Technische Universität München
Fakultät für Mathematik
Lehrstuhl für Mathematische Statistik

Simulation-based estimation of time series and stochastic volatility processes

Thiago do Rêgo Sousa

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.

Vorsitzender:		Prof. Dr. Mattias Scherer
Prüfer der Dissertation:	1.	Prof. Dr. Claudia Klüppelberg
	2.	Prof. Dr. Denis Belomestny (Universität Duisburg-Essen)
	3.	Prof. Dr. Robert Stelzer (Universität Ulm)

Die Dissertation wurde am 01.07.2019 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 09.08.2019 angenommen.

Summary

This dissertation investigates simulation-based methods to estimate time series processes and Lévy-driven stochastic volatility models.

In the first chapter, we advocate the use of an Indirect Inference method to estimate the parameter of a COGARCH(1,1) process for equally spaced observations. This requires that the true model can be simulated and a reasonable estimation method for an approximate auxiliary model. We follow previous approaches and use linear projections leading to an auxiliary autoregressive model for the squared COGARCH(1,1) returns. The asymptotic theory of the Indirect Inference estimator relies on a uniform strong law of large numbers and asymptotic normality of the parameter estimates of the auxiliary model, which require continuity and differentiability of the COGARCH(1,1) process with respect to its parameter and which we prove via Kolmogorov's continuity criterion. This leads to consistent and asymptotically normal Indirect Inference estimates under moment conditions on the driving Lévy process. A simulation study shows that the method yields a substantial finite sample bias reduction compared with previous estimators.

In the second chapter we develop new estimators for general time series observations. We estimate the parameter of a time series process by minimizing the integrated weighted mean squared error between the empirical and simulated characteristic function, when the true characteristic functions cannot be explicitly computed. Motivated by Indirect Inference, we use a Monte Carlo approximation of the characteristic function based on iid simulated blocks. As a classical variance reduction technique, we propose the use of control variates for reducing the variance of this Monte Carlo approximation. These two approximations yield two new estimators that are applicable to a large class of time series processes. We show consistency and asymptotic normality of the parameter estimators under strong mixing, moment conditions, and smoothness of the simulated blocks with respect to its parameter. In a simulation study we show the good performance of these new simulation based estimators, and the superiority of the control variates based estimator for Poisson driven time series of counts.

Finally, the third chapter is dedicated to the estimation of multivariate COGARCH(1,1) processes (MUCOGARCH(1,1)). In order to apply an estimator based on control variates as developed in Chapter 2, it is crucial to know the second order structure of the MUCOGARCH(1,1) process in closed form. Moreover, using a moment based estimator as a benchmark estimator is also desirable. This describes the problems we study, where we apply the generalized method of moments to the MUCOGARCH(1,1) process. More specifically, we obtain explicit expressions for the second order structure of the "squared returns" process observed on a discrete-time grid with fixed grid size. Under moment and strong mixing conditions, we show that the resulting estimator is weak consistent and asymptotically normal. Sufficient conditions for strong mixing, stationarity and identifiability of the model parameter are also discussed. We investigate the finite sample behavior of the estimator in a simulation study.

Zusammenfassung

Diese Dissertation schlägt neue simulationsbasierte Methoden zur Schätzung von Zeitreihenprozessen und von Lévy-getriebenen stochastischen Volatilitätsmodellen vor und untersucht ihre Eigenschaften.

Im ersten Kapitel verwenden wir eine indirekten Inferenzmethode, um die Parameter eines COGARCH(1,1) Prozesses für Beobachtungen auf einem gleichmässigen Gitter zu schätzen. Dies setzt voraus, dass das wahre Modell simuliert werden kann und dass eine vernünftige Schätzmethode für ein geeignetes approximatives Hilfsmodell vorliegt. Wir folgen früheren Ansätzen und verwenden lineare Projektionen, die zu einem autoregressiven Hilfsmodell für die quadrierten COGARCH(1,1)-Renditen führen. Die asymptotische Theorie des indirekten Inferenzschätzers benutzt das gleichmässigen starken Gesetz der großen Zahlen und die asymptotischen Normalität der Parameterschätzer des Hilfsmodells. Beides erfordert Stetigkeit und Differenzierbarkeit des COGARCH(1,1)-Prozesses in seinen Parametern. Wir beweisen beides mittels des Stetigkeitskriteriums von Kolmogorov. Dies sichert stark konsistente und asymptotisch normale indirekte Inferenzschätzer unter Momentenbedingungen für den treibenden Lévy Prozess. Eine Simulationsstudie zeigt, dass die Methode im Vergleich zu früheren Schätzern zu einer erheblichen Reduzierung der Verzerrung bei endlichen Stichproben führt.

Im zweiten Kapitel entwickeln wir neue Schätzer für allgemeine Zeitreihenbeobachtungen. Wir schätzen den Parameter eines Zeitreihenprozesses durch Minimierung des integrierten gewichteten mittleren quadratischen Fehlers zwischen der empirischen und der simulierten charakteristischen Funktion für den Fall, dass die wahre charakteristische Funktion nicht explizit berechnet werden kann. Motiviert durch die indirekte Inferenzmethode verwenden wir eine Monte-Carlo-Approximation der charakteristischen Funktion auf der Basis von simulierten unabhängigen Blöcken. Als klassische Varianzreduktionstechnik schlagen wir die Verwendung von Kontrollvariablen vor, um die Varianz dieser Monte-Carlo-Approximation zu reduzieren. Beide Approximationen ergeben zwei neue Schätzer, die auf eine große Klasse von Zeitreihenprozessen anwendbar sind. Wir beweisen starke Konsistenz und asymptotische Normalität der Parameterschätzer unter Mischungs sowie Momentenbedingungen und Glattheit in Bezug auf den Parameter der simulierten Blöcke. In einer Simulationsstudie zeigen wir relevante Eigenschaften dieser neuen simulationsbasierten Schätzer und insbesondere die Überlegenheit des auf Kontrollvariablen basierenden Schätzers für Poisson-getriebene Zeitreihen von Zähldaten.

Schließlich beschäftigen wir uns im dritten Kapitel mit der Schätzung multivariater COGARCH(1,1)-Prozesse (MUCOGARCH(1,1)). Um einen Schätzer anzuwenden, der auf den in Kapitel 2 entwickelten Kontrollvariablen basiert, benötigt man die zweite Ordnungsstruktur des MUCOGARCH(1,1)-Prozesses in geschlossener Form. Darüber hinaus ist die Verwendung eines momentenbasierenden Schätzers als Benchmark wünschenswert. Wir wenden die verallgemeinerte Momentenmethode auf den MUCOGARCH(1,1)-Prozess an. Insbesondere erhalten wir explizite Ausdrücke für die zweite Ordnungsstruktur des quadrierten Renditeprozesses, der auf einem zeitdiskreten gleichmässigen Gitter beobachtet wird. Unter Mischungs sowie Momentenbedingungen zeigen wir, dass der resultierende Schätzer schwach konsistent und asymptotisch normal ist. Ausreichende Bedingungen für stark mischend, Stationarität und Identifizierbarkeit der Modellparameter werden ebenfalls diskutiert. In einer Simulationsstudie untersuchen wir das Verhalten des Schätzers für endliche Stichproben.

Acknowledgments

First of all, I want to express my gratitude to my supervisor Claudia Klüppelberg for her help, support and encouragement. I also learned a lot from her, both scientifically and personally and benefited greatly from numerous discussions, which were both fruitful and fun. No matter how challenging a topic was, I always felt motivated to try my best. She also enabled me to travel to various scientific meetings to present my work and brought me in contact with many distinguished researches. I am also grateful to the Brazilian council for scientific and technological development (CNPq-Brasil) for support through a nationwide scholarship program and also to the TUM Graduate school for its support.

I am also grateful to my co-author and mentor Stephan Haug, for his constant support and help during this years. Many thanks to my co-authors Robert Stelzer and Richard Davis, of whom I benefited greatly both scientifically and personally. In particular I would like to thank the Statistics Department at Columbia University, who made my visiting time in New York both fruitful and pleasant.

Furthermore, I am grateful to my colleagues at the Chair of Mathematical Statistic: Nadine, Andrea, Sven, Peter, Carson, Anita, Carlos, Johannes Klepsch, Johannes Buck and Thomas. Special thanks goes to Tyson for sharing so many office hours with me and many "snack times". Thanks also to Noam Berger for the courses on stochastic analysis and Markov processes. Vielen Dank auch an allen meine Freunde am Lehrstuhl für die Geduld und Bereitschaft mir die deutsche Sprache zu erklären (beim Kaffeetrinken). Das Lernen der deutschen Sprache und Kultur hat so viel Spaß gemacht, dass ich mich im Laufe der Jahre immer mehr zu Hause gefühlt habe. Ich bin dafür unendlich dankbar.

Thanks to all my friends I made during this time, specially Jean Louis, Tom, Ludovica, Andreas, Richard, Claudio, Julian, Isabel, Nelly, Mango, Almir, Mathias, Erica and Marc.

Meus eternos agradecimentos vão para meus pais Dorinha e Zenildo e meu irmão Junior, que sempre me incentivaram a continuar estudando, mesmo em situações difíceis. A minha noiva Larizza meu eterno agradecimento, por todo amor e compreensão em todas as fases desta jornada. Obrigado também: Leda, Raizza, Matos, Elisângela, Regina, Maria, Alzira, minha avô Maria de Jesus, tios, primos, Luis, membros da familia rocha, Douglas, Eurico, Cira, Silvia, Robson, Sgualdo, Saulo, Murilo, Marcel e Alisson Barros.

Introduction

Time Series

A set of observations recorded at specific times is called a time series. They appear in almost every area of applied science, some examples being population size, number of sunspots, temperature or rainfall data, disease counts, exchange rates or stock prices, just to name a few. For a detailed overview with a range of topics on time series analysis we refer to the classical monographs Box and Jenkins [11], Brockwell and Davis [13], Hamilton [43], Lütkepohl [68] or Shumway and Stoffer [94]. In order to understand the mechanism generating the time series, it is necessary to set up a hypothetical probability model to represent the data. In this thesis we focus on parametric models, whose behavior depends on a fixed parameter $\theta \in \Theta \subset \mathbb{R}^q$ for $q \in \mathbb{N}$. In order to draw inferences from the data, a crucial step is parameter estimation, which consists basically of finding the best parameter $\theta \in \Theta$ for which the chosen model fits the data well.

The estimation of time series processes is an ongoing problem of statistical inference. Maximum likelihood estimation (MLE) has been extensively used for parameter estimation, since under weak regularity conditions it is known to be asymptotically efficient. For many models, however, MLE is not always feasible to carry out, due to a likelihood that may be intractable to compute, or maximization of the likelihood is difficult.

To overcome such problems, alternative methods have been developed, for instance, the generalized method of moments (GMM) in Hansen [44], the quasi-maximum likelihood estimation (QMLE) in White [102], composite likelihood methods in Lindsay [66] and Indirect Inference in Gourieroux et al. [39] and Smith [96]. We follow the line of research of Indirect Inference.

Indirect Inference

Indirect Inference is a simulation-based technique, which requires only that the true model can be simulated and a reasonable estimation method for an approximate auxiliary model. Originally, it was introduced for complex econometric models to overcome the estimation problem of an intractable likelihood function, as e.g. for continuous time models with stochastic volatility observed on a grid of points. Indirect Inference can be used for instance as a vehicle to produce estimators which are robust, when there are outliers in the observations. It works as follows: let $X_1, \dots, X_n$ be the observations of the true model, with true parameter $\theta_0 \in \Theta$. We first choose an auxiliary model, with parameter $\pi \in \Pi \subset \mathbb{R}^r$ for $r \in \mathbb{N}$, which we believe is able to capture some features of the observed data, and that can be easily estimated. Then, we use these observations to compute the estimator $\hat{\pi}_n$ of the auxiliary parameter π. Next, for many different $\theta \in \Theta$ we simulate $K \geq 1$ independent samples of size n of the true model $(X_i^{(k)}(\theta))_{i=1}^n$ and compute the estimators $\hat{\pi}_{n,k}(\theta)$ for $k = 1, \dots, K$. The Indirect Inference estimator of θ is then defined as

$$\hat{\theta}_{n,\mathrm{II}} = \arg\min_{\theta \in \Theta} \left\| \hat{\pi}_n - \frac{1}{K} \sum_{k=1}^{K} \hat{\pi}_{n,k}(\theta) \right\|_{\Omega}, \qquad (0.0.1)$$

where Ω is a symmetric and positive definite weight matrix. The idea behind the method is that the estimation step via simulations is actually approximating the so-called link function $\theta \mapsto \pi(\theta)$, which is a map connecting the parameters spaces Θ and Π. If this map is one-to-one, it is natural to believe that whenever $\frac{1}{K}\sum_{k=1}^K \hat{\pi}_{n,k}(\theta)$ is close to $\hat{\pi}_n$, then, the Indirect Inference estimator $\hat{\theta}_{n,\mathrm{II}}$ will be closer to θ_0 (see also Figure 1).

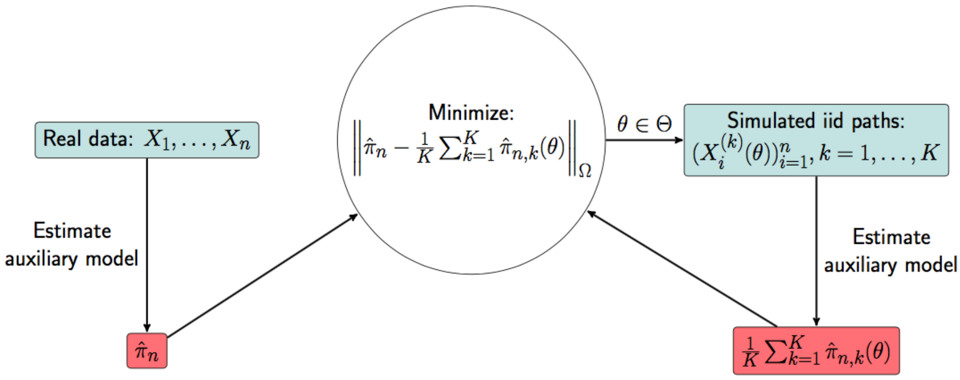


Figure 1: Indirect Inference diagram

The COGARCH(1,1) process

In this thesis, we apply the Indirect Inference estimator (0.0.1) to the continuous time GARCH(1,1) (COGARCH(1,1)) process introduced in Klüppelberg et al. [55]. The COGARCH(1,1) process generalizes the famous discrete time GARCH(1,1) process from Bollerslev [10] and Engle [29] in a natural way. GARCH processes and their extensions have found numerous applications in the field of finance. In particular, financial return data is usually characterized as uncorrelated, but not independent, heavy-tailed and with time-varying volatility, and GARCH processes exhibit such properties (the so-called stylized facts). However, over the years the analysis of stock price data in the finance industry has moved from analyzing monthly data, to intraday data. Additionally, stock market data is nowadays recorded for every transaction, and these occur irregularly in time. Therefore, it seems reasonable to consider stock prices in continuous time, giving rise to continuous time stochastic volatility models, among which the COGARCH is an important one. The COGARCH process is a stochastic volatility model defined as

$$P_t(\theta) = \int_0^t \sigma_s(\theta) \mathrm{d}L_s, \quad t \geq 0, \tag{0.0.2}$$

with parameter θ, L is a Lévy process with Lévy measure $\nu_L \not\equiv 0$ and having càdlàg sample paths. The volatility process $(\sigma_s(\theta))_{s\geq 0}$ is predictable and its stochasticity depends only on L.

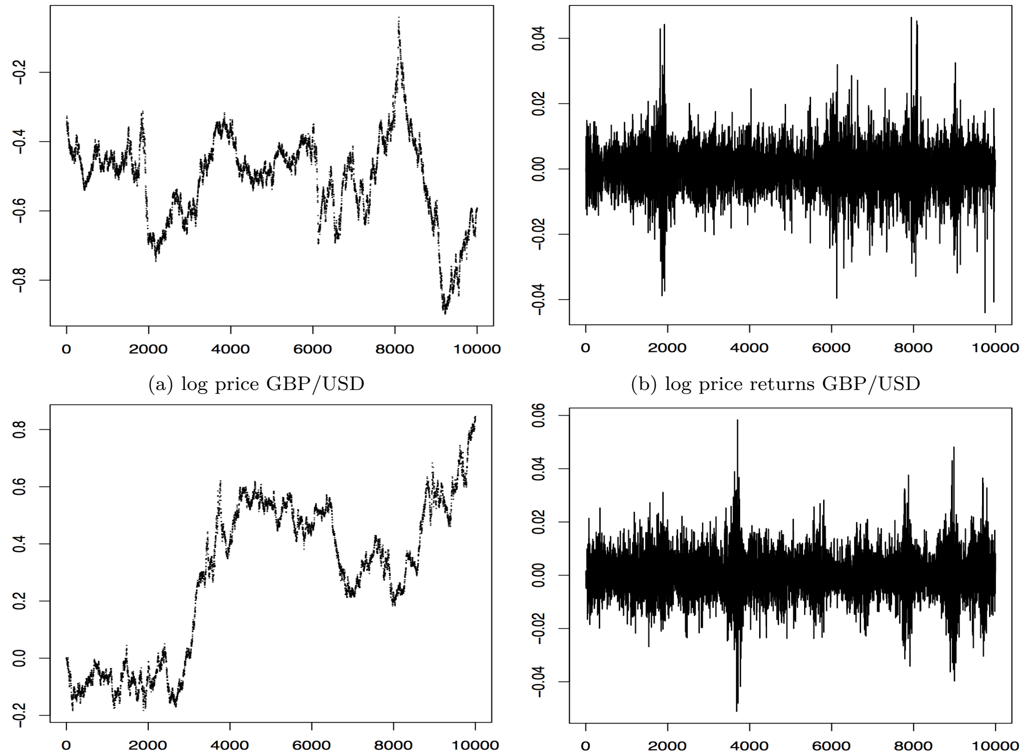
The COGARCH(1,1) process satisfies the stylized facts of financial returns and, as a continuous time model, is suited for modeling high-frequency data. Figure 2 shows a simulation for the COGARCH(1,1) price process and assesses empirically some properties of the log-returns process

$$G_i(\theta) := P_{\Delta i}(\theta) - P_{(i-1)\Delta}(\theta) = \int_{(i-1)\Delta}^{i\Delta} \sigma_s(\theta) \mathrm{d}L_s, \tag{0.0.3}$$

for $i = 1, \ldots, 10\,000$ and $\Delta = 1$. First of all, since the COGARCH(1,1) process is driven by a Lévy process, it moves away from the wrong Gaussianity in a natural way. Figures 2(a) and (b) plot the log price and returns for exchange rate data, namely GBP (British Pound)/USD (United States Dollar). In particular one sees some evidences of jumps and the formation of clusters of volatility in the returns. A nice property of the COGARCH process is that it also allows for jumps in the log price and volatility clustering in the log price returns (Figure 2(c) and (d)). Additionally, COGARCH (log) price returns are uncorrelated, while squared (log) prices returns are dependent.

Several methods have been proposed to estimate the parameter θ of a COGARCH(1,1) process. For this model, we advocate in do Rêgo Sousa et al. [27] the use of the Indirect Inference method to estimate its parameter. In particular, Indirect Inference has been shown in the literature to reduce finite sample bias (Gourieroux et al. [40, 41]), and this is our motivation. Our contribution here is the asymptotic theory of the Indirect Inference estimator for COGARCH(1,1) processes, which is completely new and, in verifying empirically that it yields a substantial finite sample bias reduction compared with previous estimators.

In order to prove consistency and asymptotic normality we need to verify that the COGARCH(1,1) process and the chosen auxiliary model satisfy some regularity condi-



(a) log price GBP/USD

(b) log price returns GBP/USD

(c) simulated log price $(P_i(\theta))$

(d) simulated log price returns $(G_i(\theta))$

Figure 2: COGARCH(1,1) process: L is a compound Poisson process with rate 1 and iid $N(0,1)$ jumps and $\Delta = 1$

tions. The starting point (A) is an appropriate auxiliary model that provides a one-to-one binding function. We follow previous approaches and use linear projections leading to an auxiliary autoregressive (AR) model of appropriate order for the squared COGARCH returns $(G_i^2(\theta))_{i\in\mathbb{N}}$ as in (0.0.3). Often the properties of the binding function are assessed via simulation, but for our models the binding function can be proved to be one-to-one.

Part (B), strong consistency and asymptotic normality of the estimator $\hat{\pi}_n$ of the AR model parameter π, is obtained in a similar way as in classical time series analysis (see e.g. Brockwell and Davis [13]), extending the theory to residuals, which may not be white noise, but an arbitrary stationary and ergodic process with finite variance. The strong law of large numbers and asymptotic normality of $\hat{\pi}_n$ will then be a consequence of the fact that $(G_i^2(\theta))_{i\in\mathbb{N}}$ is also strong mixing with appropriate mixing coefficients.

(C) is related to regularity conditions of the map $\theta \mapsto \hat{\pi}_n(\theta)$. To achieve strong consistency of the Indirect Inference estimator we need to show that

$$\sup_{\theta\in\Theta} \|\hat{\pi}_n(\theta) - \pi_\theta\| \overset{\text{a.s.}}{\to} 0, \quad n \to \infty.$$

To move from point-wise to uniform convergence we use a uniform strong law of large numbers in a compact parameter space Θ. This is applicable provided that $G_i(\theta)$ in (0.0.3) is a continuous function in θ and $\mathbb{E}\sup_{\theta\in\Theta} G_i^4(\theta) < \infty$ for all $i \in \mathbb{N}$. The continuity of this map does not follow directly from the continuity of $\sigma_s(\theta)$ for fixed s, because the Lévy process in the stochastic integral in (0.0.3) may have infinite variation. Under conditions on the moments and the characteristic exponent of the driving Lévy process, we find a version of $G_i(\theta)$ which is continuous by Kolmogorov's continuity criterion, and as a result we conclude strong consistency of the Indirect Inference estimator. A Taylor expansion of $\hat{\pi}_n(\theta)$ around the true parameter value θ_0 yields asymptotic normality by the delta method. This requires continuous differentiability of $G_i(\theta)$ in θ, which will follow from a result of Hutton and Nelson [47] together with Kolmogorov's continuity criterion. Later in this thesis we come back to a different Indirect Inference estimator, which can be applied to general time series processes.

Indirect Inference based on the empirical characteristic function and control variates

Instead of focusing on one specific model as we did previously, we develop next a new Indirect Inference estimator based on the empirical characteristic function and control variates that is applicable to a large class of time series processes.

The Indirect Inference estimator in (0.0.1) relies on an auxiliary model, which summarises the features of the true model with parameter θ in the parameter $\pi_\theta \in \mathbb{R}^r$ for $r \in \mathbb{N}$. A natural question to ask is, what happens if we decide to use an auxiliary model with more parameters, perhaps countably many, or even uncountably many? Would this result in better estimators?

This is one of the motivations for introducing an auxiliary model (auxiliary criterion) which is based on the p-dimensional characteristic function of the time series processes. In this case, the binding function would be of the form $\theta \mapsto \varphi_\theta$ where φ_θ is not a finite dimensional vector anymore, but instead a map from $\mathbb{R}^p$ to $\mathbb{C}$. This is the estimator defined in Knight and Yu [58], which we call the *oracle estimator* of θ_0 since it assumes that the binding function is known in closed form. More specifically, from the observed data, we construct the observed blocks

$$\boldsymbol{X}_j = (X_j, \dots, X_{j+p-1}), \quad j = 1, \dots, n,$$

we estimate

$$\varphi_n(t) = \frac{1}{n}\sum_{j=1}^{n} e^{i\langle t,\boldsymbol{X}_j\rangle}, \quad t \in \mathbb{R}^p, \tag{0.0.4}$$

and define the *oracle estimator*

$$\hat{\theta}_n = \mathrm{argmin}_{\theta\in\Theta} Q_n(\theta), \tag{0.0.5}$$

where

$$Q_n(\theta) = \int_{\mathbb{R}^p} |\varphi_n(t) - \varphi(t,\theta)|^2 w(t)\mathrm{d}t, \quad \theta \in \Theta, \tag{0.0.6}$$

with suitable weight function w such that the integral is well-defined, and the binding function is

$$\varphi(t,\theta) = \mathbb{E}e^{i\langle t,\boldsymbol{X}_1(\theta)\rangle}, \quad t \in \mathbb{R}^p. \tag{0.0.7}$$

In an ideal situation, the binding function $\theta \mapsto \varphi(\cdot,\theta)$ has an explicit expression and the estimator $\hat{\theta}_n$ in (2.2.3) can be used. What happens now, if the binding function is unknown? Following the ideas of Indirect Inference we approximate it by simulations and define a similar estimator.

In Davis et al. [24], we focus on time series processes for which the true characteristic function has no explicit expression and, approximate it by simulations. A natural way to approximate the binding function (0.0.7) is to proceed as we did in Indirect Inference, by simulating K iid paths of size n of the true model and for each of them compute (2.2.2) or by simulating one path of size greater than n as e.g. in Section 5.2 of Carrasco et al. [15]. But why do we proceed exactly in this way? At this point, we have the freedom to simulate the paths in various ways and approximate the binding function in the most sensible way. We choose to work with a large number of independent short paths, which we call blocks. In this way, we approximate the binding function using many iid blocks. While being unbiased, this approximation will generally have smaller variance than the approximation based on a few paths of the same length as the data, as this would result in dependent blocks. This is the motivation for the *simulation based parameter estimator* we propose here. More specifically, for many different $\theta \in \Theta$, we simulate, independent of the observed time series, an iid sample of blocks

$$\tilde{\boldsymbol{X}}_j(\theta) = (\tilde{X}_1^{(j)}(\theta), \ldots, \tilde{X}_p^{(j)}(\theta)), \quad j = 1, \ldots, H, \tag{0.0.8}$$

for $H \in \mathbb{N}$, and define the *Monte Carlo approximation* of $\varphi(\cdot,\theta)$ based on these simulations as

$$\varphi_H(t,\theta) = \frac{1}{H}\sum_{j=1}^{H} e^{i\langle t,\tilde{\boldsymbol{X}}_j(\theta)\rangle}, \quad t \in \mathbb{R}^p. \tag{0.0.9}$$

If we replace $\varphi(\cdot,\theta)$ in (2.2.4) by $\varphi_H(\cdot,\theta)$, we obtain the *simulation based parameter estimator*

$$\hat{\theta}_{n,H} = \arg\min_{\theta\in\Theta} Q_{n,H}(\theta), \tag{0.0.10}$$

where

$$Q_{n,H}(\theta) = \int_{\mathbb{R}^p} |\varphi_n(t) - \varphi_H(t,\theta)|^2 w(t)\mathrm{d}t, \tag{0.0.11}$$

with suitable weight function w such that the integral is well-defined. Indeed this gives a characteristic function approximation which yields, by minimizing the integrated distance, strongly consistent and asymptotically normal parameter estimators, which we prove. We also report their small sample properties for different models.

However, as the Monte Carlo approximation of the characteristic function is computed from iid blocks from a time series, control variates techniques provide an even more accurate approximation for the binding function. The idea of control variates is to use the knowledge of other quantities that can be computed from the model (e.g. certain moments) to get a better approximation for an unknown quantity. These known quantities are called control variates. We choose the first two terms in the Taylor expansion of the complex exponential $e^{i\langle t,\tilde{\boldsymbol{X}}_1(\theta)\rangle}$, $\langle t,\tilde{\boldsymbol{X}}_1(\theta)\rangle$ and $\langle t,\tilde{\boldsymbol{X}}_1(\theta)\rangle^2$ for $\theta\in\Theta$ as control variates. This requires knowing the mean and covariance matrix of $\tilde{\boldsymbol{X}}_1(\theta)$ for $\theta\in\Theta$, which is not a strong assumption for many time series processes.

In assessing the performance of both the Monte Carlo approximation and the control variates approximation of the characteristic function, two trends emerge. First, both the Monte Carlo and the control variates approximations work better for small values of the argument. Second, the control variates approximation performs much better than the Monte Carlo approximation, in particular, for small values of the argument. As a consequence, we propose a *control variates based parameter estimator* whose integrated mean squared error distance distinguishes between small and large values of the argument.

In a simulation study we show the good performance of these new simulation based estimators for two important models. The first one is the long-range dependence ARFIMA process. Long-range dependence models are characterized by having an autocovariance function that decays like a power function and they can be applied in numerous fields, including environmental and economic time series. Here, the *simulation based parameter estimator* already performs similarly to the *oracle estimator* in terms of bias and standard deviation, so there is no need to use control variates approximation. The second model is a Poisson driven time series of counts. It is a nonlinear model, which has been proposed in Zeger [106] and applied, for instance, for modeling disease counts (see also Campbell [14], Chan and Ledolter [16] and Davis et al. [21]). Here the binding function cannot be computed in closed form, so the *oracle estimator* is not applicable. The results show the superiority of the *control variates based parameter estimator*, when compared with the *simulation based parameter estimator*. When compared with the composite pairwise likelihood estimator in Davis and Yau [20], the *control variates based parameter estimator* has comparable or even smaller bias.

The idea of using control variates for getting better approximations of the characteristic function employed here and its good performance is a valuable result, which comes in addition to its usefulness for parameter estimation in time series processes. This combination could also be useful for approximating finite dimensional characteristic functions, distributions or densities of time series processes. Additionally, although we have focused here on time series processes, the *control variates based parameter estimator* developed here can also be applied to problems involving iid data for which the oracle estimator cannot be applied. One such situation is when estimating θ_0 from observations $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F_{\theta_0}$, where $(F_\theta, \theta \in \Theta)$ is some family of distributions parameterized by θ for which neither its density nor its characteristic function can be computed in closed-form.

The final part of this thesis deals with parameter estimation in a multivariate stochastic volatility model.

Parameter estimation for the MUCOGARCH(1,1) process

Finally, we investigate the problem of parameter estimation in the multivariate COGARCH(1,1) process (MUCOGARCH(1,1)) introduced in Stelzer [98]. It combines the features of the COGARCH(1,1) process with the ones of the multivariate BEKK GARCH(1,1) process of Engle and Kroner [30].

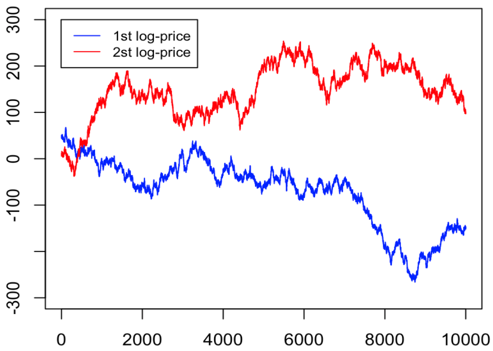
Multivariate models are necessary because in many areas of application, one has to model and understand the joint behavior of several time series. Therefore, the MUCOGARCH(1,1) is appropriate for modeling and understanding volatility and prices in several stocks and prices jointly. Market behavior is represented by a large portfolio of the joint $d-$dimensional stochastic process and it is defined as

$$P_t(\theta) = \int_0^t V_{s-}(\theta)^{1/2} \mathrm{d}L_s, \quad t \geq 0, \tag{0.0.12}$$

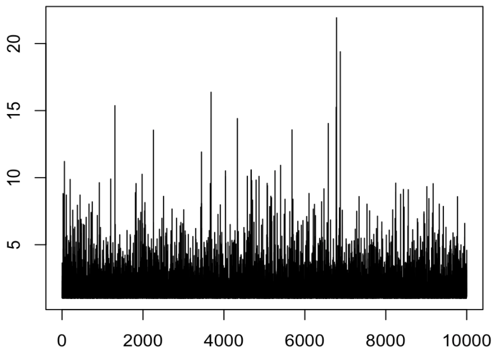
where L is a multivariate Lévy process in $\mathbb{R}^d$ with Lévy measure $\nu_L \not\equiv 0$ and having càdlàg sample paths. The matrix-valued MUCOGARCH(1,1) volatility process $(V_s)_{s\geq 0}$ depends on a parameter $\theta \in \Theta$, it is predictable and its stochasticity depends only on L.

Since it is based on a Lévy process, it allows for jumps on the volatility process and as well in the (log)-price process. The goal of the matrix process $(V_s)_{s\geq 0}$ is to model the volatility simultaneously in all components, and as well as dependence structure.

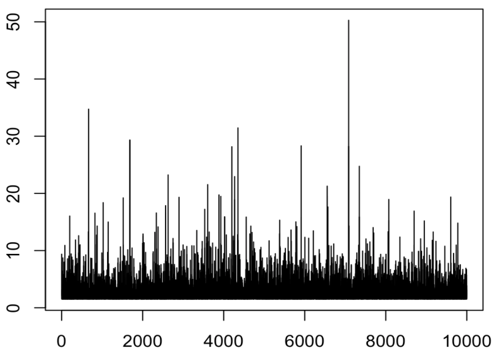
Individually, each (log) price component behaves like the COGARCH path shown in Figure 2, exhibiting volatility clustering in the (log) price returns, with (log) price returns uncorrelated, but not independent. We plot in Figure 3 the (log) price process for two components together with its volatility process. The diagonal entries of $(V_s)_{s\geq 0}$ (Figures 3(b) and (c)) correspond to the individual volatility of the first and second (log) price process from Figure 3(a), respectively. Their correlation is indicated by the off-diagonal entry of $(V_s)_{s\geq 0}$ (Figure 3(d)).



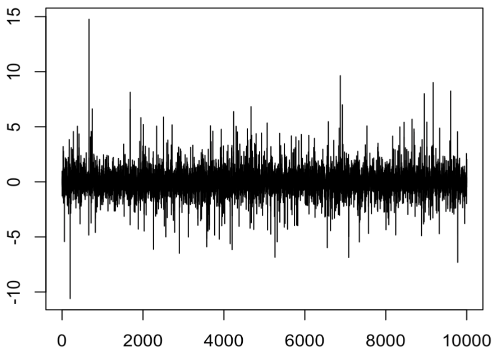
(a) bivariate (log) price $(P_i(\theta))$



(b) volatility $(V_{i,11}(\theta))$



(c) volatility $(V_{i,22}(\theta))$



(d) volatility $(V_{i,12}(\theta))$

Figure 3: MUCOGARCH(1,1) process: L is a bivariate compound Poisson process with rate 4 and iid $N(0, 1/4I_2)$ jumps, where I_2 is the 2-dimensional identity matrix

In order to apply an estimator based on control variates developed for the one dimensional model, it is crucial to know the second order structure of the MUCOGARCH(1,1) process in closed form. Moreover, using a moment based estimator as a benchmark estimator is also desirable. However, the second order structure of the 'squared returns' is not yet known.

Hence, our first challenge is to derive the second order structure of the squared returns in closed form, which already in the one-dimensional case, require complicated lengthy calculations. We accomplished this for the d-dimensional model by applying stochastic integration theory adapted for vector and matrix-valued process. In particular, we use integration by parts, Itô isometry, the compensation formula, and general inequalities to obtain upper bounds for moments of stochastic integrals with integrands involving $(V_s)_{s\geq 0}$ and integrators involving $(L_s)_{s\geq 0}$ and its quadratic variation.

Then, we apply the generalized method of moments (GMM) to the MUCOGARCH(1,1) process. Consistency and asymptotic normality of the GMM estimator is given under standard assumptions of mixing, existence of moments of the MUCOGARCH(1,1) volatility process and model identifiability. Then we give sufficient conditions under which these assumptions will hold.

We use the sufficient conditions for mixing given in Stelzer and Vestweber [99] and the conditions for asymptotic second order stationarity in Stelzer [98] to obtain consistency of the GMM estimator under rather general conditions. We also obtain asymptotic normality of the estimator under appropriate additional moment restrictions on the driving Lévy process.

The identifiability question is more delicate, since the formulas for the second order structure of the (log) price returns involve operators which are not invertible and, therefore, the strategy used for showing identifiability as in the one-dimensional COGARCH(1,1) process cannot be applied. Instead, we derive identifiability conditions which rely mainly on the autocovariance structure of the squared returns.

This describes the contribution of the paper do Rêgo Sousa and Stelzer [26], where we compute and prove the asymptotic properties of a generalized method of moment estimator for the parameters of MUCOGARCH(1,1) process.

Final remarks

The thesis contains three chapters, which are self-contained, with their own introduction. Notations and abbreviations may differ among them, since different notations and abbreviations seem reasonable in different contexts. The chapters are based on a publication (Chapter 1), a submitted preprint (Chapter 2) and a work in progress (Chapter 3), respectively:

- Chapter 1 is based on the paper do Rêgo Sousa et al. [27] that is published as:

  T. do Rêgo Sousa, S. Haug, and C. Klüppelberg. Indirect Inference for Lévy-driven continuous-time GARCH models. *Scandinavian Journal of Statistics*, 2019. To appear.

- Chapter 2 is based on the paper Davis et al. [24] that is submitted for publication as:

R.A. Davis, T. do Rêgo Sousa, and C. Klüppelberg. Indirect Inference for time series using the empirical characteristic function and control variates. 2019. Submitted. arXiv: 1904.08276.

- Chapter 3 is based on the paper do Rêgo Sousa and Stelzer [26] that is work in progress:

  T. do Rêgo Sousa and R. Stelzer. Method of moment based estimation for the multivariate COGARCH(1,1) process. 2019. In preparation.

Contents

Chapter 1: Indirect Inference for Lévy-driven continuous-time GARCH models

1.1 Introduction

The COGARCH(1,1) process was introduced in Klüppelberg et al. [55] as a continuous time analog of the discrete time GARCH(1,1) process. It is defined as

$$P_t(\boldsymbol{\theta}) = \int_0^t \sigma_s(\boldsymbol{\theta}) \mathrm{d}L_s, \quad t \geq 0, \tag{1.1.1}$$

with parameter $\boldsymbol{\theta}$ (to be specified in Section 2.2), L is a Lévy process with Lévy measure $\nu_L \not\equiv 0$ and having càdlàg sample paths. The volatility process $(\sigma_s(\boldsymbol{\theta}))_{s\geq 0}$ is predictable and its stochasticity depends only on L. The COGARCH process satisfies many stylized features of financial time series and is suited for modeling high-frequency data (see Bayracı and Ünal [2], Bibbona and Negri [7], Haug et al. [45], Klüppelberg et al. [57], Maller et al. [70], and Müller [76]).

In many practical problems, one observes the log-price process $(P_{i\Delta}(\boldsymbol{\theta}_0))_{i=1}^n$ on a fixed grid of size $\Delta > 0$ and the question of interest is how to estimate the true parameter $\boldsymbol{\theta}_0$. The data used for estimation are returns $(G_i(\boldsymbol{\theta}_0))_{i=1}^n$, where

$$G_i(\boldsymbol{\theta}_0) := P_{\Delta i}(\boldsymbol{\theta}_0) - P_{(i-1)\Delta}(\boldsymbol{\theta}_0) = \int_{(i-1)\Delta}^{i\Delta} \sigma_s(\boldsymbol{\theta}_0) \mathrm{d}L_s. \tag{1.1.2}$$

Several methods have been proposed to estimate the parameter of a COGARCH process. A method of moments was proposed in Haug et al. [45], Bibbona and Negri [7] used prediction based estimation as developed in Sørensen [97], and Maller et al. [70] proposed a pseudo maximum likelihood (PML) method which also works for non-equally spaced observations. Both moment and prediction based estimators are consistent and asymptotically normal under certain regularity conditions. The asymptotic properties of the PML estimator were studied in Iannace [49] and in Kim and Lee [53], which require that $\Delta \downarrow 0$ as $n \to \infty$. For the COGARCH process, Bayracı and Ünal [2] used Indirect Inference with an auxiliary discrete-time GARCH model with Gaussian residuals. No theoretical results were proved, but their simulation study suggests that Indirect Inference estimators achieve a similar performance as the PML estimator of Maller et al. [70] for fixed $\Delta > 0$. Furthermore, Müller [76] proposed a Markov chain Monte Carlo method, when L is a compound Poisson process.

In this paper we advocate an Indirect Inference method, different to the one suggested in Bayracı and Ünal [2], to estimate the COGARCH parameter and derive the asymptotic properties of the estimator. Such methods were introduced in Smith [95] and generalized in Gourieroux et al. [39], and they offer a way to overcome many estimation problems by a clever simulation method. In short, it only requires that the true model can be simulated and a reasonable estimation method for an approximate auxiliary model.

Indirect Inference was originally introduced for complex econometric models to overcome the estimation problem of an intractable likelihood function, as for continuous time models with stochastic volatility (see Bianchi and Cleur [6], Jiang [51], Laurini and Hotta [63], Raknerud and Skare [88], and Wahlberg et al. [101]). Indirect Inference can also be used as a vehicle to produce estimators which are robust, when there are outliers in the observations (see de Luna and Genton [25] for robust estimation of a discrete time ARMA and Fasen-Hartmann and Kimmig [31] of a continuous time ARMA). Another motivation is given in Gourieroux et al. [40, 41], where it is shown that Indirect Inference can reduce the finite sample bias considerably. This is our motivation to study the asymptotic properties of Indirect Inference estimators (IIE) in the context of COGARCH estimation.

The Indirect Inference procedure works as follows. Let $\boldsymbol{\pi}$ denote the parameter of an auxiliary model chosen for the COGARCH returns $(G_i(\boldsymbol{\theta}_0))_{i=1}^n$ or some transformed random variables. From this data we estimate $\boldsymbol{\pi}$ and obtain $\hat{\boldsymbol{\pi}}_n$. For many different $\boldsymbol{\theta} \in \Theta$ we simulate $K \geq 1$ independent samples of size n of COGARCH returns $(G_i^{(k)}(\boldsymbol{\theta}))_{i=1}^n$ and compute the estimators $\hat{\boldsymbol{\pi}}_{n,k}(\boldsymbol{\theta})$ for $k = 1, \ldots, K$. The IIE of $\boldsymbol{\theta}$ is then defined as

$$\hat{\boldsymbol{\theta}}_{n,\mathrm{II}} := \operatorname*{arg\,min}_{\boldsymbol{\theta} \in \Theta} \Big(\hat{\boldsymbol{\pi}}_n - \frac{1}{K} \sum_{k=1}^{K} \hat{\boldsymbol{\pi}}_{n,k}(\boldsymbol{\theta}) \Big)^{\top} \boldsymbol{\Omega} \Big(\hat{\boldsymbol{\pi}}_n - \frac{1}{K} \sum_{k=1}^{K} \hat{\boldsymbol{\pi}}_{n,k}(\boldsymbol{\theta}) \Big), \tag{1.1.3}$$

where $\boldsymbol{\Omega}$ is a symmetric and positive definite weight matrix. Under certain regularity conditions, IIEs are consistent and asymptotically normal. These regularity conditions are mainly related to three aspects: (A) find an auxiliary model whose parameter is connected to the COGARCH parameter through a one-to-one binding function, (B) prove strong consistency and asymptotic normality of $\hat{\boldsymbol{\pi}}_n$, and (C) prove that the estimator $\hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta})$, as a function of $\boldsymbol{\theta}$, satisfies conditions for the application of a uniform strong law of large numbers (SLLN) and a delta method for the asymptotic normality.

The starting point (A) is an appropriate auxiliary model that provides a one-to-one binding function. We follow previous approaches and use linear projections leading to an auxiliary autoregressive (AR) model of appropriate order for the squared COGARCH returns $(G_i^2(\boldsymbol{\theta}))_{i\in\mathbb{N}}$. Often the properties of the binding function are assessed via simulation (see Garcia et al. [37] and Lombardi and Calzolari [67]), but for our models the binding function can be proved to be one-to-one.

Part (B), strong consistency and asymptotic normality of the estimator $\hat{\boldsymbol{\pi}}_n$ of the AR model parameter $\boldsymbol{\pi}$, is obtained in a similar way as in classical time series analysis (see e.g. Brockwell and Davis [13]), extending the theory to residuals, which may not be white noise, but an arbitrary stationary and ergodic process with finite variance. The SLLN and asymptotic normality of $\hat{\boldsymbol{\pi}}_n$ will then be a consequence of the fact that $(G_i^2(\boldsymbol{\theta}))_{i\in\mathbb{N}}$ is also strong mixing with appropriate mixing coefficients.

(C) is related to regularity conditions of the map $\boldsymbol{\theta} \mapsto \hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta})$. To achieve strong consistency of the IIE we need to show that

$$\sup_{\boldsymbol{\theta}\in\Theta} \|\hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta}) - \boldsymbol{\pi}_{\boldsymbol{\theta}}\| \overset{\text{a.s.}}{\to} 0, \quad n \to \infty.$$

To move from point-wise to uniform convergence we use a uniform SLLN in a compact parameter space Θ. For the estimator we study here, the application of a uniform SLLN holds provided that $G_i(\boldsymbol{\theta})$ is a continuous function in $\boldsymbol{\theta}$ and $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} G_i^4(\boldsymbol{\theta}) < \infty$ for all $i \in \mathbb{N}$. The continuity of this map does not follow directly from the continuity of $\sigma_s(\boldsymbol{\theta})$ for fixed s, because the Lévy process in the stochastic integral in (1.1.2) may have infinite variation. Under conditions on the moments and the characteristic exponent of the driving Lévy process, we find a version of $G_i(\boldsymbol{\theta})$ which is continuous by Kolmogorov's continuity criterion, and as a result we conclude strong consistency of the IIE $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$. A Taylor expansion of $\hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta})$ around the true parameter $\boldsymbol{\theta}_0$ yields asymptotic normality by the delta method. This requires continuous differentiability of $G_i(\boldsymbol{\theta})$ in $\boldsymbol{\theta}$, which will follow from a result of Hutton and Nelson [47] together with Kolmogorov's continuity criterion.

Our paper is organised as follows. We start in Section 2 with the formal definition of a stationary COGARCH process as returns process, and recall its relevant properties. We also present the autoregressive (AR) auxiliary model of the squared returns and define the least squares estimator (LSE) and Yule-Walker estimator (YWE) of the AR parameter, as well as the binding function giving the link to the COGARCH parameter. In Section 2.3 we present the IIE and the conditions, which guarantee a uniform SLLN and asymptotic normality of the IIE. In Section 3 we prove strong consistency and asymptotic normality of the LSE and YWE under the non-standard conditions of stationary ergodicity and a mixing property. Section 4 is dedicated to strong consistency and asymptotic normality of the IIE of the COGARCH process. Section 5 presents a simulation study and shows that the bias reduction based on the IIE is indeed substantial compared to previous estimators. Technical results like conditions for the existence of a version of the COGARCH returns, which is continuous in its parameter and other auxiliary results are summarized in an Appendix.

Throughout we write $\|\cdot\|$ for the ℓ^1-norm in $\mathbb{R}^d$ for $d \in \mathbb{N}$ and recall that in $\mathbb{R}^d$ all norms are equivalent. For a matrix $A \in \mathbb{R}^{p\times q}$ we also write $\|A\|$ for the matrix norm generated by the ℓ^1-norm. For a vector $x \in \mathbb{R}^d$ and a $d \times d$ positive definite matrix $\mathbf{\Omega}$ we write $\|x\|_{\mathbf{\Omega}} = x^\top \Omega x$. Furthermore, we denote by $\mathcal{L}^p$ the space of p-integrable random variables, and by $\dim(A)$ the dimension of a subset A of $\mathbb{R}^d$. For a function $f(\boldsymbol{\theta})$ in $\mathbb{R}$ with $\boldsymbol{\theta} \in \mathbb{R}^q$ the gradient with respect to $\boldsymbol{\theta}$ is $\nabla_\theta f(\boldsymbol{\theta}) = (\frac{\partial}{\partial \boldsymbol{\theta}_l} f(\boldsymbol{\theta}))_{l=1}^q \in \mathbb{R}^q$, and $\nabla^2_\theta f(\boldsymbol{\theta}) = (\frac{\partial^2}{\partial \boldsymbol{\theta}_k \boldsymbol{\theta}_l} f(\boldsymbol{\theta}))_{k,l=1}^q \in \mathbb{R}^{q\times q}$ denotes the Hessian matrix.

1.2 COGARCH process, auxiliary autoregressive representation and Indirect Inference Estimation

1.2.1 Definition of the COGARCH process

For the parameter space of the COGARCH process given as $\{\boldsymbol{\theta} = (\beta, \eta, \varphi)^\top : \beta, \eta, \varphi > 0\}$, we construct a strictly stationary version of the volatility process as in Klüppelberg et al. [55]. First define the process $(Y_s(\boldsymbol{\theta}))_{s\geq 0}$ by

$$Y_s(\boldsymbol{\theta}) := \eta s - \sum_{0<u\leq s} \log(1 + \varphi(\Delta L_u)^2), \quad s \geq 0, \tag{1.2.1}$$

with Laplace transform $\mathbb{E}e^{-pY_s(\boldsymbol{\theta})} = e^{s\Psi_{\boldsymbol{\theta}}(p)}$, where

$$\Psi_{\boldsymbol{\theta}}(p) = -p\eta + \int_{\mathbb{R}} ((1 + \varphi x^2)^p - 1)\nu_L(\mathrm{d}x), \quad p \geq 0. \tag{1.2.2}$$

We shall often use the fact that for $p > 0$ by Lemma 4.1(a) in [55],

$$\mathbb{E}|L_1|^{2p} < \infty \quad \text{if and only if} \quad |\Psi_{\boldsymbol{\theta}}(p)| < \infty.$$

Define the volatility process $(\sigma^2_t(\boldsymbol{\theta}))_{t\geq 0}$ by

$$\sigma^2_t(\boldsymbol{\theta}) := \Big(\beta \int_0^t e^{Y_s(\boldsymbol{\theta})}\mathrm{d}s + \sigma^2_0(\boldsymbol{\theta})\Big) e^{-Y_{t-}(\boldsymbol{\theta})}, \quad t \geq 0, \tag{1.2.3}$$

where $Y_{t-}(\boldsymbol{\theta})$ denotes the left limit at t and $\sigma^2_0(\boldsymbol{\theta})$ the starting value of the volatility process. If $\mathbb{E}|L_1|^2 < \infty$ and $\Psi_{\boldsymbol{\theta}}(1) < 0$, then by Lemma 4.1(c) of [55], $\sigma^2_t(\boldsymbol{\theta}) \xrightarrow{\mathrm{d}} \sigma^2_\infty(\boldsymbol{\theta})$ as $t \to \infty$, where

$$\sigma^2_\infty(\boldsymbol{\theta}) \overset{\mathrm{d}}{=} \beta \int_0^\infty e^{-Y_s(\boldsymbol{\theta})}\mathrm{d}s.$$

Setting the starting value as

$$\sigma_0^2(\boldsymbol{\theta}) \stackrel{\mathrm{d}}{=} \beta \int_0^\infty e^{-Y_s(\boldsymbol{\theta})} \mathrm{d}s, \quad \text{independent of } L, \tag{1.2.4}$$

by Theorem 3.2 of [55] for such $\boldsymbol{\theta}$ the process $(\sigma_t^2(\boldsymbol{\theta}))_{t\geq 0}$ is strictly stationary. Then by Proposition 4.2 of [55] for the stationary process and $k \in \mathbb{N}$,

$$\mathbb{E}\sigma_0^{2k}(\boldsymbol{\theta}) < \infty \quad \text{if and only if} \quad \mathbb{E}L_1^{2k} < \infty \text{ and } \Psi_{\boldsymbol{\theta}}(k) < 0. \tag{1.2.5}$$

Furthermore, for $k = 1, 2$ either of this implies that the squared returns from (1.1.2) have corresponding finite moments (Proposition 5.1 of [55]). Additionally, by Corollary 3.1 of [55] the process $(P_t(\boldsymbol{\theta}))_{t\geq 0}$ defined in (3.1.1) with stationary $(\sigma_t(\boldsymbol{\theta}))_{t\geq 0}$ has stationary increments.

1.2.2 Autoregressive representation for the squared returns

We estimate the COGARCH parameter, when the log-price process is observed on a regular grid of fixed size $\Delta > 0$ based on the returns $(G_i(\boldsymbol{\theta}))_{i\in\mathbb{N}}$ as defined in (1.1.2).

We state the basic assumptions and recall some properties of the COGARCH process.

Proposition 1.2.1 (Theorems 3.1 and 3.4 in Haug et al. [45])**.** *Assume that the following hold:*

(A1) The parameter vector $\boldsymbol{\theta} = (\beta, \eta, \varphi)^\top$ *satisfies* $\beta, \eta, \varphi > 0$.

(A2) $\mathbb{E}L_1 = 0$ *and* $\mathrm{Var}L_1 = 1$.

(A3) The variance c_L *of the Brownian component of* L *is known and satisfies* $0 \leq c_L < \mathrm{Var}L_1$.

(A4) $\mathbb{E}L_1^4 < \infty$.

(A5) $\int_{\mathbb{R}} x^3 \nu_L(\mathrm{d}x) = 0$.

(A6) $\Psi_{\boldsymbol{\theta}}(2) < 0$.

Denote the expectation and variance of the squared returns process by

$$\mu_{\boldsymbol{\theta}} = \mathbb{E}G_1^2(\boldsymbol{\theta}) \quad \textit{and} \quad \gamma_{\boldsymbol{\theta}}(0) = \mathrm{Var}G_1^2(\boldsymbol{\theta}).$$

Then the following assertions hold:

(a) The autocovariance function of the squared returns process is given by

$$\gamma_{\boldsymbol{\theta}}(h) = \mathrm{Cov}(G_i^2(\boldsymbol{\theta}), G_{i+h}^2(\boldsymbol{\theta})) = \gamma_{\boldsymbol{\theta}}(0) k_{\boldsymbol{\theta}} e^{-h\rho_{\boldsymbol{\theta}}}, \quad h \in \mathbb{N}. \tag{1.2.6}$$

(b) If $\mu_{\boldsymbol{\theta}}, \gamma_{\boldsymbol{\theta}}(0), k_{\boldsymbol{\theta}}, \rho_{\boldsymbol{\theta}} > 0$, *then these parameters uniquely determine* $\boldsymbol{\theta}$.

(c) The process $(G_i(\boldsymbol{\theta}))_{i\in\mathbb{N}}$ *is* α*-mixing with exponentially decaying mixing coefficients.*

Assume that the driving Lévy process satisfies assumptions (A2)-(A5) of Proposition 1.2.1. We take as parameter space of the COGARCH process a compact set Θ satisfying the relevant conditions of Proposition 1.2.1; more precisely,

$$\Theta \subset \mathcal{M} := \{\boldsymbol{\theta} = (\beta, \eta, \varphi)^\top : \beta, \eta, \varphi > 0, \Psi_{\boldsymbol{\theta}}(2) < 0 \text{ and } \mu_{\boldsymbol{\theta}}, \gamma_{\boldsymbol{\theta}}(0), k_{\boldsymbol{\theta}}, \rho_{\boldsymbol{\theta}} > 0\}. \tag{1.2.7}$$

In what follows, we denote the true model parameter by $\boldsymbol{\theta}_0 \in \Theta$. We present the auxiliary autoregressive model using the structure of COGARCH squared returns. Define the centered squared returns for $\boldsymbol{\theta} \in \Theta$ as

$$\tilde{G}_i^2(\boldsymbol{\theta}) := G_i^2(\boldsymbol{\theta}) - \mu_{\boldsymbol{\theta}}, \quad i \in \mathbb{N}. \tag{1.2.8}$$

Proposition 1.2.2 (Auxiliary AR(r) model)**.** *Let $\boldsymbol{\theta} \in \Theta$ and $r \geq 2$ be fixed. Define*

$$U_i(\boldsymbol{\theta}) := \tilde{G}_{i+r}^2(\boldsymbol{\theta}) - P_{\mathcal{H}_i}\tilde{G}_{i+r}^2(\boldsymbol{\theta}), \quad i \in \mathbb{N},$$

where $\mathcal{H}_i = \overline{\text{sp}}\{\tilde{G}_{i+r-j}^2(\boldsymbol{\theta}), j = 1, \dots, r\}$ is the closed span in the Hilbert space $\mathcal{L}^2$ and $P_{\mathcal{H}_i}$ the projection on $\mathcal{H}_i$. Then there exist unique real numbers $a_{\boldsymbol{\theta},1}, \dots, a_{\boldsymbol{\theta},r}$ such that

$$U_i(\boldsymbol{\theta}) = \tilde{G}_{i+r}^2(\boldsymbol{\theta}) - \sum_{j=1}^{r} a_{\boldsymbol{\theta},j}\tilde{G}_{i+r-j}^2(\boldsymbol{\theta}), \quad i \in \mathbb{N}. \tag{1.2.9}$$

Moreover, the process $(U_i(\boldsymbol{\theta}))_{i\in\mathbb{N}}$ is strictly stationary with $\mathbb{E}U_i(\boldsymbol{\theta}) = 0$ and $\text{Var}U_i(\boldsymbol{\theta}) < \infty$.

Proof. We adapt the proof of Proposition 2.2 of Fasen-Hartmann and Kimmig [31] for the COGARCH process. Since $\boldsymbol{\theta} \in \Theta \subset \mathcal{M}$, by Proposition 1.2.1(a), the autocovariance function of $(\tilde{G}_i^2(\boldsymbol{\theta}))_{i\in\mathbb{N}}$ satisfies $\gamma_{\boldsymbol{\theta}}(0) > 0$ and $\gamma_{\boldsymbol{\theta}}(h) \to 0$ as $n \to \infty$. By Proposition 5.1.1 of Brockwell and Davis [13] it follows that the autocovariance matrix of $(\tilde{G}_i^2(\boldsymbol{\theta}))_{i=1}^r$ is non-singular. Hence, the numbers $a_{\boldsymbol{\theta},1}, \dots, a_{\boldsymbol{\theta},r}$ are uniquely given by

$$\begin{pmatrix} a_{\boldsymbol{\theta},1} \\ a_{\boldsymbol{\theta},2} \\ \vdots \\ a_{\boldsymbol{\theta},r} \end{pmatrix} = \begin{pmatrix} \gamma_{\boldsymbol{\theta}}(0) & \gamma_{\boldsymbol{\theta}}(1) & \dots & \gamma_{\boldsymbol{\theta}}(r-1) \\ \gamma_{\boldsymbol{\theta}}(1) & \gamma_{\boldsymbol{\theta}}(0) & \dots & \gamma_{\boldsymbol{\theta}}(r-2) \\ \vdots & \vdots & & \vdots \\ \gamma_{\boldsymbol{\theta}}(r-1) & \gamma_{\boldsymbol{\theta}}(r-2) & \dots & \gamma_{\boldsymbol{\theta}}(0) \end{pmatrix}^{-1} \begin{pmatrix} \gamma_{\boldsymbol{\theta}}(1) \\ \gamma_{\boldsymbol{\theta}}(2) \\ \vdots \\ \gamma_{\boldsymbol{\theta}}(r) \end{pmatrix} \tag{1.2.10}$$

leading to (1.2.9). □

Proposition 1.2.2 gives an AR(r) representation for $r \geq 2$ for the COGARCH squared returns from (1.2.8) by rewriting (1.2.9) as $\tilde{G}_{i+r}^2(\boldsymbol{\theta}) = \sum_{j=1}^r a_{\boldsymbol{\theta},j}\tilde{G}_{i+r-j}^2(\boldsymbol{\theta}) + U_i(\boldsymbol{\theta})$ for $i \in \mathbb{N}$. Let

$$\boldsymbol{\pi}_{\boldsymbol{\theta}} := (\mu_{\boldsymbol{\theta}}, \boldsymbol{a}_{\boldsymbol{\theta}}, \gamma_{\boldsymbol{\theta}}(0))^\top = (\mu_{\boldsymbol{\theta}}, a_{\boldsymbol{\theta},1}, \dots, a_{\boldsymbol{\theta},r}, \gamma_{\boldsymbol{\theta}}(0))^\top, \tag{1.2.11}$$

and let $C \subset \mathbb{R}^r$ be a compact subset of the set containing all possible real coefficients of a strictly stationary AR(r) process. Then we define a compact parameter space of the auxiliary model as

$$\Pi := \Big[-\frac{1}{\epsilon}, \frac{1}{\epsilon}\Big] \times C \times \Big[\epsilon, \frac{1}{\epsilon}\Big], \tag{1.2.12}$$

where ϵ is a small positive constant.

We will investigate two well-known estimators of $\boldsymbol{\pi}_{\boldsymbol{\theta}}$ in (1.2.11), namely the least squares estimator (LSE) and the Yule-Walker estimator (YWE) defined by

$$\hat{\boldsymbol{\pi}}_{n,\mathrm{LS}}(\boldsymbol{\theta}) = \begin{pmatrix} \hat{\mu}_n(\boldsymbol{\theta}) \\ \hat{\boldsymbol{a}}_{n,\mathrm{LS}}(\boldsymbol{\theta}) \\ \hat{\gamma}_n(0;\boldsymbol{\theta}) \end{pmatrix} \quad \text{and} \quad \hat{\boldsymbol{\pi}}_{n,\mathrm{YW}}(\boldsymbol{\theta}) = \begin{pmatrix} \hat{\mu}_n(\boldsymbol{\theta}) \\ \hat{\boldsymbol{a}}_{n,\mathrm{YW}}(\boldsymbol{\theta}) \\ \hat{\gamma}_n(0;\boldsymbol{\theta}) \end{pmatrix}, \tag{1.2.13}$$

respectively, whose components are given as follows.

Definition 1.2.3. *The estimators of the mean and variance are given by*

$$\hat{\mu}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} G_i^2(\boldsymbol{\theta}) \quad \textit{and} \quad \hat{\gamma}_n(0;\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}(G_i^2(\boldsymbol{\theta}) - \hat{\mu}_n(\boldsymbol{\theta}))^2.$$

(a) The LSE of $(a_{\boldsymbol{\theta},1}, \ldots, a_{\boldsymbol{\theta},r})^\top$ *is given by*

$$\hat{\boldsymbol{a}}_{n,\mathrm{LS}}(\boldsymbol{\theta}) = \arg\min_{\boldsymbol{c}\in C} S_n(\boldsymbol{c};\boldsymbol{\theta}),$$

for C as in (1.2.12)*, and*

$$S_n(\boldsymbol{c};\boldsymbol{\theta}) := \frac{1}{n-r}\sum_{i=1}^{n-r}\Big((G_{i+r}^2(\boldsymbol{\theta}) - \hat{\mu}_n(\boldsymbol{\theta})) - c_1(G_{i+r-1}^2(\boldsymbol{\theta}) - \hat{\mu}_n(\boldsymbol{\theta})) - \cdots - c_r(G_i^2(\boldsymbol{\theta}) - \hat{\mu}_n(\boldsymbol{\theta}))\Big)^2.$$

(b) The YWE of $(a_{\boldsymbol{\theta},1}, \ldots, a_{\boldsymbol{\theta},r})^\top$ *is given by*

$$\hat{\boldsymbol{a}}_{n,\mathrm{YW}}(\boldsymbol{\theta}) = \hat{\boldsymbol{\Gamma}}_n^{-1}(\boldsymbol{\theta})\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\theta}), \quad n \in \mathbb{N}, \tag{1.2.14}$$

where $\hat{\boldsymbol{\Gamma}}_n^{-1}(\boldsymbol{\theta}) = (\hat{\gamma}_n(i-j;\boldsymbol{\theta}))_{i,j=1}^r$ *and* $\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\theta}) = (\hat{\gamma}_n(1;\boldsymbol{\theta}), \ldots, \hat{\gamma}_n(r;\boldsymbol{\theta}))^\top$ *are defined in terms of the empirical autocovariance function*

$$\hat{\gamma}_n(h;\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n-h}(G_i^2(\boldsymbol{\theta}) - \hat{\mu}_n(\boldsymbol{\theta}))(G_{i+h}^2(\boldsymbol{\theta}) - \hat{\mu}_n(\boldsymbol{\theta})), \quad h, n \in \mathbb{N}, n > h.$$

We now define a function that will connect the COGARCH process to its auxiliary AR model from Proposition 1.2.2.

