \in M_N(\mathbb{R})$ is a block matrix with blocks $A_{ij} \in M_{\nu_i,\nu_j}(\mathbb{R})$ given by

$$A_{ij} = \begin{pmatrix} 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & & & & & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \alpha_{ij,1} & \cdots & \alpha_{ij,\nu_{ij}} & 0 & \cdots & 0 \end{pmatrix} + \delta_{i,j} \begin{pmatrix} 0 & & \\ \vdots & \mathbf{1}_{\nu_i - 1} & \\ 0 & & \\ 0 & \cdots & 0 \end{pmatrix}, \tag{3.4.2a}$$

(ii) $B = (b_{ij}) \in M_{N,m}(\mathbb{R})$ unrestricted,

(iii) if $\nu_i > 0$, $i = 1, \ldots, d$, then

$$C = \left(\begin{array}{cccc:cccc:c:cccc} 1 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & & & & & \\ & & & & 1 & 0 & \ldots & 0 & & & 0_{(d-1),\nu_d} & & \\ & 0_{(d-1),\nu_1} & & & & 0_{(d-2),\nu_2} & & & & 1 & 0 & \ldots & 0 \end{array}\right) \in M_{d,N}(\mathbb{R}). \quad (3.4.2b)$$

If $\nu_i = 0$, the elements of the ith row of C are also freely varying, but we concentrate here on the case where all Kronecker indices ν_i are positive. To compute $\boldsymbol{\nu}$ as well as the coefficients $\alpha_{ij,k}$ and b_{ij} for a given rational matrix function H, several numerically stable and efficient algorithms are available in the literature (see, e. g., Rózsa and Sinha, 1975, and the references therein). The orthogonal invariance inherent in spectral factorization (see Theorem 3.31) implies that this parametrization alone does not ensure identifiability. In the case $m = d$, one remedy is to restrict the parametrization to those transfer functions H satisfying $H(0) = H_0$, for a non-singular matrix H_0. To see how one must constrain the parameters $\alpha_{ij,k}, b_{ij}$ in order to ensure this normalization, we work in terms of left matrix fraction descriptions.

Theorem 3.53 (Echelon MCARMA realization, Guidorzi (1975, Section 3)) *For positive integers d and m, let $H \in M_{d,m}(\mathbb{R}\{z\})$ be a rational matrix function with Kronecker indices $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_d)$. Assume that (A, B, C) is a realization of H, parametrized as in Eqs. (3.4.2). Then a unique left matrix fraction description $P^{-1}Q$ of H is given by*

$$P(z) = \left[p_{ij}(z)\right]_{i,j=1,\ldots,d}, \quad (3.4.3a)$$

$$Q(z) = \left[q_{ij}(z)\right]_{\substack{i=1,\ldots,d \\ j=1,\ldots,m}}, \quad (3.4.3b)$$

where

$$p_{ij}(z) = \delta_{i,j} z^{\nu_i} - \sum_{k=1}^{\nu_{ij}} \alpha_{ij,k} z^{k-1}, \quad (3.4.4a)$$

$$q_{ij}(z) = \sum_{k=1}^{\nu_i} \kappa_{\nu_1+\ldots+\nu_{i-1}+k,j} z^{k-1}, \quad (3.4.4b)$$

and the coefficient $\kappa_{i,j}$ is the (i,j)th entry of the matrix $K = TB$, where $T = (T_{ij})_{i,j=1,\ldots,d} \in M_N(\mathbb{R})$

is a block matrix with blocks $T_{ij} \in M_{\nu_i,\nu_j}(\mathbb{R})$ *given by*

$$T_{ij} = \begin{pmatrix} -\alpha_{ij,2} & \dots & -\alpha_{ij,\nu_{ij}} & 0 & \dots & 0 \\ \vdots & \iddots & & & & \vdots \\ -\alpha_{ij,\nu_{ij}} & & & & & \vdots \\ 0 & & & & & \vdots \\ \vdots & & & & & \vdots \\ 0 & \dots & \dots & \dots & \dots & 0 \end{pmatrix} + \delta_{i,j} \begin{pmatrix} 0 & 0 & \dots & \dots & 0 & 1 \\ 0 & 0 & \dots & & 1 & 0 \\ \vdots & \vdots & & \iddots & & \vdots \\ \vdots & & \iddots & & \vdots & \vdots \\ 0 & 1 & & \dots & 0 & 0 \\ 1 & 0 & \dots & \dots & 0 & 0 \end{pmatrix}. \tag{3.4.5}$$

The orders p, q of the polynomials P, Q satisfy $p = \max\{\nu_1, \dots, \nu_d\}$ and $q \leqslant p-1$. Using this parametrization, there are different ways to impose the normalization $H(0) = H_0 \in M_{d,m}(\mathbb{R})$. One first observes that the special structure of the polynomials P and Q implies that $H(0) = P(0)^{-1}Q(0) = -(\alpha_{ij,1})_{ij}^{-1}(\kappa_{\nu_1+\dots+\nu_{i-1}+1,j})_{ij}$. The canonical state space parametrization (A, B, C) given by Eqs. (3.4.2) therefore satisfies $H(0) = -CA^{-1}B = H_0$ if one makes the coefficients $\alpha_{ij,1}$ functionally dependent on the free parameters $\alpha_{ij,m}$, $m = 1, \dots \nu_{ij}$ and b_{ij} by setting $\alpha_{ij,1} = -[(\kappa_{\nu_1+\dots+\nu_{k-1}+1,l})_{kl} H_0^{\sim 1}]_{ij}$, where κ_{ij} are the entries of the matrix K appearing in Theorem 3.53 and $H_0^{\sim 1}$ is a right inverse of H_0. Another possibility, which has the advantage of preserving the multi-companion structure of the matrix A, is to keep the $\alpha_{ij,1}$ as free parameters, and to restrict some of the entries of the matrix B instead. Since $|\det K| = 1$ and the matrix T is thus invertible, the coefficients b_{ij} can be written as $B = T^{-1}K$. Replacing the $(\nu_1 + \dots + \nu_{i-1} + 1, j)$th entry of K by the (i, j)th entry of the matrix $-(\alpha_{kl,1})_{kl} H_0$ makes some of the b_{ij} functionally dependent on the entries of the matrix A, and results in a state space representation with prescribed Kronecker indices and satisfying $H(0) = H_0$. This latter method has also the advantage that it does not require the matrix H_0 to possess a right inverse. In the special case that $d = m$ and $H_0 = -\mathbf{1}_d$, it suffices to set $\kappa_{\nu_1+\dots+\nu_{i-1}+1,j} = \alpha_{ij,1}$, for $i, j = 1, \dots, d$. Examples of normalized low-order canonical parametrizations are given in Tables 3.1 and 3.2.

3.4.2. Simulation study

In order to get a better feeling for how the quasi maximum likelihood estimation procedure performs in reality, we present a simulation study for a bivariate CARMA process with Kronecker indices $(1, 2)$, i. e. CARMA indices $(p, q) = (2, 1)$. As the driving Lévy process we chose a zero-mean normal-inverse Gaussian (NIG) process $(\boldsymbol{L}(t))_{t\in\mathbb{R}}$. Such processes have been found to be useful in the modelling of stock returns and stochastic volatility, as well as turbulence data (see, e. g., Barndorff-Nielsen, 1997, 1998; Barndorff-Nielsen, Blæsild and Schmiegel, 2004; Rydberg, 1997). The distribution of the increments $\boldsymbol{L}(t) - \boldsymbol{L}(t-1)$ of a

ν	$n(\nu)$	A	B	C
$(1,1)$	7	$\begin{pmatrix} \vartheta_1 & \vartheta_2 \\ \vartheta_3 & \vartheta_4 \end{pmatrix}$	$\begin{pmatrix} \vartheta_1 & \vartheta_2 \\ \vartheta_3 & \vartheta_4 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
$(1,2)$	10	$\begin{pmatrix} \vartheta_1 & \vartheta_2 & 0 \\ 0 & 0 & 1 \\ \vartheta_3 & \vartheta_4 & \vartheta_5 \end{pmatrix}$	$\begin{pmatrix} \vartheta_1 & \vartheta_2 \\ \vartheta_6 & \vartheta_7 \\ \vartheta_3 + \vartheta_5\vartheta_6 & \vartheta_4 + \vartheta_5\vartheta_7 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$
$(2,1)$	11	$\begin{pmatrix} 0 & 1 & 0 \\ \vartheta_1 & \vartheta_2 & \vartheta_3 \\ \vartheta_4 & \vartheta_5 & \vartheta_6 \end{pmatrix}$	$\begin{pmatrix} \vartheta_7 & \vartheta_8 \\ \vartheta_1 + \vartheta_2\vartheta_7 & \vartheta_3 + \vartheta_2\vartheta_8 \\ \vartheta_4 + \vartheta_5\vartheta_7 & \vartheta_6 + \vartheta_5\vartheta_8 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$
$(2,2)$	15	$\begin{pmatrix} 0 & 1 & 0 & 0 \\ \vartheta_1 & \vartheta_2 & \vartheta_3 & \vartheta_4 \\ 0 & 0 & 0 & 1 \\ \vartheta_5 & \vartheta_6 & \vartheta_7 & \vartheta_8 \end{pmatrix}$	$\begin{pmatrix} \vartheta_9 & \vartheta_{10} \\ \vartheta_1 + \vartheta_4\vartheta_{11} + \vartheta_2\vartheta_9 & \vartheta_3 + \vartheta_2\vartheta_{10} + \vartheta_4\vartheta_{12} \\ \vartheta_{11} & \vartheta_{12} \\ \vartheta_5 + \vartheta_8\vartheta_{11} + \vartheta_6\vartheta_9 & \vartheta_7 + \vartheta_6\vartheta_{10} + \vartheta_8\vartheta_{12} \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$

Table 3.1.: Canonical state space realizations (A, B, C) of normalized $(H(0) = -\mathbf{1}_2)$ rational transfer functions in $M_2(\mathbb{R}\{z\})$ with different Kronecker indices $\boldsymbol{\nu}$; the number of parameters, $n(\boldsymbol{\nu})$, includes three parameters for a covariance matrix $\Sigma^{\boldsymbol{L}}$.

ν	$n(\nu)$	$P(z)$	$Q(z)$	(p,q)
$(1,1)$	7	$\begin{pmatrix} z - \vartheta_1 & -\vartheta_2 \\ -\vartheta_3 & z - \vartheta_4 \end{pmatrix}$	$\begin{pmatrix} \vartheta_1 & \vartheta_2 \\ \vartheta_3 & \vartheta_4 \end{pmatrix}$	$(1,0)$
$(1,2)$	10	$\begin{pmatrix} z - \vartheta_1 & -\vartheta_2 \\ -\vartheta_3 & z^2 - \vartheta_4 z - \vartheta_5 \end{pmatrix}$	$\begin{pmatrix} \vartheta_1 & \vartheta_2 \\ \vartheta_6 z + \vartheta_3 & \vartheta_7 z + \vartheta_5 \end{pmatrix}$	$(2,1)$
$(2,1)$	11	$\begin{pmatrix} z^2 - \vartheta_1 z - \vartheta_2 & -\vartheta_3 \\ -\vartheta_4 z - \vartheta_5 & z - \vartheta_6 \end{pmatrix}$	$\begin{pmatrix} \vartheta_7 z + \vartheta_2 & \vartheta_8 z + \vartheta_3 \\ \vartheta_5 & \vartheta_6 \end{pmatrix}$	$(2,1)$
$(2,2)$	15	$\begin{pmatrix} z^2 - \vartheta_1 z - \vartheta_2 & -\vartheta_3 z - \vartheta_4 \\ -\vartheta_5 z - \vartheta_6 & z^2 - \vartheta_7 z - \vartheta_8 \end{pmatrix}$	$\begin{pmatrix} \vartheta_9 z + \vartheta_2 & \vartheta_{10} z + \vartheta_4 \\ \vartheta_{11} z + \vartheta_6 & \vartheta_{12} z + \vartheta_8 \end{pmatrix}$	$(2,1)$

Table 3.2.: Canonical MCARMA realizations (P, Q) with order (p, q) of normalized $(H(0) = -\mathbf{1}_2)$ rational transfer functions in $M_2(\mathbb{R}\{z\})$ with different Kronecker indices $\boldsymbol{\nu}$; the number of parameters, $n(\boldsymbol{\nu})$, includes three parameters for a covariance matrix $\Sigma^{\boldsymbol{L}}$.

bivariate normal-inverse Gaussian Lévy process is characterized by the density

$$f_{\mathrm{NIG}}(\boldsymbol{x};\boldsymbol{\mu},\alpha,\boldsymbol{\beta},\delta,\Delta)=\frac{\delta\exp(\delta\kappa)}{2\pi}\frac{\exp(\langle\boldsymbol{\beta}\boldsymbol{x}\rangle)}{\exp(\alpha g(\boldsymbol{x}))}\frac{1+\alpha g(\boldsymbol{x})}{g(\boldsymbol{x})^3},\quad \boldsymbol{x}\in\mathbb{R}^2,$$

where

$$g(\boldsymbol{x})=\sqrt{\delta^2+\langle\boldsymbol{x}-\boldsymbol{\mu},\Delta(\boldsymbol{x}-\boldsymbol{\mu})\rangle},\quad \kappa^2=\alpha^2-\langle\boldsymbol{\beta},\Delta\boldsymbol{\beta}\rangle>0,$$

and $\boldsymbol{\mu}\in\mathbb{R}^2$ is a location parameter, $\alpha\geqslant 0$ is a shape parameter, $\boldsymbol{\beta}\in\mathbb{R}^2$ is a symmetry parameter, $\delta\geqslant 0$ is a scale parameter and $\Delta\in M_2^+(\mathbb{R})$, $\det\Delta=1$, determines the dependence between the two components of $(\boldsymbol{L}(t))_{t\in\mathbb{R}}$. For our simulation study we chose parameters

$$\delta=1,\quad \alpha=3,\quad \boldsymbol{\beta}=(1,1)^T,\quad \Delta=\begin{pmatrix}5/4 & -1/2\\ -1/2 & 1\end{pmatrix},\quad \boldsymbol{\mu}=-\frac{1}{2\sqrt{31}}(3,2)^T,\tag{3.4.6}$$

resulting in a skewed, semi-heavy-tailed distribution with mean zero and covariance matrix

$$\Sigma^{\boldsymbol{L}}=\frac{1}{31^{3/2}}\begin{pmatrix}82 & -28\\ -28 & 64\end{pmatrix}\approx\begin{pmatrix}0.4751 & -0.1622\\ -0.1622 & 0.3708\end{pmatrix}.\tag{3.4.7}$$

A sample of 350 independent replicates of the bivariate CARMA$_{1,2}$ process $(\boldsymbol{Y}(t))_{t\in\mathbb{R}}$ driven by a normal-inverse Gaussian Lévy process $(\boldsymbol{L}(t))_{t\in\mathbb{R}}$ with parameters given in Eq. (3.4.6) were simulated on the equidistant grid $0,0.01,0.02,\ldots,2000$ by applying an Euler scheme to the stochastic differential equation

$$\mathrm{d}\boldsymbol{X}(t)=\begin{pmatrix}\vartheta_1 & \vartheta_2 & 0\\ 0 & 0 & 1\\ \vartheta_3 & \vartheta_4 & \vartheta_5\end{pmatrix}\boldsymbol{X}(t)\mathrm{d}t+\begin{pmatrix}\vartheta_1 & \vartheta_2\\ \vartheta_6 & \vartheta_7\\ \vartheta_3+\vartheta_5\vartheta_6 & \vartheta_4+\vartheta_5\vartheta_7\end{pmatrix}\mathrm{d}\boldsymbol{L}(t),\tag{3.4.8a}$$

$$\boldsymbol{Y}(t)=\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\end{pmatrix}\boldsymbol{X}(t),\tag{3.4.8b}$$

making use of the canonical parametrization given in Table 3.1. For the simulation, the parameter values

$$(\vartheta_1,\vartheta_2,\vartheta_3,\vartheta_4,\vartheta_5,\vartheta_6,\vartheta_7)=(-1,-2,1,-2,-3,1,2),$$

and the initial value $\boldsymbol{X}(0)=\boldsymbol{0}_3$ were used. Each realization was sampled at integer times ($h=1$). Figures 3.1 and 3.2 show one typical realization of this NIG-driven CARMA process as well as the effect of sampling the process at discrete times.

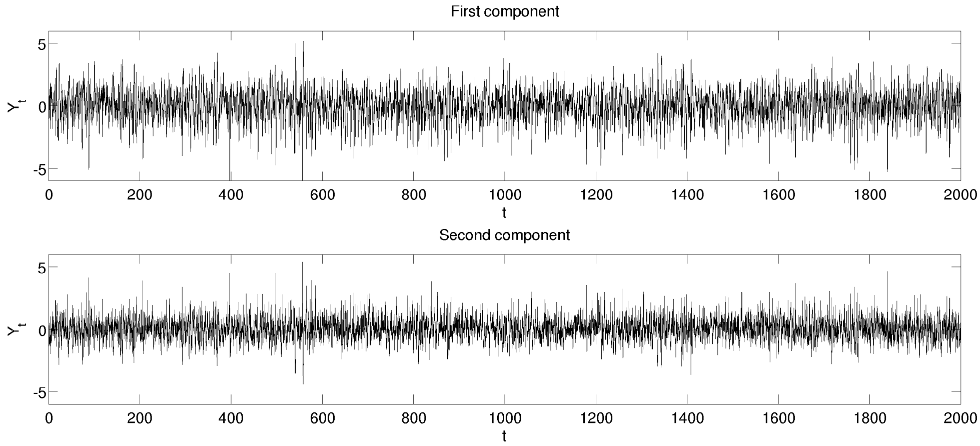


Figure 3.1.: Typical realization of the bivariate NIG-driven $\mathrm{MCARMA}_{1,2}$ process given by Eqs. (3.4.8) on the interval $[0, 2000]$

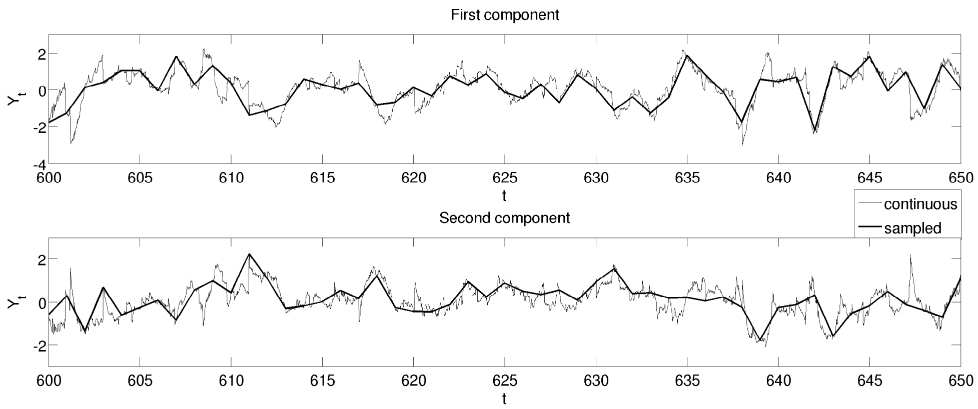


Figure 3.2.: Zoom-in of Fig. 3.1 onto the time interval $[600, 650]$. The thick line is the linear interpolation of the values of the displayed MCARMA process at integer times, illustrating the effect of sampling.

Using a differential evolution optimization routine (Price, Storn and Lampinen, 2005) in conjunction with a subspace trust-region method (Branch, Coleman and Li, 1999; Byrd, Schnabel and Shultz, 1988), we computed quasi maximum likelihood estimates of $\vartheta_1, \ldots, \vartheta_7$ as well as $(\vartheta_8, \vartheta_9, \vartheta_{10}) \coloneqq \operatorname{vech} \Sigma^{\boldsymbol{L}}$ by numerical maximization of the quasi log-likelihood function. In Table 3.3 the sample means and sampled standard deviations of the estimates are reported. Moreover, the standard deviations were estimated using the square roots of the diagonal entries of the asymptotic covariance matrix (3.2.67) with $s(L) = \lfloor L / \log L \rfloor^{1/3}$, and the estimates are also displayed in Table 3.3. One sees that the bias, the difference between the sample mean and the true parameter value, is very small in accordance with the asymptotic consistency of the estimator. Moreover, the estimated standard deviation is always slightly larger than the sample standard deviation, yet close enough to provide a useful approximation for, e. g., the construction of confidence regions. In order not to underestimate the uncertainty in the estimate, such a conservative approximation to the true standard deviations is desirable in practice. Overall, the estimation procedure performs very well in the simulation study.

3.4.3. Application to weekly bond yields

In this section we provide an illustrative data example and apply the techniques established in the preceding sections to the bivariate weekly series of Moody's seasoned Aaa and Baa corporate bond yields from October 1966 through April 2009; these data are available from the Federal Reserve Bank of St. Louis. We first took the logarithm of the data and the resulting series was seen to have a unit root in each component, so the next step in the data preparation was differencing at lag 1. Using a moving window of length 52 — corresponding to a period of one year — we removed the stochastic volatility effects displayed by the differenced time series to obtain data with no obvious departure from stationarity. Figure 3.3 shows the weekly bond log-yields after differencing and devolatilization.

We fitted bivariate CARMA processes of McMillan degrees $n = 2, 3, 4$ using the quasi maximum likelihood method described in Section 3.3.4 and employing the canonical parametrizations of Section 3.4.1. The numerical values of $\hat{\boldsymbol{\vartheta}}$ as well as their standard errors estimated by the square root of the diagonal entries in the approximate asymptotic covariance matrix $\hat{\Xi}_s^L$, defined in Eq. (3.2.67), can be found in Table 3.4. The last row displays the value of twice the negative logarithm of the Gaussian likelihood of the observations under the model corresponding to the estimated parameter value $\hat{\boldsymbol{\vartheta}}$. The quality of the fit can be assessed from Fig. 3.4 where we compare the autocorrelation functions of the fitted models with the empirical autocorrelation function of the data. One sees how the fit becomes better as one

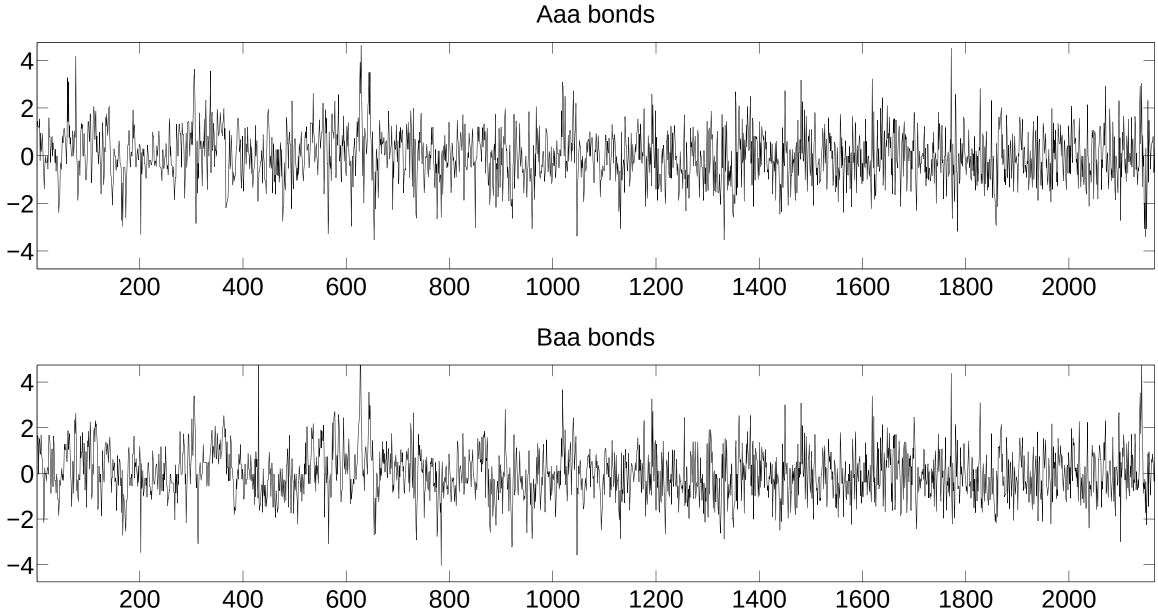


Figure 3.3.: Weekly Aaa and Baa bond yields after differencing and devolatilization

increases the model order in accordance with an increasing value of the Gaussian likelihood; in particular, the autocorrelations at higher lags are better captured by the higher order models. This phenomenon is well known from the estimation of discrete-time parametric processes where penalty terms in the likelihood together with order selection criteria like the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) are used to formalize the trade-off between goodness of fit and model complexity. Understanding their applicability in a continuous-time set-up remains a problem for future research.

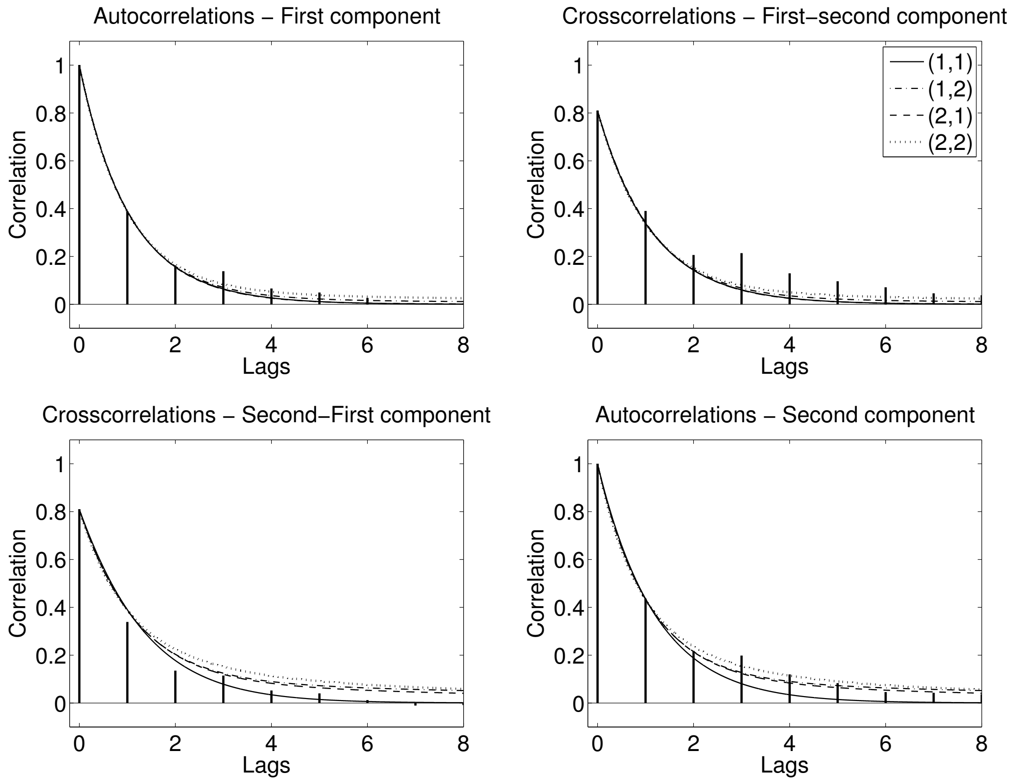


Figure 3.4.: Empirical auto- and crosscorrelations of the weekly bond data from Fig. 3.3 compared to the theoretical auto- and crosscorrelations of estimated $\mathrm{MCARMA}_{\alpha,\beta}$ models, for different Kronecker indices (α, β)

parameter	sample mean	bias	sample std. dev.	mean est. std. dev.
ϑ_1	-1.0001	0.0001	0.0354	0.0381
ϑ_2	-2.0078	0.0078	0.0479	0.0539
ϑ_3	1.0051	-0.0051	0.1276	0.1321
ϑ_4	-2.0068	0.0068	0.1009	0.1202
ϑ_5	-2.9988	-0.0012	0.1587	0.1820
ϑ_6	1.0255	-0.0255	0.1285	0.1382
ϑ_7	2.0023	-0.0023	0.0987	0.1061
ϑ_8	0.4723	-0.0028	0.0457	0.0517
ϑ_9	-0.1654	0.0032	0.0306	0.0346
ϑ_{10}	0.3732	0.0024	0.0286	0.0378

Table 3.3.: Quasi maximum likelihood estimates for the parameters of a bivariate NIG-driven $\text{CARMA}_{1,2}$ process observed at integer times over the time horizon $[0, 2000]$. The second column reports the empirical mean of the estimators as obtained from 350 independent simulated paths; the third and fourth columns contain the resulting bias and the sample standard deviation of the estimators, respectively, while the last column reports the average of the expected standard deviations of the estimators as obtained from the asymptotic normality result Theorem 3.50.

(α, β)	$(1,1)$		$(1,2)$		$(2,1)$		$(2,2)$	
	$\hat{\vartheta}_i$	$\sigma(\vartheta_i)$	$\hat{\vartheta}_i$	$\sigma(\vartheta_i)$	$\hat{\vartheta}_i$	$\sigma(\vartheta_i)$	$\hat{\vartheta}_i$	$\sigma(\vartheta_i)$
$\hat{\vartheta}_1$	-1.1326	0.1349	-1.1538	0.1401	-1.3776	0.0320	**-0.0010**	**0.0336**
$\hat{\vartheta}_2$	**0.2054**	**0.1171**	0.2307	0.1008	-2.4033	0.0197	-1.1601	0.5964
$\hat{\vartheta}_3$	0.3316	0.1206	**-0.2528**	**0.1716**	0.0228	0.0050	**-0.0098**	**0.0268**
$\hat{\vartheta}_4$	-1.0935	0.1065	**-0.0362**	**0.0472**	-4.9948	0.1096	**0.1829**	**0.7429**
$\hat{\vartheta}_5$	2.4105	0.2324	-1.2516	0.1286	-4.6276	0.1538	1.4646	0.3931
$\hat{\vartheta}_6$	2.2483	0.2061	-2.5747	0.4595	**-0.0153**	**0.0108**	1.3662	0.4039
$\hat{\vartheta}_7$	2.7055	0.2116	1.6345	0.2940	-1.2442	0.0391	-0.7438	0.2387
$\hat{\vartheta}_8$			2.8552	0.1966	0.2573	0.0492	-1.7563	0.7209
$\hat{\vartheta}_9$			3.5702	0.2151	2.4302	0.1370	-2.6936	0.6694
$\hat{\vartheta}_{10}$			4.9076	0.3888	2.9784	0.2766	1.7369	0.5381
$\hat{\vartheta}_{11}$					4.1571	0.5043	**-3.6136**	**3.0265**
$\hat{\vartheta}_{12}$							**2.8483**	**2.5122**
$\hat{\vartheta}_{13}$							4.4848	0.3327
$\hat{\vartheta}_{14}$							5.5079	0.1803
$\hat{\vartheta}_{15}$							7.0218	1.4357
$\mathscr{L}(\hat{\boldsymbol{\vartheta}}, \boldsymbol{y}^L)$	**9,893.8**		**9,850.4**		**9,853.0**		**9,840.7**	

Table 3.4.: Quasi maximum likelihood estimates of the parameters of an $\text{MCARMA}_{\alpha,\beta}$ model for weekly yields of Moody's seasoned corporate bonds. The marginal standard deviations $\sigma(\vartheta_i)$ are estimated from the diagonal elements of the asymptotic covariance matrix in Theorem 3.50. The parameters whose confidence region contains zero are marked in bold.

4. Parametric Estimation of the Driving Lévy Process of Multivariate CARMA Processes from High-Frequency Observations

4.1. Introduction

Continuous-time autoregressive moving average (CARMA) processes generalize the widely employed discrete-time ARMA process to a continuous-time setting. Heuristically, a multivariate CARMA process of order (p, q), $p > q$, can be thought of as a stationary solution $\boldsymbol{Y}$ of the linear differential equation

$$\left[\mathrm{D}^p + A_1\mathrm{D}^{p-1} + \ldots + A_p\right] \boldsymbol{Y}(t) = \left[B_0 + B_1\mathrm{D} + \ldots + B_q\mathrm{D}^q\right] \mathrm{D}\boldsymbol{L}(t), \quad \mathrm{D} = \frac{\mathrm{d}}{\mathrm{d}t}, \tag{4.1.1}$$

where $\boldsymbol{L}$ is a Lévy process and A_i, B_j are coefficient matrices, see Section 4.3 for a precise definition. They first appeared in the literature in Doob (1944), where univariate Gaussian CARMA processes were defined. Recent years have seen a rapid development in both the theory and the applications of this class of stochastic processes (see, e. g., Brockwell, 2004, and references therein). In Brockwell (2001b), the restriction of Gaussianity was relaxed and CARMA processes driven by Lévy processes with finite moments of any order greater than zero were introduced (see also Brockwell and Lindner, 2009). This extension allowed for CARMA processes to have jumps as well as a wide variety of marginal distributions, possibly exhibiting fat tails. Shortly after that, Marquardt and Stelzer (2007) defined multivariate CARMA processes and thereby made it possible to model a set of dependent time series jointly by a single continuous-time linear process. For further developments of the concept, which led to fractionally integrated CARMA (FICARMA) and superpositions of CARMA (supCARMA) processes, and allow for long-memory effects, we refer the reader to Barndorff-Nielsen and Stelzer (2011); Brockwell and Marquardt (2005); Marquardt (2007). In many contexts, continuous-time processes are particularly suitable for stochastic modelling because they allow for irregularly-spaced observations and high-frequency sampling. We refer the reader to Barndorff-Nielsen and Shephard (2001b); Benth and Šaltytė Benth (2009); Todorov and Tauchen (2006) for an overview of successful applications of CARMA processes in economics and mathematical finance.

Despite the growing interest of practitioners in using CARMA processes as stochastic models for observed time series, the statistical theory for such processes has received little attention in the past. One of the basic questions with regard to parameter inference or model selection is how to determine which particular member of a class of stochastic models best describes the characteristic statistical properties of an observed time series. If one decides to model a phenomenon by a CARMA process as in Eq. (4.1.1), which can often be argued to be a reasonable choice of model class, this problem reduces to the three tasks of choosing suitable integers p, q describing the order of the process; estimating the coefficient matrices A_i, B_j; and suggesting an appropriate model for the driving Lévy process $\boldsymbol{L}$.

In this chapter, we address the last of these three problems and develop a method to estimate a parametric model for the driving Lévy process of a multivariate CARMA process, building on an idea suggested in Brockwell et al. (2011) for the special case of a univariate CARMA process of order $(2,1)$. The strategy is to observe that the distribution of a Lévy process $\boldsymbol{L}$ is uniquely determined by the distribution of the unit increments $\Delta\boldsymbol{L}_n = \boldsymbol{L}(n) - \boldsymbol{L}(n-1)$; if one therefore had access to the increments $(\Delta\boldsymbol{L}_n)_{n=1,\ldots,N}$ over a sufficiently long time-horizon, one could easily estimate a model for $\boldsymbol{L}$ by any of several well-established methods, including parametric as well as non-parametric approaches (Figueroa-López, 2009; Gugushvili, 2009, and references therein). It is thus natural to try and express the increments of the driving Lévy process – at least approximately – in terms of the observed values of the CARMA process, and to subject this approximate sample from the unit-increment distribution to the same estimation method one would use with the true sample. One difficulty arising in this step is that one usually does not observe a CARMA processes continuously but that one instead only has access to its values on a discrete, yet possibly very fine, time grid; in fact, as we shall see in Section 4.4, it is this assumption of discrete-time observations that prevents us from exactly recovering the increments of the Lévy process from the recorded CARMA process.

In the following, we concentrate on the parametric generalized moment estimators (see, e. g., Hansen, 1982; Newey and McFadden, 1994), and we prove that the estimate based on the reconstructed increments of $\boldsymbol{L}$ has the same asymptotic distribution as the estimate based on the true increments, provided that both the length N of the observation period and the sampling frequency h^{-1} at which the CARMA process is recorded, go to infinity at the right rate. In fact we obtain the quantitative high-frequency condition that $h = h_N$ must be chosen dependent on N such that Nh_N converges to zero as N tends to infinity. The generalized method of moments (GMM) estimators contain as special cases the classical maximum likelihood estimators as well as non-linear least squares estimators that are based on fitting the empirical characteristic function of the observed sample to its theoretical counterpart. In view of the structure of the Lévy–Khintchine formula, the latter method is particularly well suited for the estimation of Lévy processes. We impose no assumptions on the driving Lévy process except for the finiteness of certain moments that depend on the particular moment function used in the GMM approach. In our main result, Theorem 4.34, we prove the consistency and asymptotic normality of a wide class of GMM estimators that satisfy a set of mild standard technical assumptions.

It seems possible to relax the assumption of uniform sampling as long as the maximal distance between two recording times in the observation interval tends to zero. More important, however, is the natural question if there exist methods to estimate the driving Lévy process of a CARMA process that do not require high-frequency sampling but still have desirable asymptotic properties. Another interesting topic for further investigation is the behaviour of non-parametric estimators for the driving Lévy process when they are used with a disturbed sample of the unit increments as described in this chapter. Maximum likelihood estimation of continuously observed Gaussian diffusions and one-dimensional Lévy-driven Ornstein–Uhlenbeck processes has been considered in Liptser and Shiryaev (2001); Mai (2009). An extension of their methods to multivariate CARMA models remains an important open problem.

Outline of the chapter The chapter is structured as follows. In Section 4.2 we take a closer look at multivariate Lévy processes and infinitely divisible distributions, the fundamental ingredients in the definition of a multivariate CARMA process. First, we briefly review their definition and some important basic properties. In Section 4.2.2 we obtain a new quantitative bound for the absolute moments of an infinitely divisible distribution in terms of its characteristic triplet, which is essential for many of the subsequent proofs. We also derive the exact polynomial time-dependence of the absolute moments of a Lévy process in Proposition 4.3. As a further preparation for the proofs of our main results, Theorem 4.4 in Section 4.2.3 establishes a Fubini-type result for double integrals with respect to a Lévy process over an unbounded domain.

The definition of multivariate CARMA processes as well as important properties, such as moments, mixing and smoothness of sample paths, are presented in Section 4.3. In Theorem 4.6 we prove an alternative state space representation for multivariate CARMA processes, called the controller canonical form, which lends itself more easily to the estimation of the driving Lévy process than the original definition.

In Section 4.4 we show that, conditional on an initial value whose influence decays exponentially fast, one can exactly recover the value of the driving Lévy process from a continuous record of the multivariate CARMA process. The functional dependence is explicit and given in Theorem 4.11.

Since such a continuous record is usually not available, Section 4.5 is devoted to discretizing the result found in Theorem 4.11. To this end, we analyse how path-wise derivatives and definite integrals of Lévy-driven CARMA processes can be approximated from observations on a discrete time grid, and we determine the asymptotic behaviour of these approximations as the mesh size tends to zero. To our knowledge, this is the first time that numerical differentiation and integration schemes are investigated quantitatively for this class of stochastic processes. The results of this section are summarized in Theorem 4.25.

In Section 4.6 we prove consistency and asymptotic normality of the generalized method of moments estimator when the sample is not i. i. d., but instead disturbed by a noise sequence, which corresponds to the discretization error from the previous section. Theorem 4.28 shows that if the sampling frequency h_N^{-1} goes to infinity fast enough with the length N of the observation interval, such that Nh_N converges to zero, then the effect of the discretization becomes asymptotically negligible, and the limiting distribution of the estimated parameter is identical to the one obtained from an unperturbed sample. Finally, in Theorem 4.34, we apply this result to give an answer to the question of how to estimate a parametric model of the driving Lévy process of a multivariate CARMA process if high-frequency observations are available.

Finally, in Section 4.7, we present the results of a simulation study for a Gamma-driven CARMA(3,1) process, which illustrate our theoretical results and demonstrate their practical applicability.

Notation We use the following notation. The natural, real, complex numbers, and the integers are denoted by $\mathbb{N}$, $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{Z}$, respectively. Vectors in $\mathbb{R}^m$ are printed in bold, and we use superscripts to denote the components of a vector, e. g., $\mathbb{R}^m \ni \boldsymbol{x} = (x^1, \ldots, x^m)$. We write $\mathbf{0}_m$ for the zero vector in $\mathbb{R}^m$, and we let $\|\cdot\|$ and $\langle\cdot\rangle$ represent the Euclidean norm and inner product, respectively. The ring of polynomial expressions in z over a ring $\mathbb{K}$ is denoted by $\mathbb{K}[z]$. The symbols $M_{m,n}(\mathbb{K})$, or $M_m(\mathbb{K})$ if $m = n$, stand for the space of $m \times n$ matrices with entries in $\mathbb{K}$. The transpose of a matrix A is written as A^T, and $\mathbf{1}_m$ and 0_m denote the identity and the zero element in $M_m(\mathbb{K})$, respectively. The symbol $\|\cdot\|$ is also used for the operator norm on $M_{m,n}(\mathbb{R})$ induced by the Euclidean vector norm. For any topological space X, the symbol $\mathscr{B}(X)$ denotes the Borel σ-algebra on X. We frequently use the following Landau notation: for two functions f and g defined on the interval $[0, 1]$, we write $f(h) = O(g(h))$ if there exists a constant C such that $\|f(h)\| \leqslant Cg(h)$ for all $h < 1$. We use the notation $\|\cdot\|_{L^p}$ for the norm on the classical L^p spaces. The symbol Leb stands for the Lebesgue measure, and the indicator function of a set B is denoted by $I_B(\cdot)$, defined to be one if the argument lies in B, and zero otherwise. We write $\xrightarrow{p}$ and $\xrightarrow{d}$ for convergence in probability and convergence in distribution, respectively, and use the symbol $\overset{d}{=}$ to denote equality in distribution of two random variables. Finally, for a positive real number α, we write $(\alpha)_0$ for the smallest even integer greater than or equal to α.

4.2. Lévy processes and infinitely divisible distributions

4.2.1. Definition and Lévy–Itô decomposition

Lévy processes are the main ingredient in the definition of a multivariate CARMA process and an important object of study in this thesis. In this section we review their definition and some elementary properties. A detailed account can be found in Applebaum (2004); Sato (1999).

Definition 4.1 (Lévy process) A (one-sided) $\mathbb{R}^m$-valued *Lévy process* $(\boldsymbol{L}(t))_{t\geqslant 0}$ is a stochastic process, defined on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$, with stationary, independent increments, continuous in probability and satisfying $\boldsymbol{L}(0) = \boldsymbol{0}_m$ almost surely.

Every $\mathbb{R}^m$-valued Lévy process $(\boldsymbol{L}(t))_{t\geqslant 0}$ can without loss of generality be assumed to be càdlàg, which means that the sample paths are right-continuous and have left limits; it is completely characterized by its characteristic function in the Lévy–Khintchine form

$$\mathbb{E}e^{i\langle \boldsymbol{u},\boldsymbol{L}(t)\rangle} = \exp\{t\psi^{\boldsymbol{L}}(\boldsymbol{u})\}, \quad \boldsymbol{u}\in\mathbb{R}^m, \quad t\geqslant 0,$$

where $\psi^{\boldsymbol{L}}$ has the special form

$$\psi^{\boldsymbol{L}}(\boldsymbol{u}) = i\langle\gamma^{\boldsymbol{L}},\boldsymbol{u}\rangle - \frac{1}{2}\langle\boldsymbol{u},\Sigma^{\mathcal{G}}\boldsymbol{u}\rangle + \int_{\mathbb{R}^m}\left[e^{i\langle\boldsymbol{u},\boldsymbol{x}\rangle} - 1 - i\langle\boldsymbol{u},\boldsymbol{x}\rangle I_{\{\|x\|\leqslant 1\}}\right]\nu^{\boldsymbol{L}}(d\boldsymbol{x}). \tag{4.2.1}$$

The vector $\gamma^{\boldsymbol{L}}\in\mathbb{R}^m$ is called the *drift*, the positive semidefinite, symmetric $m\times m$ matrix $\Sigma^{\mathcal{G}}$ is the *Gaussian covariance matrix* and $\nu^{\boldsymbol{L}}$ is a measure on $\mathbb{R}^m$, referred to as the *Lévy measure*, satisfying

$$\nu^{\boldsymbol{L}}(\{\boldsymbol{0}_m\}) = 0, \quad \int_{\mathbb{R}^m}\min(\|\boldsymbol{x}\|^2,1)\nu^{\boldsymbol{L}}(d\boldsymbol{x}) < \infty.$$

Put differently, for every $t\geqslant 0$, the distribution of $\boldsymbol{L}(t)$ is infinitely divisible with characteristic triplet $(t\gamma^{\boldsymbol{L}}, t\Sigma^{\mathcal{G}}, t\nu^{\boldsymbol{L}})$. By the Lévy–Itô decomposition, the paths of $\boldsymbol{L}$ can be decomposed almost surely into a Brownian motion with drift, a compound Poisson process, and a purely discontinuous L^2-martingale according to

$$\boldsymbol{L}(t) = \gamma^{\boldsymbol{L}}t + \left(\Sigma^{\mathcal{G}}\right)^{1/2}\boldsymbol{W}_t + \int_{\|\boldsymbol{x}\|\geqslant 1}\int_0^t \boldsymbol{x}N(ds,d\boldsymbol{x}) + \lim_{\varepsilon\searrow 0}\int_{\varepsilon\leqslant\|\boldsymbol{x}\|<1}\int_0^t \boldsymbol{x}\tilde{N}(ds,d\boldsymbol{x}), \tag{4.2.2}$$

where $\boldsymbol{W}$ is a standard m-dimensional Wiener process and $\left(\Sigma^{\mathcal{G}}\right)^{1/2}$ is the unique positive semidefinite matrix square root of $\Sigma^{\mathcal{G}}$ defined by functional calculus. The measure N is a Poisson random measure on $\mathbb{R}\times\mathbb{R}^m\backslash\{\boldsymbol{0}_m\}$, independent of $\boldsymbol{W}$ with intensity measure $\mathrm{Leb}\otimes\nu^{\boldsymbol{L}}$ describing the jumps of $\boldsymbol{L}$. More precisely, for any measurable set $B\in\mathscr{B}(\mathbb{R}\times$

$\mathbb{R}^m \backslash \{\mathbf{0}_m\})$,

$$N(B) = \#\{s \geqslant 0 : (s, \boldsymbol{L}(s) - \boldsymbol{L}(s-)) \in B\},$$

where $\boldsymbol{L}(s-) := \lim_{t \nearrow s} \boldsymbol{L}(t)$ denotes a left limit of $\boldsymbol{L}$. Finally, $\tilde{N}$ is the compensated jump measure defined by $\tilde{N}(\mathrm{d}s, \mathrm{d}\boldsymbol{x}) = N(\mathrm{d}s, \mathrm{d}\boldsymbol{x}) - \mathrm{d}s\, \nu^{\boldsymbol{L}}(\mathrm{d}\boldsymbol{x})$. We will work with two-sided Lévy processes $\boldsymbol{L} = (\boldsymbol{L}(t))_{t\in\mathbb{R}}$. These are obtained from two independent copies $(\boldsymbol{L}_1(t))_{t\geqslant 0}$, $(\boldsymbol{L}_2(t))_{t\geqslant 0}$ of a one-sided Lévy process via the construction

$$\boldsymbol{L}(t) = \begin{cases} \boldsymbol{L}_1(t), & t \geqslant 0, \\ -\boldsymbol{L}_2(-t-), & t < 0. \end{cases}$$

In the following we present some elementary facts about stochastic integrals with respect to Lévy processes, which we will use later. Comprehensive accounts of this wide field are given in the textbooks Applebaum (2004); Protter (1990). Let $f : \mathbb{R} \to M_{d,m}(\mathbb{R})$ be a measurable, square-integrable function. Under the condition that $\boldsymbol{L}(1)$ has finite second moments, the stochastic integral

$$I = \int_{\mathbb{R}} f(s)\mathrm{d}\boldsymbol{L}(s)$$

exists in $L^2(\Omega, \mathbb{P})$. Moreover, the distribution of the random variable I is infinitely divisible with characteristic triplet $(\gamma_f, \Sigma_f, \nu_f)$ which can be expressed explicitly in terms of the characteristic triplet of $\boldsymbol{L}$ via the formulæ (Rajput and Rosiński, 1989, Theorem 2.7)

$$\gamma_f = \int_{\mathbb{R}} f(s) \left[\gamma^{\boldsymbol{L}} + \int_{\mathbb{R}^d} \boldsymbol{x} \left(I_{[0,1]}(\|f(s)\boldsymbol{x}\|) - I_{[0,1]}(\|\boldsymbol{x}\|)\right) \nu^{\boldsymbol{L}}(\mathrm{d}\boldsymbol{x})\right] \mathrm{d}s, \tag{4.2.3a}$$

$$\Sigma_f = \int_{\mathbb{R}} f(s)\Sigma^{\mathcal{G}} f(s)^T \mathrm{d}s, \tag{4.2.3b}$$

and

$$\nu_f(B) = \int_{\mathbb{R}} \int_{\mathbb{R}^m} I_B(f(s)\boldsymbol{x})\nu^{\boldsymbol{L}}(\mathrm{d}\boldsymbol{x})\mathrm{d}s, \quad B \in \mathscr{B}(\mathbb{R}^d \backslash \{\mathbf{0}_d\}). \tag{4.2.3c}$$

4.2.2. Bounds for the absolute moments of infinitely divisible distributions and Lévy processes

In this short section we derive some bounds for the absolute moments of multivariate infinitely divisible distributions and Lévy processes which will turn out to be essential for the proofs of our main results later. It is well known (Sato, 1999, Corollary 25.8) that the kth absolute moment of an infinitely divisible random variable X with characteristic triplet (γ, Σ, ν) is finite if and only if the measure ν, restricted to $\{\|\boldsymbol{x}\| \geqslant 1\}$, has a finite kth absolute moment. We need the following stronger result, which establishes a quantitative bound for the absolute moments of an infinitely divisible distribution in terms of its characteristic triplet.

Lemma 4.2 *Let X be an infinitely divisible, $\mathbb{R}^m$-valued random variable with characteristic triplet (γ, Σ, ν), and let k be a positive even integer. Assume that the constants c_i, C_i, $i = 1, 2$, satisfy*

$$\int_{\|x\|<1} \|x\|^r \, \nu(\mathrm{d}x) \leqslant C_0 c_0^r, \quad r = 2, \ldots, k, \tag{4.2.4a}$$

$$\int_{\|x\|\geqslant 1} \|x\|^r \, \nu(\mathrm{d}x) \leqslant C_1 c_1^r, \quad r = 1, \ldots, k. \tag{4.2.4b}$$

Then there exists a constant $C > 0$, depending on m and k, but not on (γ, Σ, ν), such that

$$\mathbb{E}\, \|X\|^k \leqslant C \left[\|\gamma\|^k + \|\Sigma\|^{k/2} + c_0^k + c_1^k \right]. \tag{4.2.5}$$

Proof Denote by $\nu_0 = \nu|_{\{\|x\|<1\}}$ and $\nu_1 = \nu|_{\{\|x\|\geqslant 1\}}$ the restrictions of the measure ν to the unit ball of $\mathbb{R}^m$ and its complement, respectively. It follows from the Lévy–Khintchine formula (4.2.1) that we can construct a standard normal random variable $\boldsymbol{W}$ and two infinitely divisible random variables X_0, X_1, with characteristic triplets $(\mathbf{0}_m, 0_m, \nu_0)$, $(\mathbf{0}_m, 0_m, \nu_1)$, and distributions μ_0, μ_1, respectively, such that $X \overset{d}{=} \gamma + \Sigma^{1/2}\boldsymbol{W} + X_0 + X_1$. Using the notation $n!!$ for the double factorial of the natural number n as well as Bauer (2002, Eq. (4.20)) for the absolute moments of a standard normal random variable, the kth absolute moment of the Gaussian part is readily estimated as

$$\begin{aligned}
\mathbb{E} \left\| \Sigma^{1/2} \boldsymbol{W} \right\|^k &\leqslant \|\Sigma\|^{k/2} \, \mathbb{E}\, \|\boldsymbol{W}\|^k \leqslant \|\Sigma\|^{k/2} \, \mathbb{E} \left(\sum_{i=1}^m \left| W^i \right| \right)^k \\
&\leqslant \|\Sigma\|^{k/2} \, m^{k+1} \mathbb{E} \left| W^1 \right|^k \leqslant (k-1)!! \, \|\Sigma\|^{k/2} \, m^{k+1},
\end{aligned}$$

which implies that

$$\mathbb{E}\, \|X\|^k \leqslant 4^k \left[\|\gamma\|^k + m^{k+1}(k-1)!! \, \|\Sigma\|^{k/2} + \mathbb{E}\, \|X_0\|^k + \mathbb{E}\, \|X_1\|^k \right]. \tag{4.2.6}$$

The first two terms in this sum are already of the form asserted in Eq. (4.2.5). We next consider the fourth term. By construction, the characteristic function of X_1 is given by

$$\widehat{\mu}_1(\boldsymbol{u}) := \mathbb{E} \mathrm{e}^{\mathrm{i}\langle \boldsymbol{u}, X_1 \rangle} = \exp \left\{ \int_{\|x\|\geqslant 1} \left[\mathrm{e}^{\mathrm{i}\langle \boldsymbol{u}, x \rangle} - 1 \right] \nu(\mathrm{d}x) \right\}, \quad \boldsymbol{u} \in \mathbb{R}^m.$$

By assumption (4.2.4b) and Sato (1999, Corollary 25.8), the integral $\int \|x\|^k \mu_1(\mathrm{d}x)$ is finite, and Sato (1999, Proposition 2.5(ix)) shows that the mixed moments of X_1 of order k are given by

$$\mathbb{E}\left(X_1^{i_1}\cdot\ldots\cdot X_1^{i_k}\right)=\int_{\mathbb{R}^m}x^{i_1}\cdot\ldots\cdot x^{i_k}\mu_1(\mathrm{d}\boldsymbol{x})=\frac{1}{\mathrm{i}^k}\left.\frac{\partial^k}{\partial u^{i_1}\cdot\ldots\cdot\partial u^{i_k}}\widehat{\mu_1}(\boldsymbol{u})\right|_{\boldsymbol{u}=\mathbf{0}_m},\quad i_j=1,\ldots,m.$$

It is easy to see by induction that

$$\frac{\partial^k}{\partial u^{i_1}\cdot\ldots\cdot\partial u^{i_k}}\widehat{\mu_1}(\boldsymbol{u})=[\widehat{\mu_1}(\boldsymbol{u})]^k\,\mathrm{i}^k\sum_{\pi\in\mathcal{P}_k}\prod_{B\in\pi}\int_{\|\boldsymbol{x}\|\geqslant 1}\left[\prod_{j\in B}x^{i_j}\right]\mathrm{e}^{\mathrm{i}\langle\boldsymbol{u},\boldsymbol{x}\rangle}\nu(\mathrm{d}\boldsymbol{x}),$$

where $\mathcal{P}_k$ denotes the set of partitions of $\{1,2,\ldots,k\}$, a partition being a subset of the power set of $\{1,\ldots,k\}$ with pair-wise disjoint elements such that their union is equal to $\{1,\ldots,k\}$. We write $\#\pi$ for the number of sets in a partition π and $|B|$ for the number of elements in such a set. Setting $\boldsymbol{u}=\mathbf{0}_m$, specializing to $i_j=i$, and making use of the assumption that k is even, the last display yields the explicit formula

$$\mathbb{E}\left|X_1^i\right|^k=\mathbb{E}\left(X_1^i\right)^k=\sum_{\pi\in\mathcal{P}_k}\prod_{B\in\pi}\int\left(x^i\right)^{|B|}\nu(\mathrm{d}\boldsymbol{x}),\quad i=1,\ldots,m.$$

Using the fact that $x^i\leqslant\|\boldsymbol{x}\|$ for every $\boldsymbol{x}\in\mathbb{R}^m$ as well as assumption (4.2.4b), we thus obtain that

$$\begin{aligned}\mathbb{E}\left\|X_1\right\|^k\leqslant& m^{k/2}\sum_{i=1}^m\mathbb{E}\left|X_1^i\right|^k\\ \leqslant& m^{k/2+1}\sum_{\pi\in\mathcal{P}_k}\prod_{B\in\pi}\int_{\|\boldsymbol{x}\|\geqslant 1}\|x\|^{|B|}\nu(\mathrm{d}\boldsymbol{x})\leqslant c_1^k m^{k/2+1}\sum_{\pi\in\mathcal{P}_k}C_1^{\#\pi}.\end{aligned}\tag{4.2.7}$$

The third term in Eq. (4.2.6) can be analysed similarly: the characteristic function of X_0 has the form

$$\widehat{\mu_0}(\boldsymbol{u}):=\mathbb{E}\mathrm{e}^{\mathrm{i}\langle\boldsymbol{u},X_0\rangle}=\exp\left\{\int_{\|\boldsymbol{x}\|<1}\left[\mathrm{e}^{\mathrm{i}\langle\boldsymbol{u},\boldsymbol{x}\rangle}-1-\mathrm{i}\langle\boldsymbol{u},\boldsymbol{x}\rangle\right]\nu(\mathrm{d}\boldsymbol{x})\right\},\quad\boldsymbol{u}\in\mathbb{R}^m.$$

With ν_0 having bounded support, all moments of X_0 are finite, which implies that $\widehat{\mu_0}$ is infinitely often differentiable and that the mixed moments of X_0 are given by partial derivatives of $\widehat{\mu_0}$, as before. The additional compensatory term $\mathrm{i}\langle\boldsymbol{u},\boldsymbol{x}\rangle$ in the integral ensures that the first derivative of $\widehat{\mu_0}$ vanishes at zero, which leads to

$$\mathbb{E}\left|X_0^i\right|^k=\mathbb{E}\left(X_0^i\right)^k=\sum_{\substack{\pi\in\mathcal{P}_k\\ \min\{|B|,B\in\pi\}\geqslant 2}}\prod_{B\in\pi}\int_{\|\boldsymbol{x}\|<1}\left(x^i\right)^{|B|}\nu(\mathrm{d}\boldsymbol{x}),\quad i=1,\ldots,m.$$

Using assumption (4.2.4a), we can thus estimate

$$\begin{aligned}\mathbb{E}\left\|X_0\right\|^k \leqslant & m^{k/2}\sum_{i=1}^{m}\mathbb{E}\left|X_0^i\right|^k \\ \leqslant & m^{k/2+1}\sum_{\substack{\pi\in\mathcal{P}_k\\ \min\{|B|,B\in\pi\}\geqslant 2}}\prod_{B\in\pi}\int_{\|x\|<1}\|x\|^{|B|}\nu(\mathrm{d}x)\leqslant c_0^k m^{k/2+1}\sum_{\substack{\pi\in\mathcal{P}_k\\ \min\{|B|,B\in\pi\}\geqslant 2}} C_0^{\#\pi}. \quad (4.2.8)\end{aligned}$$

The bounds (4.2.6) to (4.2.8) show that the claim (4.2.5) holds with

$$C := 4^k\left[m^{k+1}(k-1)!! + m^{k/2+1}\left(\sum_{\pi\in\mathcal{P}_k} C_1^{\#\pi} + \sum_{\substack{\pi\in\mathcal{P}_k\\ \min\{|B|,B\in\pi\}\geqslant 2}} C_0^{\#\pi}\right)\right]. \qquad \square$$

Since the marginal distributions of a Lévy process $\boldsymbol{L}$ are infinitely divisible, the behaviour of their moments can be analysed by the previous Lemma 4.2. We prefer, however, to give an exact description of the time-dependence of $\mathbb{E}\left\|\boldsymbol{L}(t)\right\|^k$ for even exponents k and derive from that the asymptotic behaviour as t tends to zero.

Proposition 4.3 *Let k be a positive real number and $\boldsymbol{L}$ be a Lévy process.*

i) If k is an even integer and $\mathbb{E}\left\|\boldsymbol{L}(1)\right\|^k$ is finite, then there exist real numbers $m_1,\ldots,m_k$ such that

$$\mathbb{E}\left\|\boldsymbol{L}(t)\right\|^k = m_1 t + \ldots + m_k t^k, \quad t\geqslant 0. \qquad (4.2.9)$$

ii) If $\mathbb{E}\left\|\boldsymbol{L}(1)\right\|^{(k)_0}$ is finite, then $\mathbb{E}\left\|\boldsymbol{L}(h)\right\|^k = O(h^{k/(k)_0})$, as $h\to 0$.

Proof For the proof of i) we introduce the notation $\mathbb{K}\left(L^{i_1}(t),\ldots,L^{i_k}(t)\right)$, $1\leqslant i_1,\ldots,i_k\leqslant m$, for the mixed cumulants of $\boldsymbol{L}(t)$ of order k. They are defined in terms of the characteristic function of $\boldsymbol{L}$ as

$$\mathbb{K}\left(L^{i_1}(t),\ldots,L^{i_k}(t)\right) = \left.\frac{\partial^k}{\partial u_{i_1}\cdots\partial u_{i_k}}\log\mathbb{E}e^{\mathrm{i}\langle\boldsymbol{u},\boldsymbol{L}(t)\rangle}\right|_{\boldsymbol{u}=\boldsymbol{0}_m},$$

and are homogeneous functions of t of degree one, which is most easily seen from the Lévy–Khintchine formula. There is a close combinatoric relationship between moments and cumulants, which was used implicitly in the proof of Lemma 4.2, and which explicitly reads (see Shiryaev, 1996, §12, Theorem 6)

$$\begin{aligned}\mathbb{E}L^{i_1}(t)\cdot\cdots\cdot L^{i_k}(t) &= \sum_{\pi\in\mathcal{P}_k}\prod_{B\in\pi}\mathbb{K}\left(L^{i_j}(t): j\in B\right)\\ &= \sum_{\pi\in\mathcal{P}_k} t^{\#\pi}\prod_{B\in\pi}\mathbb{K}\left(L^{i_j}(1): j\in B\right) = \sum_{\kappa=1}^{k} m_{k,\kappa}^{i_1,\ldots,i_k} t^{\kappa},\end{aligned}$$

where

$$m_{k,\kappa}^{i_1,\ldots,i_k} = \sum_{\substack{\pi\in\mathcal{P}_k\\ \#\pi=\kappa}} \prod_{B\in\pi} \mathbb{K}\left(L^{i_j}(1) : j\in B\right).$$

Writing $k=2l$, the Multinomial Theorem implies that

$$\|\boldsymbol{L}(t)\|^k = \left[\left(L^1(t)\right)^2 + \ldots + (L^m(t))^2\right]^l = \sum_{\substack{0\leqslant l_1,\ldots,l_m\leqslant l\\ l_1+\ldots+l_m=l}} \frac{l!}{l_1!\cdot\ldots\cdot l_m!}\prod_{i=1}^m \left(L^i(t)\right)^{2l_i},$$

and thus it follows, by what was just shown and linearity of the expectation operator, that

$$\mathbb{E}\|\boldsymbol{L}(t)\|^k = \sum_{\kappa=1}^k \left[\sum_{\substack{0\leqslant l_1,\ldots,l_m\leqslant l\\ l_1+\ldots+l_m=l}} \frac{l!}{l_1!\cdot\ldots\cdot l_m!} m_{k,\kappa}^{\overbrace{1,\ldots,1}^{2l_1\text{ times}},\ldots,\overbrace{m,\ldots,m}^{2l_m\text{ times}}}\right] t^\kappa.$$

This proves Eq. (4.2.9). Assertion ii) follows for even k directly from the polynomial time-dependence of $\mathbb{E}\,\|\boldsymbol{L}(t)\|^k$ which we have just established. For general k, we use Hölder's inequality which implies that

$$\mathbb{E}\,\|\boldsymbol{L}(t)\|^k \leqslant \left(\mathbb{E}\,\|\boldsymbol{L}(t)\|^{(k)_0}\right)^{\frac{k}{(k)_0}},$$

and since $(k)_0$ is even by definition, the claim follows again from part i). □

4.2.3. A Fubini-type theorem for stochastic integrals with respect to Lévy processes

The next result is a Fubini-type theorem for a special class of stochastic integrals with respect to Lévy processes over an unbounded domain.

Theorem 4.4 (Fubini) *Let $[a,b]\subset\mathbb{R}$ be a bounded interval and $\boldsymbol{L}$ be a Lévy process with finite second moments. Assume that $F:[a,b]\times\mathbb{R}\to M_{d,m}(\mathbb{R})$ is a bounded function, and that the family $\{u\mapsto F(s,u)\}_{s\in[a,b]}$ is uniformly absolutely integrable and uniformly converges to zero as $|u|\to\infty$. It then holds that*

$$\int_a^b\int_{\mathbb{R}} F(s,u)\mathrm{d}\boldsymbol{L}(u)\mathrm{d}s = \int_{\mathbb{R}}\int_a^b F(s,u)\mathrm{d}s\mathrm{d}\boldsymbol{L}(u), \qquad (4.2.10)$$

almost surely.

Proof We first note that, since $\boldsymbol{L}$ has finite second moments and F is square-integrable, both integrals in Eq. (4.2.10) are well-defined as L^2-limits of approximating Riemann–Stieltjes sums. We start the proof by introducing the notations

$$I = \int_a^b \int_{\mathbb{R}} F(s,u)\mathrm{d}\boldsymbol{L}(u)\mathrm{d}s, \quad I_N = \int_a^b \int_{-N}^N F(s,u)\mathrm{d}\boldsymbol{L}(u)\mathrm{d}s,$$

as well as

$$\overleftrightarrow{I} = \int_{\mathbb{R}} \int_a^b F(s,u)\mathrm{d}s\mathrm{d}\boldsymbol{L}(u), \quad \overleftrightarrow{I_N} = \int_{-N}^N \int_a^b F(s,u)\mathrm{d}s\mathrm{d}\boldsymbol{L}(u).$$

It follows from Kailath, Segall and Zakai (1978, Theorem 1) (see also Protter, 1990, Theorem 64) that, for each N, $I_N = \overleftrightarrow{I_N}$ almost surely. We also write

$$\Delta_N \coloneqq I - I_N, \quad \overleftrightarrow{\Delta_N} \coloneqq \overleftrightarrow{I} - \overleftrightarrow{I_N}, \quad N > 0.$$

The strategy of the proof is to show that both Δ_N and $\overleftrightarrow{\Delta_N}$ converge to zero as N tends to infinity, and then to use the uniqueness of limits to conclude that I must equal $\overleftrightarrow{I}$. We first investigate $\mathbb{E}\left\|\Delta_N\right\|^2$. Clearly,

$$\Delta_N = \int_a^b \int_{|u|>N} F(s,u)\mathrm{d}\boldsymbol{L}(u)\mathrm{d}s. \tag{4.2.11}$$

Consequently, in order to analyse the absolute moments of Δ_N it suffices to consider the absolute moments of the infinite divisible random variables $\int_{|u|>N} F(s,u)\mathrm{d}\boldsymbol{L}(u)$, $s \in [a,b]$. By Eqs. (4.2.3), their characteristic triplets $(\gamma^s_{F,N}, \Sigma^s_{F,N}, \nu^s_{F,N})$ satisfy

$$\begin{aligned}\left\|\gamma^s_{F,N}\right\| &\leqslant \int_{|u|>N} \|F(s,u)\|\,\mathrm{d}u \left\|\gamma^{\boldsymbol{L}}\right\| + \int_{|u|>N} \|F(s,u)\| \int_{\|\boldsymbol{x}\|<1} \|\boldsymbol{x}\|\, I_{[1,\infty)}(\|F(s,u)\boldsymbol{x}\|)\nu^{\boldsymbol{L}}(\mathrm{d}\boldsymbol{x})\mathrm{d}u\\ &\quad + \int_{|u|>N} \|F(s,u)\| \int_{\|\boldsymbol{x}\|\geqslant 1} \|\boldsymbol{x}\|\, I_{[0,1]}(\|F(s,u)\boldsymbol{x}\|)\nu^{\boldsymbol{L}}(\mathrm{d}\boldsymbol{x})\mathrm{d}u\\ &\leqslant \int_{|u|>N} \|F(s,u)\|\,\mathrm{d}u \left[\left\|\gamma^{\boldsymbol{L}}\right\| + \int_{\|\boldsymbol{x}\|\geqslant 1} \|\boldsymbol{x}\|\,\nu^{\boldsymbol{L}}(\mathrm{d}\boldsymbol{x})\right],\end{aligned}$$

for all N exceeding some N_0 which satisfies $\|F(s,u)\| < 1$ for all $|u| > N_0$, $s \in [a,b]$; such an N_0 exists by assumption. Similarly, one obtains that

$$\left\|\Sigma^s_{F,N}\right\| \leqslant \left\|\Sigma^{\mathcal{G}}\right\| \int_{|u|>N} \|F(s,u)\|^2\,\mathrm{d}u \leqslant \left\|\Sigma^{\mathcal{G}}\right\| \int_{|u|>N} \|F(s,u)\|\,\mathrm{d}u, \quad \forall N > N_0,$$

and

$$\begin{aligned}\int_{\|\boldsymbol{x}\|<1} \|\boldsymbol{x}\|^2\,\nu^s_{F,N}(\mathrm{d}\boldsymbol{x}) &= \int_{|u|>N} \int_{\mathbb{R}^d} I_{[0,1]}(\|F(s,u)\boldsymbol{x}\|)\,\|F(s,u)\boldsymbol{x}\|^2\,\nu^{\boldsymbol{L}}(\mathrm{d}\boldsymbol{x})\mathrm{d}u\\ &\leqslant \int_{|u|>N} \|F(s,u)\|\,\mathrm{d}u \int_{\mathbb{R}^d} \|\boldsymbol{x}\|^2\,\nu^{\boldsymbol{L}}(\mathrm{d}\boldsymbol{x}), \quad \forall N > N_0,\end{aligned}$$

$$\begin{aligned}\int_{\|\boldsymbol{x}\|\geqslant 1}\|\boldsymbol{x}\|^r\nu^s_{F,N}(\mathrm{d}\boldsymbol{x}) &= \int_{|u|>N}\int_{\mathbb{R}^d} I_{[1,\infty)}(\|F(s,u)\boldsymbol{x}\|)\,\|F(s,(u)\boldsymbol{x}\|^r\,\nu^{\boldsymbol{L}}(\mathrm{d}\boldsymbol{x})\mathrm{d}u\\
&\leqslant \int_{|u|>N}\int_{\mathbb{R}^d} I_{[1,\infty]}\left(\|F\|_{L^\infty([a,b]\times\mathbb{R})}\|\boldsymbol{x}\|\right)\|F(s,u)\|^r\,\|\boldsymbol{x}\|^r\,\nu^{\boldsymbol{L}}(\mathrm{d}\boldsymbol{x})\mathrm{d}u\\
&\leqslant \int_{|u|>N}\|F(s,u)\|\,\mathrm{d}u\int_{\|\boldsymbol{x}\|\geqslant\max\{1,\|F\|^{-1}_{L^\infty([a,b]\times\mathbb{R})}\}}\|\boldsymbol{x}\|^2\,\nu^{\boldsymbol{L}}(\mathrm{d}\boldsymbol{x}),\quad r=1,2.\end{aligned}$$

Applying Lemma 4.2 with $k=2$ and using the assumed uniform absolute integrability of the family $\{u\mapsto F(s,u)\}_{s\in[a,b]}$, we can deduce that

$$\sup_{s\in[a,b]}\mathbb{E}\left\|\int_{|u|>N}F(s,u)\mathrm{d}\boldsymbol{L}(u)\right\|^2\to 0,\qquad\text{as } N\to\infty.$$

Together with Eq. (4.2.11) and Jensen's inequality, this implies that

$$\begin{aligned}\mathbb{E}\,\|\Delta_N\|^2 \leqslant&\mathbb{E}\left(\int_a^b\left\|\int_{|u|>N}F(s,u)\mathrm{d}\boldsymbol{L}(u)\right\|\mathrm{d}s\right)^2\\
\leqslant&\mathbb{E}\int_a^b\left\|\int_{|u|>N}F(s,u)\mathrm{d}\boldsymbol{L}(u)\right\|^2\mathrm{d}s\\
\leqslant&(b-a)\sup_{s\in[a,b]}\mathbb{E}\left\|\int_{|u|>N}F(s,u)\mathrm{d}\boldsymbol{L}(u)\right\|^2\to 0,\end{aligned}\tag{4.2.12}$$

as $N\to\infty$, showing that Δ_N converges to zero in L^2. In order to prove the same convergence also for

$$\overleftrightarrow{\Delta}_N=\overleftrightarrow{I}-\overleftrightarrow{I}_N=\int_{|u|>N}\int_a^b F(s,u)\mathrm{d}s\mathrm{d}\boldsymbol{L}(du),$$

we first define the function $\widetilde{F}:\mathbb{R}\to M_{d,m}(\mathbb{R})$ by $\widetilde{F}(u)=\int_a^b F(s,u)\mathrm{d}s$. Since, for all $u\in\mathbb{R}$, $\left\|\widetilde{F}(u)\right\|$ is smaller than $(b-a)\,\|F\|_{L^\infty([a,b]\times\mathbb{R})}$, the function $\widetilde{F}$ is bounded. It is also integrable because the normal variant of Fubini's theorem and the assumed uniform integrability of $\{F(s,\cdot)\}_{s\in[a,b]}$ imply that

$$\begin{aligned}\int_{|u|>N}\left\|\widetilde{F}(u)\right\|\mathrm{d}u\leqslant&\int_a^b\int_{|u|>N}\|F(s,u)\|\,\mathrm{d}u\mathrm{d}s\\
\leqslant&(b-a)\sup_{s\in[a,b]}\int_{|u|>N}\|F(s,u)\|\,\mathrm{d}u\to 0,\quad N\to\infty.\end{aligned}$$

Similar arguments to the ones given above then show that $\overleftrightarrow{\Delta}_N$ converges to zero in L^2 as well. It thus follows by the triangle inequality that, for every N and every ϵ,

$$\begin{aligned}\mathbb{P}\left(\left\|I-\overleftrightarrow{I}\right\|\geqslant\epsilon\right)\leqslant&\mathbb{P}\left(\left\{\|I-I_N\|\geqslant\frac{\epsilon}{2}\right\}\cup\left\{\left\|\overleftrightarrow{I}-I_N\right\|\geqslant\frac{\epsilon}{2}\right\}\right)\\
\leqslant&\mathbb{P}\left(\left\{\|I-I_N\|\geqslant\frac{\epsilon}{2}\right\}\right)+\mathbb{P}\left(\left\{\left\|\overleftrightarrow{I}-\overleftrightarrow{I}_N\right\|\geqslant\frac{\epsilon}{2}\right\}\right),\end{aligned}$$

where we have used the subadditivity of $\mathbb{P}$ as well as the fact that I_N is equal to $\overleftrightarrow{I_N}$ almost surely. Since L^2-convergence implies convergence in probability (Jacod and Protter, 2003, Theorems 17.2), it follows that the probability of the absolute difference between I and $\overleftrightarrow{I}$ exceeding ϵ is equal to zero for every positive ϵ. This means that I equals $\overleftrightarrow{I}$ almost surely, and completes the proof. □

4.3. Controller canonical form of multivariate CARMA processes

Multivariate, continuous-time autoregressive moving average (abbreviated MCARMA) processes are the continuous-time analogue of the well known vector ARMA processes. They also generalize the much-studied univariate CARMA processes to a multidimensional setting. A d-dimensional MCARMA process $\boldsymbol{Y}$, specified by an autoregressive polynomial

$$\tilde{P}(z) = z^{\tilde{p}} + \tilde{A}_1 z^{\tilde{p}-1} + \ldots + \tilde{A}_{\tilde{p}} \in M_d(\mathbb{R}[z]), \tag{4.3.1}$$

a moving average polynomial

$$\tilde{Q}(z) = \tilde{B}_0 + \tilde{B}_1 z + \ldots + \tilde{B}_{\tilde{q}} z^{\tilde{q}} \in M_{d,m}(\mathbb{R}[z]), \tag{4.3.2}$$

and driven by an m-dimensional Lévy process $\boldsymbol{L}$ is defined as a solution of the formal differential equation

$$\tilde{P}(\mathrm{D})\boldsymbol{Y}(t) = \tilde{Q}(\mathrm{D})\mathrm{D}\boldsymbol{L}(t), \quad \mathrm{D} = \frac{\mathrm{d}}{\mathrm{d}t}, \quad t \in \mathbb{R}, \tag{4.3.3}$$

the continuous-time version of the well-known ARMA equations. Equation (4.3.3) is only formal because, in general, the paths of a Lévy process are not differentiable. It has been shown in (Marquardt and Stelzer, 2007) that an MCARMA process $\boldsymbol{Y}$ can equivalently be defined by the continuous-time state space model

$$\mathrm{d}\boldsymbol{X}(t) = \mathcal{A}\boldsymbol{X}(t)\mathrm{d}t + \mathcal{B}\mathrm{d}\boldsymbol{L}(t), \quad \boldsymbol{Y}(t) = \mathcal{C}\boldsymbol{X}(t), \quad t \in \mathbb{R}, \tag{4.3.4}$$

where the matrices Ã, β and C are given by

$$\mathcal{A} = \begin{pmatrix} 0 & \mathbf{1}_d & 0 & \ldots & 0 \\ 0 & 0 & \mathbf{1}_d & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \ldots & \ldots & 0 & \mathbf{1}_d \\ -\tilde{A}_{\tilde{p}} & -\tilde{A}_{\tilde{p}-1} & \ldots & \ldots & -\tilde{A}_1 \end{pmatrix} \in M_{\tilde{p}d}(\mathbb{R}), \tag{4.3.5a}$$

$$\mathcal{B} = \begin{pmatrix} \beta_1^T & \cdots & \beta_{\tilde{p}}^T \end{pmatrix}^T \in M_{\tilde{p}d,m}(\mathbb{R}), \tag{4.3.5b}$$

where

$$\beta_{\tilde{p}-j} = -I_{\{0,\ldots,\tilde{q}\}}(j)\left[\sum_{i=1}^{\tilde{p}-j-1}\tilde{A}_i\beta_{\tilde{p}-j-i} - \tilde{B}_j\right],$$

and

$$\mathcal{C} = \begin{pmatrix} \mathbf{1}_d & 0_d & \ldots & 0_d \end{pmatrix} \in M_{d,\tilde{p}d}(\mathbb{R}). \tag{4.3.5c}$$

This is but one of several possible parametrizations of the general continuous-time state space model and is in the discrete-time literature often referred to as the *observer canonical form* (Kailath, 1980). For the purpose of estimating the driving Lévy process $\boldsymbol{L}$, it is more convenient to work with a different parametrization, which, in analogy to a canonical state space representation used in discrete-time control theory, might be called the *controller canonical form*. It is the multivariate generalization of the parametrization used for univariate CARMA processes in Brockwell et al. (2011). We first state an auxiliary lemma which we could not find in the literature.

Lemma 4.5 *Let r and s be positive integers. Assume that $R(z) = z^r + M_1 z^{r-1} + \ldots + M_r \in M_s(\mathbb{R}[z])$ is a matrix polynomial and denote by*

$$\boldsymbol{M} = \begin{bmatrix} 0 & \mathbf{1}_s & 0 & \cdots & 0 \\ 0 & 0 & \mathbf{1}_s & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \mathbf{1}_s \\ -M_r & -M_{r-1} & -M_{r-2} & \cdots & -M_1 \end{bmatrix} \in M_{rs}(\mathbb{R}) \tag{4.3.6}$$

the associated multi-companion matrix. The rational matrix function

$$S(z) = [S_{ij}(z)]_{1\leqslant i,j\leqslant r} = (z\mathbf{1}_{rs} - \boldsymbol{M})^{-1} \in M_{rs}(\mathbb{R}\{z\}), \quad S_{ij}(z) \in M_s(\mathbb{R}\{z\}), \tag{4.3.7}$$

is then given by the following formula for the block $S_{ij}(z)$:

$$S_{ij}(z) = R(z)^{-1}\begin{cases} z^{r-1+i-j}\mathbf{1}_s + \sum_{k=1}^{r-j} M_k z^{r-1-k+i-j}, & j \geqslant i, \\ -\sum_{k=r-j+1}^{r} M_k z^{r-1-k+i-j}, & j < i. \end{cases} \tag{4.3.8}$$

Proof We compute the (i,j)th block of $S(z)\,(z\mathbf{1}_{rs} - \boldsymbol{M})$. Assuming $i < j$, this block is given by

$$[S(z)\,(z\mathbf{1}_{rs} - \boldsymbol{M})]_{ij}$$

$$
\begin{aligned}
&= \sum_{k=1}^{r} S_{ik}(z)\,(z\mathbf{1}_{rs} - \boldsymbol{M})_{kj} \\
&= zS_{ij}(z) - S_{i,j-1}(z) + S_{ir}(z)M_{r-j+1} \\
&= R(z)^{-1}\left[z^{r+i-j}\mathbf{1}_s + \sum_{k=1}^{r-j} M_k z^{r-k+i-j} - z^{r+i-j}\mathbf{1}_s - \sum_{k=1}^{r-j+1} M_k z^{r-k+i-j} + z^{i-1}M_{r-j+1}\right] = 0.
\end{aligned}
$$

A similar calculation shows that for $i > j$, $[S(z)\,(z\mathbf{1}_{rs} - \boldsymbol{M})]_{ij} = 0$. For the blocks on the diagonal we obtain, for $i \geqslant 2$,

$$
\begin{aligned}
[S(z)\,(z\mathbf{1}_{rs} - \boldsymbol{M})]_{ii} &= \sum_{k=1}^{r} S_{ik}(z)\,(z\mathbf{1}_{rs} - \boldsymbol{M})_{ki} \\
&= zS_{ii}(z) - S_{i,i-1}(z) + S_{ir}(z)M_{r-i+1} \\
&= R(z)^{-1}\left[z^{r}\mathbf{1}_s + \sum_{k=1}^{r-i} M_k z^{r-k} + \sum_{k=r-i+2}^{r} M_k z^{r-k} + z^{i-1}M_{r-i+1}\right] = \mathbf{1}_s,
\end{aligned}
$$

and finally

$$
\begin{aligned}
[S(z)\,(z\mathbf{1}_{rs} - \boldsymbol{M})]_{11} = \sum_{k=1}^{r} S_{1k}(z)\,(z\mathbf{1}_{rs} - \boldsymbol{M})_{k1} &= zS_{11}(z) + S_{1r}(z)M_r \\
&= R(z)^{-1}\left[z^{r}\mathbf{1}_s + \sum_{k=1}^{r-1} M_k z^{r-k} + M_r\right] = \mathbf{1}_s.
\end{aligned}
$$

This shows that $S(z)$ is the inverse of $z\mathbf{1}_{rs} - \boldsymbol{M}$ and completes the proof. □

Theorem 4.6 (Controller form) *Assume that $\boldsymbol{L}$ is an $\mathbb{R}^m$-valued Lévy process, and that $\boldsymbol{Y}$ is a d-dimensional $\boldsymbol{L}$-driven MCARMA process with autoregressive polynomial $\tilde{P} \in M_d(\mathbb{R}[z])$ and moving average polynomial $\tilde{Q} \in M_{d,m}(\mathbb{R}[z])$. Then there exist integers $p > q > 0$ and matrix polynomials*

$$
z \mapsto P(z) = z^p + A_1 z^{p-1} + \ldots + A_p \in M_m(\mathbb{R}[z]), \tag{4.3.9a}
$$

$$
z \mapsto Q(z) = B_0 + B_1 z + \ldots + B_q z^q \in M_{d,m}(\mathbb{R}[z]) \tag{4.3.9b}
$$

satisfying $\tilde{P}(z)^{-1}\tilde{Q}(z) = Q(z)P(z)^{-1}$ for all $z \in \mathbb{C}$, and $\det P(z) = 0$ *if and only if* $\det \tilde{P}(z) = 0$. *Moreover, the process $\boldsymbol{Y}$ has the state space representation*

$$
\mathrm{d}\boldsymbol{X}(t) = \mathrm{A}\boldsymbol{X}(t)\mathrm{d}t + E_p\mathrm{d}\boldsymbol{L}(t), \quad t \in \mathbb{R}, \tag{4.3.10a}
$$

$$
\boldsymbol{Y}(t) = \underline{B}\boldsymbol{X}(t), \quad t \in \mathbb{R}, \tag{4.3.10b}
$$

where

$$\mathrm{A} = \begin{bmatrix} 0 & \mathbf{1}_m & 0 & \cdots & 0 \\ 0 & 0 & \mathbf{1}_m & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \mathbf{1}_m \\ -A_p & -A_{p-1} & -A_{p-2} & \cdots & -A_1 \end{bmatrix} \in M_{pm}(\mathbb{R}), \tag{4.3.11a}$$

$$E_p = \begin{bmatrix} 0 & 0 & \ldots & 0 & \mathbf{1}_m \end{bmatrix}^T \in M_{pm,m}(\mathbb{R}), \tag{4.3.11b}$$

and

$$\underline{B} = \begin{bmatrix} B_0 & B_1 & \cdots & B_{p-1} \end{bmatrix} \in M_{d,pm}(\mathbb{R}), \quad B_j = 0_{d,m}, \quad q+1 \leqslant j \leqslant p-1. \tag{4.3.11c}$$

Proof The existence of matrix polynomials $P \in M_m(\mathbb{R}[z])$ and $Q \in M_{m,d}(\mathbb{R}[z])$ with the asserted properties has been shown in Kailath (1980, Lemma 6.3-8). In order to prove Eqs. (4.3.10), it suffices, by Theorem 2.4, to prove that the triple $(\mathrm{A}, E_p, \underline{B})$, defined in Eqs. (4.3.11), is a realization of the right matrix fraction QP^{-1}, that is

$$\underline{B}\left[z\mathbf{1}_{pm} - \mathrm{A}\right]^{-1} E_p = Q(z)P(z)^{-1}, \quad \forall z \in \mathbb{C}.$$

Using Lemma 4.5 and the fact that right multiplication by E_p selects the last block-column, one sees that

$$\left[z\mathbf{1}_{pm} - \mathrm{A}\right]^{-1} E_p = \begin{bmatrix} 1 & z & \cdots & z^{p-1} \end{bmatrix}^T \otimes P(z)^{-1},$$

where $\otimes$ denotes the Kronecker product of two matrices. By definition it holds that

$$\underline{B}\begin{bmatrix} 1 & z & \cdots & z^{p-1} \end{bmatrix}^T = B_0 + B_1 z + \ldots + B_q z^q = Q(z),$$

and so the claim follows. □

In view of Theorem 4.6 one can assume without loss of generality that an MCARMA process $\boldsymbol{Y}$ is given by a state space representation (4.3.10) with coefficient matrices of the form (4.3.11). We make the following assumptions about the zeros of the polynomials P, Q in equations (4.3.9). The first one is a stability assumption guaranteeing the existence of a stationary solution of the state equation (4.3.10a).

Assumption A1 The zeros of the polynomial $\det P(z) \in \mathbb{R}[z]$ have strictly negative real parts.

The second assumption corresponds to the minimum-phase assumption in classical time series analysis. For a matrix $M \in M_{d,m}(\mathbb{R})$, any matrix $M^{\sim 1}$ satisfying $M^{\sim 1}M = \mathbf{1}_m$ is called a *left inverse* of M. It is easy to check that the existence of a left inverse of M is equivalent to the conditions $m \leqslant d$, $\operatorname{rank} M = m$, and that in this case $M^{\sim 1}$ can be computed as $M^{\sim 1} = (M^T M)^{-1} M^T$.

Assumption A2 The dimension m of the driving Lévy process $\boldsymbol{L}$ is smaller than or equal to the dimension of the multivariate CARMA process $\boldsymbol{Y}$, and both B_q and $B_q^T B_0$ have full rank m. The zeros of the polynomial $\det B_q^{\sim 1} Q(z) \in \mathbb{R}[z]$ have strictly negative real parts.

It is well known that every solution of Eq. (4.3.10a) satisfies

$$\boldsymbol{X}(t) = \mathrm{e}^{\mathrm{A}(t-s)}\boldsymbol{X}(s) + \int_s^t \mathrm{e}^{\mathrm{A}(t-u)} E_p \mathrm{d}\boldsymbol{L}(u), \quad s, t \in \mathbb{R}, \quad s < t.$$

Under Assumption A1, the state equation (4.3.10a) has a unique strictly stationary, causal solution given by

$$\boldsymbol{X}(t) = \int_{-\infty}^t \mathrm{e}^{\mathrm{A}(t-u)} E_p \mathrm{d}\boldsymbol{L}(u), \quad t \in \mathbb{R}. \tag{4.3.12}$$

and consequently, the multivariate CARMA process $\boldsymbol{Y}$ has the moving average representation

$$\boldsymbol{Y}(t) = \int_{-\infty}^{\infty} g(t-u) \mathrm{d}\boldsymbol{L}(u), \quad t \in \mathbb{R}; \quad g(t) = \underline{B}\mathrm{e}^{\mathrm{A}t} E_p I_{[0,\infty]}(t). \tag{4.3.13}$$

In the next section we will express the increments of the driving Lévy process $\boldsymbol{L}$ in terms of the multivariate CARMA process $\boldsymbol{Y}$. In particular, we will need to know that the paths of $\boldsymbol{Y}$ and also of the state process $\boldsymbol{X}$ are sufficiently often differentiable. We recall that we denote by $X^i(t)$ the ith component of the vector $\boldsymbol{X}(t)$, and we define, for $j = 1, \ldots, p$, the jth *m-block* of $\boldsymbol{X}$ by the formula

$$\boldsymbol{X}^{(j)}(t) = \begin{bmatrix} X^{(j-1)m+1}(t)^T & \cdots & X^{jm}(t)^T \end{bmatrix}^T, \quad t \in \mathbb{R}. \tag{4.3.14}$$

A very useful property, which the sequence of approximation errors $\left(\Delta\boldsymbol{L}_n - \widehat{\Delta\boldsymbol{L}}_n\right)_{n\in\mathbb{N}}$ might enjoy, is asymptotic independence, which heuristically means, that $\Delta\boldsymbol{L}_n - \widehat{\Delta\boldsymbol{L}}_n$ and $\Delta\boldsymbol{L}_m - \widehat{\Delta\boldsymbol{L}}_m$ are almost independent if $|n-m| \gg 1$. One possibility of making this concept precise is to introduce the notion of strong (or α-) mixing, which has first been defined in Rosenblatt (1956). Since then it has turned out to be a very powerful tool for establishing asymptotic results in the theory of inference for stochastic processes. For a stationary stochastic process $X = (X_t)_{t\in I}$, where I is either $\mathbb{R}$ or $\mathbb{Z}$, we first introduce the σ-algebras $\mathscr{F}_n^m = \sigma(X_j : j \in I, n < j < m)$, where $-\infty \leqslant n < m \leqslant \infty$. For $m \in I$, the strong mixing coefficient $\alpha(m)$ is defined as

$$\alpha(m) = \sup_{A \in \mathscr{F}_{-\infty}^0, B \in \mathscr{F}_m^\infty} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|. \tag{4.3.15}$$

The process X is called *strongly mixing* if $\lim_{m\to\infty}\alpha(m) = 0$; if $\alpha(m) = O(\lambda^m)$ for some $0 < \lambda < 1$, it is called *exponentially strongly mixing*.

Lemma 4.7 *Assume that $\boldsymbol{L}$ is a Lévy process, and that $\boldsymbol{Y}$ is an $\boldsymbol{L}$-driven multivariate CARMA process given by the state space representation (4.3.10) and satisfying Assumption A1. Then the following hold.*

(i) The process $\boldsymbol{Y}$ is strictly stationary.

(ii) The paths of $\boldsymbol{Y}$ are $p-q-1$ times differentiable. Moreover, for $j = 1,\ldots,p$, the paths of the jth m-block of the state process $\boldsymbol{X}$ are $p-j$ times differentiable.

(iii) For any $k > 0$ and any $t, \in \mathbb{R}$, finiteness of $\mathbb{E}\,\|\boldsymbol{L}(1)\|^k$ implies finiteness of both $\mathbb{E}\,\|\boldsymbol{X}(t)\|^k$ and $\mathbb{E}\,\|\boldsymbol{Y}(t)\|^k$. Conversely, finiteness of the kth moment of $\boldsymbol{X}(t)$ implies finiteness of $\boldsymbol{L}(1)$.

(iv) If $\mathbb{E}\,\|\boldsymbol{L}(1)\|^k$ is finite for some $k > 0$, then the process $\boldsymbol{Y}$ is strongly mixing with exponentially decaying mixing coefficients.

Proof The first claim is an immediate consequence of the moving average representation (4.3.13). Part ii) follows from Marquardt and Stelzer (2007, Proposition 3.30) and the observation that E_p is injective. The assertion iv) follows from Masuda (2004, Theorem 4.3), see also the proof of Marquardt and Stelzer (2007, Proposition 3.34). □

The following lemma relates strong mixing of a continuous-time process to strong mixing of functionals of the process.

Lemma 4.8 *Let $X = (X_t)_{t\in\mathbb{R}}$ be an $\mathbb{R}^d$-valued (exponentially) strongly mixing stochastic process. If, for each $n \in \mathbb{Z}$, the random variable Y_n is measurable with respect to $\sigma(X_t : n-1 \leqslant t \leqslant n)$ then the stochastic process $(Y_n)_{n\in\mathbb{Z}}$ is (exponentially) strongly mixing. In particular, if $f : \mathbb{R}^{d\times[0,1]} \to \mathbb{R}^m$ is a measurable function, then the $\mathbb{R}^m$-valued stochastic process $\left(f((X_{n-1+t})_{t\in[0,1]})\right)_{n\in\mathbb{Z}}$ is (exponentially) strongly mixing.*

Proof This follows immediately from Eq. (4.3.15), the definition of the strong mixing coefficients. □

4.4. Recovery of the driving Lévy process from continuous-time observations

In this section we address the problem of recovering the driving Lévy process of a multivariate CARMA process given by a state space representation (4.3.10), if continuous-time observations are available. We assume that the order (p,q) as well as the coefficient matrices A and $\underline{B}$ are known. If they are not, they can first be estimated by, e. g. maximization of the

Gaussian likelihood, although the precise statistical properties of this two-step estimator are beyond the scope of the present work. More precisely, we show that, conditional on the value $\boldsymbol{X}(0)$ of the state vector at time zero, one can write the value of $\boldsymbol{L}(t)$, for any $t \in [0, T]$, as a function of the continuous-time record $(\boldsymbol{Y}(t) : 0 \leqslant t \leqslant T)$. In particular, one can obtain an i. i. d. sample from the distribution of the unit increments $\boldsymbol{L}(n) - \boldsymbol{L}(n-1)$, $1 \leqslant n \leqslant T$, which, when subjected to one of several well-established estimation procedures, can be used to estimate a parametric model for $\boldsymbol{L}$. It can be argued that most of the time a continuous record of observations is not available. The results of this section will, however, serve as the starting point for the recovery of an approximate sample from the unit increment distribution based on discrete-time observation of $\boldsymbol{Y}$, which is presented in Section 4.5. The strategy is to first express the state vector $\boldsymbol{X}$ in terms of the observations $\boldsymbol{Y}$, and then to invert the state equation (4.3.10a) to obtain the driving Lévy process as a function of the state vector. We first define the *upper q-block truncation* of $\boldsymbol{X}$, denoted by $\boldsymbol{X}_q$, as

$$\boldsymbol{X}_q(t) = \begin{bmatrix} \boldsymbol{X}^{(1)}(t)^T & \cdots & \boldsymbol{X}^{(q)}(t)^T \end{bmatrix}^T, \quad t \in \mathbb{R},$$

where the m-blocks $\boldsymbol{X}^{(j)}$ have been defined in Eq. (4.3.14).

Lemma 4.9 *Assume that $\boldsymbol{L}$ is a Lévy process, and that $\boldsymbol{Y}$ is a multivariate CARMA process given as the solution of the state space equations (4.3.10). If Assumption A2 holds, the truncated state vector $\boldsymbol{X}_q$ satisfies the stochastic differential equation*

$$\mathrm{d}\boldsymbol{X}_q(t) = \mathrm{B}\boldsymbol{X}_q(t)\mathrm{d}t + E_q\boldsymbol{Y}(t)\mathrm{d}t, \quad t \in \mathbb{R}, \tag{4.4.1}$$

where

$$\mathrm{B} = \begin{bmatrix} 0 & \mathbf{1}_m & 0 & \cdots & 0 \\ 0 & 0 & \mathbf{1}_m & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & & \mathbf{1}_m \\ -B_q^{\sim 1}B_0 & -B_q^{\sim 1}B_1 & -B_q^{\sim 1}B_2 & \cdots & -B_q^{\sim 1}B_{q-1} \end{bmatrix} \in M_{mq}(\mathbb{R}), \tag{4.4.2a}$$

$$E_q = \begin{bmatrix} 0 & 0 & \cdots & 0 & \left(B_q^{\sim 1}\right)^T \end{bmatrix}^T \in M_{mq,d}(\mathbb{R}), \tag{4.4.2b}$$

and $B_q^{\sim 1}$ denotes the left inverse of B_q. Moreover, the eigenvalues of the matrix B *have strictly negative real parts.*

Proof Equation (4.4.1) follows easily from combining the first q block-rows of the state transition equation (4.3.10a) with the observation equation (4.3.10b). The assertion about the eigenvalues of B is a consequence of the well-known correspondence between the eigenvalues of a multi-companion matrix and the zeros of the associated polynomial (see,

e. g., Marquardt and Stelzer, 2007, Lemma 3.8). By this correspondence, the eigenvalues of B are exactly the zeros of the polynomial function

$$z \mapsto \det\left(\mathbf{1}_m z^q + B_q^{\sim 1} B_{q-1} z^{q-1} + \ldots + B_q^{\sim 1} B_0\right),$$

whose zeros have strictly negative real parts by Assumption A2. □

As before, we see that Eq. (4.4.1) is readily integrated to

$$\boldsymbol{X}_q(t) = \mathrm{e}^{\mathrm{B}(t-s)}\boldsymbol{X}_q(s) + \int_s^t \mathrm{e}^{\mathrm{B}(t-u)} E_q \boldsymbol{Y}(u)\mathrm{d}u, \quad s,t \in \mathbb{R}, \quad s < t. \tag{4.4.3}$$

The remaining blocks $\boldsymbol{X}^{(i)}$, $q < i \leqslant p$, are obtained from $\boldsymbol{X}_q$ and $\boldsymbol{Y}$ by differentiation. The existence of the occurring derivatives of the state process $\boldsymbol{X}$ and the MCARMA process $\boldsymbol{Y}$ is guaranteed by Lemma 4.7.

Lemma 4.10 *For $1 \leqslant n \leqslant p-q$, the block $\boldsymbol{X}^{(q+n)}$ is given by*

$$\boldsymbol{X}^{(q+n)}(t) = E_q^T\left[\mathrm{B}^n \boldsymbol{X}_q(t) + \sum_{\nu=0}^{n-1} \mathrm{B}^{n-1-\nu} E_q \mathrm{D}^\nu \boldsymbol{Y}(t)\right], \quad t \in \mathbb{R}. \tag{4.4.4}$$

Proof We first observe that Eqs. (4.3.10) and Eq. (4.4.1) imply that

$$\boldsymbol{X}^{(q+n)}(t) = \mathrm{D}\boldsymbol{X}^{(q+n-1)}(t), \qquad \mathrm{D}\boldsymbol{X}_q(t) = \mathrm{B}\boldsymbol{X}_q(t) + E_q\boldsymbol{Y}(t).$$

Therefore, the claim is true for $n = 1$. Assuming it is true for some $1 < n < p-q$, it follows that

$$\begin{aligned}
\boldsymbol{X}^{(q+n+1)}(t) =& \mathrm{D}\boldsymbol{X}^{(q+n)}(t)\\
=& \mathrm{D}E_q^T\left[\mathrm{B}^n \boldsymbol{X}_q(t) + \sum_{\nu=0}^{n-1} \mathrm{B}^{n-1-\nu} E_q \mathrm{D}^\nu \boldsymbol{Y}(t)\right]\\
=& E_q^T\left[\mathrm{B}^{n+1} \boldsymbol{X}_q(t) + \mathrm{B}^n E_q \boldsymbol{Y}(t) + \sum_{\nu=0}^{n-1} \mathrm{B}^{n-1-\nu} E_q \mathrm{D}^{\nu+1} \boldsymbol{Y}(t)\right]\\
=& E_q^T\left[\mathrm{B}^{n+1} \boldsymbol{X}_q(t) + \sum_{\nu=0}^{n} \mathrm{B}^{n-\nu} E_q \mathrm{D}^\nu \boldsymbol{Y}(t)\right].
\end{aligned}$$

□

Equations (4.4.3) and (4.4.4) allow to compute the value of $\boldsymbol{X}(t)$ based on the knowledge of the initial value $\boldsymbol{X}(0)$ and the continuous-time record $\{\boldsymbol{Y}(s) : 0 \leqslant s \leqslant t\}$. In order to obtain the value of $\boldsymbol{L}(t)$, we integrate the last block-row of the state transition equation (4.3.10a) to obtain

$$\boldsymbol{L}(t) = \boldsymbol{X}^{(p)}(t) - \boldsymbol{X}^{(p)}(0) + \underline{A}\int_0^t \boldsymbol{X}(s)\mathrm{d}s, \tag{4.4.5}$$

where $\underline{A} := \left[\begin{array}{ccc} A_p & \dots & A_1 \end{array}\right]$. We also write $\underline{A}_q := \left[\begin{array}{ccc} A_p & \dots & A_{p-q+1} \end{array}\right]$.

Theorem 4.11 (Recovery of $\boldsymbol{\Delta L_n}$) *Let $\boldsymbol{Y}$ be the multivariate CARMA process defined by the state space representation (4.3.10) and assume that Assumption A2 holds. The increment $\Delta \boldsymbol{L}_n = \boldsymbol{L}(n) - \boldsymbol{L}(n-1)$ is then given by*

$$\begin{aligned}\Delta \boldsymbol{L}_n = &\sum_{\nu=0}^{p-q-1}\left[E_q^T \mathrm{B}^{p-q-1-\nu}E_q + \sum_{k=\nu}^{p-q-2} A_{p-q-k-1}E_q^T\mathrm{B}^{k-\nu}E_q\right]\left[\mathrm{D}^\nu \boldsymbol{Y}(n) - \mathrm{D}^\nu \boldsymbol{Y}(n-1)\right]\\ &+\left[\underline{A}_q\mathrm{B}^{-1} + \sum_{k=1}^{p-q}A_{p-q-k+1}E_q^T\mathrm{B}^{k-1} + E_q^T\mathrm{B}^{p-q}\right]\left[\boldsymbol{X}_q(n) - \boldsymbol{X}_q(n-1)\right]\\ &+A_p\left[B_q^{\sim 1}B_0\right]^{-1}B_q^{\sim 1}\int_{n-1}^n \boldsymbol{Y}(s)\mathrm{d}s, \end{aligned}\tag{4.4.6}$$

and

$$\boldsymbol{X}_q(n) = \mathrm{e}^{\mathrm{B}}\boldsymbol{X}_q(n-1) + \int_{n-1}^n \mathrm{e}^{\mathrm{B}(n-u)}E_q\boldsymbol{Y}(u)\mathrm{d}u, \quad n \geqslant 1. \tag{4.4.7}$$

Proof Substituting Eq. (4.4.4) into Eq. (4.4.5) leads to

$$\begin{aligned}\Delta \boldsymbol{L}_n = &\sum_{\nu=0}^{p-q-1}\left[E_q^T \mathrm{B}^{p-q-1-\nu} + \sum_{k=\nu}^{p-q-2} A_{p-q-k-1}E_q^T\mathrm{B}^{k-\nu}\right]E_q\left[\mathrm{D}^\nu \boldsymbol{Y}(n) - \mathrm{D}^\nu \boldsymbol{Y}(n-1)\right]\\ &+E_q^T\mathrm{B}^{p-q}\left[\boldsymbol{X}_q(n) - \boldsymbol{X}_q(n-1)\right] + \left[\underline{A}_q + \sum_{k=1}^{p-q}A_{p-q-k+1}E_q^T\mathrm{B}^{k}\right]\int_{n-1}^n \boldsymbol{X}_q(s)\mathrm{d}s\\ &+\sum_{k=1}^{p-q}A_{p-q-k+1}E_q^T\mathrm{B}^{k-1}E_q\int_{n-1}^n \boldsymbol{Y}(s)\mathrm{d}s. \end{aligned}$$

Assumption A2 implies that $B_q^{\sim 1}B_0$ is invertible and, by Lemma 4.5, the matrix B is invertible as well. Thus, integration of Eq. (4.4.1) shows that

$$\int_{n-1}^n \boldsymbol{X}_q(s)\mathrm{d}s = \mathrm{B}^{-1}\left[\boldsymbol{X}_q(n) - \boldsymbol{X}_q(n-1) - E_q\int_{n-1}^n \boldsymbol{Y}(s)\mathrm{d}s\right].$$

Inserting this equation into the last expression for $\Delta \boldsymbol{L}_n$ and using the equality $\underline{A}_q\mathrm{B}^{-1}E_q = A_p\left[B_q^{\sim 1}B_0\right]^{-1}B_q^{\sim 1}$ proves Eq. (4.4.6). Equation (4.4.7) follows from setting $t = n$, $s = n-1$ in Eq. (4.4.3). □

In order to keep the notation simple, we restrict our attention to unit increments $\Delta \boldsymbol{L}$. In all our arguments and results, $\Delta \boldsymbol{L}_n$ can be replaced by $\Delta_\delta \boldsymbol{L}_n := \boldsymbol{L}(n\delta) - \boldsymbol{L}((n-1)\delta)$ for some $\delta > 0$.

4.5. Approximate recovery of the driving Lévy process from discrete-time observations

In this section we consider the question of how to obtain estimates of the increments $\Delta \boldsymbol{L}_n$ of the driving Lévy process based on a discrete-time record of the multivariate CARMA process $\boldsymbol{Y}$. The starting point is Eq. (4.4.6), which expresses the increment $\Delta \boldsymbol{L}_n$ in terms of derivatives and integrals of $\boldsymbol{Y}$. In order to approximate $\Delta \boldsymbol{L}_n$, it is therefore necessary to approximate these derivatives and integrals. We always assume that values of $\boldsymbol{Y}$ are available at the discrete times $(0, h, 2h, \ldots)$ only. For notational convenience, we also assume that $h^{-1} \in \mathbb{N}$; our results continue to hold with minor modifications if this restriction is dropped.

4.5.1. Approximation of derivatives

Throughout, we will approximate derivatives by so-called forward differences, which can be interpreted as iterated difference quotients. For a general introduction to finite difference approximations, see LeVeque (2007, Chapter 1). For any function f and any positive integer ν, we define

$$\Delta_h^\nu[f](t) := \frac{1}{h^\nu} \sum_{i=0}^{\nu} (-1)^{\nu-i} \binom{\nu}{i} f(t+ih). \tag{4.5.1}$$

It is apparent from this formula that knowledge of f on the discrete time grid $(0, h, \ldots, T)$ is sufficient to compute $\Delta_h^\nu[f](t)$ for any $t \in [0, T - \nu h] \cap h\mathbb{Z}$. The following lemma collects some useful properties of forward differences; in particular, it shows that if the function f is sufficiently smooth, then the derivative $\mathrm{D}^\nu f(t)$ is well approximated by $\Delta_h^\nu[f](t)$.

Lemma 4.12 *For $h > 0$ and a positive integer ν, let the forward differences $\Delta_h^\nu[f](t)$, $t \in \mathbb{R}$, be defined by Eq. (4.5.1). The following properties hold:*

i) For every positive integer $k < \nu$ and every function f, one has $\Delta_h^\nu[f] = \Delta_h^k\left[\Delta_h^{\nu-k}[f](\cdot)\right]$.

ii) If the function $f : \mathbb{R} \to \mathbb{R}^m$ is $\nu + 1$ times continuously differentiable on the interval $[t, t+\nu h]$, then there exist $t_i^* \in [t, t+\nu h]$, $i = 1, \ldots, m$, such that

$$\Delta_h^\nu[f](t) = \mathrm{D}^\nu f(t) - \frac{h}{2} \mathrm{D}^{\nu+1} f(\underline{t}^*), \tag{4.5.2}$$

where $\mathrm{D}^{\nu+1} f(\underline{t}^*)$ is the vector whose ith component equals the ith component of $\mathrm{D}^{\nu+1} f(t_i^*)$. In particular, for every polynomial $\mathfrak{p}$ of degree at most ν, one has $\Delta_h^\nu[\mathfrak{p}] = \mathrm{D}^\nu \mathfrak{p}$.

iii) If the $(\nu+1)$th derivative of f is not assumed to be continuous, it holds that

$$\|\Delta_h^\nu[f](t) - \mathrm{D}^\nu f(t)\| \leqslant h \sup_{s \in [t, t+\nu h]} \left\| \mathrm{D}^{\nu+1} f(s) \right\|. \tag{4.5.3}$$

Proof Property i) is immediate from the definition (4.5.1). The assertions of ii) and iii) follow from a component-wise application of Taylor's theorem (Apostol, 1974, Theorem 5.19). □

In the next lemma we will show that the supremum of an Ornstein–Uhlenbeck-type process has finite absolute kth moments if and only if the driving Lévy process has finite kth moments. This will allow us to effectively employ the error bound (4.5.3) for multivariate CARMA processes.

Lemma 4.13 *Let $(\boldsymbol{L}(t))_{t\geqslant 0}$ be an m-variate Lévy process, and let $A \in M_N(\mathbb{R})$, $B \in M_{N,m}(\mathbb{R})$ be given coefficient matrices. Assume that all eigenvalues of A have strictly negative real parts, and that $\boldsymbol{X} = (\boldsymbol{X}(t))_{t\geqslant 0}$ is the unique stationary solution of the stochastic differential equation*

$$\mathrm{d}\boldsymbol{X}(t) = A\boldsymbol{X}(t)\mathrm{d}t + B\mathrm{d}\boldsymbol{L}(t), \quad t \in \mathbb{R}. \tag{4.5.4}$$

Further denote by

$$\boldsymbol{X}^*(t) = \sup_{0\leqslant s\leqslant t} \|\boldsymbol{X}(s)\| \tag{4.5.5}$$

the supremum of $\|\boldsymbol{X}\|$ on the compact interval $[0,t]$. It then holds that, for every $t \in \mathbb{R}$ and every $k > 0$, the kth moment $\mathbb{E}\left(\boldsymbol{X}^*(t)\right)^k$ is finite if and only if $\mathbb{E}\left\|\boldsymbol{L}(1)\right\|^k$ is finite.

Proof If $\mathbb{E}\left\|\boldsymbol{L}(1)\right\|^k$ is infinite, it follows from Lemma 4.7, iii) that $\mathbb{E}\left\|\boldsymbol{X}(t)\right\|^k$ is infinite as well for every $t \in \mathbb{R}$, and that therefore $\mathbb{E}\left(\boldsymbol{X}^*(t)\right)^k$ must be infinite. The other implication requires more work.

We first note that $\boldsymbol{X}^*(t) \leqslant \sum_{i=1}^N X_i^*(t)$, where $X_i^*(t) = \sup_{0\leqslant s\leqslant t}\left|X^i(s)\right|$ is the supremum of the ith component of $\boldsymbol{X}$ over the interval $[0,t]$. Since each X^i is a semi-martingale, Protter (1990, Theorem V.2) shows that there exists a universal constant c_k such that $\mathbb{E}\left(X_i^*(t)\right)^k \leqslant c_k\left\|X^i\right\|_{\mathscr{H}_t^k}$, where the norm $\|\cdot\|_{\mathscr{H}_t^k}$ is defined by

$$\left\|X^i\right\|_{\mathscr{H}_t^k} = \inf_{X^i=\widetilde{V}_i+\widetilde{M}_i} \mathbb{E}\left(\int_0^t\left|\mathrm{d}\widetilde{V}_i(s)\right| + [\widetilde{M}_i,\widetilde{M}_i]_t^{1/2}\right)^k.$$

Here, the infimum is taken over all decompositions of X^i into a local martingale $\widetilde{M}_i$ and an adapted, càdlàg process $\widetilde{V}_i$ with finite variation and $[\cdot,\cdot]$ denotes the quadratic variation process. In our situation, Eq. (4.5.4) defines a canonical decomposition of X^i, $i = 1,\ldots,N$, into the finite variation process $V_i = (V_i(t))_{t\geqslant 0}$ given by

$$V_i(t) = \boldsymbol{e}_i^T\left[\boldsymbol{X}(0) + \int_0^t A\boldsymbol{X}(s)\mathrm{d}s + tB\mathbb{E}\boldsymbol{L}(1)\right],$$

where $\boldsymbol{e}_i$ denotes the ith unit vector in $\mathbb{R}^N$, and the martingale $M_i = (M_i(t))_{t\geqslant 0}$ given by

$$M_i(t) = \boldsymbol{e}_i^T B\left[\boldsymbol{L}(t) - t\mathbb{E}\boldsymbol{L}(1)\right].$$

Since clearly,

$$
\begin{aligned}
(\boldsymbol{X}^*(t))^k = \sup_{0\leqslant s\leqslant t}\left(X^1(s)^2+\ldots+X^N(s)^2\right)^{k/2} &\leqslant \left(X_1^*(t)^2+\ldots+X_N^*(t)^2\right)^{k/2}\\
&\leqslant N^{k/2}\max_{1\leqslant i\leqslant N} X_i^*(t)^k\\
&\leqslant N^{k/2}\sum_{i=1}^N X_i^*(t)^k,
\end{aligned}
$$

it suffices to bound the kth moments of $X_i^*(t)$ in order to obtain a bound for the kth moment of $\boldsymbol{X}^*(t)$. The former can be estimated as

$$
\mathbb{E}\left(X_i^*(t)\right)^k \leqslant c_k \left\|X^i\right\|_{\mathscr{H}_t^k} \leqslant c_k\left[\mathbb{E}\left(\int_0^t |\mathrm{d}V_i(s)|\right)^k + \mathbb{E}[M_i,M_i]_t^{k/2}\right]. \tag{4.5.6}
$$

The first term in this expression is seen to satisfy

$$
\begin{aligned}
\mathbb{E}\left(\int_0^t |\mathrm{d}V_i(s)|\right)^k \leqslant& \mathbb{E}\left(\int_0^t \left|\boldsymbol{e}_i^T A\boldsymbol{X}(s)\right|\mathrm{d}s + t\left|\boldsymbol{e}_i^T B\boldsymbol{L}(1)\right|\right)^k\\
\leqslant& 2^k\left[\|A\|^k\int_0^t \mathbb{E}\left\|\boldsymbol{X}(s)\right\|^k\mathrm{d}s + t^k\|B\|^k\,\mathbb{E}\left\|\boldsymbol{L}(1)\right\|^k\right] < \infty,
\end{aligned}
$$

where the finiteness of the integral $\int_0^t \mathbb{E}\left\|\boldsymbol{X}(s)\right\|^k \mathrm{d}s$ follows from the assumption that the absolute moment $\mathbb{E}\left\|\boldsymbol{X}(s)\right\|^k$ is finite and the strict stationarity of $\boldsymbol{X}$. For the second term in Eq. (4.5.6) one obtains the bound

$$
\begin{aligned}
\mathbb{E}[M_i,M_i]_t^{k/2} = \mathbb{E}\left(\boldsymbol{e}_i^T B[\boldsymbol{L},\boldsymbol{L}]_t B^T\boldsymbol{e}_i\right)^{k/2} \leqslant& \|B\|^k\,\mathbb{E}\left\|[\boldsymbol{L},\boldsymbol{L}]_t\right\|^{k/2}\\
\leqslant& 2^k\|B\|^k\left\{\left\|\Sigma^{\mathcal{G}}\right\|^{k/2}t^{k/2} + \mathbb{E}\left\|\int_0^t\int_{\mathbb{R}^m}\boldsymbol{x}\boldsymbol{x}^T N(\mathrm{d}s,\mathrm{d}\boldsymbol{x})\right\|^{k/2}\right\},
\end{aligned}
$$

where we have used Jacod and Shiryaev (2003, Theorem I.4.52) to compute the quadratic variation of the Lévy process $\boldsymbol{L}$ with characteristic triplet $(\gamma^{\boldsymbol{L}},\Sigma^{\mathcal{G}},\nu^{\boldsymbol{L}})$. To see that this expression is finite, we observe that

$$
\begin{aligned}
\mathbb{E}\left\|\int_0^t\int_{\mathbb{R}^m}\boldsymbol{x}\boldsymbol{x}^T N(\mathrm{d}s,\mathrm{d}\boldsymbol{x})\right\|^{k/2} \leqslant& m^{k/2}\mathbb{E}\left(\int_0^t\int_{\mathbb{R}^m}\|\boldsymbol{x}\|^2 N(\mathrm{d}s,\mathrm{d}\boldsymbol{x})\right)^{k/2}\\
=& m^{k/2}\mathbb{E}\left(\lim_{\epsilon\to 0}\int_0^t\int_{\|\boldsymbol{x}\|\geqslant\epsilon}\|\boldsymbol{x}\|^2 N(\mathrm{d}s,\mathrm{d}\boldsymbol{x})\right)^{k/2}\\
=& m^{k/2}\lim_{\epsilon\to 0}\mathbb{E}\left(\int_0^t\int_{\|\boldsymbol{x}\|\geqslant\epsilon}\|\boldsymbol{x}\|^2 N(\mathrm{d}s,\mathrm{d}\boldsymbol{x})\right)^{k/2} =: m^{k/2}\lim_{\epsilon\to 0}\mathbb{E}Y_\epsilon^{k/2},
\end{aligned}
$$

where we have applied the Monotone Convergence Theorem (Klenke, 2008, Theorem 4.20) to interchange the order of expectation and passing to the limit. By Sato (1999, Proposition 19.5), for each $\epsilon > 0$, the random variable $Y_\epsilon = \int_0^t \int_{\|x\|\geqslant\epsilon} \|x\|^2 N(\mathrm{d}s, \mathrm{d}x)$ is infinitely divisible with characteristic measure $\rho_\epsilon = (\mathrm{Leb}\,|_{[0,t]} \otimes \nu^L|_{\{\|x\|\geqslant\epsilon\}})\phi_\epsilon^{-1}$, where $\phi_\epsilon : [0,t] \times \{\|x\| \geqslant \epsilon\} \to \mathbb{R}^+$ maps (s, x) to $\|x\|^2$, and with characteristic drift $\gamma_\epsilon = \int_{\mathbb{R}} y\rho_\epsilon(\mathrm{d}y)$. From this it follows that, for every positive ϵ,

$$\int_0^\infty y^{k/2}\rho_\epsilon(\mathrm{d}y) = t\int_{\|x\|\geqslant\epsilon} \|x\|^k \nu^L(\mathrm{d}x) \leqslant t\int_{\|x\|<1} \|x\|^2 \nu^L(\mathrm{d}x) + t\int_{\|x\|\geqslant 1} \|x\|^k \nu^L(\mathrm{d}x) < \infty,$$

and

$$\gamma_\epsilon = \int_0^\infty y\rho_\epsilon(\mathrm{d}y) = t\int_{\|x\|\geqslant\epsilon} \|x\|^2 \nu^L(\mathrm{d}x) \leqslant t\int_{\|x\|<1} \|x\|^2 \nu^L(\mathrm{d}x) + t\int_{\|x\|\geqslant 1} \|x\|^2 \nu^L(\mathrm{d}x) < \infty.$$

Lemma 4.2 then implies that $\lim_{\epsilon\to 0} \mathbb{E}Y_\epsilon^{k/2}$ is finite, which completes the proof. □

Next, we consider the differentiation of integrals of functions, for which we introduce the notations

$$I_f(t) := \int_0^t f(s)\mathrm{d}s. \tag{4.5.7}$$

The corresponding approximation error is denoted by

$$\boldsymbol{e}_{I_f,n}^{(h)} := \Delta_h^1 \left[I_f\right](n) - f(n). \tag{4.5.8}$$

In the next lemma we analyse this approximation when f is a Lévy process.

Lemma 4.14 *The sequence of approximation errors $\boldsymbol{e}_{I_L}^{(h)}$ is i. i. d. Moreover, for every $\omega \in \Omega$ and for every integer n, the approximation error $\boldsymbol{e}_{I_L,n}^{(h)}$ converges to zero as $h \to 0$. If, for some positive integer k, the absolute moment $\mathbb{E}\,\|\boldsymbol{L}(1)\|^{(k)_0}$ is finite, then $\mathbb{E}\left\|\boldsymbol{e}_{I_L,n}^{(h)}\right\|^k = O(h^{k/(k)_0})$, as $h \to 0$, where the constant implicit in the $O(\cdot)$ notation does not depend on n.*

Proof We first observe that

$$\|I_L(n+h) - I_L(n) - h\boldsymbol{L}(n)\| = \left\|\int_n^{n+h} [\boldsymbol{L}(s) - \boldsymbol{L}(n)]\,\mathrm{d}s\right\| \leqslant \int_n^{n+h} \|\boldsymbol{L}(s) - \boldsymbol{L}(n)\|\,\mathrm{d}s.$$

The right continuity of $t \mapsto \boldsymbol{L}(t)$ implies that for, every integer n and each $\epsilon > 0$, there exists a $\delta_{\epsilon,n}$ such that $\|\boldsymbol{L}(n+t) - \boldsymbol{L}(n)\| \leqslant \epsilon$, for all $0 \leqslant t \leqslant \delta_{\epsilon,n}$. This means that the difference $I_L(n+h) - I_L(n) - h\boldsymbol{L}(t)$ is less than $h\epsilon$ in absolute value, provided h is smaller than $\delta_{\epsilon,n}$. Dividing by h thus proves that $\boldsymbol{e}_{I_L,n}^{(h)}$ converges to zero as h tends to zero. The proof also shows that $\boldsymbol{e}_{I_L,n}^{(h)}$ is a deterministic function of the increments $\{\boldsymbol{L}(s) - \boldsymbol{L}(n), n \leqslant s \leqslant n+h\}$. Since the increments of a Lévy process are stationary and independent, this implies that $\boldsymbol{e}_{I_L}^{(h)}$ is an i. i. d. sequence.

For the second claim about the size of the absolute moments of $e^{(h)}_{I_L,n}$ for small h, it is no restriction to assume that $n = 0$. Successive application of the triangle inequality and Hölder's inequality with the dual exponent k' determined by $1/k + 1/k' = 1$ shows that

$$\begin{aligned}\mathbb{E}\left\|e^{(h)}_{I_L,0}\right\|^k =& \frac{1}{h^k}\mathbb{E}\left\|\int_0^h \boldsymbol{L}(s)\mathrm{d}s\right\|^k\\ \leqslant& \frac{1}{h^k}\mathbb{E}\left(\int_0^h \|\boldsymbol{L}(s)\|\,\mathrm{d}s\right)^k \leqslant \frac{1}{h^k}\mathbb{E}\left(\left(\int_0^h \|\boldsymbol{L}(s)\|^k\mathrm{d}s\right)^{1/k}\left(\int_0^h 1\mathrm{d}s\right)^{1/k'}\right)^k.\end{aligned}$$

Using $k/k' = k-1$, it follows that

$$\mathbb{E}\left\|e^{(h)}_{I_L,0}\right\|^k \leqslant \frac{1}{h}\mathbb{E}\int_0^h \|\boldsymbol{L}(s)\|^k\mathrm{d}s.$$

Since $\|\boldsymbol{L}(s)\|^k$ is positive, we can interchange the expectation and integral. By Proposition 4.3, $\mathbb{E}\|\boldsymbol{L}(s)\|^k$ is of order $O(s^{k/(k)_0})$, which implies that $\left\|e^{(h)}_{I_L,0}\right\|^k = O(h^{k/(k)_0})$. □

Lemma 4.14 was dedicated to the analysis of the error of approximating the first derivative of the integral of a Lévy process. We will also need analogous results for higher order derivatives of iterated integrals of Lévy processes. The proofs are similar in spirit and only technically more complicated. For a positive integer ν, we generalize the notations (4.5.7) and (4.5.8) to

$$I_f^\nu(t) = \int_0^t I_f^{\nu-1}(s)\mathrm{d}s, \quad I_f^1(t) = \int_0^t f(s)\mathrm{d}s \tag{4.5.9}$$

for the ν-fold iterated integral of the function f, and

$$e^{\nu,(h)}_{I_f^\nu,n} := \Delta_h^\nu\left[I_f^\nu\right](n) - f(n). \tag{4.5.10}$$

Clearly, if the function f has only countably many jump discontinuities, then $\mathrm{D}^\nu I^\nu[f](t) = f(t)$ almost everywhere. We first prove the following locality property.

Lemma 4.15 *For every positive integer $\nu \geqslant 2$ and every function f, the approximation error $e^{\nu,(h)}_{I_f^\nu,n}$ is a function only of the increments $\{f(t) - f(n) : n \leqslant t \leqslant n + \nu h\}$. This function is independent of n. In particular, $e^{\nu,(h)}_{I_L^\nu}$ is an i. i. d. sequence.*

Proof The claim can be shown by direct calculations: Lemma 4.12, i) implies that

$$\begin{aligned}\Delta_h^\nu\left[I_f^\nu\right](n) =& \Delta_h^1\Delta_h^{\nu-1}\left[I_f^\nu\right](n)\\ =& \Delta_h^1\left[\frac{1}{h^{\nu-1}}\sum_{i=0}^{\nu-1}(-1)^{\nu-1-i}\binom{\nu-1}{i}I_f^\nu(\cdot+ih)\right](n)\\ =& \frac{1}{h^\nu}\sum_{i=0}^{\nu-1}(-1)^{\nu-1-i}\binom{\nu-1}{i}\int_{n+ih}^{n+(i+1)h} I_f^{\nu-1}(s)\mathrm{d}s\end{aligned}$$

$$= \frac{1}{h^\nu} \sum_{i=0}^{\nu-1} (-1)^{\nu-1-i} \binom{\nu-1}{i} \int_{n+ih}^{n+(i+1)h} \left[\int_{0 \leqslant t_{\nu-1} \leqslant \cdots \leqslant t_1 \leqslant s} f(t_{\nu-1}) \mathrm{d}t_{\nu-1} \cdots \mathrm{d}t_1 \right] \mathrm{d}s.$$

Using that the set $\{t_{\nu-1} \leqslant t_{\nu-2} \leqslant \cdots \leqslant t_1 \leqslant s\}$ is congruent to the $(\nu-2)$-dimensional simplex in the hypercube with side lengths $s - t_{\nu-1}$, and that thus

$$\int_{t_{\nu-1} \leqslant t_{\nu-2} \leqslant \cdots \leqslant t_1 \leqslant s} \mathrm{d}t_{\nu-2} \cdots \mathrm{d}t_1 = \frac{1}{(\nu-2)!} (s - t_{\nu-1})^{\nu-2},$$

we obtain that

$$\begin{aligned}
\Delta_h^\nu \left[I_f^\nu \right](n) =& \frac{1}{h^\nu (\nu-2)!} \sum_{i=0}^{\nu-1} (-1)^{\nu-1-i} \binom{\nu-1}{i} \int_{n+ih}^{n+(i+1)h} \int_0^s (s - t_{\nu-1})^{\nu-2} f(t_{\nu-1}) \mathrm{d}t_{\nu-1} \mathrm{d}s \\
=& \frac{1}{h^\nu (\nu-2)!} \int_0^n \left[\sum_{i=0}^{\nu-1} (-1)^{\nu-1-i} \binom{\nu-1}{i} \int_{n+ih}^{n+(i+1)h} (s - t_{\nu-1})^{\nu-2} \mathrm{d}s \right] f(t_{\nu-1}) \mathrm{d}t_{\nu-1} \\
&+ \frac{1}{h^\nu (\nu-2)!} \sum_{i=0}^{\nu-1} (-1)^{\nu-1-i} \binom{\nu-1}{i} \int_{n+ih}^{n+(i+1)h} \int_n^s (s - t_{\nu-1})^{\nu-2} f(t_{\nu-1}) \mathrm{d}t_{\nu-1} \mathrm{d}s.
\end{aligned}$$

It is easy to see that $\int_{n+ih}^{n+(i+1)h} (s - t_{\nu-1})^{\nu-2} \mathrm{d}s$ is equal to $\mathfrak{p}_{\nu,h}(n - t_{\nu-1} + ih)$ for some polynomial $\mathfrak{p}_{\nu,h}$ of degree $\nu - 2$. It then follows from Lemma 4.12, ii) that

$$\sum_{i=0}^{\nu-1} (-1)^{\nu-1-i} \binom{\nu-1}{i} \int_{n+ih}^{n+(i+1)h} (s - t_{\nu-1})^{\nu-2} \mathrm{d}s = \Delta_h^{\nu-1} \left[\mathfrak{p}_{\nu,h} \right] (n - t_{\nu-1}) = 0, \quad \forall t_{\nu-1} \in [0, n],$$

which implies that the first term in the last expression for $\Delta_h^\nu \left[I_f^\nu \right](n)$ vanishes. It is similarly easy to see that

$$\frac{1}{h^\nu (\nu-2)!} \sum_{i=0}^{\nu-1} (-1)^{\nu-1-i} \binom{\nu-1}{i} \int_{n+ih}^{n+(i+1)h} \int_n^s (s - t_{\nu-1})^{\nu-2} \mathrm{d}t_{\nu-1} \mathrm{d}s = 1.$$

Consequently,

$$\begin{aligned}
&\boldsymbol{e}_{I_f^\nu, n}^{\nu,(h)} = \Delta_h^\nu \left[I_f^\nu \right](n) - f(n) \\
=& \frac{1}{h^\nu (\nu-2)!} \sum_{i=0}^{\nu-1} (-1)^{\nu-1-i} \binom{\nu-1}{i} \int_{ih}^{(i+1)h} \int_0^s (s - t_{\nu-1})^{\nu-2} \left[f(n + t_{\nu-1}) - f(n) \right] \mathrm{d}t_{\nu-1} \mathrm{d}s,
\end{aligned} \tag{4.5.11}$$

which completes the proof of the first part of the lemma. The fact that Lévy processes have stationary and independent increments together with the last display implies that the sequence $\boldsymbol{e}_{I_L^\nu}^{\nu,(h)}$ is i. i. d. □

Lemma 4.16 *For every positive integer $\nu \geqslant 1$ and every integer n, the error $\boldsymbol{e}_{I_L^\nu, n}^{\nu,(h)}$ converges to zero* *as $h \to 0$. If, moreover, $\mathbb{E}\|\boldsymbol{L}(1)\|^{(k)_0}$ is finite for some $k > 0$, then $\mathbb{E}\left\|\boldsymbol{e}^{\nu,(h)}_{I^\nu_L,n}\right\|^k = O(h^{k/(k)_0})$, as $h \to 0$.*

Proof Let $\epsilon > 0$ be given. By the right-continuity of $\boldsymbol{L}$ there exists a positive number $\delta_{\epsilon,n}$ such that $\|\boldsymbol{L}(n+t) - \boldsymbol{L}(n)\| \leqslant \epsilon$ for all $t \in [0, \delta_{\epsilon,n}]$. Hence, assuming $\nu h \leqslant \delta_{\epsilon,n}$, Eq. (4.5.11) implies that

$$\begin{aligned}\left\|\boldsymbol{e}^{\nu,(h)}_{I^\nu_L,n}\right\| \leqslant & \frac{\epsilon}{h^\nu(\nu-2)!}\sum_{i=0}^{\nu-1}\binom{\nu-1}{i}\int_{ih}^{(i+1)h}\int_0^s (s-t_{\nu-1})^{\nu-2}\mathrm{d}t_{\nu-1}\mathrm{d}s\\ = & \frac{\epsilon}{\nu!}\sum_{i=0}^{\nu-1}\binom{\nu-1}{i}\left[(i+1)^\nu - i^\nu\right].\end{aligned}$$

This proves that $\left\|\boldsymbol{e}^{\nu,(h)}_{I^\nu_L,n}\right\| \to 0$, as $h \to 0$. We now turn to the absolute moments of $\boldsymbol{e}^{\nu,(h)}_{I^\nu_L,n}$. Again it entails no loss of generality to assume that $n = 0$. Equation (4.5.11) and the triangle inequality lead to

$$\begin{aligned}\mathbb{E}\left\|\boldsymbol{e}^{\nu,(h)}_{I^\nu_L,0}\right\|^k = & \mathbb{E}\left\|\frac{1}{h^\nu(\nu-2)!}\sum_{i=0}^{\nu-1}(-1)^{\nu-1-i}\binom{\nu-1}{i}\int_{ih}^{(i+1)h}\int_0^s (s-t)^{\nu-2}\boldsymbol{L}(t)\mathrm{d}t\mathrm{d}s\right\|^k\\ \leqslant & \left[\frac{1}{h^\nu(\nu-2)!}\right]^k \mathbb{E}\left(\sum_{i=0}^{\nu-1}\binom{\nu-1}{i}\int_{ih}^{(i+1)h}\int_0^s (s-t)^{\nu-2}\|\boldsymbol{L}(t)\|\,\mathrm{d}t\mathrm{d}s\right)^k.\end{aligned}$$

An application of Hölder's inequality with the dual exponent k' determined by $1/k + 1/k' = 1$ shows that the last line of the previous display is dominated by

$$\begin{aligned}&\leqslant\left[\frac{1}{h^\nu(\nu-2)!}\right]^k \mathbb{E}\left(\sum_{i=0}^{\nu-1}\binom{\nu-1}{i}^{k'}\int_{ih}^{(i+1)h}\int_0^s (s-t)^{k'(\nu-2)}\mathrm{d}t\mathrm{d}s\right)^{k/k'}\left(\int_0^{\nu h}\int_0^s \|\boldsymbol{L}(t)\|^k\,\mathrm{d}t\mathrm{d}s\right)\\ &=\frac{C}{h^2}\mathbb{E}\left(\int_0^{\nu h}(\nu h - t)\|\boldsymbol{L}(t)\|^k\,\mathrm{d}t\right),\end{aligned}$$

where the constant C depends only on ν and k and is given by

$$\frac{1}{[(\nu-2)!]^k}\left[\frac{1}{[k'(\nu-2)+2][k'(\nu-2)+1]}\sum_{i=0}^{\nu-1}\binom{\nu-1}{i}^{k'}\left[(i+1)^{k'(\nu-2)+2} - i^{k'(\nu-2)+2}\right]\right]^{k-1}.$$

Proposition 4.3 asserts the existence of a constant C' such that $\mathbb{E}\|\boldsymbol{L}(t)\|^k < C' t^{k/(k)_0}$ for all $t \leqslant \nu h$. Consequently

$$\mathbb{E}\left\|\boldsymbol{e}^{\nu,(h)}_{I^\nu_L,0}\right\|^k \leqslant \frac{CC'}{h^2}\int_0^{\nu h}(\nu h - t)t^{k/(k)_0}\mathrm{d}t = \frac{CC'\nu^{k/(k)_0+2}}{[k/(k)_0+1]\,[k/(k)_0+2]}h^{k/(k)_0},$$

showing that $\mathbb{E}\left\|e^{\nu,(h)}_{I^\nu_L,0}\right\|^k = O(h^{k/(k)_0})$ and thereby completing the proof of the lemma. □

With these auxiliary results finished, we turn to approximating derivatives of the multivariate CARMA process $\boldsymbol{Y}$. This is the first big step towards discretizing Eq. (4.4.6).

Proposition 4.17 *Let $\boldsymbol{Y}$ be an $\boldsymbol{L}$-driven multivariate CARMA process satisfying Assumption A1, let $n \geqslant 0$ be an integer, and denote by $\boldsymbol{e}^{\nu,(h)}_{\boldsymbol{Y},n} = \Delta^\nu_h[\boldsymbol{Y}](n) - \mathrm{D}^\nu \boldsymbol{Y}(n)$ the error of approximating the νth derivative of $\boldsymbol{Y}$ by the forward differences defined in Eq. (4.5.1). Assume that, for some $k > 0$, $\mathbb{E}\,\|\boldsymbol{L}(1)\|^{(k)_0} < \infty$. Then the following hold.*

i) If $1 \leqslant \nu \leqslant p-q-2$, then $\mathbb{E}\left\|\boldsymbol{e}^{\nu,(h)}_{\boldsymbol{Y},n}\right\|^k = O(h^k)$. If $\nu = p-q-1$, then $\mathbb{E}\left\|\boldsymbol{e}^{\nu,(h)}_{\boldsymbol{Y},n}\right\|^k = O(h^{k/(k)_0})$.

ii) The sequence $\boldsymbol{e}^{\nu,(h)}_{\boldsymbol{Y}}$ is strictly stationary and strongly mixing with exponentially decaying mixing coefficients.