Proposition 1.2.4 (Binding function). *Define the binding function* $\pi : \Theta \to \Pi$ *by* $\pi(\boldsymbol{\theta}) = \boldsymbol{\pi}_{\boldsymbol{\theta}}$ *as in* (1.2.11). *Then* π *is injective and continuously differentiable for* $r \geq 2$.

Proof. As in the proof of Lemma 2.5 in Fasen-Hartmann and Kimmig [31], we decompose $\pi : \Theta \to \Pi$ into three maps $\pi = \pi_1 \circ \pi_2 \circ \pi_3$. Define $\pi_1 : \Theta \to \mathbb{R}^4$ by

$$\pi_1(\boldsymbol{\theta}) = (\mu_{\boldsymbol{\theta}}, k_{\boldsymbol{\theta}}, \rho_{\boldsymbol{\theta}}, \gamma_{\boldsymbol{\theta}}(0))^\top,$$

which is by Proposition 1.2.1(b) injective. Next define $\pi_2 : \pi_1(\Theta) \to \mathbb{R}^{r+2}$ by

$$\pi_2(\mu_{\boldsymbol{\theta}}, k_{\boldsymbol{\theta}}, \rho_{\boldsymbol{\theta}}, \gamma_{\boldsymbol{\theta}}(0)) = (\mu_{\boldsymbol{\theta}}, \gamma_{\boldsymbol{\theta}}(1), \ldots, \gamma_{\boldsymbol{\theta}}(r), \gamma_{\boldsymbol{\theta}}(0))^\top.$$

By (1.2.6), $\gamma_{\boldsymbol{\theta}}(h) = \gamma_{\boldsymbol{\theta}}(0) k_{\boldsymbol{\theta}} e^{-h\rho_{\boldsymbol{\theta}}}$ for $h \in \mathbb{N}$, and simple algebra shows that $k_{\boldsymbol{\theta}}$ and $\rho_{\boldsymbol{\theta}}$ are uniquely determined by

$$k_{\boldsymbol{\theta}} = \frac{\gamma^2_{\boldsymbol{\theta}}(1)}{\gamma_{\boldsymbol{\theta}}(0)\gamma_{\boldsymbol{\theta}}(2)} \quad \text{and} \quad \rho_{\boldsymbol{\theta}} = \log\Big(\frac{\gamma_{\boldsymbol{\theta}}(1)}{\gamma_{\boldsymbol{\theta}}(2)}\Big), \tag{1.2.15}$$

and, therefore, π_2 is injective. Finally, define the map $\pi_3 : \pi_2(\pi_1(\Theta)) \to \Pi$, by

$$\pi_3(\mu_{\boldsymbol{\theta}}, \gamma_{\boldsymbol{\theta}}(1), \ldots, \gamma_{\boldsymbol{\theta}}(r), \gamma_{\boldsymbol{\theta}}(0)) = (\mu_{\boldsymbol{\theta}}, a_{\boldsymbol{\theta},1}, \ldots, a_{\boldsymbol{\theta},r}, \gamma_{\boldsymbol{\theta}}(0))^\top.$$

The map π_3 is injective, since $\gamma_{\boldsymbol{\theta}}(1), \ldots, \gamma_{\boldsymbol{\theta}}(r)$ are uniquely determined by $a_{\boldsymbol{\theta},1}, \ldots, a_{\boldsymbol{\theta},r}$ and $\gamma_{\boldsymbol{\theta}}(0)$. We need $r \geq 2$ in order to recover $(\gamma_{\boldsymbol{\theta}}(1), \gamma_{\boldsymbol{\theta}}(2))$ from $(\gamma_{\boldsymbol{\theta}}(0), a_{\boldsymbol{\theta},1}, a_{\boldsymbol{\theta},2})$ using the system of Yule-Walker equations (1.2.10), so that (1.2.15) remains valid. This implies the injectivity of the composition π.

Now we prove that π is continuously differentiable. The map π_1 is given in terms of equations (3.6)-(3.9) of Theorem 3.1 of Haug et al. [45], which are continuously differentiable maps of $\Psi_{\boldsymbol{\theta}}(1)$ and $\Psi_{\boldsymbol{\theta}}(2)$ as defined in (1.2.2). By assumption (A4) of Proposition 1.2.1 the Lévy process L has finite fourth moment and, therefore, both $\Psi_{\boldsymbol{\theta}}(1)$ and $\Psi_{\boldsymbol{\theta}}(2)$ exist and are continuously differentiable in $\boldsymbol{\theta}$. By (1.2.6), π_2 is continuously differentiable. Finally, π_3 is also continuously differentiable since it is defined recursively by means of the Yule-Walker equations (1.2.10). This proves that the composition π is continuously differentiable. □

1.2.3 Indirect Inference Estimation

Let $\hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta})$ denote an estimator of the auxiliary AR(r) model for $r \geq 2$ based on the returns $(G_i(\boldsymbol{\theta}))_{i=1}^n$, where $\boldsymbol{\theta}$ lies in a compact subspace Θ of $\mathcal{M}$ as in (1.2.7). We define now the IIE for the COGARCH process.

Definition 1.2.5. *Let* $\boldsymbol{G}_n := (G_i(\boldsymbol{\theta}_0))_{i=1}^n$ *be the returns as defined in* (1.1.2). *Let* $\hat{\boldsymbol{\pi}}_n$ *be one of the estimators given in* (1.2.13) *of* $\boldsymbol{\pi}_{\boldsymbol{\theta}_0}$ *as defined in* (1.2.11). *For arbitrary* $\boldsymbol{\theta} \in \Theta$ *and* $k = 1, \ldots, K$ *let* $\hat{\boldsymbol{\pi}}_{n,k}(\boldsymbol{\theta})$ *be estimators of* $\boldsymbol{\pi}_{\boldsymbol{\theta}}$ *based on independent simulated paths* $\boldsymbol{G}_{n,k}(\boldsymbol{\theta}) := (G_i^{(k)}(\boldsymbol{\theta}))_{i=1}^n$. *Let* $\boldsymbol{\Omega}$ *be a symmetric and positive definite weight matrix. Define the function*

$$\hat{L}_{\mathrm{II}} : \Theta \to [0, \infty) \quad \text{based on } \boldsymbol{G}_n \text{ by} \quad \hat{L}_{\mathrm{II}}(\boldsymbol{\theta}, \boldsymbol{G}_n) := \Big\| \hat{\boldsymbol{\pi}}_n - \frac{1}{K} \sum_{k=1}^{K} \hat{\boldsymbol{\pi}}_{n,k}(\boldsymbol{\theta}) \Big\|_{\boldsymbol{\Omega}}.$$

Then the IIE of $\boldsymbol{\theta}_0$ *is defined as*

$$\hat{\boldsymbol{\theta}}_{n,\mathrm{II}} := \arg\min_{\boldsymbol{\theta} \in \Theta} \hat{L}_{\mathrm{II}}(\boldsymbol{\theta}, \boldsymbol{G}_n). \tag{1.2.16}$$

Concerning the asymptotic behavior of the IIE one would hope that strong consistency and asymptotic normality of the estimator of the auxiliary model parameter also implies strong consistency and asymptotic normality of the IIE. However, as the Indirect Inference method is based on the simulation of the COGARCH process for many different parameters, we need a stronger (uniform) consistency result and also additional regularity conditions to ensure this. The following is a modification of Propositions 1 and 3 of Gourieroux et al. [39], and it is the analog of Theorem 3.2 of Fasen-Hartmann and Kimmig [31] in the context of our model.

Proposition 1.2.6. *Assume the setting of Definition 1.2.5 and* $r \geq 2$.
(a) If the uniform SLLN

$$\sup_{\boldsymbol{\theta} \in \Theta} \| \hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta}) - \boldsymbol{\pi}_{\boldsymbol{\theta}} \| \stackrel{a.s.}{\to} 0, \quad n \to \infty, \tag{1.2.17}$$

holds, then the IIE (1.2.16) *is strongly consistent:*

$$\hat{\boldsymbol{\theta}}_{n,\mathrm{II}} \stackrel{a.s.}{\to} \boldsymbol{\theta}_0, \quad n \to \infty.$$

(b) Assume additionally to (1.2.17) *that the following assumptions hold:*

(b.1) for every $n \in \mathbb{N}$ *the map* $\boldsymbol{\theta} \mapsto \hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta})$ *is continuously differentiable,*

(b.2) for every $\boldsymbol{\theta} \in \Theta$ *we have* $\sqrt{n}(\hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta}) - \boldsymbol{\pi}_{\boldsymbol{\theta}}) \stackrel{d}{\to} \mathcal{N}(0, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$ *as* $n \to \infty$, *and*

(b.3) for every sequence $(\boldsymbol{\theta}_n)_{n \in \mathbb{N}}$ *with* $\boldsymbol{\theta}_n \stackrel{\mathrm{a.s.}}{\to} \boldsymbol{\theta}_0$ *as* $n \to \infty$ *it also holds that*

$$\nabla_{\boldsymbol{\theta}} \hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta}_n) \stackrel{P}{\to} \nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}_{\boldsymbol{\theta}_0}, \quad n \to \infty,$$

and $\nabla_\theta \boldsymbol{\pi}(\boldsymbol{\theta}_0)$ *has full column rank 3.*

(b.4) The true parameter $\boldsymbol{\theta}_0$ *lies in the interior of* Θ.

Then the IIE (1.2.17) *is asymptotically normal:*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{n,\mathrm{II}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(0, \Xi_{\boldsymbol{\theta}_0}), \quad n \to \infty,$$

where the asymptotic variance is given by

$$\Xi_{\boldsymbol{\theta}_0} = (\mathcal{J}_{\boldsymbol{\theta}_0})^{-1} \mathcal{I}_{\boldsymbol{\theta}_0} (\mathcal{J}_{\boldsymbol{\theta}_0})^{-1} \tag{1.2.18}$$

with

$$\begin{aligned} \mathcal{J}_{\boldsymbol{\theta}_0} &= (\nabla_{\boldsymbol{\theta}} \pi_{\boldsymbol{\theta}_0})^\top \boldsymbol{\Omega} (\nabla_{\boldsymbol{\theta}} \pi_{\boldsymbol{\theta}_0}) \quad \textit{and} \\ \mathcal{I}_{\boldsymbol{\theta}_0} &= (\nabla_{\boldsymbol{\theta}} \pi_{\boldsymbol{\theta}_0})^\top \boldsymbol{\Omega} \Big(1 + \frac{1}{K}\Big) \boldsymbol{\Sigma}_{\boldsymbol{\theta}_0} \boldsymbol{\Omega} (\nabla_{\boldsymbol{\theta}} \pi_{\boldsymbol{\theta}_0}). \end{aligned} \tag{1.2.19}$$

Proof. Part (a) follows as a particular case of the proof of Theorem 3.2 of [31]. For part (b), we need to check assumptions (C.3)-(C.5) of that theorem. By construction of the estimator (1.2.13) the asymptotic covariance matrices in (C.3) and (C.4) are identical, so that (b.2) implies (C.3) and (C.4). Instead of verifying (C.5) we modify their argument (under (b.1) and (b.4)), when manipulating the first order condition

$$0 = \nabla_\theta \hat{L}_{\mathrm{II}}(\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}, \boldsymbol{G}_n) = 2(\nabla_\theta \hat{\boldsymbol{\pi}}_n(\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}))^T \Omega (\hat{\boldsymbol{\pi}}_n(\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}) - \hat{\boldsymbol{\pi}}_n). \tag{1.2.20}$$

We follow Theorem 3.2 of Newey and McFadden [78], and perform a Taylor expansion of order 1 around the true value $\boldsymbol{\theta}_0$ of the function $\hat{\boldsymbol{\pi}}_n(\hat{\boldsymbol{\theta}}_{n,\mathrm{II}})$ in (1.2.20). After rearranging the terms this leads to

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{n,\mathrm{II}} - \boldsymbol{\theta}_0) = -\Big((\nabla_\theta \hat{\boldsymbol{\pi}}_n(\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}))^T \Omega (\nabla_\theta \hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta}_n))\Big)^{-1} (\nabla_\theta \hat{\boldsymbol{\pi}}_n(\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}))\, \Omega\, \sqrt{n}\big(\hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta}_0) - \hat{\boldsymbol{\pi}}_n\big),$$

where $\boldsymbol{\theta}_n$ is such that $\|\boldsymbol{\theta}_n - \boldsymbol{\theta}_0\| \leq \|\hat{\boldsymbol{\theta}}_{n,\mathrm{II}} - \boldsymbol{\theta}_0\|$. The asymptotic normality follows for $n \to \infty$ from (a), (b.2) and (b.3). □

1.3 Auxiliary autoregressive model - strong consistency and asymptotic normality

Our objective is to investigate the asymptotic behavior of the IIE for the COGARCH parameter $\boldsymbol{\theta}$ using an AR(r) model for fixed $r \geq 2$ as auxiliary model. This amounts to verifying all assumptions of Proposition 1.2.6.

In a first step we investigate strong consistency and asymptotic normality of the two estimators of the auxiliary AR(r) parameter from Definition 1.2.3, which result from the projection presented in Proposition 1.2.2, and may have a non-zero mean. We recall that in classical time series theory the two estimators are asymptotically equivalent (cf. the proof of Theorem 8.1.2 in Brockwell and Davis [13]).

Here the situation is different and, to the best of our knowledge, has not yet been covered in the literature. The noise process $(U_i(\boldsymbol{\theta}))_{i\in\mathbb{N}}$ from (1.2.9) is defined as the projection errors over the finite past. Thus, $U_i(\boldsymbol{\theta})$ is orthogonal to $\mathcal{H}_i = \overline{\text{sp}}\{\tilde{G}^2_{i+r-j}(\boldsymbol{\theta}), j = 1,\dots,r\}$, but we cannot guarantee that it is also orthogonal to $\overline{\text{sp}}\{\tilde{G}^2_{i+r-j}(\boldsymbol{\theta}), j \in \mathbb{N}\}$, so it may not be a white noise process. Therefore, the classical asymptotic theory for the estimation of autoregressive processes (when data come from an AR model with white noise residuals) does not apply directly. Since the residuals are stationary and ergodic with zero mean and finite variance, and since $U_i(\boldsymbol{\theta})$ is orthogonal to $\mathcal{H}_i = \overline{\text{sp}}\{\tilde{G}^2_{i+r-j}(\boldsymbol{\theta}), j = 1,\dots,r\}$, we obtain results by modifying the classical arguments.

This section provides asymptotic results for the estimators of the auxiliary AR model for some arbitrary, but fixed COGARCH parameter $\boldsymbol{\theta}$, where the dependence on $\boldsymbol{\theta}$ is irrelevant, and we omit it for ease of notation. We define $(W_i)_{i\in\mathbb{N}} := (G_i^2)_{i\in\mathbb{N}}$ and rewrite the auxiliary AR(r) model of Proposition 1.2.2 with parameter $\boldsymbol{\pi} = (\mu, \boldsymbol{a}, \gamma(0))$ as

$$\tilde{W}_{i+r} = \sum_{j=1}^{r} a_j \tilde{W}_{i+r-j} + U_i, \quad i \in \mathbb{N},$$

where $\tilde{W}_i = \tilde{G}_i^2 = W_i - \mu, \mu = \mathbb{E}W_1$ and $\gamma(0) = \text{Var}W_1$.

1.3.1 Strong consistency of LSE and YWE

Lemma 1.3.1. *Let the assumptions of Proposition 1.2.1 hold. Then as $n \to \infty$, $\hat{\mu}_n \overset{\text{a.s.}}{\to} \mu$ and $\hat{\gamma}_n(h) \overset{\text{a.s.}}{\to} \gamma(h)$ for $h \in \mathbb{N}_0$.*

Proof. From Proposition 1.2.1 we know that $\mathbb{E}|W_1| < \infty$ and $(W_i)_{i\in\mathbb{N}}$ is ergodic, so that Birkhoff's ergodic theorem (see e.g. Theorem 4.4 in Krengel [62]) gives immediately $\hat{\mu}_n \overset{\text{a.s.}}{\to} \mu$ as $n \to \infty$. To prove almost sure convergence of the empirical autocovariance function, we first investigate it, when the mean μ is known:

$$\gamma_n^*(h) := \frac{1}{n}\sum_{i=1}^{n-h}(W_i - \mu)(W_{i+h} - \mu), \quad h \in \mathbb{N}_0. \tag{1.3.1}$$

Since $W_i W_{i+h}$ is for every $i \in \mathbb{N}$ a measurable map of finitely many values of $(W_i)_{i\in\mathbb{N}}$, the sequence $(W_i W_{i+h})_{i\in\mathbb{N}}$ is ergodic. From Proposition 1.2.1(a), $\mathbb{E}|W_1 W_{1+h}| < \infty$, so that Birkhoff's ergodic theorem gives $\gamma_n^*(h) \overset{\text{a.s.}}{\to} \gamma(h)$ as $n \to \infty$. Simple algebra shows that

$$\hat{\gamma}(h) - \gamma^*(h) = \frac{1}{n}\sum_{i=1}^{n-h}(W_i + W_{i+h} - \hat{\mu}_n - \mu)(\mu - \hat{\mu}_n). \tag{1.3.2}$$

Since as $n \to \infty$, $\hat{\mu}_n \overset{\text{a.s.}}{\to} \mu$, the difference $\gamma_n^*(h) - \hat{\gamma}_n(h) \overset{\text{a.s.}}{\to} 0$; hence, $\hat{\gamma}_n(h) \overset{\text{a.s.}}{\to} \gamma(h)$. $\square$

Theorem 1.3.2 (Consistency of LSE and YWE)**.** *Let the assumptions of Proposition 1.2.1 hold. Then as* $n \to \infty$*,* $\hat{\boldsymbol{a}}_{n,\mathrm{LS}} \overset{\text{a.s.}}{\to} \boldsymbol{a}$ *and* $\hat{\boldsymbol{a}}_{n,\mathrm{YW}} \overset{\text{a.s.}}{\to} \boldsymbol{a}$*.*

Proof. We start by proving strong consistency of the LSE, when the mean μ is known:

$$\boldsymbol{a}^*_{n,\mathrm{LS}} = \arg\min_{\boldsymbol{c}\in C} S^*_n(\boldsymbol{c}), \tag{1.3.3}$$

for C as in (1.2.12), and

$$S^*_n(\boldsymbol{c}) = \frac{1}{n-r}\sum_{i=1}^{n-r}\left((W_{i+r}-\mu) - c_r(W_{i+r-1}-\mu) - \cdots - c_1(W_i-\mu)\right)^2.$$

As in Section 8.10* of [13] we write the auxiliary AR(r) model in matrix form as

$$\tilde{\boldsymbol{Y}}_n = \tilde{\boldsymbol{W}}_n\boldsymbol{a} + \boldsymbol{U}_n, \quad n \in \mathbb{N},$$

where $\tilde{\boldsymbol{Y}}_n = (\tilde{W}_{r+1},\dots,\tilde{W}_n)^\top, \boldsymbol{U}_n = (U_1,\dots,U_{n-r})^\top$ and $\tilde{\boldsymbol{W}}_n$ is the $n\times r$ design matrix,

$$\tilde{\boldsymbol{W}}_n = \begin{pmatrix} \tilde{W}_r & \tilde{W}_{r-1} & \dots & \tilde{W}_1 \\ \tilde{W}_{r+1} & \tilde{W}_r & \dots & \tilde{W}_2 \\ \vdots & \vdots & & \vdots \\ \tilde{W}_{n-1} & \tilde{W}_{n-2} & \dots & \tilde{W}_{n-r} \end{pmatrix}. \tag{1.3.4}$$

Then notice that $(n-r)S^*_n(\boldsymbol{c}) = (\tilde{\boldsymbol{Y}}_n - \tilde{\boldsymbol{W}}_n\boldsymbol{c})^\top(\tilde{\boldsymbol{Y}}_n - \tilde{\boldsymbol{W}}_n\boldsymbol{c})$, revealing the LSE as a linear regression-type estimator given by

$$\boldsymbol{a}^*_{n,\mathrm{LS}} = (\tilde{\boldsymbol{W}}_n^\top\tilde{\boldsymbol{W}}_n)^{-1}\tilde{\boldsymbol{W}}_n^\top\tilde{\boldsymbol{Y}}_n, \tag{1.3.5}$$

provided that the $r\times r$ matrix $\tilde{\boldsymbol{W}}_n^\top\tilde{\boldsymbol{W}}_n$ is invertible. We prove that $n^{-1}\tilde{\boldsymbol{W}}_n^\top\tilde{\boldsymbol{W}}_n$ converges a.s. as $n\to\infty$ to an invertible matrix. For each fixed $u,v\in\{1,\dots,r\}$ the (u,v)-th entry of this matrix is

$$\frac{1}{n}\sum_{i=0}^{n-r-1}\tilde{W}_{r+1-u+i}\tilde{W}_{r+1-v+i}.$$

Since $\tilde{W}_{r+1-u+i}\tilde{W}_{r+1-v+i}$ is for every $i\in\mathbb{N}_0$ a measurable map of finitely many values of $(W_i)_{i\in\mathbb{N}}$, the sequence $(\tilde{W}_{r+1-u+i}\tilde{W}_{r+1-v+i})_{i\in\mathbb{N}_0}$ is ergodic. Since $\mathbb{E}W_1^2<\infty$ Birkhoff's ergodic theorem gives

$$\frac{1}{n}\sum_{i=0}^{n-r-1}\tilde{W}_{r+1-u+i}\tilde{W}_{r+1-v+i} \overset{\text{a.s.}}{\to} \mathbb{E}\tilde{W}_1\tilde{W}_{1+|u-v|}, \quad n\to\infty, \tag{1.3.6}$$

and thus $n^{-1}\tilde{\boldsymbol{W}}_n^\top\tilde{\boldsymbol{W}}_n \overset{\text{a.s.}}{\to} \boldsymbol{\Gamma}$ as $n\to\infty$, where $\boldsymbol{\Gamma}$ is the autocovariance matrix of the squared COGARCH returns, which is non-singular (cf. the proof of Proposition 1.2.2). Thus, $\boldsymbol{\Gamma}$ is invertible and, therefore, the estimator given in (1.3.5) is well defined for n large enough. With (1.3.5) we calculate

$$\begin{aligned} \boldsymbol{a}^*_{n,\mathrm{LS}} - \boldsymbol{a} &= (\tilde{\boldsymbol{W}}_n^\top \tilde{\boldsymbol{W}}_n)^{-1} \tilde{\boldsymbol{W}}_n^\top (\tilde{\boldsymbol{W}}_n \boldsymbol{a} + \boldsymbol{U}_n) - \boldsymbol{a} \\ &= n(\tilde{\boldsymbol{W}}_n^\top \tilde{\boldsymbol{W}}_n)^{-1} \frac{1}{n} \tilde{\boldsymbol{W}}_n^\top \boldsymbol{U}_n \\ &= n(\tilde{\boldsymbol{W}}_n^\top \tilde{\boldsymbol{W}}_n)^{-1} \frac{1}{n} \sum_{i=1}^n \Big(\tilde{W}_{i+r} - \sum_{j=1}^r a_j \tilde{W}_{i+r-j} \Big) \begin{pmatrix} \tilde{W}_{i+r-1} \\ \vdots \\ \tilde{W}_i \end{pmatrix} \\ &=: (n^{-1} \tilde{\boldsymbol{W}}_n^\top \tilde{\boldsymbol{W}}_n)^{-1} \frac{1}{n} \sum_{i=1}^n \boldsymbol{Z}_i. \end{aligned} \tag{1.3.7}$$

Since $\boldsymbol{Z}_i$ is for every $i \in \mathbb{N}$ a measurable map of finitely many values of $(W_i)_{i\in\mathbb{N}}$, the sequence $(\boldsymbol{Z}_i)_{i\in\mathbb{N}}$ is ergodic. According to Proposition 1.2.2, $\tilde{W}_{i+r} - \sum_{j=1}^r a_j \tilde{W}_{i+r-j}$ is uncorrelated with $\tilde{W}_i, \dots, \tilde{W}_{i+r-1}$ for all $i \in \mathbb{N}$. Since $\mathbb{E}|\boldsymbol{Z}_1| < \infty$ Birkhoff's ergodic theorem gives

$$\frac{1}{n} \sum_{i=1}^n \boldsymbol{Z}_i \overset{\text{a.s.}}{\to} \mathbb{E}\boldsymbol{Z}_1 = \boldsymbol{0}.$$

This together with the fact that the first term of (1.3.7) converges a.s. to $\boldsymbol{\Gamma}^{-1}$ shows that

$$\boldsymbol{a}^*_{n,\mathrm{LS}} \overset{\text{a.s.}}{\to} \boldsymbol{a}, \quad n \to \infty.$$

It remains to prove that $(\hat{\boldsymbol{a}}_{n,\mathrm{LS}} - \boldsymbol{a}^*_{n,\mathrm{LS}}) \overset{\text{a.s.}}{\to} \boldsymbol{0}$ as $n \to \infty$. Write the LSE in the matrix form

$$\hat{\boldsymbol{a}}_{n,\mathrm{LS}} = (\bar{\boldsymbol{W}}_n^\top \bar{\boldsymbol{W}}_n)^{-1} \bar{\boldsymbol{W}}_n^\top \bar{\boldsymbol{Y}}_n,$$

where $\bar{\boldsymbol{W}}_n$ and $\bar{\boldsymbol{Y}}_n$ denote the matrix and vector defined in Eq. (1.3.4), with entries of the form $\bar{W}_i = W_i - \hat{\mu}_n$. Using the matrix identity $A^{-1}x - C^{-1}y = A^{-1}(x-y) + A^{-1}(C-A)C^{-1}y$ gives

$$\begin{aligned} \hat{\boldsymbol{a}}_{n,\mathrm{LS}} - \boldsymbol{a}^*_{n,\mathrm{LS}} =& \Big(\frac{\bar{\boldsymbol{W}}_n^\top \bar{\boldsymbol{W}}_n}{n} \Big)^{-1} \Big(\frac{\bar{\boldsymbol{W}}_n^\top \bar{\boldsymbol{Y}}_n}{n} - \frac{\tilde{\boldsymbol{W}}_n^\top \tilde{\boldsymbol{Y}}_n}{n} \Big) \\ &+ \Big(\frac{\bar{\boldsymbol{W}}_n^\top \bar{\boldsymbol{W}}_n}{n} \Big)^{-1} \Big(\frac{\tilde{\boldsymbol{W}}_n^\top \tilde{\boldsymbol{W}}_n}{n} - \frac{\bar{\boldsymbol{W}}_n^\top \bar{\boldsymbol{W}}_n}{n} \Big) \Big(\frac{\tilde{\boldsymbol{W}}_n^\top \tilde{\boldsymbol{W}}_n}{n} \Big)^{-1} \Big(\frac{\tilde{\boldsymbol{W}}_n^\top \tilde{\boldsymbol{Y}}_n}{n} \Big). \end{aligned} \tag{1.3.8}$$

By Birkhoff's ergodic theorem $n^{-1} \bar{\boldsymbol{W}}_n^\top \bar{\boldsymbol{W}}_n, n^{-1} \tilde{\boldsymbol{W}}_n^\top \tilde{\boldsymbol{W}}_n$ and $n^{-1} \tilde{\boldsymbol{W}}_n^\top \tilde{\boldsymbol{Y}}_n$ converge a.s. to two matrices and a vector, respectively. Additionally, by (1.3.2) we can apply Birkhoff's ergodic theorem to obtain as $n \to \infty$,

$$\Big(\frac{\bar{\boldsymbol{W}}_n^\top \bar{\boldsymbol{Y}}_n}{n} - \frac{\tilde{\boldsymbol{W}}_n^\top \tilde{\boldsymbol{Y}}_n}{n} \Big) \overset{\text{a.s.}}{\to} \boldsymbol{0} \quad \text{and} \quad \Big(\frac{\tilde{\boldsymbol{W}}_n^\top \tilde{\boldsymbol{W}}_n}{n} - \frac{\bar{\boldsymbol{W}}_n^\top \bar{\boldsymbol{W}}_n}{n} \Big) \overset{\text{a.s.}}{\to} \boldsymbol{0},$$

showing that the LSE is strongly consistent. For the YWE the proof is a direct consequence of Lemma 1.3.1 and the continuous mapping theorem. □

1.3.2 Asymptotic normality of the LSE and YWE

One of the requirements for asymptotic normality of the IIE of the COGARCH parameter $\boldsymbol{\theta}$ is condition (b.2) of Proposition 1.2.6. This means we have to prove asymptotic normality of $\hat{\boldsymbol{\pi}}_{n,\mathrm{LS}}$ and $\hat{\boldsymbol{\pi}}_{n,\mathrm{YW}}$. We start with an auxiliary result.

Lemma 1.3.3. *Let the assumptions of Proposition 1.2.1 hold. Let $\boldsymbol{a}^*_{n,\mathrm{LS}}$ be the LSE defined in* (1.3.3) *and $\boldsymbol{a}^*_{n,\mathrm{YW}}$ be the modification of the YWE defined in* (1.2.14)*, when the true mean μ is known, i.e., with $\hat{\gamma}_n(\cdot)$ replaced by $\gamma^*(\cdot)$ from* (1.3.1)*. Then as $n \to \infty$,*

(a) $\sqrt{n}(\hat{\mu}_n^2 - \mu^2) \xrightarrow{P} 0$,

(b) $\sqrt{n}(\boldsymbol{a}^*_{n,\mathrm{YW}} - \hat{\boldsymbol{a}}_{n,\mathrm{YW}}) \xrightarrow{P} 0$,

(c) $\sqrt{n}(\boldsymbol{a}^*_{n,\mathrm{YW}} - \boldsymbol{a}^*_{n,\mathrm{LS}}) \xrightarrow{P} 0$,

(d) $\sqrt{n}(\boldsymbol{a}^*_{n,\mathrm{LS}} - \hat{\boldsymbol{a}}_{n,\mathrm{LS}}) \xrightarrow{P} 0$.

Proof. (a) Write $\sqrt{n}(\hat{\mu}_n^2 - \mu^2) = \sqrt{n}(\hat{\mu}_n + \mu)(\hat{\mu}_n - \mu)$ and notice that by Lemma 1.3.1 we only need to show that $\sqrt{n}(\hat{\mu}_n + \mu)$ is bounded in probability. It follows from (1.2.6) that $\gamma(h)$ decays exponentially in h and thus $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$. Let $\epsilon > 0$ be fixed and apply Chebyshev's inequality to get

$$\mathbb{P}(\sqrt{n}|\hat{\mu}_n + \mu| > \epsilon) \leq \epsilon^{-2} n \mathrm{Var}(\hat{\mu}_n) \to \epsilon^{-2} \sum_{h=-\infty}^{\infty} \gamma(h) < \infty, \quad n \to \infty,$$

where the convergence follows from Theorem 7.1.1 in Brockwell and Davis [13].
(b) Write $\boldsymbol{a}^*_{n,\mathrm{YW}} = (\boldsymbol{\Gamma}^*_n)^{-1}\boldsymbol{\gamma}^*_n$ with autocovariance function $\gamma^*(\cdot)$ defined in (1.3.1). Using properties of the inverse matrix we get

$$\begin{aligned}\sqrt{n}(\boldsymbol{a}^*_{n,\mathrm{YW}} - \hat{\boldsymbol{a}}_{n,\mathrm{YW}}) &= \sqrt{n}(\hat{\boldsymbol{\Gamma}}_n^{-1}\hat{\boldsymbol{\gamma}}_n - (\boldsymbol{\Gamma}^*_n)^{-1}\boldsymbol{\gamma}^*_n) \\ &= \hat{\boldsymbol{\Gamma}}_n^{-1}\sqrt{n}(\boldsymbol{\Gamma}^*_n - \hat{\boldsymbol{\Gamma}}_n)(\boldsymbol{\Gamma}^*_n)^{-1}\hat{\boldsymbol{\gamma}}_n + (\boldsymbol{\Gamma}^*_n)^{-1}\sqrt{n}(\hat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}^*_n).\end{aligned}$$

The estimators $\hat{\boldsymbol{\Gamma}}_n, \boldsymbol{\Gamma}^*_n$ and $\hat{\boldsymbol{\gamma}}_n$ are all bounded in probability. For fixed $h \in \mathbb{N}_0$ it follows from (1.3.2) and Lemma 1.3.3(a) that $\sqrt{n}(\hat{\gamma}_n(h) - \gamma^*_n(h)) \xrightarrow{P} 0$ as $n \to \infty$. Therefore, $\sqrt{n}(\boldsymbol{\Gamma}^*_n - \hat{\boldsymbol{\Gamma}}_n)$ and $\sqrt{n}(\hat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}^*_n)$ also converge to zero in probability as $n \to \infty$, which entails (b).
(c) This follows similarly as in the proof of Theorem 8.1.1 in Brockwell and Davis [13].
(d) By (1.3.8) and observing that $n^{-1}\bar{\boldsymbol{W}}_n^\top\bar{\boldsymbol{W}}_n$, $n^{-1}\tilde{\boldsymbol{W}}_n^\top\tilde{\boldsymbol{W}}_n$ and $n^{-1}\tilde{\boldsymbol{W}}_n^\top\tilde{\boldsymbol{Y}}_n$ are bounded in probability, we only need to show that $n^{-\frac{1}{2}}\{\bar{\boldsymbol{W}}_n^\top\bar{\boldsymbol{Y}}_n - \tilde{\boldsymbol{W}}_n^\top\tilde{\boldsymbol{Y}}_n\}$ and $n^{-\frac{1}{2}}\{\tilde{\boldsymbol{W}}_n^\top\tilde{\boldsymbol{W}}_n - \bar{\boldsymbol{W}}_n^\top\bar{\boldsymbol{W}}_n\}$ converge to zero in probability as $n \to \infty$. These terms only depend on the autocovariance function of the process $(W_i)_{i\in\mathbb{N}}$ and therefore convergence in probability to zero follows

from $\sqrt{n}(\hat{\gamma}_n(h) - \gamma_n^*(h)) \xrightarrow{P} 0$ as $n \to \infty$ as can be seen from (1.3.2), and the fact that $\sqrt{n}\hat{\mu}_n$ is bounded in probability. $\square$

The following is the main result of this Section and proves Assumption (b.2) of Proposition 1.2.6.

Theorem 1.3.4 (Asymptotic normality of the LSE and YWE)**.** *Let the assumptions of Proposition 1.2.1 hold. Assume additionally that* $\mathbb{E}|L_1|^{8+\epsilon} < \infty$ *and* $\Psi_{\boldsymbol{\theta}}(4 + \frac{\epsilon}{2}) < 0$ *for some* $\epsilon > 0$ *and that the matrix* $\boldsymbol{\Sigma}$ *defined in* (1.3.9) *is positive definite. Then, both LSE and YWE for the AR*(r) *model for* $r \geq 2$ *are asymptotically normal with covariance matrix*

$$\boldsymbol{\Sigma} = \left(\begin{array}{c|ccc|c} 1 & 0 & \dots & 0 & 0 \\ \hline 0 & & & & 0 \\ \vdots & & \boldsymbol{\Gamma}^{-1} & & \vdots \\ 0 & & & & 0 \\ \hline 0 & 0 & \dots & 0 & 1 \end{array}\right) \boldsymbol{\Sigma}^*, \tag{1.3.9}$$

where $\boldsymbol{\Gamma}$ *is the autocovariance matrix of* $(W_i)_{i=1}^r$*,*

$$\boldsymbol{\Sigma}^* = \mathbb{E}\boldsymbol{C}_1\boldsymbol{C}_1^\top + 2\sum_{i=1}^{\infty} \mathbb{E}\boldsymbol{C}_1\boldsymbol{C}_{1+i}^\top, \tag{1.3.10}$$

with $C_i \in \mathbb{R}^{r+2}$ *given by*

$$\boldsymbol{C}_i = \begin{pmatrix} \tilde{W}_i \\ (\tilde{W}_{i+r} - \sum_{j=1}^r a_j \tilde{W}_{i+r-j})\tilde{W}_{i+r-1} \\ \vdots \\ (\tilde{W}_{i+r} - \sum_{j=1}^r a_j \tilde{W}_{i+r-j})\tilde{W}_i \\ W_i^2 - \mu^2 \end{pmatrix}. \tag{1.3.11}$$

Proof. Write

$$\begin{aligned} \sqrt{n}(\hat{\boldsymbol{\pi}}_{n,\mathrm{LS}} - \boldsymbol{\pi}) &= \sqrt{n}\begin{pmatrix} \hat{\mu}_n - \mu \\ \hat{\boldsymbol{a}}_{n,\mathrm{LS}} - \boldsymbol{a} \\ \hat{\gamma}_n(0) - \gamma(0) \end{pmatrix} \\ &= \sqrt{n}\begin{pmatrix} 0 \\ \hat{\boldsymbol{a}}_{n,\mathrm{LS}} - \boldsymbol{a}^*_{n,\mathrm{LS}} \\ \mu^2 - \hat{\mu}_n^2 \end{pmatrix} + \sqrt{n}\begin{pmatrix} \hat{\mu}_n - \mu \\ \boldsymbol{a}^*_{n,\mathrm{LS}} - \boldsymbol{a} \\ \hat{\gamma}_n(0) + \hat{\mu}_n^2 - \mathbb{E}W_1^2 \end{pmatrix} \end{aligned} \tag{1.3.12}$$

and

$$\sqrt{n}(\hat{\boldsymbol{\pi}}_{n,\mathrm{YW}} - \boldsymbol{\pi}) = \sqrt{n}\begin{pmatrix} \hat{\mu}_n - \mu \\ \hat{\boldsymbol{a}}_{n,\mathrm{YW}} - \boldsymbol{a} \\ \hat{\gamma}_n(0) - \gamma(0) \end{pmatrix}$$

$$= \sqrt{n}\begin{pmatrix} 0 \\ \hat{\boldsymbol{a}}_{n,\mathrm{YW}} - \boldsymbol{a}^*_{n,\mathrm{YW}} \\ \mu^2 - \hat{\mu}_n^2 \end{pmatrix} + \sqrt{n}\begin{pmatrix} 0 \\ \boldsymbol{a}^*_{n,\mathrm{YW}} - \boldsymbol{a}^*_{n,\mathrm{LS}} \\ 0 \end{pmatrix} + \sqrt{n}\begin{pmatrix} \hat{\mu}_n - \mu \\ \boldsymbol{a}^*_{n,\mathrm{LS}} - \boldsymbol{a} \\ \hat{\gamma}_n(0) + \hat{\mu}_n^2 - \mathbb{E}W_1^2 \end{pmatrix}. \tag{1.3.13}$$

We apply Lemma 1.3.3 to the right-hand side of (1.3.12) and (1.3.13) and find that it suffices to prove that

$$\sqrt{n}\begin{pmatrix} \hat{\mu}_n - \mu \\ \boldsymbol{a}^*_{n,\mathrm{LS}} - \boldsymbol{a} \\ \hat{\gamma}_n(0) + \hat{\mu}_n^2 - \mathbb{E}W_1^2 \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}), \quad n \to \infty.$$

Using (1.3.7) we write

$$\sqrt{n}\begin{pmatrix} \hat{\mu}_n - \mu \\ \boldsymbol{a}^*_{n,\mathrm{LS}} - \boldsymbol{a} \\ \hat{\gamma}_n(0) + \hat{\mu}_n^2 - \mathbb{E}W_1^2 \end{pmatrix} = \sqrt{n}\begin{pmatrix} \frac{1}{n}\sum_{i=1}^n (W_i - \mu) \\ n(\tilde{\boldsymbol{W}}_n^\top \tilde{\boldsymbol{W}}_n)^{-1}\frac{1}{n}\sum_{i=1}^n \boldsymbol{Z}_i \\ \frac{1}{n}\sum_{i=1}^n (W_i - \hat{\mu}_n)^2 + \hat{\mu}_n^2 - \mathbb{E}W_1^2 \end{pmatrix}$$

$$= \left(\begin{array}{c|ccc|c} 1 & 0 & \dots & 0 & 0 \\ \hline 0 & & & & 0 \\ \vdots & & n(\tilde{\boldsymbol{W}}_n^\top \tilde{\boldsymbol{W}}_n)^{-1} & & \vdots \\ 0 & & & & 0 \\ \hline 0 & 0 & \dots & 0 & 1 \end{array}\right) \frac{1}{\sqrt{n}}\sum_{i=1}^n \begin{pmatrix} W_i - \mu \\ \boldsymbol{Z}_i \\ W_i^2 - \mathbb{E}W_1^2 \end{pmatrix} \tag{1.3.14}$$

$$=: \boldsymbol{B}_n \frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{C}_i.$$

For the asymptotic normality of (1.3.14) we use the Cramér-Wold device and show that

$$\sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^n \boldsymbol{\lambda}^\top \boldsymbol{C}_i\Big) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^* \boldsymbol{\lambda}), \quad n \to \infty,$$

for all vectors $\boldsymbol{\lambda} \in \mathbb{R}^{r+2}$ such that $\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^* \boldsymbol{\lambda} > 0$. It follows from Proposition 1.2.1(c) that the squared returns process $(W_i)_{i\in\mathbb{N}}$ is α-mixing with exponentially decaying mixing coefficients. Since each $\boldsymbol{C}_i$ is a measurable function of $W_i, \dots, W_{i-r}$ it follows from Remark 1.8 of Bradley [12] that $(\boldsymbol{\lambda}^\top \boldsymbol{C}_i)_{i\in\mathbb{N}}$ is also α-mixing with mixing coefficients satisfying

$\alpha_{\boldsymbol{C}}(n) \leq \alpha_{\boldsymbol{W}}(n-(r+1))$ for all $n \geq r+2$. Therefore $\sum_{n=0}^{\infty}(\alpha_{\boldsymbol{C}}(n))^{\frac{\epsilon}{2+\epsilon}} < \infty$ for all $\epsilon > 0$. Since $\mathbb{E}|L_1|^{8+\epsilon} < \infty$ and $\Psi_{\boldsymbol{\theta}}(4+\frac{\epsilon}{2}) < 0$ it follows from (1.2.5) that $\mathbb{E}|W_1|^{4+\epsilon/2} < \infty$ and, as a consequence, $\mathbb{E}|\lambda^\top \boldsymbol{C}_1|^{2+\epsilon/4} < \infty$. Thus, the CLT for α-mixing sequences applies (see Theorem 18.5.3 of Ibragimov and Linnik [50]) so that

$$\sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\lambda}^\top \boldsymbol{C}_i\Big) \stackrel{d}{\to} \mathcal{N}(0,\zeta), \quad n \to \infty,$$

where $\zeta = \mathbb{E}\boldsymbol{\lambda}^\top \boldsymbol{C}_1 \boldsymbol{C}_1^\top \boldsymbol{\lambda} + 2\sum_{i=1}^{\infty}\mathbb{E}\boldsymbol{\lambda}^\top \boldsymbol{C}_1 \boldsymbol{C}_{1+i}^\top \boldsymbol{\lambda}$. After rearranging this equation we find (1.3.10). Let $\boldsymbol{B}_n = (b_{u,v}^n)_{u,v=1}^{r+2}$ denote the matrix as defined in (1.3.14). Using (1.3.6) we get for $2 \leq u, v \leq r+1$,

$$b_{u,v}^n \stackrel{\text{a.s.}}{\to} \mathbb{E}(W_1-\mu)(W_{1+|u-v|}-\mu), \quad n \to \infty.$$

Then the inner block of the matrix $\boldsymbol{B}_n$ converges a.s. to $\boldsymbol{\Gamma}^{-1}$. This gives (1.3.9),which finishes the proof. □

1.4 IIE of the COGARCH process - strong consistency and asymptotic normality

The objective of this section is to prove strong consistency and asymptotic normality of the IIE of the COGARCH parameter $\boldsymbol{\theta}$. Let $(G_i(\boldsymbol{\theta}_0))_{i=1}^n$ be the returns originating from a COGARCH log-price process (3.1.1). As auxiliary model we use an AR(r) model for fixed $r \geq 2$ as in Proposition 1.2.2, whose parameters are estimated by one of the estimators $\hat{\boldsymbol{\pi}}_n$ from Definition 1.2.3, which we consider as functions of the COGARCH parameter $\boldsymbol{\theta}$.

1.4.1 Preliminary results

We begin with an auxiliary result, which is a consequence of Theorem 3.2 of [55].

Lemma 1.4.1. *Assume that* $\mathbb{E}|L_1|^2 < \infty$ *and* $\Psi_{\boldsymbol{\theta_i}}(1) < 0$ *for* $i = 1,\dots,d$. *Then for every* $t > 0$,

$$(\sigma_t^2(\boldsymbol{\theta}_1),\dots,\sigma_t^2(\boldsymbol{\theta}_d)) \stackrel{\text{d}}{=} (\sigma_0^2(\boldsymbol{\theta}_1),\dots,\sigma_0^2(\boldsymbol{\theta}_d)).$$

In what follows we shall need for fixed $\varphi > 0$ the stochastic process

$$K_s(\varphi) = \sum_{0<u\leq s}\frac{(\Delta L_u)^2}{1+\varphi(\Delta L_u)^2}, \quad s \geq 0. \tag{1.4.1}$$

Lemma 1.4.2. *The process* $(K_s(\varphi))_{s\geq 0}$ *is a Lévy process and* $\mathbb{E}|K_s(\varphi)|^p < \infty$ *for all* $p \in \mathbb{N}$.

Proof. That $(K_s(\varphi))_{s\geq 0}$ is a Lévy process is clear. Since

$$\sup_{s\geq 0}|\Delta K_s(\varphi)| = \sup_{s\geq 0}\frac{(\Delta L_s)^2}{1+\varphi(\Delta L_s)^2} \leq \frac{1}{\varphi} < \infty \tag{1.4.2}$$

it follows that $(K_s(\varphi))_{s\geq 0}$ has bounded jumps and, therefore, it has moments of all orders (see e.g. Theorem 2.4.7 of [1]). □

For $p \geq 1$ consider the sets

$$\Theta^{(p)} \subset \mathcal{M}^{(p)} := \{\boldsymbol{\theta} \in \mathcal{M} : \Psi_{\boldsymbol{\theta}}(p) < 0\}, \tag{1.4.3}$$

where $\Theta^{(p)}$ is compact, and recall from (1.2.5) that the condition $\Psi_{\boldsymbol{\theta}}(p) < 0$ implicitly requires $\mathbb{E}|L_1|^{2p} < \infty$.

Lemma 1.4.3. *Let $p \geq 1$ and $t \geq 0$ be fixed and consider the sets $\Theta^{(p)}$ and $\mathcal{M}^{(p)}$ as in* (1.4.3)*. Then the following hold.*
(a) There exist numbers $\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_N^* \in \mathcal{M}^{(p)}$ such that

$$\sup_{\boldsymbol{\theta}\in\Theta^{(p)}} e^{-Y_t(\boldsymbol{\theta})} \leq \sum_{j=1}^{N} e^{-Y_t(\boldsymbol{\theta}_j^*)}.$$

(b) There exists some $\sigma^* > 0$ such that $\sigma_0(\boldsymbol{\theta}) \geq \sigma^*$ a.s. for all $\boldsymbol{\theta} \in \Theta$.

Proof. (a) We use a Heine-Borel argument to control the exponential term. Since $\Theta^{(p)}$ is compact we can find a finite collection of open sets $(\Theta_j^{(p)})_{j=1}^N$ such that $\Theta^{(p)} \subseteq \cup_{j=1}^N \overline{\Theta_j^{(p)}} \subset \mathcal{M}^{(p)}$. For each fixed j the closure $\overline{\Theta_j^{(p)}}$ is a subspace of $\mathcal{M}^{(p)}$ and therefore there exists a point $\boldsymbol{\theta}_j^* = (\beta_j^*, \eta_j^*, \varphi_j^*)^\top \in \mathcal{M}^{(p)}$ such that $\eta \geq \eta_j^*, \varphi \leq \varphi_j^*$ for all $\boldsymbol{\theta} \in \Theta_j^{(p)}$. This implies that for all $\boldsymbol{\theta} \in \overline{\Theta_j^{(p)}}$:

$$Y_t(\boldsymbol{\theta}) = \eta t - \sum_{0<u\leq t}\log(1+\varphi(\Delta L_u)^2) \geq \eta_j^* t - \sum_{0<u\leq t}\log(1+\varphi_j^*(\Delta L_u)^2) = Y_t(\boldsymbol{\theta}_j^*), \quad t \geq 0.$$

This implies $\sup_{\boldsymbol{\theta}\in\Theta^{(p)}} e^{-Y_t(\boldsymbol{\theta})} \leq \sum_{j=1}^N e^{-Y_t(\boldsymbol{\theta}_j^*)}$.
(b) is Proposition 2 of Klüppelberg et al. [56]. □

Remark 1.4.4. *Both $\hat{\boldsymbol{\pi}}_{n,\mathrm{YW}}$ (by* (1.2.14)*) and $\hat{\boldsymbol{\pi}}_{n,\mathrm{LS}}$ (as in the proof of Proposition 5.6 of Fasen-Hartmann and Kimmig [31]) can be written as a map $g : \mathbb{R}^{r+2} \to \mathbb{R}^{r+2}$ for $r \geq 2$ with $g(\boldsymbol{x}) = (g_1(\boldsymbol{x}), \ldots, g_{r+2}(\boldsymbol{x}))$ for $\boldsymbol{x} = (x_1, \ldots, x_{r+2})$ applied to the vector*

$$f_n(\boldsymbol{\theta}) = \Big(\frac{1}{n}\sum_{i=1}^{n} G_i^2(\boldsymbol{\theta}), \frac{1}{n}\sum_{i=1}^{n-h} G_i^2(\boldsymbol{\theta})G_{i+h}^2(\boldsymbol{\theta}), h = 0, \ldots, r\Big), \quad n \in \mathbb{N}. \tag{1.4.4}$$

Since g involves only matrix multiplications and matrix inversion of non-singular matrices, it inherits the smoothness properties of $G_i(\boldsymbol{\theta})$ for $i \in \mathbb{N}$. Since $(G_i(\boldsymbol{\theta}))_{i\in\mathbb{N}}$ is stationary and ergodic, Birkhoff's ergodic theorem applies and $f_n(\boldsymbol{\theta})$ converges a.s. as $n \to \infty$ pointwise to

$$f(\boldsymbol{\theta}) = (\mathbb{E}G_1^2(\boldsymbol{\theta}), \mathbb{E}G_1^2(\boldsymbol{\theta})G_{1+h}^2(\boldsymbol{\theta}), h = 0, \dots, r). \tag{1.4.5}$$

Remark 1.4.5. *The results that follow are related to continuity and differentiability of the random elements $(G_i(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta)$ for $i \in \mathbb{N}$ with respect to $\boldsymbol{\theta}$. According to* (1.2.3) *and* (1.2.4) *we find*

$$G_i(\boldsymbol{\theta}) = \int_{(i-1)\Delta}^{i\Delta} \sigma_s((\beta, \eta, \varphi))\mathrm{d}L_s = \sqrt{\beta}\int_{(i-1)\Delta}^{i\Delta} \sigma_s((1, \eta, \varphi))\mathrm{d}L_s = \sqrt{\beta}\, G_i((1, \eta, \varphi)),$$

which is linear in $\sqrt{\beta}$, hence, $(G_i(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta)$ is obviously continuous in β and has a partial derivative with respect to $\beta > 0$.

1.4.2 Strong consistency of the IIE

To ensure strong consistency of $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$, we need to verify that $\hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta})$ satisfies the uniform SLLN of Proposition 1.2.6(a). The results of Lemma 1.3.1 and Theorem 1.3.2 guarantee point-wise strong consistency. Uniform strong consistency will hold by continuity of g (cf. Lemma 1.6.2), if we can apply a uniform SLLN to the sequence in (1.4.4).

Since the sequence of random elements $(G_i(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta)_{i\in\mathbb{N}}$ is stationary and ergodic, we need to show (cf. Theorem 7 in Straumann and Mikosch [100]) that $G_i(\boldsymbol{\theta})$ is for every $i \in \mathbb{N}$ a continuous function of $\boldsymbol{\theta}$ on Θ or on some compact subspace $\Theta^{(p)}$ of Θ and that

$$\mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta^{(p)}} G_i^4(\boldsymbol{\theta}) < \infty.$$

Proving that $G_i(\boldsymbol{\theta})$ is ω-wise continuous in its parameter $\boldsymbol{\theta}$ is not straightforward, since

$$G_i(\boldsymbol{\theta}) = \int_{(i-1)\Delta}^{i\Delta} \sigma_s(\boldsymbol{\theta})\mathrm{d}L_s$$

is a stochastic integral, driven by an arbitrary Lévy process, which also drives the stochastic volatility process. If L has finite variation, we can use dominated convergence to show continuity, but this is not possible when L has infinite variation sample paths; cf. Remark 1.4.9 below. However, as we shall show in the next result, applying Kolmogorov's continuity criterion, we can always find a version $(G_i^{(c)}(\boldsymbol{\theta}))_{i\in\mathbb{N}}$ of the sequence $(G_i(\boldsymbol{\theta}))_{i\in\mathbb{N}}$, which is continuous on a possibly smaller compact parameter space $\Theta^{(p)} \subseteq \Theta$ for Θ.

Theorem 1.4.6 (Hölder continuity)**.** *Assume that* $\mathbb{E}|L_1|^{2p(1+\epsilon)} < \infty$ *for some* $p > 2$ *and* $\epsilon > 0$*. Then there exists a version* $(G_i^{(c)}(\boldsymbol{\theta}))_{i\in\mathbb{N}}$ *of the random elements* $(G_i(\boldsymbol{\theta}))_{i\in\mathbb{N}}$ *which is Hölder continuous of every order* $\gamma \in [0,(p-2)/(2p))$ *on* $\Theta^{(p(1+\epsilon))}$ *as defined in* (1.4.3)*. Additionally, defining*

$$U_i = \sup_{\substack{0<\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|<1\\ \boldsymbol{\theta}_1,\boldsymbol{\theta}_2\in\Theta^{(p(1+\epsilon))}}} \frac{|G_i^{(c)}(\boldsymbol{\theta}_1) - G_i^{(c)}(\boldsymbol{\theta}_2)|}{\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|^\gamma}, \quad i\in\mathbb{N}, \tag{1.4.6}$$

we have for every $q \in [0, 2p)$

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta^{(p(1+\epsilon))}} |G_i^{(c)}(\boldsymbol{\theta})|^q < \infty \quad \textit{and} \quad \mathbb{E}U_i^q < \infty. \tag{1.4.7}$$

Proof. Without loss of generality we prove this for $i = 1$. We find a continuous version of the random element $G_1(\boldsymbol{\theta})$ on $\Theta^{(p(1+\epsilon))}$. We first prove continuity with respect to (η,φ) and assume that $\boldsymbol{\theta}_1,\boldsymbol{\theta}_2 \in \Theta^{(p(1+\epsilon))}$ with $\beta_1,\beta_2 = 1$. Using the simple inequality $|a-b|^{2p} \le |a^2-b^2|^p$, the stationarity of $\sigma_0(\boldsymbol{\theta})$ in Lemma 1.4.1, its differentiability (1.6.10) proven in Lemma 1.6.4 of the Appendix, and the mean value theorem gives

$$\begin{aligned}\int_0^\Delta \mathbb{E}|\sigma_s(\boldsymbol{\theta}_1)-\sigma_s(\boldsymbol{\theta}_2)|^{2p}\,\mathrm{d}s &\le \int_0^\Delta \mathbb{E}|\sigma_s^2(\boldsymbol{\theta}_1)-\sigma_s^2(\boldsymbol{\theta}_2)|^p\\ &= \Delta\mathbb{E}|\sigma_0^2(\boldsymbol{\theta}_1)-\sigma_0^2(\boldsymbol{\theta}_2)|^p \\ &\le \Delta\Big(\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta^{(p(1+\epsilon))}}|\nabla_{\eta,\varphi}\sigma_0^2(\boldsymbol{\theta})|^p\Big)\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|^p < \infty\end{aligned} \tag{1.4.8}$$

by Lemma 1.6.5 with $k = 1$. By (A2) of Proposition 1.2.1, $(L_t)_{t\ge0}$ is a martingale. Since $\mathbb{E}|L_1|^{2p} < \infty$ and $\int_0^\Delta \mathbb{E}|\sigma_s(\boldsymbol{\theta}_1)-\sigma_s(\boldsymbol{\theta}_2)|^{2p}\mathrm{d}s < \infty$ we can apply Theorem 66 of Ch. 5 in Protter [86] to the stochastic integral in (3.1.1) and obtain

$$\mathbb{E}|G_1(\boldsymbol{\theta}_1)-G_1(\boldsymbol{\theta}_2)|^{2p} = \mathbb{E}\Big|\int_0^\Delta(\sigma_s(\boldsymbol{\theta}_1)-\sigma_s(\boldsymbol{\theta}_2))\mathrm{d}L_s\Big|^{2p} \le c^*\int_0^\Delta \mathbb{E}|\sigma_s(\boldsymbol{\theta}_1)-\sigma_s(\boldsymbol{\theta}_2)|^{2p}\mathrm{d}s,$$

where c^* is a positive constant. This combined with (1.4.8) gives

$$\mathbb{E}|G_1(\boldsymbol{\theta}_1)-G_1(\boldsymbol{\theta}_2)|^{2p} \le c\Delta\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|^p, \tag{1.4.9}$$

where $c = c^*\Delta\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta^{(p(1+\epsilon))}}|\nabla_{\eta,\varphi}\sigma_0^2(\boldsymbol{\theta})|^p$. Since $\beta_1,\beta_2 = 1$ we show continuity with respect to (η,φ); i.e. the parameter space has dimension $d = 2$. Since $p > 2 = d$ we can apply Kolmogorov's continuity criterion (Theorem 10.1 in Schilling and Partzsch [91], or Theorem 2.5.1 of Ch. 5 in Khoshnevisan [52]). Then there exists a version $(G_1^{(c)}(\boldsymbol{\theta}),\boldsymbol{\theta}\in\Theta^{(p(1+\epsilon))})$ of $(G_1(\boldsymbol{\theta}),\boldsymbol{\theta}\in\Theta^{(p(1+\epsilon))})$ which is Hölder continuous of every order $\gamma\in[0,(p-2)/(2p))$; hence, also continuous. Since Θ is compact, Lemma 1.6.1 together with (1.4.9) gives $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta^{(p(1+\epsilon))}} |G_1^{(c)}(\boldsymbol{\theta})|^q < \infty$ for every $q \in [0, 2p)$. Finally, the second expectation in (1.4.7) is finite by Theorem 10.1 in [91].

Because of Remark 1.4.5, $G_1^{(c)}(\boldsymbol{\theta})$ is linear in $\sqrt{\beta}$ and, therefore, the results can be generalized to the map $\boldsymbol{\theta} \mapsto G_1^{(c)}(\boldsymbol{\theta})$ on $\Theta^{(p(1+\epsilon))}$. Indeed, let β^* be as in (1.6.15) and

$$\beta_* = \inf\{\beta > 0 : (\beta, \eta, \varphi) \in \Theta\} > 0.$$

For arbitrary $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta^{(p(1+\epsilon))}$ we use Remark 1.4.5, the mean value theorem for $\beta \mapsto \sqrt{\beta}$ and the Hölder continuity of order γ of $(\eta, \varphi) \mapsto G_1^{(c)}((1, \eta, \varphi))$, the definition of the ℓ^1-norm and the fact that $\gamma \in (0, 1)$ to get

$$\begin{aligned}
&|G_1^{(c)}(\boldsymbol{\theta}_1) - G_1^{(c)}(\boldsymbol{\theta}_2)| \\
&\le |G_1^{(c)}((1, \eta_1, \varphi_1)) - G_1^{(c)}((1, \eta_2, \varphi_2))|\sqrt{\beta^*} + \frac{1}{2\sqrt{\beta_*}}|\beta_1 - \beta_2| \sup_{\boldsymbol{\theta}\in\Theta^{(p(1+\epsilon))}} |G_1^{(c)}(\boldsymbol{\theta})| \\
&\le K\|(\eta_1, \varphi_1) - (\eta_2, \varphi_2)\|^\gamma \sqrt{\beta^*} + \frac{1}{2\sqrt{\beta_*}}|\beta_1 - \beta_2|^\gamma |2\beta^*|^{1-\gamma} \sup_{\boldsymbol{\theta}\in\Theta^{(p(1+\epsilon))}} |G_1^{(c)}(\boldsymbol{\theta})| \\
&\le \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^\gamma \Big(K\sqrt{\beta^*} + \frac{1}{2\sqrt{\beta_*}}|2\beta^*|^{1-\gamma} \sup_{\boldsymbol{\theta}\in\Theta^{(p(1+\epsilon))}} |G_1^{(c)}(\boldsymbol{\theta})|\Big),
\end{aligned} \tag{1.4.10}$$

showing the Hölder continuity of $\boldsymbol{\theta} \mapsto G_1^{(c)}(\boldsymbol{\theta})$ on $\Theta^{(p(1+\epsilon))}$. The first expectation in (1.4.7) is finite since $|G_1^{(c)}(\boldsymbol{\theta})| \le \beta^*|G_1^{(c)}((1, \eta, \varphi))|$. Now let $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta^{(p(1+\epsilon))}$ be such that $0 < \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| < 1$. Using the inequality in the first line of (1.4.10) and the definition of the ℓ^1-norm gives

$$\begin{aligned}
&\sup_{\substack{0<\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|<1\\ \boldsymbol{\theta}_1,\boldsymbol{\theta}_2\in\Theta^{(p(1+\epsilon))}}} \frac{|G_1^{(c)}(\boldsymbol{\theta}_1) - G_1^{(c)}(\boldsymbol{\theta}_2)|}{\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^\gamma} \\
&\le \Big(\sup_{\substack{0<\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|<1\\ \boldsymbol{\theta}_1,\boldsymbol{\theta}_2\in\Theta^{(p(1+\epsilon))}}} \frac{|G_1^{(c)}((1, \eta_1, \varphi_1) - G_1^{(c)}((1, \eta_2, \varphi_2)|}{\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^\gamma} \Big)\sqrt{\beta^*} \\
&\qquad + \sup_{\substack{0<\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|<1\\ \boldsymbol{\theta}_1,\boldsymbol{\theta}_2\in\Theta^{(p(1+\epsilon))}}} \frac{|\beta_1 - \beta_2|}{2\sqrt{\beta_*}\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^\gamma} |G_1^{(c)}(\boldsymbol{\theta})| \\
&\le \Big(\sup_{\substack{0<\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|<1\\ \boldsymbol{\theta}_1,\boldsymbol{\theta}_2\in\Theta^{(p(1+\epsilon))}}} \frac{|G_1^{(c)}((1, \eta_1, \varphi_1) - G_1^{(c)}((1, \eta_2, \varphi_2)|}{\|(\eta_1, \varphi_1) - (\eta_2, \varphi_2)\|^\gamma} \Big)\sqrt{\beta^*} + \frac{1}{2\sqrt{\beta_*}} \sup_{\boldsymbol{\theta}\in\Theta^{(p(1+\epsilon))}} |G_1^{(c)}(\boldsymbol{\theta})|
\end{aligned} \tag{1.4.11}$$

Applying the supremum and raising both sides of (1.4.11) to the power q gives the result.
□

Remark 1.4.7. *In view of Theorem 1.4.6 we will from now on work with a continuous version of the returns* $(G_i(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))})_{i\in\mathbb{N}}$.