Proof We first prove the assertions i) about the behaviour of the absolute moments of $\boldsymbol{e}^{\nu,(h)}_{\boldsymbol{Y},n}$ for small values of h. If $1 \leqslant \nu \leqslant p-q-2$, it follows from Lemma 4.7 that the paths of $\boldsymbol{Y}$ are at least $\nu+1$ times differentiable; therefore, Lemma 4.12 implies that $\left\|\boldsymbol{e}^{\nu,(h)}_{\boldsymbol{Y},n}\right\| \leqslant h \sup_{n\leqslant s\leqslant n+\nu h}\left\|\mathrm{D}^{\nu+1}\boldsymbol{Y}(s)\right\|$. To prove the claim, it is thus sufficient to show the finiteness of $\mathbb{E}\sup_{n\leqslant s\leqslant n+\nu h}\left\|\mathrm{D}^{\nu+1}\boldsymbol{Y}(s)\right\|^k$. By the defining observation equation (4.3.10b), $\boldsymbol{Y}$ is a linear combination of the first $q+1$ m-blocks of the state process $\boldsymbol{X}$; the state equation (4.3.10a) implies $\mathrm{D}\boldsymbol{X}^i = \boldsymbol{X}^{i+1}$, $i = 1,\ldots,p-1$, and since ν is assumed to be no bigger than $p-q-2$, it follows that $\mathrm{D}^{\nu+1}\boldsymbol{Y}$ is a linear combination of the first $p-1$ m-blocks of $\boldsymbol{X}$, say $\mathrm{D}^{\nu+1}\boldsymbol{Y} = \Lambda\boldsymbol{X}$, for some matrix $\Lambda \in M_{d,pm}(\mathbb{R})$. We can then apply Lemma 4.13 to estimate

$$\mathbb{E}\sup_{n\leqslant s\leqslant n+\nu h}\left\|\mathrm{D}^{\nu+1}\boldsymbol{Y}(s)\right\|^k \leqslant \|\Lambda\|^k\,\mathbb{E}\sup_{n\leqslant s\leqslant n+\nu h}\|\boldsymbol{X}(s)\|^k < \infty,$$

which proves the first claim. If $\nu = p-q-1$, we start again from the observation that $\boldsymbol{Y}$ is a linear combination of the first $q+1$ m-blocks of $\boldsymbol{X}$, namely,

$$\boldsymbol{Y}(t) = \underline{B}_q\boldsymbol{X}_q(t) + B_q\boldsymbol{X}^{(q+1)}(t), \quad t\in\mathbb{R}, \quad \underline{B}_q = \left[\begin{array}{ccc} B_0 & \cdots & B_{q-1}\end{array}\right].$$

By solving the last $p-q+1$ block-rows of the state equation (4.3.10a), one can express $\boldsymbol{X}^{(q+1)}$ as

$$\boldsymbol{X}^{(q+1)}(t) = \frac{t^{p-q-1}}{(p-q-1)!}\boldsymbol{X}^{(p)}(0) - \underline{A}I^{p-q}_{\boldsymbol{X}}(t) + I^{p-q-1}_{\boldsymbol{L}}(t),$$

where the notation I^ν_f for the ν-fold iterated integral of a function f has been introduced in Eq. (4.5.9). By linearity and the fact that $\Delta^\nu_h[\mathfrak{p}] - \mathrm{D}^\nu\mathfrak{p} = 0$ for polynomials $\mathfrak{p}$ of degree ν

(Lemma 4.12, ii)), it follows that

$$\begin{aligned}
\boldsymbol{e}_{\boldsymbol{Y},n}^{p-q-1,(h)} =& \Delta_h^{p-q-1}[\boldsymbol{Y}](n) - \mathrm{D}^{p-q-1}\boldsymbol{Y}(n)\\
=& \underline{B}_q\left[\Delta_h^{p-q-1}[\boldsymbol{X}_q](n) - \mathrm{D}^{p-q-1}\boldsymbol{X}_q(n)\right]\\
&- B_q\underline{A}\left[\Delta_h^{p-q-1}\left[I_{\boldsymbol{X}}^{p-q}\right](n) - \mathrm{D}^{p-q-1}I_{\boldsymbol{X}}^{p-q}(n)\right]\\
&+ B_q\left[\Delta_h^{p-q-1}\left[I_{\boldsymbol{L}}^{p-q-1}\right](n) - \mathrm{D}^{p-q-1}I_{\boldsymbol{L}}^{p-q-1}(n)\right].
\end{aligned}$$

Both $\boldsymbol{X}_q$ (by Lemma 4.7) and $I_{\boldsymbol{X}}^{p-q}$ are $p-q$ times differentiable so we can apply Lemma 4.12, iii) to bound the differences in the first two lines of the last display by h times the supremum of the $(p-q)$th derivative of $\boldsymbol{X}_q$ and $I_{\boldsymbol{X}}^{p-q}$, respectively. The contribution from the last line is the approximation error for the $(p-q-1)$th derivative of the $(p-q-1)$-fold iterated integral of the Lévy process $\boldsymbol{L}$, which has been investigated in Lemma 4.16. We thus obtain that

$$\begin{aligned}
\left\|\boldsymbol{e}_{\boldsymbol{Y},n}^{p-q-1,(h)}\right\| \leqslant & h\left[\left\|\underline{B}_q\right\| \sup_{n\leqslant t\leqslant n+(p-q-1)h}\left\|\mathrm{D}^{p-q}\boldsymbol{X}_q(t)\right\| + \left\|B_q\right\|\left\|\underline{A}\right\| \sup_{n\leqslant t\leqslant n+(p-q-1)h}\left\|\boldsymbol{X}(t)\right\|\right]\\
&+ \left\|B_q\right\|\left\|\boldsymbol{e}_{I_{\boldsymbol{L}}^{p-q-1},n}^{p-q-1,(h)}\right\|.
\end{aligned}$$

As before, one shows that the first term has finite kth moments which is of order $O(h^k)$. The second term has been shown in Lemma 4.16 to have finite kth moment of order $O(h^{k/(k)_0})$ which dominates the first term for $h<1$; this completes the proof of i).

In order to prove that the sequence $\boldsymbol{e}_{\boldsymbol{Y}}^{\nu,(h)}$ is strongly mixing, it is enough, by virtue of Lemma 4.7, iv) and Lemma 4.8, to show that the approximation error $\boldsymbol{e}_{\boldsymbol{Y},n}^{\nu,(h)}$ is measurable with respect to $\mathscr{Y}_n^{n+\nu h}$, the σ-algebra generated by $\{\boldsymbol{Y}(t): n\leqslant t\leqslant \nu h\}$. Clearly, $\Delta_h^\nu[\boldsymbol{Y}](t)$ is measurable with respect to the σ-algebra generated by $\{\boldsymbol{Y}(t),\boldsymbol{Y}(t+h),\ldots,\boldsymbol{Y}(t+\nu h)\}$. By the definition of derivatives as the limit of different quotients and the assumed differentiability of $t\mapsto\boldsymbol{Y}(t)$, the derivative $D_t^\nu\boldsymbol{Y}(t)$ is the ω-wise limit, as s goes to zero, of the functions $\omega\mapsto\Delta_s^\nu[\boldsymbol{Y}_\omega](t)$, where ω is an element of Ω. Each of these functions is measurable with respect to $\sigma(\boldsymbol{Y}(t),\boldsymbol{Y}(t+s),\ldots,\boldsymbol{Y}(t+\nu s))$, and therefore in particular with respect to the larger σ-algebra $\mathscr{Y}_n^{n+\nu h}$. Since point-wise limits of measurable functions are measurable (Klenke, 2008, Theorem 1.92), the claim follows.

The claim that the sequence $\boldsymbol{e}_{\boldsymbol{Y}}^{\nu,(h)}$ is strictly stationary is a consequence of the fact that the multivariate CARMA process $\boldsymbol{Y}$ is strictly stationary (Lemma 4.7, i)). By the definition of stationarity, it is enough to show that for every natural number K, all indices $n_1,\ldots,n_K\in\mathbb{Z}$, and every integer k, the two arrays $(\boldsymbol{e}_{\boldsymbol{Y},n_1}^{\nu,(h)},\ldots,\boldsymbol{e}_{\boldsymbol{Y},n_K}^{\nu,(h)})$ and $(\boldsymbol{e}_{\boldsymbol{Y},n_1+k}^{\nu,(h)},\ldots,\boldsymbol{e}_{\boldsymbol{Y},n_K+k}^{\nu,(h)})$ have the same distribution. We first observe that for each $n\in\mathbb{Z}$ and each $\omega\in\Omega$, $\boldsymbol{e}_{\boldsymbol{Y},n}^{\nu,(h)} = \lim_{s\to 0^+}\boldsymbol{e}_{\boldsymbol{Y},n}^{\nu,(h,s)}$, where $\boldsymbol{e}_{\boldsymbol{Y},n}^{\nu,(h,s)} := \Delta_h^\nu[\boldsymbol{Y}](n) - \Delta_s^\nu[\boldsymbol{Y}](n)$. In particular, since ω-wise convergence implies convergence in distribution, it holds that

$$(e_{Y,n_1}^{\nu,(h,s)},\dots,e_{Y,n_K}^{\nu,(h,s)}) \xrightarrow{d} (e_{Y,n_1}^{\nu,(h)},\dots,e_{Y,n_K}^{\nu,(s)}),$$
$$(e_{Y,n_1+k}^{\nu,(h,s)},\dots,e_{Y,n_K+k}^{\nu,(h,s)}) \xrightarrow{d} (e_{Y,n_1+k}^{\nu,(h)},\dots,e_{Y,n_K+k}^{\nu,(s)}),$$

as s tends to zero. For every finite s, the strict stationarity of Y implies that the tuple $(e_{Y,n_1}^{\nu,(h,s)},\dots,e_{Y,n_K}^{\nu,(h,s)})$ is equal in distribution to $(e_{Y,n_1+k}^{\nu,(h,s)},\dots,e_{Y,n_K+k}^{\nu,(h,s)})$. The assertion then follows from the fact that in Polish spaces weak limits are uniquely determined (Klenke, 2008, Remark 13.13). □

4.5.2. Approximation of integrals

This section is devoted to the approximations of the integrals appearing in Eq. (4.4.6), namely $\int_{n-1}^{n} Y(s)\mathrm{d}s$ and $\int_{n-1}^{n} \mathrm{e}^{\mathrm{B}(n-s)}Y(s)\mathrm{d}s$. One of the simplest approximations for definite integrals is the trapezoidal rule, see, e. g., Deuflhard and Hohmann (2008, Chapter 9) for an introduction to the topic of numerical integration. For any function $f:\mathbb{R}\to M$ with values in a metric space M, it is defined as

$$T_{[a,b]}^{K} f := \frac{b-a}{K}\left[\frac{f(a)+f(b)}{2} + \sum_{k=1}^{K-1} f(n-1+k\frac{b-a}{K})\right],\quad K\in\mathbb{N}, \tag{4.5.12}$$

and is meant to approximate the definite integral $\int_a^b f(s)\mathrm{d}s$. We will usually set $[a,b]=[n-1,n]$, $n\in\mathbb{N}$, and $=h^{-1}$. It is clear that $T_{[n-1,n]}^{h^{-1}} f$ can be computed from knowledge of the values of f on the discrete time grid $(0,h,2h,\dots)$. The following result provides a quantitative bound for the accuracy with which the trapezoidal rule approximates a definite integral if the integrand is a smooth function.

Proposition 4.18 *Let $[a,b]\subset\mathbb{R}$ be an interval, and let K be a positive integer.*

i) Assume that $f:[a,b]\to\mathbb{R}$ is a twice differentiable function. Then

$$\left|\int_a^b f(s)\mathrm{d}s - T_{[a,b]}^{K} f\right| \leqslant \frac{(b-a)^3}{12K^2}\sup_{t\in[a,b]}\left|f''(t)\right|. \tag{4.5.13}$$

ii) Assume that $F:[a,b]\to\mathbb{R}^d$ is a twice differentiable function. Then

$$\left\|\int_a^b F(s)\mathrm{d}s - T_{[a,b]}^{K} F\right\| \leqslant \frac{(b-a)^3\sqrt{d}}{12K^2}\sup_{t\in[a,b]}\left\|F''(t)\right\|. \tag{4.5.14}$$

iii) Assume that $F:[a,b]\to M_d(\mathbb{R})$ is a twice differentiable function. Then

$$\left\|\int_a^b F(s)\mathrm{d}s - T^K_{[a,b]}F\right\| \leqslant \frac{(b-a)^3 d^{3/2}}{12K^2}\sup_{t\in[a,b]}\left\|F''(t)\right\|. \tag{4.5.15}$$

Proof Part i) is Deuflhard and Hohmann (2008, Lemma 9.8). To see that ii) holds, it is enough to apply i) component-wise to obtain that

$$\begin{aligned}\left\|\int_a^b F(s)\mathrm{d}s - T^K_{[a,b]}F\right\| &\leqslant \sqrt{d}\max_{i=1,\ldots,d}\left|\left[\int_a^b F(s)\mathrm{d}s - T^K_{[a,b]}F\right]_i\right|\\ &\leqslant \frac{(b-a)^3\sqrt{d}}{12K^2}\max_{i=1,\ldots,d}\sup_{t_i\in[a,b]}\left|F_i''(t_i)\right|\\ &= \frac{(b-a)^3\sqrt{d}}{12K^2}\sup_{t\in[a,b]}\max_{i=1,\ldots,d}\left|F_i''(t)\right| \leqslant \frac{(b-a)^3\sqrt{d}}{12K^2}\sup_{t\in[a,b]}\left\|F''(t)\right\|.\end{aligned}$$

The claim (4.5.15) about matrix-valued integrands follows from the fact that $M_d(\mathbb{R})$ is canonically isomorphic to $\mathbb{R}^{d^2}$ and that the operator norm and the Euclidean vector norm induced by this isomorphism satisfy

$$\frac{1}{\sqrt{d}}\left\|M\right\|_{\mathbb{R}^{d^2}} \leqslant \left\|M\right\| \leqslant \left\|M\right\|_{\mathbb{R}^{d^2}},$$

for all $M\in M_d(\mathbb{R})$ (Stone, 1962). □

The last proposition can be used to derive properties of the approximation error of convolutions of vector-valued functions with matrix-valued kernels. For any compatible functions $f:[0,\infty]\to\mathbb{R}^d$ and $g:[0,1]\to M_d(\mathbb{R})$, we use the notation

$$\boldsymbol{\varepsilon}^{(h)}_{g\circ f,n} = T^{h^{-1}}_{[n-1,n]}g(n-\cdot)f(\cdot) - \int_{n-1}^n g(n-s)f(s)\mathrm{d}s \tag{4.5.16}$$

for the difference between the exact value of the convolution integral and the one obtained from the trapezoidal approximation with sampling interval h. In the next proposition we analyse this approximation error if f is a multivariate CARMA process; this is the second big step towards discretizing Eq. (4.4.6).

Proposition 4.19 *Assume that $\boldsymbol{L}$ is a Lévy process. Let $\boldsymbol{Y}$ be a d-dimensional $\boldsymbol{L}$-driven MCARMA process satisfying Assumption A1, let $F:[0,1]\to M_d(\mathbb{R})$ a twice continuously differentiable function, and denote by $\boldsymbol{\varepsilon}^{(h)}_{F\circ\boldsymbol{Y},n}$ the approximation error of the trapezoidal rule, defined in Eq. (4.5.16). If $\mathbb{E}\left\|\boldsymbol{L}(1)\right\|^k$ is finite, then $\mathbb{E}\left\|\boldsymbol{\varepsilon}^{(h)}_{F\circ\boldsymbol{Y},n}\right\|^k = O(h^{2k})$, as $h\to 0$. Moreover, the sequence $\boldsymbol{\varepsilon}^{(h)}_{F\circ\boldsymbol{Y}}$ is strictly stationary and strongly mixing.*

Proof By the definition of $\boldsymbol{\varepsilon}^{(h)}_{F\circ \boldsymbol{Y}}$ (Eqs. (4.5.12) and (4.5.16)), we can write

$$\boldsymbol{\varepsilon}^{(h)}_{F\circ\boldsymbol{Y},n} = h\sum_{i=0}^{h^{-1}} \alpha_i^{(h)}\boldsymbol{Y}(n-1+ih) - \int_{n-1}^{n} F(n-s)\boldsymbol{Y}(s)\mathrm{d}s,$$

where

$$\alpha_0^{(h)} = \frac{F(1)}{2}, \quad \alpha_{h^{-1}} = \frac{F(0)}{2}, \quad \alpha_i^{(h)} = F(1-ih), \quad i=1,\ldots h^{-1}-1.$$

Using Dirac's δ-distribution, which is defined by the property that $\int f(x)\delta_{x_0}(x)\mathrm{d}x = f(x_0)$ for all compactly supported smooth functions f, as well as the moving average representation (4.3.13) of $\boldsymbol{Y}$, we obtain that

$$\begin{aligned}\boldsymbol{\varepsilon}^{(h)}_{F\circ\boldsymbol{Y}} &= \int_{n-1}^{n}\left[\sum_i \alpha_i^{(h)}\delta_{n-1+ih}(s) - F(n-s)\right]\boldsymbol{Y}(s)\mathrm{d}s\\ &= \int_{n-1}^{n}\left[h\sum_i \alpha_i^{(h)}\delta_{n-1+ih}(s) - F(n-s)\right]\int_{-\infty}^{s}\underline{B}\mathrm{e}^{\mathrm{A}(s-u)}E_p\mathrm{d}\boldsymbol{L}(u)\mathrm{d}s.\end{aligned}$$

Theorem 4.4 allows us to interchange the order of integration so that we obtain

$$\begin{aligned}\boldsymbol{\varepsilon}^{(h)}_{F\circ\boldsymbol{Y},n} &= \int_{-\infty}^{n}\int_{\max\{u,n-1\}}^{n}\left[h\sum_i \alpha_i^{(h)}\delta_{n-1+ih}(s) - F(n-s)\right]\underline{B}\mathrm{e}^{\mathrm{A}(s-u)}E_p\mathrm{d}s\mathrm{d}\boldsymbol{L}(u)\\ &= \int_{-\infty}^{n-1}\int_{n-1}^{n}\left[-h\sum_i \alpha_i^{(h)}\delta_{n-1+ih}(s) - F(n-s)\right]\underline{B}\mathrm{e}^{\mathrm{A}(s-u)}E_p\mathrm{d}s\mathrm{d}\boldsymbol{L}(u)\\ &\quad + \int_{n-1}^{n}\int_{u}^{n}\left[h\sum_i \alpha_i^{(h)}\delta_{n-1+ih}(s) - F(n-s)\right]\underline{B}\mathrm{e}^{\mathrm{A}(s-u)}E_p\mathrm{d}s\mathrm{d}\boldsymbol{L}(u).\end{aligned}$$

With the notations

$$\Gamma^{(h)} := \int_0^1\left[-h\sum_i \alpha_i^{(h)}\delta_{ih}(s) - F(1-s)\right]\underline{B}\mathrm{e}^{\mathrm{A}s}\mathrm{d}s \tag{4.5.17}$$

and

$$G^{(h)}: \begin{cases}[0,1] & \to M_{d,m}(\mathbb{R}),\\ t & \mapsto \int_0^t\left[h\sum_i \alpha_i^{(h)}\delta_{t-1+ih}(s) - F(t-s)\right]\underline{B}\mathrm{e}^{\mathrm{A}s}\mathrm{d}s\, E_p,\end{cases} \tag{4.5.18}$$

we can rewrite the previous display as

$$\boldsymbol{\varepsilon}^{(h)}_{F\circ\boldsymbol{Y},n} = \Gamma^{(h)}\boldsymbol{X}(n-1) + \int_{n-1}^{n} G^{(h)}(n-u)\mathrm{d}\boldsymbol{L}(u), \tag{4.5.19}$$

where we have used the moving average representation (4.3.12) of the state vector process $\boldsymbol{X}$. This equation and the strict stationarity of $\boldsymbol{X}$ asserted in Lemma 4.7, i) immediately imply that the sequence $\varepsilon^{(h)}_{F\circ Y}$ is strictly stationary and strongly mixing. By Proposition 4.18 there exists a constant C such that $\left\|\Gamma^{(h)}\right\| \leqslant Ch^2$ and $\left\|G^{(h)}(t)\right\| \leqslant Ch^2$ for all $t \in [0,1]$, which implies that

$$\mathbb{E}\left\|\varepsilon^{(h)}_{F\circ Y}\right\|^k \leqslant h^{2k}C^k 2^k \mathbb{E}\left\|\boldsymbol{X}(n-1)\right\|^k + 2^k\mathbb{E}\left\|\int_{n-1}^n G^{(h)}(n-u)\mathrm{d}\boldsymbol{L}(u)\right\|^k.$$

The kth moment of $\boldsymbol{X}(n-1)$ is finite by Lemma 4.7, iii), so it suffices to prove that the second term is of order $O(h^{2k})$. To this end we use the fact that $\int_{n-1}^n G^{(h)}(n-u)\mathrm{d}\boldsymbol{L}(u)$ is an infinitely divisible random variable whose characteristic triplet $(\gamma_G^{(h)}, \Sigma_G^{(h)}, \nu_G^{(h)})$ can be expressed explicitly in terms of the characteristic triplet $(\gamma^{\boldsymbol{L}}, \Sigma^{\mathcal{G}}, \nu^{\boldsymbol{L}})$ of the Lévy process $\boldsymbol{L}$. Using the explicit transformation rules (4.2.3), one sees that the condition $\left\|G^{(h)}(s)\right\|_{L^\infty([0,1],\mathrm{Leb})} = O(h^2)$ implies that

$$\begin{aligned}
\left\|\gamma_G^{(h)}\right\| &= O(h^2),\\
\left\|\Sigma_G^{(h)}\right\| &= O(h^4),\\
\int_{\|\boldsymbol{x}\|<1} \|\boldsymbol{x}\|^r \nu_G^{(h)}(\mathrm{d}\boldsymbol{x}) &= O(h^{2r}), \quad r = 2,3\ldots,\\
\int_{\|\boldsymbol{x}\|\geqslant 1} \|\boldsymbol{x}\|^r \nu_G^{(h)}(\mathrm{d}\boldsymbol{x}) &= O(h^{2r}), \quad r = 2,\ldots,k,
\end{aligned}$$

so that we can apply Lemma 4.2 to conclude that $\mathbb{E}\left\|\int_{n-1}^n G(n-u)\mathrm{d}\boldsymbol{L}(u)\right\|^k = O(h^{2k})$. □

If one is willing to make the assumption that the jump part of the driving Lévy process has finite variation, then the norm of the approximation error $\varepsilon^{(h)}_{F\circ Y,n}$ can be bounded by path-wise defined quantities. In our forthcoming treatment of the general method of moments estimation in Section 4.6 we only rely on the moment bounds given in Proposition 4.19, but a path-wise understanding of the approximation error of convolution integrals of Lévy process might be useful in other contexts and is of interest in its own right. We will consider the three components in the Lévy–Itô decomposition (4.2.2) of a Lévy process separately.

For the Brownian part, the Hölder norms will play an important role, which, for any function $f : [a,b] \to M$ with values in a metric space (M,d), and any $0 < \alpha < 1$, are defined as

$$\|f\|_{\Lambda_\alpha([a,b])} = \sup_{s,t\in[a,b]} \frac{d(f(s),f(t))}{|s-t|^\alpha}.$$

Lemma 4.20 *Assume that $\boldsymbol{W}_{\mathcal{G}}$ is an m-dimensional Brownian motion with covariance matrix $\Sigma^{\mathcal{G}}$,* *and that* $F : [0,1] \to M_m(\mathbb{R})$ *is a twice continuously differentiable function. Then, for every* $\delta \in (0, 1/2)$*, there are constants* c_1, c_2 *such that*

$$\left\| \varepsilon^{(h)}_{F \circ \boldsymbol{W}_{\mathcal{G}},n} \right\| \leqslant h^{1/2-\delta} \left[c_1 \left\| \boldsymbol{W} \right\|_{L^\infty([n-1,n])} + c_2 \left\| \boldsymbol{W} \right\|_{\Lambda_{1/2-\delta}([n-1,n])} \right], \tag{4.5.20}$$

where $\boldsymbol{W} = \left(\Sigma^{\mathcal{G}}\right)^{-1/2} \boldsymbol{W}_{\mathcal{G}}$ *is a standard* m*-dimensional Brownian motion. Moreover, for every* $k > 0$*,*

$$\mathbb{E} \left\| \varepsilon^{(h)}_{F \circ \boldsymbol{W}_{\mathcal{G}},n} \right\|^k = O(h^{k(1/2-\delta)}), \quad \text{as } h \to 0. \tag{4.5.21}$$

Proof We note that it has been proven in Cruz-Uribe and Neugebauer (2002, Theorem 1.1) that for any α-Hölder continuous function $f : [a, b] \to \mathbb{R}$,

$$\left| \int_a^b f(s) \mathrm{d}s - T^K_{[a,b]} f \right| \leqslant \frac{(b-a)^{1+\alpha}}{(1+\alpha)(2K)^\alpha} \left\| f \right\|_{\Lambda_\alpha([a,b])}. \tag{4.5.22}$$

It is well known that, for any α strictly between 0 and $1/2$, paths of Brownian motion are α-Hölder continuous with probability one (Mörters and Peres, 2010, Corollary 1.20), and, consequently, so are the paths of $\boldsymbol{\mathcal{W}} : t \mapsto F(n-t) \left(\Sigma^{\mathcal{G}}\right)^{1/2} \boldsymbol{W}(t)$. We therefore only need to generalize Eq. (4.5.22) to vector valued functions to obtain

$$\begin{aligned} \left\| \varepsilon^{(h)}_{F \circ \boldsymbol{W}_{\mathcal{G}},n} \right\| = \left\| \int_{n-1}^n \boldsymbol{\mathcal{W}}(s) \mathrm{d}s - T^{h^{-1}}_{[n-1,n]} \boldsymbol{\mathcal{W}} \right\| &\leqslant \sqrt{m} \max_{i=1,\ldots,d} \left| \left[\int_{n-1}^n \boldsymbol{\mathcal{W}}(s) \mathrm{d}s - T^{h^{-1}}_{[n-1,n]} \boldsymbol{\mathcal{W}} \right]^i \right| \\ &\leqslant \frac{\sqrt{m}}{(1+\alpha)2^\alpha} h^\alpha \max_{i=1,\ldots,d} \sup_{s,t \in [a,b]} \frac{\left| \mathcal{W}^i(s) - \mathcal{W}^i(t) \right|}{|s-t|^\alpha} \\ &\leqslant \frac{\sqrt{m}}{(1+\alpha)2^\alpha} h^\alpha \left\| \boldsymbol{\mathcal{W}} \right\|_{\Lambda_\alpha([a,b])}. \end{aligned} \tag{4.5.23}$$

The α-Hölder norm of $\boldsymbol{\mathcal{W}}$ can be estimated as

$$\begin{aligned} &\left\| \boldsymbol{\mathcal{W}} \right\|_{\Lambda_\alpha([n-1,n])} \\ &= \sup_{s,t \in [n-1,n]} \frac{\left\| F(n-s) \left(\Sigma^{\mathcal{G}}\right)^{1/2} \boldsymbol{W}(s) - F(n-t) \left(\Sigma^{\mathcal{G}}\right)^{1/2} \boldsymbol{W}(t) \right\|}{|s-t|^\alpha} \\ &\leqslant \sup_{s,t \in [n-1,n]} \frac{\left\| [F(n-s) - F(n-t)] \left(\Sigma^{\mathcal{G}}\right)^{1/2} \boldsymbol{W}(s) \right\| + \left\| F(n-t) \left(\Sigma^{\mathcal{G}}\right)^{1/2} [\boldsymbol{W}(s) - \boldsymbol{W}(t)] \right\|}{|s-t|^\alpha} \\ &\leqslant \left\| F \right\|_{\Lambda_\alpha([0,1])} \left\| \left(\Sigma^{\mathcal{G}}\right)^{1/2} \right\| \left\| \boldsymbol{W} \right\|_{L^\infty([n-1,n])} + \left\| F \right\|_{L^\infty([0,1])} \left\| \left(\Sigma^{\mathcal{G}}\right)^{1/2} \right\| \left\| \boldsymbol{W} \right\|_{\Lambda_\alpha([n-1,n])}. \end{aligned} \tag{4.5.24}$$

By assumption, both $\left\| F \right\|_{\Lambda_\alpha([0,1])}$ and $\left\| F \right\|_{L^\infty([0,1])}$ are finite, so combining displays (4.5.23) and (4.5.24), and setting $\alpha = 1/2 - \delta$ proves Eq. (4.5.20) with

$$c_1 = \frac{\sqrt{m}}{(1+\alpha)2^\alpha} \|F\|_{\Lambda_\alpha([0,1])} \left\| \left(\Sigma^{\mathcal{G}}\right)^{1/2} \right\|, \quad c_2 = \frac{\sqrt{m}}{(1+\alpha)2^\alpha} \|F\|_{L^\infty([0,1])} \left\| \left(\Sigma^{\mathcal{G}}\right)^{1/2} \right\|.$$

To show the claim (4.5.21), it suffices to argue that the kth moments of the α-Hölder norm of $\mathcal{W}$ are finite. By the last display, these moments are bounded by

$$\mathbb{E} \|\mathcal{W}\|^k_{\Lambda_\alpha([n-1,n])} \leqslant \tilde{c}_1 \mathbb{E} \|\boldsymbol{W}\|^k_{L^\infty([n-1,n])} + \tilde{c}_2 \mathbb{E} \|\boldsymbol{W}\|^k_{\Lambda_\alpha([n-1,n])},$$

for some positive constants $\tilde{c}_1$ and $\tilde{c}_2$. The first term on the right is finite by Klenke (2008, Theorem 21.19). With respect to the second term it has been shown in Kwapień and Rosiński (2004, Example 3.3), using the theory of majorizing measures that, for a univariate Brownian motion W, there exists a constant C such that

$$\mathbb{E} \exp\left\{ C \sup_{s,t\in[0,1]} \frac{|W(t) - W(s)|^2}{|t-s| \log(\mathrm{e}/|t-s|)} \right\} < \infty.$$

We shall use this result to show that for any $\alpha \in (0, 1/2)$ there exists a constant C' such that the exponential moment $\mathbb{E} \exp\left\{ C' \|\boldsymbol{W}\|^2_{\Lambda_\alpha([0,1])} \right\}$ is finite, which implies that all ordinary moments of $\|\boldsymbol{W}\|_{\Lambda_\alpha([0,1])}$ are finite as well. For convenience, we write $\alpha = 1/2 - \delta$ for some $\delta \in (0, 1/2)$. We first note that, for all $\delta \in (0,1)$,

$$\frac{1}{d^{1-\delta}} \leqslant \frac{(\mathrm{e}\delta)^{-1}}{d \log(\mathrm{e}/d)}, \quad \forall d \in (0,1],$$

which can be shown by elementary calculus. Using the fact that the components of $\boldsymbol{W}$ are independent standard Brownian motions, denoted by W^i, it follows that

$$\begin{aligned}
\mathbb{E} \exp\left\{ C' \|\boldsymbol{W}\|^2_{\Lambda_{1/2-\delta}([0,1])} \right\} =& \mathbb{E} \exp\left\{ C' \sup_{s,t\in[0,1]} \frac{\|\boldsymbol{W}(t) - \boldsymbol{W}(s)\|^2}{|t-s|^{1-2\delta}} \right\} \\
\leqslant& \prod_{i=1}^m \mathbb{E} \exp\left\{ C' \sup_{s,t\in[0,1]} \frac{|W^i(t) - W^i(s)|^2}{|t-s|^{1-2\delta}} \right\} \\
\leqslant& \prod_{i=1}^m \mathbb{E} \exp\left\{ C'(2\mathrm{e}\delta)^{-1} \sup_{s,t\in[0,1]} \frac{|W^i(t) - W^i(s)|^2}{|t-s| \log(\mathrm{e}/|t-s|)} \right\} \\
<& \infty,
\end{aligned}$$

if C' is less than or equal to $2\mathrm{e}\delta C$. □

In the next lemma we derive a path-wise bound for $\left\|\boldsymbol{\varepsilon}_{F\circ f,n}^{(h)}\right\|$ for the case that the function f is a pure jump Lévy process. Every pure jump Lévy process $\boldsymbol{L}^{\mathrm{j}}$ has the representation

$$\boldsymbol{L}^{\mathrm{j}}(t) = \sum_{0<s\leqslant t} \delta\boldsymbol{L}^{\mathrm{j}}(s), \quad t \geqslant 0,$$

where $\delta\boldsymbol{L}^{\mathrm{j}}(s) = \boldsymbol{L}^{\mathrm{j}}(s) - \boldsymbol{L}^{\mathrm{j}}(s-)$ denotes the jump size of the process $\boldsymbol{L}^{\mathrm{j}}$ at time s, and $\boldsymbol{L}^{\mathrm{j}}(s-) = \lim_{r\nearrow s} \boldsymbol{L}^{\mathrm{j}}(r)$. If $\boldsymbol{L}^{\mathrm{j}}$ is of finite variation, the sum of jumps is even absolutely convergent.

Lemma 4.21 *Assume that $\boldsymbol{L}^{\mathrm{j}}$ is an m-dimensional pure jump Lévy process with Lévy measure $\nu^{\boldsymbol{L}}$ satisfying $\int_{\|\boldsymbol{x}\|\leqslant 1} \|\boldsymbol{x}\|\nu^{\boldsymbol{L}}(\mathrm{d}\boldsymbol{x}) < \infty$, and that $F : [0,1] \to M_m(\mathbb{R})$ is a twice continuously differentiable function. Then there is a constant C such that*

$$\left\|\boldsymbol{\varepsilon}_{F\circ\boldsymbol{L}^{\mathrm{j}},n}^{(h)}\right\| \leqslant Ch \sum_{n-1<s\leqslant n} \|\delta\boldsymbol{L}^{\mathrm{j}}(s)\|, \quad \delta\boldsymbol{L}^{\mathrm{j}}(s) = \boldsymbol{L}^{\mathrm{j}}(s) - \boldsymbol{L}^{\mathrm{j}}(s-). \tag{4.5.25}$$

If $\boldsymbol{L}^{\mathrm{j}}$ possesses a finite kth absolute moment, then

$$\mathbb{E}\left\|\boldsymbol{\varepsilon}_{F\circ\boldsymbol{L}^{\mathrm{j}},n}^{(h)}\right\|^k = O(h^k), \quad \text{as } h \to 0. \tag{4.5.26}$$

Moreover, there exists an i. i. d. sequence $\tilde{\boldsymbol{\varepsilon}}_{F\circ\boldsymbol{L}^{\mathrm{j}}}^{(h)}$ and a matrix $M^{(h)} \in M_m(\mathbb{R})$, such that

$$\boldsymbol{\varepsilon}_{F\circ\boldsymbol{L}^{\mathrm{j}},n}^{(h)} = \tilde{\boldsymbol{\varepsilon}}_{F\circ\boldsymbol{L}^{\mathrm{j}},n}^{(h)} + M^{(h)}\boldsymbol{L}^{\mathrm{j}}(n). \tag{4.5.27}$$

Proof It is enough to give the proof for the case $n = 1$. For some integer $1 \leqslant k \leqslant h^{-1} - 1$, we consider the interval $[kh, (k+1)h]$. Writing $\boldsymbol{L}^{\mathrm{j}}(s) = \sum_{0<t\leqslant s} \delta\boldsymbol{L}^{\mathrm{j}}(t)$ it is clear that

$$\begin{aligned} I_k &:= \int_{kh}^{(k+1)h} F(1-s)\boldsymbol{L}^{\mathrm{j}}(s)\mathrm{d}s \\ &= \int_{kh}^{(k+1)h} F(1-s)\mathrm{d}s\boldsymbol{L}^{\mathrm{j}}(kh) + \sum_{kh<t\leqslant(k+1)h} \delta\boldsymbol{L}^{\mathrm{j}}(t) \int_t^{(k+1)h} F(1-s)\mathrm{d}s \end{aligned}$$

The simple trapezoidal approximation of that integral on the other hand is given by

$$\begin{aligned} J_k :=& \frac{h}{2}\left[F(1-(k+1)h)\boldsymbol{L}^{\mathrm{j}}((k+1)h) + F(1-kh)\boldsymbol{L}^{\mathrm{j}}(kh)\right] \\ =& \frac{h}{2}F(1-(k+1)h) \sum_{kh<t\leqslant(k+1)h} \delta\boldsymbol{L}^{\mathrm{j}}(t) + \frac{F(1-(k+1)h) + F(1-kh)}{2} h\boldsymbol{L}^{\mathrm{j}}(kh), \end{aligned}$$

and it thus follows that

$$\|I_k - J_k\| = \left\| \left[\int_{kh}^{(k+1)h} F(1-s)\mathrm{d}s - \frac{F(1-(k+1)h) + F(1-kh)}{2}h\right] \boldsymbol{L}^{\mathfrak{i}}(kh) \right.$$
$$\left. + \sum_{kh<t\leqslant(k+1)h} \left[\int_{t}^{(k+1)h} F(1-s)\mathrm{d}s - \frac{h}{2}F(1-(k+1)h)\right] \delta\boldsymbol{L}^{\mathfrak{i}}(t)\right\|.$$

By Proposition 4.18 and the assumed differentiability of F, the error of the simple trapezoidal approximation in the first term in this expression is bounded by some constant C_1, which is independent of k, times h^2. For the second term we observe that for every $k = 1, \ldots, h^{-1}-1$ and every $t \in [kh, (k+1)h]$,

$$\begin{aligned}
&\left\| \int_{t}^{(k+1)h} F(1-s)\mathrm{d}s - \frac{h}{2}F(1-(k+1)h)\right\| \\
\leqslant & \int_{t}^{(k+1)h} \|F(1-s) - F(1-(k+1)h)\|\,\mathrm{d}s + \left[\left(k+\frac{1}{2}\right)h - t\right] \|F(1-(k+1)h)\| \\
\leqslant & \|F'\|_{L^\infty([0,1])} \int_{t}^{(k+1)h} [(k+1)h - s]\,\mathrm{d}s + \frac{h}{2}\|F\|_{L^\infty([0,1])} \\
= & \frac{1}{2}\|F'\|_{L^\infty([0,1])} [(k+1)h - t]^2 + \frac{h}{2}\|F\|_{L^\infty([0,1])} \\
\leqslant & \frac{1}{2}\|F'\|_{L^\infty([0,1])} h^2 + \frac{h}{2}\|F\|_{L^\infty([0,1])} \leqslant C_2 h, \quad \text{for all } h \leqslant 1,
\end{aligned}$$

where $C_2 = \frac{1}{2}[\|F\|_{L^\infty([0,1])} + \|F'\|_{L^\infty([0,1])}]$. The triangle inequality then implies that

$$\|I_k - J_k\| \leqslant C_1 h^2 \left\|\boldsymbol{L}^{\mathfrak{i}}(kh)\right\| + C_2 h \sum_{kh<t\leqslant(k+1)h} \left\|\delta\boldsymbol{L}^{\mathfrak{i}}(t)\right\|,$$

and the claim (4.5.25) follows from the following chain of inequalities:

$$\begin{aligned}
\left\|\boldsymbol{\varepsilon}^{(h)}_{F\circ\boldsymbol{L}^{\mathfrak{i}},1}\right\| = \left\|\sum_{k=1}^{h^{-1}-1}[I_k - J_k]\right\| &\leqslant \sum_{k=1}^{h^{-1}-1}\|I_k - J_k\| \\
&\leqslant C_1 h^2 \sum_{k=1}^{h^{-1}-1}\sum_{0<t\leqslant kh}\left\|\delta\boldsymbol{L}^{\mathfrak{i}}(t)\right\| + C_2 h \sum_{k=1}^{h^{-1}-1}\sum_{kh<t\leqslant(k+1)h}\|\delta\boldsymbol{L}^{\mathfrak{i}}(t)\| \\
&\leqslant C_1 h^2 \sum_{0<t\leqslant 1-h}\sum_{k=t/h}^{h^{-1}-1}\left\|\delta\boldsymbol{L}^{\mathfrak{i}}(t)\right\| + C_2 h \sum_{0<t\leqslant 1}\|\delta\boldsymbol{L}^{\mathfrak{i}}(t)\| \\
&\leqslant C_1 h \sum_{0<t\leqslant 1}[1-t]\left\|\delta\boldsymbol{L}^{\mathfrak{i}}(t)\right\| + C_2 h \sum_{0<t\leqslant 1}\|\delta\boldsymbol{L}^{\mathfrak{i}}(t)\| \\
&\leqslant C h \sum_{0<t\leqslant 1}\|\delta\boldsymbol{L}^{\mathfrak{i}}(t)\|,
\end{aligned}$$

for some constant C which only depends on C_1 and C_2. In order to quantify the absolute moments of the approximation error $\boldsymbol{\varepsilon}^{(h)}_{F\circ\boldsymbol{L}^{\mathrm{i}},1}$, we observe that the last display and the Lévy–Itô decomposition imply that

$$\begin{aligned}\mathbb{E}\left\|\boldsymbol{\varepsilon}^{(h)}_{F\circ\boldsymbol{L}^{\mathrm{i}},1}\right\|^k \leqslant& Ch^k\mathbb{E}\left(\sum_{0<t\leqslant 1}\|\delta\boldsymbol{L}^{\mathrm{i}}(t)\|\right)^k\\ =& Ch^k\int_0^1\cdot\ldots\cdot\int_0^1\int_{\mathbb{R}^d}\cdot\ldots\cdot\int_{\mathbb{R}^d}\|\boldsymbol{x}_1\|\cdot\ldots\cdot\|\boldsymbol{x}_k\|\mathbb{E}N(\mathrm{d}\boldsymbol{x}_1,\mathrm{d}s_1)\cdot\ldots\cdot N(\mathrm{d}\boldsymbol{x}_k,\mathrm{d}s_k).\end{aligned}$$

From here, the proof proceeds along the same lines as the proof of Lemma 4.2; in order not to repeat the same arguments, we only give a heuristic explanation here. Using the facts that the random variables $N(\mathrm{d}\boldsymbol{x}_i,\mathrm{d}s_i)$ and $N(\mathrm{d}\boldsymbol{x}_j,\mathrm{d}s_j)$ are independent for $s_i\neq s_j$, and that they satisfy $N(\mathrm{d}\boldsymbol{x}_i,\mathrm{d}s)N(\mathrm{d}\boldsymbol{x}_j,\mathrm{d}s)=\delta_{\boldsymbol{x}_i,\boldsymbol{x}_j}N(\mathrm{d}\boldsymbol{x}_i,\mathrm{d}s)$, it follows that the last integral can be decomposed according to which of the s_i are equal. We denote by $\mathcal{P}_k$ the set of partitions of $\{1,2,\ldots,k\}$; for $\pi\in\mathcal{P}_k$ we write $\#\pi$ for the number of blocks in π and the number of elements in such a block B is denoted by $|B|$. With these notations we can write the integral as

$$\sum_{\pi\in\mathcal{P}_k}\prod_{i=1}^{\#\pi}\mathbb{E}\int_0^1\int_{\mathbb{R}^d}\|\boldsymbol{x}_i\|^{|B_i|}N(\mathrm{d}\boldsymbol{x}_i,\mathrm{d}s).$$

The expectations in this expression are finite by the assumption of finite variation and finite kth moment of $\boldsymbol{L}^{\mathrm{i}}$. To prove Eq. (4.5.27), it is enough to note that

$$\begin{aligned}\boldsymbol{\varepsilon}^{(h)}_{F\circ\boldsymbol{L}^{\mathrm{i}},n} =& \int_{n-1}^n F(n-s)\left[\boldsymbol{L}^{\mathrm{i}}(s)-\boldsymbol{L}^{\mathrm{i}}(n)\right]\mathrm{d}s+\int_{n-1}^n F(n-s)\mathrm{d}s\boldsymbol{L}^{\mathrm{i}}(n)\\ &-h\left[\frac{F(0)\boldsymbol{L}^{\mathrm{i}}(n)-F(1)\boldsymbol{L}^{\mathrm{i}}(n-1)}{2}+\sum_{k=1}^{h^{-1}-1}F(1-kh)\boldsymbol{L}^{\mathrm{i}}(n-1+kh)\right].\end{aligned}\tag{4.5.28}$$

Introducing the approximation error

$$M^{(h)}=\int_{n-1}^n F(n-s)\mathrm{d}s-h\left[\frac{F(0)-F(1)}{2}+\sum_{k=1}^{h^{-1}-1}F(1-kh)\right],$$

which is clearly independent of n, Eq. (4.5.28) becomes

$$\begin{aligned}\boldsymbol{\varepsilon}^{(h)}_{F\circ\boldsymbol{L}^{\mathrm{i}},n} =& \int_{n-1}^n F(n-s)\left[\boldsymbol{L}^{\mathrm{i}}(s)-\boldsymbol{L}^{\mathrm{i}}(n)\right]\mathrm{d}s+M^{(h)}\boldsymbol{L}^{\mathrm{i}}(n)\\ &-h\left[\frac{-F(1)\left[\boldsymbol{L}^{\mathrm{i}}(n-1)-\boldsymbol{L}^{\mathrm{i}}(n)\right]}{2}+\sum_{k=1}^{h^{-1}-1}F(1-kh)\left[\boldsymbol{L}^{\mathrm{i}}(n-1+kh)-\boldsymbol{L}^{\mathrm{i}}(n)\right]\right].\end{aligned}$$

The sequence $\boldsymbol{\varepsilon}^{(h)}_{F\circ\boldsymbol{L}^{\mathrm{i}}}-M^{(h)}\boldsymbol{L}^{\mathrm{i}}(n)$ is therefore i. i. d. by the stationarity and the independence of the increments of $\boldsymbol{L}^{\mathrm{i}}$. □

Proposition 4.22 *Assume that* $\boldsymbol{L}$ *is a Lévy process with characteristic triplet* $(\gamma^{\boldsymbol{L}},\Sigma^{\boldsymbol{L}},\nu^{\boldsymbol{L}})$*. Let* $\boldsymbol{Y}$

be a d-dimensional $\boldsymbol{L}$-driven MCARMA(p,q) process, $F:[0,1]\to M_d(\mathbb{R})$ a twice continuously differentiable function, and denote by $\boldsymbol{\varepsilon}^{(h)}_{F\circ\boldsymbol{Y},n}$ the approximation error of the trapezoidal rule.

i) If $p-q\geqslant 3$, then there exists a constant C such that

$$\left\|\boldsymbol{\varepsilon}^{(h)}_{F\circ\boldsymbol{Y},n}\right\|\leqslant Ch^2\sum_{i=0,1,2}\left\|\mathrm{D}^i\boldsymbol{Y}\right\|_{L^\infty([n-1,n])}. \tag{4.5.29}$$

ii) If $p-q=2$, then there exists a constant C such that

$$\left\|\boldsymbol{\varepsilon}^{(h)}_{F\circ\boldsymbol{Y},n}\right\|\leqslant Ch\sum_{i=0,1}\left\|\mathrm{D}^i\boldsymbol{Y}\right\|_{L^\infty([n-1,n])}. \tag{4.5.30}$$

iii) If $p-q=1$ and $\int_{\|\boldsymbol{x}\|\leqslant 1}\|\boldsymbol{x}\|\,\nu^{\boldsymbol{L}}(d\boldsymbol{x})<\infty$, then, for every $0<\delta<1/2$, there exists a constant C such that

$$\begin{aligned}\left\|\boldsymbol{\varepsilon}^{(h)}_{F\circ\boldsymbol{Y},n}\right\|\leqslant & C\left[h\sum_{i=0,1}\left\|\mathrm{D}^i\boldsymbol{J}\right\|_{L^\infty([n-1,n])}+h\sum_{n-1<s\leqslant n}\left\|\delta\boldsymbol{L}^{\mathrm{i}}(s)\right\|\right.\\ &\left.+h^{1/2-\delta}\left[\|\boldsymbol{W}\|_{L^\infty([n-1,n])}+\|\boldsymbol{W}\|_{\Lambda_{1/2-\delta}([n-1,n])}\right]\right],\end{aligned} \tag{4.5.31}$$

where the function $\boldsymbol{J}$ is defined below in Eq. (4.5.32).

Proof If $p-q\geqslant 3$, the function $\boldsymbol{I}:s\mapsto F(n-s)\boldsymbol{Y}(s)$ is twice differentiable, so Proposition 4.18 implies that

$$\begin{aligned}\left\|\boldsymbol{\varepsilon}^{(h)}_{F\circ\boldsymbol{Y},n}\right\|\leqslant&\frac{\sqrt{d}}{12}h^2\sup_{t\in[n-1,n]}\left\|\mathrm{D}^2\boldsymbol{I}(t)\right\|\\ \leqslant&\frac{\sqrt{d}}{12}h^2\sup_{t\in[n-1,n]}\left[C_0\left\|\mathrm{D}^2\boldsymbol{Y}(t)\right\|+C_1\left\|\mathrm{D}^1\boldsymbol{Y}(t)\right\|+C_2\left\|\boldsymbol{Y}(t)\right\|\right],\end{aligned}$$

where for $i=0,1,2$, the number C_i is defined as $\left\|\mathrm{D}^iF\right\|_{L^\infty([n-1,n])}$. The claim thus follows with $C:=\sqrt{d}/12\max_{i=0.1,2}\{C_i\}$. If $p-q=2$, the paths of $\boldsymbol{I}$ are one time differentiable with discontinuous, yet bounded derivative, so they are in particular locally Lipschitz and we can apply Cruz-Uribe and Neugebauer (2002, Corollary 1.4), and obtain

$$\left|\left[\boldsymbol{\varepsilon}^{(h)}_{F\circ\boldsymbol{Y},n}\right]^i\right|\leqslant\frac{h}{8}\left[\sup_{n-1\leqslant t_i\leqslant n}\mathrm{D}I^i(t_i)-\inf_{n-1\leqslant t_i\leqslant n}\mathrm{D}I^i(t_i)\right],\quad i=1,\ldots.d,$$

for each component of $\boldsymbol{\varepsilon}^{(h)}_{F\circ\boldsymbol{Y},n}$. This implies that

$$\left\|\boldsymbol{\varepsilon}^{(h)}_{F\circ\boldsymbol{Y},n}\right\|\leqslant\sqrt{d}\max_{i=1,\ldots,d}\left|\left[\boldsymbol{\varepsilon}^{(h)}_{F\circ\boldsymbol{Y},n}\right]^i\right|$$

$$\leqslant \frac{h\sqrt{d}}{8}\max_{i=1,\ldots,d}\left[\sup_{n-1\leqslant t_i\leqslant n}\mathrm{D}I^i(t_i)-\inf_{n-1\leqslant t_i\leqslant n}\mathrm{D}I^i(t_i)\right]$$
$$\leqslant \frac{h\sqrt{d}}{4}\max_{i=1,\ldots,d}\sup_{n-1\leqslant t_1\leqslant n}\left|\mathrm{D}I^i(t_i)\right|\leqslant\frac{h\sqrt{d}}{4}\sup_{n-1\leqslant t\leqslant n}\left\|\mathrm{D}\boldsymbol{I}(t)\right\|.$$

Noting that $\|\mathrm{D}\boldsymbol{I}(t)\|\leqslant C_0\|\mathrm{D}\boldsymbol{Y}(t)\|+C_1\|\boldsymbol{Y}(t)\|$, the assertion of ii) follows with C equal to $\sqrt{d}/4\max_{i=0,1}\{C_i\}$. If $p-q=1$, the paths of $\boldsymbol{Y}$, and thus of $\boldsymbol{I}$, are in general, not differentiable. Using the abbreviation $\underline{B}_q=\left[\begin{array}{ccc}B_0 & \ldots & B_{q-1}\end{array}\right]$, we can, however, write

$$\begin{aligned}\boldsymbol{Y}(t)=&\underline{B}_q\boldsymbol{X}_q(t)+B_q\boldsymbol{X}^{(p)}(t)\\=&\underline{B}_q\boldsymbol{X}_q(t)+B_q\boldsymbol{X}^{(p)}(0)-B_q\underline{A}\int_0^t\boldsymbol{X}(s)\mathrm{d}s+B_q\boldsymbol{L}(t)\\=&\underline{B}_q\boldsymbol{X}_q(t)+B_q\boldsymbol{X}^{(p)}(0)-B_q\underline{A}\int_0^t\boldsymbol{X}(s)\mathrm{d}s+B_q\boldsymbol{\gamma}^{\boldsymbol{L}}t+B_q\Sigma^{\boldsymbol{L}}\boldsymbol{W}(t)+B_q\boldsymbol{L}^{\mathrm{i}}(t),\end{aligned}\qquad\square$$

where the last line is an application of the Lévy–Itô decomposition. We can now analyse how the different terms in the last expression contribute to the approximation error $\boldsymbol{\varepsilon}^{(h)}_{F\circ\boldsymbol{Y}}$. The function

$$\boldsymbol{J}:t\mapsto\underline{B}_q\boldsymbol{X}_q(t)+B_q\boldsymbol{X}^{(p)}(0)-B_q\underline{A}\int_0^t\boldsymbol{X}(s)\mathrm{d}s+B_q\boldsymbol{\gamma}^{\boldsymbol{L}}t\tag{4.5.32}$$

is differentiable, and one therefore sees by the same argument used to prove ii) that its contribution to $\left\|\boldsymbol{\varepsilon}^{(h)}_{F\circ\boldsymbol{Y},n}\right\|$ is bounded by a constant multiple of h times $\sum_{i=0,1}\left\|\mathrm{D}^i\boldsymbol{J}\right\|_{L^\infty([n-1,n])}$. In Lemma 4.20 it has been shown that the contribution from the Gaussian part is of order $h^{1/2-\delta}$ times the sum of the Hölder and the supremum norms of $\boldsymbol{W}$ on the interval $[n-1,n]$. Finally, the approximation error for the pure jump Lévy process $\boldsymbol{L}^{\mathrm{i}}$ with summable small jumps has been shown in Lemma 4.21 to be bounded by a constant multiple of h times the sum of the absolute sizes of the jumps in the integration interval. Combining these partial results proves part iii) of the proposition.

It remains to estimate $\boldsymbol{X}_q(n)$. In view of the AR(1) structure given in Eq. (4.4.7), we compute estimates

$$\hat{\boldsymbol{X}}^{(h)}_q(n)=\mathrm{e}^{\mathrm{B}}\hat{\boldsymbol{X}}^{(h)}_q(n-1)+\hat{I}^{(h)}_n,\quad\hat{\boldsymbol{X}}^{(h)}_q(0)=\hat{\boldsymbol{X}}^{(h)}_{q,0},\quad n\geqslant 1,\tag{4.5.33}$$

where $\hat{I}^{(h)}_n=T^{h^{-1}}_{[n-1,n]}\mathrm{e}^{\mathrm{B}(n-\cdot)}E_q\boldsymbol{Y}(\cdot)$ is the trapezoidal rule approximation to the integral $\int_{n-1}^n\mathrm{e}^{\mathrm{B}(n-s)}E_q\boldsymbol{Y}(s)\mathrm{d}s$, and $\hat{\boldsymbol{X}}^{(h)}_{q,0}$ is a deterministic or random initial value. We introduce the notation

$$\boldsymbol{e}^{(h)}_{\boldsymbol{X},n}=\hat{\boldsymbol{X}}_q(n)-\boldsymbol{X}_q(n).\tag{4.5.34}$$

It is easy to see that the sequence $\boldsymbol{e}^{(h)}_{\boldsymbol{X}}$ satisfies $\boldsymbol{e}^{(h)}_{\boldsymbol{X},n}=\mathrm{e}^{\mathrm{B}}\boldsymbol{e}^{(h)}_{\boldsymbol{X},n-1}+\boldsymbol{\varepsilon}^{(h)}_{F\circ\boldsymbol{Y},n}$, $n\in\mathbb{N}$, where $F:t\mapsto\mathrm{e}^{\mathrm{B}t}E_q$ and $\boldsymbol{\varepsilon}^{(h)}_{F\circ\boldsymbol{Y},n}$ is of the form analysed in Proposition 4.19. For the following result we recall the notion of an absolutely continuous measure. By Lebesgue's decomposition theorem (Klenke, 2008, Theorem 7.33), every measure μ on $\mathbb{R}^m$ can be uniquely decomposed as $\mu = \mu_c + \mu_s$, where μ_c and μ_s are absolutely continuous and singular, respectively, with respect to m-dimensional Lebesgue measure. If μ_c is not the zero measure, we say that μ has a non-trivial absolutely continuous component.

Proposition 4.23 *Assume that $\boldsymbol{L}$ is a Lévy process. Let $\boldsymbol{Y}$ be a d-dimensional $\boldsymbol{L}$-driven MCARMA process satisfying Assumptions A1 and A2. The sequence $\boldsymbol{e}_{\boldsymbol{X}}^{(h)}$ defined by Eqs. (4.5.33) and (4.5.34) converges almost surely to a stationary and ergodic sequence which is independent of $\hat{\boldsymbol{X}}_{q,0}^{(h)}$. If, for some integer k, $\mathbb{E}\left\|\boldsymbol{L}(1)\right\|^k$ is finite, then the absolute moment $\mathbb{E}\left\|\boldsymbol{e}_{\boldsymbol{X},n}^{(h)}\right\|^k$ is of order $O(h^{2k})$ as $h \to 0$. If, moreover, the distribution of the random variable*

$$\int_0^1 \widetilde{G}(1-s)\mathrm{d}\boldsymbol{L}(s), \quad \widetilde{G}(s) = \begin{pmatrix} G(s)^T & (\exp(\mathrm{A}s)E_p)^T \end{pmatrix}^T, \tag{4.5.35}$$

where $G(s)$ is defined in Eq. (4.5.18), has a non-trivial absolutely continuous component, then the process $\boldsymbol{e}_{\boldsymbol{X}}^{(h)}$ is exponentially strongly mixing.

Proof We first observe that

$$\boldsymbol{e}_{\boldsymbol{X},n}^{(h)} = \mathrm{e}^{(n-1)\mathrm{B}}\boldsymbol{e}_{\boldsymbol{X},1}^{(h)} + \sum_{\nu=0}^{n-2} \mathrm{e}^{\nu\mathrm{B}}\boldsymbol{\varepsilon}_{\mathrm{F}\circ\boldsymbol{Y},n-\nu}^{(h)}, \quad n \geqslant 1,$$

and define the sequence $\tilde{\boldsymbol{e}}_{\boldsymbol{X}}^{(h)}$ by

$$\tilde{\boldsymbol{e}}_{\boldsymbol{X},n}^{(h)} = \sum_{\nu=0}^{\infty} \mathrm{e}^{\nu\mathrm{B}}\boldsymbol{\varepsilon}_{\mathrm{F}\circ\boldsymbol{Y},n-\nu}^{(h)}, \quad n \in \mathbb{Z}.$$

By this definition, $\tilde{\boldsymbol{e}}_{\boldsymbol{X}}^{(h)}$ is obviously independent of $\hat{\boldsymbol{X}}_{q,0}^{(h)}$. Since $\boldsymbol{\varepsilon}_{\mathrm{F}\circ\boldsymbol{Y}}^{(h)}$ is strongly mixing by Proposition 4.19, it is in particular ergodic (Klenke, 2008, Exercise 20.5.1). The sequence $\tilde{\boldsymbol{e}}_{\boldsymbol{X}}^{(h)}$ is the unique stationary solution of the AR(1) equations

$$\tilde{\boldsymbol{e}}_{\boldsymbol{X},n}^{(h)} = \mathrm{e}^{\mathrm{B}}\tilde{\boldsymbol{e}}_{\boldsymbol{X},n-1}^{(h)} + \boldsymbol{\varepsilon}_{\mathrm{F}\circ\boldsymbol{Y},n}^{(h)}, \quad n \in \mathbb{Z},$$

and an application of Krengel (1985, Theorem 4.3) to the infinite-order moving average representation of $\tilde{\boldsymbol{e}}_{\boldsymbol{X}}^{(h)}$ shows that this last sequence is ergodic as well. It remains to prove that $\boldsymbol{e}_{\boldsymbol{X},n}^{(h)}$ converges to $\tilde{\boldsymbol{e}}_{\boldsymbol{X},n}^{(h)}$ almost surely as $n \to \infty$. This follows from

$$\left\|\boldsymbol{e}_{\boldsymbol{X},n}^{(h)} - \tilde{\boldsymbol{e}}_{\boldsymbol{X},n}^{(h)}\right\| \leqslant \left\|\mathrm{e}^{(n-1)\mathrm{B}}\right\| \left\|\boldsymbol{e}_{\boldsymbol{X},1}^{(h)}\right\| + \left\|\sum_{\nu=n-1}^{\infty} \mathrm{e}^{\nu\mathrm{B}}\boldsymbol{\varepsilon}_{\mathrm{F}\circ\boldsymbol{Y},n-\nu}^{(h)}\right\|,$$

the fact that by Lemma 4.9 the eigenvalues of the matrix B have strictly negative real parts, and the almost sure convergence of the last sum (Brockwell and Davis, 1991, Proposition 3.1.1). For the proof that the kth moments of $\boldsymbol{e}_{\boldsymbol{X},n}^{(h)}$ are of order $O(h^{2k})$ we use the following generalization of Hölder's inequality, which can be proved by induction: for any k random variables $Z_1, \ldots, Z_k$ and positive numbers $p_1, \ldots, p_k$ such that $\sum 1/p_i = 1$, it holds that

$$\mathbb{E}\left(Z_1 \cdot \ldots \cdot Z_k\right) \leqslant \prod_{i=1}^{k} \left(\mathbb{E} Z_i^{p_i}\right)^{1/p_i}. \tag{4.5.36}$$

Choosing $p_i = 1/k$, $i = 1, \ldots, k$, and using that, by Proposition 4.19, there exists a constant C, independent of n, such that $\mathbb{E}\left\|\boldsymbol{\varepsilon}_{F\circ\boldsymbol{Y},n}^{(h)}\right\|^k \leqslant Ch^{2k}$, it follows that

$$\begin{aligned}\mathbb{E}\left\|\tilde{\boldsymbol{e}}_{\boldsymbol{X},n}^{(h)}\right\|^k \leqslant& \sum_{\nu_1=0}^{\infty} \cdot \ldots \cdot \sum_{\nu_k=0}^{\infty} \left\|\mathrm{e}^{\nu_1 \mathrm{B}}\right\| \cdot \ldots \cdot \left\|\mathrm{e}^{\nu_k \mathrm{B}}\right\| \mathbb{E}\left(\left\|\boldsymbol{\varepsilon}_{F\circ\boldsymbol{Y},n-\nu_1}^{(h)}\right\| \cdot \ldots \cdot \left\|\boldsymbol{\varepsilon}_{F\circ\boldsymbol{Y},n-\nu_k}^{(h)}\right\|\right)\\ \leqslant& Ch^{2k}\left(\sum_{\nu=0}^{\infty}\left\|\mathrm{e}^{\nu \mathrm{B}}\right\|\right)^k,\end{aligned}$$

which is of order $O(h^{2k})$ because the sum is finite due to the eigenvalues of B having strictly negative real parts. In order to show that the sequence $\boldsymbol{e}_{\boldsymbol{X}}^{(h)}$ is strongly mixing, we note that the stacked process $\begin{pmatrix} \boldsymbol{e}_{\boldsymbol{X}}^{(h)T} & \boldsymbol{X}^T \end{pmatrix}^T$ satisfies the AR(1) equations

$$\begin{aligned}\begin{pmatrix} \boldsymbol{e}_{\boldsymbol{X},n}^{(h)} \\ \boldsymbol{X}(n) \end{pmatrix} &= \begin{pmatrix} \mathrm{e}^{\mathrm{B}} & \Gamma \\ 0 & \mathrm{e}^{A} \end{pmatrix} \begin{pmatrix} \boldsymbol{e}_{\boldsymbol{X},n-1}^{(h)} \\ \boldsymbol{X}(n-1) \end{pmatrix} + \boldsymbol{Z}_n, \quad n \in \mathbb{Z},\\ \boldsymbol{Z}_n &= \int_{n-1}^{n} \begin{pmatrix} G(n-s) \\ \mathrm{e}^{A(n-s)} E_p \end{pmatrix} \mathrm{d}\boldsymbol{L}(s), \quad n \in \mathbb{Z},\end{aligned}$$

where $(\boldsymbol{Z}_n)_{n\in\mathbb{Z}}$ is an i. i. d. noise sequence. An extension of the arguments leading to Mokkadem (1988, Theorem 1), which is detailed in the proof of Theorem 2.8, shows that ARMA, and in particular, AR(1) processes are strongly mixing with exponentially decaying mixing coefficients if the driving noise sequence has a non-trivial absolutely continuous component, which is precisely what is assumed in the proposition. □

Remark 4.24 Sufficient conditions for the assumption made in the previous proposition to hold can be obtained from the observation that the random variable $\int_0^1 \widetilde{G}(1-s)\mathrm{d}\boldsymbol{L}(s)$ is infinitely divisible, and that its characteristic triplet can be obtained as in Eqs. (4.2.3). Sufficient conditions for an infinitely divisible random variable to be absolutely continuous, in terms of its characteristic triplet, can be found in Tucker (1965) and Sato (1999, Section 27). Since mixing is not our primary concern in this chapter, and our results hold without it, we do not pursue this issue further here.

4.5.3. Approximation of the increments ΔL_n

If we combine what we have so far, it follows that we can obtain estimates $\widehat{\Delta \boldsymbol{L}}_n$ of the increments of the Lévy process $\boldsymbol{L}$ by discretizing Eq. (4.4.6), that is

$$\begin{aligned}\widehat{\Delta \boldsymbol{L}}_n^{(h)} &= \sum_{\nu=0}^{p-q-1}\left[E_q^T \mathrm{B}^{p-q-1-\nu}E_q + \sum_{k=\nu}^{p-q-2} A_{p-q-k-1}E_q^T\mathrm{B}^{k-\nu}E_q\right]\left[\Delta_h^\nu[\boldsymbol{Y}](n) - \Delta_h^\nu[\boldsymbol{Y}](n-1)\right]\\ &\quad + \left[\underline{A}_q \mathrm{B}^{-1} + \sum_{k=1}^{p-q} A_{p-q-k+1}E_q^T\mathrm{B}^{k-1} + E_q^T\mathrm{B}^{p-q}\right]\left[\hat{\boldsymbol{X}}_q^{(h)}(n) - \hat{\boldsymbol{X}}_q^{(h)}(n-1)\right]\\ &\quad + A_p\left[B_q^{\sim 1}B_0\right]^{-1}B_q^{\sim 1}T_{[n-1,n]}^{h^{-1}}\boldsymbol{Y},\end{aligned} \tag{4.5.37}$$

where the forward differences $\Delta_h^\nu[\boldsymbol{Y}](n)$ are defined in Eq. (4.5.1), the estimates $\hat{\boldsymbol{X}}_q^{(h)}$ are computed recursively by Eq. (4.5.33), and the formula for the trapezoidal approximation $T_{[n-1,n]}^{h^{-1}}\boldsymbol{Y}$ is given in Eq. (4.5.12). Writing

$$\widehat{\Delta \boldsymbol{L}}_n^{(h)} = \Delta \boldsymbol{L}_n + \boldsymbol{\varepsilon}_n^{(h)}, \tag{4.5.38}$$

the approximation error $\boldsymbol{\varepsilon}_n^{(h)}$ is given by

$$\begin{aligned}\boldsymbol{\varepsilon}_n^{(h)} &= \sum_{\nu=0}^{p-q-1}\left[E_q^T \mathrm{B}^{p-q-1-\nu}E_q + \sum_{k=\nu}^{p-q-2} A_{p-q-k-1}E_q^T\mathrm{B}^{k-\nu}E_q\right]\left[\boldsymbol{e}_{\boldsymbol{Y},n}^{\nu,(h)} - \boldsymbol{e}_{\boldsymbol{Y},n-1}^{\nu,(h)}\right]\\ &\quad + \left[\underline{A}_q \mathrm{B}^{-1} + \sum_{k=1}^{p-q} A_{p-q-k+1}E_q^T\mathrm{B}^{k-1} + E_q^T\mathrm{B}^{p-q}\right]\left[\boldsymbol{e}_{\boldsymbol{X},n}^{(h)} - \boldsymbol{e}_{\boldsymbol{X},n-1}^{(h)}\right]\\ &\quad + A_pB_0^{-1}B_{q-1}\boldsymbol{\varepsilon}_{\boldsymbol{Y},n}^{(h)}.\end{aligned}$$

The following theorem summarizes the results of the previous two subsections about the probabilistic properties of the sequence of approximation errors $\boldsymbol{\varepsilon}^{(h)}$, both for fixed values of h and as h tends to zero.

Theorem 4.25 (Properties of $\widehat{\Delta \boldsymbol{L}}_n^{(h)}$) *Assume that* $\boldsymbol{L}$ *is a Lévy process, and that* $\boldsymbol{Y}$ *is an* $\boldsymbol{L}$*-driven multivariate CARMA process given by the state space representation (4.3.10) and satisfying Assumptions A1 and A2. Denote by* $\Delta \boldsymbol{L}_n = \boldsymbol{L}(n) - \boldsymbol{L}(n-1)$ *the unit increments of* $\boldsymbol{L}$ *and by* $\widehat{\Delta \boldsymbol{L}}_n^{(h)}$ *the estimates of the unit increments of* $\boldsymbol{L}$ *obtained from Eq. (4.5.38). The stochastic process* $\boldsymbol{\varepsilon}^{(h)} = \widehat{\Delta \boldsymbol{L}}^{(h)} - \Delta \boldsymbol{L}$ *has the following properties:*

i) There exists a stationary, ergodic stochastic process $\tilde{\boldsymbol{\varepsilon}}^{(h)}$ *such that* $\left\|\boldsymbol{\varepsilon}_n^{(h)} - \tilde{\boldsymbol{\varepsilon}}_n^{(h)}\right\| \to 0$ *almost surely as* $n \to \infty$*. If the random variable defined in Eq. (4.5.35) has a non-trivial absolutely continuous component with respect to the Lebesgue measure, then* $\boldsymbol{\varepsilon}^{(h)}$ *is exponentially strongly mixing.*

ii) *If* $\mathbb{E}\left\|\boldsymbol{L}(1)\right\|^{(k)_0} < \infty$*, for some positive integer* k*, then there exists a constant* $C > 0$ *such that, for* $\kappa = 1, \ldots, k$ *and* $0 < h \leqslant 1$*,*

$$\sup_{n\in\mathbb{N}} \mathbb{E}\left\|\boldsymbol{\varepsilon}_n^{(h)}\right\|^{\kappa} \leqslant Ch^{1/2}. \tag{4.5.39}$$

Proof Both claims follow directly from Propositions 4.17, 4.19 and 4.23. □

For the purpose of estimating a parametric model of the Lévy process $\boldsymbol{L}$ based on the noisy observations $\widehat{\Delta\boldsymbol{L}}^{(h)}$, it is important not only to have a sound quantitative understanding of the extent to which the true increments $\Delta\boldsymbol{L}$ differ from the estimated increments $\widehat{\Delta\boldsymbol{L}}$, but also to know how strongly this difference is affected when a function is applied to the increments. This issue is investigated in the next lemma.

Lemma 4.26 *Let* $f : \mathbb{R}^m \to \mathbb{R}^q$ *be a function with bounded* k*th derivative, and let* l *be some fixed positive integer. Assume that* $\mathbb{E}\left\|\boldsymbol{L}(1)\right\|^{(kl)_0} < \infty$*, and further that, for any integer* $1 \leqslant r \leqslant k-1$ *and any integers* $1 \leqslant i_1, \ldots, i_r \leqslant m$*, the moments of the partial derivatives of* f *satisfy*

$$\mathbb{E}\left\|\partial_{i_1}\cdots\partial_{i_r} f\left(\boldsymbol{L}(1)\right)\right\|^{kl} < \infty. \tag{4.5.40}$$

It then holds that

$$\sup_{n\in\mathbb{N}} \mathbb{E}\left\|f\left(\widehat{\Delta\boldsymbol{L}}_n^{(h)}\right) - f\left(\Delta\boldsymbol{L}_n\right)\right\|^{l} = O(h^{1/2}), \tag{4.5.41}$$

where the approximate increments $\widehat{\Delta\boldsymbol{L}}_n^{(h)}$ *are defined by Eq. (4.5.37).*

Proof By Taylor's theorem (Apostol, 1974, Theorem 12.14) we have that

$$\begin{aligned} f\left(\widehat{\Delta\boldsymbol{L}}_n^{(h)}\right) - f\left(\Delta\boldsymbol{L}_n\right) =& f\left(\Delta\boldsymbol{L}_n + \boldsymbol{\varepsilon}_n^{(h)}\right) - f\left(\Delta\boldsymbol{L}_n\right) \\ =& \sum_{r=1}^{k-1}\frac{1}{r!}\mathrm{d}^{(r)}f\left(\Delta\boldsymbol{L}_n\right)\left(\boldsymbol{\varepsilon}_n^{(h)}\right)^r + R\left(\Delta\boldsymbol{L}_n;\boldsymbol{\varepsilon}_n^{(h)}\right), \end{aligned}$$

where

$$\mathrm{d}^{(r)}f\left(\Delta\boldsymbol{L}_n\right)\left(\boldsymbol{\varepsilon}_n^{(h)}\right)^r = \sum_{i_1=1}^{m}\cdots\sum_{i_r=1}^{m}\partial_{i_1}\cdots\partial_{i_r}f\left(\Delta\boldsymbol{L}_n\right)\varepsilon_n^{(h),i_1}\cdots\varepsilon_n^{(h),i_r}$$

defines the action of the rth derivative of f. We note that

$$\left\|\mathrm{d}^{(r)}f\left(\Delta\boldsymbol{L}_n\right)\left(\boldsymbol{\varepsilon}_n^{(h)}\right)^r\right\| \leqslant \sum_{i_1=1}^{m}\cdots\sum_{i_r=1}^{m}\left\|\partial_{i_1}\cdots\partial_{i_r}f\left(\Delta\boldsymbol{L}_n\right)\right\|\left\|\boldsymbol{\varepsilon}_n^{(h)}\right\|^r =: \left\|\mathrm{d}^{(r)}f\left(\Delta\boldsymbol{L}_n\right)\right\|\left\|\boldsymbol{\varepsilon}_n^{(h)}\right\|^r,$$

and assumption (4.5.40) implies that

$$\mathbb{E}\left\|\mathrm{d}^{(r)}f\left(\Delta\boldsymbol{L}_n\right)\right\|^{kl} \leqslant m^{rlk}\sum_{i_1=1}^{m}\cdots\sum_{i_r=1}^{m}\mathbb{E}\left\|\partial_{i_1}\cdots\partial_{i_r}f\left(\Delta\boldsymbol{L}_n\right)\right\|^{kl}$$
$$<\infty.$$

It follows from the boundedness of the kth derivative of f that the remainder $R\left(\Delta\boldsymbol{L}_n;\boldsymbol{\varepsilon}_n^{(h)}\right)$ satisfies

$$\left\|R\left(\Delta\boldsymbol{L}_n;\boldsymbol{\varepsilon}_n^{(h)}\right)\right\| \leqslant C\left\|\boldsymbol{\varepsilon}_n^{(h)}\right\|^{k},$$

for some constant C. In particular,

$$\begin{aligned}
&\mathbb{E}\left\|f\left(\widehat{\Delta\boldsymbol{L}}_n^{(h)}\right)-f\left(\Delta\boldsymbol{L}_n\right)\right\|^{l}\\
\leqslant&2^l\,\mathbb{E}\left(\sum_{r=1}^{k-1}\frac{1}{r!}\left\|\mathrm{d}^{(r)}f\left(\Delta\boldsymbol{L}_n\right)\left(\boldsymbol{\varepsilon}_n^{(h)}\right)^{r}\right\|\right)^{l}+2^l\,\mathbb{E}\left\|R\left(\Delta\boldsymbol{L}_n;\boldsymbol{\varepsilon}_n^{(h)}\right)\right\|^{l}\\
\leqslant&2^l\sum_{r_1=1}^{k-1}\cdot\ldots\cdot\sum_{r_l=1}^{k-1}\frac{1}{r_1!\cdot\ldots\cdot r_l!}\mathbb{E}\left(\left\|\mathrm{d}^{(r_1)}f\left(\Delta\boldsymbol{L}_n\right)\right\|\cdot\ldots\cdot\left\|\mathrm{d}^{(r_l)}f\left(\Delta\boldsymbol{L}_n\right)\right\|\left\|\boldsymbol{\varepsilon}_n^{(h)}\right\|^{r_1+\ldots+r_l}\right)\\
&\qquad+C2^l\,\mathbb{E}\left\|\boldsymbol{\varepsilon}_n^{(h)}\right\|^{kl}.
\end{aligned}$$

By Theorem 4.25, the assumption that $\boldsymbol{L}(1)$ has a finite $(kl)_0$th absolute moment implies that $\mathbb{E}\left\|\boldsymbol{\varepsilon}_n^{(h)}\right\|^{\kappa}$ is of order $O(h^{\kappa/\kappa_0})$ as $h\to 0$ for all $1\leqslant\kappa\leqslant k$, where the constant implicit in the $O(\cdot)$ notation does not depend on n. It thus follows by an application of the generalized Hölder inequality (4.5.36) with exponents $p_1=\ldots=p_l=kl$, $p_{l+1}=k/(k-1)$ that

$$\begin{aligned}
&\mathbb{E}\left\|f\left(\widehat{\Delta\boldsymbol{L}}_n^{(h)}\right)-f\left(\Delta\boldsymbol{L}_n\right)\right\|^{l}\\
\leqslant&\sum_{r_1,\ldots,r_l=1}^{k-1}\left[\prod_{i=1}^{l}\frac{2}{r_i!}\mathbb{E}\left(\left\|d^{(r_i)}f\left(\Delta\boldsymbol{L}_n\right)\right\|^{kl}\right)^{\frac{1}{kl}}\right]\underbrace{\mathbb{E}\left(\left\|\boldsymbol{\varepsilon}_n^{(h)}\right\|^{\frac{(r_1+\ldots+r_l)k}{k-1}}\right)^{\frac{k-1}{k}}}_{=O\left(h^{\frac{r_1+\ldots+r_l}{[(r_1+\ldots+r_l)k/(k-1)]_0}}\right)}+O\left(h^{\frac{kl}{(kl)_0}}\right).
\end{aligned}$$

Since for any $\alpha\in[0,2]$ and any positive integer r it holds that $(r\alpha)_0\leqslant r\alpha_0$, the dominating term in this sum is the one corresponding to $r_1=\ldots=r_l=1$, which is of order $O(h^{1/2})$. Thus Eq. (4.5.41) is shown. □

4.6. Generalized method of moments estimation with noisy data

In this section we consider the problem of estimating a parametric model $\mathbb{P}_{\boldsymbol{\vartheta}}$ if only a disturbed i. i. d. sample of the true distribution is available. More precisely, assume that Θ is some parameter space, that $(\mathbb{P}_{\boldsymbol{\vartheta}} : \boldsymbol{\vartheta} \in \Theta)$ is a family of probability distributions on $\mathbb{R}^m$, and that

$$X^N = (X_1, \ldots, X_N), \quad \mathbb{R}^m \ni X_n \sim \mathbb{P}_{\boldsymbol{\vartheta}_0}, \tag{4.6.1}$$

is an i. i. d. sample from $\mathbb{P}_{\boldsymbol{\vartheta}_0}$. The classical generalized method of moments (abbreviated as GMM) is a well-established procedure for estimating the value of $\boldsymbol{\vartheta}_0$ from the observations X^N, see for instance Hall (2005); Hansen (1982); Newey and McFadden (1994) for a general introduction. After introducing some relevant notation and taking a closer look at two particularly important special cases of this class of estimators, we state the result about the consistency and asymptotic normality of GMM estimators for easy reference in Theorem 4.27. Our goal in this section is to extend this result to the situation where the sample X^N from the distribution $\mathbb{P}_{\boldsymbol{\vartheta}_0}$ cannot be observed directly. Instead, we assume that, for each, $h > 0$ there is a stochastic process $\varepsilon^{(h)}$ not necessarily independent of X^N, which we think of as a disturbance to the i. i. d. sample X^N, and the value of $\boldsymbol{\vartheta}_0$ is to be estimated from the observations $(X_1 + \varepsilon_1^{(h)}, \ldots, X_N + \varepsilon_N^{(h)})$. In Theorem 4.28 we prove under a mild moment assumption that the asymptotic properties of the GMM estimator, as N becomes large and h becomes small, are not altered by the inclusion of the noise process $\varepsilon^{(h)}$. Finally, we use this result in Theorem 4.34 to answer the question of how to estimate a parametric model for the driving Lévy process of a multivariate CARMA process from high-frequency discrete-time observations.

Underlying the construction of any GMM estimator is the existence of a function $g : \mathbb{R}^m \times \Theta \to \mathbb{R}^q$ such that for $X_1 \sim \mathbb{P}_{\boldsymbol{\vartheta}_0}$,

$$\mathbb{E}g\left(X_1, \boldsymbol{\vartheta}\right) = 0 \Leftrightarrow \boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0. \tag{4.6.2}$$

The analogy principle, that is the philosophy that unknown population averages should be approximated by sample averages, then suggests that an estimator $\hat{\boldsymbol{\vartheta}}^N$ of $\boldsymbol{\vartheta}_0$ based on the sample X^N, given by Eq. (4.6.1), can be defined as

$$\hat{\boldsymbol{\vartheta}}^N = \operatorname{argmin}_{\boldsymbol{\vartheta}\in\Theta} \left\| \frac{1}{N} \sum_{n=1}^{N} g\left(X_n, \boldsymbol{\vartheta}\right) \right\|_{W_N}, \tag{4.6.3}$$

where W_N is a positive definite, possibly data-dependent, $q \times q$ matrix defining the norm

$$\|\cdot\|_{W_N} : \mathbb{R}^q \to \mathbb{R}^+, \quad \|x\|_{W_N} = (\boldsymbol{x}^T W_N \boldsymbol{x})^{1/2}, \quad \boldsymbol{x} \in \mathbb{R}^q.$$

As we will see shortly, the choice of W_N influences the asymptotic variance Σ of the estimator $\hat{\boldsymbol{\vartheta}}^N$, given in Eq. (4.6.5). The optimal choice of weighting matrices W_N is described in Corollary 4.33.

The advantage in considering a GMM approach to the estimation problem is that it contains many classical estimation procedures as special cases. Here, we only mention two such special cases which are particularly useful in the context of estimating a parametric model for a Lévy process. It is an immediate consequence of the Definition 4.1 that a Lévy-process $\boldsymbol{L}$ is uniquely determined by the distribution of the unit increments $\boldsymbol{L}(n) - \boldsymbol{L}(n-1)$, which, in turn, is characterized by their common characteristic function $\mathbb{E}\exp\{\mathrm{i}\langle \boldsymbol{u}, \boldsymbol{L}(n) - \boldsymbol{L}(n-1)\rangle\} = \exp\{\psi(\boldsymbol{u})\}$ in its Lévy–Khintchine form (Eq. (4.2.1)). It is therefore natural to specify a parametric model for $\boldsymbol{L}$ by parametrizing the characteristic exponents, which amounts to defining, for each $\boldsymbol{\vartheta} \in \Theta$, a function $\boldsymbol{u} \mapsto \psi_{\boldsymbol{\vartheta}}(\boldsymbol{u})$ of the form (4.2.1). A promising estimator for $\boldsymbol{\vartheta}_0$ in such a model is that value of $\boldsymbol{\vartheta}$ that best matches the characteristic function $\boldsymbol{u} \mapsto \exp\{\psi_{\boldsymbol{\vartheta}}(\boldsymbol{u})\}$ with its empirical counterpart. This leads to choosing the function g in Eq. (4.6.3) as

$$g : \mathbb{R}^m \times \Theta \to \mathbb{R}^q : \quad (X, \boldsymbol{\vartheta}) \mapsto \begin{bmatrix} \operatorname{Re}\left(\mathrm{e}^{\mathrm{i}\langle \boldsymbol{u}_k, X\rangle} - \mathrm{e}^{\psi_{\boldsymbol{\vartheta}}(\boldsymbol{u}_k)}\right) \\ \operatorname{Im}\left(\mathrm{e}^{\mathrm{i}\langle \boldsymbol{u}_k, X\rangle} - \mathrm{e}^{\psi_{\boldsymbol{\vartheta}}(\boldsymbol{u}_k)}\right) \end{bmatrix}_{k=1,\ldots,q/2},$$

where $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{q/2}$ are suitable elements of $\mathbb{R}^m$ at which the characteristic functions are to be matched. The value of $q \in 2\mathbb{N}$ as well as the particular $\boldsymbol{u}_j$ are chosen such that condition (4.6.2) holds, which means that the model is identifiable. Another special case of the generalized method of moments estimator of considerable practical importance arises if Θ is a subset of $\mathbb{R}^r$, and the parametric family of distributions $\mathbb{P}_{\boldsymbol{\vartheta}}$ is given as a family of probability densities $p_{\boldsymbol{\vartheta}}(\cdot)$. Denoting by $\nabla = \begin{pmatrix} \partial/\partial\vartheta^1 & \cdots & \partial/\partial\vartheta^r \end{pmatrix}^T$ the differential operator and choosing the moment function g as

$$g : \mathbb{R}^m \times \Theta \to \mathbb{R}^r : \quad (X, \boldsymbol{\vartheta}) \mapsto \nabla_{\boldsymbol{\vartheta}} \log p_{\boldsymbol{\vartheta}}(X),$$

one obtains the classical maximum likelihood estimator, which, under some regularity assumptions about the densities $p_{\boldsymbol{\vartheta}}$, enjoys well-known desirable asymptotic properties (van der Vaart, 1998, Section 5.5).

In order to be able to state the classical result about the asymptotic properties of the generalized method of moments estimator for a general moment function g, we introduce the notations

$$\Omega_0 = \mathbb{E} g(X_1, \boldsymbol{\vartheta}_0) g(X_1, \boldsymbol{\vartheta}_0)^T, \quad \text{and} \quad G_0 = -\mathbb{E}\nabla_{\boldsymbol{\vartheta}} g(X_1, \boldsymbol{\vartheta}_0)$$

for the covariance matrix of the moments and the generalized score matrix, respectively.

Theorem 4.27 (Newey and McFadden (1994, Theorem 2.6 and Theorem 3.4)) *Assume that $(\mathbb{P}_{\boldsymbol{\vartheta}})_{\boldsymbol{\vartheta}\in\Theta}$ is a parametric family of probability distributions, and let $X^N = (X_1, \ldots, X_N)$ be an i. i. d. sample from the distribution $\mathbb{P}_{\boldsymbol{\vartheta}_0}$ of length N. Denote by $\hat{\boldsymbol{\vartheta}}^N$ the GMM estimator based on X^N defined in Eq. (4.6.3). Assume that the following conditions hold.*

i) The domain Θ of $\boldsymbol{\vartheta}$ is a compact subset of $\mathbb{R}^r$, and $\boldsymbol{\vartheta}_0$ is in the interior of Θ.

ii) For each $\boldsymbol{\vartheta} \in \Theta$, the function $\boldsymbol{x} \mapsto g(\boldsymbol{x}, \boldsymbol{\vartheta})$ is measurable; for almost every $\boldsymbol{x} \in \mathbb{R}^m$, the function $\boldsymbol{\vartheta} \mapsto g(\boldsymbol{x}, \boldsymbol{\vartheta})$ is continuous on Θ and continuously differentiable in a neighbourhood U of $\boldsymbol{\vartheta}_0$. Moreover there exists a function $\alpha : \mathbb{R}^m \to \mathbb{R}$ satisfying $\mathbb{E}\alpha(X_1) < \infty$ such that, for every $\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2 \in U$, it holds that $\|\nabla_{\boldsymbol{\vartheta}} g(\boldsymbol{x}, \boldsymbol{\vartheta}_1) - \nabla_{\boldsymbol{\vartheta}} g(\boldsymbol{x}, \boldsymbol{\vartheta}_2)\| \leqslant \alpha(\boldsymbol{x}) \|\boldsymbol{\vartheta}_1 - \boldsymbol{\vartheta}_2\|$.

iii) $\mathbb{E} g\left(X_1, \boldsymbol{\vartheta}\right) = 0$ if and only if $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$.

iv) $\mathbb{E} \left\|g\left(X_1, \boldsymbol{\vartheta}\right)\right\|^2 < \infty$ for all $\boldsymbol{\vartheta} \in \Theta$, Ω_0 is a positive definite $q \times q$ matrix, and G_0 is a $q \times r$ matrix of rank r.

v) W_N are $q \times q$ matrices converging in probability to a positive definite matrix W.

vi) There exists a function $\alpha : \mathbb{R}^m \to \mathbb{R}$ satisfying $\mathbb{E}\alpha(X_1) < \infty$ such that $\|g(\boldsymbol{x}, \boldsymbol{\vartheta}) g(\boldsymbol{x}, \boldsymbol{\vartheta})^T\| \leqslant \alpha(\boldsymbol{x})$ and $\|\nabla_{\boldsymbol{\vartheta}} g(\boldsymbol{x}, \boldsymbol{\vartheta})\| \leqslant \alpha(\boldsymbol{x})$.

It then holds that $\hat{\boldsymbol{\vartheta}}^N$ is consistent and asymptotically normally distributed, that is

$$N^{1/2}(\hat{\boldsymbol{\vartheta}}^N - \boldsymbol{\vartheta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}_r, \Sigma), \quad N \to \infty, \tag{4.6.4}$$

where the asymptotic covariance matrix Σ is given by

$$\Sigma = \left[G_0^T W G_0\right]^{-1} G_0^T W \Omega_0 W G_0 \left[G_0^T W G_0\right]^{-1}. \tag{4.6.5}$$

A result analogous to Theorem 4.27 holds in the more general set-up, where we do not have access to the sample X^N but only to a noisy variant. We first introduce the necessary notation, which we will need in the proof. The generalized method of moments estimator $\hat{\hat{\boldsymbol{\vartheta}}}^{N,h}$ of $\boldsymbol{\vartheta}_0$ based on the disturbed sample $X^{N,h} = (X_1 + \varepsilon_1^{(h)}, \ldots, X_N + \varepsilon_N^{(h)})$ is defined as

$$\hat{\hat{\boldsymbol{\vartheta}}}^{N,h} = \operatorname{argmin}_{\boldsymbol{\vartheta}\in\Theta} Q_{N,h}(\boldsymbol{\vartheta}), \tag{4.6.6}$$

where the random criterion function $Q_{N,h} : \Theta \to \mathbb{R}^+$ has the form

$$Q_{N,h}(\boldsymbol{\vartheta}) = \|m_{N,h}(\boldsymbol{\vartheta})\|^2_{W_{N,h}} \tag{4.6.7}$$

and $m_{N,h} : \Theta \to \mathbb{R}^q$ is given by

$$m_{N,h}(\boldsymbol{\vartheta}) = \frac{1}{N} \sum_{n=1}^{N} g\left(X_n + \varepsilon_n^{(h)}, \boldsymbol{\vartheta}\right), \quad g : \mathbb{R}^m \times \Theta \to \mathbb{R}^q.$$

Again, $W_{N,h}$ is a positive definite $q \times q$ matrix, which might depend on the sample $X^{N,h}$. As before, we write $\Omega(\boldsymbol{\vartheta}) = \mathbb{E} g\left(X_1, \boldsymbol{\vartheta}\right) g\left(X_1, \boldsymbol{\vartheta}\right)^T$, and

$$\Omega_{N,h}(\boldsymbol{\vartheta}) = \frac{1}{N} \sum_{n=1}^{N} g\left(X_n + \varepsilon_n^{(h)}, \boldsymbol{\vartheta}\right) g\left(X_n + \varepsilon_n^{(h)}, \boldsymbol{\vartheta}\right)^T$$

for the covariance matrix of the moments g and its empirical counterpart. The sample analogue of the score matrix $G(\boldsymbol{\vartheta}) = -\mathbb{E}\nabla_{\boldsymbol{\vartheta}} g\left(X_1, \boldsymbol{\vartheta}\right)$ is defined as

$$G_{N,h}(\boldsymbol{\vartheta}) = -\frac{1}{N} \sum_{n=1}^{N} \nabla_{\boldsymbol{\vartheta}} g\left(X_n + \varepsilon_n^{(h)}, \boldsymbol{\vartheta}\right).$$

Theorem 4.28 (GMM with noisy data) *Assume that $(\mathbb{P}_{\boldsymbol{\vartheta}})_{\boldsymbol{\vartheta}\in\Theta}$ is a parametric family of probability distributions, that X^N is an i. i. d. sample from the distribution $\mathbb{P}_{\boldsymbol{\vartheta}_0}$ of length N, and that, for each $h > 0$, there is a stochastic process $\varepsilon^{(h)} = \left(\varepsilon_n^{(h)}\right)_{n\in\mathbb{N}}$. Denote by $\hat{\hat{\boldsymbol{\vartheta}}}^{N,h}$ the GMM estimator based on $X^{N,h}$ defined in Eq. (4.6.6). In addition to the assumptions of Theorem 4.27 assume that the following hold.*

vii) There exists a function $\beta : \mathbb{R}^+ \to \mathbb{R}^+$ satisfying $\beta(h) \to 0$, as $h \to 0$, such that

$$\sup_n \mathbb{E} \left\| g\left(X_n + \varepsilon_n^{(h)}, \boldsymbol{\vartheta}_0\right) - g(X_n, \boldsymbol{\vartheta}_0) \right\| = O\left(\beta(h)\right), \tag{4.6.8}$$

as h tends to zero.

viii) For all $\boldsymbol{\vartheta} \in \Theta$, it holds that $\sup_n \mathbb{E} \left\| g\left(X_n + \varepsilon_n^{(h)}, \boldsymbol{\vartheta}\right) - g\left(X_n, \boldsymbol{\vartheta}\right) \right\|^2 \to 0$, as $h \to 0$.

ix) For all $\boldsymbol{\vartheta} \in \Theta$, the derivative of g satisfies $\sup_n \mathbb{E} \left\| \nabla_{\boldsymbol{\vartheta}} g\left(X_n + \varepsilon_n^{(h)}, \boldsymbol{\vartheta}\right) - \nabla_{\boldsymbol{\vartheta}} g\left(X_1, \boldsymbol{\vartheta}\right) \right\| \to 0$, as $h \to 0$.

If $h = h_N$ is chosen dependent on N such that $N^{1/2}\beta(h_N) \to 0$ as $N \to \infty$, then it holds that $\hat{\hat{\boldsymbol{\vartheta}}}^{N,h_N}$ is consistent and asymptotically normally distributed with the same asymptotic covariance as $\hat{\boldsymbol{\vartheta}}^N$, given in Eq. (4.6.5).

The proof of Theorem 4.28 closely follows the arguments in Newey and McFadden (1994). We give a detailed proof in order to clarify the impact of the additional parameter h and the difficulties arising from the need to take the double limit $N \to \infty$ and $h \to 0$. We first lay ground for the proof by recalling a sequence of auxiliary lemmata.

Lemma 4.29 *For sequences $(Y_n)_{n\geqslant 1}$, $(Z_n)_{n\geqslant 1}$ of vector- or matrix-valued random variables the following hold:*

i) For every constant c, $Y_n \xrightarrow{p} c$ if and only if $Y_n \xrightarrow{d} c$.

ii) If $Y_n \xrightarrow{d} Y_\infty$ and $Z_n - Y_n \xrightarrow{p} 0$, then $Z_n \xrightarrow{d} Y_\infty$.

iii) Denote by $\operatorname{supp} Y_n$ the support of Y_n. If $Y_n \xrightarrow{d} Y_\infty$, for some Y_∞, and the function f is defined on $\bigcap_{n\geqslant 1} \operatorname{supp} Y_n$ and continuous on an open set containing $\operatorname{supp} Y_\infty$, then $f(Y_n) \xrightarrow{d} f(Y_\infty)$.

Proof Parts i) and ii) are proved in van der Vaart (1998, Theorem 2.7). Assertion iii) is Klenke (2008, Theorem 13.25)). □

The next result that we will need is a uniform version of the weak law of large numbers, given by (Newey and McFadden, 1994, Lemma 2.4).

Lemma 4.30 *Assume that, for every $\boldsymbol{\vartheta} \in \Theta$, Θ a compact subset of $\mathbb{R}^r$, there is a sequence $(Y_n(\boldsymbol{\vartheta}))_{n\geqslant 1}$ of independent identically distributed random variables with finite expectation $\psi(\boldsymbol{\vartheta}) = \mathbb{E}Y_1(\boldsymbol{\vartheta}) < \infty$. Further assume that, for each $\boldsymbol{\vartheta}' \in \Theta$, the random function $\boldsymbol{\vartheta} \mapsto Y_1(\boldsymbol{\vartheta})$ is almost surely continuous at $\boldsymbol{\vartheta}'$ and that there exists a random variable Z satisfying $\mathbb{E}Z < \infty$, such that $\sup_{\boldsymbol{\vartheta}\in\Theta} \|Y_1(\boldsymbol{\vartheta})\| \leqslant Z$. It then holds that the function $\boldsymbol{\vartheta} \mapsto \psi(\boldsymbol{\vartheta})$ is continuous, and that the time averages $\overline{Y}_N(\boldsymbol{\vartheta}) = \sum_{n=1}^N Y_n(\boldsymbol{\vartheta})$ converge uniformly in probability to $\psi(\boldsymbol{\vartheta})$, that is $\sup_{\boldsymbol{\vartheta}\in\Theta} \|\overline{Y}_N(\boldsymbol{\vartheta}) - \psi(\boldsymbol{\vartheta})\| \xrightarrow{p} 0$.*

Lemma 4.31 *For each $\boldsymbol{\vartheta} \in \Theta$, let $(Y_n(\boldsymbol{\vartheta}))_{n\geqslant 1}$ be a sequence of random variables. If $Y_n(\boldsymbol{\vartheta}) \xrightarrow{p} Y_\infty(\boldsymbol{\vartheta})$ uniformly in $\boldsymbol{\vartheta}$, the sequence $(\boldsymbol{\vartheta}_n)_{n\geqslant 1}$ of random elements of Θ converges in probability to some $\boldsymbol{\vartheta}_\infty$, and the mapping $\boldsymbol{\vartheta} \mapsto Y_\infty(\boldsymbol{\vartheta})$ is almost surely continuous at $\boldsymbol{\vartheta}_\infty$, then $Y_n(\boldsymbol{\vartheta}_n) \xrightarrow{p} Y_\infty(\boldsymbol{\vartheta}_\infty)$.*

Proof For any $\epsilon > 0$, it holds that

$$\begin{aligned}
\mathbb{P}\left(\|Y_n(\boldsymbol{\vartheta}_n) - Y_\infty(\boldsymbol{\vartheta}_\infty)\| \leqslant \epsilon\right) \geqslant& \mathbb{P}\left(\|Y_n(\boldsymbol{\vartheta}_n) - Y_\infty(\boldsymbol{\vartheta}_n)\| \leqslant \frac{\epsilon}{2} \text{ and } \|Y_\infty(\boldsymbol{\vartheta}_n) - Y_\infty(\boldsymbol{\vartheta}_\infty)\| \leqslant \frac{\epsilon}{2}\right)\\
\geqslant& \mathbb{P}\left(\|Y_n(\boldsymbol{\vartheta}_n) - Y_\infty(\boldsymbol{\vartheta}_n)\| \leqslant \frac{\epsilon}{2}\right)\\
&\quad + \mathbb{P}\left(\|Y_\infty(\boldsymbol{\vartheta}_n) - Y_\infty(\boldsymbol{\vartheta}_\infty)\| \leqslant \frac{\epsilon}{2}\right) - 1\\
\to& 1.
\end{aligned}$$

The first probability in the last line converges to one as n tends to infinity by the assumption of uniform convergence of Y_n to Y_∞, the second because Y_∞ is almost surely continuous at $\boldsymbol{\vartheta}_\infty$ and $\boldsymbol{\vartheta}_n \xrightarrow{p} \boldsymbol{\vartheta}_\infty$. □

We can now give the proof of the asymptotic properties of GMM estimators with noisy data.

Proof (of Theorem 4.28) The proof consists of four steps. In the first step we show that $N^{1/2}m_{N,h}(\boldsymbol{\vartheta}_0)$ is asymptotically normally distributed with mean zero and covariance matrix Ω_0, that $m_{N,h}(\boldsymbol{\vartheta})$, $G_{N,h}(\boldsymbol{\vartheta})$ and $\Omega_{N,h}(\boldsymbol{\vartheta})$ converge uniformly in probability to $\mathbb{E}g\left(X_1,\boldsymbol{\vartheta}\right)$, $G(\boldsymbol{\vartheta})$ and $\Omega(\boldsymbol{\vartheta})$, respectively, and that $N\,Q_{N,h}(\boldsymbol{\vartheta}_0)$ is bounded in probability. The second step consists in showing that any estimator $\tilde{\boldsymbol{\vartheta}}^{N,h}$ that approximately minimizes the criterion function $Q_{N,h}$ in the sense that $m_{N,h}(\tilde{\boldsymbol{\vartheta}}^{N,h}) \xrightarrow{p} 0$, converges in probability to $\boldsymbol{\vartheta}_0$. In step 3 we prove that stochastic boundedness of $N\,Q_{N,h}(\tilde{\boldsymbol{\vartheta}}^{N,h})$ implies the stochastic boundedness of $N^{1/2}(\tilde{\boldsymbol{\vartheta}}^{N,h}-\boldsymbol{\vartheta}_0)$. We will see that steps 2 and 3 imply the consistency of $\hat{\hat{\boldsymbol{\vartheta}}}^{N,h}$ for any sequence of weighting matrices $W_{N,h}$. In the last step the mean-value theorem is applied to the first-order condition for $\hat{\hat{\boldsymbol{\vartheta}}}^{N,h}$ to prove the asymptotic normality of $N^{1/2}(\hat{\hat{\boldsymbol{\vartheta}}}^{N,h}-\boldsymbol{\vartheta}_0)$.

Step 1 In order to prove that $N^{1/2}m_{N,h}(\boldsymbol{\vartheta}_0)$ is asymptotically normally distributed, we observe that

$$N^{1/2}m_{N,h}(\boldsymbol{\vartheta}_0)=\frac{1}{N^{1/2}}\sum_{n=1}^{N}g(X_n,\boldsymbol{\vartheta}_0)+\frac{1}{N^{1/2}}\sum_{n=1}^{N}\left[g\left(X_n+\varepsilon_n^{(h)},\boldsymbol{\vartheta}_0\right)-g(X_n,\boldsymbol{\vartheta}_0)\right].$$

The first term in this expression is asymptotically normal by the Lindeberg–Lévy Central Limit Theorem (Klenke, 2008, Theorem 15.37) since the summands $g(X_n,\boldsymbol{\vartheta}_0)$ are i. i. d. with finite variance. It therefore suffices to show that the second term converges to zero in probability as $N\to\infty$ if $h=h_N$ satisfies $N^{1/2}\beta(h_N)\to 0$. For convenience, we introduce the notation $Y_n^{(h)}=g(X_1+\varepsilon_1^{(h)},\boldsymbol{\vartheta}_0)-g(X_1,\boldsymbol{\vartheta}_0)$; by the linearity of expectation and assumption vii), it follows that

$$\mathbb{E}\left\|N^{-1/2}\sum_{n=1}^{N}Y_n^{(h)}\right\|\leqslant N^{-1/2}\sum_{n=1}^{N}\mathbb{E}\left\|Y_n^{(h)}\right\|\leqslant CN^{1/2}\beta(h),\quad\text{for some } C>0. \tag{4.6.9}$$

This proves that $N^{-1/2}\sum_{n=1}^{N}Y_n^{(h_N)}$ converges in L^1, and hence in probability, to zero, thereby showing the asymptotic normality of $N^{1/2}m_{N,h}(\boldsymbol{\vartheta}_0)$, that is

$$\Omega_0^{-1/2}N^{1/2}m_{N,h}(\boldsymbol{\vartheta}_0)=:U_{N,h}\xrightarrow{d}U\sim\mathcal{N}(\mathbf{0}_q,\mathbf{1}_q),\quad\text{as } N\to\infty, h\to 0, N^{1/2}\beta(h)\to 0. \tag{4.6.10}$$

We now turn to the uniform convergence in probability of $m_{N,h}(\boldsymbol{\vartheta})$, $G_{N,h}(\boldsymbol{\vartheta})$ and $\Omega_{N,h}(\boldsymbol{\vartheta})$: point-wise convergence of $m_{N,h}(\boldsymbol{\vartheta})$ to $\mathbb{E}g\left(X_1,\boldsymbol{\vartheta}\right)$ follows from the observation that

$$m_{N,h}(\boldsymbol{\vartheta})=\frac{1}{N}\sum_{n=1}^{N}g\left(X_n,\boldsymbol{\vartheta}\right)+\frac{1}{N}\sum_{n=1}^{N}\left[g\left(X_n+\varepsilon_n^{(h)},\boldsymbol{\vartheta}\right)-g\left(X_n,\boldsymbol{\vartheta}\right)\right].$$

As a sample average the first term converges to $\mathbb{E}g\left(X_1,\boldsymbol{\vartheta}\right)$ as $N\to\infty$ by the law of large numbers (Klenke, 2008, Theorem 5.16). As in Eq. (4.6.9) one sees that the second term converges in L^1 and therefore in probability to zero, as $N \to \infty$ and $h \to 0$. Analogously,

$$G_{N,h}(\boldsymbol{\vartheta}) = \frac{1}{N}\sum_{n=1}^{N} \nabla_{\boldsymbol{\vartheta}} g\left(X_n, \boldsymbol{\vartheta}\right) + \frac{1}{N}\sum_{n=1}^{N} \left[\nabla_{\boldsymbol{\vartheta}} g\left(X_n + \varepsilon_n^{(h)}, \boldsymbol{\vartheta}\right) - \nabla_{\boldsymbol{\vartheta}} g\left(X_n, \boldsymbol{\vartheta}\right)\right].$$

converges point-wise in probability to $G(\boldsymbol{\vartheta}) = -\mathbb{E}\nabla g\left(X_1, \boldsymbol{\vartheta}\right)$ by assumption ix). Finally,

$$\begin{aligned}\Omega_{N,h}(\boldsymbol{\vartheta}) =& \frac{1}{N}\sum_{n=1}^{N} g\left(X_n, \boldsymbol{\vartheta}\right) g\left(X_n, \boldsymbol{\vartheta}\right)^T + \frac{1}{N}\sum_{n=1}^{N} Y_n^{(h)} \left(Y_n^{(h)}\right)^T \\ &+ \frac{1}{N}\sum_{n=1}^{N} g\left(X_n, \boldsymbol{\vartheta}\right) \left(Y_n^{(h)}\right)^T + \frac{1}{N}\sum_{n=1}^{N} Y_n^{(h)} g\left(X_n, \boldsymbol{\vartheta}\right)^T,\end{aligned}$$

where we have again used the notation $Y_n^{(h)} = g\left(X_n + \varepsilon_n^{(h)}, \boldsymbol{\vartheta}\right) - g\left(X_n, \boldsymbol{\vartheta}\right)$. The first term in this expression for $\Omega_{N,h}(\boldsymbol{\vartheta})$ converges to $\Omega(\boldsymbol{\vartheta}) = \mathbb{E} g\left(X_1, \boldsymbol{\vartheta}\right) g\left(X_1, \boldsymbol{\vartheta}\right)^T$ by the law of large numbers, the second term converges to zero in L^1 and in probability due to assumption viii). An application of the Cauchy–Schwarz inequality to the third term shows that

$$\begin{aligned}\mathbb{E}\left\|\frac{1}{N}\sum_{n=1}^{N} g\left(X_n, \boldsymbol{\vartheta}\right) \left(Y_n^{(h)}\right)^T\right\| \leqslant& \frac{1}{N}\sum_{n=1}^{N} \mathbb{E}\left\|g\left(X_n, \boldsymbol{\vartheta}\right) \left(Y_n^{(h)}\right)^T\right\| \\ \leqslant& \sup_n \mathbb{E}\left\|g\left(X_n, \boldsymbol{\vartheta}\right) \left(Y_n^{(h)}\right)^T\right\| \\ \leqslant& \sup_n \mathbb{E}\left\|g\left(X_n, \boldsymbol{\vartheta}\right)\right\| \left\|Y_n^{(h)}\right\| \\ \leqslant& \sqrt{\mathbb{E}\left\|g\left(X_1, \boldsymbol{\vartheta}\right)\right\|^2} \sqrt{\sup_n \mathbb{E}\left\|Y_n^{(h)}\right\|^2}.\end{aligned}$$

The first factor is finite by assumption iv), the second one converges to zero as $h \to 0$ by assumption viii). By assumptions ii) and vi), the limiting functions $\boldsymbol{\vartheta} \mapsto \mathbb{E} g\left(X_1, \boldsymbol{\vartheta}\right)$, $\boldsymbol{\vartheta} \mapsto G(\boldsymbol{\vartheta})$ and $\boldsymbol{\vartheta} \mapsto \Omega(\boldsymbol{\vartheta})$ are continuous and dominated, and since the domain Θ is compact by assumption i), we can apply Lemma 4.30 to conclude that the convergence is uniform in $\boldsymbol{\vartheta}$. Taking into consideration the assumed convergence in probability of $W_{N,h}$ (assumption v)) as well as Eq. (4.6.10), Lemma 4.29 implies that $N\, Q_{N,h}(\boldsymbol{\vartheta}_0)$ is bounded in probability.

Step 2 In this step the consistency of any estimator $\tilde{\boldsymbol{\vartheta}}^{N,h}$ satisfying $Q_{N,h}(\tilde{\boldsymbol{\vartheta}}^{N,h}) \xrightarrow{p} 0$ is proved. In the first step we have established the uniform convergence in probability of $m_{N,h}(\boldsymbol{\vartheta})$ to $\mathbb{E} g\left(X_1, \boldsymbol{\vartheta}\right)$. In conjunction with assumption v) this implies that the sequence $\sup_{\boldsymbol{\vartheta}\in\Theta}\left|Q_{N,h}(\boldsymbol{\vartheta}) - \left\|\mathbb{E} g\left(X_1, \boldsymbol{\vartheta}\right)\right\|_W^2\right|$ converges to zero in probability. To establish consistency of $\tilde{\boldsymbol{\vartheta}}^{N,h}$, we shall show that, for any neighbourhood U of $\boldsymbol{\vartheta}_0$ and every $\epsilon > 0$, there exists an $N_\epsilon(U)$ and an $h_\epsilon(U)$ such that $\mathbb{P}\left(\tilde{\boldsymbol{\vartheta}}^{N,h} \in U\right) \geqslant 1 - \epsilon$ for all $N > N_\epsilon(U)$ and $h < h_\epsilon(U)$. For given U, we define $\delta(U) \coloneqq \inf_{\boldsymbol{\vartheta}\in\Theta\setminus U} \|\mathbb{E}g\left(X_1,\boldsymbol{\vartheta}\right)\|_W$ which is strictly positive by assumptions i) to iii). Choosing $N_\epsilon(U)$ and $h_\epsilon(U)$ such that

$$\begin{aligned}\mathbb{P}\left(Q_{N,h}(\tilde{\boldsymbol{\vartheta}}^{N,h})\leqslant\delta(U)/2\right)\geqslant&1-\epsilon/2,\\ \mathbb{P}\left(\sup_{\boldsymbol{\vartheta}\in\Theta}\left|Q_{N,h}(\boldsymbol{\vartheta})-\|\mathbb{E}g\left(X_1,\boldsymbol{\vartheta}\right)\|_W\right|\leqslant\delta(U)/2\right)\geqslant&1-\epsilon/2,\end{aligned}$$

for all $N\geqslant N_\epsilon(U)$ and $h\leqslant h_\epsilon(U)$, it follows that

$$\begin{aligned}\mathbb{P}\left(\tilde{\boldsymbol{\vartheta}}^{N,h}\in U\right)\geqslant&\mathbb{P}\left(\left\|\mathbb{E}g(X_1,\tilde{\boldsymbol{\vartheta}}^{N,h})\right\|_W\leqslant\delta(U)\right)\\ \geqslant&\mathbb{P}\left(Q_{N,h}(\tilde{\boldsymbol{\vartheta}}^{N,h})\leqslant\frac{\delta(U)}{2}\text{ and }\sup_{\boldsymbol{\vartheta}\in\Theta}\left|Q_{N,h}(\boldsymbol{\vartheta})-\|\mathbb{E}g\left(X_1,\boldsymbol{\vartheta}\right)\|_W\right|\leqslant\frac{\delta(U)}{2}\right)\\ \geqslant&1-\epsilon,\end{aligned}$$

where in the last line we used the relation $\mathbb{P}(A\cap B)\geqslant\mathbb{P}(A)+\mathbb{P}(B)-1$.

Step 3 This step is devoted to the implication that if $N\,Q_{N,h}(\tilde{\boldsymbol{\vartheta}}^{N,h})$ is bounded in probability, then the sequence $N^{1/2}(\tilde{\boldsymbol{\vartheta}}^{N,h}-\boldsymbol{\vartheta}_0)$ is bounded in probability as well. The assumption $N\,Q_{N,h}(\tilde{\boldsymbol{\vartheta}}^{N,h})=O_p(1)$ implies that $Q_{N,h}(\tilde{\boldsymbol{\vartheta}}^{N,h})\xrightarrow{p}0$ and therefore, by the previous step, that $\tilde{\boldsymbol{\vartheta}}^{N,h}\xrightarrow{p}\boldsymbol{\vartheta}_0$. By the mean-value theorem, there exist parameter values $\boldsymbol{\vartheta}_i^*\in\Theta$, $i=1,\ldots,r$, of the form $\boldsymbol{\vartheta}_i^*=\boldsymbol{\vartheta}_0+c_i(\tilde{\boldsymbol{\vartheta}}^{N,h}-\boldsymbol{\vartheta}_0)$, $0\leqslant c_i\leqslant 1$, such that we can write

$$\begin{aligned}N^{1/2}m_{N,h}(\tilde{\boldsymbol{\vartheta}}^{N,h})=&N^{1/2}m_{N,h}(\boldsymbol{\vartheta}_0)+\nabla_{\boldsymbol{\vartheta}}m_{N,h}(\underline{\boldsymbol{\vartheta}}^*)N^{1/2}(\tilde{\boldsymbol{\vartheta}}^{N,h}-\boldsymbol{\vartheta}_0)\\ =&\Omega_0^{1/2}U_{N,h}-G_{N,h}(\underline{\boldsymbol{\vartheta}}^*)N^{1/2}(\tilde{\boldsymbol{\vartheta}}^{N,h}-\boldsymbol{\vartheta}_0),\end{aligned}\tag{4.6.11}$$

where $G_{N,h}(\underline{\boldsymbol{\vartheta}}^*)$ denotes the matrix whose ith row coincides with the ith row of $G(\boldsymbol{\vartheta}_i^*)$, and $U_{N,h}$ is defined in Eq. (4.6.10). By applying the triangle inequality of the norm $\|\cdot\|_{W_{N,h}}$ to the vector

$$G_{N,h}(\underline{\boldsymbol{\vartheta}}^*)N^{1/2}(\tilde{\boldsymbol{\vartheta}}^{N,h}-\boldsymbol{\vartheta}_0)=\Omega_0^{1/2}U_{N,h}-N^{1/2}m_{N,h}(\tilde{\boldsymbol{\vartheta}}^{N,h})$$

one obtains that

$$\left\|G_{N,h}(\underline{\boldsymbol{\vartheta}}^*)N^{1/2}(\tilde{\boldsymbol{\vartheta}}^{N,h}-\boldsymbol{\vartheta}_0)\right\|_{W_{N,h}}^2\leqslant 2\left\|\Omega_0^{1/2}U_{N,h}\right\|_{W_{N,h}}^2+2N\,Q_{N,h}(\tilde{\boldsymbol{\vartheta}}^{N,h}).$$

Since $U_{N,h}$ converges in distribution to a standard normal and $W_{N,h}$ converges in probability, the first term on the right hand side of the last display converges in distribution by Lemma 4.29 and is in particular bounded in probability. By our hypothesis, $N\,Q_{N,h}(\tilde{\boldsymbol{\vartheta}}^{N,h})$ is bounded in probability, and so it follows that

$$\left\| G_{N,h}(\underline{\boldsymbol{\vartheta}}^*) N^{1/2}(\tilde{\boldsymbol{\vartheta}}^{N,h} - \boldsymbol{\vartheta}_0) \right\|_{W_{N,h}} = O_p(1). \tag{4.6.12}$$

It follows from the uniform convergence in probability of $G_{N,h}(\boldsymbol{\vartheta})$ to $G(\boldsymbol{\vartheta})$, the fact that $\boldsymbol{\vartheta}_i^* \xrightarrow{p} \boldsymbol{\vartheta}_0$, and Lemma 4.31 applied to the rows of $G_{N,h}$ that

$$G_{N,h}(\underline{\boldsymbol{\vartheta}}^*)^T W_{N,h} G_{N,h}(\underline{\boldsymbol{\vartheta}}^*) \xrightarrow{p} G_0^T W G_0;$$

this in turn implies together with Eq. (4.6.12) that $N^{1/2}(\tilde{\boldsymbol{\vartheta}}^{N,h} - \boldsymbol{\vartheta}_0)$ is bounded in probability.

Step 4 In this last step we prove that the estimator $\hat{\hat{\boldsymbol{\vartheta}}}^{N,h} = \operatorname{argmin}_{\boldsymbol{\vartheta}\in\Theta} Q_{N,h}(\boldsymbol{\vartheta})$ is asymptotically normally distributed. The definition of $\hat{\hat{\boldsymbol{\vartheta}}}^{N,h}$ implies that $Q_{N,h}(\hat{\hat{\boldsymbol{\vartheta}}}^{N,h}) \leqslant Q_{N,h}(\boldsymbol{\vartheta}_0)$. We have shown in the first step that $N\,Q_{N,h}(\boldsymbol{\vartheta}_0)$ is bounded in probability and hence so is $N\,Q_{N,h}(\hat{\hat{\boldsymbol{\vartheta}}}^{N,h})$. This implies by step 2 that $\hat{\hat{\boldsymbol{\vartheta}}}^{N,h}$ is consistent and that $N^{1/2}(\hat{\hat{\boldsymbol{\vartheta}}}^{N,h} - \boldsymbol{\vartheta}_0)$ is bounded in probability. Since $\hat{\hat{\boldsymbol{\vartheta}}}^{N,h}$ is an extremal point of $Q_{N,h}$ we obtain by setting the derivative equal to zero that $G_{N,h}(\hat{\hat{\boldsymbol{\vartheta}}}^{N,h})^T W_{N,h} N^{1/2} m_{N,h}(\hat{\hat{\boldsymbol{\vartheta}}}^{N,h}) = 0$. By combining the Taylor expansion (4.6.11) with this first-order condition it follows that

$$0 = G_{N,h}(\hat{\hat{\boldsymbol{\vartheta}}}^{N,h})^T W_{N,h} \Omega_0^{1/2} U_{N,h} - G_{N,h}(\hat{\hat{\boldsymbol{\vartheta}}}^{N,h})^T W_{N,h} G_{N,h}(\underline{\boldsymbol{\vartheta}}^*) N^{1/2}(\hat{\hat{\boldsymbol{\vartheta}}}^{N,h} - \boldsymbol{\vartheta}_0).$$

As before one sees that $G_{N,h}(\hat{\hat{\boldsymbol{\vartheta}}}^{N,h})^T W_{N,h} G_{N,h}(\underline{\boldsymbol{\vartheta}}^*)$ converges in probability to the non-singular limit $G_0^T W G_0$, which means that the vector

$$N^{1/2}(\hat{\hat{\boldsymbol{\vartheta}}}^{N,h} - \boldsymbol{\vartheta}_0) = \left[G_{N,h}(\hat{\hat{\boldsymbol{\vartheta}}}^{N,h})^T W_{N,h} G_{N,h}(\underline{\boldsymbol{\vartheta}}^*)\right]^{-1} G_{N,h}(\hat{\hat{\boldsymbol{\vartheta}}}^{N,h})^T W_{N,h} \Omega_0^{1/2} U_{N,h}$$

exists with probability approaching one. Since

$$\left[G_{N,h}(\hat{\hat{\boldsymbol{\vartheta}}}^{N,h})^T W_{N,h} G_{N,h}(\underline{\boldsymbol{\vartheta}}^*)\right]^{-1} G_{N,h}(\hat{\hat{\boldsymbol{\vartheta}}}^{N,h})^T W_{N,h} \xrightarrow{p} \left[G_0^T W G_0\right]^{-1} G_0^T W,$$

it follows from Lemma 4.29 that

$$N^{1/2}(\hat{\hat{\boldsymbol{\vartheta}}}^{N,h} - \boldsymbol{\vartheta}_0) \xrightarrow{d} \left[G_0^T W G_0\right]^{-1} G_0^T W \Omega_0^{1/2} U,$$

the limit on the right being a normally distributed random vector with covariance matrix $\Sigma = \left[G_0^T W G_0\right]^{-1} G_0^T W \Omega_0 W G_0 \left[G_0^T W G_0\right]^{-1}$. If the dimension r of the parameter space Θ is equal to the dimension q of the moment vector and the matrix G_0 is thus square or if $W = \Omega_0^{-1}$, it follows that $\Sigma = \left[G_0^T \Omega_0^{-1} G_0\right]^{-1}$. □

Remark 4.32 It seems possible to extend most aspects of the asymptotic theory of the generalized method of moments beyond the Central Limit Theorem 4.27 to deal, for example, with non-compact parameter spaces and applications to hypothesis testing based on a disturbed sample as in Theorem 6.2. We choose not to pursue these possibilities further at the present stage.

In view of Lemma 4.31, assumption v) of Theorem 4.28 is satisfied if we choose $W_{N,h} = W_{N,h}(\bar{\boldsymbol{\vartheta}}^{N,h})$, where $\bar{\boldsymbol{\vartheta}}^{N,h}$ is a consistent estimator of $\boldsymbol{\vartheta}_0$, and the functions $\boldsymbol{\vartheta} \mapsto W_{N,h}(\boldsymbol{\vartheta})$ converge uniformly in probability to $\boldsymbol{\vartheta} \mapsto W(\boldsymbol{\vartheta})$. In this way one can construct a sequence $W_{N,h}$ of weighting matrices converging in probability to Ω_0^{-1}. For this two-stage GMM estimation procedure one has the following optimality result.

Corollary 4.33 *Let $\tilde{\boldsymbol{\vartheta}}^{N,h}$ be the estimate of $\boldsymbol{\vartheta}$ obtained from maximizing the W-norm of $m_{N,h}(\boldsymbol{\vartheta})$ for any fixed $q \times q$ positive definite matrix W and let $\hat{\hat{\boldsymbol{\vartheta}}}^{N,h}$ be the estimate obtained from using the random weighting matrix*

$$\widetilde{W}_{N,h} = \Omega_{N,h}(\tilde{\boldsymbol{\vartheta}}^{N,h})^{-1} = \left[\frac{1}{N}\sum_{n=1}^{N} g\left(X_n + \varepsilon_n^{(h)}, \tilde{\boldsymbol{\vartheta}}^{N,h}\right)^T g\left(X_n + \varepsilon_n^{(h)}, \tilde{\boldsymbol{\vartheta}}^{N,h}\right)\right]^{-1}. \tag{4.6.13}$$

Under the conditions of Theorem 4.28, the estimator $\hat{\hat{\boldsymbol{\vartheta}}}^{N,h}$ is consistent and asymptotically normally distributed. In the partial order induced by positive semidefiniteness, the asymptotic covariance matrix of the limiting normal distribution, $\left[G_0^T\Omega_0^{-1}G_0\right]^{-1}$, is smaller than or equal to the covariance matrix obtained from every other sequence of weighting matrices $W_{N,h}$.

Proof It has been shown in the proof of Theorem 4.28 that the preliminary estimator $\tilde{\boldsymbol{\vartheta}}^{N,h}$ is consistent and that the sequence of functions $\boldsymbol{\vartheta} \mapsto \Omega_{N,h}(\boldsymbol{\vartheta})$ converges uniformly in probability to the function $\boldsymbol{\vartheta} \mapsto \Omega(\boldsymbol{\vartheta})$. It then follows from Lemma 4.31 that the sequence $\widetilde{W}_{N,h}$ of weighting matrices converges in probability to Ω_0^{-1}, and from Theorem 4.28 that $\hat{\hat{\boldsymbol{\vartheta}}}^{N,h}$ is asymptotically normal with asymptotic covariance matrix

$$\left[G_0^T\Omega_0^{-1}G_0\right]^{-1} G_0^T\Omega_0^{-1}\Omega_0\Omega_0^{-1}G_0\left[G_0^T\Omega_0^{-1}G_0\right]^{-1} = \left[G_0^T\Omega_0^{-1}G_0\right]^{-1}.$$

To show that this is smaller than or equal to the asymptotic covariance matrix of an estimator obtained from using a sequence of weighting matrices that converges in probability to the positive definite matrix W, we must show that the matrix

$$\Delta = \left[G_0^TWG_0\right]^{-1} G_0^TW\Omega_0WG_0\left[G_0^TWG_0\right]^{-1} - \left[G_0^T\Omega_0^{-1}G_0\right]^{-1}$$

is positive semidefinite. To see this, it is enough to note that Δ can be written as

$$\Delta = \left[\Omega_0^{1/2} W G_0 \left(G_0^T W G_0\right)^{-1}\right]^T \left[\mathbf{1}_r - \Omega_0^{-1/2} G_0 \left(G_0^T \Omega_0^{-1} G_0\right)^{-1} G_0^T \Omega_0^{-1/2}\right] \times \left[\Omega_0^{1/2} W G_0 \left(G_0^T W G_0\right)^{-1}\right].$$

Since the factor in the middle is idempotent and therefore positive semidefinite, and since semidefiniteness is preserved under conjugation, the matrix Δ is positive semidefinite. $\square$

We can now state and prove our main result about the asymptotic properties of the generalized method of moments estimation of the driving Lévy process of a multivariate CARMA process from discrete observations. This method can be used to select a suitable driving process from within a parametric family of Lévy processes as part of specifying a CARMA model for an observed time series. We assume that Θ is a parameter space and that $(\boldsymbol{L}_{\boldsymbol{\vartheta}})_{\boldsymbol{\vartheta}\in\Theta}$ is a family of Lévy processes. The process $\boldsymbol{Y}$ is an $\boldsymbol{L}_{\boldsymbol{\vartheta}_0}$-driven multivariate CARMA(p,q) process given by a state space representation of the form (4.3.10), and we assume that h-spaced observations $\boldsymbol{Y}(0), \boldsymbol{Y}(h), \ldots, \boldsymbol{Y}(N+(p-q-1)h)$ of $\boldsymbol{Y}$ are available on the discrete time grid $(0, h, \ldots, N+(p-q-1)h)$. Based on these observed values, a set of N approximate unit increments $\widehat{\Delta \boldsymbol{L}}_n^{(h)}$, $n = 1, \ldots, N$, of the driving process is computed using Eq. (4.5.37). For each integer N and each sampling frequency $h^{-1} \in \mathbb{N}$, a generalized method of moments estimator is defined as in Eq. (4.6.6) by

$$\hat{\hat{\boldsymbol{\vartheta}}}^{N,h} = \operatorname{argmin}_{\boldsymbol{\vartheta}\in\Theta} \left\| \frac{1}{N} \sum_{n=1}^{N} g\left(\widehat{\Delta \boldsymbol{L}}_n^{(h)}, \boldsymbol{\vartheta}\right) \right\|_{W_{N,h}}, \tag{4.6.14}$$

where $g : \mathbb{R}^m \times \Theta \to \mathbb{R}^q$ is a moment function, and $W_{N,h} \in M_q(\mathbb{R})$ is a positive definite weighting matrix. The following theorem asserts that the sequence $(\hat{\hat{\boldsymbol{\vartheta}}}^{N,h_N})_N$ of estimators is consistent and asymptotically normally distributed if h_N is chosen such that Nh_N converges to zero.

Theorem 4.34 (GMM with $\widehat{\Delta \boldsymbol{L}}_n^{(h)}$) *Assume that $\Theta \subset \mathbb{R}^r$ is a parameter space, that $(\boldsymbol{L}_{\boldsymbol{\vartheta}})_{\boldsymbol{\vartheta}\in\Theta}$ is a parametric family of m-dimensional Lévy processes, and that $\boldsymbol{Y}$ is an $\boldsymbol{L}_{\boldsymbol{\vartheta}_0}$-driven multivariate CARMA process satisfying Assumptions A1 and A2. Denote by $\hat{\hat{\boldsymbol{\vartheta}}}^{N,h}$ the generalized method of moments estimator defined in Eq. (4.6.14). Assume that, for some integer k, the functions $f_{\boldsymbol{\vartheta}} : \boldsymbol{x} \mapsto g(\boldsymbol{x}, \boldsymbol{\vartheta})$ possess a bounded kth derivative, that $\mathbb{E} \left\|\boldsymbol{L}_{\boldsymbol{\vartheta}_0}(1)\right\|^{2k}$ is finite, and that the partial derivatives of the functions $f_{\boldsymbol{\vartheta}}$ satisfy*

$$\mathbb{E} \left\|\partial_{i_1} \cdot \ldots \cdot \partial_{i_\kappa} f_{\boldsymbol{\vartheta}}\left(\boldsymbol{L}_{\boldsymbol{\vartheta}_0}(1)\right)\right\|^{2k} < \infty, \quad 1 \leqslant i_1, \ldots, i_\kappa \leqslant m, \quad 1 \leqslant \kappa \leqslant k-1, \quad \boldsymbol{\vartheta} \in \Theta. \tag{4.6.15}$$

Further assume that, for each $\boldsymbol{x} \in \mathbb{R}^m$, the function $\boldsymbol{\vartheta} \mapsto g(\boldsymbol{x}, \boldsymbol{\vartheta})$ is differentiable, that, for some *integer l, the functions $h_{\boldsymbol{\vartheta}} : \boldsymbol{x} \mapsto \nabla_{\boldsymbol{\vartheta}} g(\boldsymbol{x}, \boldsymbol{\vartheta})$ have a bounded lth derivative, and that the partial derivatives of $h_{\boldsymbol{\vartheta}}$ satisfy*

$$\mathbb{E}\left\|\partial_{i_1}\cdot\ldots\cdot\partial_{i_\lambda} h_{\boldsymbol{\vartheta}}\left(\boldsymbol{L}_{\boldsymbol{\vartheta}_0}(1)\right)\right\|^l < \infty, \quad 1 \leqslant i_1, \ldots, i_\lambda \leqslant m, \quad 1 \leqslant \lambda \leqslant l-1, \quad \boldsymbol{\vartheta} \in \Theta. \tag{4.6.16}$$

If, in addition, assumptions i) to vi) of Theorem 4.27 are satisfied with X_1 replaced by $\boldsymbol{L}_{\boldsymbol{\vartheta}_0}(1)$, and if $h = h_N$ is chosen dependent on N such that Nh_N converges to zero as N tends to infinity, then the estimator $\hat{\hat{\boldsymbol{\vartheta}}}^{N,h_N}$ is consistent and asymptotically normally distributed with asymptotic covariance matrix given in Eq. (4.6.5).

Proof It suffices to check conditions vii) to ix) of Theorem 4.28. All three conditions follow by assumptions (4.6.15) and (4.6.16) from Lemma 4.26, which also shows that the function β in vii) can be taken as $\beta : h \mapsto h^{1/2}$. Consequently, the assumption that $N^{1/2}\beta(h_N)$ converges to zero from Theorem 4.28 simplifies to the requirement that Nh_N converges to zero, and the result follows. □

4.7. Simulation study

In this section we illustrate the estimation procedure developed in this chapter using the example of a univariate CARMA(3,1) process Y driven by a Gamma process. A similar example was considered in (Brockwell et al., 2011) as a model for the realized volatility of DM/$ exchange rates. Gamma processes are a family of univariate infinite activity pure-jump Lévy subordinators $(\Gamma_{b,a}(t))_{t\in\mathbb{R}}$, which are parametrized by two positive real numbers a and b (see, e. g., Applebaum, 2004, Example1.3.22). Their moment generating function is given by

$$u \mapsto \mathbb{E}\mathrm{e}^{\Gamma_{b,a}(t)u} = (1 - bu)^{-at}, \quad a, b > 0,$$

and the unit increments $\Gamma_{b,a}(n) - \Gamma_{b,a}(n-1)$ follow a Gamma distribution with scale parameter b and shape parameter a. This distribution has density

$$f_{b,a}(x) = \frac{1}{\Gamma(a)b}(x/b)^{a-1}\mathrm{e}^{-x/b},$$

mean ab, and cumulative distribution function

$$F_{b,a}(x) = \int_0^x f_{b,a}(\xi)\mathrm{d}\xi = \frac{\Gamma(a; x/b)}{\Gamma(a)}, \tag{4.7.1}$$

where $\Gamma(\cdot)$ and $\Gamma(\cdot;\cdot)$ denote the complete and the lower incomplete gamma function, respectively. In contrast to the example studied in Brockwell et al. (2011), we chose to simulate a model of order $(3, 1)$ in order to demonstrate the feasibility of approximating the derivatives $\mathrm{D}^\nu Y$ which appear in Eq. (4.5.37). The dynamics of the CARMA process used in the simulations are determined by the polynomials

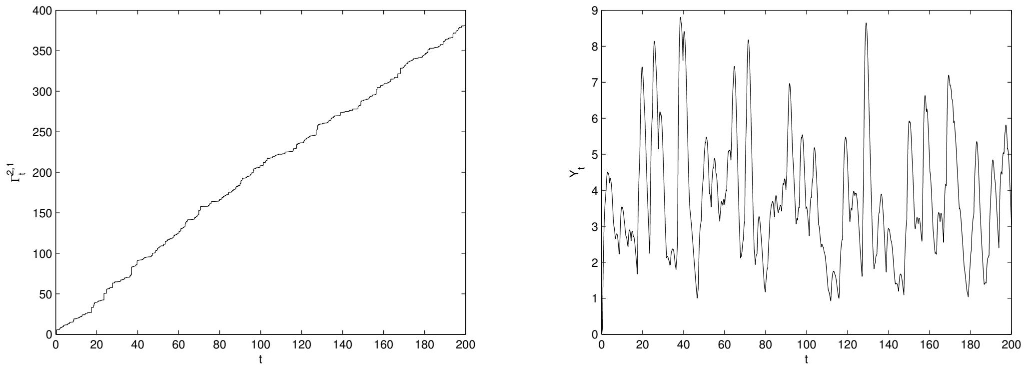
$$P(z) = z^3 + 2z^2 + \frac{3}{2}z + \frac{1}{2}, \quad \text{and} \quad Q(z) = 1 + z,$$

corresponding to autoregressive roots $\lambda_1 = -1$ and $\lambda_{2,3} = -1 \pm \mathrm{i}$. The process Y is simulated by applying an Euler scheme with step size 5×10^{-4} to the state space model (cf. Theorem 4.6)

$$\mathrm{d}\boldsymbol{X}(t) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -\frac{1}{2} & -\frac{3}{2} & -2 \end{bmatrix} \boldsymbol{X}(t)\mathrm{d}t + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \mathrm{d}\Gamma_{2,1}(t), \tag{4.7.2a}$$

$$Y(t) = \begin{bmatrix} 1 & 1 \end{bmatrix} \boldsymbol{X}(t). \tag{4.7.2b}$$

The initial value $\boldsymbol{X}(0)$ is set to zero. Another possibility would be to sample $\boldsymbol{X}(0)$ from the marginal distribution of the stationary solution of Eq. (4.7.2), but since the effect of the choice of $\boldsymbol{X}(0)$ decays at an exponential rate, this does not make a substantial difference. A typical realization of the resulting CARMA process Y on the time interval $[0, 200]$ is depicted in Fig. 4.1b. In the case of finite variation Lévy processes, there is a path-wise correspondence between a CARMA process and the driving Lévy process. Since this applies in particular to Gamma processes, it is possible to show in Fig. 4.1a the path of the driving process which generated the shown realization of Y. Such a juxtaposition is useful in that it allows to see how big jumps in the driving process can cause spikes in the resulting CARMA process.

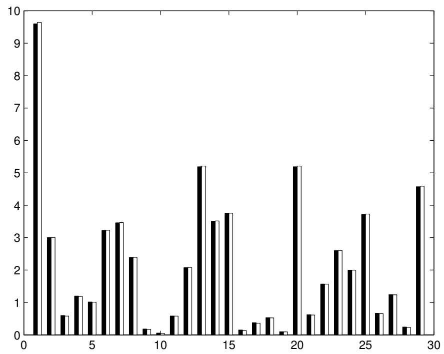


(a) $\Gamma_{2,1}$-process (b) $\Gamma_{2,1}$-driven CARMA(3,1) process

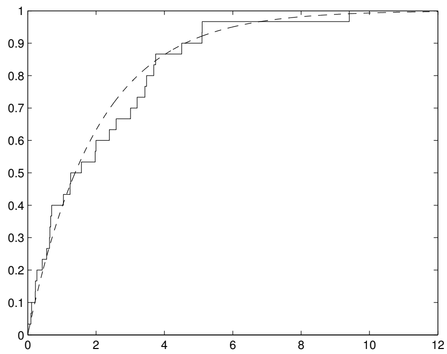
Figure 4.1.: Typical realization of a $\Gamma_{2,1}$-process and the corresponding CARMA(3,1) process with dynamics given by Eq. (4.7.2)

The first step in the implementation of our estimation procedure is to approximate the increments $\Delta\Gamma_n$ of the driving Gamma process from discrete-time observations of the

CARMA process Y. For the value $h = 0.01$ of the sampling interval, Fig. 4.2 compares the true increments with the approximations $\widehat{\Delta\Gamma}_n^{(h)}$ obtained from Eq. (4.5.37) both directly and in terms of their cumulative distribution functions. We see that the approximations $\widehat{\Delta\Gamma}_n^{(h)}$ are very good for each individual increment and that therefore the empirical distribution function of the reconstructed increments closely follows the CDF (4.7.1) of the Gamma distribution, even if the observation period is rather short.



(a) Bar chart of the increments of driving $\Gamma_{2,1}$-process. White bars represent the true increments, black bars indicate the values of the estimates.



(b) Cumulative distribution function of the increments of the driving $\Gamma_{2,1}$-process. The dashed line shows the true CDF given by Eq. (4.7.1), the solid line represents the empirical distribution function of the estimates.