Theorem 1.4.8 (Strong consistency of the IIE)**.** *Assume that* $\mathbb{E}|L_1|^{2p(1+\epsilon)} < \infty$ *for some* $p > 2$ *and* $\epsilon > 0$ *and let* $(G_i(\boldsymbol{\theta}_0))_{i=1}^n$ *be the returns* (1.1.2) *with parameter* $\boldsymbol{\theta}_0 \in \Theta^{(p(1+\epsilon))}$ *from* (1.4.1)*. Suppose that the auxiliary AR*(r) *model for* $r \geq 2$ *is estimated by the LSE or the YWE of Definition 1.2.3. Then*

$$\hat{\boldsymbol{\theta}}_{n,\mathrm{II}} \overset{a.s.}{\to} \boldsymbol{\theta}_0, \quad n \to \infty.$$

Proof. According to Proposition 1.2.6(a), strong consistency of the IIE will follow if, as $n \to \infty$,

$$\sup_{\boldsymbol{\theta}\in\Theta^{(p(1+\epsilon))}} \|\hat{\boldsymbol{\pi}}_{n,\mathrm{LS}}(\boldsymbol{\theta}) - \boldsymbol{\pi}_{\boldsymbol{\theta}}\| \overset{\text{a.s.}}{\to} 0 \quad \text{and} \quad \sup_{\boldsymbol{\theta}\in\Theta^{(p(1+\epsilon))}} \|\hat{\boldsymbol{\pi}}_{n,\mathrm{YW}}(\boldsymbol{\theta}) - \boldsymbol{\pi}_{\boldsymbol{\theta}}\| \overset{\text{a.s.}}{\to} 0.$$

By Remark 1.4.4 and Lemma 1.6.2 it suffices to prove that

$$\sup_{\boldsymbol{\theta}\in\Theta^{(p(1+\epsilon))}} \|f_n(\boldsymbol{\theta}) - f(\boldsymbol{\theta})\| \overset{\text{a.s.}}{\to} 0, \quad n \to \infty, \tag{1.4.12}$$

for f_n and f as defined in (1.4.4) and (1.4.5), respectively. The Cauchy-Schwarz inequality gives for every $h \in \mathbb{N}_0$,

$$\mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta^{(p(1+\epsilon))}} G_1^2(\boldsymbol{\theta}) G_{1+h}^2(\boldsymbol{\theta}) \leq \Big(\mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta^{(p(1+\epsilon))}} G_1^4(\boldsymbol{\theta})\Big)^{\frac{1}{2}} \Big(\mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta^{(p(1+\epsilon))}} G_{1+h}^4(\boldsymbol{\theta})\Big)^{\frac{1}{2}} < \infty. \tag{1.4.13}$$

The right-hand side of (1.4.13) is finite by Theorem 1.4.6. It also follows from the same theorem that $\mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta^{(p(1+\epsilon))}} G_1^2(\boldsymbol{\theta}) < \infty$ and, hence, by Theorem 7 in Straumann and Mikosch [100] the uniform SLLN holds and we obtain for all $h \in \mathbb{N}_0$ as $n \to \infty$,

$$\begin{aligned} &\sup_{\boldsymbol{\theta}\in\Theta^{(p(1+\epsilon))}} \Big|\frac{1}{n}\sum_{i=1}^n G_i^2(\boldsymbol{\theta}) - \mathbb{E}G_1^2(\boldsymbol{\theta})\Big| \overset{\text{a.s.}}{\to} 0 \quad \text{and} \\ &\sup_{\boldsymbol{\theta}\in\Theta^{(p(1+\epsilon))}} \Big|\frac{1}{n}\sum_{i=1}^n G_i^2(\boldsymbol{\theta}) G_{i+h}^2(\boldsymbol{\theta}) - \mathbb{E}G_1^2(\boldsymbol{\theta}) G_{1+h}^2(\boldsymbol{\theta})\Big| \overset{\text{a.s.}}{\to} 0. \end{aligned} \tag{1.4.14}$$

Hence (1.4.12) follows from (1.4.14), finishing the proof. □

Remark 1.4.9. *If the Lévy process* $(L_t)_{t\geq 0}$ *has finite variation sample paths, then the stochastic integral in* (1.1.2) *can be treated pathwise as a Riemann-Stieltjes integral, such that continuity of* $(G_i(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta)_{i\in\mathbb{N}}$ *follows from Lemma 1.4.3(c) and dominated convergence. Therefore, Theorem 1.4.6 is valid for* $\boldsymbol{\theta}_0 \in \Theta \supseteq \Theta^{(p(1+\epsilon))}$ *for* $p > 2$ *and some* $\epsilon > 0$*. Additionally, since the total variation process is also a Lévy process we can use Theorem 66 of Ch. 5 in Protter [86] to show that* $\mathbb{E}L_1^4 < \infty$ *implies* $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} G_i^4(\boldsymbol{\theta}) < \infty$ *for all* $i \in \mathbb{N}$*. Therefore, also Theorem 1.4.8 is valid for* $\boldsymbol{\theta}_0 \in \Theta \supseteq \Theta^{(p(1+\epsilon))}$.

1.4.3 Asymptotic normality of the IIE

In order to prove asymptotic normality of the IIE, we need to verify the conditions (b.1), (b.2) and (b.3) of Proposition 1.2.6. We recall that (b.2) has been proved in Theorem 1.3.4, and it remains to prove (b.1) and (b.3), which are related to the smoothness of $\hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta})$ as a function of $\boldsymbol{\theta}$.

Differentiability properties of $(G_i(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))})$

Condition (b.1) refers to the differentiability of the map $\hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta})$. By Remark 1.4.4 and the chain rule we only need to prove differentiability of $G_i(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ for every $i \in \mathbb{N}$. Since $G_i(\boldsymbol{\theta})$ is defined in terms of a stochastic integral we can not simply interchange the order of the Riemann differentiation and the stochastic integration, however, under appropriate regularity conditions formulated of Hutton and Nelson [47] this is possible.

We start by investigating the candidate for the differential of $(G_i(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))})$ with $\Theta^{(p(1+\epsilon))}$ as in (1.4.3), namely the map

$$\boldsymbol{\theta} \mapsto \int_{(i-1)\Delta}^{i\Delta} \nabla_\theta \sigma_s(\boldsymbol{\theta}) \mathrm{d}L_s := \nabla_\theta G_i(\boldsymbol{\theta}). \tag{1.4.15}$$

We show in Lemma 1.4.10 that we can find a version of the integral on the right-hand side, which is continuous on a subspace $\Theta^{(2p(1+\epsilon))}$ of $\Theta^{(p(1+\epsilon))}$. Then, Theorem 1.4.11 asserts that $G_i(\boldsymbol{\theta})$ is differentiable on $\Theta^{(2p(1+\epsilon))}$ and that its differential is indeed given by (1.4.15).

Lemma 1.4.10 (Hölder continuity of derivatives)**.** *Assume that* $\mathbb{E}|L_1|^{4p(1+\epsilon)} < \infty$ *for some* $p > 2$ *and* $\epsilon > 0$*. Then there exists a version* $(\nabla_\theta G_i^{(c)}(\boldsymbol{\theta}))_{i\in\mathbb{N}}$ *of the random elements* $(\nabla_\theta G_i(\boldsymbol{\theta}))_{i\in\mathbb{N}}$ *which is Hölder continuous of every order* $\gamma \in [0, (p-2)/p)$ *on* $\Theta^{(2p(1+\epsilon))}$ *as defined in* (1.4.3)*. Additionally, defining*

$$V_i = \sup_{\substack{0<\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|<1 \\ \boldsymbol{\theta}_1,\boldsymbol{\theta}_2\in\Theta^{(2p(1+\epsilon))}}} \frac{|\frac{\partial}{\partial\boldsymbol{\theta}_l} G_i(\boldsymbol{\theta}_1) - \frac{\partial}{\partial\boldsymbol{\theta}_l} G_i(\boldsymbol{\theta}_2)|}{\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^\gamma}, \quad i \in \mathbb{N}, \tag{1.4.16}$$

we have for every $q \in [0, p)$ *and* $l \in \{1, 2, 3\}$

$$\mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta^{(2p(1+\epsilon))}} \Big|\frac{\partial}{\partial\boldsymbol{\theta}_l} G_i(\boldsymbol{\theta})\Big|^q < \infty \quad \textit{and} \quad \mathbb{E}V_i^q < \infty.$$

Proof. Without loss of generality we consider $i = 1$. In view of Remark 1.4.5 we write (1.4.15) as

$$\Big(\frac{1}{2\sqrt{\beta}} \int_0^\Delta \sigma_s((1,\eta,\varphi))\mathrm{d}L_s, \sqrt{\beta}\int_0^\Delta \frac{\partial}{\partial\eta}\sigma_s((1,\eta,\varphi))\mathrm{d}L_s, \sqrt{\beta}\int_0^\Delta \frac{\partial}{\partial\varphi}\sigma_s((1,\eta,\varphi))\mathrm{d}L_s\Big)^\top. \tag{1.4.17}$$

From Remark 1.4.7 the first component of (1.4.17) is continuous in β even on Θ. For the remaining two components we show continuity with respect to (η, ϕ). Thus, assume that $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta^{(2p(1+\epsilon))}$ with $\beta_1 = \beta_2 = 1$. Using the distributional property of $\sigma_s(\cdot)$ in Lemma 1.4.1 and the differentiability of $\boldsymbol{\theta} \mapsto \sigma_s(\boldsymbol{\theta})$ in Lemma 1.6.4 gives for every Borel set $B \subset \mathbb{R}$ that

$$\begin{aligned}
&\mathbb{P}\Big(\frac{\partial}{\partial\eta}(\sigma_s(\boldsymbol{\theta}_1) - \sigma_s(\boldsymbol{\theta}_2)) \in B\Big) \\
= \;& \mathbb{P}\Big(\lim_{h\to 0, h\in\mathbb{Q}} \frac{[\sigma_s(\boldsymbol{\theta}_1 + (h,0)) - \sigma_s(\boldsymbol{\theta}_1)] - [\sigma_s(\boldsymbol{\theta}_2 + (h,0)) - \sigma_s(\boldsymbol{\theta}_2)]}{h} \in B\Big) \\
= \;& \lim_{h\to 0, h\in\mathbb{Q}} \mathbb{P}\Big(\frac{[\sigma_s(\boldsymbol{\theta}_1 + (h,0)) - \sigma_s(\boldsymbol{\theta}_1)] - [\sigma_s(\boldsymbol{\theta}_2 + (h,0)) - \sigma_s(\boldsymbol{\theta}_2)]}{h} \in B\Big) \\
= \;& \lim_{h\to 0, h\in\mathbb{Q}} \mathbb{P}\Big(\frac{[\sigma_0(\boldsymbol{\theta}_1 + (h,0)) - \sigma_0(\boldsymbol{\theta}_1)] - [\sigma_0(\boldsymbol{\theta}_2 + (h,0)) - \sigma_0(\boldsymbol{\theta}_2)]}{h} \in B\Big) \\
= \;& \mathbb{P}\Big(\frac{\partial}{\partial\eta}(\sigma_0(\boldsymbol{\theta}_1) - \sigma_0(\boldsymbol{\theta}_2)) \in B\Big),
\end{aligned}$$

so that

$$\frac{\partial}{\partial\eta}(\sigma_s(\boldsymbol{\theta}_1) - \sigma_s(\boldsymbol{\theta}_2)) \stackrel{\mathrm{d}}{=} \frac{\partial}{\partial\eta}(\sigma_0(\boldsymbol{\theta}_1) - \sigma_0(\boldsymbol{\theta}_2)). \tag{1.4.18}$$

Similar calculations show that (1.4.18) is also valid for $\frac{\partial}{\partial\eta}$ replaced by $\frac{\partial}{\partial\varphi}$. Thus, $(\nabla_{\eta,\varphi}\sigma_s(\boldsymbol{\theta}))_{s\geq 0}$ is stationary and it follows from its differentiability (1.6.10) proven in Lemma 1.6.4 of the Appendix, and the mean value theorem that

$$\begin{aligned}
&\int_0^\Delta \mathbb{E}\|\nabla_{\eta,\varphi}\sigma_s(\boldsymbol{\theta}_1) - \nabla_{\eta,\varphi}\sigma_s(\boldsymbol{\theta}_2)\|^p \,\mathrm{d}s \\
&= \Delta\mathbb{E}\|\nabla_{\eta,\varphi}\sigma_0(\boldsymbol{\theta}_1) - \nabla_{\eta,\varphi}\sigma_0(\boldsymbol{\theta}_2)\|^p \leq \Delta\Big(\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta^{(p(1+\epsilon))}} \|\nabla^2_{\eta,\varphi}\sigma_0(\boldsymbol{\theta})\|^p\Big)\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^p < \infty,
\end{aligned}$$

by Lemma 1.6.5 with $k = 2$. The rest of the proof follows along the same lines those in the proof of Theorem 1.4.6. □

Theorem 1.4.11 (Differentiable version of $(G_i(\boldsymbol{\theta}))_{i\in\mathbb{N}}$). *Assume the conditions of Lemma 1.4.10. Then there is a version* $(G_i(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta^{(2p(1+\epsilon))})_{i\in\mathbb{N}}$ *for* $\Theta^{(2p(1+\epsilon))}$ *as in* (1.4.3)*, which is continuously differentiable and its derivative is given a.s. by* $(\nabla_\theta G_i(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta^{(2p(1+\epsilon))})_{i\in\mathbb{N}}$.

Proof. Without loss of generality we consider $i = 1$. From Remark 1.4.5 it follows that $G_1(\boldsymbol{\theta}) = \sqrt{\beta}G_1((1,\eta,\varphi))$ so that obviously

$$\frac{\partial}{\partial\beta}G_1(\boldsymbol{\theta}) = \frac{1}{2\sqrt{\beta}}G_1((1,\eta,\varphi)) = \int_0^\Delta \frac{\partial}{\partial\beta}\sigma_s(\boldsymbol{\theta})\mathrm{d}L_s.$$

Interchanging the partial differentiation with respect to (η, ϕ) and the stochastic integral requires the four regularity conditions of Theorem 2.2 of Hutton and Nelson [47]. Let $\mathcal{F}_t := \sigma(\{L_s, 0 \le s \le t\})$, such that $(\mathcal{F}_t)_{t\ge 0}$ is the filtration generated by the Lévy process L. Condition (i) of that paper is satisfied, since $(\sigma_t(\boldsymbol{\theta}))_{t\ge 0}$ is predictable, we consider the parameter space $\mathcal{M}$ with the Borel σ-algebra, and the parameter $\boldsymbol{\theta}$ is independent of t. Since $\sigma_s(\boldsymbol{\theta}) = \sqrt{\beta}\sigma_s((1,\eta,\varphi))$ these regularity conditions need only to be checked for the map $(\eta,\varphi) \mapsto \sigma_s((1,\eta,\varphi))$. Condition (ii) requires that $\int_0^\Delta \sigma_s^2(\boldsymbol{\theta})\mathrm{d}\langle L\rangle_s < \infty$ a.s. for every $\boldsymbol{\theta} \in \Theta^{(2p(1+\epsilon))}$, where $\langle L\rangle = (\langle L\rangle_s)_{s\ge 0}$ is the characteristic of the martingale L. Since L is a square integrable Lévy process, $\langle L\rangle_s = s\mathbb{E}L_1^2$ and, thus, this condition holds since $s \mapsto \sigma_s(\boldsymbol{\theta})$ has bounded sample paths on the compact interval $[0,\Delta]$. The first part of condition (iii) requires that for every fixed s, the map $\boldsymbol{\theta} \mapsto \sigma_s(\boldsymbol{\theta})$ is absolutely continuous. From the definition of $\sigma_s^2(\boldsymbol{\theta})$ in (1.2.3) we have for $\beta_1 = \beta_2 = 1$,

$$\sigma_s^2(\boldsymbol{\theta}) = e^{-Y_{s-}(\boldsymbol{\theta})}\Big(\int_0^s e^{Y_v(\boldsymbol{\theta})}\mathrm{d}v + \int_0^\infty e^{-Y_v(\boldsymbol{\theta})}\mathrm{d}v\Big) =: h(\boldsymbol{\theta})(f(\boldsymbol{\theta}) + g(\boldsymbol{\theta})). \tag{1.4.19}$$

Then for fixed $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta^{(2p(1+\epsilon))}$ we use Lemma 1.6.4 in combination with the mean value theorem and Lemma 1.4.3(a) to get

$$\begin{aligned}
&|\sigma_s^2(\boldsymbol{\theta}_1) - \sigma_s^2(\boldsymbol{\theta}_2)|\\
\le\;& \Big|(f(\boldsymbol{\theta}_1)+g(\boldsymbol{\theta}_1))\Big(h(\boldsymbol{\theta}_1)-h(\boldsymbol{\theta}_2)+h(\boldsymbol{\theta}_2)(f(\boldsymbol{\theta}_1)-f(\boldsymbol{\theta}_2)+g(\boldsymbol{\theta}_1)-g(\boldsymbol{\theta}_2)\Big)\Big|\\
\le\;& (h(\boldsymbol{\theta}_2)+f(\boldsymbol{\theta}_1)+g(\boldsymbol{\theta}_1))\Big(|h(\boldsymbol{\theta}_1)-h(\boldsymbol{\theta}_2)|+|f(\boldsymbol{\theta}_1)-f(\boldsymbol{\theta}_2)|+|g(\boldsymbol{\theta}_1)-g(\boldsymbol{\theta}_2)|\Big)\\
\le\;& \sup_{\boldsymbol{\theta}\in\Theta}\{h(\boldsymbol{\theta})+f(\boldsymbol{\theta})+g(\boldsymbol{\theta})\}\Big(|h(\boldsymbol{\theta}_1)-h(\boldsymbol{\theta}_2)|+|f(\boldsymbol{\theta}_1)-f(\boldsymbol{\theta}_2)|+|g(\boldsymbol{\theta}_1)-f(\boldsymbol{\theta}_2)|\Big)\\
\le\;& \|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|\sum_{j=1}^N\Big\{\sup_{\boldsymbol{\theta}\in\Theta}(h(\boldsymbol{\theta})+f(\boldsymbol{\theta})+g(\boldsymbol{\theta}))\Big\}\Big\{e^{-Y_s(\boldsymbol{\theta}_j^*)}(s+K_s(\varphi_*))\\
&+\int_0^s e^{Y_v(\boldsymbol{\theta}_j^*)}(v+K_v(\varphi_*))\mathrm{d}v + \int_0^\infty e^{-Y_v(\boldsymbol{\theta}_j^*)}(v+K_v(\varphi_*))\mathrm{d}v\Big\},
\end{aligned} \tag{1.4.20}$$

where $(\boldsymbol{\theta}_j^*)_{j=1}^N \in \mathcal{M}^{(2p(1+\epsilon))}$. Since Θ is compact and for each fixed $s \ge 0$, $\boldsymbol{\theta} \mapsto \sigma_s(\boldsymbol{\theta})$ is continuous, $\sup_{\boldsymbol{\theta}\in\Theta}\{h(\boldsymbol{\theta})+f(\boldsymbol{\theta})+g(\boldsymbol{\theta})\}$ is finite. Furthermore, Lemma 1.6.3 implies that the other three random variables on the right-hand side of (1.4.20) have finite first moment, and are therefore also a.s. finite. Thus (1.4.20) implies that the map $\boldsymbol{\theta} \mapsto \sigma_s^2(\boldsymbol{\theta})$ is a.s. Lipschitz continuous on $\Theta^{(2p(1+\epsilon))}$ and, as a consequence, absolutely continuous on $\Theta^{(2p(1+\epsilon))}$. For the second part of condition (iii) we recall first that we have assumed that $\beta = 1$, such that we focus on the partial differentiation of the parameter $(\eta,\varphi)^\top$. A non-decreasing predictable process $(\lambda_t)_{t\ge 0}$ is needed such that for every t and $\boldsymbol{\theta} \in \Theta^{(2p(1+\epsilon))}$

$$\int_0^t \|\nabla_{\eta,\varphi}\sigma_s(\boldsymbol{\theta})\|^2\mathrm{d}\langle L\rangle_s < \lambda_t \quad \text{a.s.}$$

From (1.4.19), the product rule and Proposition 2 of [56] we find

$$\|\nabla_{\eta,\varphi}\sigma_s(\boldsymbol{\theta})\| \le \frac{1}{2\sigma^*}\{\|\nabla_{\eta,\varphi}h(\boldsymbol{\theta})\|(f(\boldsymbol{\theta})+g(\boldsymbol{\theta}))+h(\boldsymbol{\theta})\|\nabla_{\eta,\varphi}f(\boldsymbol{\theta})+\nabla_{\eta,\varphi}g(\boldsymbol{\theta})\|\}. \quad (1.4.21)$$

We use Lemma 1.6.4 and the definition of the process $(Y_t(\boldsymbol{\theta}))_{t\ge 0}$ in (1.2.1). First note that

$$\begin{aligned} \eta &\le \sup\{\eta>0:(\beta,\eta,\varphi)\in\Theta\} =: \eta^* < \infty, \\ \varphi &\ge \inf\{\varphi>0:(\beta,\eta,\varphi)\in\Theta\} =: \varphi_* > 0, \end{aligned} \quad (1.4.22)$$

and we get the bound

$$f(\boldsymbol{\theta}) = \int_0^s e^{Y_v(\boldsymbol{\theta})}\mathrm{d}v = \int_0^s \exp\Big\{\eta v - \sum_{0<u\le s}\log\left(1+\varphi(\Delta L_u)^2\right)\Big\}\mathrm{d}v \le s e^{\eta^* s}. \quad (1.4.23)$$

Hence it follows from (1.4.19) that

$$\begin{aligned} \|\nabla_{\eta,\varphi}f(\boldsymbol{\theta})\| &= \textstyle\int_0^s v e^{Y_v(\boldsymbol{\theta})}\mathrm{d}v + \int_0^s e^{Y_v(\boldsymbol{\theta})}K_v(\varphi)\mathrm{d}v &&\le \Big(s+K_s(\varphi_*)\Big)s e^{\eta^* s}, \\ \|\nabla_{\eta,\varphi}h(\boldsymbol{\theta})\| &= s e^{-Y_{s-}(\boldsymbol{\theta})} + e^{-Y_{s-}(\boldsymbol{\theta})}K_s(\varphi) &&\le e^{-Y_s(\boldsymbol{\theta})}\big(s+K_s(\varphi_*)\big), \\ \|\nabla_{\eta,\varphi}g(\boldsymbol{\theta})\| &= \textstyle\int_0^\infty v e^{-Y_v(\boldsymbol{\theta})}\mathrm{d}v + \int_0^\infty e^{-Y_v(\boldsymbol{\theta})}K_v(\varphi)\mathrm{d}v &&\le \int_0^\infty (v+K_v(\varphi_*))e^{-Y_v(\boldsymbol{\theta})}\mathrm{d}v. \end{aligned} \quad (1.4.24)$$

From (1.4.21) and the bounds given in (1.4.23) and (1.4.24) we obtain

$$\begin{aligned} &\|\nabla_{\eta,\varphi}\sigma_s(\boldsymbol{\theta})\| \\ \le\ & \frac{1}{2\sigma^*}e^{-Y_s(\boldsymbol{\theta})}(s+K_s(\varphi_*))\Big(s e^{\eta^* s} + \int_0^\infty e^{-Y_v(\boldsymbol{\theta})}\mathrm{d}v\Big) \\ &+\frac{1}{2\sigma^*}e^{-Y s-(\boldsymbol{\theta})}\Big\{(s+K_s(\varphi_*))s e^{\eta^* s} + \int_0^\infty e^{-Y_v(\boldsymbol{\theta})}(v+K_v(\varphi_*))\mathrm{d}v\Big\} =: l_s(\boldsymbol{\theta}). \end{aligned} \quad (1.4.25)$$

Using the compactness of $\Theta^{(2p(1+\epsilon))}$, (1.4.25) and Lemma 1.4.3(a) gives

$$\sup_{(\eta,\varphi)\in\Theta^{(2p(1+\epsilon))}} \|\nabla_{\eta,\varphi}\sigma_s(\boldsymbol{\theta})\| \le \sup_{(\eta,\varphi)\in\Theta^{(2p(1+\epsilon))}} l_s(\boldsymbol{\theta}) \le \sum_{j=1}^N l_s(\boldsymbol{\theta}_j^*),$$

where $(\boldsymbol{\theta}_j^*)_{j=1}^N$ in $\mathcal{M}^{(2p(1+\epsilon))}$. Thus,

$$\int_0^t \|\nabla_{\eta,\varphi}\sigma_s(\boldsymbol{\theta})\|^2 \mathrm{d}\langle L\rangle_s < 1 + \mathbb{E}L_1^2\int_0^t \big|\sum_{j=1}^N l_s(\boldsymbol{\theta}_j^*)\big|^2\mathrm{d}s := \lambda_t, \quad 0\le t\le\Delta,$$

which is a well defined process. Since $(\lambda_t)_{t\ge 0}$ is adapted to the filtration $(\mathcal{F}_t)_{t\ge 0}$ and continuous, it is predictable. The fourth regularity condition we need to check is that the maps

$$\boldsymbol{\theta} \mapsto \int_0^\Delta \sigma_s(\boldsymbol{\theta})\mathrm{d}L_s \quad \text{and} \quad \boldsymbol{\theta}\mapsto\int_0^\Delta \nabla_{\eta,\varphi}\sigma_s(\boldsymbol{\theta})\mathrm{d}L_s$$

are continuous, which has been proved in Theorem 1.4.6 and Lemma 1.4.10. This concludes the proof. □

Remark 1.4.12. *In view of Theorem 1.4.11 we will from now on work with returns* $(G_i(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta^{(2p(1+\epsilon))})_{i\in\mathbb{N}}$ *with* $\Theta^{(2p(1+\epsilon))}$ *as in* (1.4.3)*, which are continuously differentiable with*

$$\nabla_\theta G_i(\boldsymbol{\theta}) = \int_{(i-1)\Delta}^{i\Delta} \nabla_\theta \sigma_s(\boldsymbol{\theta}) \mathrm{d}L_s, \quad i \in \mathbb{N}.$$

As a consequence, also the map $\boldsymbol{\theta} \mapsto \hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta})$ *is continuously differentiable on* $\Theta^{(2p(1+\epsilon))}$ *and, hence condition (b.1) of Proposition 1.2.6 holds.*

Convergence of the derivatives

Finally, we prove condition (b.3) of Proposition 1.2.6.

Proposition 1.4.13 (Consistency of the derivatives)**.** *Assume that* $\mathbb{E}|L_1|^{4p(1+\epsilon)} < \infty$ *for some* $p > 2/5$ *and* $\epsilon > 0$*. Let* $\hat{\boldsymbol{\pi}}_n$ *be one of the estimators* $\hat{\boldsymbol{\pi}}_{n,\mathrm{LS}}$ *and* $\hat{\boldsymbol{\pi}}_{n,\mathrm{YW}}$ *from Definition 1.2.3. Then for every sequence* $(\boldsymbol{\theta}_n)_{n\in\mathbb{N}} \subset \Theta^{(2p(1+\epsilon))}$ *as in* (1.4.3) *and* $\boldsymbol{\theta}_n \stackrel{\text{a.s.}}{\to} \boldsymbol{\theta}_0$ *we have* $\nabla_{\boldsymbol{\theta}} \hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta}_n) \stackrel{P}{\to} \nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}_{\boldsymbol{\theta}_0}$ *as* $n \to \infty$*.*

Proof. Recall from Remark 1.4.4 that we can write each of the two estimators $\hat{\boldsymbol{\pi}}_{n,\mathrm{LS}}$ and $\hat{\boldsymbol{\pi}}_{n,\mathrm{YW}}$ as a continuously differentiable map $g : \mathbb{R}^{r+2} \to \mathbb{R}^{r+2}$, whose Jacobi matrix exists and all partial derivatives of g are continuous. Hence, $\hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta}) = (g_1(f_n(\boldsymbol{\theta})), \ldots, g_{r+2}(f_n(\boldsymbol{\theta})))^\top$ for $\boldsymbol{\theta} = (\beta, \eta, \varphi) =: (\theta_1, \theta_2, \theta_3)$, and we obtain for the partial derivatives by the chain rule

$$\frac{\partial}{\partial \theta_l} g_k(f_n(\boldsymbol{\theta})) = \Big(\frac{\partial g_k(f_n(\boldsymbol{\theta}))}{\partial x_1}, \ldots, \frac{\partial g_k(f_n(\boldsymbol{\theta}))}{\partial x_{r+2}}\Big)\Big(\frac{\partial}{\partial \theta_l} f_n(\boldsymbol{\theta})\Big). \tag{1.4.26}$$

for every $l = 1, 2, 3$ and $k = 1, \ldots, r+2$. By the continuous mapping theorem and (1.4.26) it suffices to prove that as $n \to \infty$,

$$f_n(\boldsymbol{\theta}_n) \stackrel{P}{\to} f(\boldsymbol{\theta}_0) \quad \text{and} \quad \frac{\partial}{\partial \theta_j} f_n(\boldsymbol{\theta}_n) \stackrel{P}{\to} \frac{\partial}{\partial \theta_l} f(\boldsymbol{\theta}_0), \quad l = 1, 2, 3.$$

Let $l \in \{1, 2, 3\}$ be fixed. It follows from (1.4.14) and from Lemma 1.7.2 that as $n \to \infty$,

$$\sup_{\boldsymbol{\theta} \in \Theta^{(2p(1+\epsilon))}} \|f_n(\boldsymbol{\theta}) - f(\boldsymbol{\theta})\| \stackrel{P}{\to} 0 \quad \text{and} \quad \sup_{\boldsymbol{\theta} \in \Theta^{(2p(1+\epsilon))}} \Big\|\frac{\partial}{\partial \boldsymbol{\theta}_l} f_n(\boldsymbol{\theta}) - \frac{\partial}{\partial \boldsymbol{\theta}_l} f(\boldsymbol{\theta})\Big\| \stackrel{P}{\to} 0. \tag{1.4.27}$$

Since

$$\begin{aligned} \|f_n(\boldsymbol{\theta}_n) - f(\boldsymbol{\theta}_0)\| &\le \|f_n(\boldsymbol{\theta}_n) - f(\boldsymbol{\theta}_n)\| + \|f(\boldsymbol{\theta}_n) - f(\boldsymbol{\theta}_0)\| \\ &\le \sup_{\Theta^{(2p(1+\epsilon))}} \|f_n(\boldsymbol{\theta}) - f(\boldsymbol{\theta})\| + \|f(\boldsymbol{\theta}_n) - f(\boldsymbol{\theta}_0)\|, \end{aligned} \tag{1.4.28}$$

from the continuity of f on $\Theta^{(2p(1+\epsilon))}$, the fact that $\boldsymbol{\theta}_n \xrightarrow{P} \boldsymbol{\theta}_0$, and (1.4.27) it follows that $f_n(\boldsymbol{\theta}_n) \xrightarrow{P} f(\boldsymbol{\theta}_0)$. Similar calculations as in (1.4.28) show that

$$\frac{\partial}{\partial \theta_l} f_n(\boldsymbol{\theta}_n) \xrightarrow{P} \frac{\partial}{\partial \theta_l} f(\boldsymbol{\theta}_0),$$

concluding the proof. □

We are now ready to state asymptotic normality of the IIE.

Theorem 1.4.14 (Asymptotic normality of the IIE)**.** *Assume that* $\mathbb{E}|L_1|^{4p(1+\epsilon)} < \infty$ *for some* $p > 2/5$ *and* $\epsilon > 0$*. Let* $(G_i(\boldsymbol{\theta}_0))_{i=1}^n$ *be the returns* (1.1.2) *with true parameter* $\boldsymbol{\theta}_0$*, which lies in the interior of* $\Theta^{(2p(1+\epsilon))}$ *as in* (1.4.3)*. Suppose that the auxiliary AR*(r) *model for* $r \geq 2$ *is estimated by the LSE or the YWE of Definition 1.2.3. If the matrix* $\boldsymbol{\Sigma} := \boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}$ *defined in* (1.3.9) *of Theorem 1.3.4 is positive definite and* $\nabla_\theta \boldsymbol{\pi}(\boldsymbol{\theta}_0)$ *has full column rank 3, then*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{n,\mathrm{II}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(0, \Xi_{\boldsymbol{\theta}_0}), \quad n \to \infty,$$

where $\Xi_{\boldsymbol{\theta}_0}$ *is defined in* (1.2.18)*.*

Proof. The asymptotic normality follows from Proposition 1.2.6. Since $\boldsymbol{\theta} \in \Theta^{(2p(1+\epsilon))} \subseteq \Theta^{(p(1+\epsilon))} \subseteq \Theta$, Theorem 1.4.8 implies condition (a). Conditions (b.1) and (b.3) are valid by Proposition 1.4.13 and the fact that $\nabla_\theta \boldsymbol{\pi}(\boldsymbol{\theta}_0)$ has full column rank 3. Furthermore, (b.2) holds by Theorem 1.3.4, since $\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}$ is positive definite. □

Remark 1.4.15. *We explain how to estimate the asymptotic covariance matrix* $\Xi_{\boldsymbol{\theta}_0}$ *of* $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}} fromTheorem$ 1.4.14*. First, note that it depends on* $K, \nabla_\theta \boldsymbol{\pi}_{\boldsymbol{\theta}_0}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}$ *and* Ω*. Using the map* $\boldsymbol{\theta} \mapsto \nabla_\theta \boldsymbol{\pi}_{\boldsymbol{\theta}}$ *from* (1.2.11) *we compute* $\nabla_\theta \boldsymbol{\pi}_{\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}}$*. An application of the continuous mapping theorem in combination with the continuity of* $\boldsymbol{\theta} \mapsto \nabla_\theta \boldsymbol{\pi}_{\boldsymbol{\theta}}$ *and Theorem 1.4.8 gives* $\nabla_\theta \boldsymbol{\pi}_{\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}} \xrightarrow{\text{a.s.}} \nabla_\theta \boldsymbol{\pi}_{\boldsymbol{\theta}_0}$*. Recall that* $\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}$ *depends on the inverse of the autocovariance function* $\Gamma_{\boldsymbol{\theta}_0}$ *and on* $\boldsymbol{\Sigma}^*_{\boldsymbol{\theta}_0}$ *as defined in* (1.3.10)*. A strongly consistent estimator of* $\Gamma_{\boldsymbol{\theta}_0}$ *is given by* $\Gamma_{\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}}$*. Finally, let* $\hat{C}_k$ *be as in* (1.3.11) *with* W_k *replaced by* $G_k^2(\boldsymbol{\theta}_0)$ *and* $\boldsymbol{\pi} = (\mu, \boldsymbol{a}, \gamma(0))$ *replaced by* $\hat{\boldsymbol{\pi}}_n$*. Then we estimate* $\boldsymbol{\Sigma}^*_{\boldsymbol{\theta}_0}$ *by*

$$\hat{\mu}_{n,C_1,C_1^T} + 2 \sum_{i=1}^{n-r-1} \hat{\mu}_{n,C_1,C_{1+i}^T},$$

where

$$\hat{\mu}_{n,C_1,C_{1+i}^T} = \frac{1}{n-i-r} \sum_{k=1}^{n-i-r} C_k C_{k+i}^T, \quad i = 0, \ldots, n-r-1.$$

Remark 1.4.16. *If the Lévy process $(L_t)_{t\geq 0}$ has finite variation sample paths, then the stochastic integral in* (1.1.2) *can be treated pathwise as a Riemann-Stieltjes integral, such that continuous differentiability of* $(G_i(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta)_{i\in\mathbb{N}}$ *follows by dominated convergence with dominating function as in* (1.4.25)*. Therefore, Theorem 1.4.11 is valid for* $\boldsymbol{\theta}_0 \in \Theta \supseteq \Theta^{(2p(1+\epsilon))}$ *for some* $p > 2$ *and* $\epsilon > 0$*. Additionally, since the total variation process is also a Lévy process we can use Theorem 66 of Ch. 5 in Protter [86] to show that, if* $\mathbb{E}L_1^{8+\delta} < \infty$ *for some* $\delta > 0$*, then* $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta^{(p(1+\epsilon))}} \|\nabla_\theta G_i(\boldsymbol{\theta})\|^{4+\delta/2} < \infty$ *for all* $i \in \mathbb{N}$*. This combined with Remark 1.4.9 and a dominated convergence argument can be applied to show that Lemma 1.4.13 is valid for* $\boldsymbol{\theta}_0 \in \Theta^{(2p(1+\epsilon))}$*, and, as a consequence, also Theorem 1.4.14.*

1.5 Simulation study

The data used for estimation is a sample of COGARCH squared returns $\boldsymbol{G}_n^2 = (G_i^2(\boldsymbol{\theta}_0))_{i=1}^n$ as defined in (1.1.2) with true parameter value $\boldsymbol{\theta}_0 \in \Theta$ as in (1.2.7) observed on a fixed grid of size $\Delta = 1$. We choose a pure jump Variance Gamma (VG) process as the driving Lévy process, which has infinite activity and has been used successfully for modeling stock prices (see Haug et al. [45] and reference therein). The Lévy measure of the VG process with parameter $C > 0$ has Lebesgue density

$$\nu_L(\mathrm{d}x) = \frac{C}{|x|}\exp\{-(2C)^{1/2}|x|\}\mathrm{d}x, \quad x \neq 0. \tag{1.5.1}$$

The Indirect Inference method of Gourieroux et al. [39] based on simulations was originally proposed to estimate models where the binding function is difficult or impossible to compute. However, the binding function $\boldsymbol{\theta} \mapsto \boldsymbol{\pi}_{\boldsymbol{\theta}}$ from Proposition 1.2.4 can be computed explicitly from the formulas given in Theorem 3.1 of Haug et al. [45] and the Yule-Walker equations in (1.2.10), leading to the IIE

$$\hat{\boldsymbol{\theta}}_{n,\mathrm{II}^*} := \operatorname*{arg\,min}_{\boldsymbol{\theta}\in\Theta} \|\hat{\boldsymbol{\pi}}_n - \boldsymbol{\pi}_{\boldsymbol{\theta}}\|_{\boldsymbol{\Omega}}. \tag{1.5.2}$$

We perform a simulation study to evaluate the finite sample performance of the IIE $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}^*}$ in (1.5.2) and also to compare it with the method of moments (MM) estimator $\hat{\boldsymbol{\theta}}_{n,\mathrm{MM}}$ (Algorithm 1 of Haug et al. [45]) and the optimal prediction based (OPB) estimator $\hat{\boldsymbol{\theta}}_{n,\mathrm{OPB}}$ (equation (7) of Bibbona and Negri [7]). As in the simulation studies of [7, 45], we take the VG process with true parameter value $\boldsymbol{\theta}_0 = (0.04, 0.053, 0.038)$ and $C = 1$ in (1.5.1), which implies $\Psi_{\boldsymbol{\theta}_0}(4) = -0.0261 < 0$. Under these conditions, all three estimators $\hat{\boldsymbol{\theta}}_{n,\mathrm{MM}}$ (Theorem 3.8 of [45]), $\hat{\boldsymbol{\theta}}_{n,\mathrm{OPB}}$ (Theorem 3.1 of [7]) and IIE $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}^*}$ (Theorem 1.4.8) are consistent.

The MM is based on r empirical autocovariances, OPB based on r predictors, and IIE based on an AR(r) auxiliary model. Inspection of several empirical autocovariance functions of the squared returns $\boldsymbol{G}_n^2$ with $n = 10\,000$ revealed $r = 70$ as a suitable number of lags in most of the cases, and we choose $r = 70$ for all three estimators.

We compare the three estimators in a simulation study in Section 1.5.1. Then we show in Section 1.5.2, how the IIE based on simulations defined in (1.2.16) can reduce the finite sample bias of $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}^*}$ considerably. Finally, to understand how the condition $\Psi_{\boldsymbol{\theta}_0}(4) < 0$ affects the estimation, we investigate the finite sample bias of both IIEs for two different true parameter values $\boldsymbol{\theta}_0^{(1)}$ and $\boldsymbol{\theta}_0^{(2)}$ satisfying $\Psi_{\boldsymbol{\theta}_0^{(2)}}(4) < \Psi_{\boldsymbol{\theta}_0}(4) < \Psi_{\boldsymbol{\theta}_0^{(1)}}(4) < 0$, where $\Psi_{\boldsymbol{\theta}_0^{(1)}}(4)$ is near zero.

1.5.1 Simulation results

The computations are performed using the R software (R Core Team [87]). Simulation of the COGARCH process and computation of $\hat{\boldsymbol{\theta}}_{n,\mathrm{MM}}$ and $\hat{\boldsymbol{\theta}}_{n,\mathrm{OPB}}$ are performed with the COGARCH R package from Bibbona et al. [8] (see also the YUIMA R package of Iacus et al. [48] for the simulation and estimation of higher order COGARCH models). We first compute $\hat{\boldsymbol{\theta}}_{n,\mathrm{MM}}$ based on the sample $\boldsymbol{G}_n^2$. The estimators $\hat{\boldsymbol{\theta}}_{n,\mathrm{OPB}}$ and $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}^*}$ are computed via the optimization routine *optim* in R, which requires an initial parameter value and we take $\hat{\boldsymbol{\theta}}_{n,\mathrm{MM}}$ as this value. To compute $\hat{\boldsymbol{\pi}}_n$ in (1.5.2) we use the YWE from Definition 1.2.3 and take the identity matrix for $\boldsymbol{\Omega}$ to compute $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}^*}$. In principle, there is an optimal choice of Ω (see Remark 3 of de Luna and Genton [25] and Prop. 4 of Gourieroux et al. [39]). It depends on the covariance matrix $\boldsymbol{\Sigma}$ of the auxiliary model in (1.3.9) (see also Remark 4.24(b) of [31]). This matrix depends on an infinite series and on covariances between COGARCH returns to the powers 2,4,6 and 8, and has no explicit expression. According to Remark 3 of [25] and empirical evidence reported on p. S97f of Gourieroux et al. [39] the gain of efficiency when using the optimal weight matrix is negligible, so that we only consider estimators based on the identity matrix for Ω.

We focus on the YWE for the auxiliary model, a comparison including the LSE will be given in the first author's PhD Thesis. The estimator $\hat{\boldsymbol{\theta}}_{n,\mathrm{OPB}}$ only returns a result when $\Psi_{\hat{\boldsymbol{\theta}}_{n,\mathrm{OPB}}}(4) < 0$. The estimators $\hat{\boldsymbol{\theta}}_{n,\mathrm{MM}}$ and $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}^*}$ always return a value. The results are given in Table 1.1, where we excluded those paths for which the condition $\Psi_{\hat{\boldsymbol{\theta}}_n}(4) < 0$ is not satisfied for at least one of the estimators compared here. The results are based on 1 000 independent samples of COGARCH squared returns.

The results in Table 1.1 for the estimators $\hat{\boldsymbol{\theta}}_{n,\mathrm{MM}}$ and $\hat{\boldsymbol{\theta}}_{n,\mathrm{OPB}}$ are similar to those of Table 2 of Bibbona and Negri [7]. The OPB estimator has the smallest RMSE. The MM has the smallest relative bias for the parameter φ, and the OPB the smallest for β and ϕ. The estimator $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}^*}$ performed similarly to $\hat{\boldsymbol{\theta}}_{n,\mathrm{MM}}$, but it has a large bias for the

Table 1.1: Performance assessment based on 1 000 independent samples of COGARCH squared returns $\boldsymbol{G}_n^2$ for $n = 10\,000$, sampled with parameter values $\beta_0 = 0.04$, $\eta_0 = 0.053$ and $\varphi_0 = 0.038$: mean, standard deviation (Std), root mean squared error (RMSE), and relative bias (RB). Both IIEs $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}^*}$ in (1.5.2) and $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$ in (1.5.3) used the identity matrix for $\boldsymbol{\Omega}$. The IIE $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$ is based on $K = 100$ simulated paths.

		Mean	Std	RMSE	RB
	$\hat{\beta}$	0.04698	0.02032	0.02148	0.17457
$\hat{\boldsymbol{\theta}}_{n,\mathrm{II}^*}$	$\hat{\eta}$	0.05038	0.01482	0.01504	-0.04939
	$\hat{\varphi}$	0.03243	0.00994	0.01139	-0.14663
	$\hat{\beta}$	0.05226	0.01805	0.02182	0.30658
$\hat{\boldsymbol{\theta}}_{n,\mathrm{MM}}$	$\hat{\eta}$	0.05662	0.01576	0.01616	0.06827
	$\hat{\varphi}$	0.03667	0.01023	0.01031	-0.03513
	$\hat{\beta}$	0.04439	0.01609	0.01667	0.10965
$\hat{\boldsymbol{\theta}}_{n,\mathrm{OPB}}$	$\hat{\eta}$	0.05274	0.01317	0.01317	-0.00489
	$\hat{\varphi}$	0.03583	0.00815	0.00843	-0.05712
	$\hat{\beta}$	0.04204	0.02032	0.02041	0.05105
$\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$	$\hat{\eta}$	0.05318	0.01623	0.01622	0.00336
	$\hat{\varphi}$	0.03661	0.00955	0.00965	-0.03661

parameters β and φ. This is probably due to the fact that $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}^*}$ depends on $\hat{\boldsymbol{\pi}}_n$, which is a biased estimator of $\boldsymbol{\pi}$ even for AR models with i.i.d. noise as shown in Shaman and Stine [93]. The auxiliary AR model from Proposition 1.2.2 has stationary and ergodic residuals, and certainly $\hat{\boldsymbol{\pi}}_n$ has a bias, which propagates to the IIE. As a remedy, we use the IIE based on simulations and show that it can reduce the bias of $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}^*}$, it also outperforms $\hat{\boldsymbol{\theta}}_{n,\mathrm{MM}}$ and $\hat{\boldsymbol{\theta}}_{n,\mathrm{OPB}}$.

1.5.2 Finite sample bias

In Gourieroux et al. [40, 41] it is shown that Indirect Inference based on simulations can reduce the finite sample bias considerably, in particular, when the bias originates from the estimator of the auxiliary model. The idea of the bias reduction is that the IIE

$$\hat{\boldsymbol{\theta}}_{n,\mathrm{II}} = \underset{\boldsymbol{\theta}\in\Theta}{\arg\min} \left\| \hat{\boldsymbol{\pi}}_n - \frac{1}{K}\sum_{k=1}^{K} \hat{\boldsymbol{\pi}}_{n,k}(\boldsymbol{\theta}) \right\|_{\boldsymbol{\Omega}}, \quad K \in \mathbb{N}, \tag{1.5.3}$$

from Definition 1.2.5 finds a $\boldsymbol{\theta} \in \Theta$ which minimizes the distance between two biased estimators, $\hat{\boldsymbol{\pi}}_n$ and $\frac{1}{K}\sum_{k=1}^{K} \hat{\boldsymbol{\pi}}_{n,k}(\boldsymbol{\theta})$. As they have a similar bias, they have a chance to cancel. We proceed to investigate the finite sample performance of the estimator $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$ in (1.5.3).

Table 1.2: Performance assessment based on 1 000 independent samples of COGARCH squared returns $\boldsymbol{G}_n^2$ for $n = 5\,000$ and $n = 7\,500$, sampled with parameter values $\beta_0 = 0.04$, $\eta_0 = 0.053$ and $\varphi_0 = 0.038$: mean, standard deviation (Std), root mean squared error (RMSE) and relative bias (RB). Both IIEs $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}^*}$ in (1.5.2) and $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$ in (1.5.3) used the identity matrix for $\boldsymbol{\Omega}$. The IIE $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$ is based on $K = 100$ simulated paths.

		$n = 5\,000$			
		Mean	Std	RMSE	RB
$\hat{\boldsymbol{\theta}}_{n,\mathrm{II}^*}$	$\hat{\beta}$	0.04710	0.02196	0.02307	0.17738
	$\hat{\eta}$	0.04977	0.02036	0.02061	-0.06094
	$\hat{\varphi}$	0.03168	0.01317	0.01461	-0.16637
$\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$	$\hat{\beta}$	0.04999	0.03228	0.03377	0.24968
	$\hat{\eta}$	0.05935	0.02458	0.02538	0.11974
	$\hat{\varphi}$	0.03990	0.01379	0.01391	0.04989
		$n = 7\,500$			
		Mean	Std	RMSE	RB
$\hat{\boldsymbol{\theta}}_{n,\mathrm{II}^*}$	$\hat{\beta}$	0.05093	0.02015	0.02291	0.27323
	$\hat{\eta}$	0.05401	0.01786	0.01788	0.01896
	$\hat{\varphi}$	0.03439	0.01158	0.01212	-0.09502
$\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$	$\hat{\beta}$	0.04181	0.02375	0.02381	0.04537
	$\hat{\eta}$	0.05322	0.01897	0.01896	0.00408
	$\hat{\varphi}$	0.03668	0.01093	0.01101	-0.03487

According to [41], the number of simulated paths K in (1.5.3) has to be large enough to ensure that $\mathbb{E}\,\hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta})$ is well approximated by $\frac{1}{K}\sum_{k=1}^{K}\hat{\boldsymbol{\pi}}_{n,k}(\boldsymbol{\theta})$ for all $\boldsymbol{\theta}$ appearing in the optimization algorithm. Furthermore, the asymptotic variance of the IIE decreases with K (see Eq. (3.5.10)). To compute $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$ we need to evaluate the function

$$\boldsymbol{\theta} \mapsto \frac{1}{K}\sum_{k=1}^{K}\hat{\boldsymbol{\pi}}_{n,k}(\boldsymbol{\theta}) \tag{1.5.4}$$

for all $\boldsymbol{\theta}$ giving a representation of the parameter space. To compute (1.5.4) for a fixed $\boldsymbol{\theta}$, we simulate K independent samples $\boldsymbol{G}_{n,k}(\boldsymbol{\theta}) := (G_i^{(k)}(\boldsymbol{\theta}))_{i=1}^n$ for $k = 1, \ldots, K$. For different $\boldsymbol{\theta}$ we use the same pseudo-random numbers to generate the K independent samples, which turns (1.5.4) into a deterministic function of $\boldsymbol{\theta}$ and thus suitable for optimization.

In order to save computation time when computing (1.5.4) we use for every simulated path the fact that $\boldsymbol{G}_{n,k}(\boldsymbol{\theta}) = \sqrt{\beta}\boldsymbol{G}_{n,k}((1, \eta, \phi))$ (see Remark 1.4.5) and thus it follows

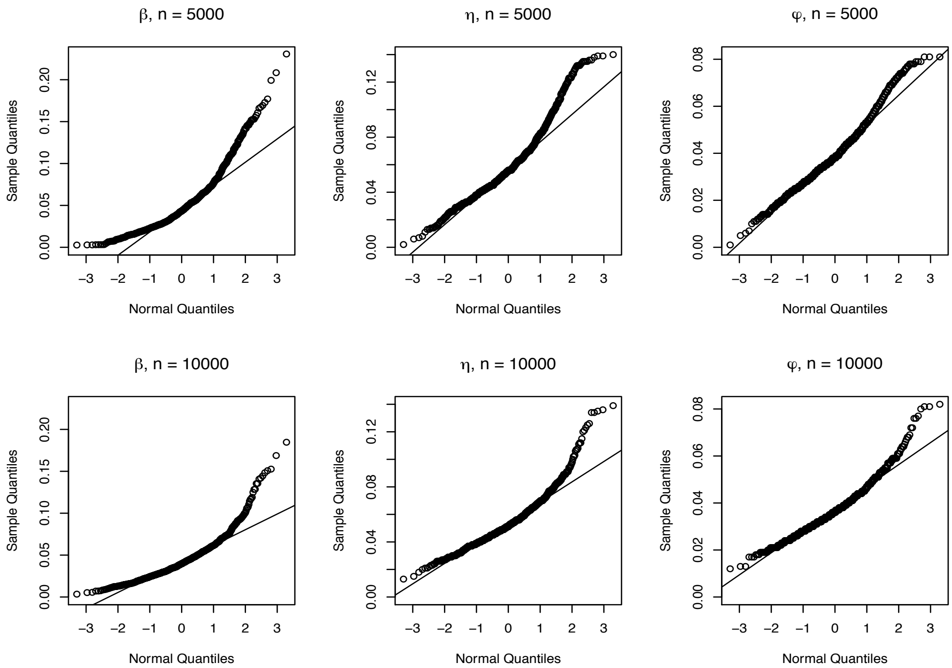


Figure 1.1: QQ plots of the estimators $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$ of $\boldsymbol{\theta}_0$ as in Table 1.1 for $n = 5\,000$ (top line) and $n = 10\,000$ (bottom line).

from Definition 1.2.3 that

$$\hat{\boldsymbol{\pi}}_{n,k}(\boldsymbol{\theta}) = \begin{pmatrix} \hat{\mu}_n(\boldsymbol{\theta}) \\ \hat{\boldsymbol{a}}_{n,k}(\boldsymbol{\theta}) \\ \hat{\gamma}_{n,k}(0; \boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} \beta\hat{\mu}_{n,k}((1,\eta,\phi)) \\ \hat{\boldsymbol{a}}_{n,k}((1,\eta,\phi)) \\ \beta^2\hat{\gamma}_{n,k}(0; (1,\eta,\phi)) \end{pmatrix}. \tag{1.5.5}$$

As it is computationally impossible to perform the optimization (1.5.3) for all $\boldsymbol{\theta} \in \Theta$, we have to restrict Θ in a reasonable way, and we restrict Θ to values in the set

$$\Theta_{\mathrm{rest}} := \{\boldsymbol{\theta} \in \Theta : \Psi_{\boldsymbol{\theta}}(4) < 0,\, \boldsymbol{\theta} \in (0, \hat{\beta}_{\max}) \times (0, \hat{\eta}_{\max}) \times (0, \hat{\varphi}_{\max})\}$$

where $\hat{\beta}_{\max}$, $\hat{\eta}_{\max}$ and $\hat{\varphi}_{\max}$ are upper bounds for the estimated parameters from Table 1.1 for all $1\,000$ independent samples $\boldsymbol{G}_n^2$ and all estimators.

For $K = 100$ and $n = 10\,000$, every evaluation of (1.5.4) takes approximately 13 minutes on a personal computer. The next goal would be to evaluate (1.5.3) using a gradient based routine. This is out of reach with respect to computation time. As a remedy we adopt the strategy of precomputing (1.5.4) on a fine grid $\Theta_{\mathrm{grid}} \subset \Theta_{\mathrm{rest}}$. The set Θ_{grid}

Table 1.3: Performance assessment based on 1 000 independent samples of COGARCH squared returns $\boldsymbol{G}_n^2$ for $n = 10\,000$, sampled with parameter values $\boldsymbol{\theta}_0^{(1)} = (0.04, 0.051, 0.040)$ and $\boldsymbol{\theta}_0^{(2)} = (0.04, 0.055, 0.036)$: mean, standard deviation (Std), root mean squared error (RMSE) and relative bias (RB). Both IIEs $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}^*}$ in (1.5.2) and $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$ in (1.5.3) used the identity matrix for $\boldsymbol{\Omega}$. The IIE $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$ is based on $K = 100$ simulated paths.

		$\boldsymbol{\theta}_0^{(1)} = (0.04, 0.051, 0.040)$			
		Mean	Std	RMSE	RB
$\hat{\boldsymbol{\theta}}_{n,\mathrm{II}^*}$	$\hat{\beta}$	0.05452	0.02341	0.02754	0.36298
	$\hat{\eta}$	0.05027	0.01294	0.01296	-0.01433
	$\hat{\varphi}$	0.03478	0.00857	0.01003	-0.13046
$\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$	$\hat{\beta}$	0.04586	0.02133	0.02211	0.14658
	$\hat{\eta}$	0.05142	0.01421	0.01421	0.00827
	$\hat{\varphi}$	0.03788	0.00872	0.00897	-0.05300
		$\boldsymbol{\theta}_0^{(2)} = (0.04, 0.055, 0.036)$			
		Mean	Std	RMSE	RB
$\hat{\boldsymbol{\theta}}_{n,\mathrm{II}^*}$	$\hat{\beta}$	0.04315	0.01858	0.01883	0.07886
	$\hat{\eta}$	0.05177	0.01603	0.01635	-0.05867
	$\hat{\varphi}$	0.03109	0.01057	0.01165	-0.13643
$\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$	$\hat{\beta}$	0.04084	0.01829	0.01830	0.02090
	$\hat{\eta}$	0.05571	0.01666	0.01667	0.01295
	$\hat{\varphi}$	0.03570	0.00948	0.00948	-0.00828

was created by generating an equally spaced grid on Θ_{rest} with componentwise distance for the parameters η and φ equal to 0.001 (resulting in about 6 000 different points). The grid for the component β was then created with $\Delta = 0.001$, but without the need to simulate the COGARCH path again by using the relation in (1.5.5). Afterwards, with COGARCH returns $\boldsymbol{G}_n^2$ generated independently from the samples $\boldsymbol{G}_{n,k}(\boldsymbol{\theta}), k = 1, \ldots, K$, applied to compute (1.5.4), we compute $\hat{\boldsymbol{\pi}}_n$, and the estimator $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$ is then simply given by

$$\underset{\boldsymbol{\theta} \in \Theta_{\mathrm{grid}}}{\arg\min} \left\| \hat{\boldsymbol{\pi}}_n - \frac{1}{K} \sum_{k=1}^{K} \hat{\boldsymbol{\pi}}_{n,k}(\boldsymbol{\theta}) \right\|_{\boldsymbol{\Omega}},$$

where we choose the identity matrix for $\boldsymbol{\Omega}$. The results are presented in the bottom line of Table 1.1. We notice a significant bias reduction for the simulation based estimator $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$ compared to $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}^*}$. The standard deviation of the estimator for η is slightly larger for $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$, but this is expected since the simulations increase the asymptotic variance by a factor of $(1 + \frac{1}{K})$ as can be seen from (3.5.10). The relative bias of the components of $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$ is also smaller than that of $\hat{\boldsymbol{\theta}}_{n,\mathrm{MM}}$ and $\hat{\boldsymbol{\theta}}_{n,\mathrm{OPB}}$. Since the standard deviations of the components of $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$ are larger than for those of $\hat{\boldsymbol{\theta}}_{n,\mathrm{OPB}}$ and the bias reduction is comparable for the parameters η and φ, the RMSE does not seem to improve, even though the bias of $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$ is smaller.

We also compare the performance of the IIE with and without simulation for different sample sizes n with $\boldsymbol{\theta}_0$ as in Table 1.1. The results are given in Table 1.2. For $n = 5\,000$ we only observe a bias reduction of $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$ for $\hat{\eta}$, whereas the bias reduction of $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$ is noticeable for all three components already for $n = 7\,500$ and of course for $n = 10\,000$; cf. Table 1.1.

We also can see in Figure 1.1 that for $n = 5\,000$ and $n = 10\,000$ the asymptotic normality of $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$ has not yet been reached, although some improvement for growing sample sizes is visible in the QQ plots of $\hat{\beta}$ and $\hat{\eta}$, however not for $\hat{\varphi}$.

To clarify if the bias reduction of $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$ depends on the choice of the true parameter values we perform a simulation study with two different values: $\boldsymbol{\theta}_0^{(1)} = (0.04, 0.051, 0.040)$ and $\boldsymbol{\theta}_0^{(2)} = (0.04, 0.055, 0.036)$. Both values are in the stationarity region with

$$\Psi_{\boldsymbol{\theta}_0^{(1)}}(4) = -0.0060, \quad \Psi_{\boldsymbol{\theta}_0}(4) = -0.0261, \quad \Psi_{\boldsymbol{\theta}_0^{(2)}}(4) = -0.0460.$$

The results are presented in Table 1.3. As for $\boldsymbol{\theta}_0$ in Table 1.1, they also show significant bias reduction for both values $\boldsymbol{\theta}_0^{(1)}$ and $\boldsymbol{\theta}_0^{(2)}$ for the estimator $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}}$ based on simulations, when compared to $\hat{\boldsymbol{\theta}}_{n,\mathrm{II}^*}$. However, the bias for $\hat{\beta}^{(1)}$ is much larger than for $\hat{\beta}$ and $\hat{\beta}^{(2)}$ reflecting the fact that $\Psi_{\boldsymbol{\theta}_0^{(1)}}(4)$ is very close to zero. The estimators $\hat{\eta}^{(1)}$ and $\hat{\varphi}^{(1)}$ seem to be robust with respect to this fact. Additionally, the relative biases for $\hat{\beta}^{(2)}$ and $\hat{\varphi}^{(2)}$ are even smaller than those for $\hat{\beta}$ and $\hat{\varphi}$ and $\hat{\beta}^{(1)}$ and $\hat{\varphi}^{(1)}$.

1.6 Appendix to Section 1.4.2

The first Lemma states important properties about moments of a continuous version of a stochastic process found via Kolmogorov's continuity criterion. The property stated in (1.6.1) is needed for the application of a uniform SLLN in Theorem 1.4.8. Lemma 1.6.4 is used to compute $\nabla_\theta \sigma_0^2(\boldsymbol{\theta})$ and $\nabla_\theta^2 \sigma_0^2(\boldsymbol{\theta})$, needed to find a continuous version of the map $\boldsymbol{\theta} \mapsto \int_{(i-1)\Delta}^{i\Delta} \sigma_s(\boldsymbol{\theta}) \mathrm{d}L_s$ in Theorem 1.4.6, and of $\boldsymbol{\theta} \mapsto \int_{(i-1)\Delta}^{i\Delta} \nabla_\theta \sigma_s(\boldsymbol{\theta}) \mathrm{d}L_s$ in Lemma 1.4.10.

Lemma 1.6.1. *Let $(X(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta)$ be a stochastic process with compact parameter space $\Theta \subset \mathbb{R}_+^d$ for $d \in \mathbb{N}$. Assume that there exist positive constants p, c, ϵ such that for all $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$:*

$$\mathbb{E}|X^{(c)}(\boldsymbol{\theta}_1) - X(\boldsymbol{\theta}_2)|^p \leq c\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^{d+\epsilon}.$$

Then there exists a continuous version $(X^{(c)}(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta)$ of $(X(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta)$ such that for $q \in [0, p)$

$$\mathbb{E} \sup_{\theta \in \Theta} |X^{(c)}(\boldsymbol{\theta})|^q < \infty. \tag{1.6.1}$$

Proof. Since Θ is compact we can use the Heine-Borel theorem to find a finite collection of open sets $(\Theta_j)_{j=1}^N$ such that $\Theta \subset \cup_{j=1}^N \Theta_j$ and $\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| \leq \delta^*$ for every $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta_j$. Choosing an arbitrary $\boldsymbol{\theta}_j \in \Theta_j \cap \Theta$ for $j = 1, \dots, N$ and using $|a-b|^q \leq 2^{q-1}|a^q - b^q|$ gives for $q < p$,

$$\begin{aligned}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|X^{(c)}(\boldsymbol{\theta})|^q &\leq \sum_{j=1}^N \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta_j}|X^{(c)}(\boldsymbol{\theta})|^q \\ &\leq \sum_{j=1}^N 2^{q-1}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta_j}\left\{|X^{(c)}(\boldsymbol{\theta}) - X^{(c)}(\boldsymbol{\theta}_j)|^q + |X^{(c)}(\boldsymbol{\theta}_j)|^q\right\} \\ &\leq 2^{q-1}\sum_{j=1}^N (1 + \mathbb{E}|X^{(c)}(\boldsymbol{\theta}_j)|^q) < \infty,\end{aligned}$$

since $\mathbb{E}|X^{(c)}(\boldsymbol{\theta})|^p < \infty$ for all $\boldsymbol{\theta} \in \Theta$. □

The following Lemma is well-known from Analysis, and can be found for instance as Exercise 6 in Ch. 15.7 of [60].

Lemma 1.6.2. *Suppose that $g : \mathbb{R}^p \to \mathbb{R}^q$ is continuous and that*

$$\sup_{\boldsymbol{\theta}\in\Theta}\|f_n(\boldsymbol{\theta}) - f(\boldsymbol{\theta})\| \stackrel{\text{a.s.}}{\to} 0, \quad n \to \infty,$$

where $(f_n(\boldsymbol{\theta}))_{n\in\mathbb{N}}$ is a sequence of random vectors in $\mathbb{R}^p$, $f : \Theta \in \mathbb{R}^d \mapsto \mathbb{R}^p$ is a deterministic function and Θ is compact. Then,

$$\sup_{\boldsymbol{\theta}\in\Theta}\|g(f_n(\boldsymbol{\theta})) - g(f(\boldsymbol{\theta}))\| \stackrel{\text{a.s.}}{\to} 0, \quad n \to \infty.$$

Lemma 1.6.3. *Let $p, b \geq 1$, $a, k \geq 0$, $\boldsymbol{\theta} \in \mathcal{M}$ be fixed and $(K_s(\tilde{\varphi}))_{s\geq 0}$ as defined in* (1.4.1) *for fixed $\tilde{\varphi} > 0$. If $\mathbb{E}|L_1|^{2p(1+\epsilon)} < \infty$ and $\Psi_{\boldsymbol{\theta}}(p(1+\epsilon)) < 0$ for some $\epsilon > 0$, then*

$$\mathbb{E}\Big(\int_0^\infty (s^a + s^k K_s^b(\tilde{\varphi}))e^{-Y_s(\boldsymbol{\theta})}\mathrm{d}s\Big)^p < \infty.$$

Proof. The proof is similar to the proof of Proposition 4.1 of Lindner and Maller [64]. For every $j \in \mathbb{N}_0$ define $Q_j(\boldsymbol{\theta}) := \int_j^{j+1}(s^a + s^k K_s^b(\tilde{\varphi}))e^{-Y_s(\boldsymbol{\theta})}\mathrm{d}s$. Then

$$\begin{aligned}
&\mathbb{E}Q_j^p(\boldsymbol{\theta}) \\
= \;& \mathbb{E}\Big(\int_j^{j+1}(s^a + s^k K_s^b(\tilde{\varphi}))e^{-Y_s(\boldsymbol{\theta})}\mathrm{d}s\Big)^p \\
\le \;& \mathbb{E}\Big(\sup_{j\le s\le j+1}(s^a + s^k K_s^b(\tilde{\varphi}))e^{-Y_s(\boldsymbol{\theta})}\Big)^p \\
\le \;& \mathbb{E}\Big(((j+1)^a + (j+1)^k K_{j+1}^b(\tilde{\varphi}))^p \sup_{j\le s\le j+1} e^{-pY_s(\boldsymbol{\theta})}\Big) \qquad (1.6.2)\\
\le \;& \Big(\mathbb{E}((j+1)^a + (j+1)^k K_{j+1}^b(\tilde{\varphi}))^{p(1+\epsilon)/\epsilon}\Big)^{\epsilon/(1+\epsilon)}\Big(\mathbb{E}\sup_{j\le s\le j+1} e^{-p(1+\epsilon)Y_s(\boldsymbol{\theta})}\Big)^{1/(1+\epsilon)}
\end{aligned}$$

by the Hölder inequality. Since by Lemma 1.4.3 $(K_s(\tilde{\varphi}))_{s\ge 0}$ is a Lévy process with moments of all orders, repeated differentiation of the characteristic function of $K_{j+1}(\tilde{\varphi})$ gives a constant $c > 0$ such that

$$\mathbb{E}((j+1)^a + (j+1)^k K_{j+1}^b(\tilde{\varphi}))^{p(1+\epsilon)/\epsilon} \le c(j+1)^{mp(1+\epsilon)/\epsilon}, \qquad (1.6.3)$$

where $m = a+k+b$. Since the process $(e^{Y_s(\boldsymbol{\theta})-s\Psi_{\boldsymbol{\theta}}(1)})_{s\ge 0}$ is a martingale, we can use Doob's martingale inequality, the Laplace transform in (1.2.2) and the fact that $\Psi_{\boldsymbol{\theta}}(1) < 0$ to get

$$\begin{aligned}
\mathbb{E}\sup_{j\le s\le j+1} e^{-p(1+\epsilon)Y_s(\boldsymbol{\theta})} &\le e^{-(j+1)p(1+\epsilon)\Psi_{\boldsymbol{\theta}}(1)}\mathbb{E}\sup_{j\le s\le j+1} e^{-p(1+\epsilon)Y_s(\boldsymbol{\theta})+sp(1+\epsilon)\Psi_{\boldsymbol{\theta}}(1)} \\
&\le e^{-(j+1)p(1+\epsilon)\Psi_{\boldsymbol{\theta}}(1)}\mathbb{E}e^{-p(1+\epsilon)Y_{j+1}(\boldsymbol{\theta})+p(1+\epsilon)(j+1)\Psi_{\boldsymbol{\theta}}(1)} \qquad (1.6.4)\\
&= \mathbb{E}e^{-p(1+\epsilon)Y_{j+1}(\boldsymbol{\theta})} \\
&= e^{(j+1)\Psi_{\boldsymbol{\theta}}(p(1+\epsilon))}.
\end{aligned}$$

Equation (1.6.2) together with (1.6.3) and (1.6.4) gives

$$\mathbb{E}Q_j^p(\boldsymbol{\theta}) \le c^*(j+1)^{mp}e^{(j+1)\Psi_{\boldsymbol{\theta}}(p(1+\epsilon))/(1+\epsilon)} < \infty, \qquad (1.6.5)$$

where $c^* = c^{\epsilon/(1+\epsilon)}$. Let $\alpha := \lfloor p \rfloor$ be the integer part of p and suppose that $p > \alpha$. Then

$$\begin{aligned}
\Big(\int_0^n (s^a + s^k K_s^b(\tilde{\varphi}))e^{-Y_s(\boldsymbol{\theta})}\mathrm{d}s\Big)^p &= \Big(\sum_{j=0}^{n-1} Q_j(\boldsymbol{\theta})\Big)^p \\
&= \sum_{j_1=0}^{n-1}\cdots\sum_{j_\alpha=0}^{n-1} Q_{j_1}(\boldsymbol{\theta})\dots Q_{j_\alpha}(\boldsymbol{\theta})\Big(\sum_{j_{\alpha+1}=0}^{n-1} Q_{j_{\alpha+1}}(\boldsymbol{\theta})\Big)^{p-\alpha} \\
&\le \sum_{j_1=0}^{n-1}\cdots\sum_{j_\alpha=0}^{n-1}\sum_{j_{\alpha+1}=0}^{n-1} Q_{j_1}(\boldsymbol{\theta})\dots Q_{j_\alpha}(\boldsymbol{\theta})Q_{j_{\alpha+1}}^{p-\alpha}(\boldsymbol{\theta}).
\end{aligned} \qquad (1.6.6)$$

If p is an integer the last sum in (1.6.6) disappears. By (1.6.5), for each $j = 1, \dots, \alpha + 1, Q_j \in \mathcal{L}^p$ so we can apply the Hölder inequality with $\frac{1}{p} + \cdots + \frac{1}{p} + \frac{p-\alpha}{p} = 1$ to the right-hand side of (1.6.6). This together with (1.6.5) gives

$$\begin{aligned} & \mathbb{E}\Big(\int_0^n (s^a + s^k K_s^b(\tilde{\varphi}))e^{-Y_s(\boldsymbol{\theta})}\mathrm{d}s\Big)^p \\ \leq\ & \sum_{j_1=0}^{n-1} \cdots \sum_{j_\alpha=0}^{n-1} \sum_{j_{\alpha+1}=0}^{n-1} (\mathbb{E}(Q_{j_1}^p(\boldsymbol{\theta}))^{\frac{1}{p}} \dots (\mathbb{E}Q_{j_\alpha}^p(\boldsymbol{\theta}))^{\frac{1}{p}} (\mathbb{E}(Q_{j_{\alpha+1}}^p(\boldsymbol{\theta}))^{\frac{p-\alpha}{p}} \qquad (1.6.7) \\ \leq\ & c^* \Big(\sum_{j=0}^{n-1} (j+1)^m e^{(j+1)/(p(1+\epsilon))\Psi_{\boldsymbol{\theta}}(p(1+\epsilon))}\Big)^\alpha \times \\ & \qquad \Big(\sum_{j=0}^{n-1} (j+1)^{m(p-\alpha)} e^{(j+1)(p-\alpha)/(p(1+\epsilon))\Psi_{\boldsymbol{\theta}}(p(1+\epsilon))}\Big). \end{aligned}$$

Since $\Psi_{\boldsymbol{\theta}}(p(1+\epsilon)) < 0$ both series in (1.6.7) converge. The monotone convergence theorem applied to the expectation in the first line of (1.6.7) gives the result. □

Lemma 1.6.4. *Let $\boldsymbol{\theta} = (\beta, \eta, \varphi)$ with $\beta, \eta, \varphi > 0$ and consider the process $(Y_s(\boldsymbol{\theta}))_{s\geq 0}$ as in* (1.2.1)*. Let $K_s(\varphi)$ be as defined in* (1.4.1)*. Then the following assertions hold:*
(a) For every fixed $s > 0$,

$$\nabla_{\eta,\varphi}(e^{-Y_s(\boldsymbol{\theta})}) = e^{-Y_s(\boldsymbol{\theta})} \begin{pmatrix} -s \\ K_s(\varphi) \end{pmatrix}. \qquad (1.6.8)$$

(b) If $\mathbb{E}|L_1|^{2(1+\epsilon)} < \infty$ and $\Psi_{\boldsymbol{\theta}}(1+\epsilon) < 0$ for some $\epsilon > 0$, then

$$\nabla_{\eta,\varphi}\Big(\int_0^\infty e^{-Y_s(\boldsymbol{\theta})}\mathrm{d}s\Big) = \int_0^\infty e^{-Y_s(\boldsymbol{\theta})} \begin{pmatrix} -s \\ K_s(\varphi) \end{pmatrix} \mathrm{d}s \qquad (1.6.9)$$

$$\nabla^2_{\eta,\varphi}\Big(\int_0^\infty e^{-Y_s(\boldsymbol{\theta})}\mathrm{d}s\Big) = \int_0^\infty e^{-Y_s(\boldsymbol{\theta})} \begin{pmatrix} s^2 & -sK_s(\varphi) \\ -sK_s(\varphi) & (d_s^2(\varphi) + d_s'(\varphi)) \end{pmatrix} \mathrm{d}s. \qquad (1.6.10)$$

Proof. (a) The partial derivatives of $Y_s(\boldsymbol{\theta}) = \eta s - \sum_{0<u\leq s} \log(1 + \varphi(\Delta L_u)^2)$ are given by

$$\frac{\partial Y_s(\boldsymbol{\theta})}{\partial \eta} = s \quad \text{and} \quad \frac{\partial Y_s(\boldsymbol{\theta})}{\partial \varphi} = -K_s(\varphi),$$

where the derivative with respect to φ follows by dominated convergence since we have the following bound independent of φ:

$$K_s(\varphi) \leq \sum_{0<u\leq s} (\Delta L_u)^2 < \infty. \qquad (1.6.11)$$

A simple application of the chain rule gives (1.6.8).
(b) It follows from Lemma 1.4.3(a) that we can find a collection of points $(\boldsymbol{\theta}_j^*)_{j=1}^N$ in $\mathcal{M}$ such that

$$\sup_{\boldsymbol{\theta}\in\Theta} e^{-Y_s(\boldsymbol{\theta})} \leq \sum_{j=1}^N e^{-Y_s(\boldsymbol{\theta}_j^*)}, \quad s \geq 0. \tag{1.6.12}$$

The first derivative of $K_s(\varphi)$ follows from dominated converge with the upper bound in (1.6.11) and is given by

$$K_s'(\varphi) = -\sum_{0<u\leq s} \frac{(\Delta L_u)^4}{(1+\varphi(\Delta L_u)^2)^2}, \quad s \geq 0.$$

Now, similar calculations as in (1.4.2) show that $|K_s'(\varphi)| \leq K_s(\varphi_*)/\varphi_*$ for φ_* as defined in (1.4.22). This combined with (1.6.12) allows us to obtain an upper bound for the sum of the bounds of the absolute values of the integrals at the r.h.s. of (1.6.9) and (1.6.10) given by

$$\sum_{j=1}^N \int_0^\infty e^{-Y_s(\boldsymbol{\theta}_j^*)}\big(s + s^2 + K_s(\varphi_*)(1 + 2s + 1/\phi_*) + d_s^2(\varphi_*)\big)\mathrm{d}s. \tag{1.6.13}$$

Since $\mathbb{E}|L_1|^{2(1+\epsilon)} < \infty$ and $\Psi_{\boldsymbol{\theta}_j^*}(1+\epsilon) < 0$ for all $j = 1, \ldots, N$ we can apply Lemma 1.6.3 with $p = 1$ to prove that the integral in (1.6.13) has finite first moment and is therefore well defined. This allows us to use dominated convergence to differentiate under the integral sign and then use the chain and product rule combined with (1.6.8) to obtain (1.6.9) and (1.6.10). □

Lemma 1.6.5. *Let $p \geq 1$ and $k \in \{1, 2\}$. If $\mathbb{E}|L_1|^{2kp(1+\epsilon)} < \infty$ for some $\epsilon > 0$, then*

$$\mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta^{(p(1+\epsilon))}} \|\nabla_{\eta,\varphi}^k \sigma_0^2(\boldsymbol{\theta})\|^p < \infty \quad \textit{and} \quad \mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta^{(kp(1+\epsilon))}} \|\nabla_{\eta,\varphi}^k \sigma_0(\boldsymbol{\theta})\|^p < \infty.$$

Proof. For $k \in \{1, 2\}$ let R_{kp} denote the integral defined in (1.6.13) with $(\boldsymbol{\theta}_j^*)_{j=1}^N \in \mathcal{M}^{(kp(1+\epsilon))}$ as in (1.4.3). By the same argument preceding (1.6.13) and from (1.2.4) we get

$$\sup_{\boldsymbol{\theta}\in\Theta^{(kp(1+\epsilon))}} \Big(\|\nabla_{\eta,\varphi}\sigma_0^2(\boldsymbol{\theta})\| + \|\nabla_{\eta,\varphi}^2\sigma_0^2(\boldsymbol{\theta})\|\Big) \leq c\beta^* R_{kp}, \quad k = 1, 2, \tag{1.6.14}$$

where $c > 0$ and

$$\beta^* = \sup\{\beta > 0 : (\beta, \eta, \varphi) \in \Theta\} < \infty. \tag{1.6.15}$$

Since from Lemma 1.4.3(b) we know that $\sigma_0(\boldsymbol{\theta}) \geq \sigma^* > 0$, the chain rule implies that

$$\|\nabla_{\eta,\varphi}\sigma_0(\boldsymbol{\theta})\| \leq \frac{1}{\sigma^*}\|\nabla_{\eta,\varphi}\sigma_0^2(\boldsymbol{\theta})\|. \tag{1.6.16}$$

Using (1.6.16) combined with the chain rule for the second order derivative gives

$$\|\nabla^2_{\eta,\varphi}\sigma_0(\boldsymbol{\theta})\| \leq \frac{1}{4\sigma^*}\|\nabla^2_{\eta,\varphi}\sigma_0^2(\boldsymbol{\theta})| + \frac{1}{8(\sigma^*)^3}\|\nabla_{\eta,\varphi}\sigma_0^2(\boldsymbol{\theta})\|^2. \tag{1.6.17}$$

Using (1.6.14) combined with (1.6.16) gives

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta^{(p(1+\epsilon))}}\|\nabla_{\eta,\varphi}\sigma_0(\boldsymbol{\theta})\|^p \leq \mathbb{E}\Big(\frac{1}{\sigma^*}c\beta^* R_p\Big)^p < \infty,$$

by an application of Lemma 1.6.3. Now, (1.6.14) combined with (1.6.17) gives

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta^{(2p(1+\epsilon))}}\|\nabla^2_{\eta,\varphi}\sigma_0(\boldsymbol{\theta})\|^p \leq \mathbb{E}\Big(\frac{1}{4\sigma^*}c\beta^* R_{2p} + \frac{1}{8(\sigma^*)^3}(c\beta^* R_{2p})^2\Big)^p < \infty,$$

by an application of the Cauchy-Schwartz inequality and Lemma 1.6.3 with p replaced by $2p$. □

1.7 Appendix to Section 1.4.3

Lemmas 1.7.2 and 1.7.1 are used in the proof of Proposition 1.4.13 to control the convergence of arithmetic means defined in terms of the sequences $(G_i(\boldsymbol{\theta}))_{i\in\mathbb{N}}$ and $(\nabla_\theta G_i(\boldsymbol{\theta}))_{i\in\mathbb{N}}$ with $\nabla_\theta G_i(\boldsymbol{\theta})$ defined in the sense of Remark 1.4.12.

Lemma 1.7.1. *Let $\boldsymbol{\theta} = (\beta,\eta,\phi) =: (\theta_1,\theta_2,\theta_3)$ with $\beta,\eta,\phi > 0$ and $\Delta > 0$. Suppose that $\mathbb{E}|L_1|^2 < \infty$ and $\Psi_{\boldsymbol{\theta}}(1) < 0$. Let $(\sigma_t(\boldsymbol{\theta}))_{t\geq 0}$ be the stationary volatility process starting with $\sigma_0(\boldsymbol{\theta})$ as in* (1.2.4) *independent of L. Then for all three components of $\boldsymbol{\theta}$ the sequences*

$$\Big(\int_{(i-1)\Delta}^{i\Delta}\sigma_s(\boldsymbol{\theta})\mathrm{d}L_s, \int_{(i-1)\Delta}^{i\Delta}\frac{\partial}{\partial\theta_j}\sigma_s(\boldsymbol{\theta})\mathrm{d}L_s\Big)_{i\in\mathbb{N}}$$

are stationary and ergodic.

Proof. Without loss of generality assume $j = 1$. Define the i.i.d. sequence $(S_k)_{k\in\mathbb{Z}}$ with

$$S_k = (\Delta L_u, (k-1)\Delta < u \leq k\Delta).$$

We consider

$$((\sigma_s(\boldsymbol{\theta}), \boldsymbol{\theta}\in\Theta), (i-1)\Delta < s \leq i\Delta) =: g(\boldsymbol{\theta}, \boldsymbol{\theta}\in\Theta, (S_k)_{k=-\infty}^i)$$

as a measurable function of all relevant jumps ΔL_u. Additionally, since limits of differentiable functions are measurable, there exists a measurable map h such that

$$(\frac{\partial}{\partial\theta_1}\sigma_s(\boldsymbol{\theta}), (i-1)\Delta < s \leq i\Delta) = h((S_k)_{k=-\infty}^i, (\boldsymbol{\theta} + (c,0,0))_{c\in\mathbb{Q}}).$$

By observing that a stochastic integral is defined as a measurable map depending on the integrand and integrator processes, we can write

$$\int_{(i-1)\Delta}^{i\Delta} \sigma_s(\boldsymbol{\theta})\mathrm{d}L_s = g((S_k)_{k=-\infty}^i, \boldsymbol{\theta})$$

and

$$\int_{(i-1)\Delta}^{i\Delta} \frac{\partial}{\partial\theta_1}\sigma_s(\boldsymbol{\theta})\mathrm{d}L_s = h((S_k)_{k=-\infty}^i, (\boldsymbol{\theta} + (c,0,0))_{c\in\mathbb{Q}}).$$

Using Proposition 5 in Straumann and Mikosch [100] (see also Theorem 2.1 in Krengel [62]) we can conclude the stationarity and ergodicity of the process $(G_i(\boldsymbol{\theta}), \nabla G_i(\boldsymbol{\theta}))_{i\in\mathbb{N}}$ based on the stationarity and ergodicity of the sequence $(S_i)_{i\in\mathbb{Z}}$ and the measurability of g and h. □

Lemma 1.7.2. *If* $\mathbb{E}|L_1|^{4p(1+\epsilon)} < \infty$ *for some* $p > 5/2$ *and* $\epsilon > 0$ *then for every* $l \in \{1,2,3\}$ *and* $h \in \mathbb{N}_0$ *we have*

$$\sup_{\boldsymbol{\theta}\in\Theta^{(2p(1+\epsilon))}} \Big|\frac{\partial}{\partial\boldsymbol{\theta}_l}\Big(\frac{1}{n}\sum_{i=1}^{n-h} G_i^2(\boldsymbol{\theta})G_{i+h}^2(\boldsymbol{\theta})\Big) - \frac{\partial}{\partial\boldsymbol{\theta}_l}(\mathbb{E}G_1^2(\boldsymbol{\theta})G_{1+h}^2(\boldsymbol{\theta}))\Big| \xrightarrow{P} 0, \quad n\to\infty. \quad (1.7.1)$$

Proof. The proof follows closely the strategy in the proof of Proposition 5.5 of Fasen-Hartmann and Kimmig [31], which divides the proof into three steps: Pointwise convergence, local Hölder continuity, and stochastic equicontinuity. Let $l \in \{1,2,3\}$ and $h \in \mathbb{N}_0$ be fixed. Write $\hat{\mu}_n(h;\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n-h} G_i^2(\boldsymbol{\theta})G_{i+h}^2(\boldsymbol{\theta})$. Then, a simple application of the chain and product rule gives

$$\frac{\partial}{\partial\boldsymbol{\theta}_l}\hat{\mu}_n(h;\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n-h}\Big[2G_i(\boldsymbol{\theta})\Big(\frac{\partial}{\partial\boldsymbol{\theta}_l}G_i(\boldsymbol{\theta})\Big)G_{i+h}^2(\boldsymbol{\theta}) + 2G_i^2(\boldsymbol{\theta})G_{i+h}(\boldsymbol{\theta})\Big(\frac{\partial}{\partial\boldsymbol{\theta}_l}G_{j+h}(\boldsymbol{\theta})\Big)\Big]. \quad (1.7.2)$$

Step 1. Pointwise convergence. Let $\boldsymbol{\theta} \in \Theta^{(2p(1+\epsilon))}$ be fixed. It follows from Lemma 1.7.1 that the sequence $(G_i(\boldsymbol{\theta}), \frac{\partial}{\partial\boldsymbol{\theta}_l}G_i(\boldsymbol{\theta}))_{i\in\mathbb{N}}$ is stationary and ergodic. Additionally, it follows from the lemma's assumptions combined with Theorem 1.4.6, Lemma 1.4.10 and the Hölder inequality with $\frac{1}{5}+\frac{2}{5}+\frac{2}{5} = 1$ that

$$\begin{aligned} &\mathbb{E}G_1(\boldsymbol{\theta})\Big(\frac{\partial}{\partial\boldsymbol{\theta}_l}G_1(\boldsymbol{\theta})\Big)G_{1+h}^2(\boldsymbol{\theta}) \\ \leq\ & (\mathbb{E}G_1^5(\boldsymbol{\theta}))^{1/5}\left(\mathbb{E}\Big(\frac{\partial}{\partial\boldsymbol{\theta}_l}G_1(\boldsymbol{\theta})\Big)^{5/2}\right)^{2/5}(\mathbb{E}G_{1+h}^5(\boldsymbol{\theta}))^{2/5} < \infty. \end{aligned} \quad (1.7.3)$$

The same calculations in (1.7.3) can be applied to show that the expectation of the second term in the summation (1.7.2) is also finite. This allows us to apply Birkhoff convergence theorem to conclude that

$$\frac{\partial}{\partial\boldsymbol{\theta}_l}\hat{\mu}_n(h;\boldsymbol{\theta}) \xrightarrow{P} \mathbb{E}G_1^2(\boldsymbol{\theta})G_{1+h}^2(\boldsymbol{\theta}), \quad n\to\infty.$$

Step 2. $\frac{\partial}{\partial \boldsymbol{\theta}_l}\hat{\mu}_n(h;\boldsymbol{\theta})$ is locally Hölder-continuous on $\Theta^{(2p(1+\epsilon))}$. For $i\in\mathbb{N}$ let U_i and V_i be as defined in (1.4.6) and (1.4.16), respectively. By stationarity of $(G_i(\boldsymbol{\theta}),\boldsymbol{\theta}\in\Theta)_{i\in\mathbb{N}}$ and $(\frac{\partial}{\partial \boldsymbol{\theta}_l}G_i(\boldsymbol{\theta}),\boldsymbol{\theta}\in\Theta)_{i\in\mathbb{N}}$, $U_i\overset{\mathrm{d}}{=}U_1$, $V_i\overset{\mathrm{d}}{=}V_1$ and for every $\boldsymbol{\theta}_1,\boldsymbol{\theta}_2\in\Theta^{(2p(1+\epsilon))}$ with $\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|<1$ it follows from Theorem 1.4.6 and Lemma 1.4.10 that there exists $\gamma\in(0,1)$ such that for all $i\in\mathbb{N}$:

$$|G_i(\boldsymbol{\theta}_1)-G_i(\boldsymbol{\theta}_2)|\le U_i\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|^\gamma \quad\text{and}\quad \Big|\frac{\partial}{\partial \boldsymbol{\theta}_l}G_i(\boldsymbol{\theta}_1)-\frac{\partial}{\partial \boldsymbol{\theta}_l}G_i(\boldsymbol{\theta}_2)\Big|\le V_i\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|^\gamma.$$

Using the inequality

$$|a_1b_1c_1^2-a_2b_2c_2^2|\le|a_1||b_1||c_1+c_2||c_1-c_2|+|a_1||c_2^2||b_1-b_2|+|b_2c_2^2||a_1-a_2|,$$

valid for every $a_1,a_2,b_1,b_2,c_1,c_2\in\mathbb{R}$ gives for all $i\in\mathbb{N}$,

$$\begin{aligned}
&\Big|G_i(\boldsymbol{\theta}_1)\Big(\frac{\partial}{\partial \boldsymbol{\theta}_l}G_i(\boldsymbol{\theta}_1)\Big)G_{i+h}^2(\boldsymbol{\theta}_1)-G_i(\boldsymbol{\theta}_2)\Big(\frac{\partial}{\partial \boldsymbol{\theta}_l}G_i(\boldsymbol{\theta}_2)\Big)G_{i+h}^2(\boldsymbol{\theta}_2)\Big|\\
&\le 2\Big(\sup_{\boldsymbol{\theta}\in\Theta^{(2p(1+\epsilon))}}|G_i(\boldsymbol{\theta})|\Big|\frac{\partial}{\partial \boldsymbol{\theta}_l}G_i(\boldsymbol{\theta})\Big||G_{i+h}(\boldsymbol{\theta})|\Big)U_{i+h}\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|^\gamma\\
&+\Big(\sup_{\boldsymbol{\theta}\in\Theta^{(2p(1+\epsilon))}}|G_i(\boldsymbol{\theta})||G_{i+h}^2(\boldsymbol{\theta})|\Big)V_i\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|^\gamma\\
&+\Big(\sup_{\boldsymbol{\theta}\in\Theta^{(2p(1+\epsilon))}}\Big|\frac{\partial}{\partial \boldsymbol{\theta}_l}G_i(\boldsymbol{\theta})\Big||G_{i+h}^2(\boldsymbol{\theta})|\Big)U_i\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|^\gamma\\
&=: I_{i,h}\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|^\gamma.
\end{aligned}\tag{1.7.4}$$

Another application of the Hölder inequality combined with (1.4.7) and an analogous result for $\nabla_\theta G_i(\boldsymbol{\theta})$ gives $\mathbb{E}I_{1,h}<\infty$. Similar calculations as in (1.7.4) can be used to show that for all $i\in\mathbb{N}$

$$\Big|G_i^2(\boldsymbol{\theta}_1)G_{i+h}(\boldsymbol{\theta}_1)\Big(\frac{\partial}{\partial \boldsymbol{\theta}_l}G_{i+h}(\boldsymbol{\theta}_1)\Big)-G_i^2(\boldsymbol{\theta}_2)G_{i+h}(\boldsymbol{\theta}_2)\Big(\frac{\partial}{\partial \boldsymbol{\theta}_l}G_{j+h}(\boldsymbol{\theta}_2)\Big)\Big|\le I_{i,h}^*\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|^\gamma,\tag{1.7.5}$$

with $\mathbb{E}I_{1,h}^*<\infty$.

Step 3. Stochastic equicontinuity. Let $\xi,\nu>0$ and $0<\delta<\min\{1,\eta\xi/\mathbb{E}(I_{1,h}+I_{1,h}^*)\}$. Then, it follows from (1.7.2), (1.7.4), (1.7.5) and Markov's inequality that

$$\mathbb{P}\Big(\sup_{\substack{0<\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|<\delta\\ \boldsymbol{\theta}_1,\boldsymbol{\theta}_2\in\Theta^{(2p(1+\epsilon))}}}\Big|\frac{\partial}{\partial \boldsymbol{\theta}_l}\hat{\mu}_n(h;\boldsymbol{\theta}_1)-\frac{\partial}{\partial \boldsymbol{\theta}_l}\hat{\mu}_n(h;\boldsymbol{\theta}_2)\Big|>\eta\Big)\le\mathbb{E}(I_{1,h}+I_{1,h}^*)\frac{\delta^\gamma}{\eta}<\xi.$$

This together with the pointwise convergence in Step 1 allow us to conclude the uniform convergence in (1.7.1) by means of Theorem 10.2 in Pollard [84]. □

Chapter 2:

Indirect Inference for Time Series Using the Empirical Characteristic Function and Control Variates

2.1 Introduction

Let $(X_j)_{j\in\mathbb{Z}}$ be a stationary time series, whose distribution depends on $\theta \in \Theta \subset \mathbb{R}^q$ for some $q \in \mathbb{N}$. Denote by $\theta_0 \in \Theta$ the true parameter, which we want to estimate from observations $X_1, \ldots, X_T$ of the time series. Maximum likelihood estimation (MLE) has been extensively used for parameter estimation, since under weak regularity conditions it is known to be asymptotically efficient. For many models, however, MLE is not always feasible to carry out, due to a likelihood that may be intractable to compute, or maximization of the likelihood is difficult, or because the likelihood function is unbounded on Θ. To overcome such problems, alternative methods have been developed, for instance, the generalized method of moments (GMM) in Hansen [44], the quasi-maximum likelihood estimation (QMLE) in White [102], and composite likelihood methods in Lindsay [66].

In a similar vein, [33] proposed an estimator based on matching the empirical characteristic function (chf) computed from blocks of the observed time series and the true chf. More specifically, given a fixed $p \in \mathbb{N}$, the observed blocks of $X_1, \ldots, X_T$ are

$$\boldsymbol{X}_j = (X_j, \ldots, X_{j+p-1}), \quad j = 1, \ldots, n, \tag{2.1.1}$$

where $n = T - p + 1$. In that paper, a finite set of points in $\mathbb{R}^p$ needs to be chosen as arguments for which the true and the empirical chf are compared. However, the practical choice of this set depends on the problem at hand and the asymptotic results derived in Feuerverger [33] do not offer practical guidance for choosing these points. To overcome this limitation [104] and Knight and Yu [58] considered a integrated weighted squared distance between the empirical and the true chfs.

This method has been used in a variety of applications; an interesting review paper, [105] contains a wealth of examples and references. More recent publications, where the method has been successfully applied to discrete-time models include Knight et al. [59], Meintanis and Taufer [73], Kotchoni [61], Milovanovic et al. [74], Francq and Meintanis [35] and Ndongo et al. [77]. The method also applies to continuous-time processes after discretization and has been used prominently for Lévy-driven models. The book [4] provides additional insight and references in this field.

All these papers assume the ideal situation that the chf has an explicit expression, as a function of $\theta \in \Theta$. We call the corresponding parameter estimator the *oracle estimator* and use it for comparison with the two new estimators we propose in this paper for models whose chf is not available in closed form. Both these estimators are constructed from a functional approximation of the chf constructed from simulated sample paths of $(X_j(\theta))_{j\in\mathbb{Z}}$.

While much attention has been given to the choice of the integrated distance used when computing such estimators, which under some regularity conditions can achieve the Cramér-Rao efficiency bound (see eq. (2.3) of Knight and Yu [58] and Proposition 4.2 of Carrasco et al. [15]), the focus of our paper is on the practical and theoretical aspects that emerge when it is required to approximate the theoretical chf for parameter estimation.

Our first estimator is computed from a simple Monte Carlo approximation to replace the true, but unknown chf. This is similar to the simulated method of moments of McFadden [72] and of the indirect inference method ([96] and Gourieroux et al. [39]). In particular, indirect inference has been successfully applied in a variety of situations: parameter estimation of continuous time models with stochastic volatility (Bianchi and Cleur [6], Jiang [51], Raknerud and Skare [88], Laurini and Hotta [63] and Wahlberg et al. [101]), robust estimation (de Luna and Genton [25] and Fasen-Hartmann and Kimmig [31]), and finite sample bias reduction (Gourieroux et al. [40, 41] and [27]).

More precisely, for many different $\theta \in \Theta$, we simulate an iid sample of blocks denoted by

$$\tilde{\boldsymbol{X}}_j(\theta) = (\tilde{X}_1^{(j)}(\theta), \dots, \tilde{X}_p^{(j)}(\theta)), \quad j = 1, \dots, H, \tag{2.1.2}$$

for $H \in \mathbb{N}$, and define a *simulation based parameter estimator*, which minimizes the integrated weighted mean squared error, which is the integrated distance we use, between the empirical chf computed from the blocks (2.1.2) of the observed time series and its simulated version computed from a large number of simulated paths of the time series.

This is in contrast to the simulation based estimator defined in Section 5.2 of Carrasco et al. [15], which is computed from one long time series path instead of the iid sample of blocks in (2.1.2). Since the Monte Carlo approximation of the chf here is computed from independent blocks, it should have smaller variance than the corresponding one for dependent blocks. By the same method in Carrasco et al. [15], Forneron [34] estimated the structural parameters and the distribution of shocks in dynamic models.

Indeed this gives a chf approximation which yields, by minimizing the integrated distance, strongly consistent and asymptotically normal parameter estimators. We also report their small sample properties for different models.

However, as the Monte Carlo approximation of the chf is computed from iid blocks from a time series, control variates techniques (see [38] and [89]) provide an even more accurate approximation for the chf. Control variates techniques are classical variance reduction methods in simulation. The idea is to use a set of control variates, which are correlated with the chf. The method then approximates the joint covariance matrix of the control variates and the chf, and uses it to construct a new Monte Carlo approximation of the chf. We choose the first two terms in the Taylor expansion of the complex exponential $e^{i\langle t, \boldsymbol{X}_1(\theta)\rangle}$, $\langle t, \boldsymbol{X}_1(\theta)\rangle$ and $\langle t, \boldsymbol{X}_1(\theta)\rangle^2$ for $\theta \in \Theta$ as control variates. This requires knowing the mean and covariance matrix of $\boldsymbol{X}_1(\theta)$ for $\theta \in \Theta$.

In assessing the performance of both the Monte Carlo approximation and the control variates approximation of the chf, two trends emerge. First, both the Monte Carlo and the control variates approximations work better for small values of the argument. Second, the control variates approximation performs much better than the Monte Carlo approximation, in particular, for small values of the argument. As a consequence, we propose a *control variates based parameter estimator* whose integrated mean squared error distance distinguishes between small and large values of the argument.

Under regularity conditions we prove strong consistency of the proposed parameter estimators and asymptotic normality of the simulation based parameter estimator. We find that the simulation based parameter estimator is asymptotically normal with asymptotic covariance matrix equal to the one of the oracle estimator as derived in [58]. From this we conclude that there cannot be any improvement in the limit law for the asymptotic normality of the control variates based estimator. However, we prove that it is computed from a better approximation of the chf. Thus, the control variates estimator improves the finite sample performance compared to the simulation based parameter estimator.

The finite sample performance of the estimators are investigated for two important models. We begin with a stationary Gaussian ARFIMA model, whose chf is explicitly known so that we can use the oracle estimator and compare its performance with the simulated based estimator. Their performance is comparable and also very close to the MLE, so in this model there is no need to use control variates. The second example is a nonlinear model for time series of counts, which has been proposed originally in Zeger [106] and applied, for instance, for modeling disease counts (see also Campbell [14], Chan and Ledolter [16] and Davis et al. [21]).

In the second example, the oracle estimator does not apply, since the chf of the vector

$\boldsymbol{X}_1(\theta)$ cannot be computed in closed form. For this model and different parameter sets, both the simulation based and the control variates based estimators perform satisfactory, and the control variates based estimator improves the performance of the simulation based estimator considerably. When compared with the composite pairwise likelihood estimator in Davis and Yau [20], the control variates based estimator has comparable or even smaller bias.

Our paper is organized as follows. In Section 2.2 we present the oracle estimator, and the estimators computed from a Monte Carlo approximation and from a control variates approximation of the chf in detail. Here we also motivate the choice of the control variates used. The asymptotic properties of the two new estimators are established in Section 2.3. As all estimators are computed from true or approximated chf's we assess their performance in Section 2.4, first for a Gaussian AR(1) process and then for the Poisson-AR process. Practical aspects of calculating the weighted least squares function are discussed in Section 2.5, as well as the estimation results for finite samples. In Section 2.5.1 we compare the oracle estimator, the simulation based parameter estimator and the MLE for a Gaussian ARFIMA model, whereas in Section 2.5.2 we compare the simulation based parameter estimator and the control variates based estimator for the Poisson-AR process. The proofs of Section 2.3 are given in the Appendix.

2.2 Parameter estimation based on the empirical characteristic function

Throughout we use the following notation. For $z \in \mathbb{C}$ we use the L^2-norm: $|z| = \sqrt{z\,\overline{z}}$, where $\overline{z}$ is the complex conjugate of z. For $x \in \mathbb{R}^d$ and $d \in \mathbb{N}$ we denote by $|x|$ the L^2-norm, but recall that in $\mathbb{R}^d$ all norms are equivalent. Furthermore, $\langle \cdot, \cdot \rangle$ denotes the usual Euclidean inner product in $\mathbb{R}^d$. For $z \in \mathbb{C}$ the symbols $\Re(z)$ and $\Im(z)$ denote its real and imaginary part. For a function $f : \mathbb{R}^q \to \mathbb{R}^p$ its gradient is given by $\nabla_\theta f(\theta) = \frac{\partial f(\theta)}{\partial \theta^T} \in \mathbb{R}^{p \times q}$ and $\nabla_\theta^2 f(\theta) = \frac{\partial \mathrm{vec}(\nabla_\theta f(\theta))}{\partial \theta^T} \in \mathbb{R}^{pq \times q}$.

2.2.1 The oracle estimator

Let $(X_j(\theta))_{j\in\mathbb{Z}}$ be a stationary time series process, whose distribution depends on $\theta \in \Theta \subset \mathbb{R}^q$ for some $q \in \mathbb{N}$. Denote by $\theta_0 \in \Theta$ the true parameter, which we want to estimate, and suppose that we observe $X_1, \dots, X_T$. Given a fixed $p \in \mathbb{N}$, define for $\theta \in \Theta$ the p-dimensional blocks

$$\boldsymbol{X}_j(\theta) = (X_j(\theta), \dots, X_{j+p-1}(\theta)), \quad j = 1, \dots, n, \tag{2.2.1}$$

where $n = T - p + 1$. The observed blocks corresponds to

$$\boldsymbol{X}_j = (X_j, \ldots, X_{j+p-1}), \quad j = 1, \ldots, n,$$

which can be used to estimate the *empirical characteristic function (chf)*, defined as

$$\varphi_n(t) = \frac{1}{n} \sum_{j=1}^{n} e^{i\langle t, \boldsymbol{X}_j \rangle}, \quad t \in \mathbb{R}^p. \tag{2.2.2}$$

Under mild conditions such as ergodicity, $\varphi_n(t)$ converges a.s. pointwise to the true chf $\varphi(t) = \mathbb{E}e^{i\langle t, \boldsymbol{X}_1 \rangle}$ for all $t \in \mathbb{R}^p$. We assume that p is chosen in such a way that $\varphi(\cdot)$ uniquely identifies the parameter of interest θ. The idea of estimating θ_0 from a single time series observation by matching the empirical chf of blocks of the observed time series and the true one has been proposed in [104] and Knight and Yu [58], and we use the one in [58], where the *oracle estimator* of θ_0 is defined as

$$\hat{\theta}_n = \text{argmin}_{\theta \in \Theta} Q_n(\theta), \tag{2.2.3}$$

where

$$Q_n(\theta) = \int_{\mathbb{R}^p} |\varphi_n(t) - \varphi(t, \theta)|^2 w(t) \mathrm{d}t, \quad \theta \in \Theta, \tag{2.2.4}$$

with suitable weight function w such that the integral is well-defined, and chf

$$\varphi(t, \theta) = \mathbb{E}e^{i\langle t, \boldsymbol{X}_1(\theta) \rangle}, \quad t \in \mathbb{R}^p.$$

In an ideal situation, $\varphi(\cdot, \theta)$ has an explicit expression, which is known for all $\theta \in \Theta$.

2.2.2 Estimator based on a Monte Carlo approximation of $\varphi(\cdot, \theta)$

Unfortunately, a closed form expression of the chf $\varphi(\cdot, \theta)$ is for many time series processes not available. However, it can be approximated by a Monte Carlo simulation, and an idea borrowed from the simulated method of moments (McFadden [72], see also [96] and Gourieroux et al. [39] for a similar idea in the context of indirect inference) is to replace $\varphi(\cdot, \theta)$ by its functional approximation constructed from simulated sample paths of $(X_j(\theta))_{j \in \mathbb{Z}}$. For many different $\theta \in \Theta$, we simulate, independent of the observed time series, an iid sample of the blocks in (2.2.1) denoted by

$$\tilde{\boldsymbol{X}}_j(\theta) = (\tilde{X}_1^{(j)}(\theta), \ldots, \tilde{X}_p^{(j)}(\theta)), \quad j = 1, \ldots, H, \tag{2.2.5}$$

for $H \in \mathbb{N}$, and define the *Monte Carlo approximation* of $\varphi(\cdot, \theta)$ based on these simulations as

$$\varphi_H(t, \theta) = \frac{1}{H} \sum_{j=1}^{H} e^{i\langle t, \tilde{\boldsymbol{X}}_j(\theta) \rangle}, \quad t \in \mathbb{R}^p. \tag{2.2.6}$$

If we replace $\varphi(\cdot,\theta)$ in (2.2.4) by $\varphi_H(\cdot,\theta)$, we obtain the *simulation based parameter estimator*

$$\hat\theta_{n,H} = \arg\min_{\theta\in\Theta} Q_{n,H}(\theta), \tag{2.2.7}$$

where

$$Q_{n,H}(\theta) = \int_{\mathbb{R}^p} |\varphi_n(t) - \varphi_H(t,\theta)|^2 w(t)\mathrm{d}t, \tag{2.2.8}$$

with suitable weight function w such that the integral is well-defined.

Remark 1. *An alternative approximation to* (2.2.6) *of the chf is based on generating one long time series path and use the empirical chf of the consecutive blocks of p-dimensional random variables constructed as in* (2.2.1)*. While being unbiased, the approximation will generally have larger variance than the approximation proposed in* (2.2.6) *using independent blocks of random variables. Nevertheless, in some cases when it is expensive to generate realizations even of size p, such as the case when a long burn-in is required to achieve stationarity, it may be computationally more efficient to generate one long series. While we do not pursue this approach here, the technical aspects of using one large realization is not much different than the estimate based on independent replicates as in* (2.2.6)*.*

Since $\varphi_H(\cdot,\theta)$ is based on H iid time series blocks, we can reduce its variance further using control variates to produce an even more accurate approximation for the chf. This will result in an improved version of $\hat\theta_{n,H}$.

2.2.3 Estimator based on a control variates approximation of $\varphi(\cdot,\theta)$

The estimator $\hat\theta_{n,H}$ in (2.2.7) requires only that the stationary time series process can be simulated, and is therefore easily applicable to a large class of models. When computing $Q_{n,H}(\theta)$ of (3.2.1), it is very important that the error

$$\xi_H(t,\theta) = |\varphi_H(t,\theta) - \varphi(t,\theta)|, \quad t\in\mathbb{R}^p, \theta\in\Theta, \tag{2.2.9}$$

in approximating the true chf is small, since it propagates to $\hat\theta_{n,H}$. In order to reduce the variance of the empirical chf $\varphi_H(\cdot,\theta)$, we use the method of control variates, as often used variance reduction technique in the context of Monte Carlo integration ([38], [79], Portier and Segers [85]).

We construct a control variates approximation of $\varphi(\cdot,\theta)$ from the iid sample $\tilde{\boldsymbol{X}}_j(\theta)$, $j = 1,\dots,H$, as in (2.2.5). We also require explicit expressions for the moments $\mathbb{E}\langle t, \boldsymbol{X}_1(\theta)\rangle^\nu$ for $\nu = 1,2$ and $\theta\in\Theta$.

Recall that $\tilde{\boldsymbol{X}}_1(\theta) \stackrel{\mathrm{d}}{=} \boldsymbol{X}_1(\theta)$ for all $\theta\in\Theta$, so that both random variables have the same moments. As in Portier and Segers [85], we denote by P_θ the distribution of the block

$\boldsymbol{X}_1(\theta)$ and by $P_{H,\theta}$ its empirical version. For example, if $f_t(x) = e^{i\langle t,x\rangle}$ for $t, x \in \mathbb{R}^p$, we want to provide a good approximation for

$$\varphi(t,\theta) = \mathbb{E}f_t(\boldsymbol{X}_1(\theta)) =: P_\theta(f_t), \quad \theta \in \Theta.$$

To apply the control variates technique, we need control functions, which are correlated with $f_t(\boldsymbol{X}_1(\theta))$ and whose expectations are known. We use the first two terms in the Taylor series of the complex function $f_t(x)$, which suggests the vector of control functions $h_{t,\theta} = (h_{1,t,\theta}, h_{2,t,\theta})^T$, where for $\nu = 1, 2$,

$$h_{\nu,t,\theta}(x) = \langle t, x\rangle^\nu - \mathbb{E}\langle t, \boldsymbol{X}_1(\theta)\rangle^\nu, \quad t \in \mathbb{R}^p,$$

so that $P_\theta(h_{t,\theta}) = 0$, the zero vector in $\mathbb{R}^2$. The Monte Carlo approximation of $\varphi(\cdot,\theta)$ based on the iid sample $\tilde{\boldsymbol{X}}_j(\theta)$, $j = 1, \dots, H$, is then

$$P_{H,\theta}(f_t) = \frac{1}{H}\sum_{j=1}^{H} f_t(\tilde{\boldsymbol{X}}_j(\theta)) = \frac{1}{H}\sum_{j=1}^{H} e^{i\langle t,\tilde{\boldsymbol{X}}_j(\theta)\rangle} = \varphi_H(t,\theta). \tag{2.2.10}$$

Since $\mathbb{E}P_{H,\theta}(f_t) = \mathbb{E}f_t(\boldsymbol{X}_1(\theta))$, the Monte Carlo approximation $\varphi_H(t,\theta)$ is unbiased and has variance

$$\mathrm{Var}[P_{H,\theta}(f_t)] = H^{-1}\sigma_\theta^2(f_t) \quad \text{with} \quad \sigma_\theta^2(f_t) = P_\theta(\{f_t - P_\theta(f_t)\}^2). \tag{2.2.11}$$

Then for every vector $\beta \in \mathbb{C}^2$, we have that $P_{H,\theta}(f_t) - \beta^T P_{H,\theta}(h_{t,\theta})$ is also an unbiased estimator of $\varphi(t,\theta)$. Since $\tilde{\boldsymbol{X}}_j(\theta)$, $j = 1, \dots, H$, is an independent sample,

$$\mathrm{Var}[P_{H,\theta}(f_t) - \beta^T P_{H,\theta}(h_{t,\theta})] = H^{-1}\sigma_\theta^2(f_t - \beta^T h_{t,\theta})$$

and, if we differentiate the map $\beta \mapsto \sigma_\theta^2(f_t - \beta^T h_{t,\theta})$ with respect to β and set it equal to zero, we obtain (cf. Approach 1 in Glynn and Szechtman [38]) the theoretical optimum

$$\beta_{\theta,f_t}^{(\mathrm{opt})}(h_{t,\theta}) = \{P_\theta(h_{t,\theta}h_{t,\theta}^T)\}^{-1}P_\theta(h_{t,\theta}f_t), \tag{2.2.12}$$

provided the inverse exists. In this case, the estimator

$$\varphi_H^{(\mathrm{cvopt})}(t,\theta) = P_{H,\theta}(f_t) - (\beta_{\theta,f_t}^{(\mathrm{opt})}(h_{t,\theta}))^T P_{H,\theta}(h_{t,\theta}) \tag{2.2.13}$$

has minimal asymptotic variance. In order to investigate the existence of the above inverse note that for each fixed $t \in \mathbb{R}^p$ and $\theta \in \Theta$,

$$\det(P_\theta(h_{t,\theta}h_{t,\theta}^T)) = \mathrm{Var}[\langle t, \tilde{\boldsymbol{X}}_1(\theta)\rangle]\mathrm{Var}[\langle t, \tilde{\boldsymbol{X}}_1(\theta)\rangle^2] - \{\mathrm{Cov}[\langle t, \tilde{\boldsymbol{X}}_1(\theta)\rangle, \langle t, \tilde{\boldsymbol{X}}_1(\theta)\rangle^2]\}^2.$$

Since by the Cauchy-Schwarz inequality,

$$\{\mathrm{Cov}[\langle t, \tilde{\boldsymbol{X}}_1(\theta)\rangle, \langle t, \tilde{\boldsymbol{X}}_1(\theta)\rangle^2]\}^2 \le \mathrm{Var}[\langle t, \tilde{\boldsymbol{X}}_1(\theta)\rangle]\mathrm{Var}[\langle t, \tilde{\boldsymbol{X}}_1(\theta)\rangle^2],$$

it follows (see e.g. Klenke [54], Theorem 5.8) that

$$\det(P_\theta(h_{t,\theta}h_{t,\theta}^T) = 0 \iff a\langle t, \tilde{\boldsymbol{X}}_1(\theta)\rangle + b\langle t, \tilde{\boldsymbol{X}}_1(\theta)\rangle^2 + c \overset{\text{a.s.}}{=} 0, \tag{2.2.14}$$

for some $a, b, c \in \mathbb{R}$ with $|a| + |b| + |c| > 0$. As the scalar product is random, universal coefficients to satisfy the right-hand side of (2.2.14) exist only in degenerate cases, which we do not consider.

Since $\beta_{\theta,f_t}^{(\text{opt})}(h_{t,\theta})$ is unknown, it needs to be estimated (e.g. by one of the methods in [38], and we use the one described in eqs. (6) and (7) in Portier and Segers [85]):

$$\begin{aligned}&\hat{\beta}_{H,\theta,f_t}(h_{t,\theta})\\ &= \{P_{H,\theta}(h_{t,\theta}h_{t,\theta}^T) - P_{H,\theta}(h_{t,\theta})P_{H,\theta}(h_{t,\theta}^T)\}^{-1}\{P_{H,\theta}(h_{t,\theta}f_t) - P_{H,\theta}(h_{t,\theta})P_{H,\theta}(f_t)\}.\end{aligned} \tag{2.2.15}$$

For the iid sample $\tilde{\boldsymbol{X}}_j(\theta), j = 1, \ldots, H$, as in (2.2.5) we obtain the *control variates approximation* of $\varphi(\cdot, \theta)$ given by

$$\varphi_H^{(\text{cv})}(t,\theta) = P_{H,\theta}(f_t) - \kappa_H(t,\theta), \quad t \in \mathbb{R}^p, \tag{2.2.16}$$

where

$$\kappa_H(t,\theta) = (\hat{\beta}_{H,\theta,f_t}(h_{t,\theta}))^T P_{H,\theta}(h_{t,\theta}). \tag{2.2.17}$$

Recall from (2.2.10) that $P_{H,\theta}(f_t) = \varphi_H(t,\theta)$, so we could simply replace $\varphi_H(t,\theta)$ in (3.2.1) by $\varphi_H^{(\text{cv})}(t,\theta)$ as given in (2.2.16). However, as we shall see in Section 2.4, the control variates approximation $\varphi_H^{(\text{cv})}(t,\theta)$ provides superior approximations of $\varphi(t,\theta)$ only for values of t, for which $\text{Var}(\langle t, \tilde{\boldsymbol{X}}_1(\theta)\rangle)$ is small. Thus, we replace $\varphi_H(t,\theta)$ in (3.2.1) by a combination of $\varphi_H(t,\theta)$ and $\varphi_H^{(\text{cv})}(t,\theta)$. More precisely, we propose the following *control variates based estimator*:

$$\hat{\theta}_{n,H,k}^{(\text{cv})} = \text{argmin}_{\theta\in\Theta} Q_{n,H,k}^{(\text{cv})}(\theta), \tag{2.2.18}$$

where for appropriate $k > 0$,

$$\begin{aligned}&Q_{n,H,k}^{(\text{cv})}(\theta)\\ &= \int_{\mathbb{R}^p} \left|\varphi_n(t) - \left(\varphi_H^{(\text{cv})}(t,\theta)1_{\{\widehat{\text{Var}}(\langle t,\boldsymbol{X}_1\rangle)<k\}} + \varphi_H(t,\theta)1_{\{\widehat{\text{Var}}(\langle t,\boldsymbol{X}_1\rangle)\geq k\}}\right)\right|^2 \frac{w(t)}{\widehat{\text{Var}}(\langle t,\boldsymbol{X}_1\rangle)}\text{d}t,\end{aligned} \tag{2.2.19}$$

with suitable weight function w such that the integral is well-defined. Note that

$$\widehat{\text{Var}}(\langle t, \boldsymbol{X}_1\rangle) = t^T\hat{\Gamma}_p t$$

where $\hat{\Gamma}_p = (\hat{\gamma}_p(i-j))_{i,j=1}^p$ with

$$\hat{\gamma}_p(h) = \frac{1}{n-h}\sum_{j=1}^{n-h}(X_j - \hat{\mu}_n)(X_{j+h} - \hat{\mu}_n), \quad h = 1, \ldots, p, \tag{2.2.20}$$

and $\hat{\mu}_n = \frac{1}{n}\sum_{j=1}^n X_j$. The choice of the indicator function $1_{\{\widehat{\mathrm{Var}}(\langle t, \boldsymbol{X}_1\rangle)<k\}}$ is justified by the fact that, when estimating the parameter θ_0, we focus on approximations of $\varphi(t,\theta)$ for θ close to θ_0.

2.3 Asymptotic behavior of the parameter estimators

Before performing the parameter estimation we need to make sure that the parameters are identifiable from the model. For the estimators we propose, we require simply that the chf uniquely identifies the parameter of interest. This will always hold true for the examples we consider later on.

The properties of the iid sample of the blocks $(\tilde{\boldsymbol{X}}_j(\theta))_{j\in\mathbb{N}}$ as a function of θ will play a crucial role for the properties of the estimators $\hat{\theta}_{n,H}$ and $\hat{\theta}_{n,H,k}^{(\mathrm{cv})}$ from (2.2.7) and (2.2.18), respectively.

In the sequel, we will make various assumptions on different aspects of the underlying process, smoothness of the model, moments of the process, and properties of the weight function. We group these assumptions into the following categories.

Assumptions A (Parameter space and time series process)**.**

(a.1) *Θ is a compact subset of $\mathbb{R}^q$ and $\theta_0 \in \Theta^{\mathrm{o}}$, the interior of Θ.*

(a.2) *$(X_j)_{j\in\mathbb{Z}}$ is a stationary and ergodic sequence.*

(a.3) *$(X_j)_{j\in\mathbb{Z}}$ is α-mixing with rate function $(\alpha_j)_{j\in\mathbb{N}}$ satisfying $\sum_{j=1}^\infty(\alpha_j)^{1/r} < \infty$ for some $r > 1$.*

Assumptions B (Continuity and differentiability in θ_0)**.**

(b.1) *For each $j \in \mathbb{N}$, the map $\theta \mapsto \tilde{\boldsymbol{X}}_j(\theta)$ is continuous on Θ.*

(b.2) *For each $j \in \mathbb{N}$, the map $\theta \mapsto \tilde{\boldsymbol{X}}_j(\theta)$ is twice continuously differentiable in an open neighborhood around θ_0.*

Assumptions C (Moments)**.**

(c.1) *$\mathbb{E}|X_1|^u < \infty$, where $u = 2r/(r-1)$ with $r > 1$ being such that (a.3) holds.*

(c.2) *$\mathbb{E}\prod_{j=1}^p |X_j|^\alpha < \infty$ for some $\alpha \in (u/2, u]$ where $u = 2r/(r-1)$ with $r > 1$ being such that (a.3) holds.*

(c.3) *$\mathbb{E}\sup_{\theta\in\Theta}|X_1(\theta)|^4 < \infty$.*

$(c.4)$ *For each* $\theta \in \Theta$, $\mathbb{E}|\nabla_\theta X_1(\theta)| < \infty$.

$(c.5)$ $\mathbb{E}\sup_{\theta\in\Theta} |\nabla_\theta X_1(\theta)|^{2(1+\epsilon)} < \infty$ *and* $\mathbb{E}\sup_{\theta\in\Theta} |\nabla^2_\theta X_1(\theta)|^{1+\epsilon} < \infty$ *for some* $\epsilon > 0$.

Assumptions D (Weight function)**.**

$(d.1)$ $\int_{\mathbb{R}^p} w(t)\mathrm{d}t < \infty$.

$(d.2)$ $\int_{\mathbb{R}^p} |t|w(t)\mathrm{d}t < \infty$.

$(d.3)$ $\int_{\mathbb{R}^p} |t|^{2(1+\epsilon)}w(t)\mathrm{d}t < \infty$ *for some* $\epsilon > 0$.

$(d.4)$ $\int_{\mathbb{R}^p} \frac{w(t)}{|t|^2}\mathrm{d}t < \infty$.

Assumption B is indeed satisfied by many linear and non-linear time series processes, in particular, when they have a representation $X_j(\theta) = f(Z_j, Z_{j-1}, \cdots; \theta)$ or $X_j(\theta) = f(Z_j, X_{j-1}(\theta), X_{j-2}(\theta), \cdots; \theta)$ for iid noise variables $(Z_j)_{j\in\mathbb{Z}}$, and $f : \mathbb{R}^\infty \times \Theta \mapsto \mathbb{R}$ is a measurable function. Prominent examples are the MA(∞) and AR(∞) representations of a causal or invertible ARMA(p, q) model (see e.g. eqs. (3.1.15) and (3.1.18) in Brockwell and Davis [13]) or the ARCH(∞) representation of a GARCH(p, q) model (see e.g. Francq and Zakoïan [36], Theorem 2.8). In this case, assumptions $(b.1)$ and $(b.2)$ will hold whenever the map f is continuously differentiable for $\theta \in \Theta$. For example, if f is Lipschitz-continuous for $\theta \in \Theta$, then the continuity assumption $(b.1)$ holds.

The key asymptotic properties, consistency and asymptotic normality of our estimates are stated in the following theorems. The proofs of these results are postponed to the appendix.

We formulate first the strong consistency results of the parameters.

Theorem 2.3.1 (Consistency of $\hat{\theta}_{n,H}$)**.** *Assume that* $(a.1)$, $(a.2)$, $(b.1)$, *and* $(d.1)$ *hold. Let* $H = H(n) \to \infty$ *as* $n \to \infty$. *Then*

$$\hat{\theta}_{n,H} \overset{\text{a.s.}}{\to} \theta_0, \quad n \to \infty.$$

Theorem 2.3.2 (Consistency of $\hat{\theta}^{(\mathrm{cv})}_{n,H,k}$)**.** *Assume that the conditions of Theorem 2.3.1 hold, and additionally* $(c.1)$, $(c.3)$, *and* $(d.4)$. *Let* $H = H(n) \to \infty$ *as* $n \to \infty$. *Then*

$$\hat{\theta}^{(cv)}_{n,H,k} \overset{\text{a.s.}}{\to} \theta_0, \quad n \to \infty.$$

The asymptotic normality of the simulation based parameter estimator reads as follows.

Theorem 2.3.3 (Asymptotic normality of $\hat{\theta}_{n,H}$)**.** *Assume that all Assumptions A and B hold, and that the moment conditions* $(c.2)$, $(c.4)$, *and* $(c.5)$ *hold. Furthermore, assume* *that the weight function satisfies* (*d*.1)*,* (*d*.2) *and* (*d*.3)*. Let* $H = H(n) := \bar{H}(n)n$ *and* $\bar{H}(n) \to \infty$ *as* $n \to \infty$*. Define for* $i = 1, \dots, q$

$$j_{1,i}(t,\theta) = \sin(\langle t, \tilde{\boldsymbol{X}}_1(\theta)\rangle)\langle t, \frac{\partial}{\partial\theta^{(i)}}\tilde{\boldsymbol{X}}_1(\theta)\rangle, \quad l_{1,i}(t,\theta) = \cos(\langle t, \tilde{\boldsymbol{X}}_1(\theta)\rangle)\langle t, \frac{\partial}{\partial\theta^{(i)}}\tilde{\boldsymbol{X}}_1(\theta)\rangle,$$

and

$$b_1^{(i)}(t) = \begin{pmatrix} -\sin(\langle t, \tilde{\boldsymbol{X}}_1(\theta_0)\rangle) \\ \cos(\langle t, \tilde{\boldsymbol{X}}_1(\theta_0)\rangle) \end{pmatrix} \langle t, \frac{\partial}{\partial\theta^{(i)}}\tilde{\boldsymbol{X}}_1(\theta_0)\rangle. \tag{2.3.1}$$

Set

$$b_1(t) = \begin{pmatrix} (b_1^{(1)}(t))^T \\ \vdots \\ (b_1^{(q)}(t))^T \end{pmatrix} \tag{2.3.2}$$

and

$$K_j(\theta) = \int_{\mathbb{R}^p} \mathbb{E}[b_1(t)] \begin{pmatrix} \cos(\langle t, \boldsymbol{X}_1\rangle) - \Re(\varphi(t,\theta_0)) \\ \sin(\langle t, \boldsymbol{X}_1\rangle) - \Im(\varphi(t,\theta_0)) \end{pmatrix} w(t)\mathrm{d}t, \quad j \in \mathbb{N}.$$

Let $Q = (Q_{k,i})_{k,i=1}^q$ *with*

$$Q_{k,i} = \int_{\mathbb{R}^p} \Big(\mathbb{E}j_{1,k}(t,\theta_0)\mathbb{E}j_{1,i}(t,\theta_0) + \mathbb{E}l_{1,k}(t,\theta_0)\mathbb{E}l_{1,i}(t,\theta_0)\Big)w(t)\mathrm{d}t. \tag{2.3.3}$$

If Q *is a non-singular matrix, then*

$$\sqrt{n}(\hat{\theta}_{n,H} - \theta_0) \xrightarrow{d} N(0, Q^{-1}WQ^{-1}), \quad n \to \infty, \tag{2.3.4}$$

where

$$W = \mathrm{Var}[K_1(\theta_0)] + 2\sum_{j=2}^{\infty} \mathrm{Cov}[K_1(\theta_0), K_j(\theta_0)] \tag{2.3.5}$$

Theorem 2.3.3 shows that $\hat{\theta}_{n,H}$ is asymptotically normal and achieves the same asymptotic efficiency as the oracle estimator from (2.2.3) (see Theorem 2.1 in [58]). Therefore, there cannot be any improvement in the limit law for the asymptotic normality of $\hat{\theta}^{(\mathrm{cv})}_{n,H,k}$. However, as we show in Section 2.4 it is based on a better approximation of the chf $\varphi(\cdot,\theta)$ than that used for $\hat{\theta}_{n,H}$. Thus, the control variates estimator $\hat{\theta}^{(\mathrm{cv})}_{n,H,k}$ improves the finite sample performance compared to the simulation based estimator $\hat{\theta}_{n,H}$.