Figure 4.2.: Comparison of the true increments of a Gamma process with parameters $(b, a) = (2, 1)$ to the estimates of the increments computed via Eq. (4.5.37) from discrete observations of the $\Gamma_{2,1}$-driven CARMA(3,1) process defined by Eq. (4.7.2) on the time grid $(0, 0.01, 0.02, \ldots, 30)$

In the next step we used the approximate increments $\widehat{\Delta\Gamma}_n^{(h)}$ and a standard numerical optimization routine to compute the maximum likelihood estimator

$$\left(\hat{b}^{N,(h)}, \hat{a}^{N,(h)}\right) = \operatorname{argmax}_{(a,b)\in\mathbb{R}^+\times\mathbb{R}^+} \prod_{n=1}^{N} f_{b,a}\left(\widehat{\Delta\Gamma}_n^{(h)}\right), \tag{4.7.3}$$

or, equivalently,

$$\left(\hat{b}^{N,(h)}, \hat{a}^{N,(h)}\right) = \operatorname{argmin}_{(a,b)\in\mathbb{R}^+\times\mathbb{R}^+} \left\| \sum_{n=1}^{N} \nabla_{(b,a)} \log f_{b,a}\left(\widehat{\Delta\Gamma}_n^{(h)}\right) \right\|.$$

In this form, the maximum likelihood estimator falls into the class of generalized moments estimators. From the explicit form of the function $g = \nabla_{(b,a)} \log f_{b,a}$, it is easy to check that the assumptions of Theorem 4.34 are satisfied. Since in the present case, and for maximum likelihood estimators in general, the dimension of the moment vector is equal to the dimension of the parameter space, the choice of the weighting matrices $W_{N,h}$ is irrelevant and the estimator is always best in the sense of Corollary 4.33.

With the goal of confirming the assertions of Theorem 4.34 we first focused on consistency and investigated the effect of finite sampling frequencies. Figure 4.3 visualizes the empirical means and marginal standard deviations of the maximum likelihood estimator (4.7.3) obtained from 500 independent realizations of the CARMA process Y from Eq. (4.7.2) simulated over the time horizon $[0, 200]$ and sampled at instants $(0, h, 2h, \ldots, N)$ for different values of h. The picture suggests that the estimator $\left(\hat{b}^{N,(h)}, \hat{a}^{N,(h)}\right)$ is biased for positive values of h, even as N tends to infinity, but that it is consistent as h tends to zero. This is in agreement with Theorem 4.34 and reflects the intuition that discrete sampling entails a loss of information compared with a genuinely continuous-time observation of a stochastic process.

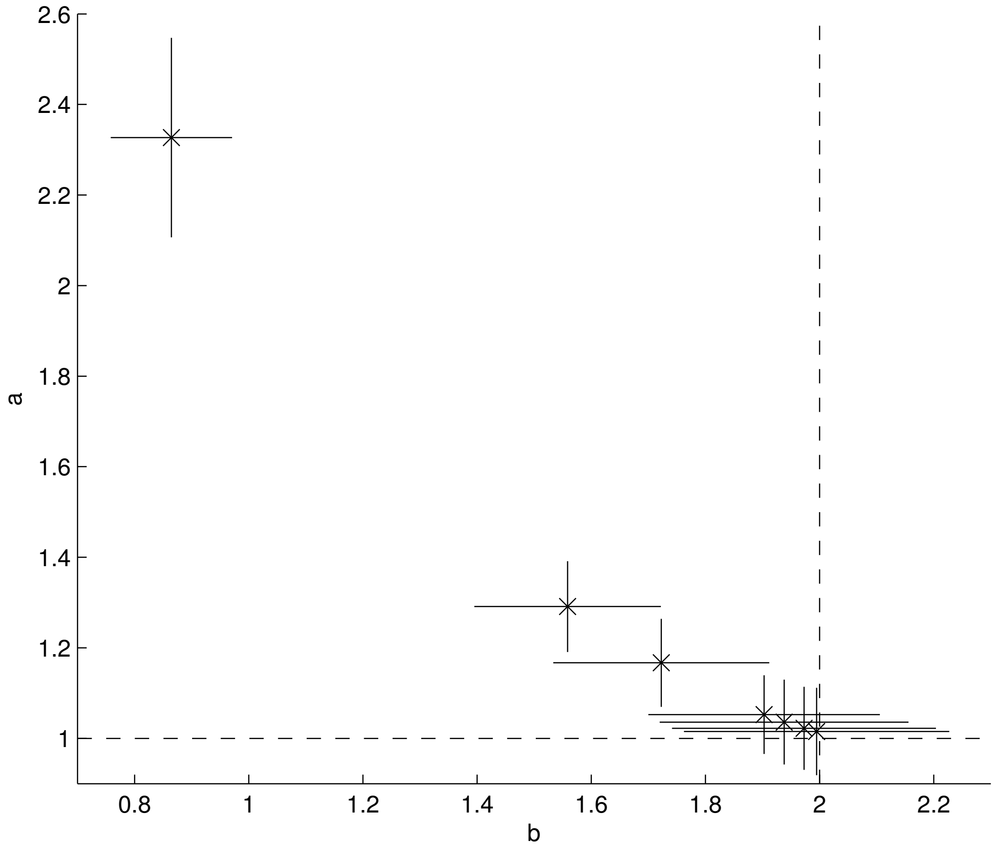


Figure 4.3.: Empirical means ($\times$) and standard deviations of the estimators $\left(\hat{b}^{200,(h)}, \hat{a}^{200,(h)}\right)$ based on 500 independent observations of the MCARMA process (4.7.2) on the time grid $(0, h, 2h, \ldots, 200)$ for $h \in \{0.5, 0.1, 0.05, 0.01, 0.005, 0.001, 0.0005\}$. The dashed lines indicate the true parameter value $(b, a) = (2, 1)$.

Finally, we conducted another Monte Carlo simulation with the goal of confirming the asymptotic normality of the maximum likelihood estimator (4.7.3). Figure 4.4 compares the empirical distribution of the estimator $\left(\hat{b}^{200,(0.001)}, \hat{a}^{200,(0.001)}\right)$ to the asymptotic normal distribution asserted by the Central Limit Theorem 4.34. The dots indicate the values of the estimates obtained from 500 independent realizations of the CARMA process (4.7.2). The dashed and solid straight lines show the empirical mean $(1.9772, 1.0217)$ of the estimates and the true values $(2, 1)$ of the parameter (b, a), respectively, which are in good agreement. The dashed and solid ellipses represent the empirical autocovariance matrix

$$\widehat{\Sigma} = \begin{pmatrix} 4.70 & -1.45 \\ -1.45 & 0.78 \end{pmatrix} \times 10^{-2}$$

of the estimates and the scaled asymptotic covariance matrix

$$\Sigma/200 \approx \begin{pmatrix} 5.11 & -1.55 \\ -1.55 & 0.78 \end{pmatrix} \times 10^{-2},$$

respectively. Their closeness, which is also reflected by the similarity of the two ellipses in Fig. 4.4, means that, even for finite observation periods and sampling frequencies, the matrix Σ/N is a good approximation of the true covariance of the estimator $\left(\hat{b}^{N,(h)}, \hat{a}^{N,(h)}\right)$, and it can thus be used for the construction of confidence regions. For the present example, the inverse of the asymptotic covariance matrix Σ, given by Eq. (4.6.5), can be computed explicitly as

$$\begin{aligned} \Sigma^{-1} &= -\mathbb{E}\left[\nabla^2_{(b,a)} \log f_{b,a}\left(\Gamma_{b,a}(1)\right)\right]_{(b,a)=(2,1)} \\ &= \left.\begin{pmatrix} a/b^2 & 1/b \\ 1/b & \psi_1(a) \end{pmatrix}\right|_{(b,a)=(2,1)} \\ &= \begin{pmatrix} 1/4 & 1/2 \\ 1/2 & \pi^2/6 \end{pmatrix}, \end{aligned}$$

where ψ_1 denotes the trigamma function, that is the second derivative of the logarithm of the gamma function. Figure 4.4 also compares histograms of $\hat{b}^{200,(0.001)}$ and $\hat{a}^{200,(0.001)}$ to the densities of the marginals of the bivariate Gaussian distribution with mean $(2, 1)$ and covariance matrix $\Sigma/200$. The agreement is very good, in accordance with the Central Limit Theorem 4.34.

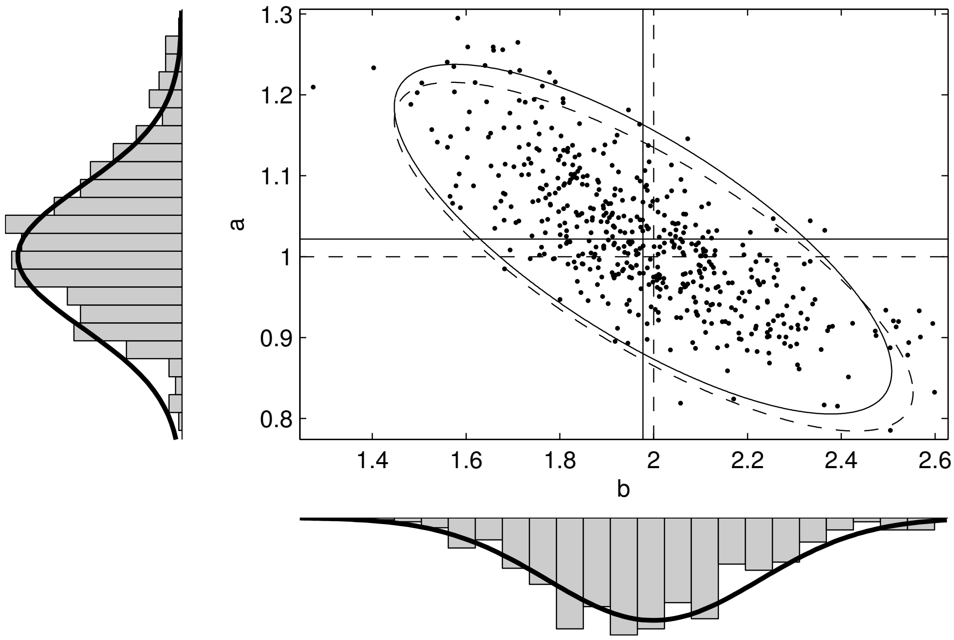


Figure 4.4.: Comparison of the empirical distribution of the estimator $\left(\hat{b}^{200,(0.001)}, \hat{a}^{200,(0.001)}\right)$ based on 500 realizations of the $\Gamma_{2,1}$-driven CARMA(3,1) process given by Eq. (4.7.2) to the asymptotic distribution implied by the Central Limit Theorem 4.34

Part II

Limiting Spectral Distributions of Random Matrix Models with Dependent Entries

5. Eigenvalue Distribution of Large Sample Covariance Matrices of Linear Processes

5.1. Introduction and main result

A typical object of interest for many statistical applications is the sample covariance matrix $(n-1)^{-1}\boldsymbol{X}\boldsymbol{X}^T$ of a data matrix $\boldsymbol{X} = (X_{i,t})_{it}$, $i = 1, \dots, p$, $t = 1, \dots, n$. The matrix $\boldsymbol{X}$ can be seen as a sample of size n of p-dimensional data vectors. For fixed p, one can show, as n tends to infinity, that under certain assumptions, such as ergodicity of the data-generating process, the eigenvalues of the sample covariance matrix converge to the eigenvalues of the true underlying covariance matrix (Anderson, 2003). However, the assumption that $p \ll n$ may not be justified if one has to deal with high-dimensional data sets; often, it is more suitable to assume that the dimension p is of the same order as the sample size n, that is $p = p_n$ tends to infinity in such a way that

$$\lim_{n\to\infty} \frac{n}{p} =: y \in (0, \infty). \tag{5.1.1}$$

For a symmetric $p \times p$ matrix A with eigenvalues $\lambda_1, \dots, \lambda_p$, we denote by

$$F^A = \frac{1}{p} \sum_{i=1}^{p} \delta_{\lambda_i}$$

the spectral distribution of A, where δ_x denotes the Dirac measure located at x. This means that $pF^A(B)$ is the number of eigenvalues of A that lie in the set B. From now on we will call $p^{-1}\boldsymbol{X}\boldsymbol{X}^T$ the sample covariance matrix. Because of Eq. (5.1.1) this change of normalization can be reversed by a simple transformation of the limiting spectral distribution. For notational convenience, we suppress the explicit dependence of the occurring matrices on n and p where this does not cause ambiguity.

It was Marchenko and Pastur (1967) who first looked at the case where the entries $\{X_{i,t}\}$ are i. i. d. random variables with finite second moments $\mathbb{E}X_{11}^2 = 1$. They showed that the empirical spectral distribution (ESD) $F^{p^{-1}\boldsymbol{X}\boldsymbol{X}^T}$ of $p^{-1}\boldsymbol{X}\boldsymbol{X}^T$ converges, as $n \to \infty$, to a non-random distribution function $\hat{F}$, called limiting spectral distribution (LSD), given by

$$\hat{F}(\mathrm{d}x) = \frac{1}{2\pi x} \sqrt{(x_+ - x)(x - x_-)} I_{\{x_- \leqslant x \leqslant x_+\}} \mathrm{d}x, \tag{5.1.2}$$

and point mass $\hat{F}(\{0\}) = 1 - y$ if $y < 1$, where $x_{\pm} = (1 \pm \sqrt{y})^2$. Here and in the following, convergence of the ESD means almost sure convergence as a random element of the space of probability measures on $\mathbb{R}$ equipped with the weak topology. In particular, as a consequence of the Marchenko–Pastur result, the eigenvalues of the sample covariance matrix of a matrix with independent entries do not converge to the eigenvalues of the true covariance matrix, which is the identity matrix and therefore only has eigenvalue one. This leads to the failure of statistics that rely on the eigenvalues of $p^{-1}\boldsymbol{X}\boldsymbol{X}^T$ which have been derived under the assumption that p is fixed, and random matrix theory is a tool to correct these statistics, see, e. g., the introduction of Johnstone (2001). In cases where the true covariance matrix is not the identity matrix, which means that the data are either dependent or have different variances, the limiting spectral distribution $\hat{F}$ can in general only be characterized in terms of a non-linear equation for its Stieltjes transform $m_{\hat{F}}$, which is defined by

$$m_{\hat{F}}(z) = \int \frac{\hat{F}(\mathrm{d}\lambda)}{\lambda - z} \quad \forall z \in \mathbb{C}^+ := \{z = u + \mathrm{i}v \in \mathbb{C} : \operatorname{Im} z = v > 0\}.$$

Conversely, the distribution $\hat{F}$ can be obtained from its Stieltjes transform $m_{\hat{F}}$ via the Stieltjes–Perron inversion formula (Bai and Silverstein, 2010, Theorem B.8), which states that

$$\hat{F}([a,b]) = \frac{1}{\pi} \lim_{\epsilon \to 0^+} \int_a^b \operatorname{Im} m_{\hat{F}}(x + \mathrm{i}\epsilon) \mathrm{d}x. \tag{5.1.3}$$

for all continuity points $a < b$ of $\hat{F}$. For a comprehensive account of random matrix theory we refer to Anderson, Guionnet and Zeitouni (2010); Bai and Silverstein (2010); Mehta (2004) and the references therein.

Our aim will be to obtain a Marchenko–Pastur type result in the case where there is dependence within the rows. More precisely, for $i = 1, \ldots, p$, the ith row of $\boldsymbol{X}$ is given by a linear process of the form

$$(X_{i,t})_{t=1,\ldots,n} = \left(\sum_{j=0}^{\infty} c_j Z_{i,t-j} \right)_{t=1,\ldots,n}, \quad c_j \in \mathbb{R}.$$

Here $(Z_{i,t})_{it}$ is an array of independent random variables satisfying

$$\mathbb{E} Z_{i,t} = 0, \quad \mathbb{E} Z_{i,t}^2 = 1, \quad \text{and} \quad \sigma_4 := \sup_{i,t} \mathbb{E} Z_{i,t}^4 < \infty, \tag{5.1.4}$$

and, for each $\epsilon > 0$,

$$\frac{1}{pn} \sum_{i=1}^{p} \sum_{j=1}^{n} \mathbb{E} \left(Z_{i,t}^2 I_{\{Z_{i,t}^2 \geqslant \epsilon n\}} \right) \to 0, \quad \text{as} \quad n \to \infty. \tag{5.1.5}$$

Clearly, the Lindeberg-type condition (5.1.5) is satisfied if all $\{Z_{i,t}\}$ are identically distributed.

The novelty of our result is that we allow for dependence within the rows, and that the equation for $m_{\hat{F}}$ is given in terms of the spectral density

$$f(\omega) = \sum_{h \in \mathbb{Z}} \gamma(h) \mathrm{e}^{-\mathrm{i}h\omega}, \quad \omega \in [0, 2\pi],$$

only, which is the Fourier transform of the autocovariance function

$$\gamma(h) = \sum_{j=0}^{\infty} c_j c_{j+|h|}, \quad h \in \mathbb{Z},$$

of the underlying linear processes X_i. Potential applications arise in contexts where data is not independent in time such that the Marchenko–Pastur law is not a good approximation. This includes areas like wireless communications (Tulino and Verdu, 2004) and mathematical finance (Plerou, Gopikrishnan, Rosenow, Amaral, Guhr and Stanley, 2002; Potters, Bouchaud and Laloux, 2005). Note that a similar question is also discussed in Bai and Zhou (2008). However, they have a different proof for the existence of the limiting spectral distribution, which relies on a moment condition to be verified. Furthermore they additionally assume that the $\{Z_{i,t}\}$ are identically distributed so that the processes within the rows are independent copies of each other. More importantly, their results do not yield concrete formulæ except in the AR(1) case and are therefore not directly applicable. In the context of free probability theory, the limiting spectral distribution of sample covariance matrices of Gaussian autoregressive moving average processes is investigated in Burda, Jarosz, Nowak and Snarska (2010).

We now present the main result of this article.

Theorem 5.1 (Limiting spectral distribution) *For each $i = 1, \ldots, p$, let $X_{i,t} = \sum_{j=0}^{\infty} c_j Z_{i,t-j}$, $t \in \mathbb{Z}$, be a linear stochastic process with continuously differentiable spectral density f. Assume that*

i) the array $(Z_{i,t})_{it}$ satisfies conditions (5.1.4) and (5.1.5),

ii) there exist positive constants C and δ such that $|c_j| \leqslant C(j+1)^{-1-\delta}$, for all $j \geqslant 0$,

iii) for almost all $\lambda \in \mathbb{R}$, $f(\omega) = \lambda$ for at most finitely many $\omega \in [0, 2\pi]$, and

iv) $f'(\omega) \neq 0$ for almost every ω.

Then the empirical spectral distribution $F^{p^{-1}\mathbf{X}\mathbf{X}^T}$ of $p^{-1}\mathbf{X}\mathbf{X}^T$ converges, as n tends to infinity, almost surely to a non-random probability distribution function $\hat{F}$ with bounded support. Moreover, there exist positive numbers λ_-, λ_+ such that the Stieltjes transform $z \mapsto m_{\hat{F}}(z)$ of $\hat{F}$ is the unique *mapping $\mathbb{C}^+ \to \mathbb{C}^+$ satisfying*

$$\frac{1}{m_{\hat{F}}(z)} = -z + \frac{y}{2\pi} \int_{\lambda_-}^{\lambda_+} \frac{\lambda}{1 + \lambda m_{\hat{F}}(z)} \sum_{\omega \in [0,2\pi]: f(\omega) = \lambda} \frac{1}{|f'(\omega)|} \mathrm{d}\lambda. \tag{5.1.6}$$

The assumptions of the theorem are met, for instance, if $(X_{i,t})_t$ is an ARMA or fractionally integrated ARMA process; see Section 5.3 for details.

Theorem 5.1, as it stands, does not contain the classical Marchenko–Pastur law as a special case. For if the entries $X_{i,t}$ of the matrix $\boldsymbol{X}$ are i. i. d., the corresponding spectral density f is identically equal to the variance of $X_{1,1}$, and thus condition iv) is not satisfied. We therefore also present a version of Theorem 5.1 that holds if the rows of the matrix $\boldsymbol{X}$ have a piecewise constant spectral density.

Theorem 5.2 (Limiting spectral distribution) *For each $i = 1, \ldots, p$, let $X_{i,t} = \sum_{j=0}^{\infty} c_j Z_{i,t-j}$, $t \in \mathbb{Z}$, be a linear stochastic process with spectral density f of the form*

$$f : [0, 2\pi] \to \mathbb{R}^+, \quad \omega \mapsto \sum_{j=1}^{k} \alpha_j I_{A_j}(\omega), \quad k \in \mathbb{N}, \tag{5.1.7}$$

for some positive real numbers α_j and a measurable partition $A_1 \cup \cdots \cup A_k$ of the interval $[0, 2\pi]$. If conditions i) and ii) of Theorem 5.1 hold, then the empirical spectral distribution $F^{p^{-1}\boldsymbol{X}\boldsymbol{X}^T}$ of $p^{-1}\boldsymbol{X}\boldsymbol{X}^T$ converges, as $n \to \infty$, almost surely to a non-random probability distribution function $\hat{F}$ with bounded support. Moreover, the Stieltjes transform $z \mapsto m_{\hat{F}}(z)$ of $\hat{F}$ is the unique mapping $\mathbb{C}^+ \to \mathbb{C}^+$ that satisfies

$$\frac{1}{m_{\hat{F}}(z)} = -z + \frac{y}{2\pi} \sum_{j=1}^{k} \frac{|A_j| \alpha_j}{1 + \alpha_j m_{\hat{F}}(z)}, \tag{5.1.8}$$

where $|A_j|$ denotes the Lebesgue measure of the set A_j. In particular, if the entries of $\boldsymbol{X}$ are i. i. d. with unit variance, one recovers the limiting spectral distribution (5.1.2) of the Marchenko–Pastur law.

Remark 5.3 In applications one often has $X_{i,t} = \mu + \sum_{j=0}^{\infty} c_j Z_{i,t-j}$ with non-zero mean $\mu \neq 0$. Denote by $x_t \in \mathbb{R}^p$ the tth column of $\boldsymbol{X}$, and define the empirical mean $\overline{x} = p^{-1} \sum_{t=1}^{n} x_t$. Then one rather considers $p^{-1} \sum_{t=1}^{n} (x_t - \overline{x})(x_t - \overline{x})^T$ instead of $p^{-1}\boldsymbol{X}\boldsymbol{X}^T$. However, by Bai and Silverstein (2010, Theorem A.44), these two matrices have the same LSD, and thus Theorems 5.1 and 5.2 remain valid in this case.

Remark 5.4 The proofs of Theorems 5.1 and 5.2 can easily be generalized to cover non-causal processes, where $X_{i,t} = \sum_{j=-\infty}^{\infty} c_j Z_{i,t-j}$ is given as a two-sided moving average. For this case one obtains the same results, except that the autocovariance function is now given by $\sum_{j=-\infty}^{\infty} c_j c_{j+|h|}$.

Remark 5.5 If one considers a matrix $\boldsymbol{X}$ with independent linear processes in the columns instead of the rows, one gets the same formulæ as in Theorems 5.1 and 5.2, except that y is replaced by y^{-1}. This is due to the fact that $\boldsymbol{X}^T\boldsymbol{X}$ and $\boldsymbol{X}\boldsymbol{X}^T$ have the same non-trivial eigenvalues.

Remark 5.6 If the assumption that the entries of the matrix $\boldsymbol{X}$ have finite second moments is dropped, the asymptotic behaviour of the eigenvalues of the sample covariance matrix $p^{-1}\boldsymbol{X}\boldsymbol{X}$ changes drastically, even in the i. i. d. case. For an introduction to the spectral analysis of heavy-tailed random matrices we refer the reader to Auffinger, Ben Arous and Péché (2009); Belinschi, Dembo and Guionnet (2009); Ben Arous and Guionnet (2008), and the references therein.

Remark 5.7 An interesting open problem is to describe the limiting distribution of the largest eigenvalue of the sample covariance matrix $p^{-1}\boldsymbol{X}\boldsymbol{X}$ if the rows of $\boldsymbol{X}$ consist of independent linear processes. It is expected to find a variant of the Tracy–Widom distribution, which describes the largest eigenvalue in the classical random matrix ensembles with i. i. d. entries (Tracy and Widom, 1994, 1996), but also occurs in a variety of seemingly unrelated stochastic growth models (Borodin, Ferrari, Prähofer and Sasamoto, 2007; Prähofer and Spohn, 2000).

Outline of the chapter In Section 5.2 we proceed with the proofs of Theorems 5.1 and 5.2. Thereafter we present some interesting examples in Section 5.3.

Notation We write $\mathbb{E}$ and $\mathbb{V}\mathrm{ar}$ for the expected value and the variance, respectively, of a random variable. The symbol $\operatorname{tr} S$ denotes the trace of a quadratic matrix S. The ESD of a matrix sequence $S = S_n$ is denoted by F^S, and their weak limit, provided that it exists, by $\hat{F}^S$. The notation $I_{\{\mathcal{E}\}}$ is used for the indicator of an expression $\mathcal{E}$, which, by definition, equals one if $\mathcal{E}$ is true, and zero otherwise.

5.2. Proofs

In this section we present a proof of Theorems 5.1 and 5.2. It turns out to be difficult to deal with infinite-order moving average processes directly, and we therefore first prove a variant of these theorems for the truncated processes $\widetilde{X}_{i,t} = \sum_{j=0}^{n} c_j Z_{i,t-j}$. We also define the matrix $\widetilde{\boldsymbol{X}} = (\widetilde{X}_{i,t})_{it}$, $i = 1, \dots, p$, $t = 1, \dots, n$.

Proposition 5.8 *Under the assumptions of Theorem 5.1 (Theorem 5.2), the empirical spectral distribution $F^{p^{-1}\widetilde{X}\widetilde{X}^T}$ of the sample covariance matrix of the truncated process $\widetilde{X}$ converges, as n tends to infinity, to a deterministic distribution with bounded support. Its Stieltjes transform is uniquely determined by Eq. (5.1.6) (Eq. (5.1.8)).*

Proof The proof starts with the observation that one can write $\widetilde{\boldsymbol{X}} = \boldsymbol{Z}H$, where $\boldsymbol{Z} = (Z_{i,t})_{it}$, $i = 1, \ldots, p$, $t = 1-n, \ldots, n$, and

$$H = \begin{pmatrix} c_n & 0 & \cdots & 0 \\ c_{n-1} & c_n & & \\ \vdots & & \ddots & \\ c_1 & & & c_n \\ c_0 & c_1 & \cdots & c_{n-1} \\ & c_0 & & \vdots \\ & & \ddots & \vdots \\ 0 & \cdots & 0 & c_0 \end{pmatrix} \in \mathbb{R}^{2n \times n}. \tag{5.2.1}$$

In particular, $\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^T = \boldsymbol{Z}HH^T\boldsymbol{Z}^T$. In order to prove convergence of the empirical spectral distribution $F^{p^{-1}\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^T}$ and obtain a characterization of the limiting distribution, it suffices, by Pan (2010, Theorem 1), to prove that the spectral distribution F^{HH^T} of HH^T converges to a non-trivial limiting distribution. This will be done in Lemma 5.9, where the LSD of HH^T is shown to be given by $\hat{F}^{HH^T} = \frac{1}{2}\delta_0 + \frac{1}{2}\hat{F}^\Gamma$; the distribution $\hat{F}^\Gamma$ is computed in Lemma 5.10 if we impose the assumptions of Theorem 5.1, respectively in Lemma 5.11 if we impose the assumptions of Theorem 5.2. Inserting that expression for $\hat{F}^{HH^T}$ into equation (1.2) of Pan (2010) shows that the ESD $F^{p^{-1}\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^T}$ converges, as $n \to \infty$, almost surely to a deterministic distribution, which is uniquely determined by the requirement that its Stieltjes transform $z \mapsto m(z)$ satisfies

$$\frac{1}{m(z)} = -z + 2y\int_{\lambda_-}^{\lambda_+} \frac{\lambda}{1+\lambda m(z)} \mathrm{d}\hat{F}^{HH^T} = -z + y\int_{\lambda_-}^{\lambda_+} \frac{\lambda}{1+\lambda m(z)} \mathrm{d}\hat{F}^{\Gamma}. \tag{5.2.2}$$

Using the explicit formulæ for the LSD $\hat{F}^\Gamma$ computed in Lemmata 5.10 and 5.11, one obtains Eqs. (5.1.6) and (5.1.8). Uniqueness of a mapping $m : \mathbb{C}^+ \to \mathbb{C}^+$ solving Eq. (5.2.2) was shown in Bai and Silverstein (2010, p. 88). We complete the proof by arguing that the LSD of $p^{-1}\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^T$ has bounded support. For this it is enough, by Bai and Silverstein (2010, Theorem 6.3), to show that the spectral norm of HH^T is bounded in n, which is also done in Lemma 5.9. □

Lemma 5.9 *Let $H = (c_{n-i+j} I_{\{0 \leqslant n-i+j \leqslant n\}})_{ij}$ be the matrix appearing in Eq. (5.2.1) and assume that there exist positive constants C, δ such that $|c_j| \leqslant C(j+1)^{-1-\delta}$ (assumption ii) of Theorem 5.1). Then the spectral norm of the matrix HH^T is bounded in n. If, moreover, the spectral distribution of $\Gamma = (\gamma(i-j))_{ij}$ converges weakly to some limiting distribution $\hat{F}^\Gamma$, then the spectral distribution F^{HH^T} converges weakly, as $n \to \infty$, to $\frac{1}{2}\delta_0 + \frac{1}{2}\hat{F}^\Gamma$.*

Proof We first introduce the notation $\mathcal{H} := HH^T \in \mathbb{R}^{2n\times 2n}$ as well as the block decomposition

$$\mathcal{H} = \begin{bmatrix} \mathcal{H}_{11} & \mathcal{H}_{12} \\ \mathcal{H}_{12}^T & \mathcal{H}_{22} \end{bmatrix}, \quad \mathcal{H}_{ij} \in \mathbb{R}^{n\times n}.$$

We prove the second part of the lemma first. There are several ways to show that the spectral distributions of two sequences of matrices converge to the same limit. In our case, it is convenient to use Bai and Silverstein (2010, Corollary A.41), which states that two sequences A_n and B_n, either of whose empirical spectral distribution converges, have the same limiting spectral distribution if $n^{-1}\operatorname{tr}(A_n - B_n)(A_n - B_n)^T$ converges to zero as $n \to \infty$. We shall employ this result twice: first to show that the LSDs of $\mathcal{H} = HH^T$ and $\widetilde{\mathcal{H}} := \operatorname{diag}(0, \mathcal{H}_{22})$ agree, and then to prove equality of the LSDs of the matrices $\mathcal{H}_{22}$ and Γ. Let

$$\Delta_{\mathcal{H}} = n^{-1}\operatorname{tr}(\mathcal{H} - \widetilde{\mathcal{H}})(\mathcal{H} - \widetilde{\mathcal{H}})^T; \tag{5.2.3}$$

a direct calculation shows that $\Delta_{\mathcal{H}} = n^{-1}\left[\operatorname{tr}\mathcal{H}_{11}\mathcal{H}_{11}^T + 2\operatorname{tr}\mathcal{H}_{12}\mathcal{H}_{12}^T\right]$ and we will consider each of the two terms in turn. From the definition of H it follows that the (i,j)th entry of $\mathcal{H}$ is given by $\mathcal{H}^{ij} = \sum_{k=1}^n c_{n-i+k}c_{n-j+k}I_{\{\max(i,j)-n\leqslant k\leqslant \min(i,j)\}}$. The trace of the square of the upper left block of $\mathcal{H}$ therefore satisfies

$$\begin{aligned}
\operatorname{tr}\mathcal{H}_{11}\mathcal{H}_{11}^T = \sum_{i,j=1}^n \left\{\mathcal{H}^{ij}\right\}^2 &= \sum_{i,j=1}^n \left[\sum_{k=1}^{\min(i,j)} c_{n-i+k}c_{n-j+k}\right]^2 \\
&\leqslant \sum_{i,j,k,l=1}^n |c_{i+k-1}||c_{j+k-1}||c_{i+l-1}||c_{j+l-1}| \\
&\leqslant C^4 \sum_{i,j,k,l=2}^{n+1} i^{-1-\delta}j^{-1-\delta}l^{-1-\delta}k^{-1-\delta} \\
&< [C\zeta(1+\delta)]^4 < \infty,
\end{aligned}$$

where $\zeta(z)$ denotes the Riemann zeta function. As a consequence, the limit of $n^{-1}\operatorname{tr}\mathcal{H}_{11}\mathcal{H}_{11}^T$, as n tends to infinity, is zero. Similarly, we obtain for the trace of the square of the off-diagonal block of $\mathcal{H}$

$$\begin{aligned}
\operatorname{tr}\mathcal{H}_{12}\mathcal{H}_{12}^T = \sum_{i=1}^n \sum_{j=n+1}^{2n} \left\{\mathcal{H}^{ij}\right\}^2 &= \sum_{i=1}^n \sum_{j=n+1}^{n+i} \left[\sum_{k=j-n}^{i} c_{n-i+k}c_{n-j+k}\right]^2 \\
&\leqslant \sum_{i=1}^n \sum_{j=1}^n \sum_{k=j}^{n-i+1} \sum_{l=j}^{n-i+1} c_{i+k-1}c_{k-j}c_{i+l-1}c_{l-j} \\
&\leqslant \sum_{i=1}^n \sum_{j=1}^n \sum_{r=0}^n \sum_{s=0}^n |c_{i+r+j-1}||c_r||c_{s+j-1}||c_s|
\end{aligned}$$

$$\begin{aligned}
&\leqslant C^4 \sum_{i,j,r,s=1}^{n+1} i^{-1-\delta} r^{-1-\delta} s^{-1-\delta} j^{-1-\delta} \\
&< [C\zeta(1+\delta)]^4 < \infty,
\end{aligned}$$

which shows that the limit of $n^{-1} \operatorname{tr} \mathcal{H}_{12}\mathcal{H}_{12}^T$, as $n \to \infty$, is zero. It thus follows that $\Delta_{\mathcal{H}}$, as defined in Eq. (5.2.3), converges to zero as $n \to \infty$, and, therefore, that the LSDs of $\mathcal{H}$ and $\widetilde{\mathcal{H}} = \operatorname{diag}(0, \mathcal{H}_{22})$ coincide. The latter is clearly given by $F^{\widetilde{\mathcal{H}}} = \frac{1}{2}\delta_0 + \frac{1}{2}F^{\mathcal{H}_{22}}$, and we show next that the LSD of $\mathcal{H}_{22}$ agrees with the LSD of $\Gamma = (\gamma(i-j))_{ij}$. As before, it suffices to show, by (Bai and Silverstein, 2010, Corollary A.41), that $\Delta_\Gamma = n^{-1} \operatorname{tr}(\mathcal{H}_{22} - \Gamma)(\mathcal{H}_{22} - \Gamma)^T$ converges to zero as $n \to \infty$. It follows from the definitions of the matrices $\mathcal{H}$ and Γ that $n\Delta_\Gamma$ is given by

$$\begin{aligned}
n\Delta_\Gamma &= \sum_{i,j=1}^{n} \left[\sum_{k=\max(i,j)}^{n} c_{k-i}c_{k-j} - \sum_{k=1}^{\infty} c_{k-1}c_{k+|i-j|-1} \right]^2 \\
&= \sum_{i,j=1}^{n} \left[\sum_{k=\max(i,j)}^{n} c_{k-i}c_{k-j} - \sum_{k=\max(i,j)}^{\infty} c_{k-i}c_{k-j} \right]^2 \\
&= \sum_{i,j=1}^{n} \sum_{k,l=1}^{\infty} c_{k+i-1}c_{k+j-1}c_{l+i-1}c_{l+j-1} \\
&\leqslant C^4 \sum_{i,j=2}^{n+1} \sum_{k,l=2}^{\infty} i^{-1-\delta} j^{-1-\delta} k^{-1-\delta} l^{-1-\delta} < [C\zeta(1+\delta)]^4 < \infty.
\end{aligned}$$

Consequently, the sequence Δ_Γ converges to zero as n goes to infinity, and it follows that $\hat{F}^{\mathcal{H}} = \frac{1}{2}\delta_0 + \frac{1}{2}\hat{F}^\Gamma$.

In order to show that the spectral norm of $\mathcal{H} = HH^T$ is bounded in n, we use Gerschgorin's circle theorem (Gerschgorin, 1931, Theorem 2), which states that every eigenvalue of $\mathcal{H}$ lies in at least one of the balls $B(\mathcal{H}^{ii}, R_i)$ with centre $\mathcal{H}^{ii}$ and radius R_i, $i = 1, \ldots, 2n$, where the radii R_i are defined as $R_i = \sum_{j \neq i} |\mathcal{H}^{ij}|$. We first note that the centres $\mathcal{H}^{ii}$ satisfy

$$\mathcal{H}^{ii} = \sum_{k=\max\{1,i-n\}}^{\min\{i,n\}} c_{n-i+k}^2 \leqslant \sum_{k=0}^{n} c_k^2 \leqslant [C\zeta(2+2\delta)]^2 < \infty.$$

To obtain a uniform bound for the radii R_i, we first assume that $i = 1, \ldots, n$. Then

$$\begin{aligned}
|R_i| &\leqslant \sum_{j=1}^{n} \sum_{k=1}^{\min\{i,j\}} |c_{n-i+k}||c_{n-j+k}| + \sum_{j=n+1}^{2n} \sum_{k=j-n}^{i} |c_{n-i+k}||c_{n-j+k}| \\
&\leqslant \sum_{j,k=1}^{n} |c_{n-i+k}||c_{j+k-1}| + \sum_{j=n+1-i}^{2n-i} \sum_{k=0}^{n-j} |c_{k+j}||c_k| \leqslant 2\,[C\zeta(1+\delta)]^2 < \infty.
\end{aligned}$$

Similarly we find that, for $i = n+1, \dots, 2n$,

$$
\begin{aligned}
|R_i| &\leqslant \sum_{j=1}^{n} \sum_{k=i-n}^{j} |c_{n-i+k}||c_{n-j+k}| + \sum_{j=n+1}^{2n} \sum_{k=\max\{i,j\}-n}^{n} |c_{n-i+k}||c_{n-j+k}| \\
&\leqslant \sum_{j=i-n}^{i-1} \sum_{k=0}^{n+1-j} |c_{k+j}||c_k| + \sum_{j=n+1}^{2n} \sum_{k=0}^{n-\max\{i,j\}} |c_k||c_{k+|j-i|}| \\
&\leqslant 3\,[C\zeta(1+\delta)]^2 < \infty
\end{aligned}
$$

is bounded, which completes the proof. □

In the following two lemmata, we argue that the distribution $\hat{F}^{\Gamma}$ exists and we prove explicit formulæ for it in the case that the assumptions of Theorem 5.1 or Theorem 5.2 are satisfied.

Lemma 5.10 *Let $(c_j)_j$ be a sequence of real numbers, $\gamma : h \mapsto \sum_{j=0}^{\infty} c_j c_{j+|h|}$ and $f : \omega \mapsto \sum_{h\in\mathbb{Z}} \gamma(h) \mathrm{e}^{-ih\omega}$. Under the assumptions of Theorem 5.1 it holds that the spectral distribution F^{Γ} of $\Gamma = (\gamma(i-j))_{ij}$ converges weakly, as $n \to \infty$, to an absolutely continuous distribution $\hat{F}^{\Gamma}$ with bounded support and density*

$$g : (\lambda_-, \lambda_+) \to \mathbb{R}^+, \quad \lambda \mapsto \frac{1}{2\pi} \sum_{\omega: f(\omega)=\lambda} \frac{1}{|f'(\omega)|}. \tag{5.2.4}$$

Proof We first note that under assumption ii) of Theorem 5.1 the autocovariance function γ is absolutely summable because

$$\sum_{h=0}^{\infty} |\gamma(h)| \leqslant \sum_{h=0}^{\infty} \sum_{j=0}^{\infty} |c_j||c_{j+h}| \leqslant C^2 \sum_{h,j=1}^{\infty} h^{-1-\delta} j^{-1-\delta} < [C\zeta(1+\delta]^2 < \infty.$$

Szegő's first convergence theorem, which can be found in Grenander and Szegő (1984) or Gray (2006, Corollary 4.1), then implies that $\hat{F}^{\Gamma}$ exists, and that the cumulative distribution function of the eigenvalues of the Toeplitz matrix Γ associated with the sequence $h \mapsto \gamma(h)$ is given by

$$G(\lambda) := \frac{1}{2\pi} \int_0^{2\pi} I_{\{f(\omega)\leqslant\lambda\}} \mathrm{d}\omega = \frac{1}{2\pi} \operatorname{Leb}(\{\omega \in [0, 2\pi] : f(\omega) \leqslant \lambda\}) \tag{5.2.5}$$

for all λ such that the level sets $\{\omega \in [0, 2\pi] : f(\omega) = \lambda\}$ have Lebesgue measure zero. By assumption iii) of Theorem 5.1, Eq. (5.2.5) holds for almost all λ. In order to prove that the LSD $\hat{F}^{\Gamma}$ is absolutely continuous with respect to the Lebesgue measure, it suffices to prove that the cumulative distribution function G is differentiable almost everywhere. Clearly, for $\Delta\lambda > 0$,

$$G(\lambda + \Delta\lambda) - G(\lambda) = \frac{1}{2\pi} \operatorname{Leb}(\{\omega \in [0, 2\pi] : \lambda < f(\omega) \leqslant \lambda + \Delta\lambda\}).$$

Due to assumption iv) of Theorem 5.1, the set of all $\lambda \in \mathbb{R}$ such that the set $\{\omega :\in [0, 2\pi] : f(\omega) = \lambda \text{ and } f'(\omega) = 0\}$ is non-empty is a Lebesgue null-set. Hence, it is enough to consider only λ for which this set is empty. Let $f^{-1}(\lambda) = \{\omega : f(\omega) = \lambda\}$ be the pre-image of λ, which is a finite set by assumption iii). The Implicit Function Theorem then asserts that, for every $\omega \in f^{-1}(\lambda)$, there exists an open interval I_ω around ω such that f restricted to I_ω is invertible. It is no restriction to assume that these I_ω are disjoint. By choosing $\Delta\lambda$ sufficiently small it can be ensured that the interval $[\lambda, \Delta\lambda]$ is contained in $\bigcap_{\omega \in f^{-1}(\lambda)} f(I_\omega)$, and from the continuity of f it follows that outside of $\bigcup_{\omega \in f^{-1}(\lambda)} I_\omega$, the values of f are bounded away from λ, so that

$$\begin{aligned}
&\lim_{\Delta\lambda \to 0} \frac{1}{\Delta\lambda} \left[G(\lambda + \Delta\lambda) - G(\lambda) \right] \\
=&\frac{1}{2\pi} \lim_{\Delta\lambda \to 0} \frac{1}{\Delta\lambda} \operatorname{Leb} \left(\bigcup_{\omega \in f^{-1}(\lambda)} \{\omega' \in I_\omega : \lambda < f(\omega') \leqslant \lambda + \Delta\lambda\} \right) \\
=&\frac{1}{2\pi} \sum_{\omega \in f^{-1}(\lambda)} \lim_{\Delta\lambda \to 0} \frac{1}{\Delta\lambda} \operatorname{Leb} \left(\{\omega' \in I_\omega : \lambda < f(\omega') \leqslant \lambda + \Delta\lambda\} \right).
\end{aligned}$$

In order to further simplify this expression, we denote the local inverse functions by $f_\omega^{-1} : f(I_\omega) \to [0, 2\pi]$. Observing that the Lebesgue measure of an interval is given by its length and that the derivatives of f_ω^{-1} are given by the inverses of the derivative of f, it follows that

$$\begin{aligned}
&\lim_{\Delta\lambda \to 0} \frac{1}{\Delta\lambda} \left[G(\lambda + \Delta\lambda) - G(\lambda) \right] \\
=&\frac{1}{2\pi} \sum_{\omega \in f^{-1}(\lambda)} \lim_{\Delta\lambda \to 0} \frac{1}{\Delta\lambda} \left| f_\omega^{-1}(\lambda + \Delta\lambda) - f_\omega^{-1}(\lambda) \right| \\
=&\frac{1}{2\pi} \sum_{\omega \in f^{-1}(\lambda)} \left| \frac{\mathrm{d}}{\mathrm{d}\lambda} f_\omega^{-1}(\lambda) \right| \\
=&\frac{1}{2\pi} \sum_{\omega \in f^{-1}(\lambda)} \frac{1}{|f'(\omega)|}.
\end{aligned}$$

This shows that the function G is differentiable Lebesgue-almost everywhere with derivative given by

$$g : \lambda \mapsto \frac{1}{2\pi} \sum_{\omega \in f^{-1}(\lambda)} \frac{1}{|f'(\omega)|}.$$

It remains to argue that the support of the limiting spectral distribution $\hat{F}^\Gamma$ is bounded. The absolute summability of the autocovariance function $\gamma(\cdot)$ implies boundedness of its Fourier transform f. The claim then follows from Eq. (5.2.5), which shows that the support of g is equal to the range of f. $\square$

Lemma 5.11 *Let $f : \omega \mapsto \sum_{j=1}^{k} \alpha_j I_{A_j}(\omega)$ be the piecewise constant spectral density of the linear process $X_t = \sum_{j=0}^{\infty} c_j Z_{t-j}$, $t \in \mathbb{R}$, and denote the corresponding autocovariance function by $\gamma : h \mapsto \sum_{j=0}^{\infty} c_j c_{j+|h|}$. Under the assumptions of Theorem 5.2 it holds that the spectral distribution F^{Γ} of the Toeplitz matrix $\Gamma = (\gamma(i-j))_{ij}$ converges weakly, as $n \to \infty$, to the distribution $\hat{F}^{\Gamma} = (2\pi)^{-1} \sum_{j=1}^{k} |A_j| \delta_{\alpha_j}$.*

Proof Without loss of generality we may assume that the numbers $\alpha_1 < \ldots < \alpha_k$ are sorted in increasing order. As in the proof of Lemma 5.10 one sees that the limiting spectral distribution $\hat{F}^{\Gamma}$ exists, and that $\hat{F}^{\Gamma}(-\infty, \lambda)$ is given by

$$G(\lambda) := \frac{1}{2\pi} \operatorname{Leb}(\{\omega \in [0, 2\pi] : f(\omega) \leqslant \lambda\}), \quad \forall \lambda \in [0, 2\pi] \setminus \bigcup_{j=1}^{k} \{\alpha_j\}.$$

The particular structure of f thus implies that $G(\lambda) = (2\pi)^{-1} \sum_{j=1}^{k_\lambda} |A_j|$, where k_λ is the largest integer such that $\alpha_{k_\lambda} \leqslant \lambda$ or zero if no such integer exists. Since G must be right-continuous, this formula holds for all λ in the interval $[0, 2\pi]$. It is easy to see that the function G is the cumulative distribution function of the discrete measure $(2\pi)^{-1} \sum_{j=1}^{k} |A_j| \delta_{\alpha_j}$, which completes the proof. □

We can now complete the proofs of our main theorems.

Proof (of Theorems 5.1 and 5.2) It is only left to show that the truncation performed in Proposition 5.8 does not alter the LSD, that is that the difference of $F^{p^{-1}\boldsymbol{X}\boldsymbol{X}^T}$ and $F^{p^{-1}\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^T}$ converges to zero almost surely. Again, by Bai and Silverstein (2010, Corollary A.42), this means that we need to show that

$$\underbrace{\frac{1}{p^2} \operatorname{tr}(\boldsymbol{X}\boldsymbol{X}^T + \widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^T)}_{=\mathrm{I}} \underbrace{\frac{1}{p^2} \operatorname{tr}((\boldsymbol{X} - \widetilde{\boldsymbol{X}})(\boldsymbol{X} - \widetilde{\boldsymbol{X}})^T)}_{=\mathrm{II}} \tag{5.2.6}$$

converges to zero. To this end we show that I has a limit and that II converges to zero, both almost surely. By the definition of $\boldsymbol{X}$ and $\widetilde{\boldsymbol{X}}$ we have

$$\mathrm{II} = \frac{1}{p^2} \sum_{i=1}^{p} \sum_{t=1}^{n} \sum_{k=n+1}^{\infty} \sum_{m=n+1}^{\infty} c_k c_m Z_{i,t-k} Z_{i,t-m}.$$

We shall first prove that the variances of II are summable. For this purpose we need the following two estimates which are implied by the Cauchy–Schwarz inequality, the assumption that $\sigma_4 = \sup_{i,t} \mathbb{E} Z_{i,t}^4 < \infty$, and the assumed absolute summability of the coefficients $(c_j)_j$:

$$\mathbb{E} \sum_{i=1}^{p} \sum_{t=1}^{n} \sum_{k,m=1}^{\infty} |c_k c_m Z_{i,t-k} Z_{i,t-m}| \leqslant pn \left(\sum_{k=1}^{\infty} |c_k| \right)^2 < \infty, \tag{5.2.7a}$$

$$\mathbb{E} \sum_{i,i'=1=1}^{p} \sum_{t,t'=1}^{n} \sum_{k,k',m,m'=1}^{\infty} |c_k c_m c_{k'} c_{m'} Z_{i,t-k} Z_{i,t-m} Z_{i',t'-k'} Z_{i',t'-m'}|.$$

$$\leqslant (np)^2 \sigma_4 \left(\sum_{k=1}^{\infty} |c_k| \right)^4 < \infty. \tag{5.2.7b}$$

Therefore we can, by Fubini's theorem, interchange expectation and summation to bound the variance

$$\mathbb{V}\mathrm{ar}(\mathrm{II}) \leqslant \frac{1}{p^4} \sum_{i,i'=1}^{p} \sum_{t,t'=1}^{n} \sum_{\substack{k,k' \\ m,m'}=n+1}^{\infty} c_k c_m c_{k'} c_{m'} \mathbb{E}(Z_{i,t-k} Z_{i,t-m} Z_{i',t'-k'} Z_{i',t'-m'}).$$

Considering separately the terms where either $i = i'$, or else $i \neq i'$, we can write

$$\begin{aligned} \mathbb{V}\mathrm{ar}(\mathrm{II}) \leqslant & \frac{1}{p^4} \sum_{\substack{i,i'=1 \\ i \neq i'}}^{p} \sum_{t,t'=1}^{n} \sum_{\substack{k,k' \\ m,m'}=n+1}^{\infty} c_k c_m c_{k'} c_{m'} \mathbb{E}(Z_{i,t-k} Z_{i,t-m} Z_{i',t'-k'} Z_{i',t'-m'}) \\ & + \frac{1}{p^4} \sum_{i=1}^{p} \sum_{t,t'=1}^{n} \sum_{\substack{k,k' \\ m,m'}=n+1}^{\infty} c_k c_m c_{k'} c_{m'} \mathbb{E}(Z_{i,t-k} Z_{i,t-m} Z_{i,t'-k'} Z_{i,t'-m'}). \end{aligned}$$

For the expectation in the first sum not to be zero, k must equal m and k' must equal m', in which case its value is unity. The expectation in the second term can always be bounded by σ_4 so that we obtain

$$\mathbb{V}\mathrm{ar}(\mathrm{II}) \leqslant \frac{p^2 - p}{p^4} n^2 \left(\sum_{k=n+1}^{\infty} c_k^2 \right)^2 + \sigma_4 \frac{pn^2}{p^4} \left(\sum_{k=n+1}^{\infty} c_k \right)^4 .$$

Due to Eq. (5.1.1) and assumption ii) that the coefficients (c_k) decay at least polynomially as $k^{-1-\delta}$, there exists a constant K such that the right hand side is bounded by $Kn^{-1-4\delta}$, which implies that

$$\sum_{n=1}^{\infty} \mathbb{V}\mathrm{ar}\,(\mathrm{II}) \leqslant K \sum_{n=1}^{\infty} n^{-1-4\delta} < \infty,$$

and therefore, by the Borel–Cantelli lemma, that II converges to a constant almost surely.To show that this constant is zero, it suffices to show that the expectation of II converges to zero. Since $\mathbb{E} Z_{i,t} = 0$, and the $\{Z_{i,t}\}$ are independent, one sees, using Eq. (5.2.7a), that

$$\mathbb{E}(\mathrm{II}) = np^{-1} \sum_{k=n+1}^{\infty} c_k^2,$$

which converges to zero because the coefficients $\{c_k\}$ are square-summable.

We now consider factor I of expression (5.2.6). With the definition $\Delta_X = \boldsymbol{X}\boldsymbol{X}^T - \widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^T$ we obtain that

$$\mathrm{I} = \underbrace{\frac{1}{p^2}\operatorname{tr}(\Delta_X)}_{=\mathrm{I}_\mathrm{a}} + 2\underbrace{\frac{1}{p^2}\operatorname{tr}(\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^T)}_{=\mathrm{I}_\mathrm{b}}. \tag{5.2.8}$$

Because of

$$(\boldsymbol{X}\boldsymbol{X}^T)_{ii} = \sum_{t=1}^{n} X_{i,t}^2 = \sum_{t=1}^{n}\sum_{k=0}^{\infty}\sum_{m=0}^{\infty} c_k c_m Z_{i,t-k} Z_{i,t-m},$$

and, similarly,

$$(\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^T)_{ii} = \sum_{t=1}^{n}\sum_{k=0}^{n}\sum_{m=0}^{n} c_k c_m Z_{i,t-k} Z_{i,t-m},$$

we have that

$$\begin{aligned}\operatorname{tr}(\Delta_X) &= \sum_{i=1}^{p}\left[(\boldsymbol{X}\boldsymbol{X}^T)_{ii} - (\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^T)_{ii}\right] \\ &= \underbrace{\sum_{i=1}^{p}\sum_{t=1}^{n}\sum_{k=n+1}^{\infty}\sum_{m=n+1}^{\infty} c_k c_m Z_{i,t-k} Z_{i,t-m}}_{=\mathrm{II}\to 0 \text{ a.s.}} + 2\sum_{i=1}^{p}\sum_{t=1}^{n}\sum_{k=n+1}^{\infty}\sum_{m=1}^{n} c_k c_m Z_{i,t-k} Z_{i,t-m}.\end{aligned} \tag{5.2.9}$$

Equation (5.2.7b) allows us to apply Fubini's theorem to compute the variance of the second term in Eq. (5.2.9) as

$$\frac{4}{p^4}\sum_{i,i'=1}^{p}\sum_{t,t'=1}^{n}\sum_{k,k'=n+1}^{\infty}\sum_{m,m'=1}^{n} c_k c_m c_{k'} c_{m'} \mathbb{E}(Z_{i,t-k} Z_{i,t-m} Z_{i',t'-k'} Z_{i',t'-m'}),$$

which is, by the same reasoning as we did for II, bounded by

$$4\sigma_4 \frac{p}{p^4} n^2 \left(\sum_{k=n+1}^{\infty} c_k\right)^2 \left(\sum_{m=1}^{n} c_m\right)^2 \leqslant K n^{-1-2\delta},$$

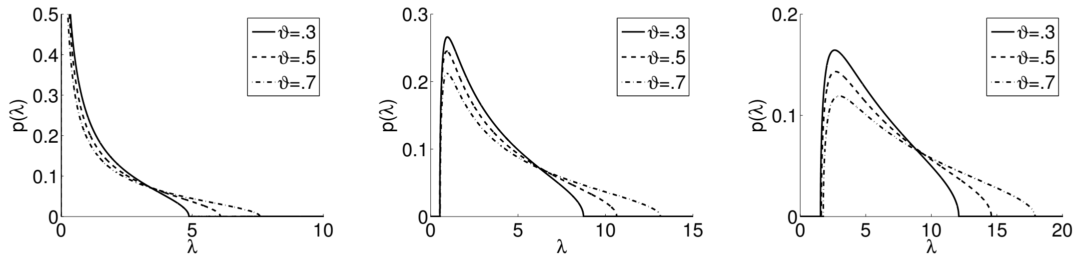
for some positive constant K. Clearly, this is summable in n. Having, by Eq. (5.2.7a), expected value zero, the second term of Eq. (5.2.9), and therefore also $\operatorname{tr}(\Delta_X)$, both converge to zero almost surely. Thus, we only have to look at the contribution of I_b in expression (5.2.8). From Proposition 5.8 we know that $F^{p^{-1}\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^T}$ converges almost surely weakly to some non-random distribution $\hat{F}$ with bounded support. Hence, denoting by $\lambda_1, \ldots, \lambda_p$ the eigenvalues of $p^{-1}\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^T$,

$$\mathrm{I}_\mathrm{b} = \frac{1}{p}\operatorname{tr}\left(\frac{1}{p}\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^T\right) = \frac{1}{p}\sum_{i=1}^{p}\lambda_i = \int \lambda \mathrm{d}F^{\frac{1}{p}\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^T} \to \int \lambda \mathrm{d}\hat{F} < \infty$$

almost surely. It follows that, in Eq. (5.2.6), factor I is bounded and factor II converges to zero, and so the proof of Theorems 5.1 and 5.2 is complete. □

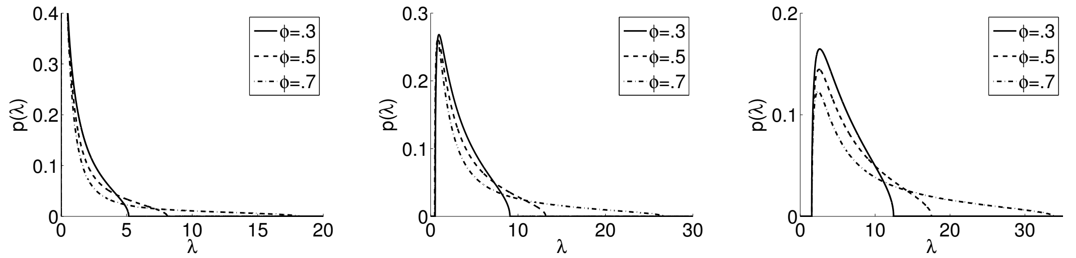
5.3. Illustrative examples

For several classes of widely employed linear processes, Theorem 5.1 can be used to obtain an explicit description of the limiting spectral distribution. In this section we consider the class of (fractionally integrated) autoregressive moving average (ARMA) processes. The distributions we obtain in the case of AR(1) and MA(1) processes can be interpreted as one-parameter deformations of the classical Marchenko–Pastur law.



(a) $y = 1$ (b) $y = 3$ (c) $y = 5$

Figure 5.1.: Limiting spectral densities $\lambda \mapsto p(\lambda)$ of $p^{-1}\boldsymbol{X}\boldsymbol{X}^T$ for the MA(1) process $X_t = Z_t + \vartheta Z_{t-1}$ for different values of ϑ and $y = n/p$



(a) $y = 1$ (b) $y = 3$ (c) $y = 5$

Figure 5.2.: Limiting spectral densities $\lambda \mapsto p(\lambda)$ of $p^{-1}\boldsymbol{X}\boldsymbol{X}^T$ for the AR(1) process $X_t = \varphi X_{t-1} + Z_t$ for different values of φ and $y = n/p$

5.3.1. Autoregressive moving average processes

Given polynomials $a : z \mapsto 1 + a_1 z + \ldots a_p z^p$ and $b : z \mapsto 1 + b_1 z + \ldots + b_q z^q$, an ARMA(p,q) process X with autoregressive polynomial a and moving average polynomial b is defined as the stationary solution to the stochastic difference equation

$$X_t + a_1 X_{t-1} + \ldots + a_p X_{t-p} = Z_t + b_1 Z_{t-1} + \ldots + b_q Z_{t-q}, \quad t \in \mathbb{Z}.$$

If the zeros of a lie outside the closed unit disk, it is well known that X has an infinite-order moving average representation $X_t = \sum_{j=0}^{\infty} c_j Z_{t-j}$, where $\{c_j\}$ are the coefficients in the power series expansion of $b(z)/a(z)$ around zero. It is also known (Brockwell and Davis, 1991, §13.2) that there exist positive constants $\rho < 1$ and K such that $|c_j| \leqslant K\rho^j$, so that assumption ii) of Theorem 5.1 is satisfied. While the autocovariance function of an ARMA process does not in general have a simple closed form, its Fourier transform, i. e. the spectral density of the ARMA process X, is given by

$$f(\omega) = \left| \frac{b\left(\mathrm{e}^{\mathrm{i}\omega}\right)}{a\left(\mathrm{e}^{\mathrm{i}\omega}\right)} \right|^2, \quad \omega \in [0, 2\pi]. \tag{5.3.1}$$

Since f is rational, assumptions iii) and iv) of Theorem 5.1 are satisfied as well. In order to compute the LSD of Γ, it is necessary, by Lemma 5.10, to find the roots of a trigonometric polynomial of possibly high degree, which can be done numerically.

We now consider the special case of the ARMA(1,1) process $X_t = \varphi X_{t-1} + Z_t + \vartheta Z_{t-1}$, $|\varphi| < 1$, for which one can obtain explicit results. By Eq. (5.3.1), the spectral density of X is given by

$$f(\omega) = \frac{1 + \vartheta^2 + 2\vartheta\cos\omega}{1 + \varphi^2 - 2\varphi\cos\omega}, \quad \omega \in [0, 2\pi].$$

Equation (5.2.4) implies that the LSD of the autocovariance matrix Γ has a density g given by

$$\begin{aligned} g(\lambda) =& \frac{1}{2\pi} \sum_{\omega\in[0,2\pi]: f(\omega)=\lambda} \frac{1}{|f'(\omega)|} \\ =& \frac{1}{\pi(\vartheta + \varphi\lambda)\sqrt{[(1+\vartheta)^2 - \lambda(1-\varphi)^2]\,[\lambda(1+\varphi)^2 - (1-\vartheta)^2]}} I_{(\lambda_-,\lambda_+)}(\lambda), \end{aligned}$$

where

$$\lambda_- = \min(\lambda^-, \lambda^+), \quad \lambda_+ = \max(\lambda^-, \lambda^+), \quad \lambda^\pm = \frac{(1 \pm \vartheta)^2}{(1 \mp \varphi)^2}.$$

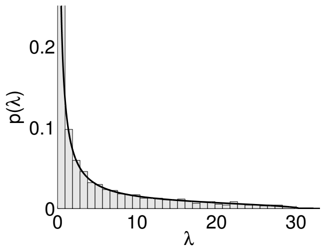
By Theorem 5.1, the Stieltjes transform $z \mapsto m_z$ of the limiting spectral distribution of $p^{-1}\boldsymbol{X}\boldsymbol{X}^T$ is the unique mapping $m : \mathbb{C}^+ \to \mathbb{C}^+$ satisfying the equation

$$\begin{aligned} \frac{1}{m_z} =& -z + y\int_{\lambda_-}^{\lambda_+} \frac{\lambda g(\lambda)}{1 + \lambda m_z}\mathrm{d}\lambda \\ =& -z + \frac{\vartheta y}{\vartheta m_z - \varphi} \\ & - \frac{(\vartheta + \varphi)(1 + \vartheta\varphi)y}{(\vartheta m_z - \varphi)\sqrt{[(1-\varphi)^2 + m_z(1+\vartheta)^2]\,[(1+\varphi)^2 + m_z(1-\vartheta)^2]}}. \end{aligned} \tag{5.3.2}$$

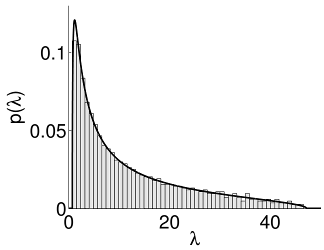
This is a quartic equation in $m_z \equiv m(z)$, which can be solved explicitly. An application of the

Stieltjes inversion formula (5.1.3) then yields the limiting spectral distribution of $p^{-1}\boldsymbol{X}\boldsymbol{X}^T$.

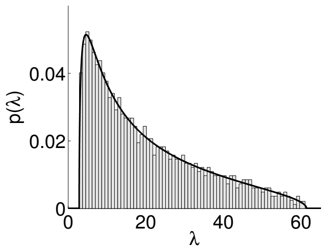
If one sets $\varphi = 0$, one obtains an MA(1) process; plots of the densities obtained in this case for different values of y and ϑ are displayed in Fig. 5.1. Similarly, the case $\vartheta = 0$ corresponds to an AR(1) process; see Fig. 5.2 for a graphical representation of the densities one obtains for different values of y and φ in this case. For the special case $\varphi = 1/2$, $\vartheta = 1$, Fig. 5.3 compares the histogram of the eigenvalues of $p^{-1}\boldsymbol{X}\boldsymbol{X}^T$ with the limiting spectral distribution obtained from Theorem 5.1 for different values of y.



(a) $y = 1$



(b) $y = 3$



(c) $y = 5$

Figure 5.3.: Histograms of the eigenvalues and limiting spectral densities $\lambda \mapsto p(\lambda)$ of $p^{-1}\boldsymbol{X}\boldsymbol{X}^T$ for the ARMA(1,1) process $X_t = \frac{1}{2}X_{t-1} + Z_t + Z_{t-1}$ for different values of $y = n/p$, $p = 1000$

Equation (5.3.2) for the Stieltjes transform of the limiting spectral distribution of the sample covariance matrix of an ARMA(1,1) process should be compared to Bai and Zhou (2008, Eq. (2.10)), where the analogous result is obtained for an autoregressive process of order one. They use the notation $c = \lim p/n$ and consider the spectral distribution of $n^{-1}\boldsymbol{X}\boldsymbol{X}^T$ instead of $p^{-1}\boldsymbol{X}\boldsymbol{X}^T$ as we do. If one observes that this difference in the normalization amounts to a linear transformation of the corresponding Stieltjes transform, one obtains their result as a special case of Eq. (5.3.2).

5.3.2. Fractionally integrated ARMA processes

In many practical situations, data exhibit long-range dependence, which can be modelled by long-memory processes. Denote by B the backshift operator and define, for $d > -1$, the (fractional) difference operator by

$$\nabla^d = (1-\mathrm{B})^d = \sum_{j=0}^{\infty}\prod_{k=1}^{j}\frac{k-1-d}{k}\,\mathrm{B}^j, \quad \mathrm{B}^j X_t = X_{t-j}.$$

A process $(X_t)_t$ is called a fractionally integrated ARMA(p,d,q) processes with $p,q \in \mathbb{N}$ and $d \in (-1/2, 1/2)$ if $(\nabla^d X_t)_t$ is an ARMA(p,q) process. These processes have a polynomially decaying autocorrelation function and therefore exhibit long-range-dependence, cf. Brockwell and Davis (1991, Theorem 13.2.2) and Granger and Joyeux (1980); Hosking (1981).

We assume that $d < 0$, and that the zeros of the autoregressive polynomial a of $(\nabla^d X_t)_t$ lie outside the closed unit disk. Then it follows that X has an infinite order moving average representation $X_t = \sum_{j=0}^{\infty} c_j Z_{t-j}$, where the $(c_j)_j$ have, in contrast to our previous examples, not an exponential decay, but satisfy $K_1(j+1)^{d-1} \leqslant c_j \leqslant K_2(j+1)^{d-1}$ for some $K_1, K_2 > 0$. Therefore, if $d < 0$, one can apply Theorem 5.1 to obtain the limiting spectral distribution of the sample covariance matrix of a FICARMA(p,q,d) process using that the spectral density of $(X_t)_t$ is given by

$$f(\omega) = \left| \frac{b\left(e^{i\omega}\right)}{a\left(e^{i\omega}\right)} \right|^2 \left|1 - e^{-i\omega}\right|^{-2d}, \quad \omega \in [0, 2\pi].$$

6. Limiting Spectral Distribution of a New Random Matrix Model with Dependence across Rows and Columns

6.1. Introduction

Random matrix theory studies the properties of large random matrices $A = (A_{i,j})_{ij} \in \mathbb{K}^{p \times n}$, where $\mathbb{K}$ is usually either $\mathbb{R}$ or $\mathbb{C}$, but can also denote a different field. In this chapter, the entries A_{ij} are supposed to be real random variables unless specified differently. Commonly, the focus is on asymptotic properties of such matrices as their dimensions tend to infinity. One particularly interesting object of study is the asymptotic distribution of the singular values. Since the squared singular values of A are the eigenvalues of AA^T, this is often done by investigating the eigenvalues of AA^T, which is called a sample covariance matrix. The spectral characteristics of a $p \times p$ matrix S are conveniently studied via the empirical spectral distribution of S, which is defined as

$$F^S = \frac{1}{p} \sum_{i=1}^{p} \delta_{\lambda_i}, \tag{6.1.1}$$

where $\{\lambda_1, \ldots, \lambda_p\}$ are the eigenvalues of S, and δ_x denotes the Dirac measure located at x. For some set $B \subset \mathbb{R}$, the figure $F^S(B)$ is the number of eigenvalues of S that lie in B. The measure F^S is considered a random element of the space of probability distributions equipped with the weak topology, and we are interested in its limit as both n and p tend to infinity such that the ratio p/n converges to a finite positive limit y.

The first result of this kind can be found in the remarkable paper Marchenko and Pastur (1967). They showed that $F^{p^{-1}AA^T}$ converges to a non-random limiting spectral distribution $\hat{F}^{p^{-1}AA^T}$ if all A_{ij} are independent, identically distributed, centred random variables with finite fourth moment. Interestingly, the Lebesgue density of $\hat{F}^{p^{-1}AA^T}$ is given by an explicit formula which only involves the ratio y and the common variance of A_{ij} and is therefore universal with respect to the distribution of the entries of A. Subsequently (Wachter, 1978; Yin, 1986), the same result has been obtained under the weaker moment condition that the entries A_{ij} have finite variance. The requirement that the entries of A be identically distributed has later been relaxed to a Lindeberg-type condition, cf. Eq. (6.2.3). For more details and a comprehensive treatment of random matrix theory we refer the reader to the text books Anderson et al. (2010); Bai and Silverstein (2010); Mehta (2004).