2.4 Assessing the quality of the estimated chf

In this section we compare the performance of both the Monte Carlo approximation $\varphi_H(\cdot,\theta)$ and the control variates approximation $\varphi_H^{(\mathrm{cv})}(\cdot,\theta)$ of the chf as defined in (2.2.6) and (2.2.16), respectively. We start with the following comparison of the two chf approximations.

Remark 2 (Comparison of $\varphi_H^{(\mathrm{cv})}(\cdot,\theta)$ and $\varphi_H(\cdot,\theta)$)**.** *Assume that* (*c*.3) *holds, and let* $\varphi_H^{(cvopt)}$ *and* $\varphi_H^{(cv)}$ *be as defined in* (2.2.13) *and* (2.2.16)*, respectively. We use that* $\hat{\beta}_{H,\theta,f_t}(h_{t,\theta}) \overset{\text{a.s.}}{\to} \beta_{\theta,f_t}^{(opt)}(h_{t,\theta})$ *as* $n \to \infty$ *with limit given in* (2.2.12)*. This follows from the representation of* $\hat{\beta}_{H,\theta,f_t}(h_{t,\theta})$ *as*

$$\hat{\beta}_{H,\theta,f_t}(h_{t,\theta}) = \hat{\beta}_{H,\theta,\Re(f_t)}(h_{t,\theta}) + i\hat{\beta}_{H,\theta,\Im(f_t)}(h_{t,\theta})$$

and the almost sure convergence of both terms.

The quantities needed to compute the estimator in (2.2.15) *are, for each* $\nu,\kappa = 1,2$*:*

$$P_{H,\theta}(f_t) = \frac{1}{H}\sum_{j=1}^{H} e^{i\langle t,\tilde{\boldsymbol{X}}_j(\theta)\rangle}, \tag{2.4.1}$$

$$\begin{aligned}
P_{H,\theta}(h_{\nu,t,\theta}) &= \frac{1}{H}\sum_{j=1}^{H}\Big(\langle t,\tilde{\boldsymbol{X}}_j(\theta)\rangle^{\nu} - \mathbb{E}\langle t,\boldsymbol{X}_1(\theta)\rangle^{\nu}\Big),\\
P_{H,\theta}(f_t h_{\nu,t,\theta}) &= \frac{1}{H}\sum_{j=1}^{H} e^{i\langle t,\tilde{\boldsymbol{X}}_j(\theta)\rangle}\Big(\langle t,\tilde{\boldsymbol{X}}_j(\theta)\rangle^{\nu} - \mathbb{E}\langle t,\boldsymbol{X}_1(\theta)\rangle^{\nu}\Big),\\
P_{H,\theta}(h_{\nu,t,\theta}h_{\kappa,t,\theta}) &= \frac{1}{H}\sum_{j=1}^{H}\Big(\langle t,\tilde{\boldsymbol{X}}_j(\theta)\rangle^{\nu} - \mathbb{E}\langle t,\boldsymbol{X}_1(\theta)\rangle^{\nu}\Big)\times\\
&\quad\Big(\langle t,\tilde{\boldsymbol{X}}_j(\theta)\rangle^{\kappa} - \mathbb{E}\langle t,\boldsymbol{X}_1(\theta)\rangle^{\kappa}\Big).
\end{aligned} \tag{2.4.2}$$

Hence, strong consistency of $\hat{\beta}_{H,\theta,f_t}(h_{t,\theta})$ *follows from the SLLN. This together with* $P_\theta(h_{t,\theta}) = 0$ *implies by Theorem 1 in Glynn and Szechtman [38] that, as* $H \to \infty$*,*

$$H^{1/2}\big(\Re(\varphi_H^{(cv)}(t,\theta) - \varphi(t,\theta))\big) \overset{d}{\to} N\big(0,\sigma_\theta^2(\Re(f_t) - [\beta_{\theta,\Re(f_t)}^{(opt)}(h_{t,\theta})]^T h_{t,\theta})\big),$$

$$H^{1/2}\big(\Im(\varphi_H^{(cv)}(t,\theta) - \varphi(t,\theta))\big) \overset{d}{\to} N\big(0,\sigma_\theta^2(\Im(f_t) - [\beta_{\theta,\Im(f_t)}^{(opt)}(h_{t,\theta})]^T h_{t,\theta})\big),$$

with

$$\sigma_\theta^2(\Re(f_t) - [\beta_{\theta,\Re(f_t)}^{(opt)}(h_{t,\theta})]^T h_{t,\theta}) \le \sigma_\theta^2(\Re(f_t))$$

and

$$\sigma_\theta^2(\Im(f_t) - [\beta_{\theta,\Im(f_t)}^{(opt)}(h_{t,\theta})]^T h_{t,\theta}) \le \sigma_\theta^2(\Im(f_t)),$$

with $\sigma_\theta^2(\cdot)$ *as defined in* (2.2.11)*. Therefore,* $\varphi_H^{(cv)}(\cdot,\theta)$ *provides an approximation of the integral* $Q_n(\theta)$ *in* (2.2.4) *with smaller variance than* $\varphi_H(\cdot,\theta)$*. As a consequence, this favors the control variates estimator* $\hat{\theta}_{n,H,k}^{(cv)}$ *over the simulation based estimator* $\hat{\theta}_{n,H}$ *for large sample sizes* $n \in \mathbb{N}$*.*

For all forthcoming examples we choose $p = 3$ and $H = 3\,000$. We begin with a stationary Gaussian AR(1) process, where we know the chf $\varphi(\cdot)$ explicitly, and then proceed to the Poisson-AR process, where we approximate the true unknown chf by a precise simulated version.

2.4.1 The AR(1) process

We start with a stationary Gaussian AR(1) process to show how the method of control variates improves the Monte Carlo approximation of its chf. Let $(X_j(\theta))_{j\in\mathbb{Z}}$ be the AR(1) process

$$X_j(\theta) = \phi X_{j-1}(\theta) + Z_j(\theta),\ j \in \mathbb{Z}, \quad (Z_j(\theta))_{j\in\mathbb{Z}} \overset{\text{iid}}{\sim} N(0,\sigma^2), \tag{2.4.3}$$

with parameter space Θ being a compact subset of $\{\theta = (\phi,\sigma) : |\phi| < 1, \sigma > 0\}$. Then the true chf of $\boldsymbol{X}_1(\theta) = (X_1(\theta), X_2(\theta), X_3(\theta))$ is given by

$$\varphi(t,\theta) = e^{-\frac{1}{2}t^T\Gamma_3(\theta)t}, \quad t \in \mathbb{R}^3,$$

where the covariance matrix $\Gamma_3(\theta)$ is explicitly known and identifies the parameter θ uniquely; see e.g. [13], Example 3.1.2. For a fixed $\theta \in \Theta$ and many $t \in \mathbb{R}^3$ we compute the absolute errors

$$\xi_H(t,\theta) = |\varphi_H(t,\theta) - \varphi(t,\theta)| \quad \text{and} \quad \xi_H^{(\text{cv})}(t,\theta) = |\varphi_H^{(\text{cv})}(t,\theta) - \varphi(t,\theta)| \tag{2.4.4}$$

where $\varphi_H(\cdot,\theta)$ is the Monte Carlo approximation of the chf of $\boldsymbol{X}_1(\theta) = (X_1(\theta), X_2(\theta), X_3(\theta))$ and $\varphi_H^{(\text{cv})}(\cdot,\theta)$ its control variates approximation. To understand how well we can approximate $\varphi(\cdot,\theta)$, we plot in Figure 2.1, $\xi_H(t,\theta)$ and $\xi_H^{(\text{cv})}(t,\theta)$ against $\sqrt{\operatorname{Var}[\langle t, \boldsymbol{X}_1(\theta)\rangle]}$ for different parameters θ. These quantities are computed from an iid sample $\boldsymbol{X}_j(\theta), j = 1,\dots,H$ as in (2.2.5). To simulate iid observations from the model (2.4.3), we use the fact that the one-dimensional stationary distribution is $X_1(\theta) \sim N(0, \sigma^2/(1-\phi^2))$, and then use the recursion in (2.4.3) to simulate $X_2(\theta)$ and $X_3(\theta)$. We chose 500 randomly generated values of t from the 3-dimensional Laplace distribution with chf given in (2.5.2).

It is clear from Figure 2.1 that both the Monte Carlo and the control variates approximations work better when $\sqrt{\operatorname{Var}[\langle t, \boldsymbol{X}_1(\theta)\rangle]}$ is small, and also that the control variates approximations are best for small values of $\sqrt{\operatorname{Var}[\langle t, \boldsymbol{X}_1(\theta)\rangle]}$. The superiority of the control variates approximation for all t and all parameter settings is clearly visible, and already expected from Remark 2.

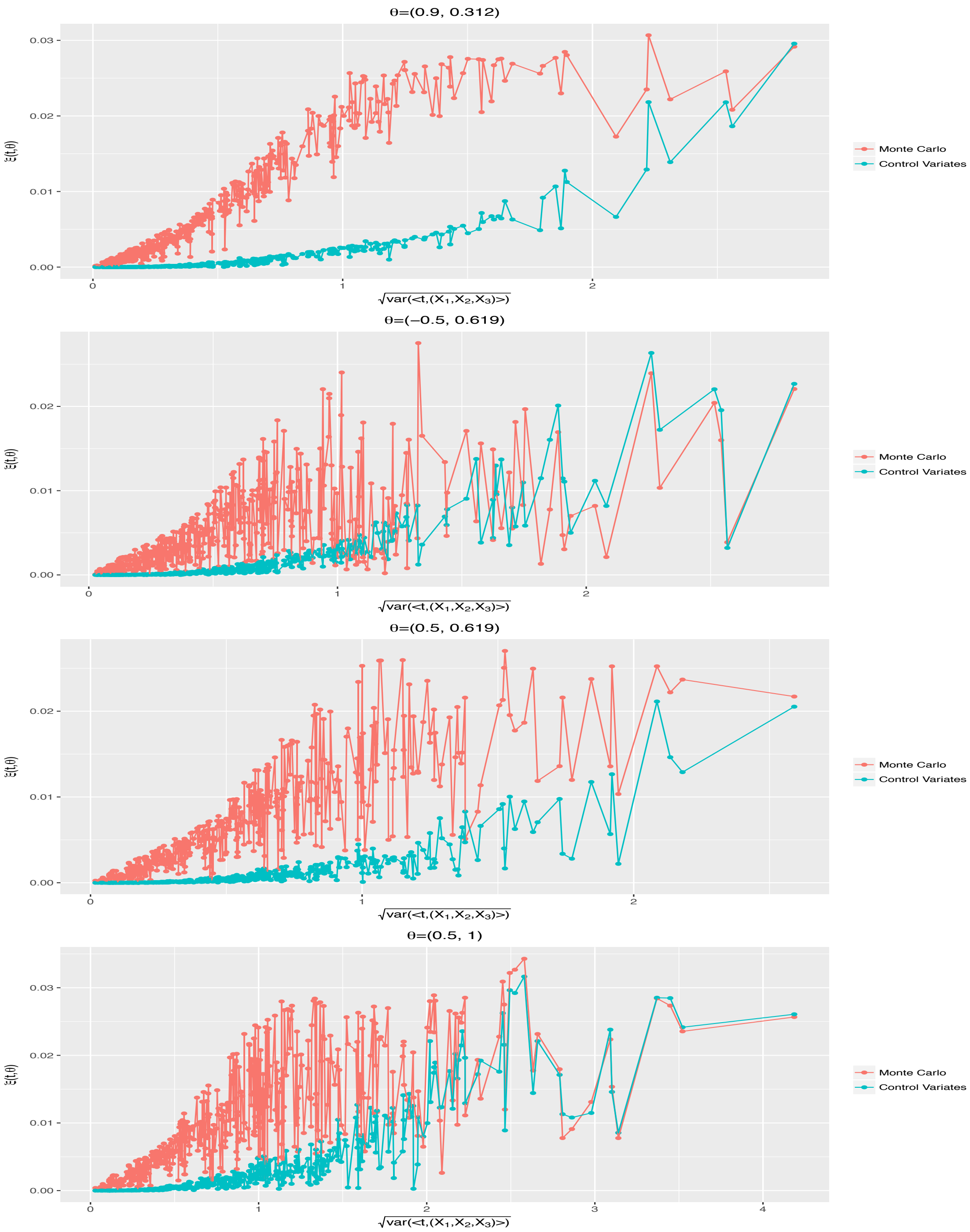


Figure 2.1: Gaussian AR(1) model: absolute error $\xi_H(t,\theta)$ and $\xi_H^{(\mathrm{cv})}(t,\theta)$ for $p = 3$ and $H = 3\,000$ as in eq. (2.4.4). We use 500 randomly generated values of $t \in \mathbb{R}^3$ from the Laplace distribution (with chf as in (2.5.2) below), which are plotted against $\sqrt{\mathrm{Var}[\langle t, \boldsymbol{X}_1(\theta)\rangle]}$.

2.4.2 The Poisson-AR model

We consider a nonlinear time series process for time series of counts, which has been proposed originally in Zeger [106]. A prototypical Poisson-AR(1) model suggested in Davis and Rodriguez-Yam [19] assumes that the observations $(X_j(\theta))_{j\in\mathbb{Z}}$ are independent and Poisson-distributed with means $e^{\beta+\alpha_j(\theta)}$ where the process $(\alpha_j(\theta))_{j\in\mathbb{Z}}$ is a latent stationary Gaussian AR(1) process, given by the equations

$$\alpha_j(\theta) = \phi\alpha_{j-1}(\theta) + \eta_j(\theta),\ j \in \mathbb{Z}, \quad (\eta_j(\theta))_{j\in\mathbb{Z}} \overset{\text{iid}}{\sim} N(0, \sigma^2),$$

with parameter space Θ being a compact subset of $\{\theta = (\beta, \phi, \sigma) : |\phi| < 1, \beta \in \mathbb{R}, \sigma > 0\}$. The parameter θ is uniquely identifiable from the second order structure, which has been computed in Section 2.1 of Davis et al. [22].

For this model, the true chf of $\boldsymbol{X}_1(\theta) = (X_1(\theta), X_2(\theta), X_3(\theta))$ cannot be computed in closed form. To mimic the assessment of the errors in eq. (2.4.4), we simulate 1 000 000 iid observations from $\boldsymbol{X}_1(\theta)$ by first simulating a Gaussian AR(1) process $(\alpha_1(\theta), \alpha_2(\theta), \alpha_3(\theta))$ (as described in Section 2.4.1) and then simulating independent Poisson random variables with means $e^{\beta+\alpha_1(\theta)}$, $e^{\beta+\alpha_2(\theta)}$ and $e^{\beta+\alpha_3(\theta)}$, respectively. From this we compute the empirical characteristic function and take it as $\varphi(\cdot, \theta)$ in the absolute error terms (2.4.4).

Now, as in Section 2.4.1, we compare the performance of both the Monte Carlo approximation and the control variates approximation of the chf. Figure 2.2 presents the results. The plots in Figure 2.2 are also in favor of the control variates approximation, when compared to the Monte Carlo approximation.

2.5 Practical aspects and estimation results for finite samples

Our objective is to obtain a simple expression of the integrated mean squared error $Q_{n,H}(\theta)$ in (3.2.1), which is needed to compute the estimator in (2.2.7). For a weight function w in (3.2.1), we write

$$\tilde{w}(x) = \int_{\mathbb{R}^p} e^{i\langle t,x\rangle} w(t)\mathrm{d}t, \quad x \in \mathbb{R}^p, \tag{2.5.1}$$

for its Fourier transform. Our preference is on weight functions such that (2.5.1) is known explicitly.

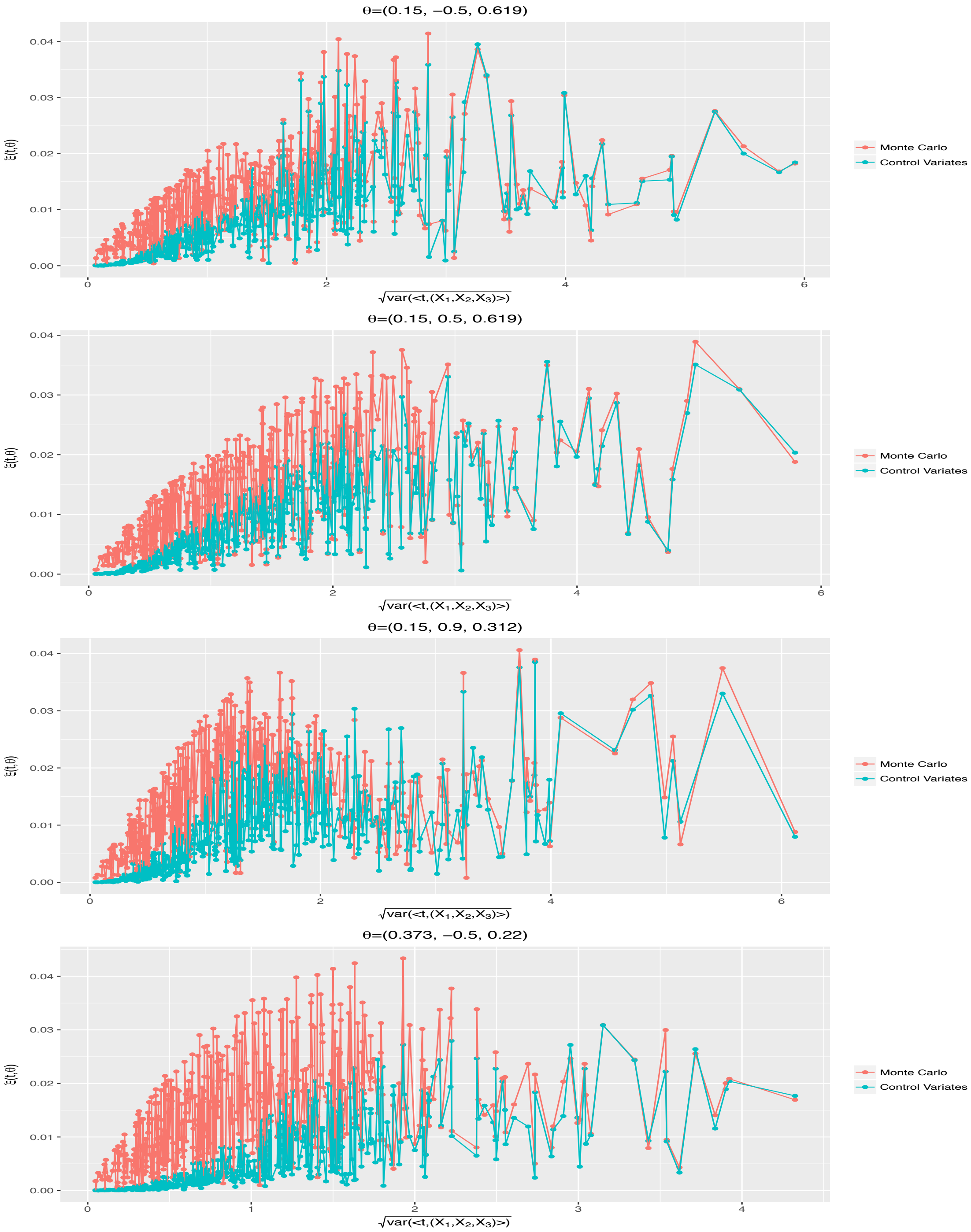


Figure 2.2: Poisson-AR model: Absolute errors $\xi_H(t,\theta)$ and $\xi_H^{(\mathrm{cv})}(t,\theta)$ for $p = 3$ and $H = 3\,000$ as in eq. (2.4.4). We use 500 randomly generated values of $t \in \mathbb{R}^3$ from the Laplace distribution (with chf as in (2.5.2) below), which are plottet against $\sqrt{\mathrm{Var}[\langle t, \boldsymbol{X}_1(\theta)\rangle]}$.

Example 2.5.1. [Weight functions and their characteristic functions]
(i) Laplace: w is a multivariate Laplace density with chf

$$\tilde{w}(t) = \frac{1}{(1 + (2\pi^2)^{-1}\, t^T t)}, \quad t \in \mathbb{R}^p. \tag{2.5.2}$$

(ii) Cauchy: w is a multivariate Cauchy density with chf

$$\tilde{w}(t) = e^{-\sqrt{t^T t}}, \quad t \in \mathbb{R}^p.$$

(iii) Gaussian: w is a standard multivariate Gaussian density with chf

$$\tilde{w}(t) = e^{-\frac{1}{2} t^T t}, \quad t \in \mathbb{R}^p. \tag{2.5.3}$$

□

Lemma 2.5.2. *Let $Q_{n,H}(\theta)$ be as in* (3.2.1) *and w a weight function with Fourier transform $\tilde{w}$. Then*

$$\begin{aligned} Q_{n,H}(\theta) \quad &= \quad \frac{1}{n^2}\sum_{k=1}^{n}\sum_{j=1}^{n} \tilde{w}(\boldsymbol{X}_j - \boldsymbol{X}_k) + \frac{1}{H^2}\sum_{j=1}^{H}\sum_{k=1}^{H} \tilde{w}(\tilde{\boldsymbol{X}}_j(\theta) - \tilde{\boldsymbol{X}}_k(\theta)) \\ &\quad -\frac{1}{Hn}\sum_{k=1}^{H}\sum_{j=1}^{n}\Big(\tilde{w}(\boldsymbol{X}_j - \tilde{\boldsymbol{X}}_k(\theta)) + \tilde{w}(\tilde{\boldsymbol{X}}_k(\theta) - \mathbf{X}_j)\Big). \end{aligned} \tag{2.5.4}$$

Proof. Since $|z|^2 = z\bar{z}$ for $z \in \mathbb{C}$, for every $\theta \in \Theta$,

$$\begin{aligned} Q_{n,H}(\theta) &= \int_{\mathbb{R}^p} \left| \frac{1}{n}\sum_{j=1}^{n} e^{i\langle t, \boldsymbol{X}_j\rangle} - \frac{1}{H}\sum_{j=1}^{H} e^{i\langle t, \tilde{\boldsymbol{X}}_j(\theta)\rangle} \right|^2 w(t)\mathrm{d}t \\ &= \int_{\mathbb{R}^p} \left(\frac{1}{n}\sum_{j=1}^{n} e^{i\langle t, \boldsymbol{X}_j\rangle} - \frac{1}{H}\sum_{j=1}^{H} e^{i\langle t, \tilde{\boldsymbol{X}}_j(\theta)\rangle}\right)\left(\frac{1}{n}\sum_{k=1}^{n} e^{-i\langle t, \boldsymbol{X}_k\rangle} - \frac{1}{H}\sum_{k=1}^{H} e^{-i\langle t, \tilde{\boldsymbol{X}}_k(\theta)\rangle}\right) w(t)\mathrm{d}t \\ &= \frac{1}{n^2}\sum_{k=1}^{n}\sum_{j=1}^{n}\int_{\mathbb{R}^p} e^{i\langle t, \boldsymbol{X}_j - \boldsymbol{X}_k\rangle} w(t)\mathrm{d}t - \frac{1}{Hn}\sum_{k=1}^{H}\sum_{j=1}^{n}\int_{\mathbb{R}^p} e^{i\langle t, \boldsymbol{X}_j - \tilde{\boldsymbol{X}}_k(\theta)\rangle} w(t)\mathrm{d}t \\ &- \frac{1}{Hn}\sum_{j=1}^{H}\sum_{k=1}^{n}\int_{\mathbb{R}^p} e^{i\langle t, \tilde{\boldsymbol{X}}_j(\theta) - \boldsymbol{X}_k\rangle} w(t)\mathrm{d}t + \frac{1}{H^2}\sum_{j=1}^{H}\sum_{k=1}^{H}\int_{\mathbb{R}^p} e^{i\langle t, \tilde{\boldsymbol{X}}_j(\theta) - \tilde{\boldsymbol{X}}_k(\theta)\rangle} w(t)\mathrm{d}t \\ &= \frac{1}{n^2}\sum_{k=1}^{n}\sum_{j=1}^{n} \tilde{w}(\boldsymbol{X}_j - \boldsymbol{X}_k) - \frac{1}{Hn}\sum_{k=1}^{H}\sum_{j=1}^{n}\Big(\tilde{w}(\boldsymbol{X}_j - \tilde{\boldsymbol{X}}_k(\theta)) + \tilde{w}(\tilde{\boldsymbol{X}}_k(\theta) - \mathbf{X}_j)\Big) \\ &+ \frac{1}{H^2}\sum_{j=1}^{H}\sum_{k=1}^{H} \tilde{w}(\tilde{\boldsymbol{X}}_j(\theta) - \tilde{\boldsymbol{X}}_k(\theta)). \end{aligned}$$

□

Formula (2.5.4) is very useful, since it avoids the computation of a p-dimensional integral. Additionally, since the first double sum on the right-hand side of (2.5.4) does not depend on the argument θ, for the optimization it can be ignored.

Remark 3. *When evaluating the integrated weighted mean squared errors* (3.2.1)*,* (2.2.19)*, or* (2.5.4) *in practice, they need to be deterministic functions of* θ*. This is enforced by taking a fixed seed for every* $j = 1, \ldots, H$*, when simulating* $\tilde{\boldsymbol{X}}_j(\theta)$ *for different values of* $\theta \in \Theta$*.*

In the following two examples we study the finite sample behavior of the estimators $\hat{\theta}_{n,H}$ and $\hat{\theta}^{(\mathrm{cv})}_{n,H,k}$. We begin with a stationary Gaussian ARFIMA model, whose chf is explicitly known so that we can use the oracle estimator from Section 2.2.1. Afterwards we come back to the Poisson-AR process. We choose $p = 3$, since the 3-dimensional chf contains sufficient information to identify the parameter of interest. We also choose $H = 3\,000$.

2.5.1 The ARFIMA model

Let $(X_j(\theta))_{j\in\mathbb{Z}}$ be the stationary Gaussian ARFIMA$(0, d, 0)$ model

$$(1 - B)^d X_j(\theta) = Z_j(\theta),\ j \in \mathbb{Z}, \quad (Z_j(\theta))_{j\in\mathbb{Z}} \overset{\text{iid}}{\sim} N(0, \sigma^2),$$

where B is the backshift operator, with parameter space Θ being a compact subset of $\{\theta = (d, \sigma) : d \in (-0.5, 0.5), \sigma > 0\}$. Then the true chf of $\boldsymbol{X}_1(\theta) = (X_1(\theta), X_2(\theta), X_3(\theta))$ is given by

$$\varphi(t, \theta) = e^{-\frac{1}{2}t^T\Gamma_3(\theta)t}, \quad t \in \mathbb{R}^3, \theta \in \Theta,$$

where the covariance matrix $\Gamma_3(\theta)$ is explicitly known and identifies the parameter θ uniquely; see e.g. Pipiras and Taqqu [83], Corollary 2.4.4.

For the long-memory case, for each value of $d \in \{0.05, \ldots, 0.45\}$ we compare the new estimators with the MLE method as implemented in the R package `arfima`. Thus, for many $\theta \in \Theta$, we generate iid Gaussian random vectors with mean zero and covariance $\Gamma_3(\theta)$ and use them to construct the simulation based estimator $\hat{\theta}_{n,H}$.

Since the chf $\varphi(\cdot, \theta)$ is known in closed form, we are able to compute the oracle estimator $\hat{\theta}_n$ from (2.2.4). For practical purpose we choose the weight function $w(t) = (2\pi)^{-3/2} e^{-\frac{1}{2}t^T t}, t \in \mathbb{R}^3$.

Then the integral in (2.2.4), which needs to be minimized with respect to the parameter

θ, can be evaluated similarly as in (2.5.4), giving for the chf being known,

$$\begin{aligned} Q_n(\theta) &= \int_{\mathbb{R}^3} \left| \frac{1}{n} \sum_{j=1}^n e^{i\langle t, \boldsymbol{X}_j \rangle} - e^{-\frac{1}{2} t^T \Gamma_3(\theta) t} \right|^2 w(t) \mathrm{d}t \\ &= \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n \exp\Big\{ - \frac{1}{2} (\boldsymbol{X}_j - \boldsymbol{X}_k)^T (\boldsymbol{X}_j - \boldsymbol{X}_k) \Big\} + \big(\det\left((2\Gamma_3(\theta) + I)^{-1} \right) \big)^{\frac{1}{2}} \\ &\quad -2 \big(\det\left((\Gamma_3(\theta) + I)^{-1} \right) \big)^{\frac{1}{2}} \frac{1}{n} \sum_{j=1}^n \exp\Big\{ - \frac{1}{2} \boldsymbol{X}_j^T (\Gamma_3(\theta) + I)^{-1} \boldsymbol{X}_j \Big\}. \quad (2.5.5) \end{aligned}$$

We compare in Table 2.2 the performance of the simulation based estimator $\hat{\theta}_{n,H}$, the oracle estimator $\hat{\theta}_n$ in (2.2.3) based on the minimization of (2.5.5), and the MLE. We notice that $\hat{\theta}_{n,H}$ is comparable to the oracle estimator, so in this model there is no need to use control variates. In particular, the RMSEs are almost the same for all $d \geq 0.20$. The MLE has a smaller RMSE, but both $\hat{\theta}_n$ and $\hat{\theta}_{n,H}$ have a smaller bias than the MLE.

	β	ϕ	σ	β	ϕ	σ	β	ϕ	σ
				$D = 10$					
TRUE	-0.613	-0.500	1.236	-0.613	0.500	1.236	-0.613	0.900	0.622
Bias($\hat{\theta}_{n,H}$)	-0.015	0.025	0.002	-0.012	0.014	-0.032	-0.016	-0.010	0.002
RMSE($\hat{\theta}_{n,H}$)	0.096	0.101	0.119	0.148	0.107	0.120	0.298	0.054	0.128
Bias($\hat{\theta}^{(\mathrm{cv})}_{n,H,k}$)	0.023	0.031	-0.007	0.006	0.002	-0.018	0.061	-0.007	-0.036
RMSE($\hat{\theta}^{(\mathrm{cv})}_{n,H,k}$)	0.102	0.129	0.122	0.138	0.098	0.098	0.285	0.049	0.132
				$D = 1$					
TRUE	0.150	-0.500	0.619	0.150	0.500	0.619	0.150	0.900	0.312
Bias($\hat{\theta}_{n,H}$)	-0.004	0.024	-0.016	-0.006	0.005	-0.023	-0.016	-0.033	0.028
RMSE($\hat{\theta}_{n,H}$)	0.057	0.144	0.088	0.074	0.141	0.081	0.147	0.084	0.095
Bias($\hat{\theta}^{(\mathrm{cv})}_{n,H,k}$)	0.003	-0.011	-0.017	0.001	0.023	-0.019	0.003	-0.009	-0.012
RMSE($\hat{\theta}^{(\mathrm{cv})}_{n,H,k}$)	0.055	0.124	0.085	0.071	0.102	0.069	0.145	0.062	0.087
				$D = 0.1$					
TRUE	0.373	-0.500	0.220	0.373	0.500	0.220	0.373	0.900	0.111
Bias($\hat{\theta}_{n,H}$)	-0.011	0.032	-0.045	-0.015	-0.322	-0.036	-0.019	-0.517	0.044
RMSE($\hat{\theta}_{n,H}$)	0.043	0.408	0.098	0.047	0.657	0.102	0.066	0.801	0.099
Bias($\hat{\theta}^{(\mathrm{cv})}_{n,H,k}$)	-0.002	0.056	-0.044	-0.003	-0.120	-0.038	-0.004	-0.310	0.031
RMSE($\hat{\theta}^{(\mathrm{cv})}_{n,H,k}$)	0.042	0.482	0.112	0.045	0.504	0.108	0.062	0.555	0.090

Table 2.1: Comparison of the simulation based estimator $\hat{\theta}_{n,H}$ of (2.2.7) and the control variates based estimator $\hat{\theta}^{(\mathrm{cv})}_{n,H,k}$ of (2.2.18) with $k = 1$, both with $H = 3\,000$. The models are classified by the index D of dispersion of $e^{\beta+\alpha_1}$. For both estimators the empirical chf has been computed with $n = 500$, $p = 3$, and w is the Laplace density as in (2.5.2). Moreover, 500 replications have been used to compute bias, standard deviation (Std) and root mean squared error (RMSE).

2.5.2 The Poisson-AR process

The Poisson-AR model has been defined in Section 2.4.2. We conduct a simulation experiment in the same setting as in Table 5 in Davis and Rodriguez-Yam [19] and Table 3 in Davis and Yau [20]. The results are shown in Table 2.1 for $n = 500$ and nine different parameter settings, where we also classify the models by the corresponding index of dispersion D of the random variable $e^{\beta+\alpha_1}$, which assumes values in $\{0.1, 1, 10\}$ as shown in Davis and Rodriguez-Yam [19].

	$d = 0.05$			$d = 0.10$			$d = 0.15$		
	Bias	Std	RMSE	Bias	Std	RMSE	Bias	Std	RMSE
$\hat{\theta}_{n,H}$	0.000	0.056	0.056	0.002	0.054	0.054	0.004	0.049	0.049
$\hat{\theta}_n$	-0.005	0.050	0.050	-0.004	0.047	0.047	-0.004	0.044	0.045
MLE	-0.015	0.040	0.043	-0.015	0.040	0.043	-0.016	0.040	0.043
	$d = 0.20$			$d = 0.25$			$d = 0.30$		
	Bias	Std	RMSE	Bias	Std	RMSE	Bias	Std	RMSE
$\hat{\theta}_{n,H}$	0.003	0.047	0.047	0.000	0.046	0.046	-0.003	0.048	0.048
$\hat{\theta}_n$	-0.004	0.045	0.045	-0.006	0.044	0.044	-0.007	0.046	0.047
MLE	-0.016	0.040	0.043	-0.017	0.039	0.043	-0.017	0.039	0.043
	$d = 0.35$			$d = 0.40$			$d = 0.45$		
	Bias	Std	RMSE	Bias	Std	RMSE	Bias	Std	RMSE
$\hat{\theta}_{n,H}$	-0.006	0.050	0.051	-0.013	0.051	0.052	-0.022	0.047	0.052
$\hat{\theta}_n$	-0.009	0.049	0.050	-0.013	0.051	0.052	-0.021	0.048	0.052
MLE	-0.019	0.039	0.043	-0.021	0.037	0.043	-0.027	0.034	0.043

Table 2.2: Comparison of the simulation based estimator $\hat{\theta}_{n,H}$ for $H = 3\,000$, the oracle estimator $\hat{\theta}_n$ and the MLE. For both estimators we have set $n = 400$, $p = 3$, and w is the Gaussian density as in (2.5.3). Moreover, 500 replications have been used to compute bias, standard deviation (Std) and root mean squared error (RMSE).

We compare both the simulation based estimator $\hat{\theta}_{n,H}$ and control variates based estimator $\hat{\theta}^{(\mathrm{cv})}_{n,H,k}$. We fix $H = 3\,000$, $p = 3$ and the 3-dimensional Laplace density as in (2.5.2) for w. To simulate iid observations of $(X_1(\theta), X_2(\theta), X_3(\theta))$ we proceed as explained in Section 2.4.2. The simulation based estimator $\hat{\theta}_{n,H}$ in (2.2.7) is computed via (2.5.4). Unfortunately, such a formula cannot be obtained for the control variates based estimator $\hat{\theta}^{(\mathrm{cv})}_{n,H,k}$, since the introduction of the correction κ_H in (2.2.17) introduces addional polynomial terms into $Q^{(\mathrm{cv})}_{n,H,k}$ in (2.2.19). Thus, we resort to numerical integration to evaluate $\hat{\theta}^{(\mathrm{cv})}_{n,H,k}$.

Our findings are as follows. For $D \in \{1, 0.1\}$, the control variates based estimator

$\hat{\theta}^{(\mathrm{cv})}_{n,H,k}$ for $k = 1$ presents smaller bias and RMSE than the simulation based estimator $\hat{\theta}_{n,H}$ in most cases, in all others it is comparable. Additionally, a significant improvement in the bias for estimating ϕ is noticeable for $\theta = (0.373, 0.500, 0.220)$ and $\theta = (0.373, 0.900, 0.111)$.

We compare now the control variates based estimator $\hat{\theta}^{(\mathrm{cv})}_{n,H,k}$ in Table 2.1 with the results for the consecutive pairwise likelihood (CPL) from Table 3 in Davis and Yau [20], which is refereed to as CPL1 in that paper. The bias of $\hat{\theta}^{(\mathrm{cv})}_{n,H,k}$ is smaller than that of CPL1 for the estimated β and σ for almost all cases, in all others it is comparable. For ϕ the bias of $\hat{\theta}^{(\mathrm{cv})}_{n,H,k}$ and CPL1 are comparable, except that $\hat{\theta}^{(\mathrm{cv})}_{n,H,k}$ has poor performance for estimating ϕ for the true parameter $(\beta, \phi, \sigma) = (0.373, 0.9, 0.111)$. This is due to the fact that the simulated sample paths contain a large number of zeros, giving very little information for the parameter estimation.

We fix $H = 3\,000$ and simulate observations of the (X_1, X_2, X_3) by using H simulated paths of length $p = 3$. We also include the CV estimator in (2.2.18), where we choose the weight function as $\tilde{w}(t) = (\widehat{\mathrm{Var}}(\langle t, \boldsymbol{X}_1\rangle))^{-1} w(t)$ and $w(t)$ as the multivariate Laplace distribution as in (2.5.2). This choice was motivated by the findings in Section 2.4 for the PDM, where the estimates of the chf were most well behaved for values of t for which $\widehat{\mathrm{Var}}(\langle t, \boldsymbol{X}_1\rangle)$ was small.

2.6 Appendix to Section 2.3

In the following we always set $H = H(n)$ and $\bar{H} = \bar{H}(n) = H(n)/n$, but omit the argument n for notational simplicity.

Throughout the letter c stands for any positive constant independent of the respective argument. Its value may change from line to line, but is not of particular interest.

For a matrix with only real eigenvalues $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue.

We often use the uniform SLLN, which guarantees for a continuous stochastic process $(Z(t))_{t\in\mathbb{R}^p}$ satisfying $\mathbb{E}\sup_{t\in K}|Z(t)| < \infty$ that $\sup_{t\in K}|Z(t) - \mathbb{E}Z(t)| \overset{\text{a.s.}}{\to} 0$ as $n \to \infty$ for every compact set $K \subset \mathbb{R}^p$. More precisely, we use the SLLN on the separable Banach space $C(K)$, the space of continuous functions on the compact set $K \subset \mathbb{R}^p$, endowed with the sup norm (see e.g. Theorem 16(a) in Ferguson [32] or Theorem 9.4 in Parthasarathey [80]).

2.6.1 Proof of Theorem 2.3.1

Let

$$Q(\theta) = \int_{\mathbb{R}^p} |\varphi(t, \theta_0) - \varphi(t, \theta)|^2 w(t)\mathrm{d}t$$

be the candidate limiting function of $Q_{n,H}(\theta)$. For $\delta > 0$ define the set

$$K_\delta = \{t \in \mathbb{R}^p : |t| \le \delta\}. \tag{2.6.1}$$

Since $|e^{i\langle t, \tilde{\boldsymbol{X}}_1(\theta)\rangle}| = 1$ for all θ and t, and the random elements $(\tilde{\boldsymbol{X}}_j(\theta), \theta \in \Theta)_{j=1}^{\infty}$ are iid, the uniform SLLN holds giving

$$\sup_{(t,\theta)\in\Theta\times K_\delta} \left| \frac{1}{H}\sum_{j=1}^{H} e^{i\langle t, \tilde{\boldsymbol{X}}_j(\theta)\rangle} - \varphi(t,\theta)\right| \overset{\text{a.s.}}{\to} 0, \quad n \to \infty. \tag{2.6.2}$$

In particular, for $\theta = \theta_0$ we also have

$$\sup_{t\in K_\delta} \left| \frac{1}{n}\sum_{j=1}^{n} e^{i\langle t, \boldsymbol{X}_j\rangle} - \varphi(t,\theta_0)\right| \overset{\text{a.s.}}{\to} 0, \quad n \to \infty. \tag{2.6.3}$$

Applying the inequality $||a|^2 - |b|^2| \le 2|a-b|$ for $a, b \in \mathbb{C}, |a|, |b| \le 1$ gives

$$\begin{aligned}
&|Q_{n,H}(\theta) - Q(\theta)| \\
&\le \int_{\mathbb{R}^p} \left| \left| \frac{1}{n}\sum_{j=1}^{n} e^{i\langle t, \boldsymbol{X}_j\rangle} - \frac{1}{H}\sum_{j=1}^{H} e^{i\langle t, \tilde{\boldsymbol{X}}_j(\theta)\rangle}\right|^2 - |\varphi(t,\theta_0) - \varphi(t,\theta)|^2 \right| w(t)\mathrm{d}t \\
&\le 2\int_{\mathbb{R}^p} \left| \frac{1}{n}\sum_{j=1}^{n} e^{i\langle t, \boldsymbol{X}_j\rangle} - \varphi(t,\theta_0) + \varphi(t,\theta) - \frac{1}{H}\sum_{j=1}^{H} e^{i\langle t, \tilde{\boldsymbol{X}}_j(\theta)\rangle}\right| w(t)\mathrm{d}t \\
&\le 2\int_{\mathbb{R}^p} \left\{ \left| \frac{1}{n}\sum_{j=1}^{n} e^{i\langle t, \boldsymbol{X}_j\rangle} - \varphi(t,\theta_0)\right| + \sup_{\theta\in\Theta}\left| \varphi(t,\theta) - \frac{1}{H}\sum_{j=1}^{H} e^{i\langle t, \tilde{\boldsymbol{X}}_j(\theta)\rangle}\right| \right\} w(t)\mathrm{d}t \\
&\le 2\sup_{(t,\theta)\in\Theta\times K_\delta} \left\{ \left| \frac{1}{n}\sum_{j=1}^{n} e^{i\langle t, \boldsymbol{X}_j\rangle} - \varphi(t,\theta_0)\right| + \left| \varphi(t,\theta) - \frac{1}{H}\sum_{j=1}^{H} e^{i\langle t, \tilde{\boldsymbol{X}}_j(\theta)\rangle}\right| \right\} \int_{K_\delta} w(t)\mathrm{d}t \\
&\quad + 8\int_{K_\delta^c} w(t)\mathrm{d}t.
\end{aligned} \tag{2.6.4}$$

Applying $\sup_{\theta\in\Theta}$ on both sides of (2.6.4), using (2.6.2) combined with (d.1), and taking the limit for $\delta \downarrow 0$ gives

$$\sup_{\theta\in\Theta} |Q_{n,H}(\theta) - Q(\theta)| \overset{\text{a.s.}}{\to} 0, \quad n \to \infty. \tag{2.6.5}$$

Now we prove that $Q(\theta) = 0$ if and only if $\theta = \theta_0$. Obviously $Q(\theta_0) = 0$. If $\theta \ne \theta_0$, then the distributions of $\boldsymbol{X}_1$ and $\tilde{\boldsymbol{X}}_1(\theta)$ are different and thus also their characteristic functions are different. Since characteristic functions are continuous, it follows that they are different at least on an interval with positive Lebesgue measure; hence $Q(\theta) > 0$. Therefore, $Q(\theta)$ is uniquely minimized at θ_0 and this fact together with (2.6.5) gives strong consistency of $\hat{\theta}_{n,H}$.

2.6.2 Proof of Theorem 2.3.2

We have that $\widehat{\mathrm{Var}}(\langle t, \boldsymbol{X}_1\rangle) = t^T\hat{\Gamma}_p t$, with $\hat{\Gamma}_p$ being the p-dimensional empirical covariance matrix of the observed time series $(X_1, \ldots, X_T)$ as in (2.2.20). Let $k > 0$ be fixed and

$$Q^{(\mathrm{cv})}(\theta) = \int_{\mathbb{R}^p} |\varphi(t,\theta_0) - \varphi(t,\theta)|^2 \frac{w(t)}{t^T\Gamma_p t}\mathrm{d}t$$

be the candidate limiting function of $Q^{(\mathrm{cv})}_{n,H,k}(\theta)$ in (2.2.19), where Γ_p is the theoretical p-dimensional covariance matrix of the time series process $(X_j)_{j\in\mathbb{Z}}$.

Based on the definition of $Q^{(\mathrm{cv})}_{n,H,k}(\theta)$ in (2.2.19), we divide the domain of integration in the integrated mean squared error $|Q^{(\mathrm{cv})}_{n,H,k}(\theta) - Q^{(\mathrm{cv})}(\theta)|$ into $\{\widehat{\mathrm{Var}}(\langle t, \boldsymbol{X}_1\rangle) < k\}$ and $\{\widehat{\mathrm{Var}}(\langle t, \boldsymbol{X}_1\rangle) \geq k\}$, equivalently into $L_n = \{t \in \mathbb{R}^p : t^T\hat{\Gamma}_p t < k\}$ and its complement L_n^c.

Recall also (2.2.16) and (2.2.17). Using $|e^{ix}| = 1$ for all $x \in \mathbb{R}$, together with $|ab-cd| \leq |b||a-c| + |c||b-d|$ for $a, b, c, d \in \mathbb{C}$ gives for the integral on L_n^c:

$$\begin{aligned}
&|Q^{(\mathrm{cv})}_{n,H,k}(\theta) - Q^{(\mathrm{cv})}(\theta)|_{L_n^c} \\
&:= \int_{L_n^c}\left|\left|\frac{1}{n}\sum_{j=1}^n e^{i\langle t,\mathbf{X}_j\rangle} - \frac{1}{H}\sum_{j=1}^H e^{i\langle t,\tilde{\boldsymbol{X}}_j(\theta)\rangle}\right|^2 \frac{1}{t^T\hat{\Gamma}_p t} - |\varphi(t,\theta_0) - \varphi(t,\theta)|^2\frac{1}{t^T\Gamma_p t}\right| w(t)\mathrm{d}t \\
&\leq \int_{L_n^c}\left|\left|\frac{1}{n}\sum_{j=1}^n e^{i\langle t,\mathbf{X}_j\rangle} - \frac{1}{H}\sum_{j=1}^H e^{i\langle t,\tilde{\boldsymbol{X}}_j(\theta)\rangle}\right|^2 - |\varphi(t,\theta_0) - \varphi(t,\theta)|^2\right| \frac{1}{t^T\hat{\Gamma}_p t} w(t)\mathrm{d}t \\
&\quad + 4\int_{L_n^c}\left|\frac{1}{t^T\hat{\Gamma}_p t} - \frac{1}{t^T\Gamma_p t}\right| w(t)\mathrm{d}t.
\end{aligned} \tag{2.6.6}$$

By $(a.3)$ and $(c.1)$ it follows from Theorem 3(a) in Section 1.2.2 of Doukhan [28] that

$$|\mathrm{Cov}(X_0, X_j)| \leq 8\alpha_j^{\frac{1}{r}} (\mathbb{E}|X_1|^u)^{\frac{2}{u}} \to 0, \quad j \to \infty. \tag{2.6.7}$$

Since $\mathrm{Var}(X_1) > 0$, it follows from (2.6.7) combined with Proposition 5.1.1 in [13] that $\det(\Gamma_p) > 0$, and therefore, the minimum eigenvalue $\lambda_{\min}(\Gamma_p)$ of Γ_p is positive. Thus, for all $t \in \mathbb{R}^p$,

$$t^T\Gamma_p t \geq \lambda_{\min}(\Gamma_p)\,|t|^2 > 0. \tag{2.6.8}$$

By $(a.2)$ and the ergodic theorem $\hat{\Gamma}_p \overset{\text{a.s.}}{\to} \Gamma_p$ and, since the eigenvalues of a matrix are continuous functions of its entries (cf. Bernstein [5], Fact 10.11.2), also $\lambda_{\min}(\hat{\Gamma}_p) \overset{\text{a.s.}}{\to} \lambda_{\min}(\Gamma_p) > 0$. It follows from (2.6.8) and from the a.s. convergence of the eigenvalues that there exists $N > 0$ such that Hence, there exists some $N \in \mathbb{N}$ such that

$$t^T\hat{\Gamma}_p t \geq |t|^2\lambda_{\min}(\hat{\Gamma}_p) \geq |t|^2\frac{\lambda_{\min}(\Gamma_p)}{2} > 0, \quad n \geq N. \tag{2.6.9}$$

Thus, for $t \in L_n^c$ we obtain

$$\left|\frac{1}{t^T\hat{\Gamma}_p t} - \frac{1}{t^T\Gamma_p t}\right| \leq \frac{2}{k\lambda_{\min}(\Gamma_p)|t|^2}|t^T(\Gamma_p - \hat{\Gamma}_p)t| \leq \frac{2|\Gamma_p - \hat{\Gamma}_p|}{k\lambda_{\min}(\Gamma_p)}. \tag{2.6.10}$$

This together with (2.6.10) gives the following upper bound for the right-hand side of (2.6.6):

$$\begin{aligned}&\int_{\mathbb{R}^p}\left|\left|\frac{1}{n}\sum_{j=1}^n e^{i\langle t,\mathbf{X}_j\rangle} - \frac{1}{H}\sum_{j=1}^H e^{i\langle t,\tilde{\mathbf{X}}_j(\theta)\rangle}\right|^2 - \left|\varphi(t,\theta_0)-\varphi(t,\theta)\right|^2\right|\frac{w(t)}{k}\mathrm{d}t \\ &+ \frac{8|\Gamma_p - \hat{\Gamma}_p|}{k\lambda_{\min}(\Gamma_p)}\int_{\mathbb{R}^p} w(t)\mathrm{d}t.\end{aligned} \tag{2.6.11}$$

The first integral can be estimated as $|Q_{n,H}(\theta) - Q(\theta)|$ in (2.6.4) which tends to 0 uniformly for $\theta \in \Theta$ provided that $(d.1)$ holds. Since $\hat{\Gamma}_p \overset{\text{a.s.}}{\to} \Gamma_p$, also the second integral in (2.6.11) tends 0 a.s. as $n \to \infty$.

We turn to the integrated mean squared error $|Q^{(\text{cv})}_{n,H,k}(\theta) - Q^{(\text{cv})}(\theta)|$ on L_n. Let $L = \{t \in \mathbb{R}^p : |t| \leq \sqrt{\frac{2k}{\lambda_{\min}(\Gamma_p)}}\}$. The control variates correction used in (2.2.19) can be regarded as a continuous function $g : \mathbb{R}^9 \mapsto \mathbb{R}^2$ whose entries are the arithmetic means defined in (2.4.1)-(2.4.2). By $(c.3)$ and the uniform SLLN, each of these arithmetic means converge a.s. uniformly on $L \times \Theta$ as $n \to \infty$ and $H \to \infty$. Thus, it follows from the continuity of g and the continuous mapping theorem that

$$\sup_{(t,\theta)\in L\times\Theta} |\kappa_H(t,\theta)|^2 \overset{\text{a.s.}}{\to} 0. \tag{2.6.12}$$

For $n \geq N$ it follows from (2.6.9) that $L_n \subseteq L$ and thus using the inequality

$$||a+b|^2c - |d|^2e| \leq ||a+b|^2 - |d|^2||c| + |d|^2|c-e| \leq (|a-d| + |b|)(4+|b|)|c| + 4|c-e|,$$

valid for $a, b, c, d, e \in \mathbb{C}$ with $|d| \leq 2$ gives

$$\begin{aligned}&\int_{L_n}\left|\left|\left(\frac{1}{n}\sum_{j=1}^n e^{i\langle t,\mathbf{X}_j\rangle} - \frac{1}{H}\sum_{j=1}^H e^{i\langle t,\tilde{\mathbf{X}}_j(\theta)\rangle}\right) + \kappa_H(t,\theta)\right|^2\frac{1}{t^T\hat{\Gamma}_p t}\right. \\ &\qquad\left. - \left|\varphi(t,\theta_0)-\varphi(t,\theta)\right|^2\frac{1}{t^T\Gamma_p t}\right|w(t)\mathrm{d}t \\ \leq &\int_L\left(\left|\frac{1}{n}\sum_{j=1}^n e^{i\langle t,\mathbf{X}_j\rangle} - \varphi(t,\theta_0)\right| + \left|\frac{1}{H}\sum_{j=1}^H e^{i\langle t,\tilde{\mathbf{X}}_j(\theta)\rangle} - \varphi(t,\theta)\right|\right. \\ &\qquad\left. + |\kappa_H(t,\theta)|\right)\Big(4 + |\kappa_H(t,\theta)|\Big)\frac{w(t)}{t^T\Gamma_p t}\mathrm{d}t + 4\int_L\left|\frac{1}{t^T\hat{\Gamma}_p t} - \frac{1}{t^T\Gamma_p t}\right|w(t)\mathrm{d}t \\ :=& I_{1,n}(\theta) + I_{2,n}(\theta).\end{aligned}$$

From (2.6.8), (2.6.12), (2.6.2), and (2.6.3) with $K_\delta = L$ for $\delta = \sqrt{2k/\lambda_{\min}(\Gamma_p))}$,and $(d.4)$ it follows that $\sup_{\theta\in\Theta} I_{1,n}(\theta) \overset{\text{a.s.}}{\to} 0$ as $n \to \infty$. Finally, $\sup_{\theta\in\Theta} I_{2,n}(\theta) \overset{\text{a.s.}}{\to} 0$ by similar arguments as used in (2.6.10) and (2.6.11), since for $t \in L$, also applying $(d.4)$,

$$\left|\frac{1}{t^T\hat{\Gamma}_p t} - \frac{1}{t^T\Gamma_p t}\right| \leq \frac{2}{(\lambda_{\min}(\Gamma_p))^2|t|^4}|t^T(\Gamma_p - \hat{\Gamma}_p)t| \leq \frac{2|\Gamma_p - \hat{\Gamma}_p|}{(\lambda_{\min}(\Gamma_p))^2|t|^2}$$

and

$$\int_{\mathbb{R}^p} \frac{w(t)}{|t|^2}\mathrm{d}t < \infty$$

.

2.6.3 Proof of Theorem 2.3.3

By the definition of $\hat{\theta}_{n,H}$ in (2.2.7) and under assumptions $(a.1)$ and $(b.2)$ we have

$$\nabla_\theta Q_{n,H}(\hat{\theta}_{n,H}) = 0.$$

A Taylor expansion of order 1 of $\nabla_\theta Q_{n,H}$ around θ_0 gives

$$0 = \nabla_\theta Q_{n,H}(\theta_0) + \nabla^2_\theta Q_{n,H}(\theta_n)(\hat{\theta}_{n,H} - \theta_0)$$

where $\theta_n \overset{\text{a.s.}}{\to} \theta_0$ as $n \to \infty$. Therefore, asymptotic normality of $\sqrt{n}(\hat{\theta}_{n,H} - \theta_0)$ will follow by the delta method, if we prove that as $n \to \infty$:

(1) $\sqrt{n}\nabla_\theta Q_{n,H}(\theta_0)$ converges weakly to a multivariate normal random variable, and

(2) $\nabla^2_\theta Q_{n,H}(\theta_n)$ converges in probability to a non-singular matrix.

We start with the first point and compute the partial derivatives of $Q_{n,H}$:

$$\begin{aligned}\frac{\partial}{\partial\theta^{(i)}}Q_{n,H}(\theta) &= \frac{\partial}{\partial\theta^{(i)}}\Big(\int_{\mathbb{R}^p} |\varphi_n(t) - \varphi_H(t,\theta)|^2 w(t)\mathrm{d}t\Big)\\ &= \int_{\mathbb{R}^p} \frac{\partial}{\partial\theta^{(i)}}\Big(\Re(\varphi_n(t) - \varphi_H(t,\theta))^2 + \Im(\varphi_n(t) - \varphi_H(t,\theta))^2\Big)w(t)\mathrm{d}t\\ &= -2\int_{\mathbb{R}^p}\Big(\Re(\varphi_n(t) - \varphi_H(t,\theta))\frac{\partial}{\partial\theta^{(i)}}\Re(\varphi_H(t,\theta))\\ &\qquad + \Im(\varphi_n(t) - \varphi_H(t,\theta))\frac{\partial}{\partial\theta^{(i)}}\Im(\varphi_H(t,\theta))\Big)w(t)\mathrm{d}t, \quad i \in 1,\dots,q.\end{aligned} \tag{2.6.13}$$

Recall that $\varphi_n(t)$ and $\varphi_H(t,\theta)$ denote the empirical characteristic functions of the observed blocks $(\boldsymbol{X}_1,\dots,\boldsymbol{X}_n)$ as in (2.2.2) and of its Monte Carlo approximation

$(\tilde{\boldsymbol{X}}_1(\theta), \dots, \tilde{\boldsymbol{X}}_H(\theta))$ as in (2.2.6), respectively. Define the partial derivatives of the real and imaginary part of $\varphi_H(t,\theta)$:

$$b_H^{(i)}(t,\theta) = \frac{1}{H}\sum_{j=1}^{H} \begin{pmatrix} -\sin(\langle t, \tilde{\boldsymbol{X}}_j(\theta)\rangle) \\ \cos(\langle t, \tilde{\boldsymbol{X}}_j(\theta)\rangle) \end{pmatrix} \langle t, \frac{\partial}{\partial \theta^{(i)}} \tilde{\boldsymbol{X}}_j(\theta)\rangle, \quad i = 1, \dots, q, \tag{2.6.14}$$

and summarize them into

$$b_H(t,\theta) = \begin{pmatrix} (b_H^{(1)}(t,\theta))^T \\ \vdots \\ (b_H^{(q)}(t,\theta))^T \end{pmatrix}. \tag{2.6.15}$$

Then consider

$$\begin{pmatrix} \Re(\varphi_n(t) - \varphi(t,\theta_0)) \\ \Im(\varphi_n(t) - \varphi(t,\theta_0)) \end{pmatrix} - \begin{pmatrix} \Re(\varphi_H(t,\theta) - \varphi(t,\theta_0)) \\ \Im(\varphi_H(t,\theta) - \varphi(t,\theta_0)) \end{pmatrix} =: g_n(t) - \tilde{g}_H(t,\theta). \tag{2.6.16}$$

Abbreviate $b_H(t) := b_H(t,\theta_0)$ and $\tilde{g}_H(t) := \tilde{g}_H(t,\theta_0)$. Then it follows from (2.6.13), (2.6.15) and (2.6.16) that

$$\nabla_\theta Q_{n,H}(\theta_0) = 2\int_{\mathbb{R}^p} b_H(t) g_n(t) w(t) \mathrm{d}t - 2\int_{\mathbb{R}^p} b_H(t) \tilde{g}_H(t) w(t) \mathrm{d}t. \tag{2.6.17}$$

We analyze the asymptotic behavior of the first term in (2.6.17) in Lemma 2.6.2. More precisely, we show there that $\int_{K_\delta} b_H(t) g_n(t) w(t)\mathrm{d}t$ for K_δ as in (2.6.1) converge in distribution to a q-dimensional Gaussian vector. Afterwards, Lemmas 2.6.3 and 2.6.4 show that as $\delta \to \infty$, componentwise in $\mathbb{R}^q$,

$$\limsup_{n\to\infty} \mathrm{Var}\Big(\int_{K_\delta^c} b_H(t)\sqrt{n} g_n(t) w(t)\mathrm{d}t\Big) \to 0 \quad \text{and} \quad \int_{K_\delta^c} \mathbb{E}[b_1(t)] G(t) w(t) \mathrm{d}t \xrightarrow{P} 0,$$

where G is a zero mean $\mathbb{R}^2$-valued Gaussian field.

We show by a standard Chebyshev argument that the second term in (2.6.17) converges in probability componentwise to 0 in (2.6.45). The convergence of the second derivatives $\nabla_\theta^2 Q_n(\theta_n)$ will be the topic of Lemma 2.6.5. For the scalar products above we use the following bounds several times below.