Recent research has focused on the question to what extent the assumption of independence of the entries of A can be relaxed without compromising the validity of the Marchenko–Pastur law. In Aubrun (2006) it was shown that for random matrices A whose rows are independent $\mathbb{R}^n$-valued random variables uniformly distributed on the unit ball of $l_q(\mathbb{R}^n)$, $q > 1$, the empirical spectral distribution $F^{p^{-1}AA^T}$ still converges to the law obtained in the i. i. d. case. The Marchenko–Pastur law is, however, not stable with respect to more substantial deviations from the independence assumptions. A very useful tool to characterize the limiting spectral distribution in random matrix models with dependent entries is the Stieltjes transform, which, for some measure μ on $\mathbb{R}$, is defined as the map

$$s_\mu : \mathbb{C}^+ \to \mathbb{C}^+, \quad s_\mu(z) = \int_{\mathbb{R}} \frac{\mu(\mathrm{d}t)}{t - z}.$$

A particular, very successful random matrix model exhibiting dependencies within the rows was investigated already in Marchenko and Pastur (1967) and later in greater generality in Pan (2010); Silverstein and Bai (1995): they modelled dependent data via a linear transformation of independent random variables, which led to the study of the eigenvalues of random matrices of the form AHA^T, where the entries of A are independent, and H is a positive semidefinite population covariance matrix whose spectral distribution converges to a non-random limit $\hat{F}^H$. They found that the Stieltjes transform of the limiting spectral distribution of $p^{-1}AHA^T$ can be uniquely characterized as the solution of a certain integral equation involving only $\hat{F}^H$ and the ratio $y = \lim p/n$. Another approach, which has been suggested in Bai and Zhou (2008) and further pursued in Chapter 5, is to model the rows of A independently as stationary linear processes with independent innovations. This structure is particularly interesting because the class of linear processes includes many practically relevant time series models, such as (fractionally integrated) ARMA processes, as special cases. The main result of Chapter 5 shows that for this model the limiting spectral distribution depends only on y and the second-order properties of the underlying linear process in form of its autocovariance function.

All results for independent rows with dependent row entries also hold with minor modifications for the case where A has independent columns with dependent column entries. This is due to the fact that AA^T and A^TA have the same non-zero eigenvalues.

In contrast to what has been said so far, there are only few results dealing with random matrix models where the entries are dependent across rows and columns. The case where A is the result of a two-dimensional linear filter applied to an array of independent complex Gaussian random variables is considered in Hachem et al. (2005). They use the fact that A can be transformed to a random matrix with uncorrelated, non-identically distributed entries. Because of the assumption of Gaussianity, the entries are in fact independent, and so an earlier result by the same authors (Hachem, Loubaton and Najim, 2006) can be used to obtain the asymptotic distribution of the eigenvalues of $p^{-1}AA^*$. In the context of operator-valued free probability theory, Rashidi Far, Oraby, Bryc and Speicher (2008) succeeded in characterizing the limiting spectral distribution of block Wishart matrices through a quadratic matrix equation for the corresponding operator-valued Stieltjes transform.

A parallel line of research focused on the spectral statistics of large symmetric or Hermitian square matrices with dependent entries, thus extending Wigner's seminal result for the i. i. d. case (Wigner, 1958). Models studied in this context include random Toeplitz, Hankel and circulant matrices (Bose, Subhra Hazra and Saha, 2009; Bryc, Dembo and Jiang, 2006; Meckes, 2007, and references therein), as well as approaches allowing for more general dependency structures (Anderson and Zeitouni, 2008; Hofmann-Credner and Stolz, 2008).

Chapter 5 dealt with sample covariance matrices of high-dimensional stochastic processes, the components of which are modelled by independent infinite-order moving average processes with identical second-order characteristics. In practice, it is often not possible to observe all components of such a high-dimensional process, and the sample covariance matrix can then not be computed. To solve this problem when only one component is observed, it seems reasonable to partition one long observation record of that observed component of length pn into p segments of length n, and to treat the different segments as if they were records of the unobserved components. We show that this approach is valid and leads to the correct asymptotic eigenvalue distribution of the sample covariance matrix if the components of the underlying process are modelled as independent moving averages.

We are thus led to investigate a model of random matrices $\boldsymbol{X}$ whose entries are dependent across both rows and columns, and which is not covered by the results mentioned above. The entries of the random matrix under consideration are defined in terms of a single linear stochastic process, see Section 6.2 for a precise definition. Without assuming Gaussianity we prove almost sure convergence of the empirical spectral distribution of $p^{-1}\boldsymbol{X}\boldsymbol{X}^T$ to a deterministic limiting measure and characterize the latter via an integral equation for its Stieltjes transform, which only depends on the asymptotic aspect ratio of the matrix and the second-order properties of the underlying linear process. Our result extends the class of random matrix models for which the limiting spectral distribution can be identified explicitly by a new, theoretically appealing model. It thus contributes to laying the ground for further research into more general random matrix models with dependent, non-identically distributed entries.

Outline of the chapter In Section 6.2 we give a precise definition of the random matrix model we investigate and state the main result about its limiting spectral distribution. The proof of the main theorem as well as some auxiliary results are presented in Section 6.3. Finally, in Section 6.4, we indicate how our result could be obtained in an alternative way from a similar random matrix model with independent rows.

Notation Throughout, we use $\mathbb{E}$ and $\mathbb{V}\mathrm{ar}$ to denote the expected value and variance, respectively, of a random variable. Where convenient, we also write $\mu_{1,X}$ and $\mu_{2,X}$ for the first and second moment, respectively, of a random variable X. The symbol $\mathbf{1}_m$, m a natural number, stands for the $m \times m$ identity matrix. For the trace of a quadratic matrix S we write $\operatorname{tr} S$. For sequences of matrices $(S_n)_n$ we will suppress the dependence on n where this does not cause ambiguity; the sequence of associated spectral distributions, see Eq. (6.1.1), is denoted by F^S, and for their weak limit, provided that it exists, we write $\hat{F}^S$. It will also be convenient to use asymptotic Landau notation: for two sequences of real numbers $(a_n)_n$, $(b_n)_n$, we write $a_n = O(b_n)$ to indicate that there exists a constant C, which is independent of n, such that $a_n \leqslant C b_n$ for all n. We denote by $\mathbb{Z}$ the set of integers, and by $\mathbb{N}$, $\mathbb{R}$, and $\mathbb{C}$ the sets of natural, real, and complex numbers, respectively. $\operatorname{Im} z$ stands for the imaginary part of a complex number z, and $\mathbb{C}^+$ is defined as $\{z \in \mathbb{C} : \operatorname{Im} z > 0\}$. The indicator of an expression $\mathcal{E}$ is denoted by $I_{\{\mathcal{E}\}}$ and defined to be one if $\mathcal{E}$ is true, and zero otherwise.

6.2. A new random matrix model

For a sequence $(Z_t)_{t\in\mathbb{Z}}$ of independent real random variables and real-valued coefficients $(c_j)_{j\in\mathbb{N}\cup\{0\}}$, the linear process $X = (X_t)_{t\in\mathbb{Z}}$ is defined by $X_t = \sum_{j=0}^{\infty} c_j Z_{t-j}$. Based upon this process the $p \times n$ matrix $\boldsymbol{X}$ is defined as

$$\boldsymbol{X} = (\boldsymbol{X}_{i,t})_{it} = (X_{(i-1)n+t})_{it} = \begin{pmatrix} X_1 & \dots & X_n \\ X_{n+1} & \dots & X_{2n} \\ \vdots & & \vdots \\ X_{(p-1)n+1} & \dots & X_{pn} \end{pmatrix} \in \mathbb{R}^{p\times n}. \tag{6.2.1}$$

The interesting feature about this matrix $\boldsymbol{X}$ is that its entries are dependent across both rows and columns. Moreover, in contrast to the models considered in Bai and Zhou (2008); Hachem et al. (2006) and Chapter 5, not all entries far away from each other are asymptotically independent. For example, the correlation between the entries $\boldsymbol{X}_{i,n}$ and $\boldsymbol{X}_{i+1,1}$, $i = 1, \dots, p-1$, does not depend on n. This makes the matrix $\boldsymbol{X}$ a fascinating object to study. We will investigate the asymptotic distribution of the eigenvalues of $p^{-1}\boldsymbol{X}\boldsymbol{X}^T$ as both p and n tend to infinity such that their ratio p/n converges to a finite, positive limit y. We assume that the sequence $(Z_t)_t$ satisfies

$$\mathbb{E}Z_t = 0, \quad \mathbb{E}Z_t^2 = 1, \quad \text{and} \quad \sigma_4 := \sup_t \mathbb{E}Z_t^4 < \infty, \tag{6.2.2}$$

and that the following Lindeberg-type condition is satisfied: for each $\epsilon > 0$,

$$\frac{1}{pn}\sum_{t=1}^{pn}\mathbb{E}\left(Z_t^2 I_{\{Z_t^2\geqslant \epsilon n\}}\right) \to 0, \quad \text{as} \quad n\to\infty. \tag{6.2.3}$$

Clearly, condition (6.2.3) is satisfied if all $\{Z_t\}$ are identically distributed, but that is not necessary. As it turns out, the limiting spectral distribution of $p^{-1}\boldsymbol{X}\boldsymbol{X}^T$ depends only on y and the second-order structure of the underlying linear process X, which we now recall: the autocovariance function $\gamma : \mathbb{Z} \to \mathbb{R}$ of X is defined by

$$\gamma(h) = \mathbb{E}X_0X_h = \mathbb{E}X_0X_{-h} = \sum_{j=0}^{\infty} c_j c_{j+|h|}, \quad h \in \mathbb{Z};$$

the spectral density $f : [0, 2\pi] \to \mathbb{R}$ of X is the Fourier transform of this function, namely

$$f(\omega) = \sum_{h\in\mathbb{Z}} \gamma(h)\mathrm{e}^{-\mathrm{i}h\omega}, \quad \omega \in [0, 2\pi].$$

The following is the main result of this chapter.

Theorem 6.1 (Limiting Spectral Distribution) *Let $X_t = \sum_{j=0}^{\infty} c_j Z_{t-j}$, $t \in \mathbb{Z}$, be a linear stochastic processes with continuously differentiable spectral density f, and let the matrix $\boldsymbol{X} \in \mathbb{R}^{p\times n}$ be given by Eq. (6.2.1). Assume that*

i) the sequence $(Z_t)_t$ satisfies conditions (6.2.2) and (6.2.3),

ii) there exist a positive constants C, δ such that $|c_j| \leqslant C(j + 1)^{-1-\delta}$ for all $j \in \mathbb{N} \cup \{0\}$,

iii) for almost all $\lambda \in \mathbb{R}$, $f(\omega) = \lambda$ for at most finitely many $\omega \in [0, 2\pi]$, and

iv) $f'(\omega) \neq 0$ for almost every ω.

Then, as both n and p tend to infinity such that the ratio p/n converges to a finite positive limit y, the empirical spectral distribution of $p^{-1}\boldsymbol{X}\boldsymbol{X}^T$ converges almost surely to a non-random probability distribution function $\hat{F}$ with bounded support. Moreover, there exist positive numbers λ_-, λ_+ such that the Stieltjes transform $z \mapsto s_{\hat{F}}(z)$ of $\hat{F}$ is the unique mapping $\mathbb{C}^+ \to \mathbb{C}^+$ satisfying

$$\frac{1}{s_{\hat{F}}(z)} = -z + \frac{y}{2\pi}\int_{\lambda_-}^{\lambda_+} \frac{\lambda}{1+\lambda s_{\hat{F}}(z)} \sum_{\omega\in[0,2\pi]:f(\omega)=\lambda} \frac{1}{|f'(\omega)|}\mathrm{d}\lambda. \tag{6.2.4}$$

Remark 6.2 The assumption that the coefficients $(c_j)_j$ decay at least polynomially is not very restrictive. In particular, it allows for the process X to be an ARMA or fractionally integrated ARMA process, which are known to exhibit long-range dependence (Granger and Joyeux, 1980; Hosking, 1981). In the latter case, the entries of the matrix $\boldsymbol{X}$ are long-range dependent as well. The proof of Theorem 6.1 also shows that the weaker assumption that the sequence $(|c_j|)_j$ be bounded by a monotone, summable sequence is sufficient for the empirical spectral distribution $F^{p^{-1}\boldsymbol{X}\boldsymbol{X}^T}$ to converge in probability to a non-random limit with Stieltjes transform satisfying Eq. (6.2.4).

Remark 6.3 It is possible to generalize the proof of Theorem 6.1 so that the result also holds for non-causal processes, where $X_t = \sum_{j=-\infty}^{\infty} c_j Z_{t-j}$. The required changes are merely notational, the only difference in the result is that the autocovariance function is then given by $\sum_{j=-\infty}^{\infty} c_j c_{j+|h|}$.

Remark 6.4 The result in Theorem 6.1 is the same as the one obtained in Chapter 5 for a random matrix model in which the rows of $\boldsymbol{X}$ are modelled as independent linear processes.

Remark 6.5 We conjecture that the same result also holds if the matrix $\boldsymbol{X}$ is defined as

$$\boldsymbol{X} = \begin{pmatrix} X_1 & \dots & X_n \\ X_{2n} & \dots & X_{n+1} \\ \vdots & & \vdots \\ X_{(p-1)n+1} & \dots & X_{pn} \end{pmatrix}.$$

The current proof cannot, however, be easily adapted to show this.

Remark 6.6 Condition iv) of Theorem 6.1 excludes a constant spectral density f, which corresponds to the classical situation of independent entries in the matrix $\boldsymbol{X}$. The Marchenko–Pastur law is therefore not a special case of Theorem 6.1. It is, however, easy to formulate a variant of our main theorem for linear processes with a piecewise constant spectral density analogously to Theorem 5.2, from which one can recover the classical result. In fact, using Szegő's limit theory about the LSD of Toeplitz matrices (see Szegő (1920, Theorem XVIII) for the original result or, e. g., Böttcher and Silbermann (1999, Sections 5.4 and 5.5) for a modern treatment), it is possible to show that conditions i) and ii) are sufficient for the Stieltjes transform $s_{\hat{F}}$ to be characterized by the equation

$$\frac{1}{s_{\hat{F}}(z)} = -z + y \int_0^{2\pi} \frac{f(\omega)}{1 + f(\omega) s_{\hat{F}}(z)} \mathrm{d}\omega. \tag{6.2.5}$$

Theorem 6.1 provides a characterization not of the limiting spectral distribution $\hat{F}$ itself, but rather of its Stieltjes transform $m_{\hat{F}}$. The LSD $\hat{F}$ can be obtained from $m_{\hat{F}}$ via the well-known inversion formula (Theorem B.8 Bai and Silverstein, 2010),

$$\hat{F}([a,b]) = \lim_{\epsilon \to 0^+} \int_a^b \operatorname{Im} s_{\hat{F}}(x + \epsilon \mathrm{i}) \mathrm{d}x,$$

which holds for all continuity points $0 < a < b$ of $\hat{F}$. In general, the analytic determination of this distribution is not feasible, but numerical approximations can be used (see, e. g., Abate, Choudhury and Whitt, 2000).

6.3. Proof of Theorem 6.1

The general strategy in the proof of Theorem 6.1 is to show that the limiting spectral distribution of $p^{-1}\boldsymbol{X}\boldsymbol{X}^T$ is stable under certain modifications of the matrix $\boldsymbol{X}$. Finally, this will imply that $\hat{F}^{p^{-1}\boldsymbol{X}\boldsymbol{X}^T}$ exists and is equal to the limiting spectral distribution of the random matrix model studied in Chapter 5. To this end we will repeatedly use the following lemma which presents sufficient conditions for the limiting spectral distributions of two sequences of matrices to be equal.

Lemma 6.7 *Let $A_{1,n}$, $A_{2,n}$ be sequences of $p \times n$ matrices, where $p = p_n$ depends on n such that $p_n \to \infty$, as $n \to \infty$. Assume that the spectral distribution $F^{p^{-1}A_{1,n}A_{1,n}^T}$ converges almost surely to a deterministic limit $\hat{F}^{p^{-1}A_{1,n}A_{1,n}^T}$, as n tends to infinity. If there exists a positive number ϵ such that*

i) $p^{-4}\mathbb{E}\left[\operatorname{tr}\left(A_{1,n} - A_{2,n}\right)\left(A_{1,n} - A_{2,n}\right)^T\right]^2 = O(n^{-1-\epsilon})$,

ii) $p^{-2}\mathbb{E}\operatorname{tr} A_{i,n}A_{i,n}^T = O(1)$, $i = 1, 2$, *and*

iii) $p^{-4}\operatorname{\mathbb{V}ar}\operatorname{tr} A_{i,n}A_{i,n}^T = O(n^{-1-\epsilon})$, $i = 1, 2$,

then the spectral distribution of $p^{-1}A_{2,n}A_{2,n}^T$ is convergent almost surely with the same limiting spectral distribution, i. e. $F^{p^{-1}A_{2,n}A_{2,n}^T}$ converges weakly to $\hat{F}^{p^{-1}A_{1,n}A_{1,n}^T}$.

Proof The first condition implies via Chebyshev's inequality and the Borel–Cantelli lemma that the expression $p^{-2}\operatorname{tr}\left(A_{1,n} - A_{2,n}\right)\left(A_{1,n} - A_{2,n}\right)^T$ converges to zero almost surely. Similarly, it follows from ii) and iii) that $p^{-2}\operatorname{tr} A_{i,n}A_{i,n}^T$ is bounded almost surely for $i = 1, 2$. These two facts together then entail almost sure convergence to zero of the product $p^{-4}\operatorname{tr}\left(A_{1,n}A_{1,n}^T + A_{2,n}A_{2,n}^T\right)\operatorname{tr}\left(A_{1,n} - A_{2,n}\right)\left(A_{1,n} - A_{2,n}\right)^T$. By Bai and Silverstein (2010, Corollary A.42), this is a sufficient condition for the assertion of the lemma. □

The following notation will be used throughout the rest of this section: with the constants C and δ from assumption ii) of Theorem 6.1 we define $\bar{c}_j := C(j+1)^{-1-\delta}$, such that $|c_j| \leqslant \bar{c}_j$ for all j. Without further reference we will repeatedly use the fact that $j \mapsto \bar{c}_j$ is monotone, that $\sum_{j=1}^{\infty} \bar{c}_j^{\alpha}$ is finite for every $\alpha \geqslant 1$, and that $\sum_{j=n}^{\infty} \bar{c}_j^{\alpha} = O(n^{1-\alpha(1+\delta)})$.

Since it is difficult to deal with infinite-order moving-averages processes directly, it is convenient to truncate the entries of the matrix $\boldsymbol{X}$ by defining $\widetilde{X}_t = \sum_{j=0}^{n} c_j Z_{t-j}$ and $\widetilde{\boldsymbol{X}} = (\widetilde{X}_{(i-1)n+t})_{it}$. The next proposition shows that this first modification to $\boldsymbol{X}$ does not alter the limiting spectral distribution of $p^{-1}\boldsymbol{X}\boldsymbol{X}^T$.

Proposition 6.8 *If the empirical spectral distribution of $p^{-1}\widetilde{X}\widetilde{X}^T$ converges to a limit, then the empirical spectral distribution of $p^{-1}XX^T$ converges to the same limit.*

Proof The proof of the proposition proceeds in two steps in which we verify conditions i) to iii) of the trace criterion, Lemma 6.7, respectively.

Step 1 The definitions of X and $\widetilde{X}$ imply that

$$\Delta_{X,\widetilde{X}} := \frac{1}{p^2}\operatorname{tr}\left(X-\widetilde{X}\right)\left(X-\widetilde{X}\right)^T = \frac{1}{p^2}\sum_{i=1}^{p}\sum_{t=1}^{n}\left[X_{it}-\widetilde{X}_{it}\right]^2$$
$$= \frac{1}{p^2}\sum_{i=1}^{p}\sum_{t=1}^{n}\sum_{k,k'=n+1}^{\infty} Z_{(i-1)n+t-k}Z_{(i-1)n+t-k'}c_kc_{k'}.$$

We shall show that the second moment of $\Delta_{X,\widetilde{X}}$ is of order at most $n^{-2-2\delta}$. Since

$$\sum_{\substack{k,k'\\m,m'}=n+1}^{\infty}\mathbb{E}\left|Z_{(i-1)n+t-k}Z_{(i-1)n+t-k'}Z_{(i'-1)n+t'-m}Z_{(i'-1)n+t'-m'}\right||c_k||c_{k'}||c_m||c_{m'}|$$
$$\leqslant \sigma_4\left[\sum_{k=0}^{\infty}|c_k|\right]^4 < \infty, \tag{6.3.1}$$

we can apply Fubini's theorem to interchange expectation and summation to obtain

$$\mu_{2,\Delta_{X,\widetilde{X}}} := \mathbb{E}\Delta^2_{X,\widetilde{X}} \tag{6.3.2}$$
$$= \frac{1}{p^4}\sum_{\substack{i,i'\\t,t'}=1}^{p,n}\sum_{\substack{k,k'\\m,m'}=n+1}^{\infty}\mathbb{E}\left[Z_{(i-1)n+t-k}Z_{(i-1)n+t-k'}Z_{(i'-1)n+t'-m}Z_{(i'-1)n+t'-m'}\right]c_kc_{k'}c_mc_{m'}.$$

Since the $\{Z_t\}$ are assumed to be independent, the expectation in that sum is non-zero only if all four Z are the same or else one can match the indices in two pairs. In the latter case we distinguish three cases according to whether the first Z is paired with the second, third, or last factor. This leads to the additive decomposition

$$\mu_{2,\Delta_{X,\widetilde{X}}} = \mu^{\sqcap\!\sqcap\!\sqcap}_{2,\Delta_{X,\widetilde{X}}} + \mu^{\sqcap\,\sqcap}_{2,\Delta_{X,\widetilde{X}}} + \mu^{\sqcap\!\!\sqcap}_{2,\Delta_{X,\widetilde{X}}} + \mu^{\overline{\sqcap}}_{2,\Delta_{X,\widetilde{X}}}, \tag{6.3.3}$$

where the ideograms indicate which of the four factors are equal. For the contribution from all four Z being equal it holds that

$$k = k', \quad m = m', \quad \text{and} \quad (i-1)n+t-k = (i'-1)n+t'-m,$$

so that

$$\mu_{2,\Delta_{X,\tilde{X}}}^{\sqcap\!\!\sqcap} = \frac{\sigma_4}{p^4} \sum_{i,i'}^{p} \sum_{t,t'=1}^{n} \sum_{m=\max\{n+1,n+1-(i-i')n-(t-t')\}}^{\infty} c_{(i-i')n+(t-t')+m}^2 c_m^2.$$

If we introduce the new summation variables $\delta_i := i - i'$ and $\delta_t := t - t'$, we obtain

$$\mu_{2,\Delta_{X,\tilde{X}}}^{\sqcap\!\!\sqcap} = \frac{\sigma_4}{p^4} \sum_{\delta_i=1-p}^{p-1} \underbrace{(p - |\delta_i|)}_{\leqslant p} \sum_{\delta_t=1-n}^{n-1} \underbrace{(n - |\delta_t|)}_{\leqslant n} \sum_{m=\max\{n+1,n+1-\delta_i n-\delta_t\}}^{\infty} c_{m+\delta_i n+\delta_t}^2 c_m^2.$$

If δ_i is positive, then $\delta_i n + \delta_t$ is positive as well; the fact that $|c_j|$ is bounded by $\bar{c}_j$ and the monotonicity of $j \mapsto \bar{c}_j$ imply that $c_{m+\delta_i n+\delta_t}^2 \leqslant \bar{c}_{(\delta_i-1)n} \bar{c}_{\delta_t+n}$ so that the contribution from $\delta_i \geqslant 1$ can be estimated as

$$\mu_{2,\Delta_{X,\tilde{X}}}^{\sqcap\!\!\sqcap,+} \leqslant \underbrace{\frac{\sigma_4 n}{p^3}}_{=O(n^{-2})} \underbrace{\sum_{\delta_i=1}^{p-1} \bar{c}_{(\delta_i-1)n}}_{=O(n^{-1-\delta})} \underbrace{\sum_{\delta_t=1}^{2n-1} \bar{c}_{\delta_t}}_{=O(1)} \underbrace{\sum_{m=n+1}^{\infty} \bar{c}_m^2}_{=O(n^{-1-2\delta})} = O(n^{-4-3\delta}). \tag{6.3.4}$$

An analogous argument shows that the contribution from $\delta_i \leqslant -1$, denoted by $\mu_{2,\Delta_{X,\tilde{X}}}^{\sqcap\!\!\sqcap,-}$, is of the same order of magnitude. The contribution to $\mu_{2,\Delta_{X,\tilde{X}}}^{\sqcap\!\!\sqcap}$ from $\delta_i = 0$ is given by

$$\begin{aligned} \mu_{2,\Delta_{X,\tilde{X}}}^{\sqcap\!\!\sqcap,\varnothing} &= \frac{\sigma_4 n}{p^3} \sum_{\delta_t=1-n}^{n-1} \sum_{m=\max\{n+1,n+1-\delta_t\}}^{\infty} c_m^2 c_{m+\delta_t}^2 \\ &\leqslant \underbrace{\frac{\sigma_4 n}{p^3}}_{=O(n^{-2})} \left[2 \underbrace{\sum_{\delta_t=1}^{n-1} \bar{c}_\delta^2}_{=O(1)} \underbrace{\sum_{m=n+1}^{\infty} \bar{c}_m^2}_{=O(n^{-1-2\delta})} + \underbrace{\sum_{m=n+1}^{\infty} \bar{c}_m^4}_{=O(n^{-3-4\delta})} \right] = O(n^{-3-2\delta}). \end{aligned} \tag{6.3.5}$$

By combining the displays (6.3.4) and (6.3.5) it follows that $\mu_{2,\Delta_{X,\tilde{X}}}^{\sqcap\!\!\sqcap} = O(n^{-3-2\delta})$. The second term in Eq. (6.3.3) corresponds to $k = k'$, $m = m'$ and $(i-1)n + t - k \neq (i'-1)n + t' - m$. The restriction that not all four factors be equal is taken into account by subtracting $\mu_{2,\Delta_{X,\tilde{X}}}^{\sqcap\!\!\sqcap}$; consequently,

$$\mu_{2,\Delta_{X,\tilde{X}}}^{\sqcap\,\sqcap} = \underbrace{\frac{1}{p^4} \sum_{i,i'=1}^{p} \sum_{t,t'=1}^{n}}_{=O(1)} \underbrace{\sum_{k,m=n+1}^{\infty} c_k^2 c_m^2}_{=O(n^{-2-4\delta})} - \mu_{2,\Delta_{X,\tilde{X}}}^{\sqcap\!\!\sqcap} = O(n^{-2-4\delta}).$$

It remains to analyse $\mu_{2,\Delta_{X,\tilde{X}}}^{\sqcap\!\!\!\sqcap}$ which, by symmetry, is equal to $\mu_{2,\Delta_{X,\tilde{X}}}^{\overline{\sqcap}}$. If the first factor is paired with the third, the condition for non-vanishment becomes

$$(i-1)n+t-k=(i'-1)n+t'-m,\quad (i-1)n+t-k'=(i'-1)n+t'-m',\quad k\neq k', m\neq m',$$

or, equivalently,

$$k=m+(i-i')n+t-t',\quad k'=m'+(i-i')n+t-t',\quad m\neq m'.$$

Again introducing the new summation variables $\delta_i := i-i'$ and $\delta_t := t-t'$, we obtain that

$$\begin{aligned}\mu_{2,\Delta_{X,\tilde{X}}}^{\sqcap\!\!\!\sqcap} = \frac{1}{p^4}\sum_{\delta_i=1-p}^{p-1}\underbrace{(p-|\delta_i|)}_{\leqslant p}\sum_{\delta_t=1-n}^{n-1}\underbrace{(n-|\delta_t|)}_{\leqslant n}\sum_{m,m'=\max\{n+1,n+1-\delta_i n-\delta_t\}}^{\infty} c_m c_{m'} c_{m+\delta_i n+\delta_t} c_{m'+\delta_i n+\delta_t}\\ -\mu_{2,\Delta_{X,\tilde{X}}}^{\sqcap\sqcap}.\end{aligned}$$

As in the analysis of $\mu_{2,\Delta_{X,\tilde{X}}}^{\sqcap\sqcap}$ we obtain the contribution from $\delta_i\neq 0$ as

$$\left|\mu_{2,\Delta_{X,\tilde{X}}}^{\sqcap\!\!\!\sqcap,+}\right| = \left|\mu_{2,\Delta_{X,\tilde{X}}}^{\sqcap\!\!\!\sqcap,-}\right| \leqslant \underbrace{\frac{n}{p^3}}_{=O(n^{-2})}\underbrace{\sum_{\delta_i=1}^{p-1}\bar{c}_{(\delta_i-1)n}}_{=O(n^{-1-\delta})}\underbrace{\sum_{\delta_t=1}^{2n-1}\bar{c}_{\delta_t}}_{=O(1)}\underbrace{\sum_{m,m'=n+1}^{\infty}\bar{c}_m\bar{c}_{m'}}_{=O(n^{-2\delta})}+\mu_{2,\Delta_{X,\tilde{X}}}^{\sqcap\sqcap} = O(n^{-3-2\delta}). \tag{6.3.6}$$

Finally, for the contribution from $\delta_i = 0$, we find that

$$\begin{aligned}\left|\mu_{2,\Delta_{X,\tilde{X}}}^{\sqcap\!\!\!\sqcap,\varnothing}\right| &\leqslant \frac{n}{p^3}\sum_{\delta_t=1-n}^{n-1}\sum_{m,m'=\max\{n+1,n+1-\delta_t\}}^{\infty}\left|c_m c_{m'} c_{m+\delta_t} c_{m'+\delta_t}\right| + \mu_{2,\Delta_{X,\tilde{X}}}^{\sqcap\sqcap}\\ &\leqslant \underbrace{\frac{n}{p^3}}_{=O(n^{-2})}\left[2\underbrace{\sum_{\delta_t=1}^{n-1}\bar{c}_{\delta_t}^2}_{=O(1)}\underbrace{\sum_{m,n'=n+1}^{\infty}\bar{c}_m\bar{c}_{m'}}_{=O(n^{-2\delta})}+\underbrace{\sum_{m,n'=n+1}^{\infty}\bar{c}_m^2\bar{c}_{m'}^2}_{=O(n^{-2-4\delta})}\right]+\mu_{2,\Delta}^{\sqcap\sqcap}\\ &= O(n^{-2-2\delta}).\end{aligned} \tag{6.3.7}$$

The last two displays (6.3.6) and (6.3.7) imply that

$$\mu_{2,\Delta_{X,\tilde{X}}}^{\sqcap\!\!\!\sqcap} = \mu_{2,\Delta_{X,\tilde{X}}}^{\sqcap\!\!\!\sqcap,-}+\mu_{2,\Delta_{X,\tilde{X}}}^{\sqcap\!\!\!\sqcap,\varnothing}+\mu_{2,\Delta_{X,\tilde{X}}}^{\sqcap\!\!\!\sqcap,+} = O(n^{-2-2\delta}).$$

Thus, $\mu_{2,\Delta_{X,\tilde{X}}} = O(n^{-2-2\delta})$, as claimed.

Step 2 Next we verify assumptions ii) and iii) of Lemma 6.7, which means that we show that both

$$\Sigma_X := \frac{1}{p^2} \operatorname{tr} \boldsymbol{X}\boldsymbol{X}^T \quad \text{and} \quad \Sigma_{\widetilde{X}} := \frac{1}{p^2} \operatorname{tr} \widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^T$$

have bounded first moment and variances of order $n^{-1-\epsilon}$, for some $\epsilon > 0$; in fact, ϵ will turn out to be one. For Σ_X we obtain

$$\begin{aligned}\mu_{1,\Sigma_X} := \mathbb{E}\Sigma_X = \frac{1}{p^2} \sum_{i=1}^{p} \sum_{t=1}^{n} \mathbb{E}\boldsymbol{X}_{it}^2 &= \frac{1}{p^2} \sum_{i=1}^{p} \sum_{t=1}^{n} \mathbb{E}X_{(i-1)n+t}^2 \\ &= \frac{1}{p^2} \sum_{i=1}^{p} \sum_{t=1}^{n} \sum_{k,k'=0}^{\infty} \mathbb{E}\left[Z_{(i-1)n+t-k} Z_{(i-1)n+t-k'}\right] c_k c_{k'} = \frac{n}{p} \sum_{k=0}^{\infty} c_k^2,\end{aligned}$$

where the change of the order of expectation and summation is valid by Fubini's theorem and the observation that, by the Cauchy–Schwarz inequality,

$$\sum_{k,k'=0}^{\infty} \mathbb{E}\left|Z_{(i-1)n+t-k} Z_{(i-1)n+t-k'}\right| |c_k||c_{k'}| \leqslant \left(\sum_k |c_k|\right)^2 < \infty.$$

The first moment μ_{1,Σ_X} converges, as n tends to infinity, to $y\,\gamma(0)$, and is, in particular, bounded. Using Eq. (6.3.1) and Fubini's theorem, the second moment of Σ_X becomes

$$\begin{aligned}\mu_{2,\Sigma_X} :=& \mathbb{E}\Sigma_X^2 \\ =& \frac{1}{p^4} \sum_{i,i'=1}^{p} \sum_{t,t'=1}^{n} \sum_{\substack{k,k' \\ m,m'}=0}^{\infty} \mathbb{E}\left[Z_{(i-1)n+t-k} Z_{(i-1)n+t-k'} Z_{(i'-1)n+t'-m} Z_{(i'-1)n+t'-m'}\right] c_k c_{k'} c_m c_{m'}.\end{aligned}$$

This sum coincides with the expression analysed in Eq. (6.3.2), except that here the k, k', m and m' sums start at zero and not at $n+1$. A straightforward adaptation of the arguments there show that

$$\mu_{2,\Sigma_X} = \frac{n^2}{p^2} \left(\sum_{k=0}^{\infty} c_k^2\right)^2 + O(n^{-2}),$$

and, consequently, that $\mathbb{V}\mathrm{ar}\,\Sigma_X = \mu_{2,\Sigma_X} - (\mu_{1,\Sigma_X})^2 = O(n^{-2})$. Analogous computations show that $\mathbb{E}\Sigma_{\widetilde{X}}$ is bounded, and that $\mathbb{V}\mathrm{ar}\,\Sigma_{\widetilde{X}} = O(n^{-2})$. Thus, conditions ii) and iii) of Lemma 6.7 are verified, and the proof of the proposition is complete. □

Due to the previous Proposition 6.8 the problem of determining the limiting spectral distribution of the sample covariance matrix $p^{-1}\boldsymbol{X}\boldsymbol{X}^T$ has been reduced to computing the LSD of the matrix $p^{-1}\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^T$, where now, for fixed n, the matrix $\widetilde{\boldsymbol{X}}$ depends on only finitely many of the noise variables Z_t. The fact that the entries of $\widetilde{\boldsymbol{X}}$ are in fact finite-order moving-average processes and therefore linearly dependent on the Z_t allows for the matrix $\widetilde{\boldsymbol{X}}$ to be written as a linear transformation of the i. i. d. matrix $\boldsymbol{Z} := (Z_{(i-2)n+t})_{i=1,\dots,p+1,t=1,\dots,n}$. We emphasize that $\boldsymbol{Z}$ — in contrast to $\boldsymbol{X}$ and $\widetilde{\boldsymbol{X}}$ — is a $(p+1)\times n$ matrix; this is necessary because the entries in the first row of $\widetilde{\boldsymbol{X}}$ depend on noise variables with negative indices, up to and including Z_{1-n}. In order to formulate the transformation that maps $\boldsymbol{Z}$ to $\widetilde{\boldsymbol{X}}$ concisely in the next lemma, we define the matrices

$$K_m = \begin{pmatrix} 0 & 0 \\ \mathbf{1}_{m-1} & 0 \end{pmatrix} \in \mathbb{R}^{m\times m}, \quad m \in \mathbb{N}, \tag{6.3.8}$$

and the polynomials

$$\chi_n(z) = c_0 + c_1 z + \ldots + c_n z^n, \tag{6.3.9a}$$
$$\bar{\chi}_n(z) = z^n \chi(1/z) = c_n + c_{n-1} z + \ldots + c_0 z^n. \tag{6.3.9b}$$

Lemma 6.9 *With $\widetilde{\boldsymbol{X}}$, $\boldsymbol{Z}$, K_m and χ_n, $\bar{\chi}_n$ defined as before, it holds that*

$$\widetilde{\boldsymbol{X}} = \begin{bmatrix} 0 & \mathbf{1}_p & \mathbf{1}_p & 0 \end{bmatrix} \begin{pmatrix} \boldsymbol{Z} & 0 \\ 0 & \boldsymbol{Z} \end{pmatrix} \begin{bmatrix} \chi_n\left(K_n^T\right) \\ \bar{\chi}_n\left(K_n\right) \end{bmatrix}. \tag{6.3.10}$$

Proof Let $s_N : \mathbb{R}^N \to \mathbb{R}^N$ be the right-shift operator which is defined by $s_N(v_1, \ldots, v_N) = (0, v_1, \ldots, v_{N-1})$ and, for positive integers r, s, denote by $\operatorname{vec}_{r,s} : \mathbb{R}^{r\times s} \to \mathbb{R}^{rs}$ the bijective linear operator that transforms a matrix into a vector by horizontally concatenating its subsequent rows, starting with the first one. The operator $S_{r,s} : \mathbb{R}^{r\times s} \to \mathbb{R}^{r\times s}$ is then defined as $S_{r,s} = \operatorname{vec}_{r,s}^{-1} \circ s_{rs} \circ \operatorname{vec}_{r,s}$. This operator shifts all entries of a matrix to the right except for the entries in the last column, which are shifted down and moved into the first column. For $k = 1, 2, \ldots$, the operator $S_{r,s}^k$ is defined as the k-fold composition of $S_{r,s}$. In the following we write $S := S_{p+1,n}$. With this notation it is clear that $\widetilde{\boldsymbol{X}} = \begin{bmatrix} 0 & \mathbf{1}_p \end{bmatrix} \chi_n(S)\boldsymbol{Z}$. In order to obtain Eq. (6.3.10), we observe that the action of S can be written in terms of matrix multiplications as

$$S\boldsymbol{Z} = K_{p+1}\boldsymbol{Z}E + \boldsymbol{Z}K_n^T,$$

where the entries of $E \in \mathbb{R}^{n\times n}$ are all zero except for a one in the lower left corner. Using the fact that $E(K_n^T)^m E$ is zero for every non-negative integer m, it follows by induction that S^k, $k = 1, \ldots, n$, acts like

$$\begin{aligned} S^k\boldsymbol{Z} &= \boldsymbol{Z}(K_n^T)^k + K_{p+1}\boldsymbol{Z}\sum_{i=1}^{k} (K_n^T)^{k-i} E (K_n^T)^{i-1} \\ &= \boldsymbol{Z}(K_n^T)^k + K_{p+1}\boldsymbol{Z}K_n^{n-k} \\ &= \begin{bmatrix} \mathbf{1}_{p+1} & K_{p+1} \end{bmatrix} \begin{pmatrix} \boldsymbol{Z} & 0 \\ 0 & \boldsymbol{Z} \end{pmatrix} \begin{bmatrix} (K_n^T)^k \\ K_n^{n-k} \end{bmatrix}. \end{aligned}$$

This implies that

$$
\begin{aligned}
\widetilde{\boldsymbol{X}} &= \begin{bmatrix} 0 & \mathbf{1}_p \end{bmatrix} \chi_n(S)\boldsymbol{Z} \\
&= \begin{bmatrix} 0 & \mathbf{1}_p \end{bmatrix} \sum_{k=0}^{n} c_k S^k \boldsymbol{Z} \\
&= \begin{bmatrix} 0 & \mathbf{1}_p \end{bmatrix} \begin{bmatrix} \mathbf{1}_{p+1} & K_{p+1} \end{bmatrix} \begin{pmatrix} \boldsymbol{Z} & 0 \\ 0 & \boldsymbol{Z} \end{pmatrix} \sum_{k=0}^{n} c_k \begin{bmatrix} (K_n^T)^k \\ K_n^{n-k} \end{bmatrix} \\
&= \begin{bmatrix} 0 & \mathbf{1}_p & \mathbf{1}_p & 0 \end{bmatrix} \begin{pmatrix} \boldsymbol{Z} & 0 \\ 0 & \boldsymbol{Z} \end{pmatrix} \begin{bmatrix} \chi_n\left(K_n^T\right) \\ \bar{\chi}_n\left(K_n\right) \end{bmatrix},
\end{aligned}
$$

and completes the proof. □

While the last lemma gives quite an explicit description of the relation between $\boldsymbol{Z}$ and $\widetilde{\boldsymbol{X}}$, it is impractical for directly determining the limiting spectral distribution of $p^{-1}\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^T$. The reason is that $\boldsymbol{Z}$ appears twice in the central block-diagonal matrix and is moreover multiplied by some deterministic matrices from both the left and the right. The LSD of the product of three random matrices has been computed in the literature (Zhang, 2006), but this result is not applicable in our situation due to the appearance of the random block matrix in Eq. (6.3.10). Sample covariance matrices derived from random block matrices have been considered in Rashidi Far et al. (2008). However, they only treat the Gaussian case and, more importantly, do not cover the case of a non-trivial population covariance matrix. We are thus not aware of any result allowing to derive the LSD of $p^{-1}\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^T$ directly from Lemma 6.9.

The next proposition allows us to circumvent this problem. It is shown that, at least asymptotically and at the cost of slightly changing the size of the involved matrices, one can simplify the structure of $\widetilde{\boldsymbol{X}}$ so that $\boldsymbol{Z}$ appears only once and is multiplied by a deterministic matrix only from the right.

Proposition 6.10 *Let $\widetilde{\boldsymbol{X}}$, $\boldsymbol{Z}$, K_n and χ_n, $\bar{\chi}_n$ be given as before and define the matrix*

$$
\widehat{\boldsymbol{X}} := \boldsymbol{Z} \begin{bmatrix} 0 & \mathbf{1}_n & \mathbf{1}_n & 0 \end{bmatrix} \begin{bmatrix} \chi_{n+1}(K_{n+1}^T) \\ \bar{\chi}_{n+1}(K_{n+1}) \end{bmatrix} \in \mathbb{R}^{(p+1)\times(n+1)}. \tag{6.3.11}
$$

If the empirical spectral distribution of $p^{-1}\widehat{\boldsymbol{X}}\widehat{\boldsymbol{X}}^T$ converges to a limit, then the empirical spectral distribution of $p^{-1}\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^T$ converges to the same limit.

Proof In order to be able to compare the limiting spectral distributions of $p^{-1}\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^T$ and $p^{-1}\widehat{\boldsymbol{X}}\widehat{\boldsymbol{X}}^T$ in spite of their dimensions being different, we introduce the matrix

$$
\overline{\boldsymbol{X}} = \begin{bmatrix} 0 & 0 \\ 0 & \widetilde{\boldsymbol{X}} \end{bmatrix} \in \mathbb{R}^{(p+1)\times(n+1)}. \tag{6.3.12}
$$

Clearly, $F^{p^{-1}\overline{X}\overline{X}^T} = \frac{1}{p+1}\delta_0 + \frac{p}{p+1}F^{p^{-1}\widetilde{X}\widetilde{X}^T}$, which implies equality of the limiting spectral distributions, i. e. $\hat{F}^{p^{-1}\overline{X}\overline{X}^T} = \hat{F}^{p^{-1}\widetilde{X}\widetilde{X}^T}$, provided that either of the two, and hence both, exists. It is therefore sufficient to show that the LSDs of $p^{-1}\widehat{X}\widehat{X}^T$ and $p^{-1}\overline{X}\overline{X}^T$ are identical; this will be done by verifying the three conditions of Lemma 6.7. We first observe that, by definition, the (i,j)th entry of $\widehat{X}$ is given by

$$\widehat{X}_{ij} = \sum_{k=1}^{j-1} Z_{(i-2)n+k}c_{j-k-1} + \sum_{k=j}^{n} Z_{(i-2)n+k}c_{j-k+n+1}, \tag{6.3.13}$$

whereas the (i,j)th entry of $\overline{X}$ has the form

$$\overline{X}_{ij} = \begin{cases} 0, & \text{if } i=1 \text{ or } j=1, \\ \sum_{k=1}^{j-1} Z_{(i-2)n+k}c_{j-k-1} + \sum_{k=j-1}^{n} Z_{(i-3)n+k}c_{j-k+n-1}, & \text{else.} \end{cases} \tag{6.3.14}$$

The remainder of the proof will be divided into two parts. In the first part we check the validity of assumption i) for the difference $\widehat{X} - \overline{X}$, whereas in the second one we consider the terms $\operatorname{tr}\widehat{X}\widehat{X}^T$ and $\operatorname{tr}\overline{X}\overline{X}^T$, which appear in conditions ii) and iii).

Step 1 Using the definitions of $\widehat{X}$, Eq. (6.3.11), and of $\overline{X}$, Eq. (6.3.12), it follows that

$$\begin{aligned}
\Delta_{\widehat{X},\overline{X}} :=& \frac{1}{p^2}\operatorname{tr}\left(\widehat{X}-\overline{X}\right)\left(\widehat{X}-\overline{X}\right)^T \\
=& \frac{1}{p^2}\sum_{i=1}^{p+1}\sum_{j=1}^{n+1}\left[\widehat{X}_{ij}-\overline{X}_{ij}\right]^2 \\
\leqslant& \frac{2}{p^2}\sum_{i=2}^{p+1}\sum_{j=2}^{n+1}\left[\sum_{k,k'=j}^{n} Z_{(i-2)n+k}Z_{(i-2)n+k'}c_{j-k+n+1}c_{j-k'+n+1}\right. \\
&\left. + \sum_{k,k'=j-1}^{n} Z_{(i-3)n+k}Z_{(i-3)n+k'}c_{j-k+n-1}c_{j-k'+n-1}\right] \\
&+ \frac{1}{p^2}\sum_{i=1}^{p+1}\sum_{k,k'=1}^{n} Z_{(i-2)n+k}Z_{(i-2)n+k'}c_{n-k+2}c_{n-k'+2} \\
&+ \frac{2}{p^2}\sum_{j=2}^{n+1}\sum_{k,k'=1}^{j-1} Z_{-n+k}Z_{-n+k'}c_{j-k-1}c_{j-k'-1} \\
&+ \frac{2}{p^2}\sum_{j=2}^{n+1}\sum_{k,k'=j}^{n} Z_{-n+k}Z_{-n+k'}c_{j-k+n+1}c_{j-k'+n+1} \\
=:& \Delta^{(1)}_{\widehat{X},\overline{X}} + \Delta^{(2)}_{\widehat{X},\overline{X}} + \Delta^{(3)}_{\widehat{X},\overline{X}} + \Delta^{(4)}_{\widehat{X},\overline{X}} + \Delta^{(5)}_{\widehat{X},\overline{X}},
\end{aligned} \tag{6.3.15}$$

where the elementary inequality $(a+b)^2 \leqslant 2a^2 + 2b^2$ was used twice. In order to show that the L^2-norm of expression (6.3.15) is summable, we consider each term in turn. For the second moment of the first term of Eq. (6.3.15) we obtain

$$\begin{aligned}\mu_{2,\Delta^{(1)}_{\widehat{X},\overline{X}}} :=& \mathbb{E}\left(\Delta^{(1)}_{\widehat{X},\overline{X}}\right)^2 \\ =& \frac{4}{p^4}\sum_{i,i'=2}^{p+1}\sum_{j,j'=2}^{n+1}\sum_{k,k'=1}^{n-j+1}\sum_{m,m'=1}^{n-j'+1}\mathbb{E}\left[Z_{(i-1)n-k+1}Z_{(i-1)n-k'+1}Z_{(i'-1)n-m+1}Z_{(i'-1)n-m'+1}\right]\times \\ & c_{j+k}c_{j+k'}c_{j'+m}c_{j'+m'}.\end{aligned}$$

As before, we consider all configurations where above expectation is not zero. The expectation equals σ_4 if $i = i'$ and k, k', m, m' are equal, hence

$$\mu^{\sqcap\!\sqcap\!\sqcap}_{2,\Delta^{(1)}_{\widehat{X},\overline{X}}} \leqslant \frac{4\sigma_4}{p^4}\sum_{i=2}^{p+1}\sum_{k=1}^{n}\left(\sum_{j=2}^{n+1}c^2_{j+k}\right)^2 \leqslant \frac{4\sigma_4}{p^3}\sum_{k=1}^{n}\bar{c}^2_k\left(\sum_{j=2}^{n+1}\bar{c}_j\right)^2 = O(n^{-3}).$$

The expectation is one if the four Z can be collected in two non-equal pairs. The first term equals the second and the third equals the fourth if $k = k'$ and $m = m'$, which implies that

$$\begin{aligned}\mu^{\sqcap\,\sqcap}_{2,\Delta^{(1)}_{\widehat{X},\overline{X}}} &= \frac{4}{p^4}\sum_{i,i'=2}^{p+1}\sum_{j,j'=2}^{n+1}\sum_{k=1}^{n-j+1}\sum_{m=1}^{n-j'+1}c^2_{j+k}c^2_{j'+m} - \mu^{\sqcap\!\sqcap\!\sqcap}_{2,\Delta^{(1)}_{\widehat{X},\overline{X}}} \\ &= \frac{4}{p^2}\left(\sum_{j=2}^{n+1}\sum_{k=1}^{n-j+1}c^2_{j+k}\right)^2 - \mu^{\sqcap\!\sqcap\!\sqcap}_{2,\Delta^{(1)}_{\widehat{X},\overline{X}}} \\ &= O(n^{-2}).\end{aligned}$$

Likewise, the contribution from pairing the first factor with the third, and the second with the fourth, can be estimated as

$$\begin{aligned}\left|\mu^{\sqcap\!\!\sqcap}_{2,\Delta^{(1)}_{\widehat{X},\overline{X}}}\right| &\leqslant \frac{4}{p^4}\sum_{i'=2}^{p+1}\sum_{j,j'=2}^{n+1}\sum_{k,k'=1}^{n}\left|c_{j+k}c_{j+k'}c_{j'+k}c_{j'+k'}\right| + \mu^{\sqcap\!\sqcap\!\sqcap}_{2,\Delta^{(1)}_{\widehat{X},\overline{X}}} \\ &\leqslant \frac{4}{p^3}\left(\sum_{j=1}^{n+1}\bar{c}_j\right)^4 + \mu^{\sqcap\!\sqcap\!\sqcap}_{2,\Delta^{(1)}_{\widehat{X},\overline{X}}} = O(n^{-3}).\end{aligned}$$

Obviously, the configuration $\mu^{\sqcap}_{2,\Delta^{(1)}_{\widehat{X},\overline{X}}}$ can be handled the same way as $\mu^{\sqcap\!\!\sqcap}_{2,\Delta^{(1)}_{\widehat{X},\overline{X}}}$ above. Thus, we have shown that the second moment of $\Delta^{(1)}_{\widehat{X},\overline{X}}$, the first term in Eq. (6.3.15), is of order n^{-2}. This can be shown for the second term in Eq. (6.3.15) in the same way. We now consider the third term in Eq. (6.3.15), whose second moment is

$$
\begin{aligned}
\mu_{2,\Delta^{(3)}_{\widehat{X},\overline{X}}} :=& \mathbb{E}\left(\Delta^{(3)}_{\widehat{X},\overline{X}}\right)^2 \\
=& \frac{1}{p^4}\sum_{i,i'=1}^{p+1}\sum_{\substack{k,k'\\m,m'}=1}^{n}\mathbb{E}\left[Z_{(i-2)n+k}Z_{(i-2)n+k'}Z_{(i'-2)n+m}Z_{(i'-2)n+m'}\right]\times \\
& c_{n-k+2}c_{n-k'+2}c_{n-m+2}c_{n-m'+2}.
\end{aligned}
$$

Distinguishing the same cases as before, we have

$$
\mu^{\sqcap\!\sqcap\!\sqcap}_{2,\Delta^{(3)}_{\widehat{X},\overline{X}}} = \sigma_4\frac{p+1}{p^4}\sum_{k=1}^{n}c^4_{n-k+2} = O(n^{-3}),
$$

and thus

$$
\mu^{\sqcap\,\sqcap}_{2,\Delta^{(3)}_{\widehat{X},\overline{X}}} = \frac{(p+1)^2}{p^4}\left(\sum_{k=1}^{n}c^2_{n-k+2}\right)^2 - \mu^{\sqcap\!\sqcap\!\sqcap}_{2,\Delta^{(3)}_{\widehat{X},\overline{X}}} = O(n^{-2}),
$$

as well as

$$
\mu^{\overline{\sqcap}}_{2,\Delta^{(3)}_{\widehat{X},\overline{X}}} = \mu^{\sqcap\!\!\sqcap}_{2,\Delta^{(3)}_{\widehat{X},\overline{X}}} = O(n^{-3}).
$$

Thus, the second moment of the third term in Eq. (6.3.15) is of order n^{-2}; repeating the foregoing arguments, it can be seen that the second moments of $\Delta^{(4)}_{\widehat{X},\overline{X}}$ and $\Delta^{(5)}_{\widehat{X},\overline{X}}$, the two last terms in Eq. (6.3.15), are of order $O(n^{-2})$ as well, so that we have shown that

$$
\frac{1}{p^4}\mathbb{E}\left[\operatorname{tr}\left(\widehat{X}-\overline{X}\right)\left(\widehat{X}-\overline{X}\right)^T\right]^2 = \mathbb{E}\left(\Delta_{\widehat{X},\overline{X}}\right)^2 \leqslant 5\sum_{i=1}^{5}\mu_{2,\Delta^{(i)}_{\widehat{X},\overline{X}}} = O(n^{-2}).
$$

Step 2 In this step we shall prove that both

$$
\Sigma_{\widehat{X}} := \frac{1}{p^2}\operatorname{tr}\widehat{X}\widehat{X}^T \quad \text{and} \quad \Sigma_{\overline{X}} := \frac{1}{p^2}\operatorname{tr}\overline{X}\overline{X}^T
$$

have bounded first moments, and that their variances are summable sequences in n, i. e. we check conditions ii) and iii) of Lemma 6.7. Since $\operatorname{tr}\overline{X}\overline{X}^T$ is equal to $\operatorname{tr}\widetilde{X}\widetilde{X}^T$, the claim about $\Sigma_{\overline{X}}$ has already been shown in the second step of the proof of Proposition 6.8. For the first term one finds, by Eq. (6.3.13), that

$$
\Sigma_{\widehat{X}} = \frac{1}{p^2}\sum_{i=1}^{p+1}\sum_{j=1}^{n+1}\left(\sum_{k=1}^{j-1}Z_{(i-2)n+k}c_{j-k-1} + \sum_{k=j}^{n}Z_{(i-2)n+k}c_{j-k+n+1}\right)^2
$$

$$
\begin{aligned}
&\leqslant \frac{2}{p^2}\sum_{i=1}^{p+1}\sum_{j=1}^{n+1}\sum_{k,k'=1}^{j-1} Z_{(i-2)n+k}c_{j-k-1}Z_{(i-2)n+k'}c_{j-k'-1} \\
&\quad + \frac{2}{p^2}\sum_{i=1}^{p+1}\sum_{j=1}^{n+1}\sum_{k,k'=j}^{n} Z_{(i-2)n+k}c_{j-k+n+1}Z_{(i-2)n+k'}c_{j-k'+n+1} =: \Sigma_{\widehat{X}}^{(1)} + \Sigma_{\widehat{X}}^{(2)}.
\end{aligned}
$$

Clearly, the first two moments of $\Sigma_{\widehat{X}}^{(1)}$ are given by

$$
\mu_{1,\Sigma_{\widehat{X}}^{(1)}} := \mathbb{E}\Sigma_{\widehat{X}}^{(1)} = \frac{2}{p^2}\sum_{i=1}^{p+1}\sum_{j=1}^{n+1}\sum_{k,k'=1}^{j-1}\mathbb{E}\left[Z_{(i-2)n+k}Z_{(i-2)n+k'}\right]c_{j-k-1}c_{j-k'-1} = \frac{2(p+1)}{p^2}\sum_{j=1}^{n+1}\sum_{k=1}^{j-1}c_{k-1}^2
$$

and

$$
\begin{aligned}
\mu_{2,\Sigma_{\widehat{X}}^{(1)}} := \mathbb{E}\left(\Sigma_{\widehat{X}}^{(1)}\right)^2 = \frac{4}{p^4}\sum_{i,i'=1}^{p+1}\sum_{j,j'=1}^{n+1}\sum_{k,k'=1}^{j-1}\sum_{m,m'=1}^{j'-1}&\mathbb{E}(Z_{(i-2)n+k}Z_{(i-2)n+k'}Z_{(i'-2)n+m}Z_{(i'-2)n+m'})\times \\
&c_{j-k-1}c_{j-k'-1}c_{j'-m-1}c_{j'-m'-1}.
\end{aligned}
$$

We separately consider the case that all four factors are equal, and the three possible pairings of the four Z. If all four Z are equal, it must hold that $i = i'$, $k = k' = m = m'$, which results in a contribution

$$
\begin{aligned}
\mu_{2,\Sigma_{\widehat{X}}^{(1)}}^{\sqcap\!\sqcap\!\sqcap} &= \frac{4\sigma_4}{p^4}\sum_{i=1}^{p+1}\sum_{j,j'=1}^{n+1}\sum_{k=1}^{\min\{j,j'\}-1} c_{j-k-1}^2 c_{j'-k-1}^2 \\
&\leqslant \frac{4\sigma_4(p+1)}{p^4}\sum_{j,j'=1}^{n+1}\bar{c}_{j-\min\{j,j'\}}\bar{c}_{j'-\min\{j,j'\}}\sum_{k=1}^{\min\{j,j'\}-1}\bar{c}_{k-1}^2 \\
&\leqslant \frac{4\sigma_4(p+1)}{p^4}\sum_{j,j'=1}^{n+1}\bar{c}_0\bar{c}_{|j-j'|}\sum_{k=1}^{n}\bar{c}_{k-1}^2.
\end{aligned}
$$

Introducing the new summation variable $\delta_j := j - j'$, we find that

$$
\mu_{2,\Sigma_{\widehat{X}}^{(1)}}^{\sqcap\!\sqcap\!\sqcap} \leqslant \frac{4\sigma_4(p+1)(n+1)}{p^4}\bar{c}_0\left[\bar{c}_0 + 2\sum_{\delta_j=1}^{n}\bar{c}_{\delta_j}\right]\sum_{k=1}^{n}\bar{c}_{k-1}^2 = O(n^{-2}). \tag{6.3.16}
$$

The first factor being paired with the second, and the third with the fourth, means that $k = k'$, $m = m'$ and $m \neq (i-i')n + k$, so that the contribution of this configuration is given by

$$
\mu_{2,\Sigma_{\widehat{X}}^{(1)}}^{\sqcap\,\sqcap} = \frac{4}{p^4}\sum_{i,i'=1}^{p+1}\sum_{j,j'=1}^{n+1}\sum_{k=1}^{j-1}\sum_{m=1}^{j'-1}c_{j-k-1}^2c_{j'-m-1}^2 - \mu_{2,\Sigma_{\widehat{X}}^{(1)}}^{\sqcap\!\sqcap\!\sqcap} = \left(\mu_{1,\Sigma_{\widehat{X}}^{(1)}}\right)^2 + O(n^{-2}). \tag{6.3.17}
$$

For the $\sqcap\!\!\!\sqcap$ pairing one has the constraint $i = i'$, $k = m$, $k' = m'$, $k \neq k'$, and the corresponding contribution is

$$\begin{aligned}
\mu_{2,\Sigma^{(1)}_{\widehat{X}}}^{\sqcap\!\!\!\sqcap} &= \frac{4}{p^4} \sum_{i=1}^{p+1} \sum_{j,j'=1}^{n+1} \sum_{k,k'=1}^{\min\{j,j'\}-1} c_{j-k-1} c_{j-k'-1} c_{j'-k-1} c_{j'-k'-1} - \mu_{2,\Sigma^{(1)}_{\widehat{X}}}^{\sqcap\sqcap} \\
&\leqslant \frac{4(p+1)}{p^4} \sum_{j,j'=1}^{n+1} \bar{c}_{j-\min\{j,j'\}} \bar{c}_{j'-\min\{j,j'\}} \sum_{k,k'=1}^{\min\{j,j'\}-1} \bar{c}_{k-1} c_{k'-1} + O(n^{-2}) \\
&\leqslant \frac{4(p+1)(n+1)}{p^4} \bar{c}_0 \left[\bar{c}_0 + 2 \sum_{\delta_j=1}^{n} \bar{c}_{\delta_j} \right] \sum_{k,k'=1}^{n} \bar{c}_{k-1} \bar{c}_{k'-1} + O(n^{-2}) = O(n^{-2}). \qquad (6.3.18)
\end{aligned}$$

Renaming the summation indices shows that $\mu_{2,\Sigma^{(1)}_{\widehat{X}}}^{\sqcap\!\!\sqcap}$ is equal to $\mu_{2,\Sigma^{(1)}_{\widehat{X}}}^{\sqcap\!\!\!\sqcap}$. Combining this observation with the displays (6.3.16) to (6.3.18), it follows that

$$\mathbb{V}\mathrm{ar}\, \Sigma^{(1)}_{\widehat{X}} = \mu_{2,\Sigma^{(1)}_{\widehat{X}}} - \left(\mu_{1,\Sigma^{(1)}_{\widehat{X}}} \right)^2 = O(n^{-2}),$$

and since a very similar reasoning can be applied to $\Sigma^{(2)}_{\widehat{X}}$, we conclude that

$$\frac{1}{p^4} \mathbb{V}\mathrm{ar}\, \mathrm{tr}\, \widehat{\boldsymbol{X}} \widehat{\boldsymbol{X}}^T \leqslant 2\, \mathbb{V}\mathrm{ar}\, \Sigma^{(1)}_{\widehat{X}} + 2\, \mathbb{V}\mathrm{ar}\, \Sigma^{(2)}_{\widehat{X}} = O(n^{-2}). \qquad \square$$

The intention behind Proposition 6.10 was to bring us into a situation where we can apply results about the limiting spectral distribution of matrices of the form $\boldsymbol{Z} H \boldsymbol{Z}^T$, where $\boldsymbol{Z}$ is an i. i. d. matrix and H is a positive-semidefinite population covariance matrix. Expressions for the Stieltjes transform of the LSD of such matrices in terms of the LSD of H have been obtained in Marchenko and Pastur (1967); Silverstein and Bai (1995), and in the most general form in Pan (2010). The next lemma shows that in the current context the population covariance matrix H has the same LSD as the autocovariance matrix Γ of the process X, which is defined by

$$\Gamma = (\gamma(i-j))_{ij}, \quad \gamma(h) = \sum_{j=0}^{\infty} c_j c_{j+|h|}, \quad h \in \mathbb{Z}. \qquad (6.3.19)$$

Lemma 6.11 *Let K_n and χ_n, $\bar{\chi}_n$ be defined by Eqs. (6.3.8) and (6.3.9). The limiting spectral distribution of $\Omega\Omega^T$, where*

$$\Omega = \begin{bmatrix} 0 & \mathbf{1}_n & \mathbf{1}_n & 0 \end{bmatrix} \begin{bmatrix} \chi_{n+1}\left(K_{n+1}^T\right) \\ \bar{\chi}_{n+1}\left(K_{n+1}\right) \end{bmatrix} \in \mathbb{R}^{n \times n+1}, \qquad (6.3.20)$$

exists and is the same as the limiting spectral distribution of Γ, defined in Eq. (6.3.19).

Proof The proof proceeds via a two-stage comparison argument. First, we define an approximate matrix square root T_n of the autocovariance matrix Γ and show that the spectral distribution $F^{T_n T_n^T}$ converges to $\hat{F}^{\Gamma}$, which exists by Lemma 5.10. In the second step we use the trace criterion to prove that $\hat{F}^{\Omega\Omega^T}$ exists and is equal to $\hat{F}^{T_n T_n^T}$.

Step 1 We define the matrix $T = T_n \in \mathbb{R}^{n\times n}$ by $T_{ij} = c_{j-i-1} I_{\{j\geqslant i+1\}}$. By a variant of the trace criterion (Bai and Silverstein, 2010, Corollary A.41), it suffices to show that $n^{-1}\,\mathrm{tr}\left[(\Gamma - TT^T)^2\right]$ converges to zero. By the definition of T we have

$$\left(TT^T\right)_{ij} = \sum_{k=\max\{i,j\}+1}^{n+1} c_{k-i-1}c_{k-j-1} = \sum_{k=0}^{n-\max\{i,j\}} c_k c_{k+|j-i|},$$

which identifies T as a truncated square root of Γ. Hence

$$\begin{aligned}\frac{1}{n}\,\mathrm{tr}\left[(\Gamma - TT^T)^2\right] &= \frac{1}{n}\sum_{i,j=1}^{n}\left(\sum_{k=n-\max\{i,j\}+1}^{\infty} c_k c_{k+|j-i|}\right)^2\\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\sum_{k=1}^{\infty} c_{k+j-1}c_{k+i-1}\right)^2\\ &\leqslant \frac{1}{n}\sum_{i=1}^{n}\bar{c}_{i-1}\sum_{j=1}^{n}\bar{c}_{j-1}\sum_{k=1}^{\infty}\bar{c}_k\sum_{k'=1}^{\infty}\bar{c}_{k'} = O(n^{-1}),\end{aligned}$$

which completes the first step.

Step 2 As can be seen easily, the matrix Ω is, except for one missing row, a circulant matrix whose entries are given by

$$\Omega_{ij} = c_{n+j-i \mod (n+1)} = c_{j-i-1 \mod (n+1)}, \quad i = 1,\dots,n, \quad j = 1,\dots,n+1.$$

By Lemma 6.7 it is sufficient to show that the expression $n^{-1}\,\mathrm{tr}(TT^T + \Omega\Omega^T)$ converges to a finite limit, and that $n^{-1}\,\mathrm{tr}\left[(\Omega - T)(\Omega - T)^T\right]$ converges to zero. By the definition of T and Ω we have that

$$\begin{aligned}\frac{1}{n}\,\mathrm{tr}\left[(\Omega - T)(\Omega - T)^T\right] &= \frac{1}{n}\sum_{j=1}^{n}\sum_{i=j}^{n} c_{j-i+n}^2 \\ &= \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n-j+1} c_{j+i}^2 \leqslant \frac{1}{n}\sum_{j=1}^{n}\bar{c}_j\sum_{i=1}^{n}\bar{c}_i = O(n^{-1}).\end{aligned} \tag{6.3.21}$$

Furthermore, we obtain that

$$\frac{1}{n}\operatorname{tr}(TT^T) = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=0}^{n-i} c_k^2 \leqslant \sum_{k=0}^{\infty} c_k^2 < \infty.$$

Thus, by Eq. (6.3.21), one has

$$\begin{aligned}\frac{1}{n}\operatorname{tr}(\Omega\Omega^T) =& \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n+1} c_{j-i-1 \mod (n+1)}\\ =& \frac{1}{n}\sum_{j=1}^{n+1}\sum_{i=j}^{n} c_{j-i+n}^2 + \frac{1}{n}\sum_{j=1}^{n+1}\sum_{i=1}^{j-1} c_{j-i-1}^2\\ \leqslant& O(n^{-1}) + \frac{n+1}{n}\sum_{s=1}^{n} c_{s-1}^2 < \infty.\end{aligned}$$

This completes the second step and thereby the proof of the lemma. □

We are now ready to prove our main result.

Proof (of Theorem 6.1) According to Eq. (6.3.11), the matrix $\widehat{\boldsymbol{X}}\widehat{\boldsymbol{X}}^T$ is of the form $\boldsymbol{Z}\Omega\Omega^T\boldsymbol{Z}^T$, where Ω is given by Eq. (6.3.20). Using Pan (2010, Theorem 1) and the fact that, by Lemma 6.11, the limiting spectral distribution of $\Omega\Omega^T$ exists, it follows that the limiting spectral distribution $\hat{F}^{p^{-1}\widehat{\boldsymbol{X}}\widehat{\boldsymbol{X}}^T}$ exists. Therefore, the combination of Propositions 6.8 and 6.10 shows that the limiting spectral distribution of $p^{-1}\boldsymbol{X}\boldsymbol{X}^T$ also exists and is the same as that of $p^{-1}\widehat{\boldsymbol{X}}\widehat{\boldsymbol{X}}^T$.

Moreover, it was shown in Lemma 6.11 that the limiting spectral distributions $\hat{F}^{\Omega\Omega^T}$ and $\hat{F}^{\Gamma}$ are the same, and so Pan (2010, equation (1.2)) implies that the Stieltjes transform $z \mapsto s_{\hat{F}^{p^{-1}\boldsymbol{X}\boldsymbol{X}^T}}(z)$ of $\hat{F}^{p^{-1}\boldsymbol{X}\boldsymbol{X}^T}$ is the unique mapping from $\mathbb{C}^+$ to $\mathbb{C}^+$ which solves

$$\frac{1}{s_{\hat{F}^{p^{-1}\boldsymbol{X}\boldsymbol{X}^T}}(z)} = -z + y\int_{\mathbb{R}} \frac{\lambda}{1+\lambda s_{\hat{F}^{p^{-1}\boldsymbol{X}\boldsymbol{X}^T}}(z)}\hat{F}^{\Gamma}(\mathrm{d}\lambda), \tag{6.3.22}$$

where $y = \lim p/n$. Under the assumptions of Theorem 6.1 it was shown in Lemma 5.10 that $\hat{F}^{\Gamma}$ is an absolutely continuous distribution with bounded support and density

$$g : (\lambda_-, \lambda_+) \to \mathbb{R}^+, \quad \lambda \mapsto \frac{1}{2\pi}\sum_{\omega: f(\omega)=\lambda} \frac{1}{|f'(\omega)|},$$

where $f : [0, 2\pi] \to \mathbb{R}$ is the Fourier transform of the sequence $(\gamma(h))_{h\in\mathbb{Z}}$. Inserting the last expression into Eq. (6.3.22) completes the proof of the theorem. □

6.4. Sketch of an alternative proof of Theorem 6.1

In this section we indicate how Theorem 6.1 could be proved alternatively using the methods employed in Chapter 5. We denote by $\widetilde{\boldsymbol{X}}_{(\alpha)}$ the matrix which is defined as in Eq. (6.2.1) but with the linear process being truncated at $\lfloor n^\alpha \rfloor$ with $0 < \alpha < 1$, i. e.

$$\widetilde{\boldsymbol{X}}_{(\alpha)} = \left(\sum_{j=0}^{\lfloor n^\alpha \rfloor} c_j Z_{(i-1)n+t-j} \right)_{it} .$$

If $1 - \alpha$ is sufficiently small, then an adaptation of the proof of Proposition 6.8 to this setting shows that $p^{-1}\boldsymbol{X}\boldsymbol{X}^T$ and $p^{-1}\widetilde{\boldsymbol{X}}_{(\alpha)}\widetilde{\boldsymbol{X}}_{(\alpha)}^T$ have the same limiting spectral distribution almost surely. The next step is to partition $\widetilde{\boldsymbol{X}}_{(\alpha)}$ into two blocks of dimensions $p \times \lfloor n^\alpha \rfloor$ and $p \times (n - \lfloor n^\alpha \rfloor)$, respectively. If we denote these two blocks by $\widetilde{\boldsymbol{X}}_{(\alpha)}^1$ and $\widetilde{\boldsymbol{X}}_{(\alpha)}^2$, i. e. $\widetilde{\boldsymbol{X}}_{(\alpha)} = \left[\widetilde{\boldsymbol{X}}_{(\alpha)}^1 \, \widetilde{\boldsymbol{X}}_{(\alpha)}^2 \right]$, then clearly $\widetilde{\boldsymbol{X}}_{(\alpha)}\widetilde{\boldsymbol{X}}_{(\alpha)}^T = \widetilde{\boldsymbol{X}}_{(\alpha)}^1 \left(\widetilde{\boldsymbol{X}}_{(\alpha)}^1 \right)^T + \widetilde{\boldsymbol{X}}_{(\alpha)}^2 \left(\widetilde{\boldsymbol{X}}_{(\alpha)}^2 \right)^T$, and an application of Bai and Silverstein (2010, Theorem A.43) yields that

$$\begin{aligned}
\sup_{\lambda \in \mathbb{R}_{\geqslant 0}} \left| F^{p^{-1}\boldsymbol{X}\boldsymbol{X}^T}([0,\lambda]) - F^{p^{-1}\widetilde{\boldsymbol{X}}_{(\alpha)}^2 \left(\widetilde{\boldsymbol{X}}_{(\alpha)}^2\right)^T}([0,\lambda]) \right| &\leqslant \frac{1}{p} \operatorname{rank} \left(\widetilde{\boldsymbol{X}}_{(\alpha)}^1 \left(\widetilde{\boldsymbol{X}}_{(\alpha)}^1 \right)^T \right) \\
&\leqslant \frac{1}{p} \min \left(\lfloor n^\alpha \rfloor, p \right) \\
&= O\left(p^{-1} n^\alpha \right) \to 0.
\end{aligned}$$

It therefore suffices to derive the limiting spectral distribution of $p^{-1}\widetilde{\boldsymbol{X}}_{(\alpha)}^2 \left(\widetilde{\boldsymbol{X}}_{(\alpha)}^2 \right)^T$. Since the matrix $\widetilde{\boldsymbol{X}}_{(\alpha)}^2$ has independent rows, this could be done by a careful adaptation of the arguments given in Chapter 5. We chose, however, to provide a self-contained proof, which also provides intermediate results of independent interest like Proposition 6.10, and we therefore omit the lengthy details of this alternative proof.

Part III

First-Passage Percolation on the Ladder Graph

7. On the Markov Transition Kernels for First-Passage Percolation on the Ladder

7.1. Introduction

The subject of first-passage percolation, introduced in Hammersley and Welsh (1965), is the study of shortest paths in random graphs and their geometric properties. Let $G = (V, E)$ be a graph with vertex set V and unoriented edges $E \subset V^2$, and assume that there is a weight function $w : E \to \mathbb{R}^+$. For vertices $u, v \in V$, a path joining u and v in G is a sequence of vertices $\boldsymbol{p}_{u\to v} = \{u = p_0, p_1, \ldots, p_{n-1}, p_n = v\}$ such that $(p_\nu, p_{\nu+1}) \in E$ for $0 \leqslant \nu < n$. The weight $w(\boldsymbol{p}_{u\to v})$ of such a path is defined as the sum of the weights of the comprising edges, i. e. $w(\boldsymbol{p}_{u\to v}) := \sum_{\nu=0}^{n-1} w((p_\nu, p_{\nu+1}))$. The first-passage time between the vertices u and v is denoted by $d_G(u, v)$ and defined as $d_G(u, v) := \inf\{w(\boldsymbol{p}), \boldsymbol{p}$ a path joining u and v in $G\}$.

First-passage percolation can be considered a model for the spread of a fluid through a random porous medium; it differs from ordinary percolation theory in that it puts special emphasis on the dynamical aspect of how long it takes for certain points in the medium to be reached by the fluid. Important applications include the spread of infectious diseases (Altmann, 1993) and the analysis of electrical networks (Grimmett and Kesten, 1984). Recently, there has also been an increased interest in first-passage percolation on graphs where not only the edge-weights, but the edge-structure itself is random. These models, including the Gilbert and Erdős-Rényi random graph, were investigated in Bhamidi, van der Hofstad and Hooghiemstra (2010); Sood, Redner and ben Avraham (2005); van der Hofstad, Hooghiemstra and Van Mieghem (2001); they were found to be a useful approximation to the internet as well as telecommunication networks.

Usually, however, the underlying graph is taken to be $\mathbb{Z}^d$, $d \geqslant 2$, and the edge weights are independent random variables with some common distribution $\mathbb{P}$, see, e. g., Kesten (1986); Smythe and Wierman (1978) and references therein. Interesting mathematical questions arising in this context include asymptotic properties of the sets $B_t := \{u \in \mathbb{Z}^d : d_{\mathbb{Z}^d}(0, u) \leqslant t\}$ (Kesten, 1987; Seppäläinen, 1998), the existence and properties of geodesics (Licea and Newman, 1996), and the limiting behaviour of $d_{\mathbb{Z}^d}((0,0),(n,0))/n$ (Kesten, 1993). The latter expression is known to converge, under weak assumptions on $\mathbb{P}$, to a deterministic constant, called the *first-passage percolation rate*. The computation of this constant has proved to be a very difficult problem and has not yet been accomplished even for the simplest choices of $\mathbb{P}$

(Graham, Grötschel and Lovász, 1995, p. 1937). An exception to this is the case when the underlying graph G is essentially one-dimensional (Flaxman, Gamarnik and Sorkin, 2006; Renlund, 2010; Schlemm, 2009).

In this chapter we consider the first-passage percolation problem on the ladder G, a particular essentially one-dimensional graph, for which the first-passage percolation rate is known (Renlund, 2010; Schlemm, 2009). We extend the existing results about the almost sure convergence of $d_G((0,0),(n,0))/n$, as $n \to \infty$, by providing a Central Limit Theorem as well as giving a complete description of the n-step transition kernels of a closely related Markov chain. Our results can be used to explicitly compute the asymptotic variance in the Central Limit Theorem. They are also the basis for the quantitative analysis of any other statistic related to first-passage percolation in this model. In particular, knowledge of the higher-order transition kernels is the starting point for the computation of the distribution of the rungs which are part of the shortest path. The ladder model is worth studying because it is one of the very few situations where a complete explicit description of the finite-time behaviour of the first-passage percolation times can be given.

Outline of the chapter The structure of the chapter is as follows: In Section 7.2 we describe the model and state our results; Section 7.3, which contains the proofs, is divided in three subsections. The first is devoted to the Central Limit Theorem, the second presents some explicit evaluations of infinite sums which are needed in Section 7.3.3, where the main theorem about the transition kernels is proven. We conclude the chapter with a brief discussion of the techniques that we used to arrive at the results which are presented in the following.

Notation We use the notation $\delta_{p,q}$ for the Kronecker symbol and $\Theta_{p,q}$ as well as $\tilde{\Theta}_{p,q}$ for versions of the discretized Heaviside step function, precisely

$$\delta_{p,q} := \begin{cases} 1, & \text{if } p = q, \\ 0, & \text{else,} \end{cases} \qquad \Theta_{p,q} := \begin{cases} 1, & \text{if } p \geqslant q, \\ 0, & \text{else,} \end{cases} \qquad \tilde{\Theta}_{p,q} := \begin{cases} 1, & \text{if } p \leqslant q, \\ 0, & \text{else.} \end{cases}$$

The symbol $(k)^!$ stands for $k!(k+1)!$. We denote by $\mathbb{R}$ the real numbers and by $\mathbb{Z}$ the integers. A superscript $+$ $(-)$ indicates restriction to the positive (negative) elements of a set. The symbol γ stands for the Euler–Mascheroni constant, $\mathbb{P}$ denotes probability, and $\mathbb{E}$ expectation.

7.2. First-passage percolation on the ladder

In this chapter we further investigate a first-passage percolation model which has been considered before in Renlund (2010) and also in Schlemm (2009). We denote by $G_n = (V_n, E_n)$,

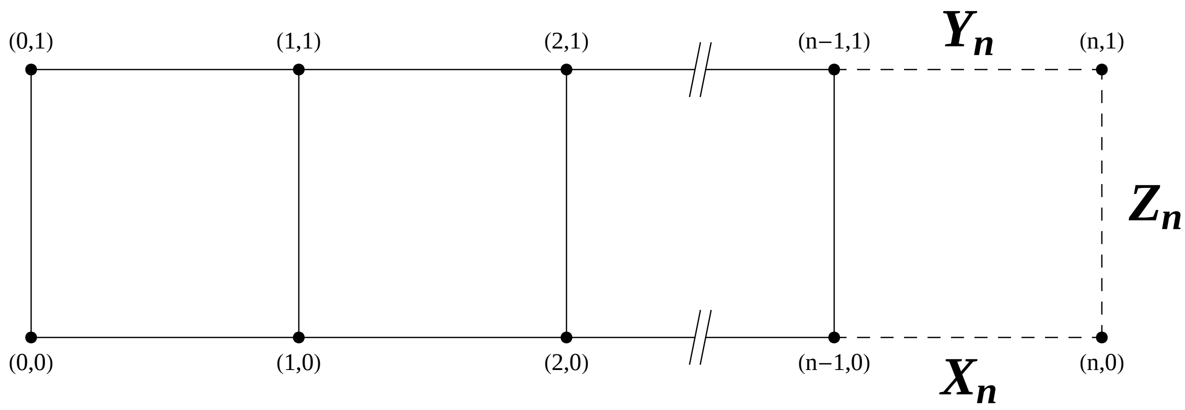


Figure 7.1.: The ladder graph G_n. The edge weights X_v, Y_v, Z_v are independent exponential random variables.

$n \in \mathbb{N}$, the graph with vertex set $V_n = \{0, \ldots, n\} \times \{0, 1\}$ and edges $E_n = \mathscr{X}_n \cup \mathscr{Y}_n \cup \mathscr{Z}_n \subset V_n^2$, where

$$\begin{aligned}
\mathscr{X}_n &= \{((i,0),(i+1,0)) : 0 \leqslant i < n\},\\
\mathscr{Y}_n &= \{((i,1),(i+1,1)) : 0 \leqslant i < n\},\\
\mathscr{Z}_n &= \{((i,0),(i,1)) : 0 \leqslant i \leqslant n\}.
\end{aligned}$$

The edge weights are independent exponentially distributed random variables which are labelled in the obvious way X_i, Y_i and Z_i, see also Fig. 7.1 for a graphical representation of the graph G_n. By time-scaling, it is no restriction to assume that the edge weights have mean one. Recently, Renlund (2011) investigated a model in which the mean of the vertical edges is different from the mean of the horizontal edges, but here we will restrict attention to the case that all edge weights have the same exponential distribution. We further denote by l_n the length of the shortest path from $(0,0)$ to $(n,0)$ in the graph G_n, by l'_n the length of the shortest path from $(0,0)$ to $(n,1)$, and by Δ_n the difference between the two, i. e. $\Delta_n = l'_n - l_n$. It has been shown in Schlemm (2009) and also, by a different method, in Renlund (2010) that $\lim_{n\to\infty} l_n/n$ almost surely exists and is equal to the constant $\chi = \frac{3}{2} - \frac{\mathrm{J}_1(2)}{2\mathrm{J}_2(2)}$, where J_ν are Bessel functions of the first kind, see Definition 7.4 or Abramowitz and Stegun (1992, Chap. 9) for a comprehensive review of Bessel functions. This constant is called the first-passage percolation rate for our model. The method employed in Schlemm (2009) to obtain this result built on Flaxman et al. (2006) and consisted in showing that there exists an ergodic $\mathbb{R} \times (\mathbb{R}^+)^{\times 3}$-valued Markov chain $\boldsymbol{M} = (\boldsymbol{M}_n)_{n \geqslant 0}$ with stationary distribution $\tilde{\pi}$ and a function $f : \mathbb{R} \times (\mathbb{R}^+)^{\times 3} \to \mathbb{R}$ such that

$$\chi = \mathbb{E} f(\boldsymbol{M}_\infty) = \int_{\mathbb{R}\times(\mathbb{R}^+)^{\times 3}} f(\boldsymbol{m}) \tilde{\pi}(\mathrm{d}\boldsymbol{m}).$$

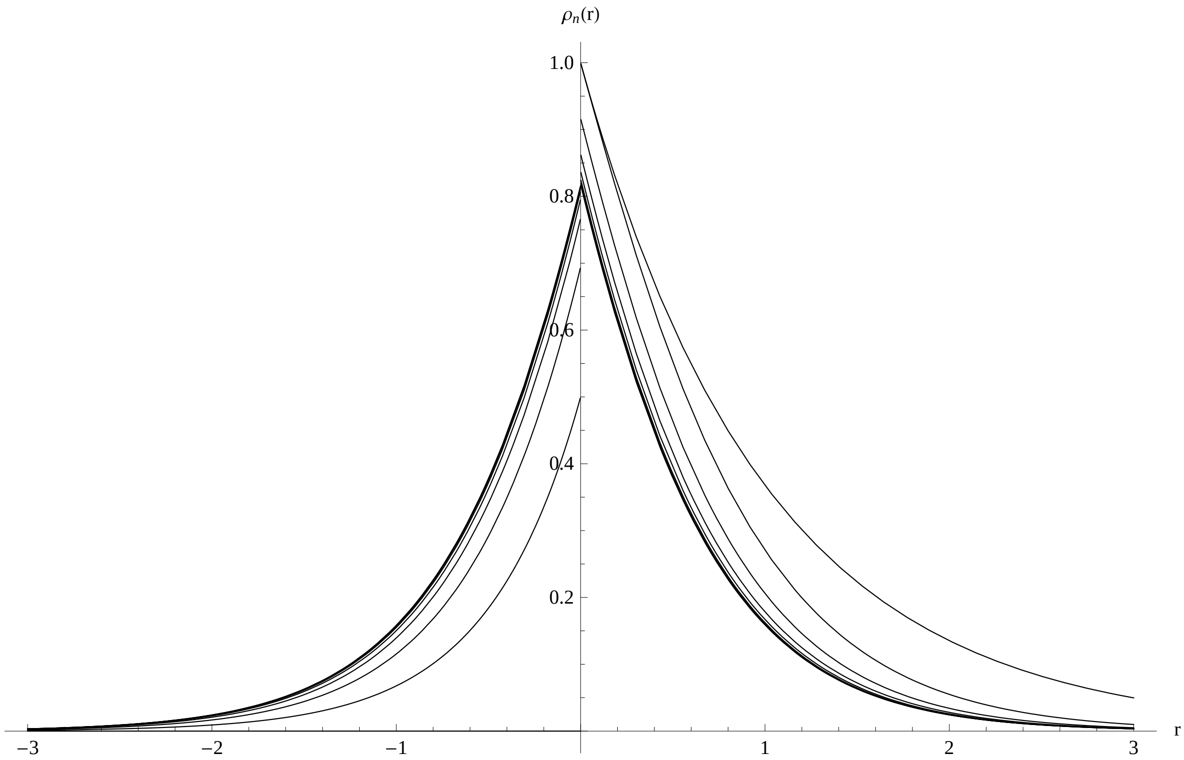


Figure 7.2.: Densities ρ_n of the random variables Δ_n for $n = 1, \dots, 8$, showing convergence towards the stationary distribution $\pi(\mathrm{d}r)$

Explicitly,

$$\boldsymbol{M}_n = (\Delta_n, X_{n+1}, Y_{n+1}, Z_{n+1}) \quad \text{and} \quad f : (r, x, y, z) \mapsto \min\{r + y + z, x\}.$$

Throughout we write $\boldsymbol{m} = (r, x, y, z)$ for some element of the state space $\mathbb{R} \times (\mathbb{R}^+)^{\times 3}$. Figure 7.2 displays the densities $\rho_n : \mathbb{R} \to \mathbb{R}^+$ of the random variables Δ_n for $n = 1, \dots, 8$; the picture confirms that the Markov chain $\Delta = (\Delta_n)_{n \geqslant 0}$ converges to a stationary distribution, which is given in Eq. (7.2.5).

In order to better understand the first-passage percolation problem on the ladder, it is important to know the higher-order transition kernels

$$\tilde{K}^n : \left(\mathbb{R} \times \left(\mathbb{R}^+\right)^{\times 3}\right) \times \left(\mathbb{R} \times \left(\mathbb{R}^+\right)^{\times 3}\right) \to \mathbb{R}^+$$

of the Markov chain $\boldsymbol{M}$. They completely determine the dynamics of the model and are defined by

$$\tilde{K}^n(\boldsymbol{m}', \boldsymbol{m})\mathrm{d}\boldsymbol{m} = \mathbb{P}\left(\boldsymbol{M}_n \in \mathrm{d}\boldsymbol{m} | \boldsymbol{M}_0 = \boldsymbol{m}'\right), \quad \boldsymbol{m}, \boldsymbol{m}' \in \mathbb{R} \times \left(\mathbb{R}^+\right)^{\times 3}. \tag{7.2.1}$$

The first result shows that it is sufficient to control the transition kernels $K^n : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$

of the Markov chain Δ, which are analogously defined by

$$K^n(r', r)\mathrm{d}r = \mathbb{P}\left(\Delta_n \in \mathrm{d}r | \Delta_0 = r'\right), \quad r, r' \in \mathbb{R}. \tag{7.2.2}$$

For convenience, we define $K^0(r', r) := \delta_{r'}(r)$, the Dirac distribution.

Proposition 7.1 *For any integer $n \geqslant 1$, denote by $\tilde{K}^n$ the n-step transition kernel of $\boldsymbol{M}$ defined in Eq. (7.2.1). Then*

$$\tilde{K}^n\left(\boldsymbol{m}', \boldsymbol{m}\right) = \mathrm{e}^{-(x+y+z)} K^{n-1}\left(\min\{r'+y', x'+z'\} - \min\{r'+y'+z', x'\}, r\right). \tag{7.2.3}$$

Moreover, the stationary distribution $\tilde{\pi}$ of $\boldsymbol{M}$ is given by

$$\tilde{\pi}(\mathrm{d}\boldsymbol{m}) = \mathrm{e}^{-(x+y+z)}\mathrm{d}^3(x, y, z)\pi(\mathrm{d}r), \tag{7.2.4}$$

where

$$\pi(\mathrm{d}r) = \frac{1}{2\,\mathrm{J}_2(2)}\mathrm{e}^{-\frac{3}{2}|r|}\,\mathrm{J}_1\left(2\mathrm{e}^{-\frac{1}{2}|r|}\right)\mathrm{d}r \tag{7.2.5}$$

is the stationary distribution of Δ.

Next, we state a Central Limit Theorem for first-passage percolation times on the ladder which has been implicit in Schlemm (2009) and which was the motivation for the present work. In Ahlberg (2009); Chatterjee and Dey (2009) a Central Limit Theorem has been obtained for first-passage times on fairly general one-dimensional graphs by a different method. The question of how to compute the asymptotic variance was, however, not addressed there. We denote by $\bar{f}$ the mean-corrected function $f - \chi$.

Theorem 7.2 (Central Limit Theorem for l_n) *For any integer $n \geqslant 0$, let l_n denote the first-passage time between $(0,0)$ and $(n,0)$ in the ladder graph G_n. Then there exists a positive constant σ^2 such that*

$$\frac{l_n - n\chi}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \sigma^2), \tag{7.2.6}$$

where $\mathcal{N}(0, \sigma^2)$ is a normally distributed random variable with mean zero and variance σ^2, and $\xrightarrow{d}$ denotes convergence in distribution. Moreover,

$$\sigma^2 = \int_{\mathbb{R}\times(\mathbb{R}^+)^{\times 3}} \bar{f}(\boldsymbol{m})^2\tilde{\pi}(\mathrm{d}\boldsymbol{m}) + 2\sum_{n=1}^{\infty}\int_{\mathbb{R}\times(\mathbb{R}^+)^{\times 3}} \bar{f}(\boldsymbol{m})P^n\bar{f}(\boldsymbol{m})\tilde{\pi}(\mathrm{d}\boldsymbol{m}), \tag{7.2.7}$$

where

$$P^n\bar{f}(\boldsymbol{m}) = \mathbb{E}\left[\bar{f}(\boldsymbol{M}_n)|\boldsymbol{M}_0 = \boldsymbol{m}\right] = \int_{\mathbb{R}\times(\mathbb{R}^+)^{\times 3}} \bar{f}(\boldsymbol{m}')\tilde{K}^n(\boldsymbol{m}, \boldsymbol{m}')\mathrm{d}\boldsymbol{m}'. \tag{7.2.8}$$

Equation (7.2.7) shows that in order to evaluate the asymptotic variance of the first-passage times, one must know the transition kernels $\tilde{K}^n$. In the next theorem we therefore explicitly describe the structure of the transition kernels K^n and, thus, by Proposition 7.1, the structure of $\tilde{K}^n$. To state the formulæ in a compact way, we define the five functions

$$\mathrm{S}^1(z) = \frac{z - 2\,\mathrm{J}_2(2\sqrt{z})}{z}, \tag{7.2.9a}$$

$$\mathrm{S}^2(z) = \frac{2(z-1) + 2\,\mathrm{J}_0(2\sqrt{z})}{z}, \tag{7.2.9b}$$

$$\mathrm{G}(z) = -\frac{z^3}{4}\left[5 - 4\gamma + 2\pi\frac{\mathrm{Y}_2(2\sqrt{z})}{\mathrm{J}_2(2\sqrt{z})} - 2\log z\right], \tag{7.2.9c}$$

$$\alpha(z) = \frac{z^4}{4\sqrt{z}\,\mathrm{J}_2(2\sqrt{z})\left[\sqrt{z}\,\mathrm{J}_0(2\sqrt{z}) + (z-1)\,\mathrm{J}_1(2\sqrt{z})\right]}, \tag{7.2.9d}$$

and

$$\mathrm{H}(z) = \frac{z^2}{2(1-z)}\left[3z - 3 + 2\pi\,\mathrm{Y}_0(2\sqrt{z}) + \mathrm{J}_0(2\sqrt{z})\,(5 - 4\gamma - 2\log z)\right]. \tag{7.2.9e}$$

The Bessel functions J_ν and Y_ν are defined in Definition 7.4 and treated comprehensively in Abramowitz and Stegun (1992).

Theorem 7.3 (Transition kernels) *The transition kernels K^n, $n \geqslant 1$, defined in Eq. (7.2.2), satisfy $K^n(r', r) = K^n(-r', -r)$. For $r \geqslant 0$, the values $K^n(r', r)$ are given by*

$$K^n(r',r) = \begin{cases} \sum_{p,q=0}^{n} a^n_{p,q} \mathrm{e}^{pr'-(q+2)r}, & r' \leqslant 0, \\ \frac{(-1)^{n-1}\mathrm{e}^{r'-(n+1)r}}{(n-1)!} + \sum_{p=0}^{n-2} \frac{(-1)^n r' \mathrm{e}^{-p(r'-r)-nr}}{(p)!(n-p-2)!} + \sum_{p,q=0}^{n} b^n_{p,q} \mathrm{e}^{-pr'-(q+2)r}, & 0 < r' \leqslant r, \\ \frac{(-1)^{n-1}\mathrm{e}^{-(n-1)r'-r}}{(n-1)!} + \sum_{p=0}^{n-2} \frac{(-1)^n r \mathrm{e}^{-p(r'-r)-nr}}{(p)!(n-p-2)!} + \sum_{p,q=0}^{n} c^n_{p,q} \mathrm{e}^{-pr'-(q+2)r}, & r' > r, \end{cases} \tag{7.2.10}$$

where the coefficients $a^n_{p,q}$, $b^n_{p,q}$, and $c^n_{p,q}$ are determined by their generating functions:

i) the generating functions $\mathrm{A}_{p,q}(z) = \sum_{n=1}^{\infty} a^n_{p,q} z^n$, $p, q \geqslant 0$, are given by

$$\mathrm{A}_{1,q}(z) = \frac{(-z)^q}{(q)!}\alpha(z), \tag{7.2.11a}$$

$$\mathrm{A}_{p,q}(z) = \frac{2(-z)^{p-1}}{(p)!}\,\mathrm{A}_{1,q}(z), \quad p \geqslant 2, \tag{7.2.11b}$$

$$\mathrm{A}_{0,q}(z) = \frac{\mathrm{S}^2(z)}{1-z}\,\mathrm{A}_{1,q}(z); \tag{7.2.11c}$$

ii) the generating functions $\mathrm{B}_{p,q}(z) = \sum_{n=1}^{\infty} b^n_{p,q} z^n$, $p, q \geqslant 0$, are given by

$$\mathrm{B}_{1,q}(z) = \frac{(-z)^q}{(q)!}\,\mathrm{G}(z) - \mathrm{A}_{1,q}(z), \tag{7.2.12a}$$

p \ q	0	1	2	3
0	$\frac{115}{96}$	$-\frac{17}{36}$	$\frac{1}{24}$	0
1	$\frac{11}{36}$	$-\frac{1}{36}$	$\frac{1}{12}$	$-\frac{1}{144}$
2	$-\frac{11}{108}$	$\frac{1}{12}$	$-\frac{1}{72}$	0
3	$\frac{1}{72}$	$\frac{1}{144}$	0	0
4	$-\frac{1}{1440}$	0	0	0

(a) Values of $a^4_{p,q}$

p \ q	0	1	2	3
0	$\frac{721}{432}$	$-\frac{13}{12}$	$\frac{1}{12}$	0
1	$-\frac{241}{540}$	$\frac{17}{48}$	$\frac{1}{36}$	0
2	$\frac{3}{16}$	$\frac{1}{36}$	0	0
3	$\frac{1}{216}$	0	0	0
4	0	0	0	0

(b) Values of $b^4_{p,q}$

p \ q	0	1	2	3
0	$\frac{721}{432}$	$-\frac{13}{12}$	$\frac{5}{18}$	0
1	$-\frac{241}{540}$	$\frac{17}{48}$	$\frac{1}{36}$	0
2	$-\frac{1}{144}$	$\frac{1}{36}$	0	0
3	$\frac{1}{216}$	0	0	0
4	0	0	0	0

(c) Values of $c^4_{p,q}$

Table 7.1.: Coefficients of the four-step transition kernel K^4, as given in Eq. (7.2.10)

$$\mathrm{B}_{p,q}(z) = \frac{2(-z)^{p-1}}{(p)!}\,\mathrm{B}_{1,q}(z) + \frac{(-z)^{p+q+2}}{(p)!(q)!}\sum_{k=2}^{p}\frac{2k+1}{k(k+1)}, \qquad p \geqslant 2, \tag{7.2.12b}$$

$$\mathrm{B}_{0,q}(z) = \frac{\mathrm{S}^2(z)}{1-z}\,\mathrm{B}_{1,q}(z) + \frac{(-z)^q}{(q)!}\,\mathrm{H}(z); \tag{7.2.12c}$$

iii) the generating functions $\mathrm{C}_{p,q}(z) = \sum_{n=1}^{\infty} c^n_{p,q} z^n$, $p, q \geqslant 0$, *are given by*

$$\mathrm{C}_{0,q}(z) = d_q z^{q+2} + \mathrm{B}_{0,q}(z) - \frac{(-z)^{q+2}}{(q)!}, \tag{7.2.13a}$$

$$\mathrm{C}_{1,q}(z) = \left[\mathrm{S}^1(z) + \frac{z\,\mathrm{S}^2(z)}{2(1-z)}\right]\mathrm{A}_{1,q}(z) - \frac{z}{2}\,\mathrm{C}_{0,q}(z), \tag{7.2.13b}$$

$$\mathrm{C}_{p,q}(z) = \frac{2(-z)^{p-1}}{(p)!}\,\mathrm{C}_{1,q}(z), \qquad p \geqslant 2, \tag{7.2.13c}$$

and the numbers d_q are determined by their generating function $D(z) = \sum_{q=0}^{\infty} d_q z^q$, which is given by

$$D(z) = \frac{1}{z}\left[\sqrt{z}\,\mathrm{J}_1(2\sqrt{z})\,(2\gamma + \log z) - \pi\sqrt{z}\,\mathrm{Y}_1(2\sqrt{z}) - 1\right]. \tag{7.2.14}$$

By the properties of generating functions, the coefficients $a^n_{p,q}$, $b^n_{p,q}$, and $c^n_{p,q}$ are determined by the derivatives of the functions $z \mapsto \mathrm{A}_{p,q}(z)$, $z \mapsto \mathrm{B}_{p,q}(z)$, and $z \mapsto \mathrm{C}_{p,q}(z)$, evaluated at zero. These derivatives are routinely calculated to any order with the help of computer algebra systems such as Mathematica®. Table 7.1 exemplifies Theorem 7.3 by reporting the non-zero values of the coefficients $a^n_{p,q}$, $b^n_{p,q}$, $c^n_{p,q}$ in the case $n = 4$. Using the results of Theorem 7.3, the expression (7.2.7) for the asymptotic variance σ^2 can be evaluated explicitly in terms of certain integrals of hypergeometric functions; the computations, however, are quite involved and the final result rather lengthy, so we decided not to include them here.

7.3. Proofs

7.3.1. Proofs of Lemma 7.1 and Theorem 7.2

In this section we present the proofs of the relation between the Markov chains $\boldsymbol{M}$ and Δ, and of the Central Limit Theorem.

Proof (of Proposition 7.1) Since

$$\begin{aligned}\Delta_n =& l'_n - l_n\\ =& \min\{l'_{n-1} + Y_n, l_{n-1} + X_n + Z_n\} - \min\{l'_{n-1} + Y_n + Z_n, l_{n-1} + X_n\}\\ =& \min\{\Delta_{n-1} + Y_{n-1}, X_{n-1} + Z_{n-1}\} - \min\{\Delta_{n-1} + Y_{n+1} + Z_{n-1}, X_{n-1}\},\end{aligned}$$

it follows at once that $\boldsymbol{M}_0$ being equal to some $m' = (r', x', y', z')$ implies that $\Delta_1 = \min\{r' + y', x' + z'\} - \min\{r' + y' + z', x'\}$; thus, the Markov property of Δ together with the independence of the edge weights implies that, for any integer $n > 1$, the conditional probability $\mathbb{P}\left(\boldsymbol{M}_n \in \mathrm{d}\boldsymbol{m} | \boldsymbol{M}_0 = \boldsymbol{m}'\right)$ is given by

$$\mathrm{e}^{-(x+y+z)}\mathrm{d}^3(x,y,z)\mathbb{P}\left(\Delta_n \in \mathrm{d}r | \Delta_1 = \min\{r' + y', x' + z'\} - \min\{r' + y' + z', x'\}\right).$$

The homogeneity of the Markov chain Δ then implies Eq. (7.2.10) because, for $n = 1$, we clearly have

$$\mathbb{P}\left(\boldsymbol{M}_1 \in \mathrm{d}\boldsymbol{m} | \boldsymbol{M}_0 = \boldsymbol{m}'\right) = \mathrm{e}^{-(x+y+z)}\delta_{\min\{r'+y', x'+z'\} - \min\{r'+y'+z', x'\}}(r)\mathrm{d}\boldsymbol{m}.$$

Equation (7.2.4) about the stationary distribution of $\boldsymbol{M}$ is a direct consequence of the fact that the edge weights X_{n+1}, Y_{n+1} ,Z_{n+1} are independent of Δ_n, and expression (7.2.5) was derived in Schlemm (2009, Proposition 5.5). □

Next, we prove the Central Limit Theorem and the formula for the asymptotic variance σ^2.

Proof (of Theorem 7.2) We apply the general result Chen (1999, Theorem 4.3) for functionals of ergodic Markov chains on general state spaces. We first note that the second moment of f with respect to $\tilde{\pi}$ is finite. In fact it can be shown that

$$\int_{\mathbb{R}\times(\mathbb{R}^+)^{\times 3}} f(\boldsymbol{m})^2\tilde{\pi}(\boldsymbol{m}) = \frac{2\,\mathrm{J}_1(2) - 3\,\mathrm{J}_0(2) + {}_2\mathrm{F}_3\left(\{1,1\},\{2,2,2\};-1\right) - 1}{\mathrm{J}_2(2)},$$

where the Bessel functions J_ν and the hypergeometric function ${}_2\mathrm{F}_3$ are introduced in detail below in Definitions 7.4 and 7.5. It then suffices to prove that the Markov chain $\boldsymbol{M}$ is uniformly ergodic, which is equivalent to showing that the Markov chain Δ is uniformly ergodic. We use the drift criterion Aldous, Lovász and Winkler (1997, Theorem B), which asserts that if there is a sufficiently strong drift towards the centre of the state space of a Markov chain, it is uniformly ergodic. Using the Lyapunov function $V(r) = 1 - \mathrm{e}^{-|r|}$ as well as the explicit formula

$$K(r', r) = \begin{cases} \mathrm{e}^{-|r|}, & \text{if } r' < r < 0 \vee r' > r > 0, \\ \mathrm{e}^{-|r'-2r|}, & \text{else,} \end{cases} \tag{7.3.1}$$

for the one step transition kernel of Δ (Schlemm, 2009, Proposition 5.1), we obtain

$$\begin{aligned} \psi(r) \coloneqq \mathbb{P}^1 V(r) \coloneqq \mathbb{E}_r \psi(\Delta_1) =& 1 - \mathbb{E}_r \mathrm{e}^{-|\Delta_1|} \\ =& 1 - \int_{\mathbb{R}} \mathrm{e}^{-|\rho|} K(r, \rho) \mathrm{d}\rho = \frac{1}{6}\left[3 - 2\mathrm{e}^{-|r|} + \mathrm{e}^{-2|r|}\right]. \end{aligned}$$

It is easy to check that

$$\psi(r) \leqslant V(r) - \frac{1}{10}, \quad |r| \geqslant 1, \quad \text{and} \quad \sup_{|r| \leqslant 1} \psi(r) \leqslant \sup_{r \in \mathbb{R}} \psi(r) = \frac{1}{2} < \infty.$$

Since the interval $[-1, 1]$ is compact and has positive invariant measure, it is a small set (Nummelin and Tuominen, 1982, Remark 2.7), and it follows that Δ is uniformly ergodic, which completes the proof. □

7.3.2. Summation formulæ

In this section we derive some summation formulæ which we will use in the proofs in Section 7.3.3. Some of them are well known, others can be checked with computer algebra systems such as Mathematica, a few (Formulæ 8, 9, 11 and 12) seem to be new. The sums will be evaluated explicitly in terms of Bessel and generalized hypergeometric functions, which we now define.

Definition 7.4 (Bessel function) Let λ be a real number in $\mathbb{R}\backslash\mathbb{Z}^-$. The *Bessel function of the first kind of order* λ, denoted by J_λ, is defined by the series representation

$$\mathrm{J}_\lambda(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{\Gamma(k)\,\Gamma(k+\lambda+1)} \left(\frac{x}{2}\right)^{2k+\lambda}. \tag{7.3.2}$$

For any integer ν, the *Bessel function of the second kind of order* ν, denoted by Y_ν, is defined as

$$\mathrm{Y}_\nu(x) = \lim_{\lambda \to \nu} \frac{\mathrm{J}_\lambda(x) \cos \lambda\pi - \mathrm{J}_{-\lambda}(x)}{\sin \lambda\pi}. \tag{7.3.3}$$

It is well known (Wolfram Research, Inc., 2010, Formula 03.01.17.0002.01) that Bessel functions satisfy the recurrence equation $\mathrm{J}_\nu(x) = \frac{2(\nu-1)}{x}\mathrm{J}_{\nu-1}(x) - \mathrm{J}_{\nu-2}(x)$, which we use without further mentioning to simplify various expressions.

Definition 7.5 (Generalized hypergeometric function) For two non-negative integers $p \leqslant q$ and sets of complex numbers $\boldsymbol{a} = \{a_1, \ldots, a_p\}$ and $\boldsymbol{b} = \{b_1, \ldots, b_q\}$, $b_j \notin \mathbb{Z}^-$, the *generalized hypergeometric function of order* (p, q) *with coefficients* $\boldsymbol{a}$, $\boldsymbol{b}$, denoted by ${}_p\mathrm{F}_q\left(\boldsymbol{a}, \boldsymbol{b}; \cdot\right)$, is defined by the series representation

$$
{}_p\mathrm{F}_q\left(\boldsymbol{a}, \boldsymbol{b}; x\right) = \sum_{k=0}^{\infty} \frac{(a_1)_k \cdots (a_p)_k}{(b_1)_k \cdots (b_q)_k} \frac{x^k}{k!}, \tag{7.3.4}
$$

where $(z)_k$ denotes the rising factorial defined by $(z)_k = \Gamma(z+k)/\Gamma(z)$.

In particular, we will encounter the regularized confluent hypergeometric functions ${}_0\tilde{\mathrm{F}}_1$, which are defined by ${}_0\tilde{\mathrm{F}}_1\left(\{\}, \{b\}; -z\right) = \frac{1}{\Gamma(b)} {}_0\mathrm{F}_1\left(\{\}, \{b\}; -z\right)$; in the next lemma we relate their derivative with respect to b to certain sums involving the harmonic numbers $\mathrm{H}_k \coloneqq \sum_{n=1}^{k} 1/n$.

Lemma 7.6 *Denote by* ${}_0\tilde{\mathrm{F}}_1$ *the regularized version of the hypergeometric function* ${}_0\mathrm{F}_1$*. It then holds that*

i) for every positive integer ν*,*

$$
\frac{\mathrm{d}}{\mathrm{d}b} {}_0\tilde{\mathrm{F}}_1\left(\{\}, \{b\}; -z\right)\bigg|_{b=\nu} = \gamma z^{\frac{1-\nu}{2}} \mathrm{J}_{\nu-1}\left(2\sqrt{z}\right) - \sum_{k=0}^{\infty} \frac{(-z)^k}{k!(k+\nu-1)!} \mathrm{H}_{k+\nu-1}, \tag{7.3.5a}
$$

ii) for every positive integer ν*,*

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}b} {}_0\tilde{\mathrm{F}}_1\left(\{\}, \{b\}; -z\right)\bigg|_{b=-\nu} =& (-1)^{\nu-1} \gamma z^{\frac{\nu+1}{2}} \mathrm{J}_{\nu+1}\left(2\sqrt{z}\right) - \sum_{k=0}^{\infty} \frac{(-z)^{k+\nu+1}}{k!(k+\nu+1)!} \mathrm{H}_k \\
&+ (-1)^{\nu} \sum_{k=0}^{\nu} \frac{z^k (\nu-k)!}{k!}.
\end{aligned} \tag{7.3.5b}
$$

Proof For part i), we differentiate the series representation (7.3.4) term by term. Using the definition of the Digamma function F as the logarithmic derivative of the Gamma function Γ as well as the relation $F(k) = -\gamma + \mathrm{H}_{k-1}$, $k \in \mathbb{N}$ (Wolfram Research, Inc., 2010, Formula 06.14.27.0003.01), we get

$$
\frac{\mathrm{d}}{\mathrm{d}b} \frac{1}{\Gamma(k+b)}\bigg|_{b=\nu} = -\frac{F(k+\nu)}{\Gamma(k+\nu)} = \frac{\gamma - \mathrm{H}_{k+\nu-1}}{(k+\nu-1)!},
$$

and, thus,

$$
\frac{\mathrm{d}}{\mathrm{d}b} {}_0\tilde{\mathrm{F}}_1\left(\{\}, \{b\}; -z\right)\bigg|_{b=\nu} = \gamma \sum_{k=0}^{\infty} \frac{(-z)^k}{k!(k+\nu-1)!} - \sum_{k=0}^{\infty} \frac{(-z)^k}{k!(k+\nu-1)!} \mathrm{H}_{k+\nu-1},
$$

which concludes the proof of the first part of the lemma. Part ii) is shown in a completely analogous way, using the relation $\lim_{\mu\to -m} F(\mu)/\Gamma(\mu) = (-1)^{m+1}m!$, for every non-negative integer m, which follows from the fact (Wolfram Research, Inc., 2010, Formula 06.05.04.0004.01) that the Gamma function has a simple pole at $-m$ with residue $(-1)^m/m!$.□

Formula 1

$$\sum_{k=1}^{\infty} \frac{(-z)^k}{k!(k+2)!} = \frac{2\,\mathrm{J}_2(2\sqrt{z}) - z}{2z} = -\frac{1}{2}\,\mathrm{S}^1(z).$$

Proof Immediate from Definition 7.4 of Bessel functions. □

Formula 2

$$\sum_{k=1}^{\infty} \frac{(-z)^k}{(k+1)!^2} = \frac{1 - z - \mathrm{J}_0(2\sqrt{z})}{z} = -\frac{1}{2}\,\mathrm{S}^2(z).$$

Proof This is also clear from Definition 7.4 of Bessel functions. □

Formula 3

$$\mathrm{S}^3(z) = \sum_{n=q+2}^{\infty} \frac{(-z)^{n-q}}{(n-q-2)!(n-q)^2} = 1 - \mathrm{J}_2\left(2\sqrt{z}\right) - \sqrt{z}\,\mathrm{J}_1\left(2\sqrt{z}\right).$$

Proof Shifting the index of summation n by $q+2$ we obtain

$$\begin{aligned}
\mathrm{S}^3(z) &= \sum_{n=0}^{\infty} \frac{(-z)^{n+2}}{n!(n+1)!(n+2)^2} = \sum_{n=0}^{\infty} (-z)^{n+2} \left[\frac{1}{(n+1)!(n+2)!} - \frac{1}{(n+2)!^2}\right] \\
&= \sum_{n=0}^{\infty} \frac{(-z)^{n+1}}{(n)!} + z - \sum_{n=0}^{\infty} \frac{(-z)^n}{n!^2} + 1 - z \\
&= 1 - \mathrm{J}_2\left(2\sqrt{z}\right) - \sqrt{z}\,\mathrm{J}_1\left(2\sqrt{z}\right)
\end{aligned}$$

by Definition 7.4 of Bessel functions. □

Formula 4

$$\Sigma^1(\zeta z) = \sum_{q=0}^{\infty} \sum_{k=0}^{q-2} \frac{(-\zeta z)^q}{(k)!(q-k-2)!(k+2)^2} = \frac{\mathrm{J}_1\left(2\sqrt{\zeta z}\right)}{\sqrt{\zeta z}}\,\mathrm{S}^3(\zeta z).$$

Proof By Fubini's theorem, we can interchange the order of summation and then shift the summation index q by $k+2$ to obtain

$$\Sigma^1(\zeta z) = \sum_{k=0}^{\infty} \frac{(-\zeta z)^{k+2}}{(k)!(k+2)^2} \sum_{q=0}^{\infty} \frac{(-\zeta z)^q}{(q)!}.$$

By Eq. (7.3.2), the second sum equals $\mathrm{J}_1\left(2\sqrt{\zeta z}\right)/\sqrt{\zeta z}$, and so the claim follows with Formula 3. □

Formula 5

$$\Sigma^2(\zeta z) = \sum_{q=0}^{\infty}\sum_{k=1}^{q-2}\frac{2(-\zeta z)^q}{k!(k+2)!(q-k-2)!} = \left[2\,\mathrm{J}_2\left(2\sqrt{\zeta z}\right) - \zeta z\right]\sqrt{\zeta z}\,\mathrm{J}_1\left(2\sqrt{\zeta z}\right).$$

Proof We can interchange the order of summation and shift the index q by $k+2$ to obtain that

$$\Sigma^2(\zeta z) = \sum_{k=1}^{\infty}\frac{2(-\zeta z)^{k+2}}{k!(k+2)!}\sum_{q=0}^{\infty}\frac{(-\zeta z)^q}{(q)!}.$$

The first factor equals $2\zeta z\,\mathrm{J}_2\left(2\sqrt{\zeta z}\right) - (\zeta z)^2$ and the second factor equals $\mathrm{J}_1\left(2\sqrt{\zeta z}\right)/\sqrt{\zeta z}$, both by Eq. (7.3.2), and so the claim follows. □

Formula 6

$$\mathrm{T}^1(\zeta z) = \sum_{q=0}^{\infty}\frac{(-\zeta z)^q}{(q)!} = \frac{\mathrm{J}_1\left(2\sqrt{\zeta z}\right)}{\sqrt{\zeta z}}.$$

Proof The proof is clear from Eq. (7.3.2). □

Formula 7

$$\mathrm{T}^2(\zeta z) = \sum_{q=0}^{\infty}\frac{(-\zeta z)^{q+1}q}{(q+1)!^2} = 1 - \mathrm{J}_0\left(2\sqrt{\zeta z}\right) - \sqrt{\zeta z}\,\mathrm{J}_1\left(2\sqrt{\zeta z}\right).$$

Proof This follows from the decomposition $\frac{q}{(q+1)!^2} = \frac{1}{(q)!} - \frac{1}{(q+1)!^2}$ and Eq. (7.3.2). □

Formula 8

$$\begin{aligned}\mathrm{U}^1(z) &= \sum_{k=1}^{\infty}\frac{(-z)^{k+2}}{k!(k+2)!}\sum_{l=2}^{k}\frac{2l+1}{l(l+1)}\\ &= \frac{3z^2}{4} - z - 2 - \pi z\,\mathrm{Y}_2\left(2\sqrt{z}\right) + \frac{\sqrt{z}}{2}\mathrm{J}_1\left(2\sqrt{z}\right)\left[4\gamma - 3 + 2\log z\right]\\ &\quad + \mathrm{J}_0\left(2\sqrt{z}\right)\left[1 + \frac{5z}{2} - 2\gamma z - z\log z\right].\end{aligned}$$

Proof First we note that

$$\sum_{l=2}^{k}\frac{2l+1}{l(l+1)} = \sum_{l=2}^{k}\left[\frac{1}{l} + \frac{1}{l+1}\right] = 2\,\mathrm{H}_k - \frac{5k+3}{2(k+1)}, \tag{7.3.6}$$

where H_k denotes the kth harmonic number. The first contribution to $\mathrm{U}^1(z)$ can therefore be computed as

$$\frac{1}{2}\sum_{k=1}^{\infty}\frac{(5k+3)(-z)^k}{k!(k+2)!(k+1)} = -\frac{1}{z}-\frac{3}{4}-\frac{1}{z}\sum_{k=0}^{\infty}\frac{(-z)^k}{(k)!}+\frac{5}{2}\sum_{k=0}^{\infty}\frac{(-z)^k}{k!(k+2)!}$$
$$=\frac{1}{4z^{3/2}}\left[4\,\mathrm{J}_1\left(2\sqrt{z}\right)+10\sqrt{z}\,\mathrm{J}_2\left(2\sqrt{z}\right)-3z^{3/2}-4\sqrt{z}\right].$$

The other contribution to $\mathrm{U}^1(z)$ is, with the help of Lemma 7.6, ii), obtained as

$$2\sum_{k=1}^{\infty}\frac{(-z)^k}{k!(k+2)!}\mathrm{H}_k = \frac{2}{z^2}\left[\gamma z\,\mathrm{J}_2\left(2\sqrt{z}\right)-z-1-\frac{\mathrm{d}}{\mathrm{d}b}{}_0\tilde{\mathrm{F}}_1\left(\{\},\{b\};-z\right)\Big|_{b=-1}\right].$$

By known properties of the regularized confluent hypergeometric function (see Wolfram Research, Inc., 2010, Formula 07.18.20.0015.01),

$$\frac{\mathrm{d}}{\mathrm{d}b}{}_0\tilde{\mathrm{F}}_1\left(\{\},\{b\};-z\right)\Big|_{b=-1} = \frac{\pi z\,\mathrm{Y}_2\left(2\sqrt{z}\right)-\sqrt{z}\,\mathrm{J}_1\left(2\sqrt{z}\right)\left[2+\log z\right]+\mathrm{J}_0\left(2\sqrt{z}\right)\left[z\log z-1\right]}{2},$$

and the result follows upon combining the last four displayed equations. □

Formula 9
$$\mathrm{Y}^1(\zeta z) = \sum_{q=0}^{\infty}\sum_{k=1}^{q-2}\frac{(-\zeta z)^q}{(q-k-2)!k!(k+2)!}\sum_{l=2}^{k}\frac{2l+1}{l(l+1)} = \frac{\mathrm{J}_1\left(2\sqrt{\zeta z}\right)}{\sqrt{\zeta z}}\,\mathrm{U}^1(\zeta z).$$

Proof Interchanging the order of the first two summations and shifting the index q by $k+2$, we find that

$$\mathrm{Y}^1(\zeta z) = \sum_{k=1}^{\infty}\frac{(-\zeta z)^{k+2}}{k!(k+2)!}\sum_{q=0}^{\infty}\frac{(-\zeta z)^q}{(q)!}\sum_{l=2}^{k}\frac{2l+1}{l(l+1)}.$$

As before, the sum in the middle equals $\mathrm{J}_1\left(2\sqrt{\zeta z}\right)/\sqrt{\zeta z}$, and so the claim follows with Formula 8. □

Formula 10
$$\mathrm{S}^4(z) = \sum_{n=q+2}^{\infty}\frac{(-z)^{n-q}}{(n-q-2)!(n-q-1)^2} = z^2{}_2\mathrm{F}_3\left(\{1,1\},\{2,2,2\};-z\right).$$

Proof After shifting the index n by $q+2$, the proof follows immediately from the definition of the hypergeometric function, given in Eq. (7.3.4). □

Formula 11

$$\begin{aligned}\mathrm{U}^2(z) =& \sum_{k=1}^{\infty} \frac{(-z)^{k+2}}{(k+1)!^2} \sum_{l=2}^{k} \frac{2l+1}{l(l+1)}\\ =&\frac{z}{2}\left[3z-5+2\pi\,\mathrm{Y}_0\left(2\sqrt{z}\right)-2z{}_2\mathrm{F}_3\left(\{1,1\},\{2,2,2\},-z\right)+\mathrm{J}_0\left(2\sqrt{z}\right)\left[5-4\gamma-2\log z\right]\right].\end{aligned}$$

Proof The proof is analogous to that of Formula 8. Using Eq. (7.3.6), the first contribution to $\mathrm{U}^2(z)$ is

$$\begin{aligned}\frac{1}{2}\sum_{k=1}^{\infty}\frac{(5k+3)(-z)^{k+2}}{(k+1)!^2(k+1)} =&\frac{5-3z}{2z}-\frac{5}{2z}\sum_{k=0}^{\infty}\frac{(-z)^k}{k!^2}-\sum_{k=0}^{\infty}\frac{(-z)^k}{(k+1)!^2(k+1)^2}\\ =&\frac{1}{2z}\left[5-3z-5\,\mathrm{J}_0\left(2\sqrt{z}\right)-2z{}_2\mathrm{F}_3\left(\{1,1\},\{2,2,2\};-z\right)\right], \qquad (7.3.7)\end{aligned}$$

where we used Eqs. (7.3.2) and (7.3.4). For the remaining part, we first use Lemma 7.6, i) and the identity $\mathrm{J}_1\left(2\sqrt{z}\right)/\sqrt{z}-\mathrm{J}_2\left(2\sqrt{z}\right)=\mathrm{J}_0\left(2\sqrt{z}\right)$ to compute

$$\begin{aligned}&\sum_{k=1}^{\infty}\frac{(-z)^k}{(k+1)!^2}\mathrm{H}_{k+2}\\ =&\sum_{k=1}^{\infty}\frac{(-z)^k}{(k+1)!(k+2)!}\mathrm{H}_{k+2}+\sum_{k=1}^{\infty}\frac{(-z)^k}{k!(k+2)!}\mathrm{H}_{k+2}\\ =&-\frac{3}{2}+\frac{1}{z}-\frac{1}{z}\left[\sum_{k=0}^{\infty}\frac{(-z)^k}{(k)!}\mathrm{H}_{k+1}-z\sum_{k=0}^{\infty}\frac{(-z)^k}{k!(k+2)!}\mathrm{H}_{k+2}\right]\\ =&-\frac{3}{2}+\frac{1}{z}-\frac{\gamma}{z}\mathrm{J}_0\left(2\sqrt{z}\right)+\frac{1}{z}\left[\frac{\mathrm{d}}{\mathrm{d}b}{}_0\tilde{\mathrm{F}}_1\left(\{\},\{b\};-z\right)\Big|_{b=2}-z\,\frac{\mathrm{d}}{\mathrm{d}b}{}_0\tilde{\mathrm{F}}_1\left(\{\},\{b\};-z\right)\Big|_{b=3}\right].\end{aligned}$$

It then follows from the explicit characterization of $\frac{\mathrm{d}}{\mathrm{d}b}{}_0\tilde{\mathrm{F}}_1\left(\{\},\{b\};-z\right)\Big|_{b=-\nu}$, $\nu\in\mathbb{N}$ (Wolfram Research, Inc., 2010, Formula 07.18.20.0013.01), that

$$\sum_{k=1}^{\infty}\frac{(-z)^k}{(k+1)!^2}\mathrm{H}_{k+2}=\frac{\sqrt{z}\left[2-3z+\pi\,\mathrm{Y}_0\left(2\sqrt{z}\right)\right]-2\,\mathrm{J}_1\left(2\sqrt{z}\right)-\sqrt{z}\,\mathrm{J}_0\left(2\sqrt{z}\right)\left[2\gamma+\log z\right]}{2z^{3/2}}.$$

Using this, we obtain

$$\begin{aligned}\sum_{k=1}^{\infty}\frac{(-z)^k}{(k+1)!^2}\mathrm{H}_k =&\sum_{k=1}^{\infty}\frac{(-z)^k}{(k+1)!^2}\mathrm{H}_{k+2}-\sum_{k=1}^{\infty}\frac{(-z)^k}{(k+1)!^2(k+2)}-\sum_{k=1}^{\infty}\frac{(-z)^k}{(k+1)!^2(k+1)}\\ =&\frac{\sqrt{z}\left[2-3z+\pi\,\mathrm{Y}_0\left(2\sqrt{z}\right)\right]-2\,\mathrm{J}_1\left(2\sqrt{z}\right)-\sqrt{z}\,\mathrm{J}_0\left(2\sqrt{z}\right)\left[2\gamma+\log z\right]}{4z^{3/2}}\\ &+\left[\frac{1}{4}-\frac{1}{2z}+\frac{\mathrm{J}_1\left(2\sqrt{z}\right)}{2z^{3/2}}\right]+\frac{1}{2}-\frac{1}{2}{}_2\mathrm{F}_3\left(\{1,1\},\{2,2,2\};-z\right). \qquad (7.3.8)\end{aligned}$$

Combining Eqs. (7.3.7) and (7.3.8) completes the proof. □

Formula 12

$$\Upsilon^2(\zeta z) = \sum_{q=0}^{\infty}\sum_{k=1}^{q-1} \frac{(-\zeta z)^q}{(q-k-1)!(k+1)!^2} \sum_{l=2}^{k} \frac{2l+1}{l(l+1)} = -\frac{\mathrm{J}_1\left(2\sqrt{\zeta z}\right)}{(\zeta z)^{3/2}}\, \mathrm{U}^2(\zeta z).$$

Proof Interchanging the order of the first two summations and shifting the index q by $k+1$, we find that

$$\Upsilon^2(\zeta z) = \sum_{k=1}^{\infty} \frac{(-\zeta z)^{k+2}}{(k+1)!^2} \sum_{q=0}^{\infty} \frac{(-\zeta z)^{q-1}}{(q)!} \sum_{l=2}^{k} \frac{2l+1}{l(l+1)}.$$

By Eq. (7.3.2), the middle sum is equal to $-\mathrm{J}_1\left(2\sqrt{\zeta z}\right)/(\zeta z)^{3/2}$, and so the result follows readily from Formula 11. □

Formula 13

$$\Sigma^3(\zeta z) = \sum_{q=0}^{\infty}\sum_{k=1}^{q-1} \frac{2(-\zeta z)^q}{(q-k-1)!(k+1)!^2} = \frac{2\left[\zeta z - 1 + \mathrm{J}_0\left(2\sqrt{\zeta z}\right)\right]\mathrm{J}_1\left(2\sqrt{\zeta z}\right)}{\sqrt{\zeta z}}.$$

Proof The proof is the same as that of Formula 5, and so we omit it. □

Formula 14

$$\Sigma^4(\zeta z) = \sum_{q=0}^{\infty}\sum_{k=0}^{q-1} \frac{(-\zeta z)^q}{(k)!(q-k-1)!(k+1)^2} = -\sqrt{\zeta z}\,\mathrm{J}_1\left(2\sqrt{\zeta z}\right) {}_2\mathrm{F}_3\left(\{1,1\},\{2,2,2\};-\zeta z\right).$$

Proof Using Fubini's theorem, we obtain

$$\Sigma^4(\zeta z) = \sum_{k=0}^{\infty} \frac{(-\zeta z)_k}{(k)!(k+1)^2} \sum_{q=k+1}^{\infty} \frac{(-\zeta z)^{q-k}}{(q-k-1)!}.$$

The first factor is equal to ${}_2\mathrm{F}_3\left(\{1,1\},\{2,2,2\};-\zeta z\right)$ by definition (7.3.4), and the second factor equals $-\sqrt{\zeta z}\,\mathrm{J}_1\left(2\sqrt{\zeta z}\right)$ by Eq. (7.3.2). □

Having evaluated these sums we now turn to relations between them which we will also need, and which are proved by writing out the relevant expressions and straightforward computations. The first one only involves the five functions from Eqs. (7.2.9) which appear in the statement of Theorem 7.3.

Lemma 7.7 *For almost every complex number z with respect to the Lebesgue measure on the complex plane,*

$$\left[\frac{\mathrm{S}^2(z)}{2(1-z)} + \frac{1}{z}\right]\mathrm{G}(z) + \left[\frac{\mathrm{S}^1(z)}{z} + \frac{\mathrm{S}^2(z)}{1-z} + \frac{1}{z}\right]\alpha(z) + \frac{\mathrm{H}(z)}{2} + \frac{3z^2}{4} = 0. \tag{7.3.9}$$

Lemma 7.8 *For almost every complex number z with respect to the Lebesgue measure on the complex plane,*

$$\mathrm{U}^1(z) + \mathrm{S}^3(z) + \left[\frac{\mathrm{S}^1(z) - 1}{z}\right] \mathrm{G}(z) = \frac{3}{4}z^2 - z - 1. \tag{7.3.10}$$

7.3.3. Proof of Theorem 7.3

In this section we prove the main result, Theorem 7.3. First, however, we take a closer look at the coefficients $a^n_{p,q}$, $b^n_{p,q}$, and $c^n_{p,q}$ which are defined implicitly through the generating functions (7.2.11) to (7.2.13).

Lemma 7.9 *For any integers $p, q \geqslant 0$, the following hold:*

$$a^1_{p,q} = \delta_{p,1}\delta_{q,0}, \quad b^1_{p,q} = 0, \quad c^1_{p,q} = 0. \tag{7.3.11}$$

Proof This follows from evaluating the derivatives of the generating functions $\mathrm{A}_{p,q}$, $\mathrm{B}_{p,q}$, and $\mathrm{C}_{p,q}$ at zero. □

Next we derive some useful relations between the coefficients $a^n_{p,q}$, $b^n_{p,q}$, and $c^n_{p,q}$. These will be the main ingredient in our inductive proof of Theorem 7.3. The general strategy in proving the equality of two sequences $(s_n)_{n\geqslant 1}$ and $(\tilde{s}_n)_{n\geqslant 1}$ will be to compute their generating functions $\sum_{n\geqslant 1} s_n z^n$ and $\sum_{n\geqslant 1} \tilde{s}_n z^n$, and to show that they coincide for every z. The validity of this approach follows from the well-known bijection between sequences of real numbers and generating functions, see, e. g., Wilf (2006) for an introductory treatment. We will constantly be making use of the convolution property of generating functions. By this we mean the simple fact that if $(s_n)_{n\geqslant 1}$ is a real sequence with generating function $S(z)$, and $(t_n)_{n\geqslant 1}$ is another such sequence with generating function $T(z)$, then the sequence of partial sums $(\sum_{\nu=1}^{n-1} t_{n-\nu} s_\nu)_{n\geqslant 1}$ has generating function $S(z)T(z)$. We also encounter generating functions of sequences indexed by q instead of n. In this case we denote the formal variable by ζ instead of z, and sums are understood to be indexed from zero to infinity.

Lemma 7.10 *For all integers $n \geqslant 1$ and $q \geqslant 0$, the coefficients defined by the generating functions given in Theorem 7.3 satisfy the relation*

$$\sum_{k=0}^{q-2} \frac{b^n_{k,q-k-2}}{k+2} - \sum_{k=0}^{q-2} \frac{c^n_{k,q-k-2}}{k+2} = \delta_{q,n}\left[\frac{(-1)^n(n-1)}{n!^2} - \sum_{k=0}^{n-2} \frac{(-1)^n}{(k)!(n-k-2)!(k+2)^2}\right]. \tag{7.3.12}$$

Proof These equations are true for all $n \geqslant 1$ if and only if the corresponding generating functions coincide. Multiplying both sides by z^n, summing over n, and using the recursive definitions of the generating functions as well as the fact that, by Formulæ 1 and 2,

$$\sum_{k=1}^{\infty} \frac{(-z)^k}{k!(k+2)!} = -\frac{1}{2}\,\mathrm{S}^1(z), \qquad \sum_{k=1}^{\infty} \frac{(-z)^k}{(k+1)!^2} = -\frac{1}{2}\,\mathrm{S}^2(z),$$

we find that the claim of the lemma is equivalent to

$$\begin{aligned} 0 = &\sum_{k=1}^{q-2} \frac{(-z)^q}{k!(k+2)!(q-k-2)!} \sum_{l=1}^{k} \frac{2l+1}{l(l+1)} + \frac{1}{2}\Theta_{q,2}\left[\frac{(-z)^q}{(q-2)!} - z^q d_{q-2}\right] \\ &+ \frac{z^q}{2}\sum_{k=1}^{q-2} \frac{2(-1)^{k-1}}{k!(k+2)!} d_{q-k-2} - \Theta_{q,1}\left[\frac{(-z)^q(q-1)}{q!^2} - \sum_{k=0}^{q-2} \frac{(-z)^q}{(k)!(q-k-2)!(k+2)^2}\right] \\ &+ \frac{5}{4}z^3 \sum_{k=1}^{q-2} \frac{2(-z)^{q-3}}{k!(k+2)!(q-k-2)!}, \end{aligned}$$

where we have used Lemma 7.7 to simplify the coefficient of the last sum. To show this equality for all non-negative integers q, we compare the q-generating functions and must then show that

$$0 = \mathrm{Y}^1(\zeta z) + \frac{(\zeta z)^2}{2}\left[\mathrm{T}^1(\zeta z) + \left((\mathrm{S}^1(\zeta z) - 1\right) D(\zeta z)\right] - \mathrm{T}^2(\zeta z) + \Sigma^1(\zeta z) + \frac{5}{4}\Sigma^2(\zeta z), \quad (7.3.13)$$

where closed-form expressions for

$$\begin{aligned} \Sigma^1(\zeta z) &:= \sum_{q=0}^{\infty}\sum_{k=0}^{q-2} \frac{(-\zeta z)^q}{(k)!(q-k-2)!(k+2)^2}, \\ \Sigma^2(\zeta z) &:= \sum_{q=0}^{\infty}\sum_{k=1}^{q-2} \frac{2(-\zeta z)^q}{k!(k+2)!(q-k-2)!}, \\ \mathrm{T}^1(\zeta z) &:= \sum_{q=0}^{\infty} \frac{(-\zeta z)^q}{(q)!}, \\ \mathrm{T}^2(\zeta z) &:= \sum_{q=0}^{\infty} \frac{(-\zeta z)^{q+1} q}{(q+1)!^2}, \end{aligned}$$

and

$$\mathrm{Y}^1(\zeta z) := \sum_{q=0}^{\infty}\sum_{k=1}^{q-2} \frac{(-\zeta z)^q}{(q-k-2)!k!(k+2)!} \sum_{l=2}^{k} \frac{2l+1}{l(l+1)}$$

are derived in Formulæ 4 to 7 and 9. Using these closed-form formulæ, Eq. (7.3.13) is seen to be identically true by simple algebra. □

Lemma 7.11 *For all integers $p, q \geqslant 0$, the sequences of coefficients defined by the generating func-*

tions given in Theorem 7.3 satisfy the recursion

$$
\begin{aligned}
a_{p,q}^{n+1} =&\delta_{p,0}\sum_{k=0}^{n}\frac{a_{k,q}^{n}}{k+1}-\Theta_{p,1}\frac{a_{p-1,q}^{n}}{p(p+1)}+\delta_{p,1}\Bigg[\tilde{\Theta}_{q,n-2}\frac{(-1)^n}{(n-q-2)!(q)!(n-q)^2}\\
&+\delta_{q,n-1}\frac{(-1)^{n-1}}{(n-1)!}-\delta_{q,n}\left[\sum_{k=0}^{n-2}\frac{(-1)^n}{(k)!(n-k-2)!(k+2)^2}+\frac{(-1)^{n-1}n}{(n-1)!(n+1)!}\right]\\
&+\sum_{k=0}^{n}\frac{b_{k,q}^{n}}{k+2}-\sum_{k=0}^{q-2}\frac{b_{k,q-k-2}^{n}}{k+2}+\sum_{k=0}^{q-2}\frac{c_{k,q-k-2}^{n}}{k+2}\Bigg],\quad n\geqslant 1.
\end{aligned}
\tag{7.3.14}
$$

Proof Applying Lemma 7.10 and computing the generating functions of both sides of the asserted equality, we find that the claim of the lemma is equivalent to

$$
\begin{aligned}
\frac{\mathrm{A}_{p,q}(z)}{z} =&\delta_{p,0}\frac{\mathrm{S}^2(z)}{z(1-z)}\mathrm{A}_{1,q}(z)-\Theta_{p,1}\frac{\mathrm{A}_{p-1,q}(z)}{p(p+1)}\\
&+\delta_{p,1}\frac{(-z)^q}{(q)!}\left[\left[\frac{\mathrm{S}^2(z)}{2(1-z)}+\frac{\mathrm{S}^1(z)}{z}\right][\mathrm{G}(z)-\alpha(z)]+\frac{\mathrm{H}(z)}{2}+\mathrm{S}^3(z)+\mathrm{U}^1(z)+z+1\right],
\end{aligned}
\tag{7.3.15}
$$

where explicit expressions for

$$
\mathrm{S}^3(z) \coloneqq \sum_{n=q+2}^{\infty}\frac{(-z)^{n-q}}{(n-q-2)!(n-q)^2}\quad\text{and}\quad \mathrm{U}^1(z) \coloneqq \sum_{k=1}^{\infty}\frac{(-z)^{k+2}}{k!(k+2)!}\sum_{l=2}^{k}\frac{2l+1}{l(l+1)}
$$

are derived in Formulæ 3 and 8. For $p=0$ and $p>1$, Eq. (7.3.15) follows immediately from the defining equations (7.2.11b) and (7.2.11c). For $p=1$, the claim follows from combining Lemmata 7.7 and 7.8. □

Lemma 7.12 *For all integers $p,q\geqslant 0$, the sequences of coefficients defined by the generating functions given in Theorem 7.3 satisfy the recursion*

$$
\begin{aligned}
b_{p,q}^{n+1} =&\delta_{p,1}\sum_{k=0}^{n}\frac{a_{k,q}^{n}}{k+2}+\delta_{p,0}\Bigg[\delta_{q,n-1}\frac{(-1)^{n-1}}{(n-1)!}+\tilde{\Theta}_{q.n-2}\frac{(-1)^n}{(n-q-2)!(q)!(n-q-1)^2}\\
&+\sum_{k=0}^{n}\frac{b_{k,q}^{n}}{k+1}\Bigg]-\Theta_{p,1}\left[\frac{b_{p-1,q}^{n}}{p(p+1)}+\delta_{q,n-p-1}\frac{(-1)^n(2p+1)}{(p)!(n-p-1)!p(p+1)}\right],
\end{aligned}
\tag{7.3.16}
$$

for all integers $n\geqslant 1$.