Lemma 2.6.1. *Let $\nu \geq 1$, $t \in \mathbb{R}^p$, $k, i \in \{1, \dots, q\}$ and $j \in \mathbb{Z}$ be fixed and assume that (b.2) holds. Then the following bounds hold true.*

(a) If $\mathbb{E}|\nabla_\theta X_1(\theta)|^\nu < \infty$ for $\theta \in \Theta$, then there exists a constant $c > 0$ such that

$$\mathbb{E}\Big|\langle t, \frac{\partial}{\partial \theta^{(k)}} \tilde{\boldsymbol{X}}_j(\theta)\rangle\Big|^\nu \leq c|t|^\nu \mathbb{E}|\nabla_\theta X_1(\theta)|^\nu, \quad t \in \mathbb{R}^p. \tag{2.6.18}$$

(b) If $\mathbb{E}|\nabla_\theta^2 X_1(\theta)|^\nu < \infty$ *for* $\theta \in \Theta$, *then there exists a constant* $c > 0$ *such that*

$$\mathbb{E}\Big|\langle t, \frac{\partial}{\partial\theta^{(k)}\partial\theta^{(i)}}\tilde{\boldsymbol{X}}_j(\theta)\rangle\Big|^\nu \le c|t|^\nu \mathbb{E}|\nabla_\theta^2 X_1(\theta)|^\nu, \quad t \in \mathbb{R}^p. \tag{2.6.19}$$

The same bounds hold uniformly, taking expectations over $\sup_{\theta\in\Theta}$ *or over* $\sup_{t\in K}$ *for some compact* $K \subset \mathbb{R}^p$ *at both sides of* (2.6.18) *and* (2.6.19), *provided the corresponding expectations exist.*

Proof. (a) Applying the Cauchy-Schwarz inequality for the inner product, the fact that $(\tilde{\boldsymbol{X}}_j(\theta), \theta \in \Theta) \stackrel{\mathrm{d}}{=} (\tilde{\boldsymbol{X}}_1(\theta), \theta \in \Theta) \stackrel{\mathrm{d}}{=} (\boldsymbol{X}_1(\theta), \theta \in \Theta)$, bounding the L^2-norm by the L^1-norm, employing the inequality $|\sum_{j=1}^p \beta_j|^\nu \le p^{\nu-1}\sum_{j=1}^p |\beta_j|^\nu$ valid for $\beta_1, \dots, \beta_p \in \mathbb{R}$ and $\nu \ge 1$ gives

$$\begin{aligned}
&\mathbb{E}\Big|\langle t, \frac{\partial}{\partial\theta^{(k)}}\tilde{\boldsymbol{X}}_j(\theta)\rangle\Big|^\nu \le |t|^\nu \mathbb{E}\Big|\frac{\partial}{\partial\theta^{(k)}}\tilde{\boldsymbol{X}}_j(\theta)\Big|^\nu = |t|^\nu \mathbb{E}\Big|\frac{\partial}{\partial\theta^{(k)}}\boldsymbol{X}_1(\theta)\Big|^\nu \\
&\le |t|^\nu \mathbb{E}\Big(\sum_{r=1}^p \Big|\frac{\partial}{\partial\theta^{(k)}} X_r(\theta)\Big|\Big)^\nu \le p^{\nu-1}|t|^\nu \sum_{r=1}^p \mathbb{E}\Big|\frac{\partial}{\partial\theta^{(k)}} X_r(\theta)\Big|^\nu \\
&\le p^{\nu-1}|t|^\nu \sum_{r=1}^p \mathbb{E}|\nabla_\theta X_r(\theta)|^\nu = p^\nu |t|^\nu \mathbb{E}|\nabla_\theta X_1(\theta)|^\nu =: c|t|^\nu \mathbb{E}|\nabla_\theta X_1(\theta)|^\nu.
\end{aligned} \tag{2.6.20}$$

Part (b) follows by analogous calculations. □

Lemma 2.6.2. *Under assumptions* (a.2), (b.2), (a.3), (c.2) *and* (c.4) *we have on the Borel sets of* $\mathbb{R}^q$,

$$\int_{K_\delta} b_H(t)\sqrt{n} g_n(t) w(t) \mathrm{d}t \stackrel{d}{\to} \int_{K_\delta} \mathbb{E}[b_1(t)] G(t) w(t) \mathrm{d}t, \quad n \to \infty, \tag{2.6.21}$$

where G *is an* $\mathbb{R}^2$*-valued Gaussian field.*

Proof. Under assumptions (a.3) and (c.2), it follows from Lemma 4.1(2) in Davis et al. [23] that $\sqrt{n}(\varphi_n(\cdot) - \varphi(\cdot, \theta_0))$ convergences in distribution on compact subsets of $\mathbb{R}^p$ to a complex-valued Gaussian field $\tilde{G}$, equivalently the vector of real and imaginary part converge to a bivariate Gaussian field G. Since the random elements $(\tilde{\boldsymbol{X}}_j(\theta), \theta \in \Theta)_{j\in\mathbb{N}}$ are iid and the partial derivatives exist by (b.2), also $(\tilde{\boldsymbol{X}}_j(\theta_0), \nabla_\theta \tilde{\boldsymbol{X}}_j(\theta_0))_{j\in\mathbb{N}}$ are iid. Then it follows from the definitions (2.6.14), (2.6.15), and Lemma 2.6.1 with $K = K_\delta$) in combination with (c.4) that

$$\mathbb{E}\sup_{t\in K_\delta} |b_1(t)| \le c \sup_{t\in K_\delta} |t| \mathbb{E}|\nabla_\theta X_1(\theta_0)| \le c|\delta| \mathbb{E}|\nabla_\theta X_1(\theta_0)| < \infty. \tag{2.6.22}$$

Hence, the uniform SLLN guarantees that

$$\sup_{t\in K_\delta} |b_H(t) - \mathbb{E}b_1(t)| \stackrel{\text{a.s.}}{\to} 0, \quad n \to \infty.$$

Slutsky's theorem gives then $b_H(\cdot)\sqrt{n}g_n(\cdot,\theta_0)$ convergences in distribution on compact subsets of $\mathbb{R}^p$ to $\mathbb{E}[b_1(\cdot)]G(\cdot)$ as $n\to\infty$. The result in (2.6.21) follows from the continuity of the integral by another application of the continuous mapping theorem on $C(K_\delta)$. □

Lemma 2.6.3. *Under assumptions* $(b.2)$, $(c.4)$ *and* $(d.2)$ *we have componentwise in* $\mathbb{R}^q$,

$$\limsup_{n\to\infty}\operatorname{Var}\Big(\int_{K_\delta^c} b_H(t)\sqrt{n}g_n(t)w(t)\mathrm{d}t\Big)\to 0,\quad \delta\to\infty. \tag{2.6.23}$$

Proof. Since $b_H(\cdot)$ and $g_n(\cdot)$ are independent and $\mathbb{E}g_n(t)=0$, we have $\mathbb{E}[b_H(t)g_n(t)]=0$ for all $t\in\mathbb{R}^p$. An application of the Cauchy-Schwartz inequality for integrals gives

$$\begin{aligned}\operatorname{Var}\Big(\int_{K_\delta^c} b_H(t)\sqrt{n}g_n(t)w(t)\mathrm{d}t\Big) &= \mathbb{E}\Big(\int_{K_\delta^c} b_H(t)\sqrt{n}g_n(t)w(t)\mathrm{d}t\Big)^2\\ &\le \Big(\mathbb{E}\int_{K_\delta^c}|b_H(t)|^2 n|g_n(t)|^2 w(t)\mathrm{d}t\Big)\Big(\int_{K_\delta^c} w(t)\mathrm{d}t\Big).\end{aligned} \tag{2.6.24}$$

We first obtain a bound for the product between the first component $g_{n,1}(\cdot)$ of $g_n(\cdot)$ and the first component $b_{H,1}^{(i)}(\cdot)$ of $b_H^{(i)}(\cdot)$. Define for $t\in\mathbb{R}^p$ and $j\in\mathbb{Z}$

$$U_j(t)=\cos(\langle t,\boldsymbol{X}_j\rangle)-\Re(\varphi(t,\theta_0)),\quad V_j(t)=-\sin(\langle t,\tilde{\boldsymbol{X}}_j(\theta_0)\rangle)\langle t,\frac{\partial}{\partial\theta^{(i)}}\tilde{\boldsymbol{X}}_j(\theta_0)\rangle. \tag{2.6.25}$$

Then,

$$g_{n,1}(t)=\frac{1}{n}\sum_{j=1}^n U_j(t)\quad\text{and}\quad b_{H,1}^{(i)}(t)=\frac{1}{H}\sum_{j=1}^H V_j(t),\quad t\in\mathbb{R}^p.$$

Under $(a.3)$ it follows from Theorem 3(a) in Section 1.2.2 of Doukhan [28] that for fixed t,

$$|\operatorname{Cov}(U_0(t),U_j(t))|\le 8\alpha_j^{\frac{1}{r}}\left(\mathbb{E}|U_0(t)|^u\right)^{\frac{2}{u}},\quad j\in\mathbb{N}, \tag{2.6.26}$$

where $u=\frac{2r}{(r-1)}$ and, thus, it follows from the stationarity of $(U_j(t))_{j\in\mathbb{N}}$ combined with (2.6.26) and the fact that $|U_0(t)|\le 2$ that

$$\begin{aligned} n\mathbb{E}\Big|\frac{1}{n}\sum_{j=1}^n U_j(t)\Big|^2 &= \frac{1}{n}\sum_{j=1}^n \mathbb{E}U_j^2(t)+\frac{2}{n}\sum_{j=1}^{n-1}\Big(1-\frac{k}{n}\Big)\mathbb{E}|U_0(t)U_j(t)|\\ &\le \mathbb{E}U_0^2(t)+16\big(\mathbb{E}|U_0(t)|^u\big)^{\frac{2}{u}}\sum_{j=1}^\infty \alpha_j^{1/r}\\ &\le 4+64\sum_{j=1}^\infty\alpha_j^{1/r}<\infty,\end{aligned} \tag{2.6.27}$$

where the bound is independent of t. Recall that $H = H(n) = \bar{H}(n)n$. Under (c.4), it follows from the iid property of $(V_j(t))_{j\in\mathbb{N}}$

$$
\begin{aligned}
&n\mathbb{E}\Big|\frac{1}{H}\sum_{j=1}^{H}V_j(t)-\mathbb{E}V_0(t)\Big|^2 = n\mathrm{Var}\Big(\frac{1}{\bar{H}n}\sum_{j=1}^{\bar{H}n}V_j(t)\Big)\\
&=\frac{\mathbb{E}V_1^2(t)}{\bar{H}(n)}\leq\frac{c|t|^2\mathbb{E}|\nabla_\theta X_1(\theta_0)|^2}{\bar{H}(n)}\leq\frac{c|t|^2}{\bar{H}(n)}.
\end{aligned}
\tag{2.6.28}
$$

Using the fact that $\Big|\frac{1}{n}\sum_{j=1}^n U_j(t)\Big| \leq 2$, adding and subtracting $\mathbb{E}V_0(t)$ with the inequality $|a+b|^2 \leq 2(|a|^2+|b|^2)$, and (2.6.28) gives

$$
\begin{aligned}
&n\mathbb{E}\Big|\frac{1}{n}\sum_{j=1}^{n}U_j(t)\Big|^2\Big|\frac{1}{H}\sum_{j=1}^{H}V_j(t)\Big|^2\\
&\leq 2n\mathbb{E}\Big|\frac{1}{n}\sum_{j=1}^{n}U_j(t)\Big|^2(\mathbb{E}V_0(t))^2+8n\mathbb{E}\Big|\frac{1}{H}\sum_{j=1}^{H}V_j(t)-\mathbb{E}V_0(t)\Big|^2\\
&\leq c\Big(1+\frac{|t|^2}{\bar{H}(n)}\Big).
\end{aligned}
\tag{2.6.29}
$$

The calculations in (2.6.27), (2.6.28), and (2.6.29) can now be applied to show that for all $n\in\mathbb{N}$,

$$
n\mathbb{E}|g_n(t)|^2|b_H(t)|^2\leq c\Big(1+\frac{|t|^2}{\bar{H}(n)}\Big)
$$

and, thus, it follows from (2.6.24) together with (d.1) and (d.3) that

$$
\begin{aligned}
&\limsup_{n\to\infty}\mathrm{Var}\Big(\int_{K_\delta^c}b_H(t)\sqrt{n}g_n(t)w(t)\mathrm{d}t\Big)\\
&\leq\limsup_{n\to\infty}\frac{c}{\bar{H}(n)}\int_{K_\delta^c}(1+|t|^2)w(t)\mathrm{d}t\int_{K_\delta^c}w(t)\mathrm{d}t\to 0,\quad \delta\to\infty.
\end{aligned}
\tag{2.6.30}
$$

□

Lemma 2.6.4. *Under assumptions* (b.2)*,* (d.2) *and* (c.4)

$$
\int_{K_\delta^c}\mathbb{E}[b_1(t)]G(t)w(t)\mathrm{d}t\xrightarrow{P}0,\quad\delta\to\infty.
$$

Proof. It follows from (2.6.14), (2.6.15), (c.4), and (2.6.22) $\mathbb{E}|b_1(t)| \leq c|t|\mathbb{E}|\nabla_\theta X_1(\theta_0)| < \infty$. Now we find an upper bound for the variance of each component of $G(t)$ for a fixed t. Let $U_j(t)$ be as defined at the left-hand side of (2.6.25) and notice that the first component of $G(t)$ is the distributional limit of $\frac{1}{\sqrt{n}}\sum_{j=1}^n U_j(t)$. Since $(U_j(t))_{j\in\mathbb{N}}$ is α-mixing by (a.3), we can apply the CLT in [50] (Theorem 18.5.3 with $\delta = 2/(r-1)$) and find that the variance of the first component of $G(t)$ is given by

$$\sigma_U^2 = \mathbb{E}[U_0^2(t)] + 2\sum_{j=1}^{\infty} \mathbb{E}[U_0(t)U_j(t)].$$

This combined with Theorem 3(a) in Section 1.2.2 of [28] and the fact that $\mathbb{E}U_j(t) = 0$ and $|U_j(t)| \leq 2$ for all $j \in \mathbb{N}$ gives by $(a.3)$ and (2.6.26)

$$|\sigma_U^2| \leq 4 + \sum_{j=1}^{\infty} |\mathrm{Cov}(U_0(t), U_j(t))| \leq 4 + 8\sum_{j=1}^{\infty}(2\alpha_j)^{1/r}\left(\mathbb{E}|U_0(t)|^u\right)^{\frac{2}{u}} \leq 4 + 64\sum_{j=1}^{\infty}(2\alpha_j)^{1/r}.$$

A similar calculation shows that the variance of the second component of $G(t)$ is also bounded by a finite constant, which does not depend on t. Therefore, $\mathbb{E}|G(t)| \leq c$. This combined with (2.6.22) and assumption $(d.2)$ gives

$$\mathbb{E}\Big|\int_{K_\delta^c} \mathbb{E}[b_1(t)]G(t)w(t)\mathrm{d}t\Big| \leq c\mathbb{E}|\nabla_\theta X_1(\theta_0)| \int_{K_\delta^c} |t|w(t)\mathrm{d}t \to 0, \quad \delta \to \infty.$$

Since L^1-convergence implies convergence in probability the result follows. □

This proves part (1) of the delta method.

We now turn to part (2). In order to calculate the second derivatives of $Q_{n,H}(\theta)$, which exist by $(b.2)$, we rewrite (2.6.13) as

$$\begin{aligned}
&\frac{\partial}{\partial\theta^{(i)}} Q_{n,H}(\theta) \\
= \quad & -2\int_{\mathbb{R}^d} \Big\{\Big(\frac{1}{n}\sum_{j=1}^{n}\cos(\langle t, \boldsymbol{X}_j(\theta)\rangle) - \frac{1}{H}\sum_{j=1}^{H}\cos(\langle t, \tilde{\boldsymbol{X}}_j(\theta)\rangle)\Big)\frac{\partial}{\partial\theta^{(i)}}\Re(\varphi_H(t,\theta)) \\
& +\Big(\frac{1}{n}\sum_{j=1}^{n}\sin(\langle t, \boldsymbol{X}_j(\theta)\rangle) - \frac{1}{H}\sum_{j=1}^{H}\sin(\langle t, \tilde{\boldsymbol{X}}_j(\theta)\rangle)\Big)\frac{\partial}{\partial\theta^{(i)}}\Im(\varphi_H(t,\theta))\Big\}w(t)\mathrm{d}t \\
=: \quad & 2\int_{\mathbb{R}^d}\Big\{i_{n,H}(t,\theta)j_{H,i}(t,\theta) - k_{n,H}(t,\theta)l_{H,i}(t,\theta)\Big\}w(t)\mathrm{d}t.
\end{aligned}$$

For the second derivatives we calculate for every $i, k \in \{1, \ldots, q\}$,

$$\begin{aligned}
\frac{\partial}{\partial\theta^{(k)}\partial\theta^{(i)}} Q_{n,H}(\theta) = 2\int_{\mathbb{R}^p} \Big\{& j_{H,k}(t,\theta)j_{H,i}(t,\theta) + i_{n,H}(t,\theta)g_{H,k,i}(t,\theta) \\
&+ l_{H,k}(t,\theta)l_{H,i}(t,\theta) - k_{n,H}(t,\theta)h_{H,k,i}(t,\theta)\Big\}w(t)\mathrm{d}t,
\end{aligned} \tag{2.6.31}$$

where we summarize all quantities used in the following list:

$$\begin{aligned}
i_{n,H}(t,\theta) &= \frac{1}{n}\sum_{j=1}^{n}\cos(\langle t,\boldsymbol{X}_j\rangle) - \frac{1}{H}\sum_{j=1}^{H}\cos(\langle t,\tilde{\boldsymbol{X}}_j(\theta)\rangle)\\
j_{H,i}(t,\theta) &= \frac{\partial}{\partial\theta^{(i)}} i_{n,H}(t,\theta) = \frac{1}{H}\sum_{j=1}^{H}\sin(\langle t,\tilde{\boldsymbol{X}}_j(\theta)\rangle)\langle t,\frac{\partial}{\partial\theta^{(i)}}\tilde{\boldsymbol{X}}_j(\theta)\rangle\\
g_{H,k,i}(t,\theta) &= \frac{\partial}{\partial\theta^{(k)}} j_{H,i}(t,\theta)\\
&= \frac{1}{H}\sum_{j=1}^{H}\cos(\langle t,\tilde{\boldsymbol{X}}_j(\theta)\rangle)\langle t,\frac{\partial}{\partial\theta^{(k)}}\tilde{\boldsymbol{X}}_j(\theta)\rangle\langle t,\frac{\partial}{\partial\theta^{(i)}}\tilde{\boldsymbol{X}}_j(\theta)\rangle\\
&\qquad + \sin(\langle t,\tilde{\boldsymbol{X}}_j(\theta)\rangle)\langle t,\frac{\partial}{\partial\theta^{(k)}\partial\theta^{(i)}}\tilde{\boldsymbol{X}}_j(\theta)\rangle\\
k_{n,H}(t,\theta) &= \frac{1}{n}\sum_{j=1}^{n}\sin(\langle t,\boldsymbol{X}_j\rangle) - \frac{1}{H}\sum_{j=1}^{H}\sin(\langle t,\tilde{\boldsymbol{X}}_j(\theta)\rangle)\\
l_{H,i}(t,\theta) &= -\frac{\partial}{\partial\theta^{(i)}} k_{n,H}(t,\theta) = \frac{1}{H}\sum_{j=1}^{H}\cos(\langle t,\tilde{\boldsymbol{X}}_j(\theta)\rangle)\langle t,\frac{\partial}{\partial\theta^{(i)}}\tilde{\boldsymbol{X}}_j(\theta)\rangle\\
h_{H,k,i}(t,\theta) &= \frac{\partial}{\partial\theta^{(k)}} l_{H,i}(t,\theta)\\
&= \frac{1}{H}\sum_{j=1}^{H} -\sin(\langle t,\tilde{\boldsymbol{X}}_j(\theta)\rangle)\langle t,\frac{\partial}{\partial\theta^{(k)}}\tilde{\boldsymbol{X}}_j(\theta)\rangle\langle t,\frac{\partial}{\partial\theta^{(i)}}\tilde{\boldsymbol{X}}_j(\theta)\rangle\\
&\qquad + \cos(\langle t,\tilde{\boldsymbol{X}}_j(\theta)\rangle)\langle t,\frac{\partial}{\partial\theta^{(k)}\partial\theta^{(i)}}\tilde{\boldsymbol{X}}_j(\theta)\rangle.
\end{aligned}$$

Lemma 2.6.5. *If the assumptions* (a.2), (b.1), (b.2), (c.5), (d.3) *hold and* $(\theta_n)_{n\in\mathbb{N}} \subset \Theta$ *satisfying* $\theta_n \overset{\text{a.s.}}{\to} \theta_0$, *then for every* $k, i \in \{1,\ldots,q\}$, *as* $n\to\infty$

$$\frac{\partial}{\partial\theta^{(k)}\partial\theta^{(i)}} Q_{n,H}(\theta_n) \overset{P}{\to} \int_{\mathbb{R}^p}\Big(\mathbb{E}j_{1,k}(t,\theta_0)\mathbb{E}j_{1,i}(t,\theta_0) + \mathbb{E}l_{1,k}(t,\theta_0)\mathbb{E}l_{1,i}(t,\theta_0)\Big)w(t)\mathrm{d}t. \tag{2.6.32}$$

Proof. We first prove

$$\int_{\mathbb{R}^p} i_{n,H}(t,\theta_n) g_{H,k,i}(t,\theta_n) w(t)\mathrm{d}t \overset{P}{\to} \int_{\mathbb{R}^p}\mathbb{E}i_{1,1}(\theta_0,t)\mathbb{E}g_{1,k,i}(\theta_0,t)w(t)\mathrm{d}t, \quad n\to\infty. \tag{2.6.33}$$

Step 1: Uniform convergence on Θ: It follows from the iid property of the random elements $(\tilde{\boldsymbol{X}}_j(\theta), \theta\in\Theta)_{j\in\mathbb{N}}$ that the sequence $(\tilde{\boldsymbol{X}}_j(\theta), \nabla_\theta\tilde{\boldsymbol{X}}_j(\theta), \nabla_\theta^2\tilde{\boldsymbol{X}}_j(\theta), \theta\in\Theta)_{j\in\mathbb{N}}$ is iid. Lemma 2.6.1 together with (c.5) gives the uniform bound

$$\mathbb{E}\sup_{\theta\in\Theta}|g_{1,k,i}(t,\theta)| \le c\Big(|t|^2\mathbb{E}\sup_{\theta\in\Theta}|\nabla_\theta X_1(\theta)|^2 + |t|\mathbb{E}\sup_{\theta\in\Theta}|\nabla_\theta^2 X_1(\theta)|\Big) < \infty,$$

and it follows from the uniform SLLN that for every fixed $t \in \mathbb{R}^p$

$$\sup_{\theta\in\Theta} |g_{H,k,i}(t,\theta) - \mathbb{E}g_{1,k,i}(t,\theta)| \overset{\text{a.s.}}{\to} 0, \quad n \to \infty. \tag{2.6.34}$$

Similarly,

$$\sup_{\theta\in\Theta} \left| \frac{1}{H} \sum_{j=1}^{H} \cos(\langle t, \tilde{\boldsymbol{X}}_j(\theta)\rangle) - \Re(\varphi(t,\theta)) \right| \overset{\text{a.s.}}{\to} 0, \quad n \to \infty. \tag{2.6.35}$$

Because of $(a.2)$ the ergodic theorem gives

$$\frac{1}{n} \sum_{j=1}^{n} \cos(\langle t, \boldsymbol{X}_j\rangle) \overset{\text{a.s.}}{\to} \Re(\varphi(t,\theta_0)), \quad n \to \infty. \tag{2.6.36}$$

Therefore, (2.6.35) combined with (2.6.36) and the triangle inequality imply

$$\sup_{\theta\in\Theta} |i_{n,H}(t,\theta) - \mathbb{E}i_{1,1}(t,\theta)| \overset{\text{a.s.}}{\to} 0, \quad n \to \infty. \tag{2.6.37}$$

Step 2: Pointwise convergence of $i_{n,H}(t,\theta_n)g_{H,k,i}(t,\theta_n)$: The triangle inequality implies

$$\begin{aligned}
&|i_{n,H}(t,\theta_n)g_{H,k,i}(t,\theta_n) - \mathbb{E}i_{1,1}(\theta_0,t)\mathbb{E}g_{1,k,i}(\theta_0,t)| \\
\le& |i_{n,H}(t,\theta_n)g_{H,k,i}(t,\theta_n) - \mathbb{E}i_{1,1}(t,\theta_n)\mathbb{E}g_{1,k,i}(t,\theta_n)| \\
&+ |\mathbb{E}i_{1,1}(t,\theta_n)\mathbb{E}g_{1,k,i}(t,\theta_n) - \mathbb{E}i_{1,1}(\theta_0,t)\mathbb{E}g_{1,k,i}(\theta_0,t)| \\
\le& \sup_{\theta\in\Theta} \{|i_{n,H}(t,\theta)g_{H,k,i}(t,\theta) - \mathbb{E}i_{1,1}(t,\theta)\mathbb{E}g_{1,k,i}(t,\theta)|\} \\
&+ |\mathbb{E}i_{1,1}(t,\theta_n)\mathbb{E}g_{1,k,i}(t,\theta_n) - \mathbb{E}i_{1,1}(\theta_0,t)\mathbb{E}g_{1,k,i}(\theta_0,t)|.
\end{aligned} \tag{2.6.38}$$

Since $\theta_n \overset{\text{a.s.}}{\to} \theta_0$ and the map $\theta \mapsto \mathbb{E}i_{1,1}(t,\theta)\mathbb{E}g_{1,k,i}(t,\theta)$ is continuous in Θ, (by $(b.2)$ and $(c.5)$) it follows that the second term on the right-hand side of (2.6.38) converges a.s. to zero. Additionally, since the uniform convergences on (2.6.34) and (2.6.37) imply the uniform convergence of the product $i_{n,H}(t,\theta)g_{H,k,i}(t,\theta)$ on Θ it follows that the first term on the right-hand side of (2.6.38) also converges a.s. to zero.

Step 3: L^1-convergence: Since we have already shown a.s. convergence, it follows from Theorems 6.25(iii) and 6.19 in [54] (with $H(x) = |x|^{1+\varepsilon}$) that L^1-convergence follows provided that

$$\sup_{n\in\mathbb{N}} \mathbb{E}|i_{n,H}(t,\theta_n)g_{H,k,i}(t,\theta_n)|^{1+\epsilon} < \infty$$

for some $\epsilon > 0$. Using the fact that $|i_{n,H}(t,\theta_n)| \le 2$ and the inequality $|\frac{1}{n}\sum_{j=1}^{n}\beta_j|^{1+\epsilon} \le$

$\frac{1}{n}\sum_{j=1}^{n}|\beta_j|^{1+\epsilon}$, $\beta_1,\dots,\beta_n \in \mathbb{R}$, we obtain

$$\begin{aligned}
&\mathbb{E}|i_{n,H}(t,\theta_n)g_{H,k,i}(t,\theta_n)|^{1+\epsilon}\\
&\leq 2^{1+\epsilon}\mathbb{E}|g_{H,k,i}(t,\theta_n)|^{1+\epsilon}\\
&\leq 2^{1+\epsilon}\frac{1}{H}\sum_{j=1}^{H}\mathbb{E}\Big|\cos(\langle t,\tilde{\boldsymbol{X}}_j(\theta_n)\rangle)\langle t,\frac{\partial}{\partial\theta^{(k)}}\tilde{\boldsymbol{X}}_j(\theta_n)\rangle\langle t,\frac{\partial}{\partial\theta^{(i)}}\tilde{\boldsymbol{X}}_j(\theta_n)\rangle\\
&\qquad+\sin(\langle t,\tilde{\boldsymbol{X}}_j(\theta_n)\rangle)\langle t,\frac{\partial}{\partial\theta^{(k)}\partial\theta^{(i)}}\tilde{\boldsymbol{X}}_j(\theta_n)\rangle\Big|^{1+\epsilon}\\
&\leq 2^{1+\epsilon}\frac{1}{H}\sum_{j=1}^{H}\mathbb{E}\Big|\langle t,\frac{\partial}{\partial\theta^{(k)}}\tilde{\boldsymbol{X}}_j(\theta_n)\rangle\langle t,\frac{\partial}{\partial\theta^{(i)}}\tilde{\boldsymbol{X}}_j(\theta_n)\rangle+\langle t,\frac{\partial}{\partial\theta^{(k)}\partial\theta^{(i)}}\tilde{\boldsymbol{X}}_j(\theta_n)\rangle\Big|^{1+\epsilon},
\end{aligned}\tag{2.6.39}$$

since $|\cos(\cdot)|,|\sin(\cdot)|\leq 1$. Now we use the inequality $|a+b|^{1+\epsilon}\leq 2^{\epsilon}(|a|^{1+\epsilon}+|b|^{1+\epsilon})$ for $a,b\in\mathbb{R}$, assumption $(c.5)$ for the uniform bound in Lemma 2.6.1 and the fact that the sequence $(\tilde{\boldsymbol{X}}_j(\theta),\nabla_\theta\tilde{\boldsymbol{X}}_j(\theta),\nabla^2_\theta\tilde{\boldsymbol{X}}_j(\theta),\theta\in\Theta)_{j\in\mathbb{N}}$ is iid to continue

Now we use the inequality $|a+b|^{1+\epsilon}\leq 2^{\epsilon}(|a|^{1+\epsilon}+|b|^{1+\epsilon})$ for $a,b\in\mathbb{R}^q$, apply the Cauchy-Schwarz inequality for the inner product, and use the fact that $(\tilde{\boldsymbol{X}}_j(\theta),\nabla_\theta\tilde{\boldsymbol{X}}_j(\theta),\nabla^2_\theta\tilde{\boldsymbol{X}}_j(\theta),\theta\in\Theta)_{j\in\mathbb{N}}$ are iid and assumption $(c.5)$ for the uniform bound in Lemma 2.6.1 to continue

$$\begin{aligned}
&\leq 2^{1+2\epsilon}\frac{1}{H}\sum_{j=1}^{H}\Big(\mathbb{E}\Big|\langle t,\frac{\partial}{\partial\theta^{(k)}}\tilde{\boldsymbol{X}}_j(\theta_n)\rangle\langle t,\frac{\partial}{\partial\theta^{(i)}}\tilde{\boldsymbol{X}}_j(\theta_n)\rangle\Big|^{1+\epsilon}+\mathbb{E}\Big|\langle t,\frac{\partial}{\partial\theta^{(k)}\partial\theta^{(i)}}\tilde{\boldsymbol{X}}_j(\theta_n)\rangle\Big|^{1+\epsilon}\Big)\\
&\leq c\frac{1}{H}\sum_{j=1}^{H}\left(|t|^{2(1+\epsilon)}\mathbb{E}|\nabla_\theta X_1(\theta_n)|^{2(1+\epsilon)}+|t|^{1+\epsilon}\mathbb{E}|\nabla^2_\theta X_1(\theta_n)|^{1+\epsilon}\right)\\
&\leq c\Big(|t|^{2(1+\epsilon)}\mathbb{E}\sup_{\theta\in\Theta}|\nabla_\theta X_1(\theta)|^{2(1+\epsilon)}+|t|^{1+\epsilon}\mathbb{E}\sup_{\theta\in\Theta}|\nabla^2_\theta X_1(\theta)|^{1+\epsilon}\Big):=v(t)<\infty.
\end{aligned}\tag{2.6.40}$$

Step 4: Convergence of the random integrals: Define the sequence of functions

$$v_n(t)=\mathbb{E}|i_{n,H}(t,\theta_n)g_{H,k,i}(t,\theta_n)-\mathbb{E}i_{1,1}(\theta_0,t)\mathbb{E}g_{1,k,i}(\theta_0,t)|,\quad t\in\mathbb{R}^p,$$

and recall that from the L^1-convergence showed in Step 3, for every $t\in\mathbb{R}^p$ we have $v_n(t)\to 0$ as $n\to\infty$. From the definition of the function v in the last line of (2.6.40) it follows that $\sup_{n\in\mathbb{N}}v_n(t)\leq 2v(t)$. Additionally, assumption $(d.3)$ implies that

$$\int_{\mathbb{R}^p}v(t)w(t)\mathrm{d}t<\infty.$$

Therefore, it follows from Fubini's Theorem and dominated convergence that

$$\begin{aligned} &\mathbb{E}\Big| \int_{\mathbb{R}^p} (i_{n,H}(t,\theta_n) g_{H,k,i}(t,\theta_n) - \mathbb{E} i_{1,1}(\theta_0,t) \mathbb{E} g_{1,k,i}(\theta_0,t)) w(t) \mathrm{d}t \Big| \\ &\leq \mathbb{E} \int_{\mathbb{R}^p} | i_{n,H}(t,\theta_n) g_{H,k,i}(t,\theta_n) - \mathbb{E} i_{1,1}(\theta_0,t) \mathbb{E} g_{1,k,i}(\theta_0,t) | w(t) \mathrm{d}t \\ &= \int_{\mathbb{R}^p} v_n(t) w(t) \mathrm{d}t \to 0, \quad n \to \infty, \end{aligned} \tag{2.6.41}$$

and therefore the convergence in probability of (2.6.33) follows from the L^1-convergence in (2.6.41).

The proofs for the other three remaining integrals on the right-hand side of (2.6.31) follow along the same lines. The result in (2.6.32) is then a consequence of the fact that for all $t \in \mathbb{R}^p$, $\mathbb{E} i_{1,1}(t,\theta_0) = \mathbb{E} k_{1,1}(t,\theta_0) = 0$. □

Proof of Theorem 2.3.3: We handle each term in (2.6.17) separately. As a direct consequence of Theorem 2.3.1 and Lemmas 2.6.2, 2.6.3, 2.6.4 and 2.6.5,

$$-2(\nabla^2_\theta Q_{n,H}(\theta_n))^{-1} \int_{\mathbb{R}^p} b_H(t) \sqrt{n} g_n(t) w(t) \mathrm{d}t \xrightarrow{d} N(0, Q^{-1} W Q^{-1}), \quad n \to \infty,$$

with Q as in (2.3.3),

$$W = \mathrm{Var}\Big(\int_{\mathbb{R}^p} \mathbb{E}[b_1(t)] G(t) w(t) \mathrm{d}t \Big)$$

and G being the $\mathbb{R}^2$-valued Gaussian field from Lemma 2.6.2. For arbitrary $k, r \in \{1, \dots, q\}$ we have

$$\begin{aligned} W_{k,r} &= \mathrm{Cov}\Big(\int_{\mathbb{R}^p} \mathbb{E}[b_1^{(k)}(t)]^T G(t) w(t) \mathrm{d}t, \int_{\mathbb{R}^p} \mathbb{E}[b_1^{(r)}(t)]^T G(t) w(t) \mathrm{d}t \Big) \\ &= \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \mathbb{E}[b_1^{(k)}(t)]^T \mathbb{E}[G(t) G(s)^T] \mathbb{E}[b_1^{(k)}(s)] w(t) w(s) \mathrm{d}t \mathrm{d}s. \end{aligned} \tag{2.6.42}$$

Since $(X_j)_{j\in\mathbb{N}}$ is α-mixing by $(a.3)$, we can apply the CLT in [50] (Theorem 18.5.3 with $\delta = 2/(r-1)$) and find that

$$\mathbb{E}[G(t)G(s)^T] = \mathbb{E}[F_1(t)F_1(s)^T] + 2 \sum_{j=2}^{\infty} \mathbb{E}[F_1(t) F_j(s)^T], \tag{2.6.43}$$

where

$$F_j(t) = \begin{pmatrix} \cos(\langle t, \boldsymbol{X}_j \rangle) - \Re(\varphi(t,\theta_0)) \\ \sin(\langle t, \boldsymbol{X}_j \rangle) - \Im(\varphi(t,\theta_0)) \end{pmatrix}. \tag{2.6.44}$$

Substituting (2.6.43) and (2.6.44) into (2.6.42) gives with Fubini's Theorem

$$\begin{aligned} W_{k,r} &= \int_{\mathbb{R}^p}\int_{\mathbb{R}^p} \mathbb{E}[b_1^{(k)}(t)]^T \Big(\mathbb{E}[F_1(t)F_1(s)^T] \\ &\quad + 2\sum_{j=2}^{\infty} \mathbb{E}[F_1(t)F_j(s)^T] \Big) \mathbb{E}[b_1^{(k)}(s)] w(t)w(s) \mathrm{d}t \mathrm{d}s \\ &= \int_{\mathbb{R}^p}\int_{\mathbb{R}^p} \mathbb{E}[b_1^{(k)}(t)]^T \mathbb{E}[F_1(t)F_1(s)^T] \mathbb{E}[b_1^{(k)}(s)] w(t)w(s) \mathrm{d}t \mathrm{d}s \\ &\quad + 2\sum_{j=2}^{\infty} \int_{\mathbb{R}^{p^2}} \mathbb{E}[b_1^{(k)}(t)]^T \mathbb{E}[F_1(t)F_j(s)^T] \mathbb{E}[b_1^{(k)}(s)] w(t)w(s) \mathrm{d}t \mathrm{d}s \\ &= \mathbb{E}\Big(\int_{\mathbb{R}^p} \mathbb{E}[b_1^{(k)}(t)]^T F_1(t) w(t) \mathrm{d}t \Big)^2 \\ &\quad + 2\sum_{j=2}^{\infty} \mathbb{E}\Big[\Big(\int_{\mathbb{R}^p} \mathbb{E}[b_1^{(k)}(t)]^T F_1(t) w(t) \mathrm{d}t \Big) \Big(\int_{\mathbb{R}^p} \mathbb{E}[b_1^{(k)}(s)]^T F_j(t) w(s) \mathrm{d}s \Big) \Big], \end{aligned}$$

which gives (2.3.5).

The second term in (2.6.17) is, up to a constant,

$$\int_{\mathbb{R}^p} b_H(t) \tilde{g}_H(t) w(t) \mathrm{d}t.$$

It follows from the fact that $(\tilde{\boldsymbol{X}}_j(\theta_0))_{j\in\mathbb{N}} \stackrel{\mathrm{d}}{=} (\boldsymbol{X}_j)_{j\in\mathbb{N}}$ combined with (2.6.30) that

$$\begin{aligned} &\operatorname{Var}\Big(\int_{\mathbb{R}^p} b_H(t) \sqrt{n} \tilde{g}_H(t) w(t) \mathrm{d}t \Big) \\ &\leq \frac{c}{\bar{H}(n)} \Big(\int_{\mathbb{R}^p} (1+|t|^2) w(t) \mathrm{d}t \Big) \Big(\int_{\mathbb{R}^p} w(t) \mathrm{d}t \Big) =: \frac{c}{\bar{H}(n)} \to 0, \end{aligned} \tag{2.6.45}$$

as $n \to \infty$. Thus (2.3.4) follows from Chebyshev's inequality. □

Chapter 3:

Method of moment based estimation for the multivariate COGARCH(1,1) process

3.1 Introduction

The modeling of financial data has received much attention over the last decades, where several models have been proposed for capturing its stylized facts. Prominent models are the class of ARCH (autoregressive conditionally heteroskedastic) and GARCH (generalized ARCH) processes (see Engle [29] and Bollerslev [10], respectively). They are able to capture most of these stylized facts about financial data (see Cont [18] and Guillaume et al. [42]).

To model and understand the behavior of stochastic volatility, it is most natural to consider it as a process in continuous time, specially when dealing with high-frequency financial data. The COGARCH process is a natural generalization of the discrete time GARCH process to continuous time. It exhibits many stylized features of financial time series and is well suited for modeling high-frequency data (see Bayracı and Ünal [2], Bibbona and Negri [7], Haug et al. [45], Maller et al. [70], Klüppelberg et al. [57] and Müller [76]).

The MUCOGARCH process introduced in Stelzer [98] is a multivariate extension of the COGARCH process. It combines the features of the continuous time GARCH processes with the ones of the multivariate BEKK GARCH process of Engle and Kroner [30]. Multivariate models are necessary because in many areas of application, one has to model and understand the joint behavior of several time series (Lütkepohl [68, Section 1.1]). The MUCOGARCH is therefore, as a multivariate model, appropriate for modeling and understanding volatility in several assets jointly. It is a $d-$dimensional stochastic process and it is defined as

$$G_t = \int_0^t V_{s-}^{1/2} \mathrm{d}L_s, \quad t \geq 0, \tag{3.1.1}$$

where L is an $\mathbb{R}^d$-valued Lévy process with Lévy measure $\nu_L \not\equiv 0$ and having càdlàg sample paths. The matrix valued volatility process $(V_s)_{s\geq 0}$ depends on a parameter $\theta_0 \in \Theta \subset \mathbb{R}^q$, it is predictable and its stochasticity depends only on L. As it is common, we assume that we have a sample of size n of the log-price process (3.1.1) observed on a fixed grid of size $\Delta > 0$, and compute the log-price returns

$$\boldsymbol{G}_i = \int_{(i-1)\Delta}^{i\Delta} V_{s-}^{1/2} \mathrm{d}L_s, \quad i = 1, \dots, n. \tag{3.1.2}$$

Therefore, the question of interest is how to estimate the true parameter θ_0 from the observations $(\boldsymbol{G}_i)_{i=1}^n$. In the univariate case, several methods have been proposed to estimate the parameters of the COGARCH process (Haug et al. [45], Maller et al. [70], Bayracı and Ünal [2], Bibbona and Negri [7] and do Rêgo Sousa et al. [27]). All these methods rely on the fact that the COGARCH process is, under certain regularity conditions, ergodic and strongly mixing.

Recently, Stelzer and Vestweber [99] introduced sufficient conditions for the existence of a unique stationary distribution, for the geometric ergodicity, and for the finitenss of moments of the stationary distribution in the MUCOGARCH process. These results imply ergodicity and strong mixing of the log-price process $(\boldsymbol{G}_i)_{i=1}^\infty$, thus paving the way for statistical inference. We will employ the results of this paper to use the generalized method of moments (GMM) for estimating the parameter of the MUCOGARCH process.

Our first challenge is to compute the second-order structure of the squared returns in closed form, which already in the one-dimensional case, require rather lengthy calculations. This will be the topic of Lemmas 3.4.2 and 3.4.3.

Consistency and asymptotic normality of the GMM estimator is given under standard assumptions of mixing, existence of moments of the MUCOGARCH volatility process and model identifiability. Then we give sufficient conditions under which these assumptions will hold.

Sufficient conditions for mixing and the existence of $p \geq 1$ moments of the MUCOGARCH volatility process are already given in Stelzer and Vestweber [99]. We use their conditions and the conditions for asymptotic second-order stationarity in Stelzer [98] to obtain consistency of the GMM estimator under rather general conditions. Asymptotic normality of the estimator is also obtained under appropriate additional moment restrictions on the driving Lévy process.

The identifiability question is more delicate, since the formulas for the second-order structure of the log-price returns involve operators which are not invertible and, therefore, the strategy used for showing identifiability as in the one-dimensional COGARCH process cannot be in general applied. Instead, we derive identifiability conditions which rely mainly on the autocovariance structure of the squared returns.

Our paper is organized as follows. In Section 2, we fix the notation and briefly introduce Lévy processes. In Section 3 we define the MUCOGARCH process, and obtain in Section

4 its second-order structure. Section 5 introduces the GMM estimator and discusses sufficient conditions for stationarity, mixing and identifiability of the model parameters. In Section 3.6, we study the finite sample behavior of the estimators in a simulation study. Finally, Section 3.7 presents the proofs for results in Section 3.4.

3.2 Preliminaries

3.2.1 Notation

Denote the set of non-negative real numbers by $\mathbb{R}^+$. For $z \in \mathbb{C}$ the symbols $\Re(z)$ and $\Im(z)$ denote its real and imaginary part. We denote by $M_{m,d}(\mathbb{R})$, the set of real $m \times d$ matrices and write $M_d(\mathbb{R})$ for $M_{d,d}(\mathbb{R})$. The group of invertible $d \times d$ matrices is denoted by $GL(\mathbb{R})$, the linear subspace of symmetric matrices by $\mathbb{S}_d$, the (closed) positive semidefinite cone by $\mathbb{S}_d^+$ and the (open) positive definite cone by $\mathbb{S}_d^{++}$. We write I_d for the $d \times d$ identity matrix. The tensor (Kronecker) product of two matrices A, B is written as $A \otimes B$. The vec operator denotes the well-known vectorization operator that maps the set of $d \times d$ matrices to $\mathbb{R}^{d^2}$ by stacking the columns of the matrices below one another. Similarly, vech stacks the entries on and below the main diagonal of a square matrix. For more information regarding the tensor product, vec and vech operators we refer to Horn and Johnson [46] and Bernstein [5]. The spectrum of a square matrix is denoted by $\sigma(\cdot)$. Finally, A^* denotes the transpose of a matrix $A \in M_{m,d}(\mathbb{R})$ and $A_{(i,j)}$ denotes the entry at the ith line and jth column of A. Norms of vectors or matrices are denoted by $\|\cdot\|$, and $\|\cdot\|_2$ denotes the operator norm on $M_{d^2}(\mathbb{R})$ associated with the usual Euclidean norm. If the norm is not specified, then it is irrelevant which particular norm is used. The symbol c stands for any positive constant, whose value may change from line to line, but is not of particular interest.

Additionally, we employ an intuitive notation with respect to (stochastic) integration with matrix-valued integrators, referring to any of the standard texts (for example, Protter [86]) for a comprehensive treatment of the theory of stochastic integration. Let $(A_t)_{t\in\mathbb{R}^+}$ in $M_{m,d}(\mathbb{R})$ and $(B_t)_{t\in\mathbb{R}^+}$ in $M_{r,u}(\mathbb{R})$ be càdlàg and adapted processes and $(L_t)_{t\in\mathbb{R}^+}$ in $M_{d,r}(\mathbb{R})$ be a semimartingale. We then denote by $\int_0^t A_{s-}\mathrm{d}L_s B_{s-}$ the matrix $C_t \in M_{m,u}(\mathbb{R})$ which has ij-th entry $\sum_{k=1}^d \sum_{l=1}^r \int_0^t A_{ik,s-} B_{lj,s-} \mathrm{d}L_{kl,s}$. If $(X_t)_{t\in\mathbb{R}^+}$ is a semimartingale in $\mathbb{R}_m$ and $(Y_t)_{t\in\mathbb{R}^+}$ one in $\mathbb{R}_d$, then the quadratic variation $([X,Y]_t)_{t\in\mathbb{R}^+}$ is defined as the finite variation process in $M_{m,d}(\mathbb{R})$ with ij-th entry $[X_i, Y_j]_t$ for $t \in \mathbb{R}^+$, $i = 1, \ldots, m$ and $j = 1, \ldots, d$. We also refer to Lemma 2.2 in Behme [3] for a collection of basic properties related to integration with matrix-valued integrators. Lastly, let $\boldsymbol{Q} : M_{d^2}(\mathbb{R}) \mapsto M_{d^2}(\mathbb{R})$ be the linear map defined by

$$(\boldsymbol{Q}X)_{(k-1)d+l,(p-1)d+q} = X_{(k-1)d+p,(l-1)d+q} \quad \text{for all } k,l,p,q = 1,\ldots,d, \tag{3.2.1}$$

which has the property that $\boldsymbol{Q}(\mathrm{vec}(X)\,\mathrm{vec}(Z)^T) = X \otimes Z$ for all $X, Z \in \mathbb{S}_d$ (Pigorsch and Stelzer [82, Theorem 4.3]). Let K_d be the commutation matrix which can be characterized by $K_d\,\mathrm{vec}(A) = \mathrm{vec}(A^*)$ for all $A \in M_d(\mathbb{R})$ (see Magnus and Neudecker [69] for more details). Now, define $\mathcal{Q} \in M_{d^4}(\mathbb{R})$ as the matrix associated with the linear map $\mathrm{vec} \circ \boldsymbol{Q} \circ \mathrm{vec}^{-1}$ on $\mathbb{R}^{d^4}$, and $\mathcal{K}_d \in M_{d^4}(\mathbb{R})$ as the matrix associated with the linear map $\mathrm{vec}(K_d\,\mathrm{vec}^{-1}(x))$ for $x \in \mathbb{R}^{d^4}$.

3.2.2 Lévy processes

A Lévy process $L = (L_t)_{t\in\mathbb{R}^+}$ in $\mathbb{R}^d$ is defined by its Lévy-Khintchine form $\mathbb{E}e^{i\langle u,L_t\rangle} = \exp\{t\psi_L(u)\}$ for $t \in \mathbb{R}^+$ with

$$\psi_L(u) = i\langle \gamma_L, u\rangle - \frac{1}{2}\langle u, \Gamma_L u\rangle + \int_{\mathbb{R}^d} (e^{i\langle u,x\rangle} - 1 - i\langle u, x\rangle I_{[0,1]}(\|x\|))\nu_L(dx), \quad u \in \mathbb{R}^d,$$

where $\gamma_L \in \mathbb{R}^d, \Gamma_L \in \mathbb{S}_d^+$ and the Lévy measure ν_L is a measure on $\mathbb{R}^d$ satisfying $\nu_L(\{0\}) = 0$ and $\int_{\mathbb{R}^d} (\|x\|^2 \wedge 1)\,\nu_L(dx) < \infty$. We assume L to have càdlàg paths. The discontinuous part of the quadratic variation of L is denoted by $([L, L]_t^{\mathfrak{d}})_{t\in\mathbb{R}^+}$ and it is also a Lévy process. It has finite variation, zero drift and Lévy measure $\nu_{[L,L]^{\mathfrak{d}}}(B) = \int_{\mathbb{R}^d} I_B\,(xx^*)\,\nu_L(dx)$ for all Borel sets $B \subseteq \mathbb{S}_d$. For more details on Lévy processes we refer to Applebaum [1] and Sato [90].

3.3 The MUCOGARCH process

Throughout, we assume that all random variables and processes are defined on a given filtered probability space $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\in T})$, with $T = \mathbb{N}$ in the discrete-time case and $T = \mathbb{R}^+$ in the continuous-time one. In the continuous-time setting, we assume the usual conditions (complete, right-continuous filtration) to be satisfied. Moreover, we implicitly assume that the given filtered probability space is enlarged (Protter [86, p. 293]), to allow for arbitrary initial conditions of stochastic differential equations. We start with the definition of the MUCOGARCH process.

Definition 3.3.1. *(MUCOGARCH(1,1) - Stelzer [98, Definition 3.1]) Let L be an $\mathbb{R}^d$-valued Lévy process and $A, B \in M_d(\mathbb{R})$ and $C \in \mathbb{S}_d^{++}$. The process $G = (G_t)_{t\in\mathbb{R}^+}$ solving*

$$\mathrm{d}G_t = V_{t-}^{1/2}\mathrm{d}L_t \tag{3.3.1}$$

$$V_t = C + Y_t \tag{3.3.2}$$

$$\mathrm{d}Y_t = (BY_{t-} + Y_{t-}B^*)\mathrm{d}t + AV_{t-}^{1/2}\mathrm{d}[L, L]_t^{\mathfrak{d}} V_{t-}^{1/2} A^* \tag{3.3.3}$$

with initial values G_0 in $\mathbb{R}^d$ and Y_0 in $\mathbb{S}_d^+(\mathbb{R})$ is called a MUCOGARCH(1,1) process. The process $Y = (Y_t)_{t\geq 0}$ is called a MUCOGARCH(1,1) volatility process. Hereafter we will always write MUCOGARCH for short.

The interpretation of the model parameters B and C is the following. If $\sigma(B) \in (-\infty, 0) + i\mathbb{R}$, the process V, as long as no jump occurs, returns to the level C at an exponential rate determined by B. Additionally, since all jumps are positive semidefinite, C is not a mean level but, instead, a lower bound for V.

According to Theorems 3.2 and 4.4 in Stelzer [98], in order for the MUCOGARCH process to be well-defined, we need that its initial value $Y_0^+ \in \mathbb{S}_d^+$. In this case, the solution $(Y_t)_{t\in\mathbb{R}^+}$ is locally bounded and of finite variation. Additionally, the process $(G_t, Y_t)_{t\in\mathbb{R}^+}$ and its volatility process $(Y_t)_{t\in\mathbb{R}^+}$ are time homogeneous strong Markov processes on $\mathbb{R}^d \times \mathbb{S}_d^+$ and $\mathbb{S}_d^+$, respectively.

3.4 Second-order structure of the MUCOGARCH process

In this section, we compute the second-order structure of the squared returns process $(\boldsymbol{G}_i\boldsymbol{G}_i^*)_{i\in\mathbb{N}}$ defined in terms of (3.1.2), which will be used in Section 3.5 bellow to estimate the parameters A, B and C of the MUCOGARCH.

First, we will state various assumptions on different aspects of the underlying Lévy process L, and the model parameter. We group these assumptions into the following categories.

Assumptions a (Lévy process)**.**

$(a.1)$ *L is an $\mathbb{R}^d$-valued Lévy process with non-zero Lévy measure ν_L.*

$(a.2)$
$$\int_{\mathbb{R}^d} x_i x_j x_k \, \nu_L(dx) = 0, \quad \textit{for all } i, j, k \in \{1, \ldots, d\}.$$

$(a.3)$ $\mathbb{E}\|L_1\|^4 < \infty$.

$(a.4)$ $\mathbb{E}L_1 = 0$.

$(a.5)$ $\mathrm{Var}(L_1) = (\sigma_W + \sigma_L)I_d$, *with* $\sigma_W \geq 0$ *and* $\sigma_L > 0$.

$(a.6)$ *There exists a constant $\rho_L > 0$ such that*
$$\mathbb{E}[\mathrm{vec}([L, L^*]^{\mathfrak{d}}), \mathrm{vec}([L, L^*]^{\mathfrak{d}})^*]_1^{\mathfrak{d}} = \rho_L(I_{d^2} + K_d + \mathrm{vec}(I_d)\,\mathrm{vec}(I_d)^*).$$

Assumptions b (Parameter space)**.**

$(b.1)$ *The matrix*
$$\mathcal{B} = B \otimes I + I \otimes B + \sigma_L(A \otimes A), \tag{3.4.1}$$
is invertible.

(*b*.2) *The matrix*

$$\mathcal{C} = \mathcal{B} \otimes I_{d^2} + I_{d^2} \otimes \mathcal{B} + \mathcal{A}\mathcal{R} \tag{3.4.2}$$

is invertible, where $\mathcal{A} = (A \otimes A) \otimes (A \otimes A)$ *and* $\rho_L(\mathcal{Q} + \mathcal{K}_d\mathcal{Q} + I_{d^4})$.

Assumption c (Second-order stationarity)**.**

(*c*.1) $(Y_t)_{t\in\mathbb{R}_+}$ *is a second-order stationary MUCOGARCH volatility process.*

We recall now the expressions for the second-order structure of the process Y and of the log returns process $(\boldsymbol{G}_i)_{i\in\mathbb{N}}$.

Proposition 3.4.1 (Stelzer [98, Theorems 4.8, 4.11, Corollary 4.19 and Proposition 5.2])**.** *If assumptions* (*a*.1)*,* (*a*.3)*-*(*a*.6)*, all of* ***b*** *and* (*c*.1) *hold, then*

$$\begin{aligned}
\mathbb{E}(\mathrm{vec}(Y_0)) &= -\sigma_L \mathcal{B}^{-1}(A\otimes A)\,\mathrm{vec}(C) && (3.4.3)\\
\mathrm{acov}_Y(h) &= e^{\mathcal{B}h}\mathrm{Var}(\mathrm{vec}(Y_0)) && (3.4.4)\\
\mathrm{Var}(\mathrm{vec}(Y_0)) &= -\mathcal{C}^{-1}\big[(\sigma_L^2\mathcal{C}(\mathcal{B}^{-1}\otimes\mathcal{B}^{-1})\mathcal{A} + \mathcal{A}\mathcal{R})(\mathrm{vec}(C)\otimes\mathrm{vec}(C)) \\
&\quad +(\sigma_L(A\otimes A)\otimes I_{d^2} + \mathcal{A}\mathcal{R})\,\mathrm{vec}(C)\otimes\mathbb{E}(\mathrm{vec}(Y_0)) \\
&\quad +(\sigma_L I_{d^2}\otimes(A\otimes A) + \mathcal{A}\mathcal{R})\mathbb{E}(\mathrm{vec}(Y_0))\otimes\mathrm{vec}(C)\big] && (3.4.5)\\
\mathbb{E}(\boldsymbol{G}_1) &= 0 \\
\mathrm{Var}(\boldsymbol{G}_1) &= (\sigma_L+\sigma_W)\Delta\mathbb{E}(C+Y_0) && (3.4.6)\\
\mathrm{acov}_{\boldsymbol{G}}(h) &= 0 \quad \textit{for all } h\in\mathbb{Z}/\{0\}.
\end{aligned}$$

Based on Proposition 3.4.1, we state now the second-order properties of the MUCOGARCH process. The proofs are postponed to Section 3.7. In the following, for a second-order stationary $\mathbb{R}^d$-valued process, its autocovariance function $\mathrm{acov}_X : \mathbb{R} \mapsto M_d(\mathbb{R})$ is given by $\mathrm{acov}_X(h) = \mathrm{cov}(X_h, X_0) = E(X_h X_0^*) - E(X_0)E(X_0)^*$ for $h \geq 0$ and by $\mathrm{acov}_X(h) = (\mathrm{acov}_X(-h))^*$ for $h < 0$. For matrix-valued processes $(Z_t)_{t\in\mathbb{R}}$, we set $\mathrm{acov}_Z = \mathrm{acov}_{\mathrm{vec}(Z)}$.

Lemma 3.4.2. *If all assumptions* ***a****,****b*** *and* **c** *hold and the matrix* A *is invertible, then*

$$\begin{aligned}
\mathrm{acov}_{\boldsymbol{G}\boldsymbol{G}^*}(h) = e^{\mathcal{B}\Delta h}&\mathcal{B}^{-1}(I_{d^2} - e^{-\mathcal{B}\Delta})(\sigma_L+\sigma_W)\mathrm{Var}(\mathrm{vec}(V_0)) \\
&\times(e^{\mathcal{B}^*\Delta} - I_{d^2})[(\sigma_W+\sigma_L)(\mathcal{B}^*)^{-1} - 2((A\otimes A)^*)^{-1}]
\end{aligned} \tag{3.4.7}$$

Lemma 3.4.3. *If the conditions of Lemma 3.4.2 hold, then*

$$\begin{aligned}
&\mathbb{E}\,\mathrm{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*)\,\mathrm{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*)^* \\
&= \Delta\rho_L(\boldsymbol{Q} + K_d\boldsymbol{Q} + I_{d^2})\mathbb{E}\,\mathrm{vec}(V_0)\,\mathrm{vec}(V_0)^*(I_{d^2}+K_d)\boldsymbol{Q}(D^*)(I_{d^2}+K_d) + D + D^*,
\end{aligned} \tag{3.4.8}$$

with $\boldsymbol{Q}$ *as in* (3.2.1), K_d *is the commutation matrix,*

$$D := (\sigma_L + \sigma_W)\Big(\frac{1}{2}(\sigma_L + \sigma_W)\Delta^2 \mathbb{E}\operatorname{vec}(V_0)\mathbb{E}\operatorname{vec}(V_0)^* + \operatorname{Var}(\operatorname{vec}(V_0))\tilde{\mathcal{B}}\Big) \tag{3.4.9}$$

$$\tilde{\mathcal{B}} := \big[(\mathcal{B}^*)^{-1}(e^{\mathcal{B}^*\Delta} - I_{d^2}) - I_{d^2}\Delta\big]\big[(\sigma_W + \sigma_L)(\mathcal{B}^*)^{-1} - 2((A \otimes A)^*)^{-1}\big] \tag{3.4.10}$$

Remark 3.4.4. *In the one-dimensional case* ($d = 1$)*, using the parametrization* $\varphi = A \otimes A$, $\Xi = -2B$, $\beta = -2BC$*, gives the COGARCH(1,1) in Klüppelberg et al. [55] and simple algebra shows that* (3.4.7) *and* (3.4.8) *agree with the formulas given in (2.6) and (2.5), respectively in Haug et al. [45].*

3.5 Moment based estimation of the MUCOGARCH process

In this section, we consider the matrices $A_\theta, B_\theta \in M_d(\mathbb{R})$ and $C_\theta \in \mathbb{S}_d^{++}$ from Definition 3.3.1 as depending on a parameter $\theta \in \Theta \subset \mathbb{R}^q$ for $q \in \mathbb{N}$. Throughout, we assume that all Assumptions **a** and **b** from Section 3.4 hold.

The data used for estimation is a sample of d-dimensional log-price process $(\boldsymbol{G}_i)_{i=1}^n$ as defined in (3.1.2) with true parameter $\theta_0 \in \Theta$. We assume that the quantities σ_L, σ_W and ρ_L defined in Assumptions ($a.5$) and ($a.6$) are known. These assumptions are not very restrictive and are comparable to assuming iid standard noise in the discrete time multivariate GARCH process, which is very common (see eq. (11.6) in Francq and Zakoïan [36]).

Based on the observations $(\boldsymbol{G}_i)_{i=1}^n$ and a fixed $H < n$, the sample moments are defined as

$$\hat{k}_{n,H} = \frac{1}{n}\sum_{i=1}^{n-H} D_i = \frac{1}{n}\sum_{i=1}^{n-H}\begin{pmatrix} \operatorname{vec}(\boldsymbol{G}_i\boldsymbol{G}_i^*) \\ \operatorname{vec}(\operatorname{vec}(\boldsymbol{G}_i\boldsymbol{G}_i^*)\operatorname{vec}(\boldsymbol{G}_i\boldsymbol{G}_i^*)^*) \\ \vdots \\ \operatorname{vec}(\operatorname{vec}(\boldsymbol{G}_i\boldsymbol{G}_i^*)\operatorname{vec}(\boldsymbol{G}_{i+H}\boldsymbol{G}_{i+H}^*)^*) \end{pmatrix}. \tag{3.5.1}$$

The value H needs to be chosen in such a way that the model parameters are identifiable and also to ensure a good fit of the autocovariance structure to the data. For each $\theta \in \Theta$, for which the matrices A_θ, B_θ and C_θ satisfy the assumptions required in Lemma 3.4.2, let

$$k_{\theta,H} = \begin{pmatrix} \mathbb{E}_\theta \operatorname{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*) \\ \mathbb{E}_\theta \operatorname{vec}(\operatorname{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*)\operatorname{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*)^*) \\ \vdots \\ \mathbb{E}_\theta \operatorname{vec}(\operatorname{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*)\operatorname{vec}(\boldsymbol{G}_{1+H}\boldsymbol{G}_{1+H}^*)^*) \end{pmatrix}, \tag{3.5.2}$$

where the expectations are defined as in (3.4.6), (3.4.7) and (3.4.8) by replacing A, B and C by A_θ, B_θ and C_θ, respectively. Then, the GMM estimator of θ_0 is given by

$$\hat{\theta}_n = \underset{\theta\in\Theta}{\arg\min}\Big\{(\hat{k}_{n,H} - k_{\theta,H})^T\Omega(\hat{k}_{n,H} - k_{\theta,H})\Big\}, \tag{3.5.3}$$

where Ω is a positive definite weight matrix. Additionally to Assumption **a** and **b** we provide additional assumptions required for proving consistency and asymptotic normality of $\hat{\theta}_n$.

Assumptions d (Parameter space and log-price process)**.**

(d.1) *The parameter space* Θ *is a compact subset of* $\mathbb{R}^q$.

(d.2) *The true parameter* θ_0 *lies in the interior of* Θ.

(d.3) *The matrix* A_θ *is invertible for all* $\theta \in \Theta$.

(d.4) *[Identifiability]. Let* $H > 1$ *be fixed. For any* $\theta \neq \tilde{\theta} \in \Theta$ *we have* $k_{\theta,H} \neq k_{\tilde{\theta},H}$.

(d.5) *The map* $\theta \mapsto (A_\theta, B_\theta, C_\theta)$ *is twice continuously differentiable.*

(d.6) *The sequence* $(\boldsymbol{G}_n)_{n\in\mathbb{N}}$ *is stationary and strongly mixing with exponentially decaying mixing coefficients* $\alpha_{\boldsymbol{G}}$.

Assumptions e (Moments)**.**

(e.1) $\mathbb{E}\|\boldsymbol{G}_1\|^4 < \infty$.

(e.2) *There exists a positive constant* $\delta > 0$ *such that* $\mathbb{E}\|\boldsymbol{G}_1\|^{8+\delta} < \infty$.

Assumptions (e.1) and (e.2) can be written in terms of moments of L and Y_0.

Lemma 3.5.1. *If* $\mathbb{E}\|Y_0\|^r < \infty$ *and* $\mathbb{E}\|L_1\|^{2r} < \infty$ *for some* $r \geq 1$, *then* $\mathbb{E}\|\boldsymbol{G}_1\|^{2r} < \infty$.

Proof. It follows from Stelzer [98, Proposition 4.7] (with $k = r$) that $\mathbb{E}\|Y_t\|^r < \infty$ for all $t \in \mathbb{R}^+$ and $t \mapsto \mathbb{E}\|Y_t\|^r$ is locally bounded. Then an application of Protter [86, Theorem 66 of Ch. 5] together with the fact that $\mathbb{E}\|L_1\|^{2r} < \infty$ and the definition of $(V_t)_{t\in\mathbb{R}_+}$ in (3.3.2) gives

$$\mathbb{E}\|\boldsymbol{G}_1\|^{2r} = \mathbb{E}\Big\|\int_0^\Delta V_{s-}^{1/2}\mathrm{d}L_s\Big\|^{2r} \leq c\int_0^\Delta \mathbb{E}\|V_{s-}^{1/2}\|^{2r}\mathrm{d}s \leq c\int_0^\Delta \mathbb{E}\|C + Y_{s-}\|^r\mathrm{d}s < \infty.$$

□

We are now ready to state the strong consistency of the empirical moments.

Lemma 3.5.2. *If* $(d.6)$ *and* $(e.1)$ *hold, then* $\hat{k}_{n,H} \overset{\text{a.s.}}{\to} k_{\theta_0}$ *as* $n \to \infty$.

Proof. It follows from $(d.6)$ that the log-price process $(\boldsymbol{G}_n)_{n\in\mathbb{N}}$ is ergodic. By $(e.1)$ both $\mathbb{E}\|\operatorname{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*)\|$ and $\mathbb{E}\|\operatorname{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*)\operatorname{vec}(\boldsymbol{G}_{1+h}\boldsymbol{G}_{1+h}^*)^*\|$ are finite, and thus we can apply Birkhoff's ergodic theorem (Krengel [62, Theorem 4.4]) to conclude the result. □

Next, we state the weak consistency property of the GMM estimator.

Theorem 3.5.3. *If* $(d.1)$*,* $(d.3)$*,* $(d.4)$*,* $(d.5)$*,*$(d.6)$ *and* $(e.1)$ *hold, then the GMM estimator defined in* (3.5.3) *is weakly consistent.*

Proof. We check assumptions 1.1-1.3 in Mátyás et al. [71] that ensure weak consistency of the GMM estimator in (3.5.3). Assumption 1.1 is also satisfied due to our identifiability condition $(d.4)$. It follows from (3.5.3) combined with Lemma 3.5.2 that

$$\sup_{\theta\in\Theta} \|\hat{k}_{n,H} - k_{\theta,H} - (k_{\theta_0,H} - k_{\theta,H})\| = \|\hat{k}_{n,H} - k_{\theta_0,H}\| \overset{\text{a.s.}}{\to} 0, \quad n \to \infty, \tag{3.5.4}$$

which is assumption 1.2 of [71]. Since the weight matrix Ω in (3.5.3) is non-random, their assumption 1.3 is automatically satisfied, completing the proof. □

In order to prove asymptotic normality of the GMM estimator, we need some auxiliary results.

Lemma 3.5.4. *If* $(d.1)$ *and* $(d.5)$ *hold, then the map* $\Theta \mapsto k_{\theta,H}$ *in* (3.5.2) *is continuously differentiable.*

Proof. The the map $\Theta \mapsto k_{\theta,H}$ depends on the moments given in (3.4.6), (3.4.7) and (3.4.8). These moments are given in terms of products and Kronecker products involving the quantities $A_\theta, A_\theta^{-1}, \mathcal{B}_\theta, \mathcal{B}_\theta^{-1}, e^{-\alpha\mathcal{B}_\theta}, \alpha > 0, C_\theta, \mathcal{C}_\theta$ and $\mathcal{C}_\theta^{-1}$. From $(d.5)$ we obtain the continuous differentiability of $\mathcal{B}_\theta, \mathcal{B}_\theta^{-1}, \mathcal{C}_\theta, \mathcal{C}_\theta^{-1}$ and $(A_\theta \otimes A_\theta)^{-1}$ on Θ. Let $i \in \{1, \ldots, q\}$ be fixed. According to (2.1) in Wilcox [103], the matrix exponential is differentiable and

$$\frac{\partial}{\partial\theta_i} e^{-\alpha\mathcal{B}_\theta} = -\int_0^\alpha e^{-(\alpha-u)\mathcal{B}_\theta}\left(\frac{\partial}{\partial\theta_i}\mathcal{B}_\theta\right)e^{-u\mathcal{B}_\theta}\mathrm{d}u. \tag{3.5.5}$$

Using the definition of $\mathcal{B}_\theta$ in (3.4.1) combined with $(d.1)$ and $(d.5)$ gives

$$\sup_{\theta\in\Theta}\|\mathcal{B}_\theta\| \le 2\left(\sup_{\theta\in\Theta}\|B_\theta\|\right)\|I_d\| + \sigma_L\left(\sup_{\theta\in\Theta}\|A_\theta\|^2\right) < \infty. \tag{3.5.6}$$

Additionally, an application of the chain rule to $\frac{\partial}{\partial\theta_i}\mathcal{B}_\theta$ combined with $(d.1)$ and $(d.5)$ gives $\sup_{\theta\in\Theta}\|\frac{\partial}{\partial\theta_i}\mathcal{B}_\theta\| < \infty$ and, therefore,

$$\sup_{\theta\in\Theta}\left\|e^{-(\alpha-u)\mathcal{B}_\theta}\left(\frac{\partial}{\partial\theta_i}\mathcal{B}_\theta\right)e^{-u\mathcal{B}_\theta}\right\| \le \sup_{\theta\in\Theta} e^{(|\alpha-u|+|u|)\|\mathcal{B}_\theta\|}\left(\sup_{\theta\in\Theta}\left\|\frac{\partial}{\partial\theta_i}\mathcal{B}_\theta\right\|\right), \quad u \in [0,\alpha]. \tag{3.5.7}$$

Thus, the continuous differentiability of the map in (3.5.5) follows by dominated convergence with dominating function as in (3.5.7). Another application of the chain rule shows that the map $\theta \mapsto k_{\theta,H}$ is continuously differentiable on Θ. $\square$

Lemma 3.5.5. *Assume that* $(d.2)$, $(d.6)$ *and* $(e.2)$ *hold and let*

$$\Sigma \;=\; \mathbb{E}(F_1 F_1^*) + 2\sum_{i=1}^{\infty} \mathbb{E}(F_1 F_{1+i}^*) \tag{3.5.8}$$

with $F_i = D_i - k_{\theta_0,H}$ *and* D_i *as defined in* (3.5.1). *Then for* $H \in \mathbb{N}_0$

$$\sqrt{n}(\hat{k}_{n,H} - k_{\theta_0,H}) \xrightarrow{d} \mathcal{N}(0,\Sigma), \quad n \to \infty. \tag{3.5.9}$$

Proof. For the asymptotic normality of (3.5.1) we use the Cramér-Wold device and show that

$$\sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n-H} \lambda^* F_i\Big) \xrightarrow{d} \mathcal{N}(0,\lambda^*\Sigma\lambda), \quad n \to \infty,$$

for all vectors $\lambda \in \mathbb{R}^{d^2+(H+1)d^4}$. Since each F_i is a measurable function of $G_i, \ldots, G_{i+H}$ it follows from $(d.6)$ and Remark 1.8 of Bradley [12] that $(\lambda^* F_i)_{i\in\mathbb{N}}$ is α-mixing with mixing coefficients satisfying $\alpha_F(n) \le \alpha_{\boldsymbol{G}}(n-(r+1))$ for all $n \ge r+2$. Therefore, $\sum_{n=0}^{\infty}(\alpha_F(n))^{\frac{\epsilon}{2+\epsilon}} < \infty$ for all $\epsilon > 0$. From $(e.2)$ we obtain $\mathbb{E}\|\lambda^* F_1\|^{2+\epsilon/4} < \infty$ for some $\epsilon > 0$. Thus, the CLT for α-mixing sequences applies (see Theorem 18.5.3 of Ibragimov and Linnik [50]), so that

$$\sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n-H} \lambda^* F_i\Big) \xrightarrow{d} \mathcal{N}(0,\zeta), \quad n \to \infty,$$

where

$$\zeta = \mathbb{E}\lambda^* F_1 F_1^* \lambda + 2\sum_{i=1}^{\infty} \mathbb{E}\lambda^* F_1 F_{1+i}^* \lambda.$$

After rearranging this equation we find (3.5.8). $\square$

Theorem 3.5.6. *Assume the conditions of Theorem 3.5.3 and that the matrix* Σ *in* (3.5.8) *is positive definite. If additionally Assumption* $(e.2)$ *holds, then the GMM estimator defined in* (3.5.3) *is asymptotically normal with covariance matrix* $(\mathcal{J}_{\theta_0})^{-1}\mathcal{I}_{\theta_0}(\mathcal{J}_{\theta_0})^{-1}$ *with*

$$\mathcal{J}_{\theta_0} = (\nabla_\theta k_{\theta_0,H})^\top \Omega (\nabla_\theta k_{\theta_0,H}) \quad \text{and} \quad \mathcal{I}_{\theta_0} = (\nabla_\theta k_{\theta_0,H})^\top \Omega \Sigma_{\theta_0} \Omega (\nabla_\theta k_{\theta_0,H}) \tag{3.5.10}$$

Proof. We check Assumptions 1.7-1.9 of Theorem 1.2 in Mátyás et al. [71]. Since from Lemma 3.5.4 the map $\theta \mapsto k_{\theta,H}$ is continuously differentiable, Assumption 1.7 is valid. Now, for any sequence $\tilde{\theta}_n$ such that $\tilde{\theta}_n \xrightarrow{P} \theta_0$ as $n \to \infty$, it follows from the continuous mapping theorem by the continuity of the map $\Theta \mapsto \frac{\partial}{\partial\theta} k_{\theta,H}$ in Lemma 3.5.4 that $\frac{\partial}{\partial\theta}(\hat{k}_{n,H} - k_{\theta_n}) \xrightarrow{P} (k_{\theta_0} - \frac{\partial}{\partial\theta} k_{\theta_0})$ as $n \to \infty$. Therefore, Assumption 1.8 is also satisfied. Since Lemma 3.5.5 implies Assumption 1.9, we conclude the result. $\square$

3.5.1 Sufficient conditions for strict stationarity and mixing

Sufficient conditions for the existence of a unique stationary distribution, geometric ergodicity and for the finiteness of moments of order p of the stationary distribution have recently been given in Stelzer and Vestweber [99]. We state these conditions in the next theorem, which are conditions (i), (iv) and (v) of Theorem 4.3 in [99].

Theorem 3.5.7. *[Geometric Ergodicity - [99, Theorem 4.3]] Let Y be a MUCOGARCH volatility process which is μ-irreducible with support having non-empty interior and aperiodic. Assume that one of the following conditions is satisfied:*

(i) setting $p = 1$ there exists an $\Xi \in \mathbb{S}_d^{++}$ such that

$$\Xi B + B^\top \Xi + \sigma_L A^\top \Xi A \in -\mathbb{S}_d^{++}, \tag{3.5.11}$$

where σ_L is as defined in $(a.5)$.

(ii) there exist $p \in [1, \infty)$ and an $\Xi \in \mathbb{S}_d^{++}$ such that

$$\int_{\mathbb{R}^d} \left(2^{p-1}\left(1 + K_{\Xi,A}\|y\|_2^2\right)^p - 1\right) \nu_L(dy) + pK_{\Xi,B} < 0, \tag{3.5.12}$$

where $K_{\Xi,B} := \max_{X \in \mathbb{S}_d^+, \operatorname{tr}(X)=1} \frac{\operatorname{tr}\left((\Xi B + B^\top \Xi)X\right)}{\operatorname{tr}(\Xi X)}$ and
and $K_{\Xi,A} := \max_{X \in \mathbb{S}_d^+, \operatorname{tr}(X)=1} \frac{\operatorname{tr}\left(A^\top \Xi A X\right)}{\operatorname{tr}(\Xi X)}$,

(iii) there exist $p \in [1, \infty)$ and an $\Xi \in \mathbb{S}_d^{++}$ such that

$$\max\left\{2^{p-2}, 1\right\} K_{\Xi,A} \int_{\mathbb{R}^d} \|y\|_2^2 \left(1 + \|y\|_2^2 K_{\Xi,A}\right)^{p-1} \nu_L(dy) + K_{\Xi,B} < 0 \tag{3.5.13}$$

where $K_{\Xi,B}, K_{\Xi,A}$ are as in (ii).

Then a unique stationary distribution for the MUCOGARCH volatility process Y exists, Y is positive Harris recurrent, geometrically ergodic and the stationary distribution has a finite p-th moment.

A consequence of Theorem 3.5.7 is that the process Y is exponentially β-mixing. This will also be true for the log-price process as we state next.

Corollary 3.5.8. *If the conditions of Theorem 3.5.7 hold, then the log-price process $(\boldsymbol{G}_i)_{i \in \mathbb{N}}$ is exponentially β-mixing, and as a consequence also ergodic.*

Proof. Since Y is exponentially β-mixing, homogeneous strong Markov ([98, Theorem 4.4]), and driven only by the discrete part of the quadratic variation of L, the proof follows by the same arguments as that of Theorem 3.4 in Haug et al. [45]. □

For the consistency of the GMM estimator, we need to ensure that the stationary distribution of the volatility process has finite second moment. However, conditions (3.5.12) or (3.5.13) of Theorem 3.5.7 with $p = 2$ might be very restrictive. For example, condition (3.5.13) with $p > 1$ requires that L is a compound Poisson ([99, Remark 4.4]). Therefore, it is desirable to ensure the existence of second moments of the stationary distribution under less restrictive conditions. The next theorem takes care of this issue, giving consistency of the GMM estimator under rather weak conditions. Before we state it, we give the definition of asymptotic second-order stationarity which will be used in its proof. A stochastic process $X \in \mathbb{S}_d$ is said to be asymptotically second-order stationary with mean $\mu \in \mathbb{R}^{d^2}$, variance $\Sigma \in \mathbb{S}_{d^2}^+$ and autocovariance function $f : \mathbb{R}^+ \mapsto M_{d^2}(\mathbb{R})$ if it has finite second moments and

$$\begin{aligned} &\lim_{t\to\infty} E\left(X_t\right) = \mu, \qquad \lim_{t\to\infty} \operatorname{Var}\left(\operatorname{vec}\left(X_t\right)\right) = \Sigma \\ &\lim_{t\to\infty} \sup_{h\in\mathbb{R}^+} \left\{\left\|\operatorname{Cov}\left(\operatorname{vec}\left(X_{t+h}\right), \operatorname{vec}\left(X_t\right)\right) - f(h)\right\|\right\} = 0. \end{aligned} \tag{3.5.14}$$

Theorem 3.5.9. *Assume that the matrices $B, \mathcal{B}, \mathcal{C}$ are such that $\sigma(B), \sigma(\mathcal{B}), \sigma(\mathcal{C}) \subset (-\infty, 0) + i\mathbb{R}$, that the MUCOGARCH volatility process is μ-irreducible with support having non-empty interior and aperiodic and that there exists an $\Xi \in \mathbb{S}_d^{++}$ such that $\Xi B + B^\top \Xi + \sigma_L A^\top \Xi A \in -\mathbb{S}_d^{++}$. If additionally,* $(d.1)$, $(d.3)$, $(d.4)$ *and* $(d.5)$ *hold, then the GMM estimator defined in* (3.5.3) *is weakly consistent.*

Proof. Let $D \in \mathbb{S}_d^+$ be a constant matrix, and let the MUCOGARCH process $(Y_t)_{t\in\mathbb{R}^+}$ have starting value D. Then, a combination of assumptions $(a.5)$, $(a.6)$ with the fact that the starting value D is non-random and the hypothesis imposed on the matrices $B, \mathcal{B}, \mathcal{C}$ allow us to apply Theorem 4.20(ii) in Stelzer [98] to conclude that the process $(Y_t)_{t\in\mathbb{R}^+}$ is asymptotically second-order stationary. Additionally, Theorem 3.5.7(i) ensures that the process $(Y_t)_{t\in\mathbb{R}^+}$ has a unique stationary distribution, is geometrically ergodic and its stationary distribution has finite first moment, i.e., $\mathbb{E}\|Y_0\| < \infty$. Since

$$\mathbb{E}\|Y_t\|^2 \leq \operatorname{Var}(Y_t) + (\mathbb{E}\|Y_t\|)^2, \quad t > 0, \tag{3.5.15}$$

and both maps $t \mapsto \mathbb{E}\|Y_t\|$ and $t \mapsto \operatorname{Var}(Y_t)$ are continuous ([98, eqs. (4.7) and (4.16)]), it follows from (3.5.15) that $\limsup_{t\geq 0} \mathbb{E}\|Y_t\|^2 < \infty$. Since Theorem 3.5.7(i) implies convergence of the transition probabilities in total variation, which in turns implies weak convergence (e.g. Klenke [54, Exercise 13.2.2]), we have that $Y_t \xrightarrow{d} Y_0$ as $t \to \infty$. Hence, we can use the continuous mapping theorem and Billingsley [9, Theorem 25.11] to conclude that $\mathbb{E}\|Y_0\|^2 < \infty$. Finally, the result follows by an application of Lemma 3.5.1, Corollary 3.5.8 and Theorem 3.5.3. □

Next, we state a result which gives sufficient conditions for the irreducibility of the MUCOGARCH process, which is one of the necessary conditions for the geometric ergodicity result from Theorem 3.5.7.

Theorem 3.5.10. *[Irreducibility and Aperiodicity 1 - [98, Theorem 5.1]] Let Y be a MUCOGARCH volatility process driven by a compound Poisson process L with $A \in GL_d(\mathbb{R})$ and $\Re(\sigma(B)) < 0$. If the jump distribution of L has a non-trivial absolutely continuous component w.r.t. to the Lebesgue measure on $\mathbb{R}^d$ restricted to an open neighborhood of zero, then Y in (3.3.3) is irreducible w.r.t. the Lebesgue measure restricted to an open neighborhood of zero in $\mathbb{S}_d^+$ and aperiodic.*

For example, if L is a compound Poisson process with jump distribution having a density, then it satisfies the conditions of Theorem 3.5.11. The conditions on the jump distribution of the compound Poisson process can also be relaxed.

Theorem 3.5.11. *[Irreducibility and Aperiodicity 2 - [98, Corollary 5.2]] Let Y be a MUCOGARCH volatility process driven by a compound Poisson process L with $A \in GL_d(\mathbb{R})$ and $\Re(\sigma(B)) < 0$. If the jump distribution of L has a non-trivial absolutely continuous component equivalent to the Lebesgue measure on $\mathbb{R}^d$ restricted to an open neighborhood of zero, then Y in (3.3.3) is irreducible w.r.t. the Lebesgue measure restricted to an open neighborhood of zero in $\mathbb{S}_d^+$ and aperiodic.*

Recall that for the asymptotic normality result, we need to ensure that the stationarity distribution of the MUCOGARCH volatility process has more than 4 moments ($(e.2)$) and that it is also strongly mixing ($(d.6)$). For $p > 1$, the conditions of Theorem 3.5.7 are in general stronger than the one appearing in [98, Theorem 4.5] for the existence of a stationary distribution and of its p-th moments (see Remark 4.9 in [99]). Thus, if the matrix $B \in M_d(\mathbb{R})$ is diagonalizable with $S \in GL_d(\mathbb{C})$ such that $S^{-1}BS$ is diagonal, it is better to use Theorem 3.5.7 for $p = 1$ to ensure mixing and [98, Theorem 4.5] to ensure the finiteness of $p > 4$ moments.

Theorem 3.5.12. *([98, Theorem 4.5]) Assume the conditions of Theorem 3.5.7, that $B \in M_d(\mathbb{R})$ is diagonalizable with $S \in GL_d(\mathbb{C})$ such that $S^{-1}BS$ is diagonal and that*

$$\int_{\mathbb{R}^d} ((1 + \alpha_1 \| \operatorname{vec}(yy^*)\|_{B,S})^p - 1)\nu_L(\mathrm{d}y) < -2\lambda p, \quad p > 1, \tag{3.5.16}$$

where

$$\|X\|_{B,S} = \|(S^{-1} \otimes S^{-1})X(S \otimes S))\|_2, \quad X \in M_{d^2}(\mathbb{R}), \tag{3.5.17}$$

$\|x\|_{B,S} = \|(S^{-1} \otimes S^{-1})x\|_2$ *on* $\mathbb{R}^{d^2}$, $\lambda = \max(\Re(\sigma(B)))$,
$\alpha_1 = \|S\|_2^2 \|S^{-1}\|_2^2 K_{2,B} \|A \otimes A\|_{B,S}$ *and* $K_{2,B} = \max_{X \in \mathbb{S}_d^+, \|X\|_2 = 1} \left(\frac{\|X\|_2}{\| \operatorname{vec}(X)\|_{B,S}} \right)$. *Then* $\mathbb{E}\|Y_0\|^p < \infty$.

Proof. The assertion follows by the same arguments of [98, Theorem 4.5] combined with Lindner and Maller [65, Proposition 4.1]. □

Since (3.5.16) with $p > 4$ implies condition (3.5.11), it suffices to check (3.5.16) with $p > 4$ in order to ensure that assumptions $(d.6)$ and $(e.2)$, required for the asymptotic normality of the GMM estimator are valid.

3.5.2 Sufficient conditions for identifiability

We start with the identifiability of the matrix C_θ.

Lemma 3.5.13. *Assume the conditions of Lemma 3.4.2 and that* $\sigma(B_\theta) \subset (-\infty, 0) + i\mathbb{R}$. *If the matrices* A_θ *and* B_θ *are known, then* $\mathbb{E}_\theta(\boldsymbol{G}_1\boldsymbol{G}_1^*)$ *uniquely determines* C_θ.

Proof. Since $\sigma(B_\theta \otimes I + I \otimes B_\theta) = \sigma(B_\theta) + \sigma(B_\theta) \subset (-\infty, 0) + i\mathbb{R}$, the matrix $B_\theta \otimes I + I \otimes B_\theta$ is invertible. The rest of the proof follows by noticing that from (3.4.3) and (3.4.6) it follows that

$$\operatorname{vec}(C_\theta) = (\sigma_L + \sigma_W)^{-1}\Delta^{-1}(B_\theta \otimes I + I \otimes B_\theta)^{-1}\mathcal{B}_\theta\mathbb{E}_\theta(\boldsymbol{G}_1\boldsymbol{G}_1^*). \tag{3.5.18}$$

□

For the identification of the matrices A_θ and B_θ we need to use the autocovariance and variance of the squared returns process in Lemmas 3.4.2 and 3.4.3, respectively. We first state two auxiliary results, which provide conditions such that we can identify the components of the autocovariance function in (3.4.7).

Lemma 3.5.14. *Assume that* $B \in M_d(\mathbb{R})$ *is diagonalizable with* $S \in GL_d(\mathbb{C})$ *such that* $S^{-1}BS$ *is diagonal. If*

$$\left|\frac{\sigma_L - \sigma_W}{2\sigma_L}\right| \|A \otimes A\|_{B,S} < -2\max\{\Re(\sigma(B))\}, \tag{3.5.19}$$

with $\|\cdot\|_{B,S}$ *as in* (3.5.17) *then the matrix*

$$(\sigma_W + \sigma_L)(\mathcal{B}^*)^{-1} - 2((A \otimes A)^*)^{-1}$$

is invertible.

Proof. From [5, fact 2.16.14], $X^{-1} + Y^{-1}$ is non-singular if and only if $X + Y$ is non-singular and X, Y are non-singular. Setting $X = \frac{\mathcal{B}}{(\sigma_L + \sigma_W)}$, $Y = -\frac{1}{2}(A \otimes A)$ and using the definition of $\mathcal{B}$ in (3.4.1) we get

$$X + Y = \frac{1}{(\sigma_L + \sigma_W)}\Big((B \otimes I + I \otimes B) + \frac{(\sigma_L - \sigma_W)}{2}(A \otimes A)\Big).$$

Since B is diagonlizable, we can use Bernstein [5, Proposition 7.1.6] to obtain

$$B \otimes I + I \otimes B = (S \otimes S)(S^{-1}BS \otimes I)(S^{-1} \otimes S^{-1}),$$

which guarantees that $B \otimes I + I \otimes B$ is also diagonlizable. Now we rewrite the first equation on p. 106 in [98] with the matrix $\mathcal{B}$ replaced by $(B \otimes I + I \otimes B) + \frac{(\sigma_L - \sigma_W)}{2}(A \otimes A)$ and apply the Bauer-Fikes Theorem ([46, Theorem 6.3.2]) to see that (3.5.19) implies that all eigenvalues of $(X+Y)(\sigma_L+\sigma_W)$ are in $(-\infty, 0)+i\mathbb{R}$ and, therefore, $X+Y$ is invertible. □

Lemma 3.5.15. *If $A \in M_d(\mathbb{R})$ is such that $A_{(1,j)} > 0$ for some $j \in \{1, \dots, d\}$, then the map $X \mapsto AXA^T$ for $X \in \mathbb{S}_d$ identifies A.*

Proof. Assume first that $A_{(1,1)} > 0$. For each $i \in \{1, \dots, d\}$, let e_i be the ith column unit vector in $\mathbb{R}^d$ and define the matrix $E^{(i,j)} = e_i e_j^T$. The first line of the matrix $AE^{(1,1)}A^T$ is

$$(A^2_{(1,1)}, A_{(1,1)}A_{(2,1)}, \dots, A_{(1,1)}A_{(d,1)}). \tag{3.5.20}$$

Since $A_{(1,1)} > 0$, (3.5.20) allows us to identify first $A_{(1,1)}$ and then $A_{(2,1)}, \dots, A_{(d,1)}$. Now, for each $k \in \{2, \dots, d\}$, notice that $E^{(1,k)} + E^{(k,1)}$ is symmetric. Simple calculations reveal that the first line of the matrix $A(E^{(1,k)} + E^{(k,1)})A^T$ is

$$(2A_{(1,1)}A_{(1,k)}, A_{(1,1)}A_{(2,k)} + A_{(1,k)}A_{(2,1)}, \dots, A_{(1,1)}A_{(d,k)} + A_{(1,k)}A_{(d,1)}). \tag{3.5.21}$$

Since $A_{(1,1)} > 0$, we identify $A_{(1,k)}$ from the first entry of (3.5.21). Now, since also $A_{(2,1)}, \dots, A_{(d,1)}$ are known, we can identify $A_{(2,k)}, \dots, A_{(d,k)}$. Thus, all entries of A can be identified. The proof for the cases $A_{(1,j)} > 0$ for some $j > 1$ is achieved by the same arguments and replacing $AE^{(1,1)}A^T$ by $AE^{(j,j)}A^T$ and $A(E^{(1,k)} + E^{(k,1)})A^T$ by $A(E^{(j,k)} + E^{(k,j)})A^T$ for $k \neq j$, respectively. □

Lemma 3.5.16. *Assume the conditions of Lemma 3.5.14 and, furthermore that for all $\theta \in \Theta$, $\sigma(\mathcal{B}_\theta) \subset \{z \in \mathbb{C} : -\pi < \Im(z)\Delta < \pi, \Re(z) < 0\}$ and that $\mathrm{Var}_\theta(\mathrm{vech}(V_0))$ is invertible. Based on (3.4.7), write $\mathrm{acov}_{\theta,\boldsymbol{GG}^*}(h) = e^{\mathcal{B}_\theta \Delta h} M_\theta$ for $h \in \mathbb{N}$, where $M_\theta = (e^{\mathcal{B}\Delta})^{-1}\mathrm{acov}_{\theta,\boldsymbol{GG}^*}(1)$. Then, $\mathrm{acov}_{\theta,\boldsymbol{GG}^*}(1)$ and $\mathrm{acov}_{\theta,\boldsymbol{GG}^*}(2)$ uniquely identify $\mathcal{B}_\theta$ and M_θ.*

Proof. Since M_θ is given in terms $\mathcal{B}_\theta$ and $\mathrm{acov}_{\theta,\boldsymbol{GG}^*}(1)$, we only need to identify $\mathcal{B}_\theta$. Observe that we are using the vec operator only for convenience, as it interacts nicely with tensor products of matrices and thus gives nicely looking formulae. However, the volatility and "squared returns" processes takes values in $\mathbb{S}_d$ which is a $d(d+1)/2$-dimensional vector space, whereas the vec operator assumes values in a d^2-dimensional vector space. Instead of using the vech operator and cumbersome notation, we take an abstract point of view. The variance of a random element of $\mathbb{S}_d$ is a symmetric positive semi-definite linear operator from $\mathbb{S}_d$ to itself. Likewise, the autocovariance of $\boldsymbol{G}_1\boldsymbol{G}_1^*$ and $\boldsymbol{G}_{1+h}\boldsymbol{G}_{1+h}^*$ is a linear operator from $\mathbb{S}_d$ to itself. The condition that $\mathrm{Var}_\theta(\mathrm{vech}(V_0))$ is invertible is equivalent to the invertibility of the linear operator, which is the variance of V_0. Similarly all other $d^2 \times d^2$ matrices in

$$\begin{aligned} &e^{\mathcal{B}\Delta h}\mathcal{B}^{-1}(I_{d^2} - e^{-\mathcal{B}\Delta})(\sigma_L + \sigma_W)\mathrm{Var}(\mathrm{vec}(V_0))(e^{\mathcal{B}^*\Delta} - I_{d^2}) \\ &\quad\times [(\sigma_W + \sigma_L)(\mathcal{B}^*)^{-1} - 2((A\otimes A)^*)^{-1}] \end{aligned} \tag{3.5.22}$$

are representing linear operators from $\mathbb{S}_d$ to itself. Under the assumptions made, the above product involves only invertible linear operators. Hence $\mathrm{acov}_{\theta,\boldsymbol{GG}^*}(h)$ is invertible (over $\mathbb{S}_d$) for every $h > 0$. Thus,

$$e^{\mathcal{B}\Delta} = \mathrm{acov}_{\theta,\boldsymbol{GG}^*}(2)[\mathrm{acov}_{\theta,\boldsymbol{GG}^*}(1)]^{-1}. \tag{3.5.23}$$

By the assumptions on the eigenvalues of $\mathcal{B}_\theta$ there is a unique logarithm for $e^{\mathcal{B}_\theta\Delta}$ (see [46, Section 6.4] or [92, Lemma 3.11]), so $\mathcal{B}_\theta\Delta$ and thus $\mathcal{B}_\theta$ is identified. Finally, note that the matrices in the vec representations are uniquely identified by the employed linear operators on $\mathbb{S}_d$ due to Pigorsch and Stelzer [81, Proposition 3.1] and Lemma 3.5.15.

□

Lemma 3.5.17. *(Identifiability of A_θ, B_θ and C_θ) Assume the conditions of Lemma 3.5.14 and that the entries of the matrices A_θ and B_θ satisfy: $A_{(1,d),\theta} > 0$, $A_{(1,2),\theta} \neq A_{(2,1),\theta}$ and $B_{(1,2),\theta} = B_{(2,1),\theta}$. Then the map $\theta \mapsto k_{\theta,2}$ in* (3.5.2) *is injective.*

Proof. For the sake of clarity we omit θ in the notation and assume wlog that $\sigma_L = 1$. Because of Lemma 3.5.13, we only need to show the identification of A and B.
Assume first that $d = 2$. Then the 4×4-matrix $\mathcal{B}$ from (3.4.1) becomes

$$\begin{pmatrix} 2B_{(1,1)} + A_{(1,1)}^2 & B_{(1,2)} + A_{(1,1)}A_{(1,2)} & B_{(1,2)} + A_{(1,1)}A_{(1,2)} & A_{(1,2)}^2 \\ B_{(2,1)} + A_{(1,1)}A_{(2,1)} & B_{(1,1)} + B_{(2,2)} + A_{(1,1)}A_{(2,2)} & A_{(1,2)}A_{(2,1)} & B_{(1,2)} + A_{(1,2)}A_{(2,2)} \\ B_{(2,1)} + A_{(1,1)}A_{(2,1)} & A_{(1,2)}A_{(2,1)} & B_{(1,1)} + B_{(2,2)} + A_{(1,1)}A_{(2,2)} & B_{(1,2)} + A_{(1,2)}A_{(2,2)} \\ A_{(2,1)}^2 & B_{(2,1)} + A_{(2,1)}A_{(2,2)} & B_{(2,1)} + A_{(2,1)}A_{(2,2)} & 2B_{(2,2)} + A_{(2,2)}^2 \end{pmatrix}. \tag{3.5.24}$$

Using the entry at position $(1,4)$ and the fact that $A_{(1,2)} > 0$ allow us to identify $A_{(1,2)}$. Then, we use the entry at position $(2,3)$ to identify $A_{(2,1)}$. Now, we use the entries at positions $(1,2)$ and $(2,1)$ together with the fact that $A_{(1,2)} \neq A_{(2,1)}$ and $B_{(1,2)} = B_{(2,1)}$ to write $A_{(1,1)} = (\mathcal{B}_{(1,2)} - \mathcal{B}_{(2,1)})/(A_{(1,2)} - A_{(2,1)})$. Similarly we use the entries at positions $(3,4), (4,3)$ to get $A_{(2,2)} = (\mathcal{B}_{(3,4)} - \mathcal{B}_{(4,3)})/(A_{(1,2)} - A_{(2,1)})$. Now, since all the entries of A are known, we can use the entries at positions $(1,1), (1,2)$ and $(2,2)$ to identify the entries of B.
Now assume that $d > 2$. Write the matrix $\mathcal{B}_\theta$ from (3.4.1) in the following block form:

$$\mathcal{B}_\theta = B\otimes I + I\otimes B + A\otimes A = \begin{pmatrix} \mathcal{B}^{(1,1)} & \cdots & \mathcal{B}^{(1,d)} \\ \vdots & \ddots & \vdots \\ \mathcal{B}^{(d,1)} & \cdots & \mathcal{B}^{(d,d)} \end{pmatrix}, \tag{3.5.25}$$

where $\mathcal{B}^{(i,j)} \in M_d(\mathbb{R})$ for all $i, j = 1, \ldots, d$. First, we have that

$$\mathcal{B}^{(1,d)} =$$

$$\begin{pmatrix} B_{(1,d)} + A_{(1,d)}A_{(1,1)} & A_{(1,d)}A_{(1,2)} & A_{(1,d)}A_{(1,3)} & \cdots & A_{(1,d)}A_{(1,d)} \\ A_{(1,d)}A_{(2,1)} & B_{(1,d)} + A_{(1,d)}A_{(2,2)} & A_{(1,d)}A_{(2,3)} & \cdots & A_{(1,d)}A_{(1,d)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{(1,d)}A_{(d,1)} & A_{(1,d)}A_{(d,2)} & A_{(1,d)}A_{(d,3)} & \cdots & B_{(1,d)} + A_{(1,d)}A_{(d,d)} \end{pmatrix}, \tag{3.5.26}$$

Since $A_{(1,d)} > 0$ we can identify it from (3.5.26), because $\mathcal{B}^{(1,d)}_{(1,d)} = A^2_{(1,d)}$. Then we use the off-diagonal entries of the matrix $\mathcal{B}^{(1,d)}$ in (3.5.26) together with $A_{(1,d)}$ to identify all the off-diagonal entries of the matrix A. Next we identify the diagonal entries of A. It follows from (3.5.25) that

$$\begin{cases} \mathcal{B}^{(k,k)}_{(1,2)} = B_{(1,2)} + A_{(k,k)}A_{(1,2)} \\ \mathcal{B}^{(k,k)}_{(2,1)} = B_{(2,1)} + A_{(k,k)}A_{(2,1)} \end{cases}, \quad k = 1, \ldots, d. \tag{3.5.27}$$

Since $A_{(1,2)} - A_{(2,1)} \neq 0$ and $B_{(1,2)} = B_{(2,1)}$, the system of equations (3.5.27) gives

$$A_{(k,k)} = (\mathcal{B}^{(k,k)}_{(1,2)} - \mathcal{B}^{(k,k)}_{(2,1)})/(A_{(1,2)} - A_{(2,1)}), \quad k = 1, \ldots, d.$$

Finally, since the matrix A is now completely known, we can use (3.5.25) to identify all entries of B. □

In Lemma 3.5.17 we identify the matrices A_θ and B_θ only from $\mathcal{B}_\theta$ and, therefore, some restrictions on the off-diagonal entries of B_θ appear. In order to avoid those restrictions, we take the structure of $\mathbb{E}_\theta \operatorname{vec}(\operatorname{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*) \operatorname{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*)^*)$ in (3.4.8) into account. Already in the 2-dimensional case the identification conditions are quite involved, and this has mainly to do with the fact that the linear operator $(\boldsymbol{Q} + K_d\boldsymbol{Q} + I_{d^2})$ at the right hand side of (3.4.8) is not one-to-one in the space of matrices of the form $\mathbb{E}_\theta \operatorname{vec}(V_0) \operatorname{vec}(V_0)^*$. We state this in the next lemma.

Lemma 3.5.18. *(Identifiability for $d = 2$ with B_θ not necessarily symmetric). Assume the conditions of Lemma 3.5.14. Let $Z := (z_{\theta,ij})_{i,j=1}^4 := (\sigma_W + \sigma_L)(\mathcal{B}_\theta^*)^{-1}$ and*

$$\operatorname{Var}_\theta(\operatorname{vec}(V_0)) = \begin{pmatrix} \tau_{\theta,1} & \tau_{\theta,2} & \tau_{\theta,2} & \tau_{\theta,3} \\ \tau_{\theta,2} & \tau_{\theta,4} & \tau_{\theta,4} & \tau_{\theta,5} \\ \tau_{\theta,1} & \tau_{\theta,4} & \tau_{\theta,4} & \tau_{\theta,5} \\ \tau_{\theta,3} & \tau_{\theta,5} & \tau_{\theta,5} & \tau_{\theta,6} \end{pmatrix}. \tag{3.5.28}$$

Let $\theta = (\theta_1, \ldots, \theta_{10})$*,* $A_\theta = \begin{pmatrix} \theta_1 & 0 \\ 0 & \theta_2 \end{pmatrix}$ *with* $\theta_1 > 0$*,* $\theta_2 \neq 0$*,* $B_\theta = \begin{pmatrix} \theta_3 & \theta_4 \\ \theta_5 & \theta_6 \end{pmatrix}$ *and* $C_\theta = \begin{pmatrix} \theta_7 & \theta_8 \\ \theta_9 & \theta_{10} \end{pmatrix}$*. Assume that* $B_\theta \otimes I + I \otimes B_\theta$ *and* $A_\theta \otimes A_\theta$ *commute and that either*

$$\begin{pmatrix} -\tau_{\theta,1} & -2z_{\theta,41} \\ -\tau_{\theta,2} & z_{\theta,21} + z_{\theta,31} \end{pmatrix} \quad \textit{is invertible and } z_{\theta,24} + z_{\theta,34} \neq 0, \tag{3.5.29}$$

or

$$\begin{pmatrix} -\tau_{\theta,5} & z_{\theta,24} + z_{\theta,34} \\ -\tau_{\theta,6} & z_{\theta,14} \end{pmatrix} \quad \textit{is invertible and } z_{\theta,41} \neq 0. \tag{3.5.30}$$

Then the map $\theta \mapsto k_{\theta,2}$ *in* (3.5.2) *is injective.*

Proof. Since $(B_\theta \otimes I + I \otimes B_\theta)$ and $(A_\theta \otimes A_\theta)$ commute, it follows that $(e^{\mathcal{B}_\theta^* \Delta} - I_4)$ and $(\sigma_W + \sigma_L)(\mathcal{B}_\theta^*)^{-1} - 2((A_\theta \otimes A_\theta)^*)^{-1}$ commute. From Lemma 3.5.16 we can identify $\mathcal{B}_\theta$ and

$$\mathcal{B}_\theta^{-1}(I_4 - e^{-\mathcal{B}_\theta \Delta})(\sigma_L + \sigma_W)\operatorname{Var}_\theta(\operatorname{vec}(V_0))(e^{\mathcal{B}_\theta^* \Delta} - I_4)[(\sigma_W + \sigma_L)(\mathcal{B}^*)^{-1} - 2((A_\theta \otimes A_\theta)^*)^{-1}]$$

from $k_{\theta,2}$. Using the invertibility of $\mathcal{B}_\theta$ we identify

$$\operatorname{Var}_\theta(\operatorname{vec}(V_0))[(\sigma_W + \sigma_L)(\mathcal{B}_\theta^*)^{-1} - 2((A_\theta \otimes A_\theta)^*)^{-1}]. \tag{3.5.31}$$

Plugging (3.5.31) into $\mathbb{E}_\theta \operatorname{vec}(\boldsymbol{G}_1 \boldsymbol{G}_1^*) \operatorname{vec}(\boldsymbol{G}_1 \boldsymbol{G}_1^*)^*$ as defined in (3.4.8) allows us to identify

$$(\boldsymbol{Q} + K_d \boldsymbol{Q} + I_4)\mathbb{E}_\theta \operatorname{vec}(V_0) \operatorname{vec}(V_0)^*. \tag{3.5.32}$$

Using the definition of the linear map $\boldsymbol{Q}$ and the matrix commutation matrix K_2, and writing $\operatorname{vec}(V_0) = (V_{0,1}\, V_{0,2}\, V_{0,2}\, V_{0,3})^T$, gives:

$$\begin{aligned}
&(\boldsymbol{Q} + K_2 \boldsymbol{Q} + I_4)\mathbb{E}_\theta \operatorname{vec}(V_0) \operatorname{vec}(V_0)^* \\
&= \mathbb{E}_\theta \begin{pmatrix} 3V_{0,1}^2 & 3V_{0,1}V_{0,2} & 3V_{0,1}V_{0,2} & 2V_{0,1}^2 + 3V_{0,1}V_{0,3} \\ 3V_{0,1}V_{0,2} & 2V_{0,1}^2 + 3V_{0,1}V_{0,3} & 2V_{0,1}^2 + 3V_{0,1}V_{0,3} & 3V_{0,2}V_{0,3} \\ 3V_{0,1}V_{0,2} & 2V_{0,1}^2 + 3V_{0,1}V_{0,3} & 2V_{0,1}^2 + 3V_{0,1}V_{0,3} & 3V_{0,2}V_{0,3} \\ 2V_{0,1}^2 + 3V_{0,1}V_{0,3} & 3V_{0,2}V_{0,3} & 3V_{0,2}V_{0,3} & 3V_{0,3}^2 \end{pmatrix}.
\end{aligned} \tag{3.5.33}$$

From (3.5.33) we identify, $\mathbb{E}_\theta V_{0,1}^2$, $\mathbb{E}_\theta V_{0,3}^2$, $\mathbb{E}_\theta V_{0,1}V_{0,2}$, $\mathbb{E}_\theta V_{0,2}V_{0,3}$ and $\mathbb{E}_\theta(2V_{0,1}^2 + 3V_{0,1}V_{0,3})$. Since $\mathbb{E}_\theta V_0$ is known from $\mathbb{E}_\theta \operatorname{vec}(\boldsymbol{G}_1 \boldsymbol{G}_1^*)$, then $\operatorname{Var}_\theta(\operatorname{vec}(V_0))$ is partially known. This, together with (3.5.28) allow us to identify $\tau_{\theta,1}, \tau_{\theta,2}, \tau_{\theta,5}, \tau_{\theta,6}$ and $\tau_{\theta,3} + 2\tau_{\theta,4} := \nu_\theta$. Thus, we can write $\tau_{\theta,3} = \nu_\theta - 2\tau_{\theta,4}$, and the only unknown in (3.5.28) is $\tau_{\theta,4}$. Write (3.5.31) as

$$\operatorname{Var}_\theta(\operatorname{vec}(V_0))[Z - \sqrt{2}(A_\theta^*)^{-1} \otimes \sqrt{2}(A_\theta^*)^{-1}] =: P, \tag{3.5.34}$$

with $P \in M_2(\mathbb{R})$, Z being known, $\mathrm{Var}_\theta(\mathrm{vec}(V_0))$ partially known and $\sqrt{2}(A_\theta^*)^{-1}$ unknown. Writing (3.5.34) as a linear system of equations with unknowns $\tau_{\theta,4}$, θ_1 and θ_2, allow us to identify them provided that one of the matrices in (3.5.29) or (3.5.30) are invertible. Since θ_1 and θ_2 are now known, the identification of $\theta_3, \ldots, \theta_6$ follows from (3.5.24) and

$$\mathcal{B}_\theta = \begin{pmatrix} 2\theta_3 + \theta_1^2 & \theta_4 & \theta_4 & 0 \\ \theta_5 & \theta_3 + \theta_6 + \theta_1\theta_2 & 0 & \theta_4 \\ \theta_5 & 0 & \theta_3 + \theta_6 + \theta_1\theta_2 & \theta_4 \\ 0 & \theta_5 & \theta_5 & 2\theta_6 + \theta_2^2 \end{pmatrix}. \tag{3.5.35}$$

Finally, the identification of $\theta_7, \ldots, \theta_{10}$ follows by an application of Lemma 3.5.13. □

Remark 3.5.19. *It is worth noticing that the proof of Lemma 3.5.18 could not have been achieved just by using the matrix $\mathcal{B}_\theta$, since then,* (3.5.35) *would give a system with only* 3 *equations and 4 unknowns. Additionally, the commutativity condition imposed on the matrices $B_\theta \otimes I + I \otimes B_\theta$ and $A_\theta \otimes A_\theta$ in Lemma 3.5.18 seem to be essential for the proof as they lead to a system of equations involving only the entries of A_θ.*

Since commutativity is a quite strong condition, we prefer to work with the class of MUCOGARCH processes, which are identifiable by Lemma 3.5.17. The exponential decay of the autocovariance function of the model is still quite flexible, because of the interplay between the matrices A_θ and B_θ (see (3.5.24), for instance).

In the next section, we investigate the finite sample performance of the estimators in a simulation study.

3.6 Simulation study

To assess the performance of the GMM estimator, we will focus on the MUCOGARCH model in dimension $d = 2$. We fix $L_t = L_t^{\mathfrak{d}} + B_t$ for $t \in \mathbb{R}^+$ where $L^{\mathfrak{d}}$ is a bivariate compound Poisson process (CPP) and B is a standard bivariate Brownian motion, independent of $L^{\mathfrak{d}}$. We choose $L^{\mathfrak{d}}$ as a CPP, since it allows to simulate the MUCOGARCH volatility process V exactly. Thus, we only need to approximate the Brownian part of the (log) price process G in (3.1.1), which is done by an Euler scheme. Setting $L^{\mathfrak{d}}$ as a CPP is not a very crucial restriction, since for an infinite activity Lévy process one would need to approximate it using only finitely many jumps. For example by using a CPP for the big jumps component of $L^{\mathfrak{d}}$ and an appropriate Brownian motion for its small jumps component (see Cohen and Rosinski [17]). In applications, a CPP has also been used in combination with the univariate COGARCH(1,1) process for modeling high frequency data (see Müller [76]). The jumps of $L^{\mathfrak{d}}$ are $N(0, 1/4 I_2)$, and its rate is 4, so that $\mathrm{Var}(L_1) = 2I_2$ and

$$\mathbb{E}[\mathrm{vec}([L, L^*]^{\mathfrak{d}}), \mathrm{vec}([L, L^*]^{\mathfrak{d}})^*]_1^{\mathfrak{d}} = 1/4(I_4 + K_2 + \mathrm{vec}(I_2)\,\mathrm{vec}(I_2)^*).$$

In this case, the chosen Lévy process L satisfies the hypothesis of Theorem 3.5.11 and all assumptions **a** from Section 3.4 (with $\sigma_L = \sigma_W = 1$ and $\rho_L = 1/4$).

Based on the identification Lemma 3.5.17, we assume that the model is parameterized with $\theta = (\theta^{(1)}, \ldots, \theta^{(11)})$, and the matrices A_θ, B_θ and C_θ are defined as:

$$A_\theta = \begin{pmatrix} \theta^{(1)} & \theta^{(2)} \\ \theta^{(3)} & \theta^{(4)} \end{pmatrix}, \quad B_\theta = \begin{pmatrix} \theta^{(5)} & \theta^{(6)} \\ \theta^{(6)} & \theta^{(7)} \end{pmatrix} \quad \text{and} \quad C_\theta = \begin{pmatrix} \theta^{(8)} & \theta^{(9)} \\ \theta^{(10)} & \theta^{(11)} \end{pmatrix}. \tag{3.6.1}$$

The data used for estimation is a sample of the log-price process $\boldsymbol{G} = (\boldsymbol{G}_i)_{i=1}^n$ as defined in (3.1.2) with true parameter value $\theta_0 \in \Theta \subset \mathbb{R}^{11}$ observed on a fixed grid of size $\Delta = 0.1$ (the grid size for the Euler approximation of the Gaussian part is 0.01).

We experiment with two different settings, namely:

Example 3.6.1. θ_0 *is such that*

$$A_{\theta_0} = \begin{pmatrix} 0.85 & 0.10 \\ -0.10 & 0.75 \end{pmatrix}, \quad B_{\theta_0} = \begin{pmatrix} -2.43 & 0.05 \\ 0.05 & -2.42 \end{pmatrix} \quad \textit{and} \quad C_{\theta_0} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1.5 \end{pmatrix}. \tag{3.6.2}$$

Example 3.6.2. θ_0 *is such that* A_{θ_0} *and* C_{θ_0} *are as in Example 3.6.1 and*

$$B_{\theta_0} = \frac{1}{4} \begin{pmatrix} -2.43 & 0.05 \\ 0.05 & -2.42 \end{pmatrix}. \tag{3.6.3}$$

In Example 3.6.1, θ_0 is chosen in such a way that the asymptotic normality of $\hat{\theta}_n$ can be verified. Then, In Example 3.6.2 we rescale B_{θ_0} from Example 3.6.1 in such a way that weak consistency is satisfied, but not asymptotic normality.

Due to the identifiability Lemma 3.5.17 we need to choose $H \geq 2$. For comparison purposes, we perform the estimation for each lag $H \in \{2, 5, 10\}$ and sample size $n \in \{1\,000, 5\,000, 10\,000, 20\,000, 50\,000, 100\,000\}$. The computations are performed with the `optim` routine in combination with the `Nelder-Mead` algorithm in R (R Core Team [87]). Initial values for the estimation were found by the `DEoptim` routine on a neighborhood around the true parameter θ_0. This algorithm implements a differential evolution algorithm (for more details see Mullen et al. [75]), and it is very useful for finding good initial values in optimization problems. We only consider estimators based on the identity matrix for the weight matrix Ω in (3.5.3). The results are based on 500 independent samples of MUCOGARCH returns.

In the following we report the finite sample results of the GMM for Examples 3.6.1 and 3.6.2.

3.6.1 Simulation results for Example 3.6.1

We can check numerically that C_{θ_0} is positive definite, A_{θ_0} is invertible and $\sigma(B_{\theta_0}), \sigma(\mathcal{B}_{\theta_0})$ and $\sigma(\mathcal{C}_{\theta_0}) \in (-\infty, 0) + i\mathbb{R}$. For $\Xi = I_2$ in Theorem 3.5.9, the eigenvalues of the matrix $B_{\theta_0} + B^*_{\theta_0} + \sigma_L A^*_{\theta_0} \Xi A_{\theta_0}$ are -4.067 and -4.328, so it is negative definite. Hence, we can apply Theorem 3.5.9 to conclude that the GMM estimator is weakly consistent. Now we use Theorem 3.5.12 to ensure asymptotic normality by checking (3.5.16) with $p > 4$. For our choice of θ_0 we have that B_{θ_0} is diagonalizable with $B_{\theta_0} = S_{\theta_0} D_{\theta_0} S^{-1}_{\theta_0}$, where

$$S_{\theta_0} = \begin{pmatrix} -0.671 & -0.741 \\ -0.741 & 0.671 \end{pmatrix} \quad \text{and} \quad D_{\theta_0} = \begin{pmatrix} -2.375 & 0 \\ 0 & -2.475 \end{pmatrix}. \tag{3.6.4}$$

In addition, for $p = 4.001$,

$$\int_{\mathbb{R}^2} ((1 + \alpha_1 \| \operatorname{vec}(yy^*) \|_{B_{\theta_0}, S_{\theta_0}})^p - 1)\nu_L(\mathrm{d}y) + 2\lambda p = -0.024 < 0, \tag{3.6.5}$$

where α_1, λ and the norm $\| \cdot \|_{B,S}$ is defined in Theorem 3.5.12. Therefore, also $(e.2)$ is satisfied and the GMM estimator is asymptotically normal. We also note that the chosen matrix B_{θ_0} is very close to not satisfying assumption (3.6.5). For instance, if we set B_{θ_0} to $\begin{pmatrix} -2.4 & 0.05 \\ 0.05 & -2.4 \end{pmatrix}$, then the integral at the left hand side of (3.6.5) becomes 0.164 and, therefore, one could not ensure the existence of more than 8 moments of the stationary distribution of the price process, required for asymptotic normality. Of course, in this case we would still have a weakly consistent estimator.

We investigate the behavior of the bias and standard deviation in Figures 3.1 and 3.2, where we excluded those paths for which the algorithm did not converged successfully (around 10 percent of the paths for $n = 1\,000$ and less than 3 percent for larger n). Figures 3.1 and 3.2 show the absolute values of the bias and standard deviation for different lags H and varying n. As expected, they decay when n increases. Additionally, the results favor the choice of $H = 10$. It is also worth noticing that the estimation of the parameters in the matrix B_{θ_0} is more difficult than the other parameters, specially for $n \in \{1\,000, 5\,000, 10\,000\}$.

Figures 3.3 and 3.4 assess asymptotic normality though normal QQ-plots. Based on the previous findings we fix $H = 10$, since it gave the best results. This might have to do with the fact that using just a few lags for the autocovariance function ($H = 2$ or $H = 5$) was not sufficient for a good fit. We also restrict ourselves to $n \in \{5\,000, 20\,000, 100\,000\}$, since they already allow us to confirm the convergence to the normal distribution. Here we do not exclude those paths for which the algorithm did not converge (these are denoted by large red points in the normal QQ-plots in Figures 3.3 and 3.4). These plots clearly indicate asymptotic normality of the estimators. It is worth noting that the tails corresponding to the estimates of B_{θ_0} deviate from the ones of a normal distribution for values of $n \in \{5\,000, 20\,000\}$, but they get closer to a normal distribution for $n = 100\,000$. The left tail of the plots for $A_{21,\hat{\theta}_n}$ in Figure 3.3 is not close to a normal (although the plots show it is converging). This is due to identifiability condition in Lemma 3.5.17 which requires $A_{21,\theta} > 0$ but $A_{21,\theta_0} = 0.1$ is very close to the boundary. For $n = 5\,000$, there are very large negative outliers for the estimates of the diagonal entries of B_{θ_0}, which affects the bias substantially.

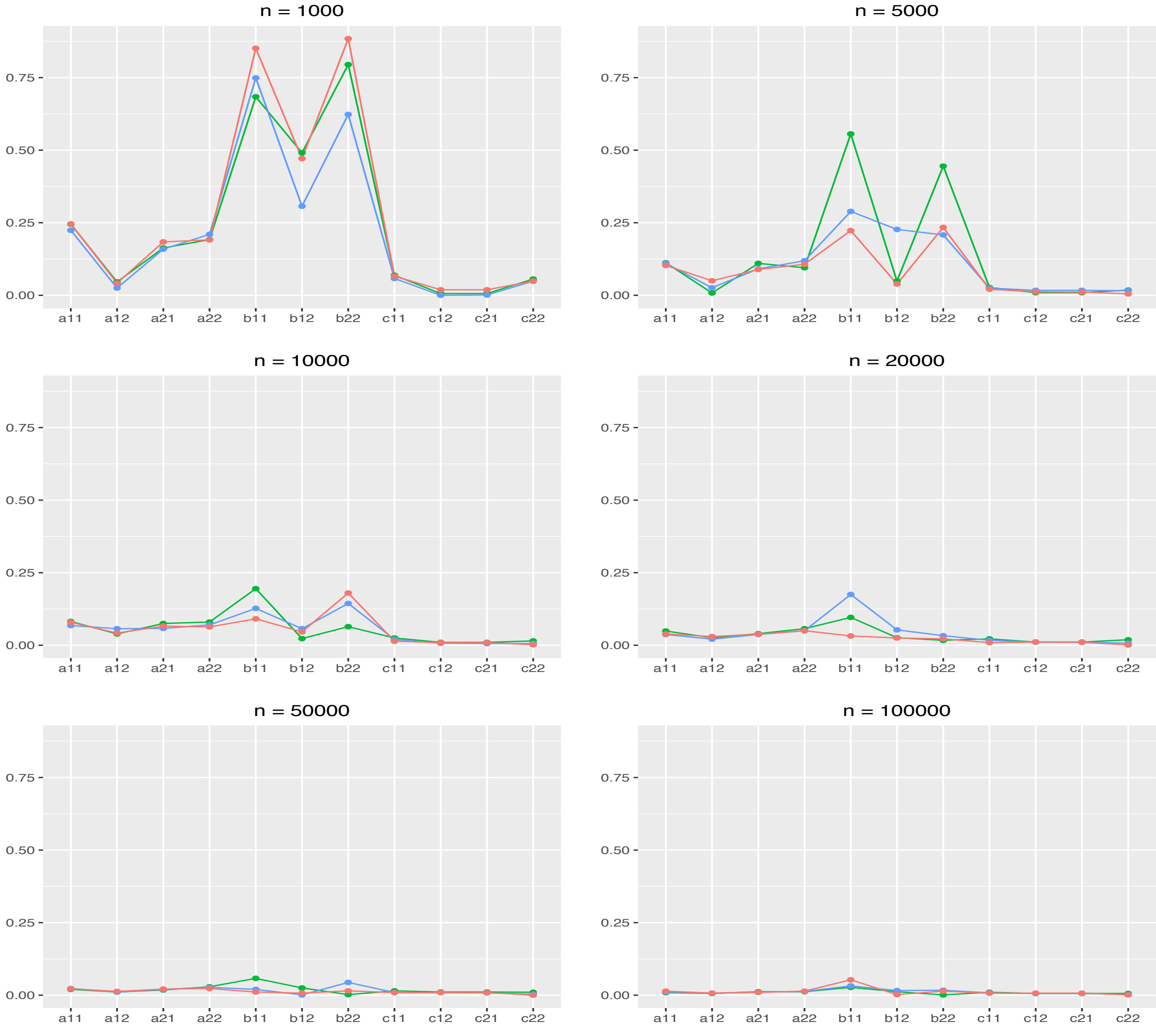


Figure 3.1: Absolute values of the bias of $\hat{\theta}_{n,H}$. The colors green, blue and red correspond to $H = 2, 5$ and 10, respectively.

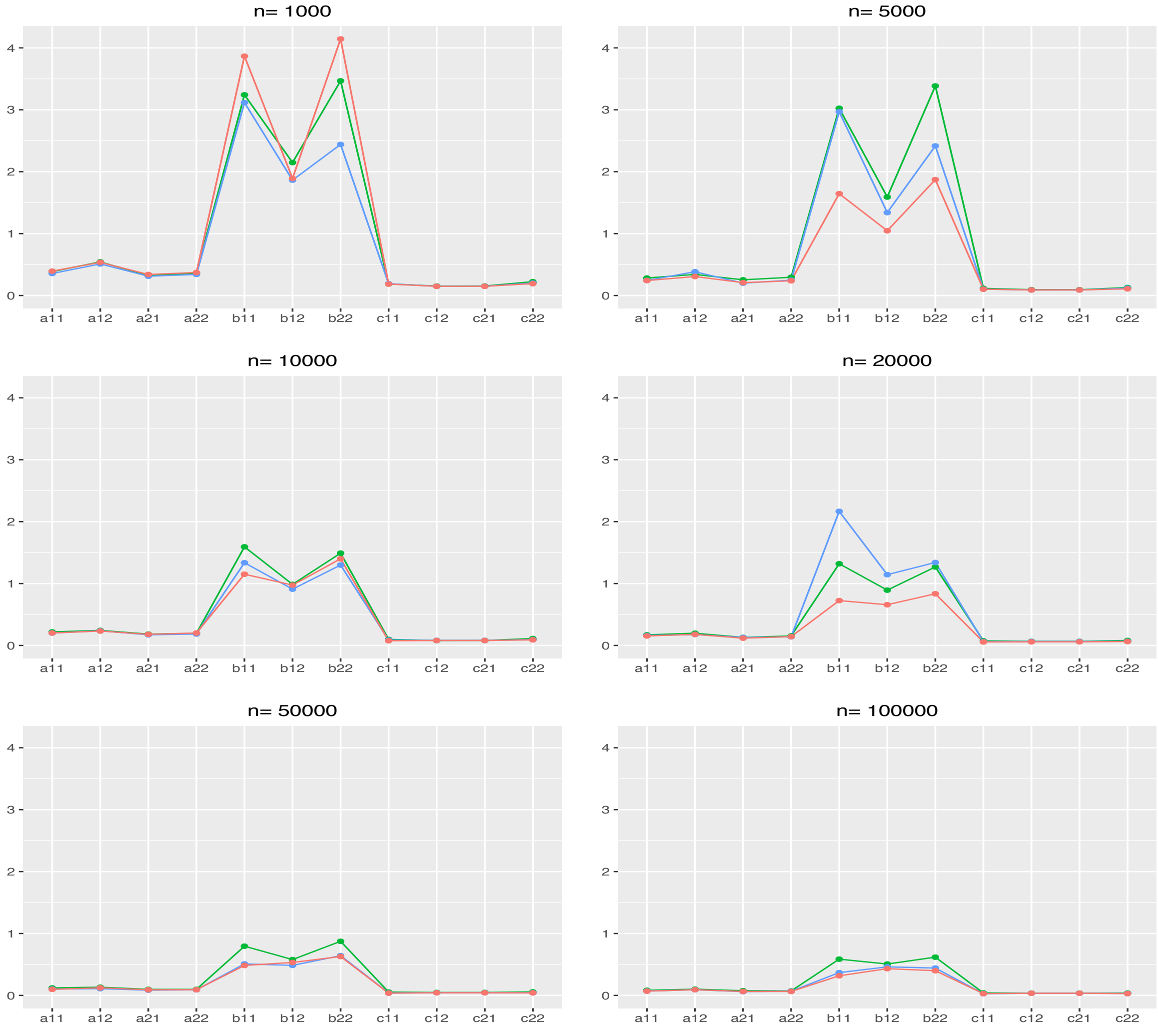


Figure 3.2: Standard deviation (std) of $\hat{\theta}_{n,H}$. The colors green, blue and red correspond to $H = 2, 5$ and 10, respectively.

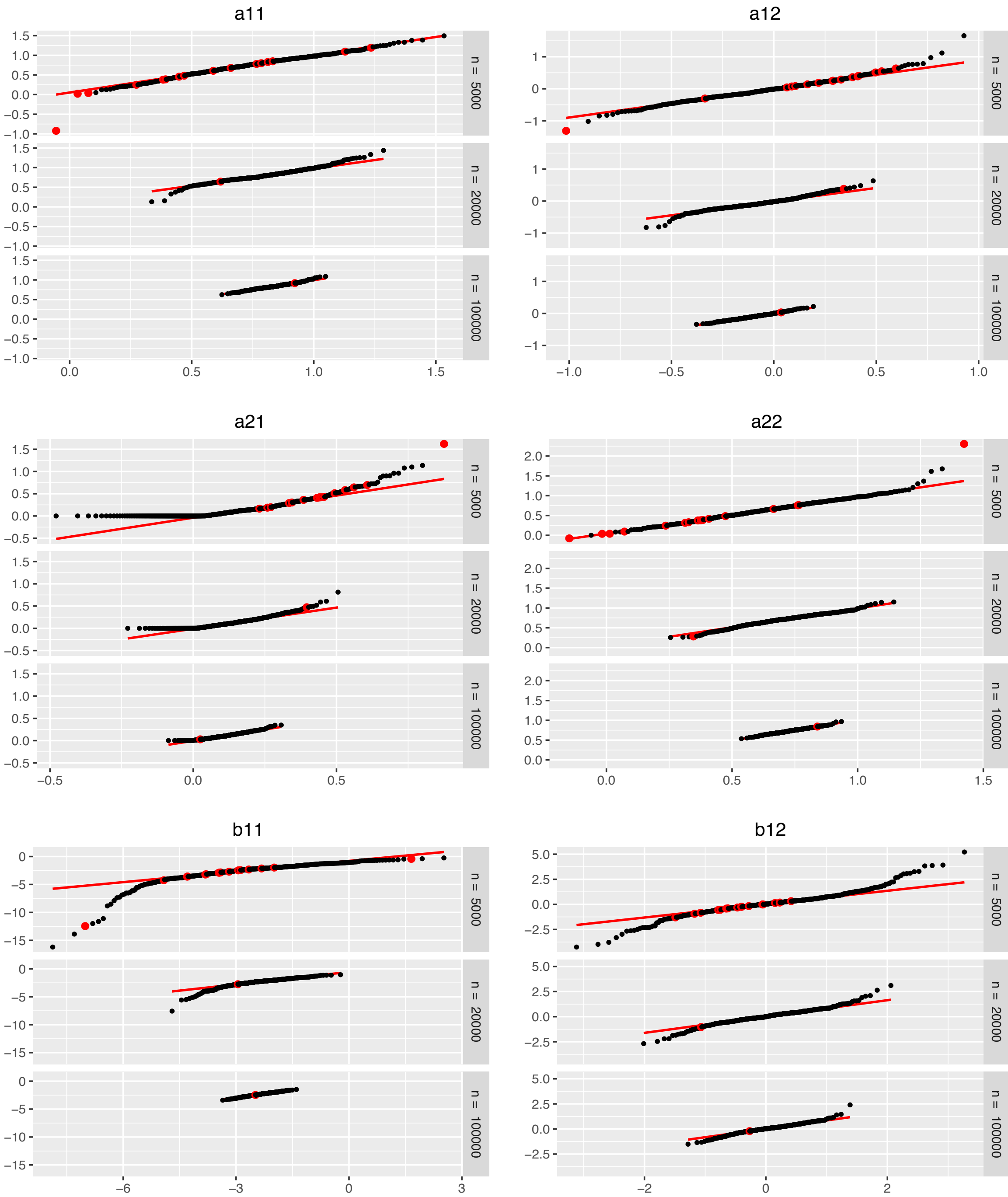


Figure 3.3: Normal QQ-plots of $\hat{\theta}_{n,10}$ for θ_0 as in (3.6.3). The red dots are values for which the algorithm did not converged.