Proof We proceed as in the proof of Lemma 7.11 and show the equality for every n by showing equality of the corresponding generating functions. We find that the claim is equivalent to

$$\begin{aligned}\frac{\mathrm{B}_{p,q}(z)}{z} =&\delta_{p,1}\left[\frac{\mathrm{S}^1(z)}{z}+\frac{\mathrm{S}^2(z)}{2(1-z)}\right]\mathrm{A}_{1,q}(z)+\delta_{p,0}\frac{(-z)^{q+1}}{(q)^!}\Bigg[\mathrm{S}^4(z)+\mathrm{U}^2(z)+z\\ &+\frac{\mathrm{S}^2(z)}{z(1-z)}\left[\mathrm{G}(z)-\alpha(z)\right]+\mathrm{H}(z)\Bigg]\\ &-\Theta_{p.1}\left[\frac{\mathrm{B}_{p-1,q}(z)}{p(p+1)}+\frac{(-z)^{p+q+1}(2p+1)}{(p)^!(q)^!p(p+1)}\right], \end{aligned} \tag{7.3.17}$$

where the functions

$$\mathrm{S}^4(z) := \sum_{n=q+2}^{\infty}\frac{(-z)^{n-q}}{(n-q-2)^!(n-q-1)^2} \quad \text{and} \quad \mathrm{U}^2(z) := \sum_{k=1}^{\infty}\frac{(-z)^{k+2}}{(k+1)!^2}\sum_{l=2}^{k}\frac{2l+1}{l(l+1)}$$

are evaluated in Formulæ 10 and 11. For $p=0$, Eq. (7.3.17) follows from the observation that

$$\mathrm{H}(z) = z\left[\mathrm{S}^4(z)+\mathrm{U}^2(z)+z+\mathrm{H}(z)\right].$$

Next we observe that Eq. (7.2.12b) implies that

$$\begin{aligned}\mathrm{B}_{p,q}(z)+\frac{z\,\mathrm{B}_{p-1,q}(z)}{p(p+1)} &=\frac{(-z)^{p+q+2}}{(p)^!(q)^!}\sum_{k=2}^{p}\frac{2k+1}{k(k+1)}-\frac{(-z)^{p+q+2}}{(p)^!(q)^!}\sum_{k=2}^{p-1}\frac{2k+1}{k(k+1)}\\ &=\frac{(-z)^{p+q+2}}{(p)^!(q)^!}\frac{2p+1}{p(p+1)},\end{aligned}$$

and, thus, Eq. (7.3.17) also holds for $p>1$. Finally, for $p=1$, we need to show that

$$\frac{\mathrm{B}_{1,q}(z)}{z} = \left[\frac{\mathrm{S}^1(z)}{z}+\frac{\mathrm{S}^2(z)}{2(1-z)}\right]\mathrm{A}_{1,q}(z)-\frac{\mathrm{B}_{0,q}(z)}{2}-\frac{3(-z)^{q+1}}{4(q)^!},$$

which, after using the defining equations (7.2.11) and (7.2.12) several times, amounts to showing that

$$\left[\frac{\mathrm{S}^2(z)}{2(1-z)}+\frac{1}{z}\right]\mathrm{G}(z)+\left[\frac{\mathrm{S}^1(z)}{z}+\frac{\mathrm{S}^2(z)}{1-z}+\frac{1}{z}\right]\alpha(z)+\frac{\mathrm{H}(z)}{2}+\frac{3z^2}{4}=0,$$

which is exactly what Lemma 7.7 asserts. □

Lemma 7.13 *For all integers $p,q\geqslant 0$, the sequences of coefficients defined by the generating func-*

tions given in Theorem 7.3 satisfy the recursion

$$
\begin{aligned}
c^{n+1}_{p,q} =&\delta_{p,1}\sum_{k=0}^{n}\frac{a^n_{k,q}}{k+2}-\Theta_{p,1}\frac{c^n_{p-1,q}}{p(p+1)}+\delta_{p,0}\Bigg[\tilde{\Theta}_{q,n-2}\frac{(-1)^n}{(n-q-2)^!(q)^!(n-q-1)^2}\\
&+\delta_{q,n-1}\left(\sum_{k=0}^{n-2}\frac{(-1)^n}{(k)^!(n-k-2)^!(k+1)^2}-\frac{(-1)^{n-1}}{n!^2}\right)\\
&+\sum_{k=0}^{n}\frac{b^n_{k,q}}{k+1}-\sum_{k=0}^{q-1}\frac{b^n_{k,q-k-1}}{k+1}+\sum_{k=0}^{q-1}\frac{c^n_{k,q-k-1}}{k+1}\Bigg],\quad n\geqslant 1.
\end{aligned}
\tag{7.3.18}
$$

Proof The proof follows along the same lines as the previous proofs. Equating the generating functions of both sides and using Lemma 7.7, we need to show that

$$
\begin{aligned}
\frac{\mathrm{C}_{p,q}(z)}{z} =&\delta_{p,1}\left[\frac{\mathrm{S}^1(z)}{z}+\frac{\mathrm{S}^2(z)}{2(1-z)}\right]\mathrm{A}_{1,q}(z)-\Theta_{p,1}\frac{\mathrm{C}_{p-1,q}(z)}{p(p+1)}\\
&+\delta_{p,0}\Bigg[\frac{(-z)^q}{(q)^!}\left(\frac{\mathrm{S}^2(z)}{z(1-z)}(\mathrm{G}(z)-\alpha(z))+\mathrm{H}(z)+\mathrm{S}^4(z)+\mathrm{U}^2(z)+\frac{z}{q+1}\right)\\
&+\frac{5z}{4}\sum_{k=1}^{q-1}\frac{2(-z)^q}{(q-k-1)^!(k+1)!^2}-\sum_{k=0}^{q-1}\frac{(-z)^{q+1}}{(k)^!(q-k-1)^!(k+1)^2}\\
&-\sum_{k=1}^{q-1}\frac{(-z)^{q+1}}{(q-k-1)^!(k+1)!^2}\sum_{l=2}^{k}\frac{2l+1}{l(l+1)}-\frac{z^{q+1}}{2}\sum_{k=1}^{q-1}\frac{2(-1)^{k-1}}{(k+1)!^2}d_{q-k-1}\\
&+\Theta_{q,1}\left(z^{q+1}d_{q-1}-\frac{(-z)^{q+1}}{(q-1)^!}\right)\Bigg].
\end{aligned}
\tag{7.3.19}
$$

For $p\geqslant 1$, Eq. (7.3.19) is immediately clear from Eqs. (7.2.13b) and (7.2.13c). For $p=0$, we show that the q-generating functions coincide. Doing this, we find, after some algebra, that Eq. (7.3.18) is equivalent to

$$
\begin{aligned}
0 =& \left[\frac{z-1}{z}\mathrm{H}(z)+\mathrm{S}^4(z)+z+\mathrm{U}^2(z)-\zeta z^2\right]\mathrm{T}^1(\zeta z)+z\left[1-\frac{\mathrm{S}^2(z)}{2}\right]\\
&+\left[\zeta z^2\left(1-\frac{\mathrm{S}^2(\zeta z)}{2}\right)-z\right]D(\zeta z)+\frac{5z}{4}\Sigma^3(\zeta z)+z\left[\Sigma^4(\zeta z)+\Upsilon^2(\zeta z)\right],
\end{aligned}
\tag{7.3.20}
$$

where the functions

$$
\begin{aligned}
\Sigma^3(\zeta z) &= \sum_{q=0}^{\infty}\sum_{k=1}^{q-1}\frac{2(-\zeta z)^q}{(q-k-1)^!(k+1)!^2},\\
\Sigma^4(\zeta z) &= \sum_{q=0}^{\infty}\sum_{k=0}^{q-1}\frac{(-\zeta z)^q}{(k)^!(q-k-1)^!(k+1)^2},
\end{aligned}
$$

and

$$\mathrm{Y}^2(\zeta z) = \sum_{q=0}^{\infty}\sum_{k=1}^{q-1} \frac{(-\zeta z)^q}{(q-k-1)!(k+1)!^2} \sum_{l=2}^{k} \frac{2l+1}{l(l+1)}$$

are given in Formulæ 12 to 14. Since all functions occurring in Eq. (7.3.20) are explicitly known, the result follows from basic algebra. □

We can now prove our main theorem.

Proof (of Theorem 7.3) The Chapman–Kolmogorov equation implies the recursion

$$K^n(r',r) = \int_{\mathbb{R}} K(r',s)K^{n-1}(s,r)\mathrm{d}s, \quad r,r' \in \mathbb{R}, \quad n > 1, \tag{7.3.21}$$

where K is the one-step transition kernel of Δ given in Eq. (7.3.1) From this we can first prove the asserted symmetry $K^n(r',r) = K^n(-r',-r)$ by induction on n. For $n = 1$, this is clearly true, so assuming that it holds for some $n \geqslant 0$, we conclude that $K^{n+1}(r',r)$ is equal to

$$\begin{aligned}\int_{\mathbb{R}} K(r',s)K^n(s,r)\mathrm{d}s &= \int_{\mathbb{R}} K(-r',-s)K^n(-s,-r)\mathrm{d}s \\ &= \int_{\mathbb{R}} K(-r',s)K^n(s,-r)\mathrm{d}s = K^{n+1}(-r',-r).\end{aligned}$$

In the next step we prove Eq. (7.2.10), also by induction on n. For $n = 1$, the claim is true by Lemma 7.9. We now assume that Eq. (7.2.10) holds for some $n \geqslant 1$. It then follows that, for $r \geqslant 0$,

$$\begin{aligned}K^{n+1}(r',r) &= \int_{\mathbb{R}} K(r',s)K^n(s,r)\mathrm{d}s \\ &= \int_{-\infty}^{0} K(r',s)K^n(s,r)\mathrm{d}s + \int_0^r K(r',s)K^n(s,r)\mathrm{d}s + \int_r^{\infty} K(r',s)K^n(s,r)\mathrm{d}s \\ &= \sum_{p,q=0}^{n} a^n_{p,q}\mathrm{e}^{-(q+2)r}\int_{-\infty}^{0} K(r',s)\mathrm{e}^{ps}\mathrm{d}s + \frac{(-1)^{n-1}\mathrm{e}^{-(n+1)r}}{(n-1)!}\int_0^r K(r',s)\mathrm{e}^{s}\mathrm{d}s \\ &\quad + \sum_{p=0}^{n-1} \frac{(-1)^n\mathrm{e}^{-(n-p)r}}{(p)!(n-p-2)!}\int_0^r K(r',s)s\mathrm{e}^{-ps}\mathrm{d}s \\ &\quad + \sum_{p,q=0}^{n} b^n_{p,q}\mathrm{e}^{-(q+2)r}\int_0^r K(r',s)\mathrm{e}^{-ps}\mathrm{d}s + \frac{(-1)^{n-1}\mathrm{e}^{-r}}{(n-1)!}\int_r^{\infty} K(r',s)\mathrm{e}^{-(n-1)s}\mathrm{d}s \\ &\quad + \sum_{p=0}^{n-1} \frac{(-1)^n\mathrm{e}^{-(n-p)r}r}{(p)!(n-p-2)!}\int_r^{\infty} K(r',s)\mathrm{e}^{-ps}\mathrm{d}s \\ &\quad + \sum_{p,q=0}^{n} c^n_{p,q}\mathrm{e}^{-(q+2)r}\int_r^{\infty} K(r',s)\mathrm{e}^{-ps}\mathrm{d}s.\end{aligned}$$

The five types of integrals occurring in this expression are easily evaluated to give, for $p \geqslant 0$,

$$\int_{-\infty}^{0} K(r',s)\mathrm{e}^{ps}\mathrm{d}s = \begin{cases} \frac{1}{p+1} - \frac{\mathrm{e}^{(p+1)r'}}{(p+1)(p+2)}, & r' \leqslant 0, \\ \frac{\mathrm{e}^{-r'}}{p+2}, & r' > 0, \end{cases}$$

$$\int_{0}^{r} K(r',s)\mathrm{e}^{s}\mathrm{d}s = \begin{cases} \mathrm{e}^{r'} - \mathrm{e}^{r'-r}, & r' \leqslant 0, \\ 1 - \mathrm{e}^{r'-r} + r', & 0 < r' \leqslant r, \\ r, & r' > r, \end{cases}$$

$$\int_{0}^{r} K(r',s)s\mathrm{e}^{-ps}\mathrm{d}s = \begin{cases} -\frac{\mathrm{e}^{r'-(p+2)r}r}{p+2} + \frac{\mathrm{e}^{r'} - \mathrm{e}^{r'-(p+2)r}}{(p+2)^2}, & r' \leqslant 0, \\ \frac{1}{(p+1)^2} - \frac{\mathrm{e}^{-(p+1)r'}(2p+3)}{(p+1)^2(p+2)^2} - \frac{\mathrm{e}^{-(p+1)r'}r'}{(p+1)(p+2)} - \frac{\mathrm{e}^{r'-(p+2)r}}{(p+2)^2} - \frac{\mathrm{e}^{r'-(p+2)r}r}{p+2}, & 0 < r' \leqslant r, \\ -\frac{\mathrm{e}^{-(p+1)r}r}{p+1} + \frac{1-\mathrm{e}^{-(p+1)r}}{(p+1)^2}, & r' > r, \end{cases}$$

$$\int_{0}^{r} K(r',s)\mathrm{e}^{-ps}\mathrm{d}s = \begin{cases} \frac{\mathrm{e}^{r'} - \mathrm{e}^{r'-(p+2)r}}{p+2}, & r' \leqslant 0, \\ \frac{1}{p+1} - \frac{\mathrm{e}^{r'-(p+2)r}}{p+2} - \frac{\mathrm{e}^{-(p+1)r'}}{(p+1)(p+2)}, & 0 < r' \leqslant r, \\ \frac{1-\mathrm{e}^{-(p+1)r}}{p+1}, & r' > r, \end{cases}$$

and

$$\int_{r}^{\infty} K(r',s)\mathrm{e}^{-ps}\mathrm{d}s = \begin{cases} \frac{\mathrm{e}^{r'-(p+2)r}}{p+2}, & r' \leqslant r, \\ \frac{\mathrm{e}^{-(p+1)r}}{p+1} - \frac{\mathrm{e}^{-(p+1)r'}}{(p+1)(p+2)}, & r' > r. \end{cases}$$

This implies that, for $r' \leqslant 0$, the function K^{n+1} is given by

$$K^{n+1}(r',r) = \sum_{p,q=0}^{n} \tilde{a}_{p,q}^{n+1}\mathrm{e}^{pr'-(q+2)r}, \tag{7.3.22}$$

where

$$\begin{aligned} \tilde{a}_{p,q}^{n+1} =& \delta_{p,0}\sum_{k=0}^{n}\frac{a_{k,q}^{n}}{k+1} - \Theta_{p,1}\frac{a_{p-1,q}^{n}}{p(p+1)} + \delta_{p,1}\Bigg[\delta_{q,n-1}\frac{(-1)^{n-1}}{(n-1)!} - \delta_{q,n}\frac{(-1)^{n-1}}{(n-1)!} \\ &+\tilde{\Theta}_{q,n-2}\frac{(-1)^{n}}{(n-q-2)!(q)!(n-q)^2} \\ &-\delta_{q,n}\sum_{k=0}^{n-2}\frac{(-1)^{n}}{(k)!(n-k-2)!(p+2)^2} \\ &+\sum_{k=0}^{n}\frac{b_{k,q}^{n}}{k+2} - \sum_{k=0}^{q-2}\frac{b_{k,q-k-2}^{n}}{k+2} + \sum_{k=0}^{q-2}\frac{c_{k,q-k-2}^{n}}{k+2} + \delta_{q,n}\frac{(-1)^{n-1}}{(n-1)!(n+1)}\Bigg], \end{aligned} \tag{7.3.23}$$

By Lemma 7.11, $\tilde{a}_{p,q}^{n+1}$ is equal to $a_{p,q}^{n+1}$. Similarly, for $0 < r' \leqslant r$, the function K^{n+1} takes the form

$$K^{n+1}(r',r) = \sum_{q=0}^{n} \beta_q^{n+1} e^{r'-(q+2)r} + \sum_{p=0}^{n-1} \frac{(-1)^{n+1} r' e^{-p(r'-r)-(n+1)r}}{(p)!(n-p-1)!} + \sum_{p=0}^{n+1}\sum_{q=0}^{n+1} \tilde{b}_{p,q}^{n+1} e^{-pr'-(q+2)r}, \tag{7.3.24}$$

where

$$\beta_q^{n+1} = \delta_{q,n}\left[\frac{(-1)^{n-1}}{(n-1)!(n+1)} - \frac{(-1)^{n-1}}{(n-1)!} - \sum_{k=0}^{n-2} \frac{(-1)^n}{(k)!(n-k-2)!(k+2)^2}\right] - \sum_{k=0}^{q-2} \frac{b_{k,q-k-2}^n}{k+2} + \sum_{k=0}^{q-2} \frac{c_{k,q-k-2}^n}{k+2}$$

and

$$\tilde{b}_{p,q}^{n+1} = \delta_{p,1}\sum_{k=0}^{n} \frac{a_{k,q}^n}{k+2} - \Theta_{p,1}\delta_{q,n-p-1}\frac{(-1)^n(2p+1)}{(p-1)!(n-p-1)!p^2(p+1)^2} + \delta_{p,0}\left[\delta_{q,n-1}\frac{(-1)^{n-1}}{(n-1)!} + \tilde{\Theta}_{q,n-2}\frac{(-1)^n}{(n-q-2)!(q)!(n-q-1)^2} + \sum_{k=0}^{n}\frac{b_{k,q}^n}{k+1}\right]. \tag{7.3.25}$$

Lemma 7.10 implies that $\beta_q^{n+1} = \delta_{q,n}\frac{(-1)^n}{(n)!}$ and, by Lemma 7.12, $\tilde{b}_{p,q}^{n+1}$ is equal to $b_{p,q}^{n+1}$. Finally, for $r' > r$, the function K^{n+1} becomes

$$K^{n+1}(r',r) = \frac{(-1)^n e^{-nr'-r}}{(n)!} + \sum_{p=0}^{n-1} \frac{(-1)^{n+1} r e^{-p(r'-r)-(n+1)r}}{(p)!(n-p-1)!} + \sum_{p=0}^{n+1}\sum_{q=0}^{n+1} \tilde{c}_{p,q}^{n+1} e^{-pr'-(q+2)r}, \tag{7.3.26}$$

where

$$\tilde{c}_{p,q}^{n+1} = \delta_{p,1}\sum_{k=0}^{n}\frac{a_{k,q}^n}{k+2} - \Theta_{p,1}\frac{c_{p-1,q}^n}{p(p+1)} + \delta_{p,0}\left[\tilde{\Theta}_{q,n-2}\frac{(-1)^n}{(n-q-2)!(q)!(n-q-1)^2} + \delta_{q,n-1}\sum_{k=0}^{n-2}\frac{(-1)^n}{(k)!(n-k-2)!(k+1)^2} + \sum_{k=0}^{n}\frac{b_{k,q}^n}{k+1} - \sum_{k=0}^{q-1}\frac{b_{k,q-k-1}^n}{k+1} - \delta_{q.n-1}\frac{(-1)^{n-1}}{n!^2} + \sum_{k=0}^{q-1}\frac{c_{k,q-k-1}^n}{k+1}\right]. \tag{7.3.27}$$

In Lemma 7.13 it was shown that $\tilde{c}_{p,q}^{n+1}$ equals $c_{p,q}^{n+1}$. Combining Eqs. (7.3.22), (7.3.24) and (7.3.26) proves the theorem because it follows that, for $r \geqslant 0$, the values $K^{n+1}(r',r)$

are given by

$$\begin{cases} \sum_{p,q=0}^{n} a_{p,q}^{n+1} \mathrm{e}^{pr'-(q+2)r}, & r' \leqslant 0, \\ \frac{(-1)^n \mathrm{e}^{r'-(n+2)r}}{(n)^!} + \sum_{p=0}^{n-1} \frac{(-1)^{n+1} r' \mathrm{e}^{-p(r'-r)-(n+1)r}}{(p)^!(n-p-1)^!} + \sum_{p=0}^{n+1}\sum_{q=0}^{n+1} b_{p,q}^{n+1} \mathrm{e}^{-pr'-(q+2)r}, & 0 < r' \leqslant r, \\ \frac{(-1)^n \mathrm{e}^{-nr'-r}}{(n)^!} + \sum_{p=0}^{n-1} \frac{(-1)^{n+1} r \mathrm{e}^{-p(r'-r)-(n+1)r}}{(p)^!(n-p-1)^!} + \sum_{p=0}^{n+1}\sum_{q=0}^{n+1} c_{p,q}^{n+1} \mathrm{e}^{-pr'-(q+2)r}, & r' > r. \end{cases} \quad \square$$

7.4. Discussion

The way in which Theorem 7.3 was proved gives little insight into how one arrives at the expressions (7.2.11) to (7.2.13) for the generating functions $\mathrm{A}_{p,q}$, $\mathrm{B}_{p,q}$, and $\mathrm{C}_{p,q}$ in the first place. It appears pertinent to briefly comment on how we derived these formulæ. The first step was to compute the kernels K^n for low values of n from the Chapman–Kolmogorov equation (7.3.21), and to observe that they have the form asserted in Theorem 7.3. In the next step we guessed the expression for the part of $K^n(r', r)$ not involving the coefficients $a_{p,q}^n$, $b_{p,q}^n$, and $c_{p,q}^n$, so that the problem was reduced to solving the recurrence equations (7.3.23), (7.3.25) and (7.3.27). Assuming the validity of Lemma 7.10, it turns out that the first two of these recurrence equation can be relatively easily solved first for $\mathrm{G}(z)$, which is, up to the factor $(-z)^q/(q)^!$, the generating function of $(a_{1,q}^n + b_{1,q}^n)_{n\geqslant 1}$, then for $\mathrm{A}_{p,q}$, and finally for $\mathrm{B}_{p,q}$. The third recursion for $(c_{p,q}^n)$ was simplified by the empirical observation that

$$\sum_{k=0}^{q-1} \frac{b_{k,q-k-1}^n}{k+1} - \sum_{k=0}^{q-1} \frac{c_{k,q-k-1}^n}{k+1} = \delta_{q,n+1}\left[\frac{(-1)^{n-1}}{n!^2} + \sum_{k=0}^{q-1} \frac{(-1)^q}{(k)^!(q-k-1)^!(k+1)^2} - d_q\right]$$

for some real numbers d_q, $q \geqslant 0$, and then solved for $\mathrm{C}_{p,q}$. The educated guesses made in the course of this derivation are justified *ex posteriori* by the proofs presented in this chapter.

Our original motivation was to derive an explicit expression for the asymptotic variance (7.2.7). For this purpose, knowledge of the generating function of the coefficients of the n-step transition kernel, as opposed to knowledge of the coefficients themselves, is sufficient. In order to evaluate the infinite sum appearing in Eq. (7.2.7), one is primarily interested in sums of the form $\sum_{n\geqslant 1} a_{p,q}^n$, which is equal to $\mathrm{A}_{p,q}(1)$, provided this number is finite. Carrying out the computations, however, turns out to be quite subtle, and the results will be reported elsewhere.

It is natural to ask whether the results presented in this chapter can be extended to the first-passage percolation problem on $\mathbb{N} \times \{0, 1, \ldots, k\}$, $k \geqslant 2$. Conceptually, our approach carries over to this setting only if one considers semi-directed percolation in which the horizontal edges may be traversed in only one direction; the combinatorics involved in computing the one-step transition kernel of the Markov chain Δ as well as the explicit iteration of the Chapman–Kolmogorov equation (7.3.21), however, soon become unmanageable for larger values of k. For the undirected first-passage percolation problem, there is the possibility that the shortest path $\{(0,0) = p_0, p_1, \ldots, p_{N-1}, p_N = (n,0)\}$, $p_i = (x_i, y_i)$, between $(0,0)$ and $(n,0)$ *backtraces*, by which we mean that there exist indices $0 \leqslant i < j \leqslant N$ such that $x_j < x_i$. The possible occurrence of such configurations prevents an extension of our recursive method to broader graphs in the undirected setting. One might also wonder if similar results can be obtained for more general class of edge-weight distributions $\mathbb{P}$. It is easy to see that the Markov property of Δ does not depend on the choice of $\mathbb{P}$, and an analysis of our proofs shows that the validity of the Central Limit Theorem 7.2 as well as expression (7.2.7) for the asymptotic variance is not affected by choosing a different edge-weight distribution either, provided one can prove that the stationary distribution $\tilde{\pi}$ and the one-step kernel K satisfy the moment and mixing conditions used in the proof of Theorem 7.2. It is however, very difficult, to evaluate the formula for the n-step transition kernel explicitly, if $\mathbb{P}$ is not the exponential distribution, although our approach via generating functions remains likewise applicable.

Bibliography

Abate, J., Choudhury, G. L. and Whitt, W. (2000). An introduction to numerical transform inversion and its application to probability models, *in* W. K. Grassmann (ed.), *Computational Probability*, Vol. 24, Kluwer Academic Publishers, Norwell, pp. 257–323.

Abramowitz, M. and Stegun, I. A. (eds) (1992). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, Dover Publications Inc., New York. Reprint of the 1972 edition.

Ahlberg, D. (2009). Asymptotics of first-passage percolation on 1-dimensional graphs, *Preprint - Department of Mathematical Sciences, Chalmers University of Technology and Göteborg University* **39**.

Aït-Sahalia, Y. and Yu, J. (2009). High frequency market microstructure noise estimates and liquidity measures, *Ann. Appl. Stat.* **3**(1): 422–457.

Akaike, H. (1977). On entropy maximization principle, *Applications of statistics (Proc. Sympos., Wright State Univ., Dayton, Ohio, 1976)*, North-Holland, Amsterdam, pp. 27–41.

Aldous, D., Lovász, L. and Winkler, P. (1997). Mixing times for uniformly ergodic Markov chains, *Stoch. Process. Their Appl.* **71**(2): 165–185.

Altmann, M. (1993). Reinterpreting network measures for models of disease transmission, *Social Networks* **15**(1): 1–17.

Amihud, Y., Mendelson, H. and Pedersen, L. (2006). *Liquidity and asset prices*, Now Publishers, Boston.

Anderson, G. W., Guionnet, A. and Zeitouni, O. (2010). *An introduction to random matrices*, Vol. 118 of *Cambridge Studies in Advanced Mathematics*, Cambridge University Press, Cambridge.

Anderson, G. W. and Zeitouni, O. (2008). A law of large numbers for finite-range dependent random matrices, *Comm. Pure Appl. Math.* **61**(8): 1118–1154.

Anderson, T. W. (2003). *An introduction to multivariate statistical analysis*, Wiley Series in Probability and Statistics, third edn, Wiley-Interscience, Hoboken.

Ansley, C. F. and Kohn, R. (1985). Estimation, filtering, and smoothing in state space models with incompletely specified initial conditions, *Ann. Statist.* **13**(4): 1286–1316.

Apostol, T. M. (1974). *Mathematical analysis*, second edn, Addison-Wesley Publishing Co., Reading–London–Don Mills.

Applebaum, D. (2004). *Lévy processes and stochastic calculus*, Vol. 93 of *Cambridge Studies in Advanced Mathematics*, Cambridge University Press, Cambridge.

Åström, K. J. (1970). *Introduction to stochastic control theory*, Vol. 70 of *Mathematics in Science and Engineering*, Academic Press, New York.

Åström, K. J., Hagander, P. and Sternby, J. (1984). Zeros of sampled systems, *Automatica J. IFAC* **20**(1): 31–38.

Athreya, K. B. and Pantula, S. G. (1986). Mixing properties of Harris chains and autoregressive processes, *J. Appl. Probab.* **23**(4): 880–892.

Aubrun, G. (2006). Random points in the unit ball of l_p^n, *Positivity* **10**(4): 755–759.

Auffinger, A., Ben Arous, G. and Péché, S. (2009). Poisson convergence for the largest eigenvalues of heavy tailed random matrices, *Ann. Inst. Henri Poincaré Probab. Stat.* **45**(3): 589–610.

Bai, Z. D. and Zhou, W. (2008). Large sample covariance matrices without independence structures in columns, *Stat. Sinica* **18**(2): 425–442.

Bai, Z. and Silverstein, J. W. (2010). *Spectral analysis of large dimensional random matrices*, Springer Series in Statistics, second edn, Springer, New York.

Bar-Ness, Y. and Langholz, G. (1975). Preservation of controllability under sampling, *Internat. J. Control* **22**(1): 39–47.

Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling, *Scand. J. Stat.* **24**(1): 1–13.

Barndorff-Nielsen, O. E. (1998). Processes of normal inverse Gaussian type, *Finance Stoch.* **2**(1): 41–68.

Barndorff-Nielsen, O. E., Blæsild, P. and Schmiegel, J. (2004). A parsimonious and universal description of turbulent velocity increments, *Eur. Phys. J. B Condens. Matter Phys.* **41**(3): 345–363.

Barndorff-Nielsen, O. E., Maejima, M. and Sato, K. (2006). Some classes of multivariate infinitely divisible distributions admitting stochastic integral representations, *Bernoulli* **12**(1): 1–33.

Barndorff-Nielsen, O. E. and Shephard, N. (2001a). Modelling by Lévy processes for financial econometrics, *in* O. E. Barndorff-Nielsen, T. Mikosch and S. I. Resnick (eds), *Lévy Processes: Theory and Applications*, Birkhäuser, Basel, pp. 283–318.

Barndorff-Nielsen, O. E. and Shephard, N. (2001b). Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics, *J. R. Stat. Soc. Ser. B* **63**(2): 167–241.

Barndorff-Nielsen, O. E. and Stelzer, R. (2011). Multivariate supOU processes, *Ann. Appl. Probab.* **21**(1): 140–182.

Bauer, D. (2009). Estimating ARMAX systems for multivariate time series using the state approach to subspace algorithms, *J. Multivar. Anal.* **100**(3): 397–421.

Bauer, D., Deistler, M. and Scherrer, W. (1999). Consistency and asymptotic normality of some subspace alg orithms for systems without observed inputs, *Automatica J. IFAC* **35**(7): 1243–1254.

Bauer, H. (2002). *Wahrscheinlichkeitstheorie*, de Gruyter Lehrbuch, fifth edn, Walter de Gruyter & Co., Berlin.

Belinschi, S., Dembo, A. and Guionnet, A. (2009). Spectral measure of heavy tailed band and covariance random matrices, *Comm. Math. Phys.* **289**(3): 1023–1055.

Ben Arous, G. and Guionnet, A. (2008). The spectrum of heavy tailed random matrices, *Comm. Math. Phys.* **278**(3): 715–751.

Benth, F. E. and Šaltytė Benth, J. (2009). Dynamic pricing of wind futures, *Energy Economics* **31**(1): 16–24.

Bernstein, D. S. (2005). *Matrix mathematics*, Princeton University Press, Princeton. Theory, facts, and formulas with application to linear systems theory.

Bertoin, J. (1996). *Lévy processes*, Vol. 121 of *Cambridge Tracts in Mathematics*, Cambridge University Press, Cambridge.

Bhamidi, S., van der Hofstad, R. and Hooghiemstra, G. (2010). First passage percolation on random graphs with finite mean degrees, *Ann. Appl. Probab.* **20**(5): 1907–1965.

Birkhoff, G. D. (1931). Proof of the ergodic theorem, *Proc. Nat. Acad. Sci. USA* **17**(12): 656–660.

Bodnarchuk, S. V. and Kulik, A. M. (2008). Conditions for existence and smoothness of the distribution density for an Ornstein–Uhlenbeck process with Lévy noise. Preprint: available at arXiv:0806.0442v1.

Borodin, A., Ferrari, P. L., Prähofer, M. and Sasamoto, T. (2007). Fluctuation properties of the TASEP with periodic initial configuration, *J. Stat. Phys.* **129**(5-6): 1055–1080.

Bose, A., Subhra Hazra, R. and Saha, K. (2009). Limiting spectral distribution of circulant type matrices with dependent inputs, *Electron. J. Probab.* **14**(86): 2463–2491.

Böttcher, A. and Silbermann, B. (1999). *Introduction to large truncated Toeplitz matrices*, Universitext, Springer-Verlag, New York.

Boubacar Mainassara, B. and Francq, C. (2011). Estimating structural VARMA models with uncorrelated but non-independent error terms, *J. Multivar. Anal.* **102**(3): 496–505.

Bradley, R. C. (2007). *Introduction to strong mixing conditions. Vol. 1*, Kendrick Press, Heber City.

Branch, M. A., Coleman, T. F. and Li, Y. (1999). A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems, *SIAM J. Sci. Comput.* **21**(1): 1–23 (electronic).

Brockwell, P. J. (2001a). Continuous-time ARMA processes, *Stochastic processes: theory and methods*, Vol. 19 of *Handbook of Statistics*, North-Holland, Amsterdam, pp. 249–276.

Brockwell, P. J. (2001b). Lévy-driven CARMA processes, *Ann. Inst. Stat. Math.* **53**(1): 113–124.

Brockwell, P. J. (2004). Representations of continuous-time ARMA processes, *J. Appl. Probab.* **41A**: 375–382. Stochastic methods and their applications.

Brockwell, P. J. and Davis, R. A. (1991). *Time series: theory and methods*, Springer Series in Statistics, second edn, Springer-Verlag, New York.

Brockwell, P. J., Davis, R. A. and Yang, Y. (2011). Estimation for nonnegative Lévy-driven CARMA processes, *J. Bus. Econ. Stat.* **29**(2): 250–259.

Brockwell, P. J. and Lindner, A. (2009). Existence and uniqueness of stationary Lévy-driven CARMA processes, *Stoch. Process. Their Appl.* **119**(8): 2660–2681.

Brockwell, P. J. and Schlemm, E. (2011). Parametric estimation of the driving Lévy process of multivariate CARMA processes from discrete observations. Preprint: Available at `http://www-m4.ma.tum.de`.

Brockwell, P. and Marquardt, T. G. (2005). Lévy-driven and fractionally integrated ARMA processes with continuous time parameter, *Stat. Sinica* **15**(2): 477–494.

Brown, B. M. and Hewitt, J. I. (1975). Asymptotic likelihood theory for diffusion processes, *J. Appl. Probab.* **12**(2): 228–238.

Bryc, W., Dembo, A. and Jiang, T. (2006). Spectral measure of large random Hankel, Markov and Toeplitz matrices, *Ann. Probab.* **34**(1): 1–38.

Burda, Z., Jarosz, A., Nowak, M. A. and Snarska, M. (2010). A random matrix approach to VARMA processes, *New Journal of Physics* **12**: 075036.

Byrd, R. H., Schnabel, R. B. and Shultz, G. A. (1988). Approximate solution of the trust region problem by minimization over two-dimensional subspaces, *Math. Program.* **40**(1): 247–263.

Caines, P. E. (1988). *Linear stochastic systems*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York.

Chatterjee, S. and Dey, P. S. (2009). Central limit theorem for first-passage percolation time across thin cylinders. Preprint: available at arXiv:0911.5702v2.

Chen, X. (1999). Limit theorems for functionals of ergodic Markov chains with general state space, *Mem. Am. Math. Soc.* **139**(664): xiv+203.

Chiuso, A. (2006). Asymptotic variance of closed-loop subspace identification methods, *IEEE Trans. Automat. Control* **51**(8): 1299–1314.

Chiuso, A. and Picci, G. (2005). Consistency analysis of some closed-loop subspace identification methods, *Automatica J. IFAC* **41**(3): 377–391.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues, *Quant. Financ.* **1**(2): 223–236.

Cruz-Uribe, D. and Neugebauer, C. J. (2002). Sharp error bounds for the trapezoidal rule and Simpson's rule, *JIPAM. J. Inequal. Pure Appl. Math.* **3**(4): Article 49, 22 pp.

Davydov, Y. A. (1968). Convergence of distributions generated by stationary stochastic processes, *Theory Probab. Appl.* **13**(4): 691–696.

Dedecker, J., Doukhan, P., Lang, G., León R., J. R., Louhichi, S. and Prieur, C. (2007). *Weak dependence: with examples and applications*, Vol. 190 of *Lecture Notes in Statistics*, Springer, New York.

Deistler, M. (1983). The properties of the parameterization of ARMAX systems and their relevance for structural estimation and dynamic specification, *Econometrica* **51**(4): 1187–1207.

Deuflhard, P. and Hohmann, A. (2008). *Numerische Mathematik. 1*, de Gruyter Lehrbuch, fourth edn, Walter de Gruyter & Co., Berlin. Eine algorithmisch orientierte Einführung.

Dieudonné, J. (1968). *Éléments d'analyse. Tome II: Chapitres XII à XV*, Cahiers Scientifiques, Fasc. XXXI, Gauthier-Villars, Éditeur, Paris.

Doob, J. L. (1944). The elementary Gaussian processes, *Ann. Math. Statistics* **15**(3): 229–282.

Doukhan, P. (1994). *Mixing: Properties and examples*, Vol. 85 of *Lecture Notes in Statistics*, Springer-Verlag, New York.

Duncan, T. E., Mandl, P. and Pasik-Duncan, B. (1999). A note on sampling and parameter estimation in linear stochastic systems, *IEEE Trans. Autom. Control* **44**(11): 2120–2125.

Durrett, R. (2010). *Probability: theory and examples*, Cambridge Series in Statistical and Probabilistic Mathematics, fourth edn, Cambridge University Press, Cambridge.

Fan, H., Söderström, T., Mossberg, M., Carlsson, B. and Zon, Y. (1998). Continuous-time AR process parameter estimation from discrete-time data, *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 4, pp. 1232–1244.

Feigin, P. D. (1976). Maximum likelihood estimation for continuous-time stochastic processes, *Adv. Appl. Probab.* **8**(4): 712–736.

Ferguson, T. S. (1996). *A course in large sample theory*, Texts in Statistical Science Series, Chapman & Hall, London.

Figueroa-López, J. E. (2009). Nonparametric estimation of Lévy models based on discrete-sampling, *Optimality*, Vol. 57 of *IMS Lecture Notes Monogr. Ser.*, Inst. Math. Stat., Beachwood, pp. 117–146.

Flaxman, A., Gamarnik, D. and Sorkin, G. (2006). First-passage percolation on a width-2 strip and the path cost in a VCG auction, *Internet and Network Economics* **4286**: 99–111.

Francq, C. and Zakoïan, J.-M. (1998). Estimating linear representations of nonlinear processes, *J. Stat. Plan. Infer.* **68**(1): 145–165.

Gerschgorin, S. (1931). Über die Abgrenzung der Eigenwerte einer Matrix, *Bulletin de l'Académie des Sciences de l'URSS. Classe des sciences mathématiques et naturelles* **6**: 749–754.

Gevers, M. (1986). ARMA models, their Kronecker indices and their McMillan degree, *Int. J. Control* **43**(6): 1745–1761.

Gevers, M. and Wertz, V. (1983). Overlapping parametrizations for the representation of multivariate stationary time series, *Geometry and identification (Weston, Mass., 1981)*, Lie Groups: Hist., Frontiers and Appl. Ser. B: Systems Inform. Control, 1, Math Sci Press, Brookline, pp. 73–99.

Gevers, M. and Wertz, V. (1984). Uniquely identifiable state-space and ARMA parametrizations for multivariable linear systems, *Automatica J. IFAC* **20**(3): 333–347.

Gillberg, J. and Ljung, L. (2009). Frequency-domain identification of continuous-time ARMA models from sampled data, *Automatica* **45**(6): 1371–1378.

Graham, R. L., Grötschel, M. and Lovász, L. (eds) (1995). *Handbook of combinatorics. Vol. 2*, Elsevier Science B.V., Amsterdam.

Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing, *J. Time Ser. Anal.* **1**(1): 15–29.

Gray, R. M. (2006). *Toeplitz and circulant matrices: A review*, Now Publishers, Boston.

Grenander, U. and Szegő, G. (1984). *Toeplitz forms and their applications*, second edn, Chelsea Publishing, New York.

Grimmett, G. and Kesten, H. (1984). First-passage percolation, network flows and electrical resistances, *Z. Wahrsch. Verw. Gebiete* **66**(3): 335–366.

Gugushvili, S. (2009). Nonparametric estimation of the characteristic triplet of a discretely observed Lévy process, *J. Nonparametr. Stat.* **21**(3): 321–343.

Guidorzi, R. P. (1975). Canonical structures in the identification of multivariable systems, *Automatica—J. IFAC* **11**(4): 361–374.

Guidorzi, R. P. (1981). Invariants and canonical forms for systems: structural and parametric identification, *Automatica—J. IFAC* **17**(1): 117–133.

Hachem, W., Loubaton, P. and Najim, J. (2005). The empirical eigenvalue distribution of a Gram matrix: from independence to stationarity, *Markov Process. Related Fields* **11**(4): 629–648.

Hachem, W., Loubaton, P. and Najim, J. (2006). The empirical distribution of the eigenvalues of a Gram matrix with a given variance profile, *Ann. Inst. H. Poincaré Probab. Statist.* **42**(6): 649–670.

Hagiwara, T. and Araki, M. (1988). Controllability indices of sampled-data systems, *Internat. J. Systems Sci.* **19**(12): 2449–2457.

Hall, A. R. (2005). *Generalized method of moments*, Advanced Texts in Econometrics, Oxford University Press, Oxford.

Halmos, P. R. (1950). *Measure Theory*, D. Van Nostrand Company, Inc., New York.

Halmos, P. R. (1974). *Finite-dimensional vector spaces*, Undergraduate Texts in Mathematics, second edn, Springer-Verlag, New York.

Hamilton, J. D. (1994). *Time series analysis*, Princeton University Press, Princeton.

Hammersley, J. M. and Welsh, D. J. A. (1965). First-passage percolation, subadditive processes, stochastic networks, and generalized renewal theory, *Proc. Internat. Res. Semin., Statist. Lab., Univ. California, Berkeley, Calif*, Springer, New York, pp. 61–110.

Hannan, E. J. (1969). The estimation of mixed moving average autoregressive systems, *Biometrika* **56**(3): 579–593.

Hannan, E. J. (1970). *Multiple time series*, John Wiley and Sons, Inc., New York–London–Sydney.

Hannan, E. J. (1971). The identification problem for multiple equation systems with moving average errors, *Econometrica* **39**(5): 751–765.

Hannan, E. J. (1973). The asymptotic theory of linear time-series models, *J. Appl. Probab.* **10**: 130–145, corrections, ibid. 10 (1973), 913.

Hannan, E. J. (1975). The estimation of ARMA models, *Ann. Statist.* **3**(4): 975–981.

Hannan, E. J. (1976). The identification and parametrization of ARMAX and state space forms, *Econometrica* **44**(4): 713–723.

Hannan, E. J. (1979). A note on autoregressive-moving average identification, *Biometrika* **66**(3): 672–674.

Hannan, E. J. (1980). The estimation of the order of an ARMA process, *Ann. Statist.* **8**(5): 1071–1081.

Hannan, E. J. and Deistler, M. (1988). *The statistical theory of linear systems*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York.

Hannan, E. J. and Kavalieris, L. (1984a). A method for autoregressive-moving average estimation, *Biometrika* **71**(2): 273–280.

Hannan, E. J. and Kavalieris, L. (1984b). Multivariate linear time series models, *Adv. Appl. Probab.* **16**(3): 492–561.

Hannan, E. J. and Rissanen, J. (1982). Recursive estimation of mixed autoregressive-moving average order, *Biometrika* **69**(1): 81–94.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators, *Econometrica* **50**(4): 1029–1054.

Hansen, L. P. and Sargent, T. J. (1983). The dimensionality of the aliasing problem in models with rational spectral densities, *Econometrica* **51**(2): 377–387.

Hansen, P. and Lunde, A. (2006). Realized variance and market microstructure noise, *J. Bus. Econ. Stat.* **24**(2): 127–161.

Haug, S. and Stelzer, R. (2011). Multivariate ECOGARCH processes, *Econometric Theory* **27**(2): 344–371.

Hautus, M. L. J. (1969). Controllability and observability conditions of linear autonomous systems, *Indag. Math.* **31**: 443–448.

Herrndorf, N. (1984). A functional central limit theorem for weakly dependent sequences of random variables, *Ann. Probab.* **12**(1): 141–153.

Higham, N. J. (2008). *Functions of matrices*, Society for Industrial and Applied Mathematics, Philadelphia. Theory and computation.

Hofmann-Credner, K. and Stolz, M. (2008). Wigner theorems for random matrices with dependent entries: ensembles associated to symmetric spaces and sample covariance matrices, *Electron. Commun. Probab.* **13**: 401–414.

Horn, R. A. and Johnson, C. R. (1994). *Topics in matrix analysis*, Cambridge University Press, Cambridge. Corrected reprint of the 1991 original.

Hosking, J. R. M. (1981). Fractional differencing, *Biometrika* **68**(1): 165–176.

Hu, Y. and Long, H. (2009). Least squares estimator for Ornstein–Uhlenbeck processes driven by α-stable motions, *Stoch. Process. Their Appl.* **119**(8): 2465–2480.

Hyndman, R. J. (1993). Yule–Walker estimates for continuous-time autoregressive models, *J. Time Ser. Anal.* **14**(3): 281–296.

Ibragimov, I. A. (1962). Some limit theorems for stationary processes, *Theory Probab. Appl.* **7**: 349–382.

Jacod, J. and Protter, P. (2003). *Probability essentials*, Universitext, second edn, Springer-Verlag, Berlin.

Jacod, J. and Shiryaev, A. N. (2003). *Limit theorems for stochastic processes*, Vol. 288 of *Grundlehren der Mathematischen Wissenschaften*, second edn, Springer-Verlag, Berlin.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis, *Ann. Statist.* **29**(2): 295–327.

Jones, R. H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations, *Technometrics* **22**(3): 389–395.

Kailath, T. (1980). *Linear systems*, Prentice-Hall Information and System Sciences Series, Prentice-Hall Inc., Englewood Cliffs.

Kailath, T., Segall, A. and Zakai, M. (1978). Fubini-type theorems for stochastic integrals, *Sankhyā Ser. A* **40**(2): 138–143.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems, *Journal of Basic Engineering* **82**(1): 35–45.

Kesten, H. (1986). Aspects of first passage percolation, *École d'été de probabilités de Saint-Flour, XIV—1984*, Vol. 1180 of *Lecture Notes in Math.*, Springer, Berlin, pp. 125–264.

Kesten, H. (1987). Percolation theory and first-passage percolation, *Ann. Probab.* **15**(4): 1231–1271.

Kesten, H. (1993). On the speed of convergence in first-passage percolation, *Ann. Appl. Probab.* **3**(2): 296–338.

Kingman, J. F. C. (1968). The ergodic theory of subadditive stochastic processes, *J. R. Stat. Soc. Ser. B* **30**(3): 499–510.

Klein, A., Mélard, G. and Saidi, A. (2008). The asymptotic and exact Fisher information matrices of a vector ARMA process, *Stat. Probab. Lett.* **78**(12): 1430–1433.

Klein, A. and Neudecker, H. (2000). A direct derivation of the exact Fisher information matrix of Gaussian vector state space models, *Linear Alg. Appl.* **321**(1-3): 233–238.

Klenke, A. (2008). *Probability theory*, Universitext, Springer-Verlag London Ltd., London.

Klüppelberg, C. and Mikosch, T. (1993). Spectral estimates and stable processes, *Stoch. Process. Their Appl.* **47**(2): 323–344.

Krengel, U. (1985). *Ergodic theorems*, Vol. 6 of *de Gruyter Studies in Mathematics*, Walter de Gruyter & Co., Berlin. With a supplement by Antoine Brunel.

Kulikova, M. V. and Semoushin, I. V. (2006). Score evaluation within the extended square-root information filter, *Lecture Notes in Computer Science* **3991**: 473–481.

Kwapień, S. and Rosiński, J. (2004). Sample Hölder continuity of stochastic processes and majorizing measures, *Seminar on Stochastic Analysis, Random Fields and Applications IV*, Vol. 58 of *Progr. Probab.*, Birkhäuser, Basel, pp. 155–163.

Lahalle, E., Fleury, G. and Rivoira, A. (2004). Continuous ARMA spectral estimation from irregularly sampled observations, *Proceedings of the 21st IEEE Instrumentation and Measurement Technology Conference*, Vol. 2, pp. 923– 927.

Lancaster, P. and Rodman, L. (1995). *Algebraic Riccati equations*, The Clarendon Press/Oxford University Press, New York – Oxford.

Larsson, E. K. (2005). Limiting sampling results for continuous-time ARMA systems, *Internat. J. Control* **78**(7): 461–473.

Larsson, E. K., Mossberg, M. and Söderström, T. (2006). An overview of important practical aspects of continuous-time ARMA system identification, *Circuits Systems Signal Process.* **25**(1): 17–46.

Larsson, E. K. and Söderström, T. (2002). Identification of continuous-time AR processes from unevenly sampled data, *Automatica J. IFAC* **38**(4): 709–718.

Lax, P. D. (2002). *Functional analysis*, Pure and Applied Mathematics, Wiley-Interscience, New York.

Lebesgue, H. (1904). Une propriété caractéristique des fonctions de classe 1, *Bull. Soc. Math. France* **32**: 229–242.

Leneman, O. and Lewis, J. (1966). Random sampling of random processes: Mean-square comparison of various interpolators, *IEEE Trans. Autom. Control* **11**(3): 396–403.

LeVeque, R. J. (2007). *Finite difference methods for ordinary and partial differential equations*, Society for Industrial and Applied Mathematics, Philadelphia. Steady-state and time-dependent problems.

Licea, C. and Newman, C. M. (1996). Geodesics in two-dimensional first-passage percolation, *Ann. Probab.* **24**(1): 399–410.

Lii, K. S. and Masry, E. (1995). On the selection of random sampling schemes for the spectral estimation of continuous time processes, *J. Time Ser. Anal.* **16**(3): 291–311.

Liptser, R. S. and Shiryaev, A. N. (2001). *Statistics of random processes. II*, Vol. 6 of *Applications of Mathematics*, expanded edn, Springer-Verlag, Berlin. Applications, Translated from the 1974 Russian original by A. B. Aries.

Luenberger, D. G. (1967). Canonical forms for linear multivariable systems, *IEEE Trans. Automatic Control* **12**(3): 290–293.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*, Springer-Verlag, Berlin.

Lütkepohl, H. and Poskitt, D. S. (1996). Specification of echelon-form VARMA models, *J. Bus. Econ. Stat.* **14**(1): 69–79.

Mai, H. (2009). Maximum-Likelihood-Estimation of Lévy driven Ornstein–Uhlenbeck processes. Presented at the Workshop on Statistical Inference for Lévy processes, EURANDOM.

Marchenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices, *Mat. Sb. (N.S.)* **72(114)**(4): 507–536.

Mari, J., Stoica, P. and McKelvey, T. (2000). Vector ARMA estimation: a reliable subspace approach, *IEEE Trans. Signal Process.* **48**(7): 2092–2103.

Marquardt, T. (2007). Multivariate fractionally integrated CARMA processes, *J. Multivar. Anal.* **98**(9): 1705–1725.

Marquardt, T. and Stelzer, R. (2007). Multivariate CARMA processes, *Stoch. Process. Their Appl.* **117**(1): 96–120.

Masry, E. (1978). Poisson sampling and spectral estimation of continuous-time processes, *IEEE Trans. Inf. Theory* **24**(2): 173–183.

Masuda, H. (2004). On multidimensional Ornstein–Uhlenbeck processes driven by a general Lévy process, *Bernoulli* **10**(1): 97–120.

McCrorie, J. R. (2003). The problem of aliasing in identifying finite parameter continuous time stochastic models, *Proceedings of the Eighth Vilnius Conference on Probability Theory and Mathematical Statistics, Part II (2002)*, Vol. 79, pp. 9–16.

Meckes, M. W. (2007). On the spectral norm of a random Toeplitz matrix, *Electron. Comm. Probab.* **12**: 315–325.

Mehta, M. L. (2004). *Random matrices*, Pure and Applied Mathematics, third edn, Elsevier/Academic Press, Amsterdam.

Mikosch, T., Gadrich, T., Klüppelberg, C. and Adler, R. J. (1995). Parameter estimation for ARMA models with infinite variance innovations, *Ann. Stat.* **23**(1): 305–326.

Mokkadem, A. (1988). Mixing properties of ARMA processes, *Stoch. Process. Their Appl.* **29**(2): 309–315.

Mörters, P. and Peres, Y. (2010). *Brownian motion*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge.

Muirhead, R. J. (1982). *Aspects of multivariate statistical theory*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York.

Na, S. S. and Rhee, H. K. (2002). An experimental study for property control in a continuous styrene polymerization reactor using a polynomial ARMA model, *Chem. Eng. Sci.* **57**(7): 1165–1173.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing, *Handbook of econometrics, Vol. IV*, Vol. 2 of *Handbooks in Econom.*, North-Holland, Amsterdam, pp. 2111–2245.

Nielsen, J. N., Madsen, H. and Young, P. C. (2000). Parameter estimation in stochastic differential equations: an overview, *Annu. Rev. Control* **24**: 83–94.

Nummelin, E. and Tuominen, P. (1982). Geometric ergodicity of Harris recurrent Markov chains with applications to renewal theory, *Stoch. Process. Their Appl.* **12**(2): 187–202.

Pan, G. (2010). Strong convergence of the empirical distribution of eigenvalues of sample covariance matrices with a perturbation matrix, *J. Multivar. Anal.* **101**(6): 1330–1338.

Patel, R. V. (1981). Computation of matrix fraction descriptions of linear time-invariant systems, *IEEE Trans. Automat. Control* **26**(1): 148–161.

Peternell, K., Scherrer, W. and Deistler, M. (1996). Statistical analysis of novel subspace identification methods, *Signal Processing* **52**(2): 161–177.

Pfaffel, O. and Schlemm, E. (2011). Eigenvalue distribution of large sample covariance matrices of linear processes, *Probab. Math. Stat.* **31**(2).

Pfaffel, O. and Schlemm, E. (2012). Limiting spectral distribution of a new random matrix model with dependence across rows and columns, *Linear Alg. Appl.* . To appear.

Pham, T. D. (1977). Estimation of parameters of a continuous time Gaussian stationary process with rational spectral density, *Biometrika* **64**(2): 385–399.

Pham, T. D. and Tran, L. T. (1985). Some mixing properties of time series models, *Stoch. Process. Their Appl.* **19**(2): 297–303.

Picard, J. (1996). On the existence of smooth densities for jump processes, *Probab. Theory Relat. Fields* **105**(4): 481–511.

Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L. A. N., Guhr, T. and Stanley, H. E. (2002). Random matrix approach to cross correlations in financial data, *Physical Review E* **65**(6): 66126.

Poskitt, D. S. (1992). Identification of echelon canonical forms for vector linear processes using least squares, *Ann. Statist.* **20**(1): 195–215.

Potters, M., Bouchaud, J.-P. and Laloux, L. (2005). Financial applications of random matrix theory: old laces and new pieces, *Acta Phys. Polon. B* **36**(9): 2767–2784.

Prähofer, M. and Spohn, H. (2000). Statistical self-similarity of one-dimensional growth processes, *Phys. A* **279**(1-4): 342–352.

Price, K. V., Storn, R. M. and Lampinen, J. A. (2005). *Differential evolution*, Natural Computing Series, Springer-Verlag, Berlin. A practical approach to global optimization.

Priola, E. and Zabczyk, J. (2009). Densities for Ornstein–Uhlenbeck processes with jumps, *Bull. Lond. Math. Soc.* **41**(1): 41–50.

Protter, P. (1990). *Stochastic integration and differential equations*, Vol. 21 of *Applications of Mathematics*, Springer-Verlag, Berlin. A new approach.

Rajput, B. S. and Rosiński, J. (1989). Spectral representations of infinitely divisible processes, *Probab. Theory Relat. Field* **82**(3): 451–487.

Rashidi Far, R., Oraby, T., Bryc, W. and Speicher, R. (2008). On slow-fading MIMO systems with nonseparable correlation, *IEEE Trans. Inform. Theory* **54**(2): 544–553.

Reinsel, G. C. (1997). *Elements of multivariate time series analysis*, Springer Series in Statistics, second edn, Springer-Verlag, New York.

Renlund, H. (2010). First-passage percolation with exponential times on a ladder, *Comb. Probab. Comput.* **19**(4): 593–601.

Renlund, H. (2011). First-passage percolation on ladder-like graphs with heterogeneous exponential times. Preprint: available at arXiv:1102.4744v1.

Rissanen, J. (1983). A universal prior for integers and estimation by minimum description length, *Ann. Statist.* **11**(2): 416–431.

Rissanen, J. (1986). Order estimation by accumulated prediction errors, *J. Appl. Probab.* (Special Vol. 23A): 55–61. Essays in time series and allied processes.

Rivoira, A., Moudden, Y. and Fleury, G. (2002). Real time continuous AR parameter estimation from randomly sampled observations, *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 2, pp. 1725–1728.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition, *Proc. Nat. Acad. Sci. U. S. A.* **42**(1): 43–47.

Rosenbrock, H. H. (1970). *State-space and multivariable theory*, John Wiley & Sons, Inc., New York.

Rozanov, Y. A. (1967). *Stationary random processes*, Holden-Day Inc., San Francisco. Translated from the Russian by A. Feinstein.

Rózsa, P. and Sinha, N. K. (1975). Minimal realization of a transfer function matrix in canonical forms, *Internat. J. Control* **21**(2): 273–284.

Rydberg, T. (1997). The normal inverse Gaussian Lévy process: simulation and approximation, *Stoch. Models* **13**(4): 887–910.

Sato, K. (1999). *Lévy processes and infinitely divisible distributions*, Vol. 68 of *Cambridge Studies in Advanced Mathematics*, Cambridge University Press, Cambridge.

Sato, K. (2006). Additive processes and stochastic integrals, *Ill. J. Math.* **50**(1-4): 825–851.

Sato, K. and Yamazato, M. (1983). Stationary processes of Ornstein–Uhlenbeck type, *Probability theory and mathematical statistics (Tbilisi, 1982)*, Vol. 1021 of *Lecture Notes in Math.*, Springer, Berlin, pp. 541–551.

Sato, K. and Yamazato, M. (1984). Operator-self-decomposable distributions as limit distributions of processes of Ornstein–Uhlenbeck type, *Stoch. Process. Their Appl.* **17**(1): 73–100.

Schlemm, E. (2009). First-passage percolation rates on width-two stretches with exponential link weights, *Electron. Commun. Probab.* **140**: 424–434.

Schlemm, E. (2011). On the Markov transition kernels for first-passage percolation on the ladder, *J. Appl. Probab.* **48**(2): 366–388.

Schlemm, E. and Stelzer, R. (2011). Quasi maximum likelihood estimation for strongly mixing linear state space models and multivariate CARMA processes. Preprint: Available at `http://www-m4.ma.tum.de`.

Schlemm, E. and Stelzer, R. (2012). Multivariate CARMA processes, continuous-time state space models and complete regularity of the innovations of the sampled processes, *Bernoulli* . To appear.

Schwarz, G. (1978). Estimating the dimension of a model, *Ann. Statist.* **6**(2): 461–464.

Segal, M. and Weinstein, E. (1989). A new method for evaluating the log-likelihood gradient, the Hessian, and the Fisher information matrix for linear dynamic systems, *IEEE Trans. Inform. Theory* **35**(3): 682–687.

Seppäläinen, T. (1998). Exact limiting shape for a simplified model of first-passage percolation on the plane, *Ann. Probab.* **26**(3): 1232–1250.

Shiryaev, A. N. (1996). *Probability*, Vol. 95 of *Graduate Texts in Mathematics*, second edn, Springer-Verlag, New York. Translated from the first (1980) Russian edition by R. P. Boas.

Silverstein, J. W. and Bai, Z. D. (1995). On the empirical distribution of eigenvalues of a class of large-dimensional random matrices, *J. Multivar. Anal.* **54**(2): 175–192.

Simon, T. (2010). On the absolute continuity of multidimensional Ornstein–Uhlenbeck processes, *Probab. Theory Relat. Field* . online first, DOI:10.1007/s00440-010-0296-5.

Smythe, R. T. and Wierman, J. C. (1978). *First-passage percolation on the square lattice*, Vol. 671 of *Lecture Notes in Mathematics*, Springer, Berlin.

Söderström, T. (1990). On zero locations for sampled stochastic systems, *IEEE Trans. Automat. Control* **35**(11): 1249–1253.

Söderström, T., Fan, H., Carlsson, B. and Mossberg, M. (1997). Some approaches on how to use the delta operator when identifying continuous-time processes, *Proceedings of the 36th IEEE Conference on Decision and Control*, Vol. 1, pp. 890–895.

Sontag, E. D. (1998). *Mathematical control theory*, Vol. 6 of *Texts in Applied Mathematics*, second edn, Springer-Verlag, New York. Deterministic finite-dimensional systems.

Sood, V., Redner, S. and ben Avraham, D. (2005). First-passage properties of the Erdős-Renyi random graph, *J. Phys. A* **38**(1): 109–123.

Spiliopoulos, K. (2009). Method of moments estimation of Ornstein–Uhlenbeck processes driven by general Lévy process, *Ann. I.S.U.P.* **53**(2-3): 3–17.

Stoffer, D. S. and Wall, K. D. (1991). Bootstrapping state-space models: Gaussian maximum likelihood estimation and the Kalman filter, *J. Am. Stat. Assoc.* **86**(416): 1024–1033.

Stone, B. J. (1962). Best possible ratios of certain matrix norms, *Numer. Math.* **4**: 114–116.

Sun, J. (1998). Sensitivity analysis of the discrete-time algebraic Riccati equation, *Linear Alg. Appl.* **275/276**: 595–615.

Szegő, G. (1920). Beiträge zur Theorie der Toeplitzschen Formen, *Math. Z.* **6**(3-4): 167–202.

Tao, T. and Vu, V. (2010). Random covariance matrices: Universality of local statistics of eigenvalues, *Ann. Probab.* . To appear.

Todorov, V. and Tauchen, G. (2006). Simulation methods for Lévy-driven continuous-time autoregressive moving average (CARMA) stochastic volatility models, *J. Bus. Econ. Stat.* **24**(4): 455–469.

Tong, H. (1990). *Nonlinear time series*, Vol. 6 of *Oxford Statistical Science Series*, The Clarendon Press/Oxford University Press, New York – Oxford. A dynamical system approach.

Tracy, C. A. and Widom, H. (1994). Level-spacing distributions and the Airy kernel, *Comm. Math. Phys.* **159**(1): 151–174.

Tracy, C. A. and Widom, H. (1996). On orthogonal and symplectic matrix ensembles, *Comm. Math. Phys.* **177**(3): 727–754.

Tucker, H. G. (1965). On a necessary and sufficient condition that an infinitely divisible distribution be absolutely continuous, *Trans. Am. Math. Soc.* **118**: 316–330.

Tulino, A. M. and Verdu, S. (2004). *Random Matrix Theory and Wireless Communications*, Now Publishers, Boston.

van der Hofstad, R., Hooghiemstra, G. and Van Mieghem, P. (2001). First-passage percolation on the random graph, *Probab. Engrg. Inform. Sci.* **15**(2): 225–237.

van der Vaart, A. W. (1998). *Asymptotic statistics*, Vol. 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press, Cambridge.

van Overschee, P. and De Moor, B. (1996). *Subspace identification for linear systems*, Kluwer Academic Publishers, Boston. Theory—implementation—applications.

Volkonskiĭ, V. A. and Rozanov, Y. A. (1959). Some limit theorems for random functions. I, *Theory Probab. Appl.* **4**(2): 178–197.

Wachter, K. W. (1978). The strong limits of random matrix spectra for sample matrices of independent elements, *Ann. Probab.* **6**(1): 1–18.

Wigner, E. P. (1958). On the distribution of the roots of certain symmetric matrices, *Ann. of Math. (2)* **67**(2): 325–327.

Wilf, H. S. (2006). *generatingfunctionology*, third edn, A K Peters Ltd., Wellesley.

Wold, H. (1954). *A study in the analysis of stationary time series*, second edn, Almqvist and Wiksell, Stockholm.

Wolfram Research, Inc. (2010). The wolfram functions site, `http://functions.wolfram.com/`.

Yin, Y. Q. (1986). Limiting spectral distribution for a class of random matrices, *J. Multivar. Anal.* **20**(1): 50–68.

Zhang, L. (2006). *Spectral Analysis of large dimensional random matrices*, PhD thesis, National University of Singapore.

General Notation

Symbols involving letters in alphabetical order:

$\mathscr{B}(\cdot)$	Borel σ-algebra
$\mathbb{C}$	complex numbers
$\mathbb{C}\mathrm{ov}$	covariance
$\delta_{i,j}$	Kronecker symbol
δ_x	Dirac measure
e	Euler number
$\mathbb{E}$	expectation
im	image of a matrix
Im	imaginary part
i	imaginary unit
$I_{\{\mathcal{E}\}}$	indicator of the expression $\mathcal{E}$
$I_A(\cdot)$	indicator function of the set A; $I_A(x) = I_{\{x \in A\}}$
$\mathbb{K}[X]$	polynomial expressions in X over $\mathbb{K}$
$\mathbb{K}\{X\}$	rational expressions in X over $\mathbb{K}$
ker	kernel of a matrix
$L^p(\cdot)$	Lebesgue space
$M_{m,n}(\cdot)$	$m \times n$ matrices
$M_n(\cdot)$	$n \times n$ matrices
$\mathbb{N}$	natural numbers
$O(\cdot)$	of the same order, that is $a_n = O(b_n)$, if $\exists\, C > 0$ such that $\|a_n\| \leqslant Cb_n$ for all n.
$o(\cdot)$	of smaller order, that is $a_n = o(b_n)$, if $\lim_{n\to\infty} a_n/b_n = 0$.
$\mathbb{P}(\cdot)$	probability
rank	rank of a matrix
Re	real part
$\mathbb{R}$	real numbers
$\mathbb{R}^+$	non-negative real numbers
$\mathbb{S}_n(\cdot)$	symmetric $n \times n$ matrices
$\mathbb{S}_n^+(\cdot)$	positive semidefinite $n \times n$ matrices
$\mathbb{S}_n^{++}(\cdot)$	positive definite $n \times n$ matrices
$\sigma(\cdot)$	spectrum of a matrix

$(\cdot)^T$	transpose of a matrix
$\mathbb{V}\mathrm{ar}$	variance
vec	vectorization operator
$\mathbb{Z}$	integers
$\mathbb{Z}^-$	negative integers

Other symbols in alphabetical order of their meaning:

$[\cdot,\cdot]$	closed interval
$\xrightarrow{d}$	convergence in distribution
$\xrightarrow{p}$	convergence in probability
$\sim$	distributed as
$\stackrel{d}{=}$	equality in distribution
$\langle\cdot,\cdot\rangle$	Euclidean inner product
$\mathbf{1}_n$	$n \times n$ identity matrix
$\otimes$	Kronecker product
$\lVert\cdot\rVert$	norm
$(\cdot,\cdot)$	open interval
$0_{m,n}$	$m \times n$ zero matrix
0_n	$n \times n$ zero matrix
$\mathbf{0}_n$	zero vector in $\mathbb{R}^n$

Abbreviations

ARMA	autoregressive moving average
a. s.	almost surely
CARMA	continuous-time autoregressive moving average
cf.	confer
CLT	Central Limit Theorem
e. g.	for example
ESD	empirical spectral distribution
Eq.	equation
et al.	et alii
i. e.	that is
i. i. d.	independent and identically distributed
LSD	limiting spectral distribution
RMT	random matrix theory
QML	quasi maximum likelihood