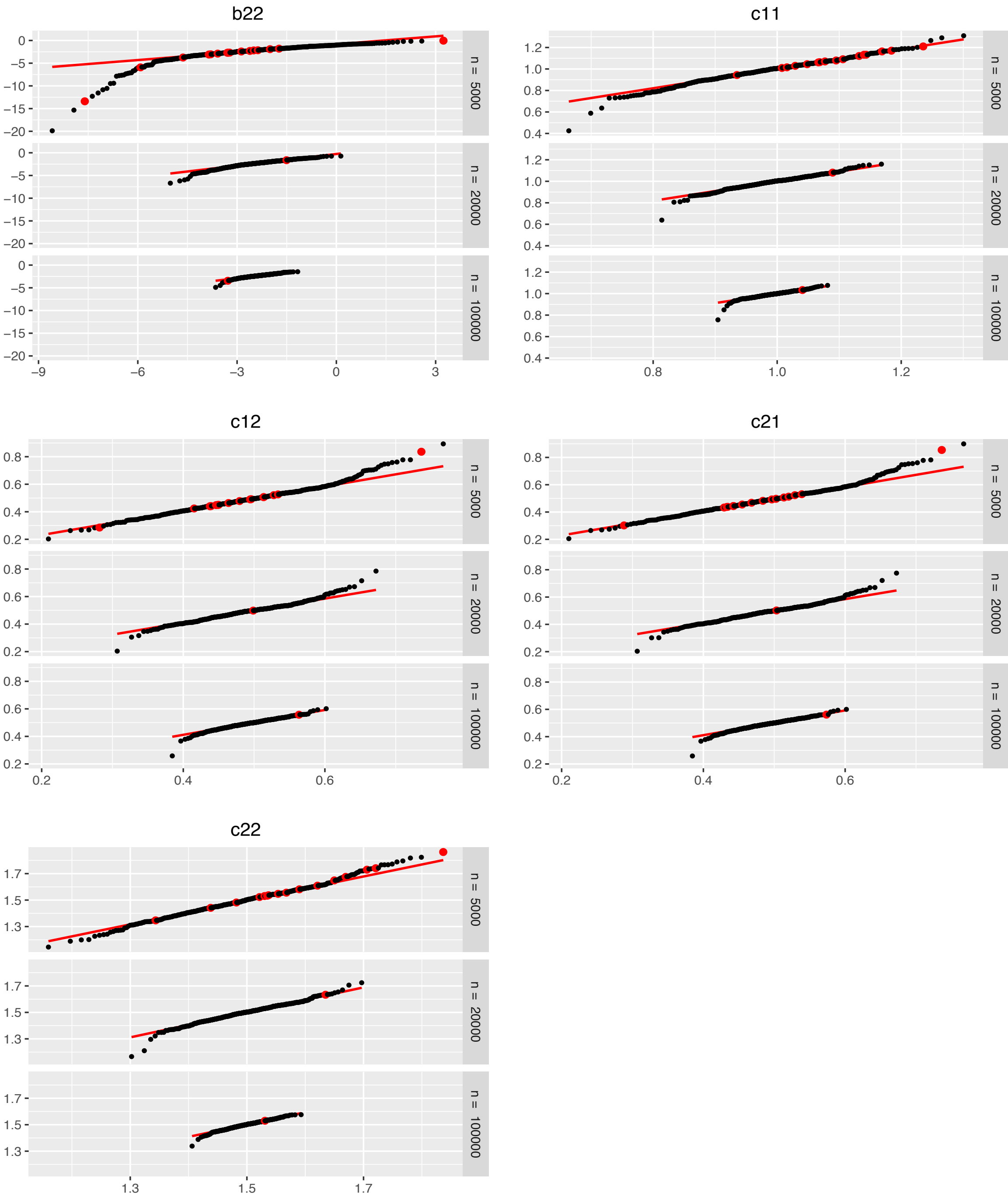


Figure 3.4: Normal QQ-plots of $\hat{\theta}_{n,10}$ for θ_0 as in (3.6.3). The red dots are values for which the algorithm did not converged.

Bias

	$n = 1000$	$n = 5000$	$n = 10000$	$n = 20000$	$n = 50000$	$n = 100000$
$\theta^{(1)}$	-0.112	-0.050	-0.030	-0.020	-0.011	-0.003
$\theta^{(2)}$	0.028	0.035	0.038	0.039	0.036	0.035
$\theta^{(3)}$	0.050	-0.005	-0.012	-0.017	-0.032	-0.036
$\theta^{(4)}$	-0.124	-0.058	-0.052	-0.034	-0.026	-0.024
$\theta^{(5)}$	-0.054	0.002	-0.005	-0.006	-0.001	-0.003
$\theta^{(6)}$	0.011	0.002	-0.004	-0.010	-0.007	-0.007
$\theta^{(7)}$	-0.109	-0.072	-0.038	-0.039	-0.022	-0.018
$\theta^{(8)}$	0.046	0.025	0.018	0.013	0.022	0.013
$\theta^{(9)}$	0.037	0.018	0.018	0.025	0.023	0.022
$\theta^{(10)}$	0.034	0.017	0.017	0.024	0.022	0.022
$\theta^{(11)}$	0.112	0.072	0.043	0.029	0.022	0.016

Std

	$n = 1000$	$n = 5000$	$n = 10000$	$n = 20000$	$n = 50000$	$n = 100000$
$\theta^{(1)}$	0.173	0.118	0.105	0.093	0.081	0.077
$\theta^{(2)}$	0.203	0.142	0.126	0.110	0.092	0.087
$\theta^{(3)}$	0.188	0.172	0.100	0.099	0.061	0.055
$\theta^{(4)}$	0.222	0.130	0.134	0.114	0.092	0.086
$\theta^{(5)}$	0.322	0.189	0.180	0.162	0.137	0.131
$\theta^{(6)}$	0.286	0.195	0.175	0.157	0.134	0.130
$\theta^{(7)}$	0.415	0.239	0.231	0.207	0.162	0.169
$\theta^{(8)}$	0.434	0.296	0.282	0.228	0.179	0.178
$\theta^{(9)}$	0.327	0.244	0.214	0.175	0.168	0.152
$\theta^{(10)}$	0.324	0.244	0.209	0.175	0.168	0.153
$\theta^{(11)}$	0.442	0.309	0.268	0.224	0.187	0.162

Table 3.1: Bias and std for the GMM estimator $\hat{\theta}_{n,10}$. Estimation based on 500 replications with θ_0 as in Example 3.6.2.

3.6.2 Simulation results for Example 3.6.2

In this section we analyze the behavior of the GMM estimator when the consistency conditions are valid, but asymptotic normality do not. Here, we have that C_{θ_0} is positive definite, A_{θ_0} is invertible and $\sigma(B_{\theta_0}), \sigma(\mathcal{B}_{\theta_0})$, $\sigma(\mathcal{C}_{\theta_0})$ and $\sigma(B_{\theta_0}/4 + B^*_{\theta_0}/4 + \sigma_L A^*_{\theta_0} A_{\theta_0}) \in (-\infty, 0) + i\mathbb{R}$. Thus, Theorem 3.5.9 applies and gives weakly consistency of the GMM estimator. On the other hand, for $p = 4.001$ the integral in (3.6.5) is $14.22 > 0$, and thus, we cannot apply Theorem 3.5.6 to ensure asymptotic normality.

The results for Example 3.6.2 is given in Table 3.1. We only present the results for $H = 10$, since this choice gave in general, smaller bias and standard deviation when, compared with $H = 2$ and $H = 5$. The bias and standard deviation decreases in general as n grows, showing consistency of the estimators. When compared with Example 3.6.1, the estimation of the entries of B_{θ_0} does not seem to be substantially more difficult than the entries of A_{θ_0} and C_{θ_0}. For $n \geq 20\,000$, one observes the standard deviations for estimating the entries of B_{θ_0} and C_{θ_0} is approximately twice as much the ones when estimating the entries of A_{θ_0}.

3.7 Proofs

3.7.1 Auxiliary results

Several results related to the algebra of multivariate stochastic integrals will be used here, for which we refer to Lemma 2.1 in Behme [3]. Other additional Lemmas are stated bellow.

Fact 3.7.1. *(See (4.15) in [98]) Let X, Z in $\mathbb{S}_d$. Then there exist an invertible linear operator $\boldsymbol{Q}$ such that*

$$\boldsymbol{Q}(\operatorname{vec}(X)\operatorname{vec}(Z)^*) = X \otimes Z.$$

Fact 3.7.2. *Let $(A_t)_{t\in\mathbb{R}^+}$ in $M_{m,n}(\mathbb{R})$, $(B_t)_{t\in\mathbb{R}^+}$ in $M_{n,1}(\mathbb{R}), (C_t)_{t\in\mathbb{R}^+}$ in $M_{1,p}(\mathbb{R})$ and $(D_t)_{t\in\mathbb{R}^+}$ in $M_{p,q}(\mathbb{R})$. Then*

$$\left[\int_0^\cdot A_s \mathrm{d}B_s, \int_0^\cdot \mathrm{d}C_s D_s\right]_t = \int_0^t A_s \mathrm{d}[B, C]_s D_s, \quad t \geq 0. \tag{3.7.1}$$

Fact 3.7.3. *(Lemma 6.9 in [98] with drift) Assume that $(X_t)_{t\in\mathbb{R}^+}$ is an adapted cadlag $M_{d,d}(\mathbb{R})$-valued process satisfying $\mathbb{E}(\|X_t\|) < \infty$ for all $t \in \mathbb{R}^+$, $t \mapsto \mathbb{E}(\|X_t\|)$ is locally bounded and $(L_t)_{s\in\mathbb{R}_+}$ is an $\mathbb{R}^d$-valued Lévy process of finite variation with $\mathbb{E}(\|L_1\|) < \infty$. Then*

$$\mathbb{E}\int_0^\Delta X_{s-}\mathrm{d}L_s = \int_0^\Delta \mathbb{E}(X_{s-})\mathbb{E}(L_1)\mathrm{d}s.$$

Fact 3.7.4. *Let $(A_t)_{t\in\mathbb{R}^+}$ in $M_{d^2,d^2}(\mathbb{R})$, $(B_t)_{t\in\mathbb{R}^+}$ in $M_{1,d^2}(\mathbb{R})$ be adapted caglad processes satisfying $\mathbb{E}\|A_t \operatorname{vec}(I_d)B_t\| < \infty$ for all $t \in \mathbb{R}^+$, $t \mapsto \mathbb{E}\|A_t \operatorname{vec}(I_d)B_t\|$ is locally bounded and $(L_t)_{s\in\mathbb{R}_+}$ be an R^d valued Lé process satisfying Assumption 5.2 in [98]. Then,*

$$\mathbb{E}\int_0^t A_s \mathrm{d}(\operatorname{vec}([L,L]_s))B_s = (\sigma_W + \sigma_L)\int_0^t \mathbb{E}[A_s \operatorname{vec}(I_d)B_s]\mathrm{d}s.$$

We need a lemma on the finiteness of the relevant moments.

Lemma 3.7.1. *(Propositions 4.7-5.2 and 5.4 in [98]) Assume that $(a.3)$, $(a.4)$ and $(c.1)$ hold. Then, for $k \in \{1,2\}$:*

(a) $\mathbb{E}\|Y_t\|^k < \infty$ for all $t \in \mathbb{R}^+$ and $t \mapsto \mathbb{E}\|Y_t\|^k$ is locally bounded.

(b) $\mathbb{E}\|G_t\|^{2k} < \infty$ for all $t \in \mathbb{R}^+$ and $t \mapsto \mathbb{E}\|G_t\|^{2k}$ is locally bounded.

Proof. (a) See Proposition 4.7 in [98].
(b) The finiteness of the moments is proved in Propositions 5.2 and 5.4 in [98]. We prove local boundedness. Let $k = 1$. It follows from (5.6) in [98] that for every $\Delta > 0, \mathbb{E}(\boldsymbol{G}_1\boldsymbol{G}_1^*) = (\sigma_L+\sigma_W)\Delta\mathbb{E}(V_0)$. Thus, $\mathbb{E}\|G_t\|^2 \le (\sigma_L+\sigma_W)t\sum_{j=1}^d \mathbb{E}(V_{0,j,j})$, so that $t \mapsto \mathbb{E}\|G_t\|^2$ is locally bounded. Let $k = 2$. We give an upper bound for $\mathbb{E}[G_i,G_i]_t^2, i = 1,\dots,d$. It follows from (3.2) in [98] that

$$\begin{aligned}
&[G_i,G_i]_t^2\\
&= \Big(\sum_{k,l=1}^d \int_0^t V_{ik,s-}^{1/2}V_{li,s-}^{1/2} d[L_k,L_l]_s\Big)^2\\
&= \Big(\sum_{k,l=1}^d \Big[\int_0^t V_{ik,s-}^{1/2}V_{li,s-}^{1/2} d([L_k,L_l]_s - \mathbb{E}[L_k,L_l]_s) + \int_0^t V_{ik,s-}^{1/2}V_{li,s-}^{1/2} d(\mathbb{E}[L_k,L_l]_s)\Big]\Big)^2\\
&=: (M_t+N_t)^2.
\end{aligned} \tag{3.7.2}$$

Under $(a.3)$, $\mathbb{E}([L_k,L_l]_s) = \mathbb{1}_{k=l}(\mathbb{E}L_k^2)$. This combined with Jensen's inequality and $(c.1)$ gives,

$$\begin{aligned}
\mathbb{E}N_t^2 &= \mathbb{E}\Big(\sum_{k=1}^d \int_0^t V_{ik,s-}^{1/2}V_{li,s-}^{1/2}(\mathbb{E}L_k^2)ds\Big)^2 \le 2^{d-1}t\Big(\sum_{k=1}^d(\mathbb{E}L_k^2)\Big)\mathbb{E}\int_0^t (V_{ik,s-}^{1/2}V_{ki,s-}^{1/2})^2 ds\\
&\le 2^{d-1}t^2\mathbb{E}\|V_0\|^2\Big(\sum_{k=1}^d(\mathbb{E}L_k^2)\Big).
\end{aligned} \tag{3.7.3}$$

Using the Burkholder-Davis-Gundy inequality for the martingale M_t, using $(a.3)$, Fact 3.7.1(a) and $((c.1))$ gives $\mathbb{E}M_t^2 \leq ct$. This combined with (3.7.3) proves that $t \mapsto E[G_i, G_i]_t^2$ is locally bounded. □

Lemma 3.7.2. *Assume that* $(a.2)$, $(a.3)$, $(a.4)$, $(a.5)$, $(b.1)$ *and* $(c.1)$ *hold. If* $A \in GL(\mathbb{R})$, *then*

$$\begin{aligned} &\operatorname{Cov}(\operatorname{vec}(Y_\Delta), \operatorname{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*)) = \operatorname{Cov}(\operatorname{vec}(Y_\Delta), \operatorname{vec}(G_\Delta G_\Delta^*)) \\ &= \operatorname{Var}(\operatorname{vec}(V_0))(e^{\mathcal{B}^*\Delta} - I_{d^2})[(\sigma_W + \sigma_L)(\mathcal{B}^*)^{-1} - 2((A \otimes A)^*)^{-1}], \quad \Delta \geq 0. \end{aligned} \tag{3.7.4}$$

Proof. Since $(a.3)$, $(a.4)$ and $(c.1)$ hold, we can apply Fact 3.7.1 with $k = 2$ to conclude that both $\|\operatorname{vec}(Y_\Delta)\|$ and $\|\boldsymbol{G}_1\boldsymbol{G}_1^*\|$ are square integrable random variables and thus, the covariance at the left hand side of (3.4.1) is finite. Use integration by parts formula at the end of p. 111 in [98] to write

$$G_\Delta G_\Delta^* = \int_0^\Delta V_{s-}^{1/2} \mathrm{d}L_s G_{s-}^* + \int_0^\Delta G_{s-} \mathrm{d}L_s^* V_{s-}^{1/2} + \int_0^\Delta V_{s-}^{1/2} \mathrm{d}[L, L^*]_s V_{s-}^{1/2} := A_\Delta + A_\Delta^* + C_\Delta. \tag{3.7.5}$$

It follows from Fact 3.7.1(a) and (b) together with the Cauchy-Schwarz inequality that

$$\int_0^t \mathbb{E}\|V_{s-}^{1/2}\|\|G_{s-}\|\mathrm{d}s \leq \int_0^t \left(\mathbb{E}\|V_{s-}\|\right)^{1/2} \left(\mathbb{E}\|G_{s-}\|^2\right)^{1/2} \mathrm{d}s < \infty \tag{3.7.6}$$

and therefore $(A_t)_{t\geq 0}$ is a martingale and $A_t \in L^2$ for all $t \geq 0$. Thus, the integration by parts formula, the formula $\mathrm{d}(\operatorname{vec}(A_s))^* = \mathrm{d}L_s^*(G_{s-}^* \otimes V_{s-}^{1/2})$ (Lemma 2.1(vi) in [3]) and the Itô isometry imply

$$\begin{aligned} &\operatorname{Cov}(\operatorname{vec}(Y_\Delta), \operatorname{vec}(A_\Delta)) \\ &= \mathbb{E}[\operatorname{vec}(Y_\Delta)(\operatorname{vec}(A_\Delta))^*] - \mathbb{E}\operatorname{vec}(Y_\Delta)\mathbb{E}(\operatorname{vec}(A_\Delta))^* \\ &= \mathbb{E}\Big(\int_0^\Delta \operatorname{vec}(Y_{s-})\mathrm{d}(\operatorname{vec}(A_s))^* + \int_0^\Delta \mathrm{d}\operatorname{vec}(Y_s)(\operatorname{vec}(A_{s-}))^* + [\operatorname{vec}(Y), (\operatorname{vec}(A))^*]_\Delta\Big) - 0 \\ &= 0 + \mathbb{E}\int_0^\Delta \mathrm{d}\operatorname{vec}(Y_s)(\operatorname{vec}(A_{s-}))^* + \mathbb{E}([\operatorname{vec}(Y), (\operatorname{vec}(A))^*]_\Delta). \end{aligned} \tag{3.7.7}$$

Let $\tilde{C} := (B \otimes I + I \otimes B)$ and recall from p.84 in [98] that

$$\mathrm{d}\operatorname{vec}(Y_s) = \tilde{C}\operatorname{vec}(Y_{s-})\mathrm{d}s + (A \otimes A)(V_{s-}^{1/2} \otimes V_{s-}^{1/2})\mathrm{d}\operatorname{vec}([L, L]_s^{\mathfrak{d}}) \tag{3.7.8}$$

Using (3.7.8), the bilinearity of the quadratic covariation process, Facts 3.7.2, 3.7.1, 3.7.3,

$(a.2)$ and (3.7.6) and Itô isometry gives

$$\begin{aligned}&\mathbb{E}[\operatorname{vec}(Y),(\operatorname{vec}(A))^*]_\Delta\\&=\mathbb{E}\Big[\int_0^\cdot \tilde{C}\operatorname{vec}(Y_{s-})\mathrm{d}s+\int_0^\cdot (A\otimes A)(V_{s-}^{1/2}\otimes V_{s-}^{1/2})\mathrm{d}\operatorname{vec}([L,L]_s^{\mathfrak{d}}),\int_0^\cdot \mathrm{d}L_s^*(G_{s-}^*\otimes V_{s-}^{1/2})\Big]_\Delta\\&=\mathbb{E}\int_0^\Delta (A\otimes A)(V_{s-}^{1/2}\otimes V_{s-}^{1/2})\mathrm{d}[\operatorname{vec}([L,L]^{\mathfrak{d}}),L^*]_s(G_{s-}^*\otimes V_{s-}^{1/2})=0.\end{aligned} \tag{3.7.9}$$

Let $l_s := \mathbb{E}\operatorname{vec}(Y_s)(\operatorname{vec}(A_s))^*$ and notice that it follows from Fact 3.7.1 and the Cauchy-Schwarz inequality that $\mathbb{E}\|l_s\| < \infty$ and $s \mapsto \mathbb{E}\|l_s\|$ is locally bounded. Use (3.7.8), (3.7.9), the compensation formula, Proposition 7.1.9 in [5], $\mathbb{E}\operatorname{vec}(V_s)\operatorname{vec}(A_{s-}) = l_s$ and Itô isometry and $(a.5)$ to get

$$\begin{aligned}l_\Delta &= \mathbb{E}\int_0^\Delta \mathrm{d}\operatorname{vec}(Y_s)(\operatorname{vec}(A_{s-}))^*\\&=\mathbb{E}\int_0^\Delta [\tilde{C}\operatorname{vec}(Y_{s-})\mathrm{d}s+(A\otimes A)(V_{s-}^{1/2}\otimes V_{s-}^{1/2})\mathrm{d}(\operatorname{vec}([L,L]_s^{\mathfrak{d}})](\operatorname{vec}(A_{s-}))^*\\&=\tilde{C}\int_0^\Delta \mathbb{E}\operatorname{vec}(Y_{s-})(\operatorname{vec}(A_{s-}))^*\mathrm{d}s\\&+\sigma_L\int_0^\Delta \mathbb{E}[(A\otimes A)(V_{s-}^{1/2}\otimes V_{s-}^{1/2})\operatorname{vec}(I_d)(\operatorname{vec}(A_{s-}))^*]\mathrm{d}s\\&=(\tilde{C}+\sigma_L(A\otimes A))\int_0^\Delta l_s\mathrm{d}s.\end{aligned} \tag{3.7.10}$$

Solving the matrix valued integral equation in (3.7.10) and using that $A_0 = 0$ implies $l_0 = 0$, gives $l_s = 0$ for all $s \geq 0$ (see [45]). Thus, it follows from (3.7.7)-(3.7.10) that

$$\operatorname{Cov}(\operatorname{vec}(Y_\Delta),\operatorname{vec}(A_\Delta)) = 0, \tag{3.7.11}$$

and, as a consequence $\operatorname{Cov}(\operatorname{vec}(Y_\Delta),\operatorname{vec}(A_\Delta^*)) = 0$. Let $\mathcal{V}_{s-} := V_{s-}^{1/2}\otimes V_{s-}^{1/2}$. Then,

$$\begin{aligned}&\operatorname{vec}(C_\Delta)\\&=\int_0^\Delta \mathcal{V}_{s-}\mathrm{d}\operatorname{vec}([L,L^*]_s)=\int_0^\Delta \mathcal{V}_{s-}\mathrm{d}\operatorname{vec}([L,L^*]_s^{\mathfrak{d}})+\sigma_W\int_0^\Delta (V_{s-}^{1/2}\otimes V_{s-}^{1/2})\operatorname{vec}(I_d)\mathrm{d}s\\&=\int_0^\Delta \mathcal{V}_{s-}\mathrm{d}\operatorname{vec}([L,L^*]_s^{\mathfrak{d}})+\sigma_W\int_0^\Delta \operatorname{vec}(V_{s-})\mathrm{d}s.\end{aligned} \tag{3.7.12}$$

Using the compensation formula, Fact 3.7.3 and the stationarity of $(V_s)_{s\in\mathbb{R}_+}$ we get

$$\mathbb{E}\int_0^\Delta \mathcal{V}_{s-}\mathrm{d}\operatorname{vec}([L,L^*]_s)=(\sigma_W+\sigma_L)\int_0^\Delta \mathbb{E}\mathcal{V}_{s-}\operatorname{vec}(I_d)\mathrm{d}s=\Delta(\sigma_W+\sigma_L)\mathbb{E}\operatorname{vec}(V_0). \tag{3.7.13}$$

Additionally, it follows from Fact 3.7.1 that $\mathbb{E}\|\operatorname{vec}(V_s)\operatorname{vec}(Y_\Delta)^*\| < \infty$ for all $s \geq 0$ and that $s \mapsto \mathbb{E}\|\operatorname{vec}(V_s)\operatorname{vec}(Y_\Delta)^*\|$ is locally bounded. Then,

$$\begin{aligned}&\mathbb{E}\Big(\int_0^\Delta \operatorname{vec}(V_{s-})\mathrm{d}s\,(\operatorname{vec}(Y_\Delta))^*\Big) = \int_0^\Delta \mathbb{E}\operatorname{vec}(V_{s-})(\operatorname{vec}(Y_\Delta))^*\mathrm{d}s \\ &= \Delta\operatorname{vec}(C)(\mathbb{E}\operatorname{vec}(Y_0))^* + \int_0^\Delta \mathbb{E}\operatorname{vec}(Y_s)(\operatorname{vec}(Y_\Delta))^*\mathrm{d}s.\end{aligned} \tag{3.7.14}$$

Now it follows from the invertibility of $(A \otimes A)$ and from the second equation following (3.5) in [98] that

$$\begin{aligned}&\int_0^\Delta \mathcal{V}_{s-}\mathrm{d}\operatorname{vec}([L,L^*]_s^{\mathfrak{d}}) \\ &= (A\otimes A)^{-1}\Big(\operatorname{vec}(Y_\Delta) - \operatorname{vec}(Y_0) - \int_0^\Delta (B\otimes I + I\otimes B)\operatorname{vec}(Y_{s-})\mathrm{d}s\Big).\end{aligned} \tag{3.7.15}$$

The representation in (3.7.15) gives

$$\begin{aligned}&\mathbb{E}\Big[\Big(\int_0^\Delta \mathcal{V}_{s-}\mathrm{d}\operatorname{vec}([L,L^*]_s^{\mathfrak{d}})\Big)(\operatorname{vec}(Y_\Delta))^*\Big] \\ &= \mathbb{E}\Big[(A\otimes A)^{-1}\Big(\operatorname{vec}(Y_\Delta) - \operatorname{vec}(Y_0) - \int_0^\Delta (B\otimes I + I\otimes B)\operatorname{vec}(Y_{s-})\mathrm{d}s\Big)(\operatorname{vec}(Y_\Delta))^*\Big] \\ &= (A\otimes A)^{-1}\Big[\mathbb{E}\operatorname{vec}(Y_\Delta)(\operatorname{vec}(Y_\Delta))^* - \mathbb{E}\operatorname{vec}(Y_0)(\operatorname{vec}(Y_\Delta))^* \\ &\quad - (B\otimes I + I\otimes B)\int_0^\Delta \mathbb{E}\operatorname{vec}(Y_{s-})(\operatorname{vec}(Y_\Delta))^*\mathrm{d}s\Big].\end{aligned} \tag{3.7.16}$$

Using the definition of C_Δ in (3.7.5), together with (3.7.12), (3.7.13) and (3.7.16) gives

$$\begin{aligned}&\operatorname{Cov}(\operatorname{vec}(C_\Delta),\operatorname{vec}(Y_\Delta)) = (A\otimes A)^{-1}\Big[\mathbb{E}\operatorname{vec}(Y_\Delta)(\operatorname{vec}(Y_\Delta))^* - \mathbb{E}\operatorname{vec}(Y_0)(\operatorname{vec}(Y_\Delta))^* \\ &\qquad - (B\otimes I + I\otimes B)\Big(\int_0^\Delta \mathbb{E}\operatorname{vec}(Y_{s-})(\operatorname{vec}(Y_\Delta))^*\mathrm{d}s\Big)\Big] \\ &+ \Delta\sigma_W\operatorname{vec}(C)(\mathbb{E}\operatorname{vec}(Y_0))^* + \sigma_W\int_0^\Delta \mathbb{E}\operatorname{vec}(Y_{s-})(\operatorname{vec}(Y_\Delta))^*\mathrm{d}s \\ &- \Delta(\sigma_W+\sigma_L)\mathbb{E}\operatorname{vec}(V_0)\mathbb{E}(\operatorname{vec}(Y_\Delta))^* \\ &= [\sigma_W I_{d^2} - (A\otimes A)^{-1}(B\otimes I + I\otimes B)]\int_0^\Delta \mathbb{E}\operatorname{vec}(Y_s)(\operatorname{vec}(Y_\Delta))^*\mathrm{d}s \\ &+ (A\otimes A)^{-1}[\operatorname{Var}(\operatorname{vec}(Y_0)) - \operatorname{Cov}(\operatorname{vec}(Y_0),\operatorname{vec}(Y_\Delta))] - \Delta\sigma_L\operatorname{vec}(C)\mathbb{E}(\operatorname{vec}(Y_0))^* \\ &- \Delta(\sigma_W+\sigma_L)\mathbb{E}\operatorname{vec}(Y_0)\mathbb{E}(\operatorname{vec}(Y_0))^*,\end{aligned} \tag{3.7.17}$$

where the last inequality follows from $V_0 = C + Y_0$ and the stationarity of $(Y_s)_{s\in\mathbb{R}_+}$. Using (3.4.3) it follows first that

$$\begin{aligned}\int_0^\Delta \mathbb{E}\operatorname{vec}(Y_s)(\operatorname{vec}(Y_\Delta))^* \mathrm{d}s &= \int_0^\Delta e^{\mathcal{B}(\Delta-s)}\operatorname{Var}(\operatorname{vec}(Y_0))\mathrm{d}s + \Delta\mathbb{E}\operatorname{vec}(Y_0)\mathbb{E}(\operatorname{vec}(Y_0))^* \\ &= \mathcal{B}^{-1}(e^{\mathcal{B}\Delta} - I_{d^2})\operatorname{Var}(\operatorname{vec}(Y_0)) + \Delta\mathbb{E}\operatorname{vec}(Y_0)\mathbb{E}(\operatorname{vec}(Y_0))^*,\end{aligned} \tag{3.7.18}$$

and second that

$$\operatorname{Var}(\operatorname{vec}(Y_0)) - \operatorname{Cov}(\operatorname{vec}(Y_0), \operatorname{vec}(Y_\Delta)) = -(e^{\mathcal{B}\Delta} - I_{d^2})\operatorname{Var}(\operatorname{vec}(Y_0)). \tag{3.7.19}$$

Substituting $B\otimes I + I\otimes B = \mathcal{B} - \sigma_L(A\otimes A)$, using (3.7.18), (3.7.19), (b.1) and the formula for $\mathbb{E}\operatorname{vec}(Y_0)$ in (3.4.3) gives

$$\begin{aligned}&\operatorname{Cov}(\operatorname{vec}(C_\Delta), \operatorname{vec}(Y_\Delta)) \\ &= \big[\sigma_W I_{d^2} - (A\otimes A)^{-1}(\mathcal{B} - \sigma_L(A\otimes A))\big]\big[\mathcal{B}^{-1}(e^{\mathcal{B}\Delta} - I_{d^2})\operatorname{Var}(\operatorname{vec}(Y_0)) \\ &+ \Delta\mathbb{E}\operatorname{vec}(Y_0)\mathbb{E}(\operatorname{vec}(Y_0))^*\big] \\ &- (A\otimes A)^{-1}(e^{\mathcal{B}\Delta} - I_{d^2})\operatorname{Var}(\operatorname{vec}(Y_0)) - \Delta\sigma_L\operatorname{vec}(C)\mathbb{E}(\operatorname{vec}(Y_0))^* \\ &- \Delta(\sigma_W + \sigma_L)\mathbb{E}\operatorname{vec}(Y_0)\mathbb{E}(\operatorname{vec}(Y_0))^* \\ &= \big[(\sigma_W + \sigma_L)\mathcal{B}^{-1} - 2(A\otimes A)^{-1}\big](e^{\mathcal{B}\Delta} - I_{d^2})\operatorname{Var}(\operatorname{vec}(Y_0)) \\ &\quad - \big[(A\otimes A)^{-1}\mathcal{B}\mathbb{E}\operatorname{vec}(Y_0) + \sigma_L\operatorname{vec}(C)\big]\Delta\mathbb{E}(\operatorname{vec}(Y_0))^* \\ &= \big[(\sigma_W + \sigma_L)\mathcal{B}^{-1} - 2(A\otimes A)^{-1}\big](e^{\mathcal{B}\Delta} - I_{d^2})\operatorname{Var}(\operatorname{vec}(Y_0)) \\ &\quad - \big[(A\otimes A)^{-1}\mathcal{B}(-\sigma_L\mathcal{B}^{-1}(A\otimes A)\operatorname{vec}(C)) + \sigma_L\operatorname{vec}(C)\big]\Delta\mathbb{E}(\operatorname{vec}(Y_0))^* \\ &= \big[(\sigma_W + \sigma_L)\mathcal{B}^{-1} - 2(A\otimes A)^{-1}\big](e^{\mathcal{B}\Delta} - I_{d^2})\operatorname{Var}(\operatorname{vec}(Y_0)).\end{aligned} \tag{3.7.20}$$

Finally, the result of the Lemma given in (3.7.4) follows from (3.7.5), (3.7.11), (3.7.20) and the fact that

$$\operatorname{Cov}(\operatorname{vec}(Y_\Delta), \operatorname{vec}(G_\Delta G_\Delta^*)) = (\operatorname{Cov}(\operatorname{vec}(G_\Delta G_\Delta^*), \operatorname{vec}(Y_\Delta)))^* = (\operatorname{Cov}(\operatorname{vec}(C_\Delta), \operatorname{vec}(Y_\Delta)))^*.$$

□

3.7.2 Proofs of Lemmas 3.4.2 and 3.4.3

The proof of Lemma 3.4.2 follows directly from Lemma 3.7.2 combined with (5.7) in [98].

3.7.3 Proof of Lemma 3.4.3

An application of the Cauchy-Schwarz inequality combined with Fact 9.9.61 in [5] and Fact 3.7.1(b) gives $\mathbb{E}\|\operatorname{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*)\operatorname{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*)^*\|_2 \le c\mathbb{E}\|\boldsymbol{G}_1\|_2^4 < \infty$ for some $c > 0$. Let

$a_s := \operatorname{vec}(G_s G_s^*)$, $s \in [0, \Delta]$ and use the integration by parts formula to write

$$\begin{aligned} a_\Delta a_\Delta^* &= \int_0^\Delta a_{s-} \mathrm{d}(a_s^*) + \int_0^\Delta \mathrm{d}a_s (a_{s-}^*) + [a, a^*]_\Delta \\ &= \Big(\int_0^\Delta \mathrm{d}a_s (a_{s-}^*) \Big)^* + \int_0^\Delta \mathrm{d}a_s (a_{s-}^*) + [a, a^*]_\Delta, \end{aligned} \tag{3.7.21}$$

which means that we only need to compute

$$\mathbb{E} \int_0^\Delta \mathrm{d}a_s (a_{s-}^*) \quad \text{and} \quad \mathbb{E}[a, a^*]_\Delta.$$

From (3.7.5), Lemma 2.1(vi) in Behme [3] and the symmetry of $(V_t)_{t \in \mathbb{R}_+}$ it follows that

$$\begin{aligned} \mathrm{d}a_t &= \mathrm{d}(\operatorname{vec}(G_t G_t^*)) \\ &= \mathrm{d}\Big(\operatorname{vec}\Big(\int_0^t V_{s-}^{1/2} \mathrm{d}L_s G_{s-}^* + \int_0^t G_{s-} \mathrm{d}L_s^* V_{s-}^{1/2} + \int_0^t V_{s-}^{1/2} \mathrm{d}[L, L^*]_s V_{s-}^{1/2} \Big)\Big) \\ &= \mathrm{d}\Big(\int_0^t (G_{s-} \otimes V_{s-}^{1/2}) \mathrm{d}L_s + \int_0^t (V_{s-}^{1/2} \otimes G_{s-}) \mathrm{d}L_s + \int_0^t (V_{s-}^{1/2} \otimes V_{s-}^{1/2}) \mathrm{d}\operatorname{vec}([L, L^*]_s) \Big) \\ &= (G_{t-} \otimes V_{t-}^{1/2} + V_{t-}^{1/2} \otimes G_{t-}) \mathrm{d}L_t + (V_{t-}^{1/2} \otimes V_{t-}^{1/2}) \mathrm{d}\operatorname{vec}([L, L^*]_t), \quad t \geq 0. \end{aligned} \tag{3.7.22}$$

Thus it follows from (3.7.22), the Itô isometry, the fact that $[L, L^*]_t = [L, L^*]_t^{\mathfrak{d}} + \sigma_w I_d t$, facts 3.7.4 and 3.7.1 that

$$\begin{aligned} &\mathbb{E} \int_0^\Delta \mathrm{d}a_s (a_{s-}^*) \\ &= \mathbb{E}\Big(\int_0^\Delta (G_{s-} \otimes V_{s-}^{1/2} + V_{s-}^{1/2} \otimes G_{s-}) \mathrm{d}L_s a_{s-}^* + \int_0^\Delta (V_{s-}^{1/2} \otimes V_{s-}^{1/2}) \mathrm{d}\operatorname{vec}([L, L^*]_s) a_{s-}^* \Big) \\ &= (\sigma_L + \sigma_W) \Big(\int_0^\Delta \mathbb{E}((V_{s-}^{1/2} \otimes V_{s-}^{1/2}) \operatorname{vec}(I_d) a_{s-}^*) \mathrm{d}s \Big) \\ &= (\sigma_L + \sigma_W) \int_0^\Delta \mathbb{E}(\operatorname{vec}(V_{s-}) a_{s-}^*) \mathrm{d}s. \end{aligned} \tag{3.7.23}$$

It follows from (5.6) in [98] that

$$\int_0^\Delta \mathbb{E} a_{s-}^* \mathrm{d}s = \int_0^\Delta \big(\operatorname{vec}((\sigma_L + \sigma_W) s \mathbb{E} V_0) \big)^* \mathrm{d}s = \frac{1}{2} (\sigma_L + \sigma_W) \Delta^2 \mathbb{E} \operatorname{vec}(V_0)^*. \tag{3.7.24}$$

Since we assumed here that all hypothesis for using Lemma 3.7.2 are valid, we can use (3.7.4) with $\Delta = s$ to get

$$\begin{aligned}&\int_0^\Delta \mathrm{Cov}(\mathrm{vec}(Y_{s-}), a_{s-})\mathrm{d}s\\&= \mathrm{Var}(\mathrm{vec}(Y_0))\Big(\int_0^\Delta (e^{\mathcal{B}^* s} - I_{d^2})\mathrm{d}s\Big)[(\sigma_W + \sigma_L)(\mathcal{B}^*)^{-1} - 2((A \otimes A)^*)^{-1}]\\&= \mathrm{Var}(\mathrm{vec}(Y_0))\tilde{\mathcal{B}},\end{aligned} \tag{3.7.25}$$

where $\tilde{\mathcal{B}}$ is defined in (3.4.10). Using (3.7.23), $(c.1)$ (3.7.24), (3.7.25) gives

$$\begin{aligned}\int_0^\Delta \mathbb{E}\,\mathrm{vec}(V_{s-})a_{s-}^*\mathrm{d}s &= \int_0^\Delta \mathrm{Cov}(\mathrm{vec}(V_{s-}), a_{s-})\mathrm{d}s + (\mathbb{E}\,\mathrm{vec}(V_s))\int_0^\Delta \mathbb{E}(a_{s-}^*)\mathrm{d}s\\&= \int_0^\Delta \mathrm{Cov}(\mathrm{vec}(Y_{s-}), a_{s-})\mathrm{d}s + (\mathbb{E}\,\mathrm{vec}(V_0))\int_0^\Delta \mathbb{E}(a_{s-}^*)\mathrm{d}s\\&= \frac{1}{2}(\sigma_L + \sigma_W)\Delta^2\mathbb{E}\,\mathrm{vec}(V_0)E\,\mathrm{vec}(V_0)^* + \mathrm{Var}(\mathrm{vec}(Y_0))\tilde{\mathcal{B}}\\&= (\sigma_L + \sigma_W)^{-1}D,\end{aligned} \tag{3.7.26}$$

where D is defined in (3.4.9). Let $f_s := (G_{s-} \otimes V_{s-}^{1/2} + V_{s-}^{1/2} \otimes G_{s-}), s \geq 0$ and recall $\mathcal{V}_{s-} = V_{s-}^{1/2} \otimes V_{s-}^{1/2}$. Using (3.7.5), Lemma 2.1(vi) in [3] and the symmetry of $V_{s-}^{1/2}$ gives

$$\begin{aligned}&[a, a^*]_\Delta\\&= \Big[\mathrm{vec}\Big(\int_0^\cdot V_{s-}^{1/2}\mathrm{d}L_s G_{s-}^* + \int_0^\cdot G_{s-}\mathrm{d}L_s^* V_{s-}^{1/2} + \int_0^\cdot V_{s-}^{1/2}\mathrm{d}[L, L^*]_s V_{s-}^{1/2}\Big),\\&\Big(\mathrm{vec}\Big(\int_0^\cdot V_{s-}^{1/2}\mathrm{d}L_s G_{s-}^* + \int_0^\cdot G_{s-}\mathrm{d}L_s^* V_{s-}^{1/2} + \int_0^\cdot V_{s-}^{1/2}\mathrm{d}[L, L^*]_s V_{s-}^{1/2}\Big)\Big)^*\Big]_\Delta\\&= \Big[\int_0^\cdot f_{s-}\mathrm{d}L_s + \int_0^\cdot \mathcal{V}_{s-}\mathrm{d}\,\mathrm{vec}([L, L^*]_s), \int_0^\cdot \mathrm{d}L_s^* f_{s-}^* + \int_0^\cdot \mathrm{d}(\mathrm{vec}([L, L^*]_s)^*)\mathcal{V}_{s-}\Big]_\Delta\\&= \int_0^\Delta f_{s-}\mathrm{d}[L, L^*]_s f_{s-}^* + \int_0^\Delta f_{s-}\mathrm{d}[L, \mathrm{vec}([L, L^*])^*]_s\mathcal{V}_{s-}\\&+ \int_0^\Delta \mathcal{V}_{s-}\mathrm{d}[\mathrm{vec}([L, L^*]), L^*]_s f_{s-}^* + \int_0^\Delta \mathcal{V}_{s-}\mathrm{d}[\mathrm{vec}([L, L^*]), \mathrm{vec}([L, L^*])^*]\mathcal{V}_{s-}\\&:= I_1 + I_2 + I_3 + I_4.\end{aligned} \tag{3.7.27}$$

By Fact 3.7.1 we have $\mathbb{E}\|\mathcal{V}_{s-}\|\|f_{s-}\| < \infty$ and the map $s \mapsto \mathbb{E}\|\mathcal{V}_{s-}\|\|f_{s-}\|$ is locally bounded. The same is true for the random variable $\|\mathcal{V}_{s-}\|\|\mathcal{V}_{s-}\|$. Thus, it follows from $(a.2)$ that we have $\mathbb{E}I_2 = \mathbb{E}I_3 = 0$. Using the second-order stationarity of $(V_s)_{s\in\mathbb{R}_+}$ in

$(c.1)$, the compensation formula and the formulas at p. 108 in [98]

$$
\begin{aligned}
\mathbb{E}I_4 &= \mathbb{E}\Big(\int_0^\Delta \mathcal{V}_{s-}\mathrm{d}[\mathrm{vec}([L,L^*]), \mathrm{vec}([L,L^*])^*]\mathcal{V}_{s-}\Big) \\
&= \mathbb{E}\Big(\int_0^\Delta \mathcal{V}_{s-}\mathrm{d}[\mathrm{vec}([L,L^*]^{\mathfrak{d}}), (\mathrm{vec}([L,L^*]^{\mathfrak{d}}))^*]^{\mathfrak{d}}\mathcal{V}_{s-}\Big) \\
&= \int_0^\Delta \mathbb{E}(\mathcal{V}_{s-}\rho_L[I_{d^2} + K_d + \mathrm{vec}(I_d)\,\mathrm{vec}(I_d)^*]\mathcal{V}_{s-})\mathrm{d}s \\
&= \rho_L \int_0^\Delta (\boldsymbol{Q} + K_d\boldsymbol{Q} + I_{d^2})\mathbb{E}(\mathrm{vec}(V_s)\,\mathrm{vec}(V_s)^*)\mathrm{d}s \\
&= \Delta\rho_L(\boldsymbol{Q} + K_d\boldsymbol{Q} + I_{d^2})\mathbb{E}\,\mathrm{vec}(V_0)\,\mathrm{vec}(V_0)^*.
\end{aligned}
\tag{3.7.28}
$$

To compute $\mathbb{E}I_1$ we will need the following matrix identity, which is based on Fact 7.4.30(xiv) in Bernstein [5] and Fact 3.7.1. Let $A \in M_{d,1}(\mathbb{R})$ and $B, B^2 \in M_{d,d}(\mathbb{R})$ be symmetric matrices. Then,

$$
\begin{aligned}
&(A\otimes B + B\otimes A)(A\otimes B + B\otimes A)^* = (A\otimes B + K_d(A\otimes B))(A\otimes B + K_d(A\otimes B))^* \\
&= (I+K_d)(A\otimes B)(A^*\otimes B)(I+K_d) = (I+K_d)\boldsymbol{Q}\,\mathrm{vec}(AA^*)\,\mathrm{vec}(B^2)(I+K_d).
\end{aligned}
\tag{3.7.29}
$$

Write $b_s := \mathbb{E}\,\mathrm{vec}(G_sG_s^*)\,\mathrm{vec}(V_s)^*$, which is finite by Fact 3.7.1. Using the compensation formula, (3.7.29) and the definition of f_s gives

$$
\begin{aligned}
&\mathbb{E}\Big(\int_0^\Delta f_{s-}\mathrm{d}[L,L^*]_s f_{s-}^*\Big) \\
&= (\sigma_L + \sigma_W)\int_0^\Delta \mathbb{E}(f_s f_s^*)\mathrm{d}s \\
&= (\sigma_L + \sigma_W)\int_0^\Delta \mathbb{E}(G_{s-}\otimes V_{s-}^{1/2} + V_{s-}^{1/2}\otimes G_{s-})(G_{s-}^*\otimes V_{s-}^{1/2} + V_{s-}^{1/2}\otimes G_{s-}^*)\mathrm{d}s \\
&= (\sigma_L + \sigma_W)\int_0^\Delta (I+K_d)\boldsymbol{Q}b_s(I+K_d)\mathrm{d}s \\
&= (\sigma_L + \sigma_W)(I+K_d)\boldsymbol{Q}\Big(\int_0^\Delta b_s\mathrm{d}s\Big)(I+K_d).
\end{aligned}
\tag{3.7.30}
$$

Finally, it follows from (3.7.26) that

$$
\int_0^\Delta b_s^*\mathrm{d}s = \int_0^\Delta \mathbb{E}\,\mathrm{vec}(V_s)a_{s-}^*\mathrm{d}s = (\sigma_L + \sigma_W)^{-1}D. \tag{3.7.31}
$$

The result now is a direct consequence of (3.7.21), (3.7.26), (3.7.27), (3.7.28), (3.7.30) and (3.7.31).

Bibliography

[1] D. Applebaum. *Lévy Processes and Stochastic Calculus*. Cambridge University Press, Cambridge, 2009.

[2] S. Bayracı and G. Ünal. Stochastic interest rate volatility modeling with a continuous-time GARCH(1,1) model. *Journal of Computational and Applied Mathematics*, 259:464–473, 2014.

[3] A. Behme. Moments of MGOU processes and positive semidefinite matrix processes. *Journal of Multivariate Analysis*, 111:183–197, 2012.

[4] D. Belomestny, F. Comte, V. Genon-Catalot, H. Masuda, and M. Reiß. *Estimation for Discretely Observed Lévy Processes*. Lévy Matters IV (LN in Mathematics, Vol. 2128). Springer, Cham, 2015.

[5] D. S. Bernstein. *Matrix Mathematics: Theory, Facts, and Formulas*. Princeton University Press, Princeton, 2009.

[6] C. Bianchi and E. M. Cleur. Indirect estimation of stochastic differential equation models: some computational experiments. *Computational Economics*, 9(3):257–274, 1996.

[7] E. Bibbona and I. Negri. Higher moments and prediction-based estimation for the COGARCH(1,1) model. *Scandinavian Journal of Statistics*, 42(4):891–910, 2015.

[8] E. Bibbona, I. Negri, A. Arcagni, and L. Mercuri. *COGARCH: Simulation and inference for COGARCH(1,1) processes*, 2014. URL `http://R-Forge.R-project.org/projects/cogarch/`. R package version 1.3-1/r32.

[9] P. Billingsley. *Probability and Measure*. Wiley, New York, 2008.

[10] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

[11] G. E. Box and Jenkins. *Time Series Analysis: Forecasting and Control*. Holden Day, San Francisco, 1970.

[12] R. Bradley. *Introduction to Strong Mixing Conditions*, volume 1. Kendrick Press, Heber City, 2007.

[13] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer, New York, 2013.

[14] M. Campbell. Time series regression for counts: an investigation into the relationship between sudden infant death syndrome and environmental temperature. *Journal of the Royal Statistical Society. Series A*, pages 191–208, 1994.

[15] M. Carrasco, M. Chernov, J.-P. Florens, and E. Ghysels. Efficient estimation of general dynamic models with a continuum of moment conditions. *Journal of Econometrics*, 140(2):529–573, 2007.

[16] K. Chan and J. Ledolter. Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association*, 90:242–252, 1995.

[17] S. Cohen and J. Rosinski. Gaussian approximation of multivariate lévy processes with applications to simulation of tempered stable processes. *Bernoulli*, 13(1):195–210, 2007.

[18] R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1:223–236, 2001.

[19] R. A. Davis and G. Rodriguez-Yam. Estimation for state-space models based on a likelihood approximation. *Statistica Sinica*, 15:381–406, 2005.

[20] R. A. Davis and C. Y. Yau. Comments on pairwise likelihood in time series models. *Statistica Sinica*, 21:255–277, 2011.

[21] R. A. Davis, W. T. Dunsmuir, and Y. Wang. Modeling time series of count data. In S. Ghosh, editor, *Asymptotics, Nonparametrics, and Time Series*, pages 63–114. Marcel Dekker, New York, 1999.

[22] R. A. Davis, W. T. Dunsmuir, and Y. Wang. On autocorrelation in a Poisson regression model. *Biometrika*, 87(3):491–505, 2000.

[23] R. A. Davis, M. Matsui, T. Mikosch, and P. Wan. Applications of distance correlation to time series. *Bernoulli*, 24(4A):3087–3116, 2018.

[24] R. A. Davis, T. do Rêgo Sousa, and C. Klüppelberg. Indirect inference for time series using the empirical characteristic function and control variates. *arXiv:1904.08276*, 2019.

[25] X. de Luna and M. G. Genton. Robust simulation-based estimation of ARMA models. *Journal of Computational and Graphical Statistics*, 10(2):370–387, 2001.

[26] T. do Rêgo Sousa and R. Stelzer. Method of moment based estimation for the multivariate cogarch(1,1) process. 2019. In preparation.

[27] T. do Rêgo Sousa, S. Haug, and C. Klüppelberg. Indirect inference for Lévy-driven continuous-time GARCH models. *Scandinavian Journal of Statistics*, 2019. To appear.

[28] P. Doukhan. *Mixing: Properties and Examples.* Springer, Heidelberg, 1994. Springer LN in Statistics, Vol. 85.

[29] R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

[30] R. F. Engle and K. F. Kroner. Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11(1):122–150, 1995.

[31] V. Fasen-Hartmann and S. Kimmig. Robust estimation of continuous-time ARMA models via indirect inference. *arXiv:1804.00849*, 2018.

[32] T. Ferguson. *A Course in Large Sample Theory.* Springer, Dordrecht, 1996.

[33] A. Feuerverger. An efficiency result for the empirical characteristic function in stationary time-series models. *The Canadian Journal of Statistics*, 18(2):155–161, 1990.

[34] J.-J. Forneron. *Essays on Simulation-based Estimation.* PhD thesis, Columbia University, New York, 2018.

[35] C. Francq and S. G. Meintanis. Fourier-type estimation of the power GARCH model with stable-Paretian innovations. *Metrika*, 79(4):389–424, 2016.

[36] C. Francq and J.-M. Zakoïan. *GARCH Models: Structure, Statistical Inference and Financial Applications.* Wiley, West Sussex, 2011.

[37] R. Garcia, E. Renault, and D. Veredas. Estimation of stable distributions by Indirect Inference. *Journal of Econometrics*, 161(2):325–337, 2011.

[38] P. W. Glynn and R. Szechtman. Some new perspectives on the method of control variates. In K. Fang, F. Hickernell, and H. Niederreiter, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2000*, pages 27–49. Springer, 2002.

[39] C. Gourieroux, A. Monfort, and E. Renault. Indirect inference. *Journal of Applied Econometrics*, 8:S85–S118, 1993.

[40] C. Gourieroux, E. Renault, and N. Touzi. Calibration by simulation for small sample bias correction. In R. Mariano, T. Schuermann, and M. J. Weeks, editors, *Simulation-based Inference in Econometrics: Methods and Applications*, pages 328–358. Cambridge University Press, Cambridge, 2000.

[41] C. Gourieroux, P. C. B. Phillips, and J. Yu. Indirect Inference for dynamic panel models. *Journal of Econometrics*, 157(1):68–77, 2010.

[42] D. M. Guillaume, M. M. Dacorogna, R. R. Davé, U. A. Müller, R. B. Olsen, and O. V. Pictet. From the bird's eye to the microscope: A survey of new stylized facts of the intra-daily foreign exchange markets. *Finance and stochastics*, 1(2):95–129, 1997.

[43] J. D. Hamilton. *Time series analysis*. Princeton University Press, Princeton, 1994.

[44] L. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054, 1982.

[45] S. Haug, C. Klüppelberg, A. Lindner, and M. Zapp. Method of moment estimation in the COGARCH(1,1) model. *Econometrics Journal*, 10(2):320–341, 2007.

[46] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, Cambridge, 1991.

[47] J. Hutton and P. Nelson. Interchanging the order of stochastic integration and ordinary differentiation. *Stochastic Processes and their Applications*, 18:371–377, 1984.

[48] S. Iacus, L. Mercuri, and E. Rroji. COGARCH(p, q): Simulation and Inference with the YUIMA package. *Journal of Statistical Software*, 80(4), 2017.

[49] M. Iannace. *COGARCH Processes: Theory and Asymptotics for the Pseudo-Maximum Likelihood Estimator*. PhD thesis, University of Milano-Bicocca, 2014.

[50] I. Ibragimov and Y. Linnik. *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen, 1971.

[51] G. J. Jiang. Estimation of jump-diffusion processes based on Indirect Inference. *IFAC Proceedings Volumes*, 31(16):385–390, 1998.

[52] D. Khoshnevisan. *Multiparameter Processes: An Introduction to Random Fields.* Springer, New York, 2002.

[53] M. Kim and S. Lee. On the maximum likelihood estimator for irregularly observed time series data from COGARCH(1,1) models. *Revstat Statistical Journal*, 11(2): 135–168, 2013.

[54] A. Klenke. *Probability Theory: a Comprehensive Course.* Springer, Berlin, 2013.

[55] C. Klüppelberg, A. Lindner, and R. Maller. A continuous-time GARCH process driven by a Lévy process: stationarity and second-order behaviour. *Journal of Applied Probability*, 41(3):601–622, 2004.

[56] C. Klüppelberg, A. Lindner, and R. Maller. Continuous time volatility modelling: COGARCH versus Ornstein-Uhlenbeck models. In Y. Kabanov, R. Liptser, and J. Stoyanov, editors, *The Shiryaev Festschrift: From Stochastic Calculus to Mathematical Finance*, pages 393–419. Springer, Berlin, 2006.

[57] C. Klüppelberg, R. Maller, and A. Szimayer. The COGARCH: a review, with news on option pricing and statistical inference. In J. Blath, P. Imkeller, and S. Roelly, editors, *Survey in Stochastic Processes. Proc. of the 33rd SPA Conference in Berlin*, pages 29–50. EMS Series of Congress Reports, EMS Publishing House, Zürich, 2011.

[58] J. Knight and J. Yu. Empirical characteristic function in time series estimation. *Econometric Theory*, 18(3):691–721, 2002.

[59] J. L. Knight, S. E. Satchell, and J. Yu. Estimation of the stochastic volatility model by the empirical characteristic function method. *Australian & New Zealand Journal of Statistics*, 44(3):319–335, 2002.

[60] K. Königsberger. *Analysis 1.* Springer, Berlin, 6 edition, 2004.

[61] R. Kotchoni. Applications of the characteristic function-based continuum GMM in finance. *Computational Statistics & Data Analysis*, 56(11):3599–3622, 2012.

[62] U. Krengel. *Ergodic Theorems.* Walter de Gruyter, Berlin, 1985.

[63] M. P. Laurini and L. K. Hotta. Indirect inference in fractional short-term interest rate diffusions. *Mathematics and Computers in Simulation*, 94:109–126, 2013.

[64] A. Lindner and R. Maller. Lévy integrals and the stationarity of generalised Ornstein-Uhlenbeck processes. *Stochastic Processes and their Applications*, 115(10): 1701–1722, 2005.

[65] A. Lindner and R. Maller. Lévy integrals and the stationarity of generalised Ornstein-Uhlenbeck processes. *Stochastic Processes and their Applications*, 115(10): 1701–1722, 2005.

[66] B. G. Lindsay. Composite likelihood methods. *Contemporary mathematics*, 80(1): 221–239, 1988.

[67] M. J. Lombardi and G. Calzolari. Indirect estimation of alpha-stable distributions and processes. *Econometrics Journal*, 11(1):193–208, 2008.

[68] H. Lütkepohl. *New Introduction to Multiple Time Series Analysis*. Springer, Berlin Heidelberg, 2005.

[69] J. R. Magnus and H. Neudecker. The commutation matrix: some properties and applications. *The Annals of Statistics*, 7:381–394, 1979.

[70] R. A. Maller, G. Müller, and A. Szimayer. GARCH modelling in continuous time for irregularly spaced time series data. *Bernoulli*, 14(2):519–542, 2008.

[71] L. Mátyás et al. *Generalized Method of Moments Estimation*. Cambridge University Press, Cambridge, 1999.

[72] D. McFadden. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, 57:995–1026, 1989.

[73] S. G. Meintanis and E. Taufer. Inference procedures for stable-Paretian stochastic volatility models. *Mathematical and Computer Modelling*, 55(3-4):1199–1212, 2012.

[74] G. Milovanovic, B. Popovic, and V. Stojanovic. An application of the ECF method and numerical integration in estimation of the stochastic volatility models. *Facta Universitatis, Mathematics and Informatics*, 29(3):295–311, 2014.

[75] K. Mullen, D. Ardia, D. Gil, D. Windover, and J. Cline. DEoptim: An R package for global optimization by differential evolution. *Journal of Statistical Software*, 40 (6):1–26, 2011.

[76] G. Müller. MCMC estimation of the COGARCH(1,1) model. *Journal of Financial Econometrics*, 8(4):481–510, 2010.

[77] M. Ndongo, A. Kâ Diongue, A. Diop, and S. Dossou-Gbété. Estimation for seasonal fractional ARIMA with stable innovations via the empirical characteristic function method. *Statistics*, 50(2):298–311, 2016.

[78] W. K. Newey and D. McFadden. Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245, 1994.

[79] C. J. Oates, M. Girolami, and N. Chopin. Control functionals for Monte Carlo integration. *Journal of the Royal Statistical Society: Series B*, 79(3):695–718, 2017.

[80] K. Parthasarathey. *Probability Measures on Metric Spaces*. Academic Press, New York, 1967.

[81] C. Pigorsch and R. Stelzer. On the definition, stationary distribution and second order structure of positive semi-definite Ornstein-Uhlenbeck type processes. *Bernoulli*, 15:754–773, 2009.

[82] C. Pigorsch and R. Stelzer. A multivariate Ornstein-Uhlenbeck type stochastic volatility model. *Technical Report: www.uni-ulm.de/en/mawi/finmath/people/stelzer/publications.html*, 2009.

[83] V. Pipiras and M. S. Taqqu. *Long-Range Dependence and Self-Similarity*. Cambridge University Press, Cambridge, 2017.

[84] D. Pollard. *Empirical Processes: Theory and Applications*, volume 2. Institute of Mathematical Statistics, Hayward, CA, 1990.

[85] F. Portier and J. Segers. Monte Carlo integration with a growing number of control variates. *arXiv:1801.01797v3*, 2018.

[86] P. Protter. *Stochastic Integration and Differential Equations*. Springer, New York, 2 edition, 2005.

[87] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015. URL `https://www.R-project.org/`.

[88] A. Raknerud and Ø. Skare. Indirect inference methods for stochastic volatility models based on non-Gaussian Ornstein-Uhlenbeck processes. *Computational Statistics & Data Analysis*, 56(11):3260–3275, 2012.

[89] C. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, New York, 2004.

[90] K. Sato. *Lévy processes and Infinitely Divisible Distributions.* Cambridge university press, Cambridge, 1999.

[91] R. L. Schilling and L. Partzsch. *Brownian Motion: An Introduction to Stochastic Processes.* De Gruyter, Berlin/Boston, 2014.

[92] E. Schlemm and R. Stelzer. Quasi maximum likelihood estimation for strongly mixing state space models and multivariate Lévy-driven CARMA processes. *Electronic Journal of Statistics*, 6:2185–2234, 2012.

[93] P. Shaman and R. A. Stine. The bias of autoregressive coefficient estimators. *Journal of the American Statistical Association*, 83(403):842–848, 1988.

[94] R. H. Shumway and D. S. Stoffer. *Time Series Analysis and Its Applications.* Springer, New York, 3 edition, 2011.

[95] A. A. Smith. Estimating nonlinear time series models using simulated vector autoregressions. *Journal of Applied Econometrics*, 8:S63–S84, 1993.

[96] A. A. Smith. Estimating nonlinear time series models using simulated vector autoregressions. *Journal of Applied Econometrics*, 8(S1):S63–S84, 1993.

[97] M. Sørensen. Prediction-based estimating functions. *Econometrics Journal*, 3:123–147, 2000.

[98] R. Stelzer. Multivariate COGARCH(1,1) processes. *Bernoulli*, 16(1):80–115, 2010.

[99] R. Stelzer and J. Vestweber. Geometric ergodicity of the multivariate COGARCH(1,1) process. *arXiv:1701.07859*, 2019.

[100] D. Straumann and T. Mikosch. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *Annals of Statistics*, 34(5):2449–2495, 2006.

[101] B. Wahlberg, J. Welsh, and L. Ljung. Identification of stochastic Wiener systems using Indirect Inference. *IFAC-PapersOnLine*, 48(28):620–625, 2015.

[102] H. White. Maximum likelihood estimation of misspecified models. *Econometrica*, 50:1–26, 1982.

[103] R. M. Wilcox. Exponential operators and parameter differentiation in quantum physics. *Journal of Mathematical Physics*, 8(4):962–982, 1967.

[104] J. Yu. *Empirical characteristic function in time series estimation and a test statistic in financial modelling.* PhD thesis, Faculty of Graduate Studies, University of Western Ontario, 1998.

[105] J. Yu. Empirical characteristic function estimation and its applications. *Econometric Reviews*, 23(2):93–123, 2004.

[106] S. L. Zeger. A regression model for time series of counts. *Biometrika*, 75(4):621–629, 1988.